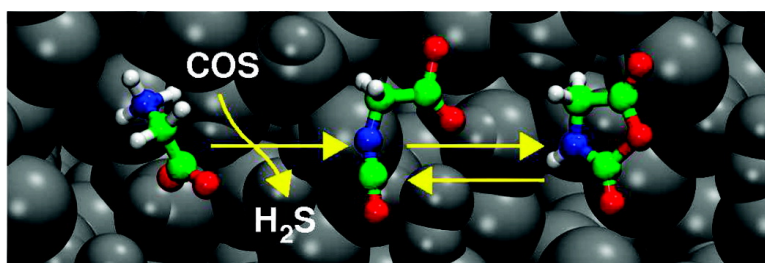**Article**

# Peptide Synthesis in Aqueous Environments: The Role of Extreme Conditions on Amino Acid Activation

Nisanth N. Nair, Eduard Schreiner, and Dominik Marx

## More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

# Peptide Synthesis in Aqueous Environments: The Role of Extreme Conditions on Amino Acid Activation

Nisanth N. Nair,* Eduard Schreiner,† and Dominik Marx

*Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, 44780 Bochum, Germany*

Received April 1, 2008; E-mail: nisanth.nair@theochem.rub.de

***Abstract:*** The free energy surfaces and reaction mechanisms underlying the activation of amino acids by COS in bulk water at ambient conditions as well as extreme temperature−pressure thermodynamic conditions were studied using accelerated ab initio molecular dynamics. The results for the reaction sequence leading from glycine to its activated form, a so-called Leuchs anhydride or α-amino acid *N*-carboxyanhydride (NCA), suggest that extreme conditions not far from the critical point of water may favor the formation of this activated species. This is traced back to appropriately affecting relative stabilities of neutral versus charged or zwitterionic molecular species which shifts equilibria, affects relative barriers, and thus modifies reaction rates. Furthermore, it is shown that the *N*-carboxyanhydride of glycine is not formed from *N*-thiocarboxyl glycine by its direct cyclization, but instead an indirect mechanism, the so-called isocyanate route, is clearly preferred at both conditions. The work quantitatively underpins the impact of extreme solvent conditions on the investigated organic reactions in aqueous media which implies that the presented results are of relevance to fields such as prebiotic chemistry and green chemistry.

## 1. Introduction

Chemical synthesis of peptide bonds[1] at ambient aqueous conditions by direct condensation, i.e., involving carboxyl and amino groups stemming from amino acids and/or peptides, is typically a very slow process in the absence of suitable catalysts.[2] There are two major points to consider connected to such a direct condensation approach. In the first place, there is the equilibrium between the charged and neutral forms of the reacting amino acids or peptides. As is well-known, it is the neutral form which is required for such condensation reactions, whereas amino acids and peptides prefer the charged zwitterionic form at neutral pH and aqueous conditions.[2] The standard approach to address this issue is to shift the dissociation equilibrium toward the neutral form by suitably buffering the pH of the aqueous solution to an appropriate value. The second issue connected to condensation is the hydrolysis of the formed peptides;[3] in particular when the destruction of peptide bonds is faster than their formation, there will be no net production of peptides. One of the generally used procedures to accelerate the production of peptides over hydrolysis is to convert the rather inert amino acid to an another, activated form which is more reactive than the amino acid itself. The vast family of α-amino

acid *N*-carboxyanhydrides (NCAs)[4−6] are the classic example for such activated forms of amino acids used in peptide synthesis.

Since their discovery in 1906 by Leuchs,[7] NCAs also known as "Leuchs anhydrides" proved to be useful for the stepwise synthesis of block copolymers and oligopeptides.[8] Excellent authoritative overviews of the NCA chemistry are provided in refs 5 and 6 and in an early book.[4] Chemically, NCAs offer the advantage of possessing an activated CO group (C5) while simultaneously protecting the amino group. In addition to employing NCAs in a wealth of preparative organic synthesis strategies in the laboratory, it has been suggested that NCAs might have played a crucial role in putative prebiotic chemistry and molecular evolution as discussed in section ten of Kricheldorf's review[5] and, in much broader scope and detail, in a specialized review.[9] In particular, amino acids could be activated by $CO_2/HCO_3^-$ thus leading to NCAs which "would provide an ideal prebiotic, activated amino acid".[10] Several other investigations showed that NCAs can also be formed in reactions of CO-activated amino acids, such as esters,[11,12] with $CO_2/HCO_3^-$. Apart from $CO_2$ other simple agents like $NCO^-$, $NO/O_2$ [13,14] and COS (carbonyl sulfide) both, at ambient

† Present address: Theoretical and Computational Biophysics Group, Beckman Institute, Urbana, IL 61801.

(1) Sewald, N.; Jakubke, H.-D. *Peptides: Chemistry and Biology*; Wiley-VCH Verlag GmbH: Weinheim, 2002; pp 135−268.

(2) Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry*, 5th ed.; W. H. Freeman and Company: New York, 2002; pp 43−53.

(3) Radzicka, A.; Wolfenden, R. *J. Am. Chem. Soc.* **1996**, *118*, 6105–6109.

(4) Kricheldorf, H. R. α-*Aminoacid-N -Carboxy-Anhydrides and Related Hetrocycles*; Springer-Verlag: Berlin Heidelberg, 1987; pp 3−157.

(5) Kricheldorf, H. R. *Angew. Chem., Int. Ed.* **2006**, *45*, 5752−5784.

(6) Deming, T. J. *Adv. Polym. Sci.* **2006**, *202*, 1–18.

(7) Leuchs, H. *Ber. Dtsch. Chem. Ges.* **1906**, *39*, 857–861.

(8) Denkewalter, R. G.; Schwam, H.; Strachan, R. G.; Beesley, T. E.; Veber, D. F.; Schoenewaldt, E. F.; Barkemeyer, H., Jr.; Jacob, T. A.; Hirschmann, R. *J. Am. Chem. Soc.* **1966**, *88*, 3163–3164.

(9) Pascal, R.; Boiteau, L.; Commeyras, A. *Top. Curr. Chem.* **2005**, *259*, 69–122.

(10) Ehler, K. W.; Orgel, L. *Biochim. Biophys. Acta* **1976**, *434*, 233–243.

(11) Korshak, V. V.; Poroshin, K. T.; Kozarenko, T. D. *Izv. Akad. Nauk SSSR, Otdel Khim. Nauk* **1954**, 663–669.

(12) Wieland, T.; Jaenicke, F. *Justus Liebigs Ann. Chem.* **1956**, *599*, 125–130.

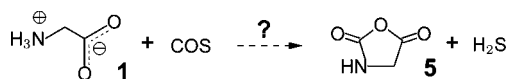(13) Collet, H.; Bied, C.; Mion, L.; Taillades, J.; Commeyras, A. *Tetrahedron Lett.* **1996**, *37*, 9043–9046.

**Figure 1.** Investigated activation mechanism of glycine **1** by COS to form the corresponding α-amino acid *N*-carboxyanhydride (NCA) **5** at aqueous reaction conditions.

aqueous conditions[15] and extreme hydrothermal conditions,[16] were identified to provide synthetic pathways to these activated amino acids. It was recently demonstrated that aminoacyl–phosphate anhydrides can be synthesized from the reaction between phosphate and NCA formed in situ by the activation of amino acids with COS in the presence of the oxidizing agent ferricyanide.[17] This extends the relevance of NCAs within the realm of putative prebiotic chemistry to aminoacylation of ribonucleotides, thus addressing the coupling of the protein and the RNA worlds and the emergence of translation.[18] Apart from that, there is a multitude of reasons to consider NCAs as an important activated form of amino acids relevant for prebiotic chemistry in aqueous media as argued in an excellent review.[9]

Mechanistic insights into the formation of NCAs from amino acids and COS are limited to speculations in footnote 16 of ref 19 and in ref 15 based on the observation of ureido acids as side products. Thus, despite impressive experimental progress,[15,16] the key step of the formation of NCAs with the participation of COS in aqueous environment as depicted in Figure 1 is still hypothetical and needs verification according to the discussion of Scheme 39 in ref 5.

Interestingly, in the Fuchs–Farthing method, which is widely used for the preparation of NCAs,[4] the treatment of amino acids with excess phosgene and HCl leads to a rearrangement of the NCA to an isocyanate acid chloride as revealed by infrared spectroscopic measurements.[20] Such an intimate relation between NCAs and the corresponding isocyanate derivatives is indeed observed in a variety of reactions, including the thermal polymerization of NCAs.[4] Additionally, it was found that strong bases with low nucleophilicity tend to favor the conversion of NCAs to isocyanates, which is supported by the observation of urea side products and the appearance of isocyanate specific infrared peaks at 2230 cm$^{-1}$.[21] The mechanism is thought to start with an *N*-deprotonation of the NCA to its anion, which is in equilibrium with the isocyanate. Discussions on an isocyanate route to the formation of NCAs can also be found in earlier work[10,22] on carbonyldiimidazole activation of aminoacids and peptide formation in aqueous conditions.

Usually NCAs are prepared and used in organic solvents[4] whereas the behavior of NCA-activated peptide synthesis in

water is expected to be rather different. Unlike in organic solvents with high nucleophilicity and donor ability such as DMF (*N,N*-dimethylformamide), NMP (*N*-methylpyrrolidone), DMSO (dimethyl sulfoxide),[23] or pyridine[24,25] polymerization of NCAs in ambient water yields linear polypeptides.[5] What remains fully unexplored, however, are the consequences of extreme temperature and pressure conditions on such reactions, which might open up avenues for more efficient amino acid activation and peptidization in aqueous reaction environments. This is despite the fact that hot-pressurized water obtained much attention recently as it reacts differently with organic molecules compared to ambient water and noncatalytic chemical reactions can be induced efficiently under such extreme thermodynamic conditions.[26−29] As a result, water and hot-pressurized water in particular moved into focus of synthetic organic chemistry as an alternative reaction medium "for green sustainable organic synthesis".[30] In this context, it must be appreciated that the dielectric constant of water at its critical point (where $T_c = 647.1$ K and $p_c = 22.1$ MPa) is dramatically reduced to only $\varepsilon \approx 6$ instead of about 80 at ambient conditions.[28] In general it is well appreciated that due to the much lower dielectric constant (zwitter)ionic species are destabilized wrt their neutral counterparts and that nonpolar compounds become better soluble in water with raising temperature and pressure,[27] which brings about advantages for many synthetic applications.[30,31] At the same time, the ion product of water is raised and the viscosity is lowered by several orders of magnitude at criticality.[28] Thus, water at such extreme conditions behaves indeed like a different substance, which implies that well-known concepts from organic chemistry might rather be misleading than helpful here.[32] At this point, it is worth pointing out that the extreme temperature and pressure conditions found in hydrothermal vent environments[33−36] (typically $T \approx 500$ K and $p \approx 20$ MPa with $\varepsilon \approx 32$, see ref 37) are not very far from critical conditions. From an experimental perspective, excellent overviews on many classes of organic reactions in sub- and supercritical water can be found in refs 27, 29, 38, and 39. Such investigations are not only

