Step by Step Strecker Amino Acid Synthesis from ab Initio Prebiotic Chemistry

Théo Magrino, Fabio Pietrucci, A. Marco Marco Saitta

Step by Step Strecker Amino Acid Synthesis from ab Initio Prebiotic Chemistry

Théo Magrino,* Fabio Pietrucci,* and A. Marco Saitta*

Sorbonne Université, Muséum National d'Histoire Naturelle, UMR CNRS 7590, IMPMC, Paris, France

E-mail: theo.magrino@sorbonne-universite.fr; fabio.pietrucci@sorbonne-universite.fr; marco.saitta@sorbonne-universite.fr

Abstract

The amino acids synthesis from elementary precursors in abiotic conditions is traditionally described according to the Strecker reaction, thoroughly invoked to justify the observation of amino acids in extraterrestrial samples, and their emergence in the primordial Earth. To this day, however, a quantitative microscopic description of the mechanism, thermodynamics and kinetics of the multi-step Strecker reaction is still lacking. In the present work we tackle this study by adopting a state-of-the-art *ab initio* computational approach, combining an efficient scheme of exploration of the relevant chemical networks with a rigorous determination of the underlying free energy and transition states. We determine the step-by-step chemical pathway from "Strecker precursors" to glycine in solution, and calculate the corresponding full free energy landscape. Our results agree well with the scarce available experimental data, and complete them, thus providing the first end-to-end study of this complex reaction, a crucial bottleneck for the emergence of life.

Keywords : Strecker reaction, free energy landscape, ab initio molecular dynamics, glycine, prebiotic synthesis

TOC graphical entry.
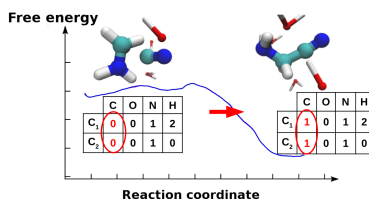
Understanding the prebiotic appearance of amino acids in the primordial Earth has received enormous attention since the Miller experiments in the 1950's,[1,2] becoming one of the major issues in prebiotic chemistry and origin of life studies[3]. However, the variety of geochemical environments in which amino acids may have appeared, also including extraterrestrial ones, makes the identification of accurate and realistic synthetic pathways for their formation extremely challenging. Amino acids formation environments may have included for instance meteorites, water vapour-rich volcanic eruptions, interstellar media, meteorites, or high-velocity cometary impacts[4–11] . This complexity is enhanced by the number of molecular species, possibly present in the early Earth, which could have been precursors in amino acid synthesis, including (but not restricted to) water, ammonia, hydrogen cyanide, carbon dioxide, molecular nitrogen, carbon monoxide, methane, molecular hydrogen.

Ever since the seminal Miller experiment, the observed spark-discharge-driven synthesis of amino acids was preceded by the appearance, in the aqueous solution in contact with the gas phase, of reaction intermediates such as aldehydes and cyanide.[1] This observation led to interpret Miller experiments, and in general the prebiotic synthesis of amino acids, in terms of the "Strecker-cyanohydrin" multi-component reaction[12] (see figure 1). This synthesis involves water, formaldehyde, cyanide and one catalytic ammonia to lead to glycine, the simplest amino acid, but also larger amino acids starting from larger aldehydes. Miller studies further supported the Strecker mechanism hypothesis,[2] and gave, to some extent, thermodynamics and kinetics experimental data to benchmark the Strecker synthesis of glycine.[13] However, recent crucial analysis of meteorite content and experimental work stressed that (i) β- and γ- amino acids are found in meteorites, while Strecker mechanism can explain only α-amino acids formation and (ii) formation of amino acids could occur in experiments without cyanide addition (*i.e.* with only ammonia and aldehyde, hence not through Strecker mechanism).[14] Similarly, recent *ab initio* computer simulations addressed amino acid formation under extreme conditions (intense electric fields, meteorite impacts), suggesting the possible existence of alternative mechanisms.[15–18] For instance, electric field may allow original redox

processes and hydrogen transfers, leading to glycine formation with a formamide intermediate.[16] In the case of meteorite impacts, C-N bonded oligomers may have formed and their decomposition may have led to glycine derivatives.[18] When it comes to milder conditions, on the other hand, accurate quantum chemistry studies were generally focused on solventless static calculations in the low-temperature limit, thus neglecting entropic contributions apart from vibro-rotational harmonic corrections.[19,20] Thus, there is a lack of theoretical studies addressing more widespread, non-extreme environments, i.e., synthesis in solution of relevance to hydrothermal vents and parent bodies of meteorites.