(14) Commeyras, A.; Collet, H.; Boiteau, L.; Taillades, J.; Vandenabeele-Trambouze, O.; Cottet, H.; Biron, J.-P.; Plasson, R.; Mion, L.; Lagrille, O.; Martin, H.; Selsis, F.; Dobrijevic, M. *Polym. Int.* **2002**, *51*, 661–665.
(15) Leman, L.; Orgel, L.; Ghadiri, M. R. *Science* **2004**, *306*, 283–286.
(16) Huber, C.; Wächtershäuser, G. *Science* **1998**, *281*, 670–672.
(17) Leman, L. J.; Orgel, L. E.; Ghadiri, M. R. *J. Am. Chem. Soc.* **2006**, *128*, 20–21.
(18) Biron, J.-P.; Parkes, A. L.; Pascal, R.; Sutherland, J. D. *Angew. Chem., Int. Ed.* **2005**, *44*, 6731–6734.
(19) Deweye, R. S.; Schoenewaldt, E. F.; Joshua, H., Jr.; Schwam, H.; Barkemeyer, H.; Arison, B. H.; Veber, D. F.; Strachan, R. G.; Milkowski, J.; Denkewaltear, R. G.; Hirschmann, R. *J. Org. Chem.* **1971**, *36*, 49–59.
(20) Iwakura, Y.; Uno, K.; Kang, S. *J. Org. Chem.* **1965**, *30*, 1158–1161.
(21) Kopple, K. D. *J. Am. Chem. Soc.* **1957**, *79*, 6442–6446.
(22) Ehler, K. W. *J. Org. Chem.* **1976**, *41*, 3041–3042.
(23) Kricheldorf, H. R.; von Lossow, C.; Schwarz, G. *Macromolecules* **2005**, *38*, 5513–5518.

(24) Bilek, L.; Derkosch, J.; Michl, H.; Wessely, F. *Monatsh. Chem.* **1953**, *84*, 717–740.
(25) Kricheldorf, H. R.; von Lossow, C.; Schwarz, G. *J. Polym. Sci. Part A: Polym. Chem* **2006**, *44*, 4680–4695.
(26) Nagai, Y.; Morooka, S.; Matubayasi, N.; Nakahara, M. *J. Phys. Chem. A* **2004**, *108*, 11635–11643.
(27) Akiya, N.; Savage, P. E. *Chem. Rev.* **2002**, *102*, 2725–2750.
(28) Weingärtner, H.; Frank, E. U. *Angew. Chem., Int. Ed.* **2005**, *44*, 2672–2692.
(29) Li, C.-J. *Chem. Rev.* **2005**, *105*, 3095–3165.
(30) Sheldon, R. A. *Green Chem.* **2005**, *7*, 267–278.
(31) Hailes, H. C. *Org. Process Res. Dev.* **2007**, *11*, 114–120.
(32) Hazen, R. M.; Boctor, N.; Brandes, J. A.; Cody, G. D.; Hemley, R. J.; Sharma, A.; Yoder, H. S., Jr. *J. Phys.: Condens. Matter* **2002**, *14*, 11489–11494.
(33) Amend, J. P.; Shock, E. L. *Science* **1998**, *281*, 1659–1662.
(34) Imai, E.; Honda, H.; Hatori, K.; Brack, A.; Matsuno, K. *Science* **1999**, *283*, 831–833.
(35) Singh, S. C.; Crawford, W. C.; Carton, H.; Seher, T.; Combier, V.; Cannat, M.; Canales, J. P.; Düsünür, D.; Escartin, J.; Miranda, J. M. *Nature* **2006**, *442*, 1029–1032.
(36) Wächtershäuser, G. *Chem. Biodivers.* **2007**, *4*, 584–602.
(37) Uematsu, M.; Franck, E. U. *J. Phys. Chem. Ref. Data* **1980**, *9*, 1291–1306.
(38) Simoneit, B. R. T. *Origins Life Evol. Biospheres* **1995**, *25*, 119–140.
(39) Katritzky, A. R.; Nichols, D. A.; Siskin, M.; Murugan, R.; Balasubramanian, M. *Chem. Rev.* **2001**, *101*, 837–892.
(40) Plankensteiner, K.; Reiner, H.; Rode, B. M. *Curr. Org. Chem.* **2005**, *9*, 1107–1114.
(41) Chen, Q. W.; Chen, C. L. *Curr. Org. Chem.* **2005**, *9*, 989–998.
(42) Leach, S.; Smith, I. W. M.; Cockell, C. S. *Phil. Trans. R. Soc. B* **2006**, *361*, 1675–1679.

interesting in the framework of organic chemistry at extreme aqueous conditions, rather they also contribute to ideas revolving around putative prebiotic chemistry.[36] However, reviewing even part of the ideas in this fascinating area of research is far beyond the scope of the present work so that we refer the interested reader to a broad selection of reviews[9,40−46] and books.[47−49] But despite these interesting prospects, the present authors are not aware of any theoretical studies of amino acid activation via NCAs or peptide synthesis in hot-pressurized water with the exception of our own preliminary efforts.[50]

The present study focuses on the general question of *activation of amino acids in aqueous environment via formation of NCAs* using glycine as the simplest example as sketched in Figure 1. As will be shown, studying the formation of NCAs involves a sequence of five chemical reactions to be investigated in condensed phase environments. We note in passing that we address exclusively the direct activation of aminoacids by COS, although alternative ways involving for instance transition metals, alkylating, or oxidizing agents are known. At this stage, we restrict ourselves to the study of the simpler direct activation, while aspects of the more complex cases, including the effects of mineral surfaces, will be the subject of our future studies. In order to understand the crucial influence of extreme thermodynamic conditions on the free energy barriers and, possibly, on the reaction mechanism, all these chemical reactions have been studied separately in ambient water and in hot-pressurized water (at $T \approx 500$ K and $p \approx 20$ MPa) while keeping all other parameters identical. In order to explore the complex free energy landscapes all simulations were performed in the framework of accelerated ab initio molecular dynamics sampling as outlined in the next section.

## 2. Methods and Models

Ab initio molecular dynamics[51] (which is also called, e.g. "first principles", "on-the-fly", "quantum", or "direct" molecular dynamics)[51] is an efficient class of simulation techniques in order to study chemical reactions in aqueous solution, in particular in those cases where solvent molecules can be involved as reactive species, e.g. via protolysis. However, standard ab initio molecular dynamics cannot overcome reaction (free) energy barriers of the order of many $k_{\mathrm{B}}T$ on the currently accessible picosecond time scale for sufficiently large systems. It is thus mandatory to employ accelerated sampling techniques in order to surmount these barriers and to map free energy pathways using explicit and reactive solvent. Among the

various methods used to study "rare events" such as thermodynamic integration, umbrella sampling, and transition path sampling to name but a few (see e.g. various contributions collected in ref 52), the metadynamics sampling technique[53] is particularly well-suited to be combined with ab initio molecular dynamics[54] as reviewed in refs 55 and 56. In particular, it allows one to explore efficiently free energy landscapes by driving several generalized coordinates simultaneously, thus sampling multidimensional free energy surfaces without specifying a priori the products of the studied reactions, which is mandatory in the present investigation.

All calculations were performed using the Car−Parrinello approach to ab initio molecular dynamics[57] together with Nosé−Hoover chain[58] thermostats for both nuclei and electronic orbitals within spin-restricted Kohn−Sham density functional theory in its plane wave/pseudopotential formulation using the CPMD program package.[51,59] The PBE[60,61] exchange-correlation functional was chosen and the core electrons were taken into account using Vanderbilt's ultrasoft pseudopotentials;[62] in the case of sulfur, the pseudopotentials contained additional *d*-projectors.[63,64] Previous tests and experience[63,64] showed that a plane wave cutoff of 25 Ry is sufficient for the systems studied if these ultrasoft pseudopotentials are used. A time step of 0.145 fs was used for the integration of the equations of motion and the fictitious mass for the orbitals was set to 700 au. No mass corrections are applied for orbital inertia effects that effectively renormalize the nuclear masses (see e.g. section 2.4.3 in ref 51 for a discussion of this drag or dressing effect and ref 65 for an early application of a simple correction). All hydrogen masses in the system were substituted by deuterium masses in order to allow for a larger fictitious orbital mass and thus a larger time step resulting into more efficient molecular dynamics propagation.[51]

In the present work, the free energy landscapes and the reaction mechanisms of all individual steps leading from glycine to NCA, i.e., the reaction sequence finally presented in Figure 9 as a graphical synopsis (cf. Figure 1), are investigated in aqueous bulk environment. As it will turn out, these are five independent steps involving a direct and an indirect synthesis route. In addition, independent simulations are carried out at two different thermodynamic conditions, ambient (ABW) and hot-pressurized (HPW) bulk water, in order to probe the effect of extreme reaction conditions, which amounts to a total of 10 distinct ab initio simulations. Condensed phase conditions were imposed in these calculations by the periodicity of the simulation supercell at a defined temperature and particle number density, i.e., canonical ensemble conditions. For the cases of ABW and HPW conditions, periodic supercells used are orthorhombic with fixed dimensions of $10.8 \times 10.8 \times 9.5$ and

(43) Herdewijn, P.; Kisakürek, M. V. *Chem. Biodivers.* **2007**, *4*, 539–540.
(44) Martin, W.; Russell, M. J. *Phil. Trans. R. Soc. B* **2007**, *362*, 1887–1925.
(45) Eschenmoser, A. *Tetrahedron* **2007**, *63*, 12821–12844.
(46) Milner-White, E. J.; Russell, M. J. *Biol. Direct* **2008**, *3*.
(47) Bernal, J. D. *The Physical Basis of Life*; Routledge and Kegan Paul: London, 1951; p 6–80.
(48) Deamer, D. W.; Fleischaker, G. R. *Origins of Life: The Central Concepts*; Jones and Bartlett Publishers, Inc.: Boston, 1994; pp 133−223.
(49) Hazen, R. M. *Genesis: The Scientific Quest for Life's Origin*; Joseph Henry Press: WA, 2005; pp 191−243.
(50) Schreiner, E.; Nair, N. N.; Marx, D. *J. Am. Chem. Soc.* **2008**, *130*, 2768–2770.
(51) Marx, D.; Hutter, J., Ab initio molecular dynamics: Theory and Implementation. In *Modern Methods and Algorithms of Quantum Chemistry*; Grotendorst, J., Ed.; John von Neumann Institute for Computing (NIC), Forschungszentrum Jülich: Germany, 2000; Vol. 3, pp 301−449.
(52) Ciccotti G.; Binder K.; Ferrario M., Introduction: Condensed Matter Theory by Computer Simulation. In *Computer Simulations In Condensed Matter: From Materials to Chemical Biology*; Ferrario, M., Ciccotti, G., Binder, K., Eds.; Springer Verlag: Berlin, Heidelberg, 2006; Vol. 1, pp 1−11.