Up to now, the Strecker synthesis is therefore the most (if not the only) accepted and studied chemical pathway towards amino acids synthesis in prebiotic conditions. However, a thorough study of the whole pathway, and the quantitative determination of the relevant thermodynamics and kinetics of each elementary step is, quite surprisingly, still lacking. This is particularly true with respect to the kinetics of the fastest chemical steps (the addition and elimination reactions yielding to glycinonitrile, the first experimentally stable intermediate ((**4**) in fig. 1). In fact, the glycinonitrile synthesis (reactions (**1**)→(**2**), (**2**)→(**3**), (**3**)→(**4**))) occurs quickly on the experimental timescale, thus without allowing isolation/identification of intermediate metastable states. The kinetics of this specific synthesis is determined only indirectly, using *e.g.* Taft empirical relationship.[21–24] Symmetrically, thermodynamic data lack for the rate limiting part of the synthesis, *e.g.* the nitrile hydration and the amide hydrolysis (reactions (**4**)→(**5**) and (**5**)→(**6**)). To our best knowledge, despite the crucial role of this synthesis in origins of life scenarios, few experimental studies have been undertaken to achieve a full, step-by-step, quantitative description of the kinetics and thermodynamics of the Strecker amino acids synthesis, and even more so in realistic prebiotic conditions.

To "grasp the nettle", we undertook a detailed theoretical study of the chemistry of the Strecker synthesis of glycine. We assessed the free energy differences and barriers involved in this mechanism, and compared them with experimental data. Eventually, we filled existing gaps in experimental data. We determined the full microscopic reaction mechanism, and
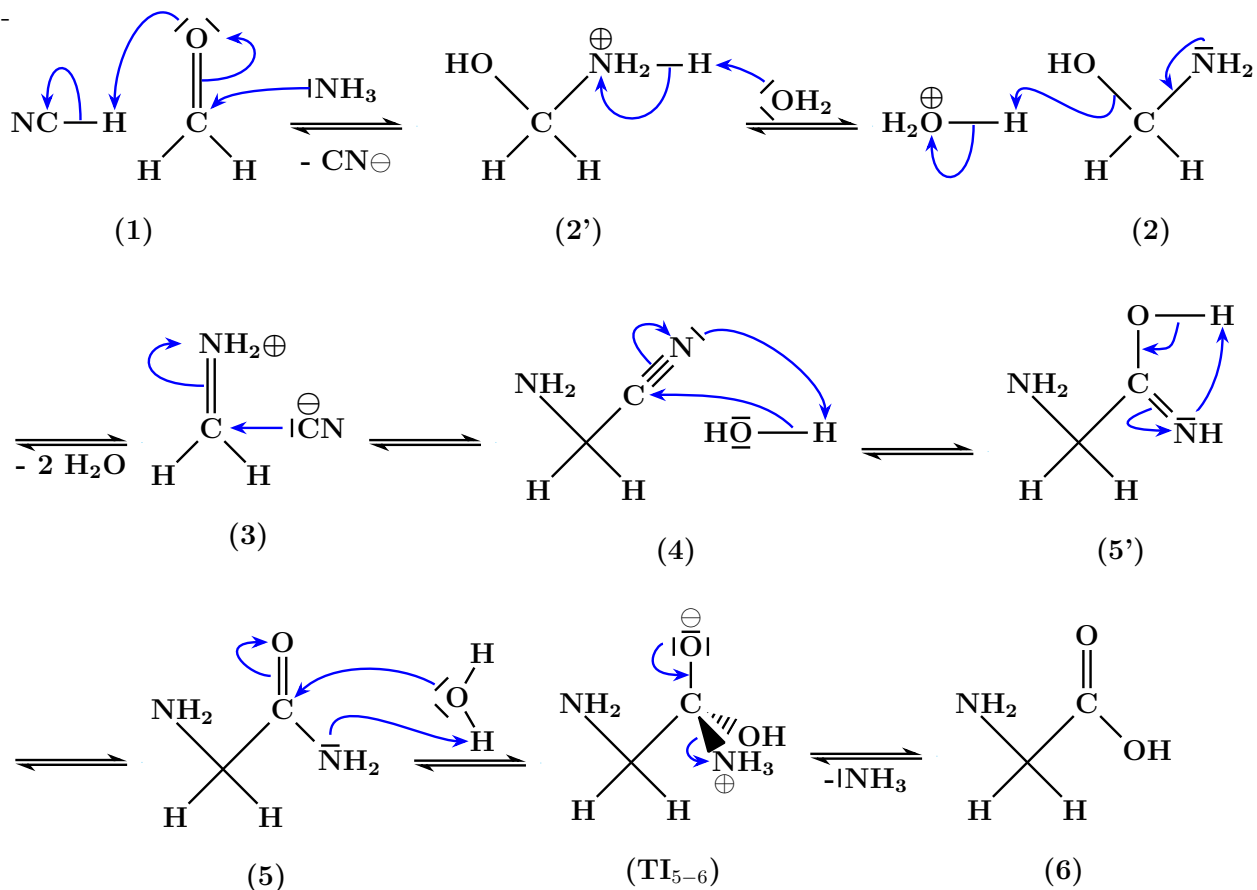
Figure 1: The Strecker synthesis of glycine, according to the detailed mechanism obtained in the present work. Labels denote a set of species solvated in 81 explicit waters. Neither acid or base is added to the simulation box. In the mechanism representation above, molecules not intervening in the chemical step are omitted (*e.g.* waters in (3)). Labels denote (1) formaldehyde, hydrogen cyanide, ammonia, water; (2) aminomethanol, cyanide, water, $H^+_{(aq)}$; (3) protonated methanimine, cyanide, two waters; (4) glycinonitrile (2-amino ethanenitrile) and two waters; (5) glycinamide (2-amino ethanamide), water; (6) glycine (2-amino ethanoic acid) and ammonia. In addition, both a metastable state (2') (N-protonated aminomethanol, cyanide, water) and a metastable state (5') (tautomeric aminoamide (NH_2CH_2COHNH), water) are met as intermediates in reactions (1)→(2) and (4)→(5), respectively. We characterised the zwitterionic tetrahedral intermediate (TI_{5-6}) as a transient intermediate of reaction (5)→(6). Proton exchanges may happen through a water molecule relay during the simulations. We emphasise that the origin and the fate of protons are not decided beforehand in simulations but emerge as a result of our unsupervised protocol of pathway exploration.

thus provided the first quantum-based *ab initio* framework of this fundamental reaction.