(53) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci.* **2002**, *99*, 12562–12566.
(54) Iannuzzi, M.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2003**, *90*, 238302-1–238302-4.
(55) Ensing, B.; Vivo, M. D.; Liu, Z.; Moore, P.; Klein, M. L. *Acc. Chem. Res.* **2006**, *39*, 73–81.
(56) Laio, A.; Parrinello, M. Computing Free Energies and Accelerating Rare Events with Metadynamics. In *Computer Simulations in Condensed Matter: From Materials to Chemical Biology*; Ferrario, M., Ciccotti, G., Binder, K., Eds.; Springer Verlag: Berlin, Heidelberg, 2006; Vol. 1, pp 315−347.
(57) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471–2474.
(58) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635–2643.
(59) *CPMD*; IBM Corp and MPI für Festkörperforschung Stuttgart; 1990; see also http://www.cpmd.org.
(60) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
(61) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1997**, *78*, 1396–1396.
(62) Vanderbilt, D. *Phys. Rev. B* **1990**, *41*, 7892–7895.
(63) Pollet, R.; Boehme, C.; Marx, D. *Origins Life Evol. Biospheres* **2006**, *36*, 363–379.
(64) Nair, N. N.; Schreiner, E.; Marx, D. *J. Am. Chem. Soc.* **2006**, *128*, 13815–13826.
(65) Marx, D.; Hutter, J.; Parrinello, M. *Chem. Phys. Lett.* **1995**, *241*, 457–462.

$10.8 \times 10.8 \times 10.8$ Å$^3$, respectively, hosting 36 water molecules in each case. For the ABW setup, the density of water is $\approx 1.00$ g/cm$^3$ and $p \approx 0.1$ MPa and the whole system is thermostatted to 300 K. At HPW conditions, the density of water is $\approx 0.85$ g/cm$^3$, which corresponds to $p \approx 20$ MPa according to the experimental equation of state[66] at the thermostatted temperature of 500 K. Concerning the equation-of-state (or density), we note that H versus D quantum effects are neglected consistently throughout this study being based on simulations relying on using classical nuclei within the Born−Oppenheimer approximation. Classical nuclear dynamics in the electronic ground-state is expected to be sufficient for the present purpose, although there are several ab initio molecular dynamics techniques available to transcend these approximations as explained in ref 51. Furthermore, temperatures are likely to be somewhat underestimated and dynamics seems to be slower when compared to experiment, although it is not clear if this is mainly due to inaccuracies of the used density functionals, or if this is in part because of neglecting quantum effects on nuclear motion, not working in the complete basis set limit, or other limitations.[67−73] In view of this situation, neither a priori nor a posteriori empirical corrections are applied to temperatures.

The initial configurations for each simulation were obtained either by using the final structure of the previous reaction within the series of runs or, alternatively, by the following equilibration protocol. In a pre-equilibrated water system, one water molecule was replaced by the solute and equilibrated at the target temperature for 1 ps while the atomic positions of the solute were kept fixed. Subsequently, these constraints were taken off and all degrees of freedom were thermostatted individually according to the "massive thermostatting" protocol for typically 1−3 ps where each Cartesian degree of freedom of each nucleus is thermostatted separately. This ensures energy equipartition in the total system even in the presence of stiff vibrational modes prior to the metadynamics simulations in which one thermostat for all nuclei was used as usual. Depending on the conditions, i.e., ABW or HPW, and the particular reaction studied, metadynamics trajectories of at least 150 ps and up to about 200 ps were collected. In total, about two nanoseconds of ab initio metadynamics trajectories were produced to provide the results presented in this paper.

The basic idea behind the metadynamics technique[53] is a reduction of the dimensionality of the underlying reaction space by an appropriate a priori selection of a suitable set of collective coordinates, $\{S_\alpha(\mathbf{R})\}$, which discriminate reactant and product states and describe the slowest modes in the process of interest. The $S_\alpha(\mathbf{R})$ can be understood as generalized coordinates being functions of the Cartesian coordinates $\mathbf{R}$ of the nuclei. We used three different types of collective coordinates in our study. The simplest one is just the distance

$$d[A-B] = |\mathbf{R}_A - \mathbf{R}_B| \qquad (1)$$

between two specified atoms A and B. Another type was the coordination number[74] $c[A-B]$ between an atom A with respect to a set of other atoms B consisting of the same species, which is defined as

$$c[A-B] = \sum_{I \in B} \frac{1 - (R_{AI}/R_{AB}^0)^6}{1 - (R_{AI}/R_{AB}^0)^{12}} \qquad (2)$$

Here $R_{AI}$ is the distance between atom A and atom $I$ of the set of B atoms and $R_{AB}^0$ 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, the function $c[A-B]$ is nearly unity when the actual bond distance $R_{AI} < R_{AB}^0$ and approaches zero rapidly when $R_{AI} > R_{AB}^0$. The used cutoff distances $R_{AB}^0$ depend on the atomic species A and B and we chose $R_{XH}^0 = 1.4$ Å for X = N, O, and S, while $R_{CO}^0 = 1.8$ Å. The third type of collective coordinate is a more general coordination number, $c^{tot}[A-B]$, which measures the total coordination between two sets of atoms A and B, defined as

$$c^{tot}[A-B] = \sum_{J \in A} \sum_{I \in B} \frac{1 - (R_{IJ}/R_{AB}^0)^6}{1 - (R_{IJ}/R_{AB}^0)^{12}} \qquad (3)$$

The specific sets of A and B atoms depend on the reaction studied and will be defined when presenting the results.

In the extended Lagrangian formulation of metadynamics[54] used here, a set of auxiliary (fictitious) particles with positions $\{s_\alpha\}$, velocities $\{\dot{s}_\alpha\}$, and associated masses $\{M_\alpha\}$ equal in number to the selected collective coordinates $N_s$ extends the physical degrees of freedom. In addition, their fictitious motion is coupled to that of the collective coordinates, $\{S_\alpha\}$, by a harmonic potential, thus restraining the dynamics of the latter. Starting from the Car−Parrinello Lagrangian,[51,57] $L_{CP}$, the metadynamics Lagrangian reads

$$L^{MTD} = L_{CP} + \frac{1}{2}\sum_\alpha^{N_s} M_\alpha \dot{s}_\alpha^2 - \frac{1}{2}\sum_\alpha^{N_s} k_\alpha [S_\alpha(\mathbf{R}) - s_\alpha]^2 - V(t, \mathbf{s}) \qquad (4)$$

where $k_\alpha$ is the force constant of the corresponding harmonic potential involving the auxiliary particle $s_\alpha$. The last term, $V(t, \mathbf{s})$, represents a repulsive time-dependent potential, which is slowly built up as the dynamics proceeds in order to escape from free energy minima in the subspace spanned by the selected set of collective coordinates $S_\alpha$ and to eventually map the underlying free energy surface; see refs 55 and 56 for general reviews. In the present work, $V(t, \mathbf{s})$ was chosen to be a multivariate Gaussian potential with height $H(t_i)$ and width $\delta s(t_i)$,

$$V(t, \mathbf{s}) = \sum_{t_i < t} H(t_i) \exp\left\{ -\frac{[\mathbf{s}(t) - \mathbf{s}(t_i)]^2}{2[\mathbf{w}(t_i)\delta s(t_i)]^2} \right\} \qquad (5)$$

which is dropped at times $t_0, t_1, ..., t_i$ thus defining the metadynamics time step $\Delta t^{MTD}$.

The height $H$ of the Gaussian bias potential was manually adapted according to the depth of the free energy wells. For efficient filling of free energy basins which are nonspherical in their topology, anisotropic scaling factors $\mathbf{w} = \{w_\alpha\}$ were multiplied with the width of the Gaussians $\delta s(t_i)$ according to eq 5. Depending on the specific reaction and on the respective fluctuations the height was varied between 0.6 and 4.0 $k_BT$. The dimensionless parameter $\delta s(t_i)$ was set to 0.05 and the individual scaling factors $w_\alpha$ were adjusted in such a way that

$$w_\alpha = \frac{1}{4}\max(|S_\alpha - \langle S_\alpha \rangle|) \ \forall \ \alpha \qquad (6)$$

(66) 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 cm³/g for 200 bar pressure and 225 °C temperature in Table 5 (page 62) is corrected to 1.1778 cm³/g.

(67) Fernandez-Serra, M. V.; Artacho, E. *J. Chem. Phys.* **2004**, *121*, 11136–11144.

(68) Kuo, I. F. W.; Mundy, C. J.; McGrath, M. J.; Siepmann, J. I.; VandeVondele, J.; Sprik, M.; Hutter, J.; Chen, B.; Klein, M. L.; Mohamed, F.; Krack, M.; Parrinello, M. *J. Phys. Chem. B* **2004**, *108*, 12990–12998.

(69) Fernandez-Serra, M. V.; Ferlat, G.; Artacho, E. *Mol. Simul.* **2005**, *31*, 361–366.

(70) Sit, P. H. L.; Marzari, N. *J. Chem. Phys.* **2005**, *122*, 204510.

(71) McGrath, M. J.; Siepmann, J. I.; Kuo, I. F. W.; Mundy, C. J.; VandeVondele, J.; Hutter, J.; Mohamed, F.; Krack, M. *ChemPhysChem* **2005**, *6*, 1894–1901.

(72) Mantz, Y. A.; Chen, B.; Martyna, G. J. *J. Phys. Chem. B* **2006**, *110*, 3540–3554.

(73) Lee, H. S.; Tuckerman, M. E. *J. Chem. Phys.* **2007**, *126*, 164501.

(74) Sprik, M. *Chem. Phys.* **2000**, *258*, 139–150.

(75) Ensing, B.; Laio, A.; Parrinello, M.; Klein, M. L. *J. Phys. Chem. B* **2005**, *109*, 6676–6687.
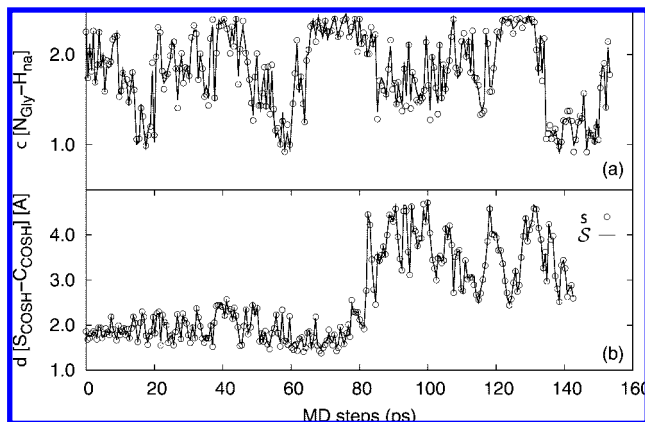
**Figure 2.** Collective coordinate $\mathcal{S}$ (solid lines) and corresponding auxiliary variable $\mathcal{S}$ (open circles) during the metadynamics simulations of (a) the reaction of COS and glycine to $N$-thiocarboxyl glycine in HPW and (b) the ring closure of $N$-thiocarboxyl glycine to yield Leuchs anhydride in HPW. In part a, the coordination number $c[N_{Gly}-H_{na}]$ of the glycine nitrogen to all nonaliphatic hydrogen atoms in the system, and in part b, the distance $d[S_{COSH}-C_{COSH}]$ between the sulfur and carbon in the COSH part of the $N$-thiocarboxyl glycine is shown; see text for further details.

is satisfied, which is estimated during the preparation runs without adding any Gaussians.