Quantitatively addressing the question of how, in general, chemical products form in solution, starting from given reactants, demands a versatile and efficient theoretical approach. In fact, it is indispensable to explicitly include the solvent molecules, which are instead often neglected or oversimplified in more standard computational chemistry studies, particularly in the prebiotic context. In this work, we undertake the first-ever (to our best knowledge) step-by-step ab initio study of the Strecker reaction in solution. To this end, we use *ab initio* molecular dynamics at the Density Functional Theory (DFT) level, including explicitly tens of solvent water molecules, thus simulating realistic solvation at a realistic concentration of 0.7 mol/L for initial formaldehyde, hydrogen cyanide, or final glycine.

In addition to the cost due to the explicit inclusion of the solvent, a foremost limiting factor in the predictive power of *ab initio* simulations is represented by the enormous gap between the experimental time scales and the ones that can be simulated. To tackle the time scale issue, we have developed an approach[25] based on state-of-the-art enhanced sampling techniques, combined with a flexible definition of the reaction coordinates (collective variables (CVs)) based on the topology of the atomic network. Our method, which includes path-based CVs,[26] has been successfully employed to study chemical reactions in the past, generally limited to a single or to isolated elementary acts.[27–30]

This approach handles the mechanistic analysis in a fully agnostic way, as, for a given elementary act, only the reactants and the products are constrained. Therefore, no input about possible transition states is given, and the system is thus free to find the most favourable way to react. In addition, our description takes the solvent into account, leaving the system free to choose which water molecules might interact with the solute and, for example, exchange H^+/OH^- with it. Here we deploy our approach for the first time to a complex multi-step synthesis, and a very meaningful one, as the Strecker reaction in aqueous solution. The technical details are described in the Methods section.

(a) Nitrile hydration transition state, step **(4)**→**(5')**. $\Delta F^{\ddagger} = 23$ kcal/mol.

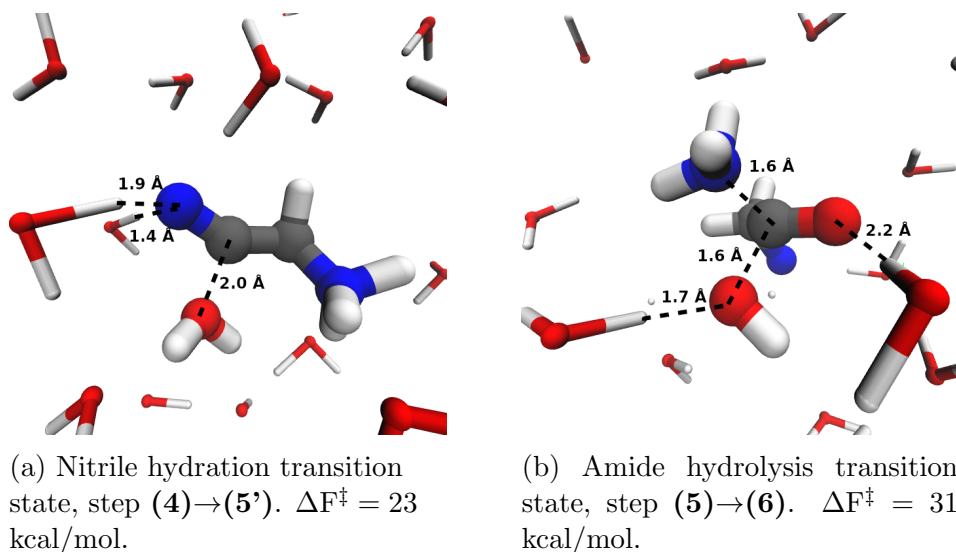(b) Amide hydrolysis transition state, step **(5)**→**(6)**. $\Delta F^{\ddagger} = 31$ kcal/mol.

Figure 2: Representative transition state structures for the reaction steps with the largest barriers. The geometries are extracted from umbrella sampling simulations, and the distances displayed are average values.

In order to introduce our results, we first represent, in figure 2, the transition states with the two highest activation free energies: nitrile hydration and amide hydrolysis. It is worth noting how explicit solvent molecular dynamics gives structural details about solvation that (i) are out of reach of current experimental methods and (ii) show fine and specific solvent/transition state interactions that are crucial for a realistic quantitative description. We note for instance that the negative charge on the nitrile moiety, due to the nucleophilic attack of water, is stabilised by a strong H-bond solvation, as shown by the short (1.4 Å) hydrogen-nitrogen distance in figure 2(a).

Concerning amide hydrolysis to glycine, the metadynamics exploration yielded an unexpected mechanism, with the concerted addition of an hydroxyde anion and the departure of an ammonia molecule as shown in figure 2(b). The mechanism expected from literature proceeds instead *via* a well-known tetrahedral intermediate.[31] However, umbrella sampling simulations in the barrier region display fast interconversion, on a picosecond timescale, between the concerted transition state and the tetrahedral intermediate, resulting in the

7

"volcano"-like shape of the free energy profile in figure 3(g), the intermediate lying within the shallow minimum at $s \approx 0.6$. This observation strengthens our results, as our optimisation of the reaction pathways was able to find, unsupervisedly, both the actual transition state, and the close-by tetrahedral intermediate.

We present in the following the full outcome of our study, *i.e.* the accurate quantitative free energy landscape for the Strecker reaction in aqueous solution. We calculate the free energy level for each (metastable) intermediate state, and the activation free energy for each elementary act of the mechanism.

In the following, we analyse the overall reaction by focusing on two main steps, the synthesis of glycinonitrile (steps (1) to (4)) and then the pathway to glycine (steps (4) to (6)), and we compare to (sparse) available experimental data whenever possible. Before starting, we underline that when comparing data from different sources, the formaldehyde hydration reaction must be considered: $H_2CO + H_2O \rightarrow H_2C(OH)_2$,with $\Delta_r H_{aq}^{\ominus} = -8.4$ kcal/mol, $\Delta_r S_{aq}^{\ominus} = -13$ cal/mol/K, $\Delta_r G_{aq}^{\ominus} = -4.5$ kcal/mol.[32] All values presented in this paper consider bare formaldehyde as displayed in figure 1, state **(1)**.

For the multi-step reaction **(1)→(4)**, we report a cyanide addition mechanism that proceeds *via* a protonated imine (iminium), consistent with the mechanism proposed in the experimental literature.[13] We note that the experimental difference of activation parameters of the forward and backward reactions in literature[13] is not equal to the thermodynamic difference $\Delta_r G_{aq}^{\ominus}(\mathbf{(1)}{\rightarrow}\mathbf{(4)})$, because the rate limiting steps are not the same in both directions. The experimental activation parameters for the overall reaction **(1)→(4)** are actually "effective" data, since elementary acts **(1)→(2')**, to **(3)→(4)** are confounded in experiments. The effective experimental activation free enthalpy is $\Delta G_{effective}^{\ddagger}(\mathbf{(1)}{\rightarrow}\mathbf{(4)}) = 10$ kcal/mol.[13] Few experimental data exist on this global result, which makes difficult a detailed comparison to our numerical value. Our observed rate limiting step for the overall process **(1)→(4)** has a barrier of 15 kcal/mol (elementary act **(2')→(2)**). The 5 kcal/mol discrepancy is quite satisfying considering the typical PBE-DFT error,[33] see below.
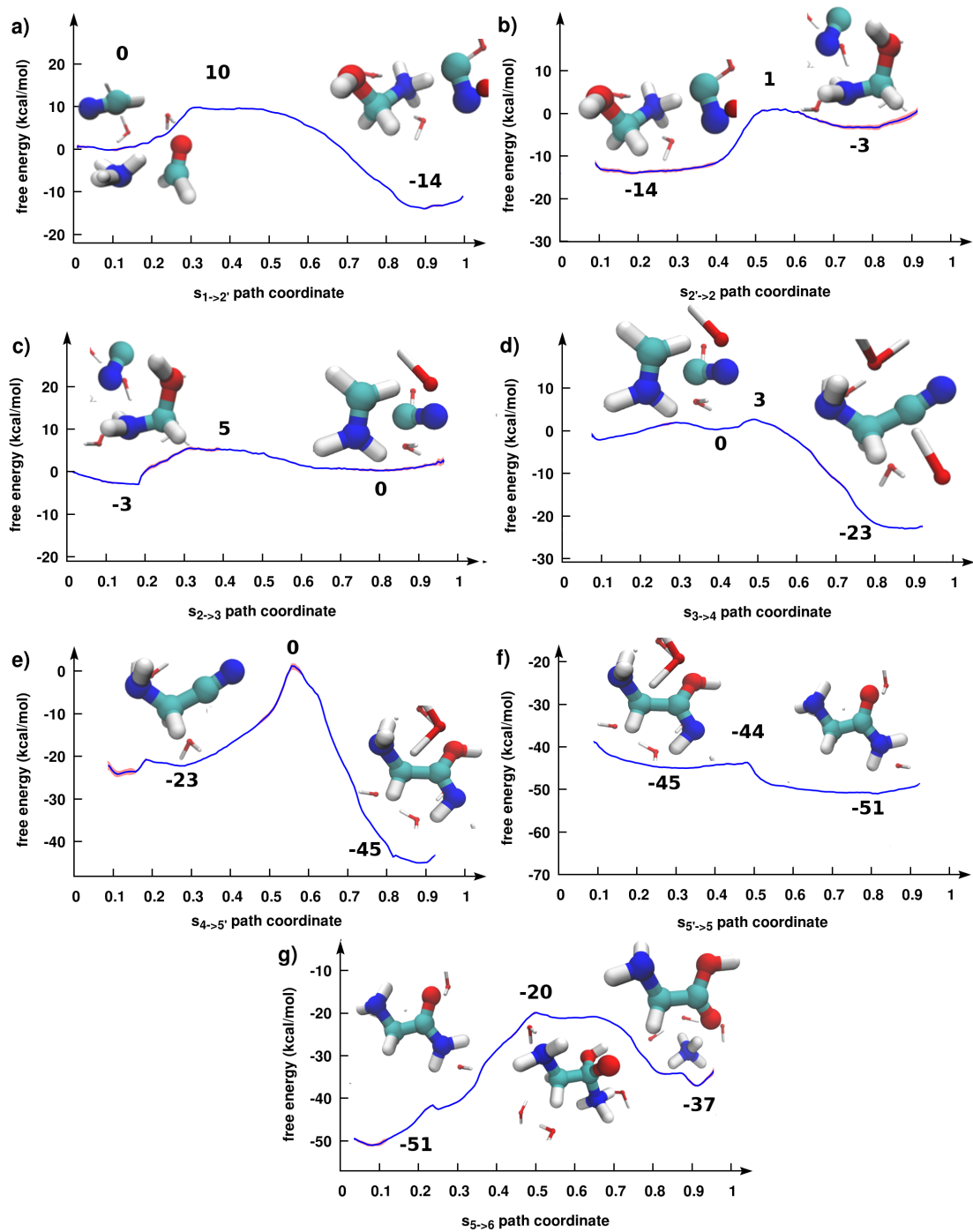
8