The coupling constants $k_\alpha$ and masses $M_\alpha$ depend on the type of the collective coordinate used and were chosen to be 0.4 au and 50.0 amu, respectively, for distances and 2.0 au and 50.0 amu, respectively, for coordination numbers. This choice allows the collective coordinate $\mathcal{S}_\alpha(\mathbf{R})$ and the associated auxiliary variable $s_\alpha$ to move close to each other such that the dynamics of the latter maintains adiabatic separation from the fictitious electronic degrees of freedom. Figure 2 depicts the dynamics of the collective coordinates and associated auxiliary variables for the distance $d(S_{COSH}-C_{COSH})$ between the sulfur and the carbon in the COS moiety of the $N$-thiocarboxyl glycine and for the coordination number $c(N_{Gly}-H_{na})$ of the glycine nitrogen to all nonaliphatic hydrogen atoms. These two representative trajectories were monitored during the cyclization of $N$-thiocarboxyl glycine ("thiocarbamate") to NCA ("Leuchs anhydride") and during the reaction of COS and glycine to $N$-thiocarboxyl glycine, respectively, as explained later in the results sections. It is clear that for the parameters chosen as described above the collective coordinates and the auxiliary variables are moving close to each other as required for proper metadynamics sampling.

The metadynamics time step $\Delta t^{MTD}$ 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

$$|\mathbf{s}(t) - \mathbf{s}(t_i)| = 3/2\delta s \qquad (7)$$

which is necessary to avoid so-called "hill-surfing" problems as discussed in ref 75. Note that steepness in the rough biasing potential may result in heating up the system, because the auxiliary variables pull on the nuclear 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 as defined by the motion of the nuclei. Simple velocity scaling is used to restrict the fluctuations of the fictitious kinetic energy of the auxiliary particles corresponding to a temperature window $\pm 200$ K with respect to the target temperature, while the nuclear degrees of freedom are controlled by a Nosé–Hoover chain thermostat[58] in order to properly establish the canonical ensemble.

Having the accumulated information on the positions, widths, and heights of the Gaussians deposited during such a metadynamics simulation, reconstruction of the free energy landscape underlying
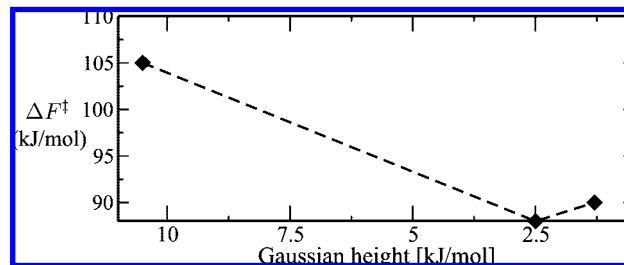


**Figure 3.** Convergence of free energy barriers with respect to the height of the Gaussians used for metadynamics sampling of multidimensional free energy surfaces.[64]

the dynamics is technically straightforward. The Helmholtz (or *NVT*) free energy $F(\mathbf{s})$ can be obtained[76] from

$$F(\mathbf{s}) = -\lim_{t \to \infty} V(t, \mathbf{s}) + \text{constant} \qquad (8)$$

so that relative free energy surfaces and free energy barriers can be extracted readily from metadynamics.[53−55]

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 deposited Gaussians. Our previous experience[64] 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 $\approx 2-4k_BT$ 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 $\leq 1k_BT$. The error of the free energy estimates obtained by this procedure has been shown to amount to about $1k_BT$. The convergence of this protocol to computing free energy barriers is depicted in Figure 3 with respect to the Gaussian height for one of the systems studied.[64] Since smaller biasing potentials are used during the refinement, the system has more time to relax, thus mapping the underlying potential energy surface more accurately, 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 observed mechanism during this procedure, to detect if any crucial slow modes are missing in the selected set of collective coordinates. This allows one to see if the subspace spanned by these coordinates actually accommodates the true reaction coordinate.

In order to restrict the sampling in relevant regions of the space spanned by the collective coordinates, additional repulsive potentials (so-called "wall potentials") were appropriately applied on the auxiliary variables. As described in ref 75, this can introduce artifical deepening of the free energy surface near the positions where such repulsive potentials are applied. Therefore, these regions are carefully avoided for the computation of free energy barriers. Still, in most of the reactions we have studied there are some very slow and essentially unbound diffusive modes which require long simulation times in order to fill the associated spacious free energy basins fully. A representative example for this is the direct cyclization of $N$-thiocarboxyl glycine to NCA; see the depth of the corresponding free energy minimum **3** in Figure 6b and the volume spanned by the coordinate $d(C_{COS}-O_{COOH})$.

Although the particular chemistry is not discussed at this stage, it is worthwhile to comment already on Figure 6b from a methodological point of view. It demonstrates nicely a crucial point of the employed metadynamics technique, namely the efficient exploration of the free energy landscape which can lead to uncovering new and unexpected minima and interconnecting pathways, i.e., chemically speaking it can lead to the discovery of novel intermediates, products, and reaction mechanisms. In the

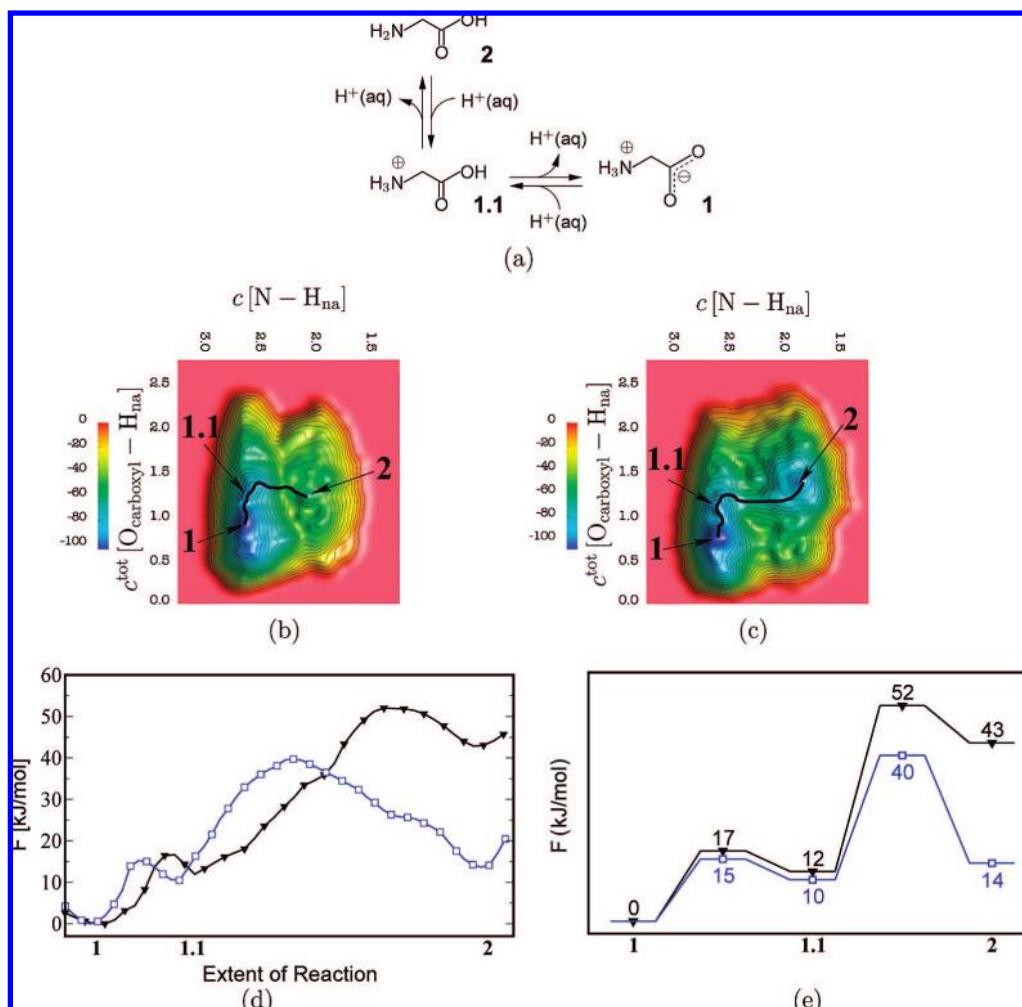(76) Bussi, G.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2006**, *96*, 090601.

**Figure 4.** Interconversion of neutral and zwitterionic isomers of glycine in water. (a) Reaction mechanism for the interconversion between the neutral (**2**) and zwitterionic (**1**) isomers of glycine in aqueous solution at ambient and extreme thermodynamic conditions. The reconstructed free energy surfaces at (b) ambient conditions and (c) in hot-pressurized water (free energy values are reported in kJ/mol) together with a sketch of the minimum free energy pathway to guide the eye (thick black curve). (d) Minimum free energy profiles extracted from the reconstructed free energy surfaces and (e) corresponding free energy schemes for the equilibrium between **2** and **1** at ambient (black, filled triangles down) and extreme thermodynamic conditions (blue, open squares).

particular case shown in Figure 6b, minimum **4** corresponds to an isocyanate structure that formed spontaneously during metadynamics sampling after the ring opening of the NCA. Only subsequent to this finding, more detailed exploration of this most interesting feature of the free energy surface led to the discovery of the more efficient indirect route from *N*-thiocarboxyl glycine via isocyanate to NCA, as discussed in section 3.3.

## 3. Results and Discussion

**3.1. Protonation Equilibrium of Glycine at Extreme versus Ambient Conditions.** All studied reactions take place in aqueous environments, either at ambient (ABW) or extreme temperature and pressure (HPW) conditions as defined in section 2. Glycine, as every amino acid, exists in different protonation states in water. In the neutral state of the molecule, both the amino group and the carboxyl group are uncharged, whereas in the zwitterionic form, the amino group is protonated and the carboxyl group is deprotonated. It is well-known that the zwitterionic form is the thermodynamically preferred species in water at ambient conditions. However, for chemical reactions in which either the amino or the carboxyl group participate, the relative concentrations of the two isomers, which are governed by their relative stabilities as well as the activation

barriers for their interconversion play a crucial role. This equilibrium, in turn, is ruled by the properties of water which change greatly as a function of temperature and pressure as mentioned in the Introduction.[28] Therefore, it is paramount to understand the influence of elevated temperature and pressure on the equilibrium between the zwitterionic and the neutral form as indicated in step **A** within the full reaction sequence that is depicted in terms of a graphical synopsis in Figure 9.