Figure 3: Umbrella sampling free energy profiles (blue) and error bars (red) for each elementary act. The tetrahedral intermediate ($\mathbf{TI}_{5-6}$) is also shown. Relative free energies are indicated with respect to the Strecker reactants. Error bars are estimated from block analysis.
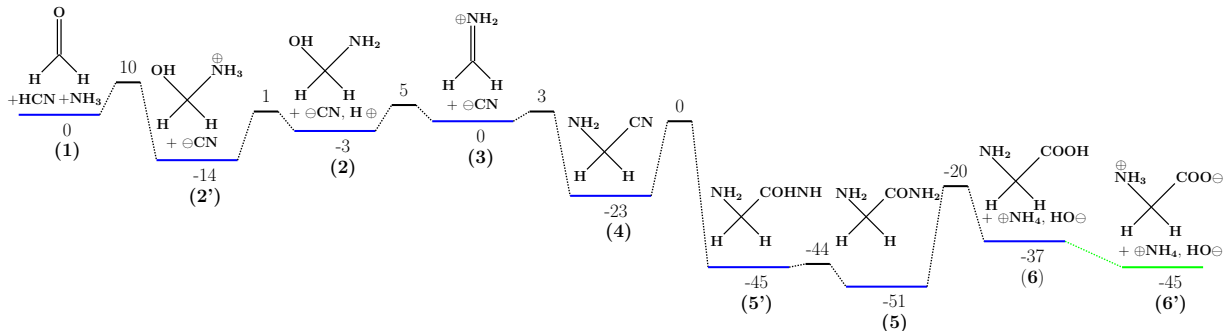
Figure 4: Free energy diagram of the Strecker reaction. Values next to metastable and transition states are our umbrella sampling free energies relatively to **(1)**, in kcal/mol, except for state **(6')**, that takes literature acido-basic equilibrium data into account, since simulations led to non-zwitterionic glycine and ammonium **(6)**.

Another relevant experimental difficulty in this reaction concerns the addition of ammonia to formaldehyde and the imine formation without hydrogen cyanide **(1)**→**(3)**. Ammonia and formaldehyde indeed spontaneously lead to a polycyclic oligomer, the so-called hexamine.[34,35] As a consequence, the authors of reference[35] gave only Taft-type estimate for kinetics of the bare ammonia to formaldehyde reaction.[21,22]

Another experimental work yields an Arrhenius activation energy of 4.5 kcal/mol for the low temperature aminomethanol formation on ice.[36] Furthermore, a recent computational study[34] adopting an implicit solvent approach studied the free energy differences and barriers for ammonia addition to formaldehyde (corresponding to **(1)**→**(2)** in this work) and the subsequent imine formation reaction (corresponding to **(2)**→**(3)**), calculating for the latter one a rate limiting barrier of 21.6 kcal/mol. This value is significantly worse than ours in comparison to experiments, thus confirming the importance of adopting an explicit solvent approach. For the second part of the glycinonitrile **(4)** formation, experimental kinetic data are provided in Ref.[13] The decomposition reaction $(NH_2)CH_2(CN) \rightarrow (NH_2^+)CH_2 + CN^-$, opposite of reaction **(3)**→**(4)**, has an experimental activation free enthalpy of 27 kcal/mol. Our computed value is 26 kcal/mol, in excellent agreement. Other data about cyanide addition to imine **(3)**→**(4)** rely upon gas-phase calculations or small size solvent clusters (three water molecules[37]). This reference proposes a rate limiting barrier of 35 kcal/mol for **(4)**→**(3)**,

again significantly worse than ours in comparison to experiments. Furthermore, this reference[37] has shown dramatic effect of explicit water molecules addition on the calculated activation barriers, supporting our fully explicit solvent description.

For the pathway leading from glycinonitrile to glycine, in the Strecker mechanism, the main experimental difficulties are twofold: (i) the high activation barriers hamper equilibrium uncatalyzed studies in neutral water, and (ii) the equilibrium is strongly displaced towards products, which makes the nitrile detection difficult.[38,39] The precise thermodynamics of nitrile hydration and of the subsequent amide hydrolysis is therefore an open question. It is generally considered that nitrile hydration is favoured thermodynamically and proceeds through a tautomeric amide intermediate (imidic acid),[13,38] and that the amide hydrolysis to carboxylic acid is thermodynamically disfavoured. Our results are consistent with this general knowledge (see steps **(4)→(5)** and **(5)→(6)** figure 4). Furthermore, for nitrile hydration to amide, the water addition step is known to be rate limiting compared to the tautomerisation step, also in agreement with our results (compare barriers **(4)→(5')** and **(5')→(5)** in figure 4). Again, since quantitative results are lacking, our study fills these important gaps in the literature. In particular, our explicit solvent approach is crucial to reveal, with respect to literature,[39] that the thermodynamically relevant sub-step of the nitrile hydration is the water addition and not the imidic acid/amide tautomeric equilibrium (compare free energy levels of species **(4)**, **(5)** and **(5')**).

Kinetic data, on the other hand, are available for glycinonitrile hydration from Ref.[13] glycinonitrile hydration has experimental activation free enthalpies of 29 kcal/mol and 20.5 kcal/mol, under acid limit mechanism or basic limit mechanism, respectively. Our calculated value is 23 kcal/mol, consistently with experiments. In fact, the theoretical transition state is characterised by a water molecule close to a unprotonated nitrile (see figure 2(a)), which is intermediate between the two literature limit mechanisms, and consistent with a barrier comprised between the two limiting values.

The glycinamide hydrolysis second-order kinetic parameters are also given in the case

of both acid limit mechanism or basic limit mechanism.[13,40] The acid-catalysed activation free enthalpy is 25 kcal/mol, while it is only 19.5 kcal/mol in the basic case. In our simulation of the elementary act $(\mathbf{5})\rightarrow(\mathbf{6})$, the mechanism corresponds to the acid-catalysed limit mechanism (addition of water and ammonia leaving group). The calculated barrier is 31 kcal/mol, that compares quite well with the 25 kcal/mol experimental result, within the typical PBE-DFT errors.[33]

The full comparison of our results to the few existing experimental thermodynamics data is summarized in table 1. We made the approximation, common in condensed phase chemistry, that PV terms in enthalpy undergo small variations compared to energy variations, due to the small molar volumes of condensed phases. Thus, (free) enthalpy *differences* can be approximately compared to (free) energy *differences*. The total cumulative error on the free energy difference between our calculations and the experimental data for the overall synthesis is 24 kcal/mol, which results from the error accumulation on the 7 elementary acts studied. If one examine these discrepancies in details, the free energy differences between two states reported in experimental papers are: $\Delta_r\mathrm{G}^{\ominus}((\mathbf{1})\rightarrow(\mathbf{2}))$=4 kcal/mol, $\Delta_r\mathrm{G}^{\ominus}((\mathbf{2})\rightarrow(\mathbf{3}))$=-5.5 kcal/mol, $\Delta_r\mathrm{G}^{\ominus}((\mathbf{1})\rightarrow(\mathbf{4}))$=-13/-16 kcal/mol.

The corresponding values herein computed are: $\Delta_r\mathrm{G}^{\ominus}((\mathbf{1})\rightarrow(\mathbf{2}))$=-3 kcal/mol, $\Delta_r\mathrm{G}^{\ominus}((\mathbf{2})\rightarrow(\mathbf{3}))$=3 kcal/mol, $\Delta_r\mathrm{G}^{\ominus}((\mathbf{1})\rightarrow(\mathbf{4}))$=-23 kcal/mol. In other words, these discrepancies are only respectively 7, 8, and 7/10 kcal/mol, which is reasonable (at most approximately two H bonds[41]), and largely within the typical PBE-DFT error on heat of formation, of about 10 kcal/mol.[33] The global -24 kcal/mol error on $\Delta_r\mathrm{G}^{\ominus}((\mathbf{1})\rightarrow(\mathbf{6}))$ can be explained by the PBE-DFT overestimation of this specific reaction enthalpy, by comparing database gas phase formation enthalpies from experiments[42] and PBE-DFT.[43] Using the best available basis sets of Pople-type and Dunning-type (resp. 6-311+G(3df,2p) and aug-cc-pVDZ) in database[43] for our system with the PBE functional, the glycine Strecker synthesis enthalpy reaction errors are respectively -22 and -19 kcal/mol in gas phase compared to experiments.[42] This intrinsic DFT discrepancy appears to explain most of the present work error.