The chosen collective coordinates, which capture the relevant free energy subspace, are the coordination numbers[74] of the nitrogen and the carboxyl oxygen atoms to all nonaliphatic hydrogen atoms in the system, $c[N-H_{na}]$ and $c^{tot}[O_{carboxyl}-H_{na}]$, respectively. By excluding only the two inert hydrogen atoms at $\alpha$-C, this most general choice allows for deprotonation and reprotonation of the species involving possibly any protons from the solvent as well as those at the carboxyl and amino group of the amino acid itself, without introducing any a priori bias. Using these generalized coordinates, the neutral state **2** is characterized by the coordination numbers $c[N-H_{na}] \approx 2$ and $c^{tot}[O_{carboxyl}-H_{na}] \approx 1$, whereas for the zwitterionic state **1** $c[N-H_{na}] \approx 3$ and $c^{tot}[O_{carboxyl}-H_{na}] \approx 0$. Thus, trajectories in the two-dimensional reaction subspace spanned by
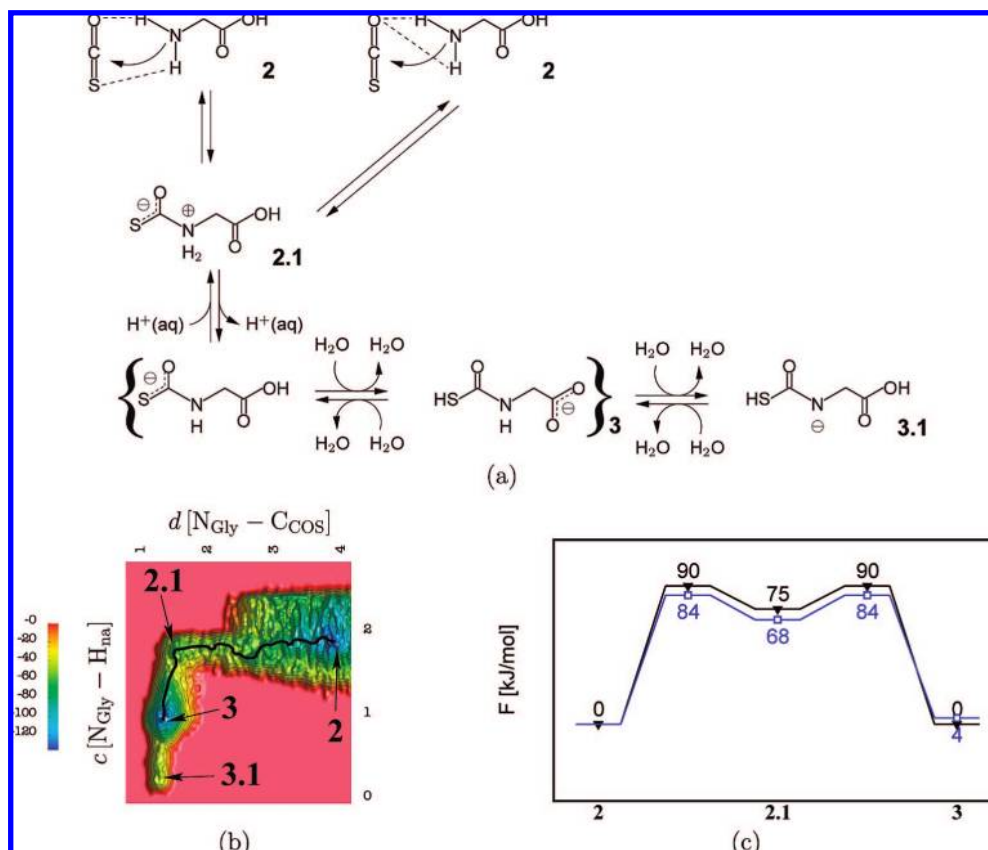
**Figure 5.** Reaction of glycine with COS. (a) Reaction mechanisms at ambient (left) and at hot-pressurized (right) conditions. Curly brackets show alternative protonation states at carboxyl/thiocarboxyl oxygen/sulfur atoms that were observed in the simulations. (b) Reconstructed free energy surface for the reaction in hot-pressurized water (free energy values are reported in kJ/mol) together with the sketch of the minimum energy pathway (black curve $2 \rightarrow 2.1 \rightarrow 3$). Note that the free energy surface for the reaction at ambient condition looks very similar to that at hot-pressurized condition. (c) Free energy schemes at ambient (black line and filled triangles down) and at extreme thermodynamic conditions (blue line and open squares).

$c[N-H_{na}]$ and $c^{tot}[O_{carboxyl}-H_{na}]$ allow to monitor changes of the protonation state in a very general way.

The reconstructed free energy surfaces, the resulting minimum free energy profiles, the simplified free energy scheme, and the extracted mechanisms for the interconversion of the two glycine isomers are collected in Figure 4 for both ambient conditions and hot-pressurized water. The initial configuration at both thermodynamic conditions was the neutral form of glycine **2**. Both variants of deprotonation/reprotonation processes, i.e., a stepwise and a concerted one, were observed at both conditions where water plays an active role of a proton relay (see Figure 4a). However, a pure intramolecular proton transfer channel, i.e., a direct proton transfer between the carboxyl and amino groups without involving any water molecule, was observed additionally at HPW conditions, but only after water-mediated interconversion between **1** and **2**. Even though the barrier for this process could not be properly obtained from the reconstructed free energy surface, the above observation gives a hint that a direct proton transfer is less preferred over the water-mediated interconversion.

The reconstructed free energy surfaces (see Figure 4b and c) show that at both conditions the zwitterionic species **1** is more stable than the neutral isomer **2** which is particularly clear in the one-dimensional free energy profile in panel d. Note that the values of the coordination numbers $c^{tot}[O_{carboxyl}-H_{na}]$ are much larger (mostly between 0.5 and 1.0) during the simulation than their ideal value of 0.0 when both carboxyl oxygens were deprotonated; see for instance the position of the minimum **1** in Figure 4b and c. This is because the total coordination number

included substantial contributions by hydrogen atoms in close vicinity to both carboxylate oxygens thus indicating relatively strong hydrogen bonding between them.

The minimum free energy pathways (sketched as black curves on the top of the free energy surfaces in Figure 4b and c) indicate that the conversion from zwitterion **1** to neutral **2** at ABW and HPW conditions begins with a proton transfer to one of the carboxyl oxygen atom. This corresponds to the shallow local minimum **1.1** near the free energy minimum for **1** as clearly seen from the one-dimensional free energy profile in panel d. It is also apparent that the relative stabilities of the two states change significantly with raising temperature and pressure. This allows one to compare the thermodynamics of such an equilibrium at extreme conditions relative to ambient in the absence of any other chemical reaction, which also applies to the other individual equilibria to be discussed in the subsequent sections. Consistent with a lower dielectric constant of water upon increasing both temperature and pressure,[28] the stabilization of the zwitterionic species **1** is reduced to only 14 kJ/mol in HPW while at ABW conditions it is 43 kJ/mol more stable than the neutral form **2** according to the simplified free energy scheme in panel e. It is interesting to descry that high temperature and pressure favor the conversion of the zwitterionic glycine into its neutral isomer, lowering the barrier from about 52 kJ/mol at ambient to nearly 40 kJ/mol at extreme conditions, i.e., from $\approx 21k_B T_{300}$ to $\approx 10k_B T_{500}$.[77] Note that it is useful to report relative stabilities also in units of $k_B T$ at the respective temperature, i.e., in the units of thermal energy fluctuations, in order to compare thermodynamic stabilities more easily at different temperatures.

In addition, reaction rates are approximately governed by $\sim\exp[-\Delta F^{\ddagger}/k_{B}T]$ so that reporting free energy barriers in units of $k_{B}T$ allow for a direct assessment of differences. In this way, direct insights into the change of kinetics as a result of changing the pressure/temperature conditions from ambient to extreme is provided for this and all subsequent reaction steps. Coming back to the de/reprotonation equilibria at ambient versus extreme conditions, it is observed that the reverse reaction, i.e., converting the neutral form to the zwitterion, is hampered by extreme conditions. The free energy barrier of this reaction is raised to about 26 kJ/mol (corresponding to $\approx 6 k_{B}T_{500}$) at HPW conditions which amounts to a factor of 3 relative to ambient where this barrier is about 9 kJ/mol (or $\approx 4 k_{B}T_{300}$).

For ABW, the obtained relative stabilities and activation barriers are consistent with previous ab initio simulation results[78] as well as with experimental data.[79] The activation barrier for the conversion of the zwitterionic into the neutral isomer was calculated previously to be 53.1 kJ/mol,[78] and the reported experimental value is 59.8 kJ/mol[79] compared to our estimate of about 52 kJ/mol. The obtained free energy difference between the neutral and zwitterionic forms of 43 kJ/mol is also in accord with that of the previous simulation[78] (46.8 kJ/mol) wheres the experimental value is 30.4 kJ/mol.[79] Note that, in contrast to the Helmholtz free energies $\Delta F$ computed in our simulations as well as those reported in ref 78, the cited experimental work refers to Gibbs free energies $\Delta G$.

Overall, the influence of temperature and pressure on the equilibrium between different protonation states of glycine in water is dramatic and in line with general expectations.[28] The neutral form is not only stabilized at extreme conditions but also the barrier for the interconversion from the neutral to the zwitterionic isomer raises with temperature and pressure. Both effects lead to a higher concentration of the neutral isomer in hot-pressurized water in comparison to ambient. This will have important ramifications on the activation of amino acids at HPW conditions relative to ambient, since it is only their neutral form that is suitable for nucleophilic attack on COS as will be shown in the following.

**3.2. Reaction of Glycine by Carbonyl Sulfide.** The next step in the activation process is the reaction between an amino acid and COS (carbonyl sulfide) to yield an *N*-thiocarboxyl amino acid ("thiocarbamate") denoted as step **B** in the complete synthesis sequence as summarized in Figure 9. To capture this process in simulations, the distance between the nitrogen atom of glycine, $N_{Gly}$, and the carbon atom of COS, $d[N_{Gly}-C_{COS}]$, together with the coordination number of $N_{Gly}$ to all but the aliphatic hydrogen atoms, $c[N_{Gly}-H_{na}]$, were chosen as collective coordinates. Upon reaction, the $N_{Gly}-C_{COS}$ distance decreases to a typical covalent distance of $\approx 1.3$ Å and the coordination number $c[N_{Gly}-H_{na}]$ changes from about 2 to $\approx 1$. Since only an amino group but not an ammonium group can attack COS as a nucleophile, a glycine molecule in its neutral form **2** as well as a COS molecule were used to start the simulations. This is in the spirit of experiments where basic pH, which preserves the amino group from being protonated, has been used in studies on COS-activated prebiotic synthesis of peptides at ambient reaction conditions.[15] Thus, in order to prevent the amino group from being protonated but still to allow

for hydrogen-bonding, a repulsive potential was added if the coordination number exceeded a certain cutoff value, $c[N_{Gly}-H_{na}] \geq 2.4$, which would otherwise lead to protonation of glycine. Effectively, it means that the issue of relative concentrations of neutral versus zwitterionic glycine species has been disconnected from reaction step **B** since it was already considered separately in step **A**. On the basis of the same line of arguments, an additional such wall potential was placed if $d[N_{Gly}-C_{COS}]$ exceeded 4.2 Å to make sure that the reactants do not diffuse too far from each other. Sampling such diffusive modes sufficiently would be prohibitively slow even if a rather small periodic supercell is used. It is noted in passing that both potentials were introduced only to sample the relevant part of the free energy surface exclusively; not adding them would be computationally demanding, and in addition, it is necessary only for either sampling explicitly the protonation equilibrium involving one of the reactants together with the reaction of interest or to include purely diffusive processes of both reactants. Of course, the effect of such wall potentials must be taken into account in the analysis of the free energies by not considering the sampled regions exceeding 4 Å, where a certain artificial deepening of the free energy surface due to the wall potentials was observed.