We remark the importance of careful, systematic assessment of the reaction steps for the prebiotic synthesis of biological building blocks, since the complexity of chemical networks starting from small organic compounds grows quicker than exponentially with the number of consecutive reactions, as pointed out by a recent high impact experimental study of the amino acids synthesis in complex mixtures,[44] as well as recent screening study of prebiotic networks.[45] However, this kind of studies provide only plausible mechanistic networks and not their quantitative free energy characterisation. In this context, quantitative studies like the present one are needed to understand why some compounds have been selected, and which pathway could actually lead to them.

Table 1: For state (4) we indicate two values for its free energy difference with (1) since use of calculated $\Delta_r H_{aq}^{\ominus}$ and $\Delta_r S_{aq}^{\ominus}$ from[13] gives $\Delta_r G^{\ominus}((1){\to}(4))$=-13 kcal/mol, while direct extrapolation from temperatures close to 300 K in the same source gives $\Delta_r G^{\ominus}((1){\to}(4))$=-16 kcal/mol. Reference[46] retained -13 kcal/mol from.[13] We felt giving the experimental value most suitable to the temperature we used is simulations was relevant, and it gives an idea of the error that can be expected from literature data.

state	Thermodynamic data, with respect to state $\mathbf{1}$		
	G_{aq} (lit.)	F_{aq}	$\Delta(F_{aq}\text{-}G_{aq}$ (lit.))
(1)	0 (-105)^a	0	0
$(2')$	NA	-14	NA
(2)	4^b	-3	-7
(3)	-1^c	0	+1
(4)	-13/-16^d	-23	-10/-7
$(5')$	NA	-45	NA
(5)	NA	-51	NA
(6)	-13^e	-37	-24
$(6')$	-21^e	-45 (-8 w.r.t. $\mathbf{6}$)	idem

a^{42} , b^{13,23,24} , c^{13,21,22} , d^{13,32,46} , e$^{42,47-51}$

In this work we tackled a fully *ab initio* study of the end-to-end Strecker reaction in aqueous conditions, by using an explicit solvent description and a state-of-the-art approach for the exploration of chemical networks, the unsupervised identification of best pathways, and the accurate determination of relative free-energies, activation barriers, and transition states. In accomplishing this challenging task, we obtained results in very good agreement

with the scarce and sparse experimental data, typically focused on specific elementary steps of this complex, 7-step reaction, filling the experimental gaps. In particular, disclosing relative free energy values for all elementary acts in an explicitly simulated complex medium opens perspective in terms of comparison to actual prebiotic relevant systems. Similar simulations may allow for instance to interpret meteoritic abundances, where thermodynamic equilibrium can be reached. In this context, it is worth noting glycinotirile (**4**) has been detected in the Sgr B2(N) molecular gas cloud.[52] The identification of a compound such as amino-amide (**5**) as a thermodynamic trap of Strecker synthesis, could be considered another signature of Strecker Mechanism. Our study allows, for the first time, a global and at the same time accurate view of the fundamental Strecker reaction, which is widely regarded as a central tenet in prebiotic chemistry and origins of life studies.

Computational Methods

Born-Oppenheimer Molecular Dynamics (BOMD).

Ab initio molecular dynamics were based on the Density Functional Theory with Perdew-Burke-Ernzerhof functional and Grimme's van der Waals corrections,[53,54] and Martins-Troullier pseudopotentials[55] for C, N, O and H atoms (with a cutoff of 70 Ry and 10^{-5} convergence cutoff for orbitals), with the CPMD software.[56] BOMD simulations were performed in the canonical ensemble (NVT) at a temperature of 300 K , with a time step of 0.48 fs, in a periodically repeated cubic box of 13.4 Å, including 2 carbon, 2 nitrogen, 6 hydrogen, 1 oxygen atoms, and 81 water molecules. The density of the box is therefore 1.058 kg/L, and the concentration is 0.7 mol/L for initial formaldehyde, hydrogen cyanide or final glycine.

Enhanced sampling simulations.

Metadynamics and umbrella sampling were performed with the PLUMED free energy plugin.[57] The former approach was employed for discovery of pathways, the latter for precise

free energy profile determination. We used path-based collective variables s and z.[26] The s coordinate represents the progress along a putative pathway defined as a sequence reference structures. The z coordinate represents the distance from the putative pathway. For path-based collective variables definition, we used a topological metric based on coordination patterns, suited for chemical reactions in solution.[27–29]

Metadynamics simulations were performed with only two reference structures, corresponding to reactant and products. To identify transition states, we extracted a set of configurations from metadynamics trajectories close to the putative transition state. Starting from such configurations, we launched independent short unbiased BOMD trajectories and we verified whether they fell into reactants and products (committor analysis). A transition state is identified when approximately half of the trajectories are committed to both reactants and products.

To resolve more precisely the transition pathways, we used additional reference structures (12 in total including reactants and products), extracted from unbiased committor analysis trajectories, to define path-based collective variables for umbrella sampling. The latter technique was applied to each elementary reaction by restraining s at different values with parabolic potentials along reaction pathways. 30 to 60 windows were used for each elementary reaction.

Free energy profiles were eventually computed with two different codes, by A. Grossfield[58] and J. Goings[59] based respectively on Weighted Histogram Analysis Method and Multistate Bennett Acceptance Ratio. Statistical error bars were determined by cutting every umbrella sampling trajectory in four parts: the two last parts were used to determine free energy profiles, while the convergence error is estimated as the difference between these two profiles. Typical statistical errors are of about 1 kcal/mol, smaller than DFT error.