The full mechanism of this reaction step as well as the corresponding free energy scheme are depicted in Figure 5a and c, respectively, whereas the underlying free energy surface is shown in panel b; note that only the free energy surface for HPW conditions is shown here as the one corresponding to ABW is qualitatively similar. As the first step of the reaction the amino group of *neutral* glycine **2** attacks the carbon atom of COS resulting in the adduct **2.1**. The situation prior to this nucleophilic attack is slightly different at ABW and HPW conditions. In the former case, the two amino hydrogen atoms form hydrogen bonds with the sulfur and the oxygen atoms of COS (marked schematically in Figure 5a by dotted lines in the left structure **2**). These noncovalent interactions orient the COS molecule into an appropriate position in ABW, whereas at HPW the two hydrogen bonds are accepted by the same oxygen atom, i.e., $O_{COS}$ (see dotted lines in the right structure **2**). Since the basic chemistry is the same, the barriers at ABW and HPW conditions are similar as well, 90 and 84 kJ/mol, respectively (see Figure 5c). But, this reaction will be clearly faster in HPW largely due to the temperature effect since these barriers are $36 k_{B}T_{300}$ and $20 k_{B}T_{500}$ in the relevant thermal energy units. Thus, the relative free energy barrier at extreme conditions is effectively only half the barrier at ambient conditions. The adduct **2.1** deprotonates at both conditions over a barrier of $\approx 15$ kJ/mol at the nitrogen atom to give the *N*-thiocarboxyl glycine **3** which implies again a faster reaction at HPW conditions, $4 k_{B}T_{500}$ versus $6 k_{B}T_{300}$ due to increased thermal fluctuations. Note that we observe two different protonation states of *N*-thiocarboxyl glycine while sampling the product configurations: deprotonation at the sulfur together with protonation at the carboxylic group oxygen atoms and vice versa. Both these structures belong to the free energy basin of the products and are represented within a curly bracket in Figure 5a since they are sampled therein together.

On the basis of the resulting simplified free energy scheme Figure 5c, it can be judged that **2.1** clearly is an intermediate, which reacts however forward to **3** and backward to **2** with equal probability. Moreover, the two barriers involved in the reverse reaction ($3 \rightarrow 2.1 \rightarrow 2 + COS$) are the same as for the forward reaction ($2 + COS \rightarrow 2.1 \rightarrow 3$). This makes clear the point that

(77) Note that some free energies reported here deviate insignificantly from data published in our preliminary investigation[50] due to a refined analysis applied here.

(78) Leung, K.; Rempe, S. B. *J. Chem. Phys.* **2005**, *122*, 184506.

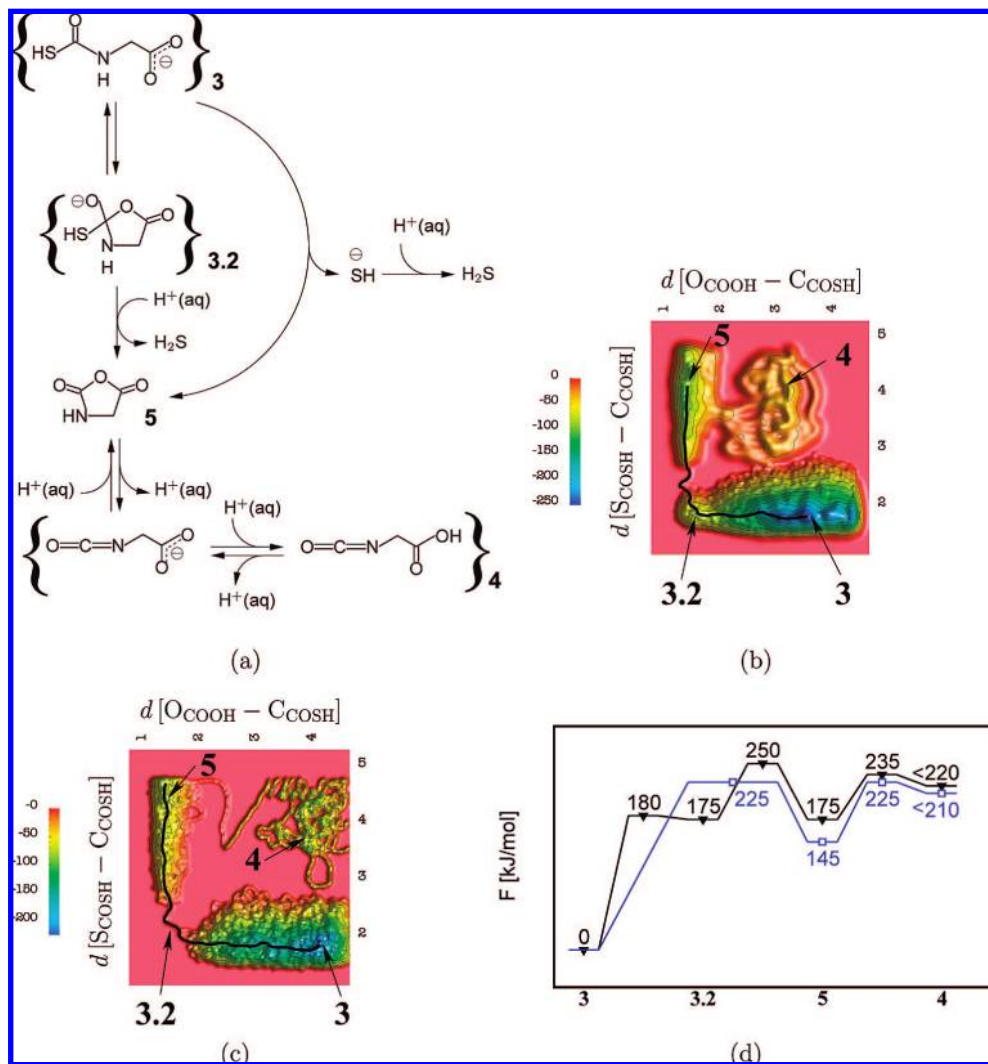(79) Slifkin, M. A.; Ali, S. M. *J. Mol. Liq.* **1984**, *28*, 215–221.

**Figure 6.** Direct formation of glycine NCA from the corresponding thiocarbamate. (a) Reaction mechanisms at ambient conditions and in hot-pressurized water. The long-curved arrow from **3** to **5** shows the concerted cyclization pathway in hot-pressurized water. Curly brackets show alternative protonation states at carboxyl/thiocarboxyl oxygen/sulfur atoms that were observed in the simulations. The reconstructed free energy surfaces for (b) ambient and (c) hot-pressurized conditions (free energies are reported in kJ/mol) are shown together with the sketch of the minimum free energy pathway (black curve). (b) Obtained free energy schemes at ambient (black line and filled triangles down) and at extreme thermodynamic conditions (blue line and open squares).

higher temperature and pressure do not lead to an enhanced net production of the $N$-thiocarboxyl glycine **3** but rather lead to a more dynamic equilibrium between reactants and products at HPW conditions.

At both conditions an additional minimum, or plateau, **3.1** could be identified with a $c[N_{Gly}-H_{na}]$ value near to 0 and a $N_{Gly}-C_{COS}$ distance about 1.3 Å; see Figure 5b for the corresponding free energy surface at HPW conditions. This class of structures represents an $N$-thiocarboxyl glycine which is deprotonated at the nitrogen atom. The barrier for this deprotonation is about 90 and 60 kJ/mol at HPW and ABW conditions, respectively. The species **3.1**, obtained in this reaction merely as a side product, will be shown later to act as a key intermediate along the reaction pathway from $N$-thiocarboxyl glycine **3** to isocyanato carboxylic acid **4**; i.e., step **C** in Figure 9 as discussed in section 3.3.2.

To conclude, the rate-limiting elementary step in process **B** is the nucleophilic attack of the amino group on COS at both conditions. The total activation barriers are estimated to be $36k_BT_{300}$ and $20k_BT_{500}$ at ABW and HPW conditions and are similar for forward and backward reactions, showing that

extreme conditions make the equilibrium between glycine and its thiocarbamate more dynamic without shifting it in a certain direction.

**3.3. Formation of Glycine NCA.** The third and crucial step on the way to an activated amino acid is the reaction of the thiocarbamate of an amino acid, i.e., $N$-thiocarboxyl glycine, to the corresponding NCA or Leuchs anhydride, which is believed to be rate-limiting for the *entire* peptide synthesis cycle.[15] However, as reviewed in section 1 the mechanism for the formation of NCAs from such thiocarbamates is still not known, see Scheme 39 in ref 5, although there are speculations that an isocyanate species should be involved. Thus, two alternative hypothetical[15,19] mechanisms must be investigated at this point. The first one is the direct cyclization of the $N$-thiocarboxyl glycine (route **C′** in Figure 9) while the second involves the formation of an isocyanato carboxylic acid intermediate (route **C**), which reacts subsequently to NCA (route **D**). Here we recall in passing the discussion from section 2 that the indirect isocyanate route was actually found by the metadynamics technique without imposing it.
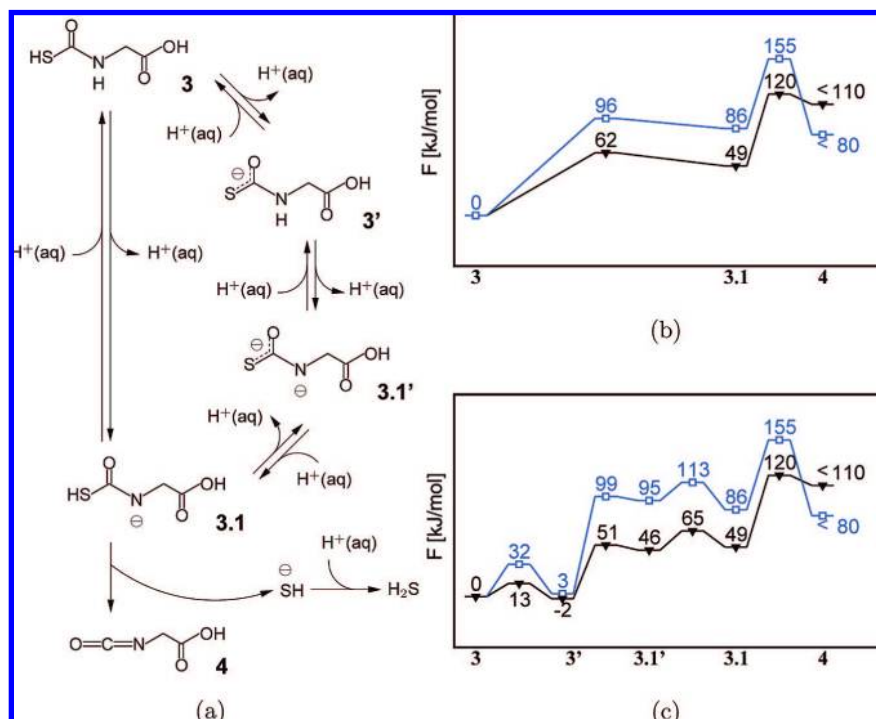
**Figure 7.** Conversion of N-thiocarboxyl glycine to α-isocyanato carboxylic acid ("isocyanate" species). (a) Reaction mechanisms. The left branch, **3** → **3.1**, shows the mechanism corresponding to the lowest free energy path while on the right, **3** → **3'** → **3.1'** → **3.1**, an alternative mechanism is presented, which is higher in free energy. (b) Obtained free energy schemes for the lowest free energy path. (c) Free energy schemes for the alternative mechanism at ambient (black line and filled triangles down) and at extreme thermodynamic conditions (blue line and open squares).

**3.3.1. Direct Cyclization of N-Thiocarboxyl Glycine.** For the direct cyclization of **3** (route **C'** in Figure 9), the chosen collective coordinates involve the distances between one of the oxygen atoms of the carboxyl group and the carbon atom of the COS moiety, $d[O_{COOH}-C_{COSH}]$, and between the carbon and the sulfur atoms within the COS group, $d[S_{COSH}-C_{COSH}]$. As initial configuration, the neutral form of the N-thiocarboxyl glycine **3** was used and the initial values for the collective coordinates were $d[O_{COOH}-C_{COSH}] \approx 4$ Å and $d[S_{COSH}-C_{COSH}] \approx 1.8$ Å. To prevent a significant diffusion of the H$_2$S from the generated NCA, a repulsive potential was applied along the $S_{COSH}-C_{COSH}$ distance coordinate at 4.8 Å similar to the procedure described above when studying step **B**.

The corresponding reaction mechanisms together with the associated free energy landscapes and the free energy schemes at ABW and HPW conditions are shown in Figure 6. The reactant state is the N-thiocarboxyl glycine **3** in its various possible protonation states of the carboxyl and thiocarboxyl groups, which are not distinguished within the chosen set of collective coordinates as symbolized by the curly brackets around structure **3** in Figure 6a (see also Figure 5a). At ABW conditions, the reaction starts with the addition of one of the two equivalent carboxyl(ate) oxygens, $O_{COOH}$, to $C_{COSH}$. The barrier for this step from **3** to **3.2** is about 180 kJ/mol (or $\approx 72k_BT_{300}$). The resulting cyclic intermediate **3.2**, which is also observed in several protonation states as indicated by the curly brackets, is only marginally lower in energy than the transition state ($\approx 175$ kJ/mol versus $\approx 180$ kJ/mol). Subsequently, H$_2$S is eliminated after the protonation of the SH moiety in **3.2** by a hydronium ion which finally yields NCA **5**. The activation barrier for this step is about 75 kJ/mol (i.e., $\approx 30k_BT_{300}$).

This mechanism changes its character if the reaction is carried out at extreme conditions: at HPW conditions, NCA formation is concerted, **3** → **5**, whereas it is a two-step process at ABW,

**3** → **3.2** → **5**. In the HPW case, **3.2** is only a transition state according to the free energy scheme in Figure 6d, and SH$^-$ is eliminated instead of H$_2$S, which is readily protonated by water. This difference between ABW and HPW is clearly seen in the topology of the free energy surfaces too (compare Figure 6b to c). In particular, structure **3.2** corresponds to a (shallow) minimum in ambient water (b) before reaching the rate-determining transition state from **3** to **5**, whereas **3.2** is a saddle point connecting the two deep basins stemming from **3** and **5** at extreme conditions (c). The barrier for the formation of **5** is $\approx 225$ kJ/mol or equivalently about $54k_BT_{500}$. The nature of this reaction at HPW conditions is in accord with general expectations that the much lower dielectric constant of water at extreme thermodynamic conditions causes a less effective stabilization of charge-separated and even more so of charged species like **3.2** and thus favors concerted mechanisms that avoid charge-separated intermediates.

The large barriers for the initial step seem to be mainly due to large entropic contributions to the free energy basin corresponding to reactants. Both, the carboxyl as well as the COS group are observed in different relative distances and orientations as well as in different protonation states featuring a myriad of possible hydrogen bonding patterns. The total barrier for the cyclization of N-thiocarboxyl glycine is about $100k_BT_{300}$ at ABW. Although the barrier remains still substantial, the extreme thermodynamic conditions lower it by a factor of 2, to $54k_BT_{500}$, which is predominantly an effect due to increased thermal fluctuations at the elevated temperature. The relative stability of the NCA with respect to the reactants could not be obtained directly because no complete sampling of the spacious free energy basin of **4** was achieved. However, the reverse reaction involves an H$_2$S attack, which is expected to have a very high barrier associated to it so that one can regard the forward reaction as being nearly irreversible.

Interestingly, the generated NCA **5** decays at both conditions to an isocyanato carboxylic acid **4** with a barrier of ca. 60 kJ/mol (i.e., $\approx 24k_BT_{300}$) and about 80 kJ/mol (i.e., $\approx 19k_BT_{500}$) at ABW and HPW conditions, respectively. Clearly, the barriers for this reaction can only be regarded as upper bounds since the N–H coordination number is missing in the collective coordinates, which is crucial for this process. The accidental formation of an isocyanate **4** by ring opening of NCA **5** during this simulation, once **5** was formed through this direct cyclization route, was a first strong hint that **4** might indeed be an important intermediate species in the formation of NCA **5** out of the *N*-thiocarboxyl glycine **3**, which will be fully explored in the next section.

**3.3.2. Indirect Isocyanate Route.** An alternative mechanism that leads to NCAs is hypothesized to proceed through isocyanato carboxylic acids. In the previous section, it was shown that it can indeed be obtained as a *decay* product in metadynamics simulations *after* the NCA has been formed. However, it might also be *generated* from the *N*-thiocarboxyl glycine **3** via the indirect route **C** according to Figure 9 in order to facilitate NCA synthesis as suggested in ref 19. The NCA **5** itself, in turn, could form from this isocyanate intermediate via route **D**. To investigate this possibility, the reaction of *N*-thiocarboxyl glycine to isocyanate and its subsequent cyclization to NCA were studied in detail in terms of free energy landscapes and reaction mechanisms, again at normal and extreme conditions.

**Formation of the Isocyanate.** During formation of the isocyanate from the *N*-thiocarboxyl glycine, coordination numbers of nitrogen and sulfur atoms with respect to nonaliphatic hydrogen atoms, $c[N–H_{na}]$ and $c[S–H_{na}]$, as well as the distance between carbon and sulfur atoms within the COS moiety, $d[S–C]$, change. Thus, to capture the entire process all these three collective coordinates were chosen simultaneously to span the reaction coordinate space. Again, repulsive potentials were applied to restrict $d[S–C]$ as well as $c[N–H_{na}]$ not to exceed 4.5 Å and 1.5, respectively, as explained before. Although **3** was chosen as the initial structure at both conditions, it deprotonated readily at ambient conditions at the sulfur site to yield **3′** during equilibration. The reaction mechanisms along with the corresponding free energy schemes are shown in Figure 7; note that the full free energy landscapes are not depicted since these three-dimensional surfaces are difficult to represent due to the inherent topological complexity which can be gathered from panels b and c.

For the sake of clarity and better understanding, it is appropriate to divide the mechanism into two parts: the formation of **3.1** and its conversion to **4**. For the formation of **3.1**, two alternative routes could be identified at both thermodynamic conditions. The first route is the direct *N*-deprotonation of **3**. The free energies of activation are about 62 kJ/mol (i.e., $\approx 25k_BT_{300}$) and nearly 96 kJ/mol (i.e., $\approx 23k_BT_{500}$) at ABW and HPW conditions, respectively. In ambient water, the resulting species **3.1** is approximately 49 kJ/mol less stable than **3** and the reverse barrier is only about 13 kJ/mol (or $\approx 5k_BT_{300}$). As can be expected from the lower dielectric constant of water at high temperature, the destabilization of **3.1** is even larger at HPW than at ambient conditions. This is reflected in the high relative free energy of about 86 kJ/mol of **3.1** and a barrier for reprotonation of nearly 10 kJ/mol (i.e., $\approx 2k_BT_{500}$).

Before proceeding with the second part of the total mechanism, it is worthwhile to examine the alternative route to **3.1**. The difference to the previously described one is the presence of two additional species, **3′** and **3.1′**. The former is the result
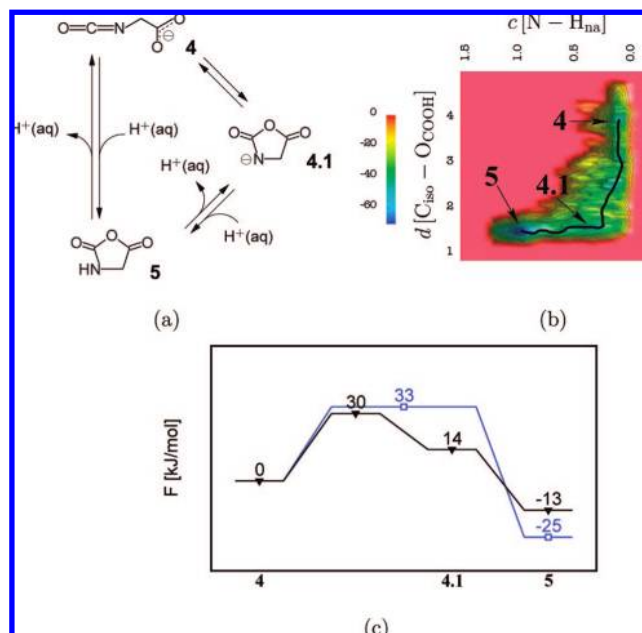


**Figure 8.** Cyclization of the isocyanate. (a) Reaction mechanisms for the cyclization of isocyanate at ambient conditions (right: **4 → 4.1 → 5**) and in hot-pressurized water (left: **4 → 5**). (b) Reconstructed free energy surface for the reaction in ambient water together with the sketch of the minimum free energy pathway (black curve); free energies are reported in kJ/mol. (c) Obtained free energy schemes at ambient (black line and filled triangles down) and at extreme thermodynamic conditions (blue line and open squares).

of an S-deprotonation of **3**. This process is associated with a free energy of activation of approximately 13 kJ/mol (i.e., $\approx 5k_BT_{300}$) in ambient water and about 32 kJ/mol (i.e., $\approx 8k_BT_{500}$) at HPW conditions. Note that **3′** is almost isoenergetic to **3**. As shown in Figure 7a, species **3′** looses another proton, this time at the nitrogen atom, surmounting a barrier of about 51 kJ/mol (i.e., $\approx 20k_BT_{300}$) at ABW conditions and $\approx 100$ kJ/mol (i.e., $\approx 24k_BT_{500}$) in hot-pressurized water. The 2-fold negatively charged species **3.1′** is extremely unstable as underlined by the low reverse barrier of only about 5 kJ/mol at both conditions. In the last step of this alternative route, **3.1′** is reprotonated at the sulfur atom yielding **3.1**. The barrier for this protonation is about 19 kJ/mol ($\approx 8k_BT_{300}$) and approximately 18 kJ/mol (i.e., $\approx 4k_BT_{500}$) at ABW and HPW conditions, respectively. Overall, the direct deprotonation **3 → 3.1** is preferred over the stepwise pathway via **3′** and **3.1′**.

The conversion of **3.1** to **4** is the second part of this reaction sequence from **3** to **4**. Analysis of the trajectory reveals that species **3.1** has been formed many times, but has been reprotonated by water. Along the metadynamics trajectory, soon before the transition of **3.1** to **4**, we have noticed a breakage of hydrogen bonds between the surrounding water molecules and the nitrogen. The transition of **3.1** to **4** is accompanied by the release of a thiolate ion. The elimination of SH⁻ is expected to be the bottleneck because of the poor leaving nature of this group, which is underpinned by the noticeable acceleration in the peptide formation by alkylating and oxidizing agents in experiments.[15] Indeed, at both conditions the system must surmount a barrier of about $\approx 70$ kJ/mol (corresponding to about $28k_BT_{300}$ and $17k_BT_{500}$ at ABW and HPW conditions, respectively) to eliminate the thiolate ion. The solvated SH⁻ is finally protonated to yield $H_2S$. The observed barrier for this process is about 15 kJ/mol (i.e., $\approx 6k_BT_{300}$) in ambient water whereas the protonation of SH⁻ is barrierless at HPW extreme conditions.
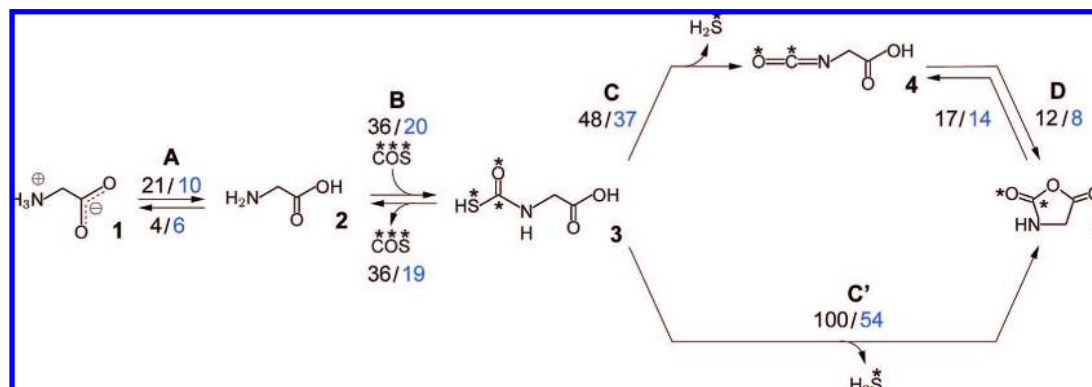
**Figure 9.** Complete reaction sequence describing the activation of glycine **1** by COS leading to the formation of NCA **5** via the direct cyclization route (lower pathway) and via the indirect isocyanate route (upper pathway). All atoms of the activating COS molecule are marked by stars throughout. The calculated free energy barriers for the two thermodynamic conditions investigated are given (in units of $k_BT$) in the sequence (color code) "ambient bulk water (black)/hot-pressurized water (blue)".

Since at both conditions, the thiolate elimination is the bottleneck of the entire reaction, the total activation barriers for the conversion of the *N*-thiocarboxyl glycine **3** to isocyanate **4** is about 120 kJ/mol (i.e., $\approx 48k_BT_{300}$) and 155 kJ/mol (i.e., $\approx 37k_BT_{500}$) at ABW and HPW conditions, respectively. The increase in the barrier at extreme conditions could be mainly due to the fact that the elimination of charged species like SH⁻ is quite unfavorable in HPW due to the low dielectric constant of the solvent. This is supported by the fact that protonation of SH⁻ is immediately occurring after the elimination in HPW in contrast to ABW where protonation occurs only after a rather long time. Compared with the free energy barrier for the SH⁻ elimination, the differences in the activation energies along the two routes to **3.1** appear rather small (see Figure 7a) indicating an equilibrium between the species **3**, **3′**, **3.1′**, and **3.1**. This observation implies that both routes contribute to the formation of **3.1**. It should be stressed at this point that in a real experiment this reaction is nearly irreversible since effectively H₂S is eliminated which escapes from an open reaction system.

**Cyclization of the Isocyanate.** In the final step on the way to NCAs, the synthesized isocyanate **4** must undergo a cyclization reaction according to path **D** in Figure 9. In order to describe the cyclization process, the distance between the carbon atom of the isocyanate group and an oxygen atom of the carboxyl group, $d[C_{iso}-O_{COOH}]$, and the coordination number of the nitrogen atom with respect to all nonaliphatic hydrogen atoms, $c[N-H_{na}]$, were chosen as collective coordinates. During the cyclization reaction, $d[C_{iso}-O_{COOH}]$ should decrease and $c[N-H_{na}]$ should raise from zero to approximately unity; as explained before repulsive potentials were imposed along both coordinates. On the basis of the experience from the investigation of the direct cyclization of the *N*-thiocarboxyl glycine, such wall potentials were applied for distances at $d[S_{iso}-O_{COOH}] > 4.3$ Å and at $c[N-H_{na}] > 1.4$ in case of the coordination number.

The mechanism and the corresponding free energy schemes are shown in Figure 8a and c, respectively, for both ABW and HPW, whereas the reconstructed free energy surface is presented in Figure 8b only at ambient conditions. At ABW conditions the reaction is a stepwise process involving two elementary steps. The first step is the intramolecular attack of $O_{COOH}$ at the carbon atom of the isocyanate group and is associated with a barrier of about 30 kJ/mol (i.e., $\approx 12k_BT_{300}$). The resulting cyclic structure **4.1** can be identified as a plateau at about 14 kJ/mol on the free energy surface from which the system relaxes readily in a barrierless protonation process to NCA **5**. From the free

energy scheme in Figure 8c, it is evident that this product is about 13 kJ/mol more stable than the reactants and the reverse barrier is about 43 kJ/mol (or $\approx 17k_BT_{300}$).[77]

In HPW, the mechanism is again different in the sense that extreme thermodynamic conditions favor concerted processes which minimize the transient existence of charged intermediates. Thus, the nitrogen atom is immediately protonated after the nucleophilic attack of $O_{COOH}$ at the isocyanate group and the barrier separating **4** and **5** is approximately 33 kJ/mol (or $\approx 8k_BT_{500}$). Because the activation barrier for the reverse reaction is about 58 kJ/mol (i.e., $\approx 14k_BT_{500}$), **5** is about 25 kJ/mol more stable than the isocyanate **4**, suggesting that the equilibrium is clearly shifted toward the production of **5** also at extreme thermodynamic conditions.

## 4. Conclusions and Outlook

The presented investigations assess both the detailed mechanistic aspects and the free energies involved in the activation of amino acids via the formation of "Leuchs anhydrides" or NCAs ($\alpha$-amino acid *N*-carboxyanhydrides) in ambient bulk water (ABW) and at hot-pressurized bulk water (HPW) extreme thermodynamic conditions using COS (cf. Figure 1). Such activation is believed to be a crucial step in the direct formation of oligopeptides out of amino acids in aqueous environments. In addition to this synthetic aspect, NCA activation has been proposed to be relevant in the framework of putative prebiotic peptide synthesis, noting that many competing ideas can be found in the vast literature as discussed in the Introduction. The key results of this computational study can be summarized in Figure 9, but in addition to this overview, the mechanistic and energetic details of all reaction pathways involved have been elucidated at both ambient and extreme conditions.

The first important observation is that extreme conditions stabilize the required neutral forms of reactants, such as amino acids, and intermediates compared to ambient, thus increasing their concentration and thereby enhancing the productivity of activation at neutral pH conditions. This is qualitatively consistent with the significant lowering of the dielectric constant of water upon heating and compression thus making the solvation of charged or charge-separated species, intermediates, and transition states less favorable. Second, due to the increased temperature the speed of all the reaction steps is increased in HPW, which is made evident by a comparison of barriers in the appropriate units of $k_BT$ (see Figure 9). On the other hand,

the destabilization of charged species increases the free energy barrier for the elimination of SH⁻ in HPW compared to ambient conditions. Third, HPW conditions favor concerted mechanisms over stepwise reaction pathways found in ABW which is indeed observed at several instances. Though our simulation results provide valuable insights into the influence of extreme conditions on both thermodynamic and kinetic aspects of NCA formation according to Figure 9, only a *global* free energy hypersurface (that is spanned *simultaneously by all reactions coordinates* needed in *all reaction steps*) supplemented by microkinetic network modeling (taking into account all possible intermediates and side reactions including all partial fluxes) would allow for a full assessment. Unfortunately, this is currently far beyond both methodological and computational possibilities.

Another important result of this work is that glycine NCA is not formed from *N*-thiocarboxyl glycine ("thiocarbamate") via direct cyclization but, rather, by taking the more complicated but (free) energetically preferred so-called isocyanate route. This strongly supports those mechanistic speculations that involve such isocyanate intermediate species[10,15,19] over other ideas. Furthermore, the simulations clearly show that the formation of the isocyanate species is the rate-determining step for the activation of glycine at both ambient and extreme aqueous conditions. From the computed free energy barriers underlying this rate-limiting step, it can be crudely estimated that the time scale of NCA formation, and thus of amino acid activation, decreases from the order of years in ambient water to roughly minutes at extreme temperature and pressure conditions! This, however, remains for future experimental studies probing explicitly the suggested effects of extreme temperatures and pressures on this interesting reaction network.

When it comes to peptide synthesis in aqueous environments based on amino acids as a next step, these results show that the crucial activation, i.e., NCA formation, is indeed sped up considerably when going from the usual ambient reaction conditions to hot-pressurized water as a medium. This, in turn, will alter the relative stability of reactants in favor of NCA production at neutral pH conditions when applying elevated temperatures and pressures. Clearly, a comprehensive understanding of the NCA-mediated pathway in the hypothetic prebiotic formation of peptides entails scrutinizing alternative reaction routes to NCA formation itself, side and back reactions affecting the productivity of the peptide synthesis cycle, and the influence of catalytic minerals and extreme thermodynamic conditions on all these reactions. Work along these lines is progressing in our laboratory.

Transcending the specific reactions studied and their relevance to widely discussed prebiotic chemistry scenarios, the presented study of chemical reactions at extreme conditions will also be of interest to organic chemistry in aqueous high-pressure and high-temperature environments in more general terms. This is because our investigation contributes detailed molecular insights on the typical influence of temperature and pressure on mechanisms underlying several classes of chemical reactions in terms of relative stabilities of intermediates versus transition states, zwitterionic versus neutral reactive species, concerted versus stepwise processes, and thus on the reaction pathways that connect reactants with products.