Each metadynamics simulation lasted about 30 ps, committor analysis trajectories 3 ps, and umbrella sampling trajectories 15 ps or more. The global cost for an elementary act is therefore dominated by umbrella sampling, corresponding to about 600 ps in the case of 40

windows. The cumulative time simulated in this work is of about 4 ns.

Acknowledgement

Supporting Information Available:

Supporting information are provided in a separate file.

SI figure S1: Schematic algorithm depicting the simulation protocol developed in this study.

SI figure S2: Illustration of the use of path collective variables together with the protocol developed in this study.

SI figure S4-S10: Umbrella sampling comprehensive results.

References

(1) Miller, S. L. A Production of Amino Acids Under Possible Primitive Earth Conditions. *Science* **1953**, *117*, 528–529.

(2) Miller, S. L. The mechanism of synthesis of amino acids by electric discharges. *Biochimica et Biophysica Acta* **1957**, *23*, 480–489.

(3) Bada, J. L.; Lazcano, A. Prebiotic soup–revisiting the Miller experiment. *Science* **2003**, *300*, 745–746.

(4) Pizzarello, S.; Cronin, J. Non-racemic amino acids in the Murray and Murchison meteorites. *Geochimica et Cosmochimica Acta* **2000**, *64*, 329–338.

(5) Walker, J. C. Carbon dioxide on the early Earth. *Origins of Life and Evolution of the Biosphere* **1985**, *16*, 117–127.

(6) Holland, H. D. *The chemistry of the atmosphere and oceans-(v. 1)*; New York, NY (USA) Wiley-Interscience, 1978.

(7) Johnson, A. P.; Cleaves, H. J.; Dworkin, J. P.; Glavin, D. P.; Lazcano, A.; Bada, J. L. The Miller volcanic spark discharge experiment. *Science* **2008**, *322*, 404–404.

(8) Bernstein, M. P.; Dworkin, J. P.; Sandford, S. A.; Cooper, G. W.; Allamandola, L. J. Racemic amino acids from the ultraviolet photolysis of interstellar ice analogues. *Nature* **2002**, *416*, 401–403.

(9) Sandford, S. A.; Nuevo, M.; Bera, P. P.; Lee, T. J. Prebiotic astrochemistry and the formation of molecules of astrobiological interest in interstellar clouds and protostellar disks. *Chemical reviews* **2020**, *120*, 4616–4659.

(10) Ioppolo, S.; Fedoseev, G.; Chuang, K.-J.; Cuppen, H.; Clements, A.; Jin, M.; Garrod, R.; Qasim, D.; Kofman, V.; van Dishoeck, E., *et al.* A non-energetic mechanism for glycine formation in the interstellar medium. *Nature Astronomy* **2021**, *5*, 197–205.

(11) Ferus, M.; Rimmer, P.; Cassone, G.; Knížek, A.; Civiš, S.; Šponer, J. E.; Ivanek, O.; Šponer, J.; Saeidfirozeh, H.; Kubelík, P., *et al.* One-pot hydrogen cyanide-based prebiotic synthesis of canonical nucleobases and glycine initiated by high-velocity impacts on early Earth. *Astrobiology* **2020**, *20*, 1476–1488.

(12) Strecker, A. Ueber einen neuen aus Aldehyd-Ammoniak und Blausäure entstehenden Körper. *European Journal of Organic Chemistry* **1854**, *91*, 349–351.

(13) van Trump, J. E. The Strecker Synthesis and Its Prebiological Importance. Ph.D. thesis, University of California, San Diego, 1975; UCSD permalink : `https://roger.ucsd.edu/record=b6916426~S9`.

(14) Koga, T.; Naraoka, H. A new family of extraterrestrial amino acids in the Murchison meteorite. *Scientific reports* **2017**, *7*, 1–8.

(15) Carrascoza Mayén, J. F.; Rydzewski, J.; Szostak, N.; Blazewicz, J.; Nowak, W. Prebiotic Soup Components Trapped in Montmorillonite Nanoclay Form New Molecules: Car-Parrinello Ab Initio Simulations. *Life* **2019**, *9*, 46.

(16) Saitta, A. M.; Saija, F. Miller experiments in atomistic computer simulations. *Proceedings of the National Academy of Sciences* **2014**, *111*, 13768–13773.

(17) Martins, Z.; Price, M. C.; Goldman, N.; Sephton, M. A.; Burchell, M. J. Shock synthesis of amino acids from impacting cometary and icy planet surface analogues. *Nature Geoscience* **2013**, *6*, 1045.

(18) Goldman, N.; Reed, E. J.; Fried, L. E.; Kuo, I.-F. W.; Maiti, A. Synthesis of glycine-containing complexes in impacts of comets on early Earth. *Nature Chemistry* **2010**, *2*, 949–954.

(19) Rimola, A.; Sodupe, M.; Ugliengo, P. Deep-space glycine formation via Strecker-type reactions activated by ice water dust mantles. A computational approach. *Physical Chemistry Chemical Physics* **2010**, *12*, 5285–5294.

(20) Rimola, A.; Sodupe, M.; Ugliengo, P. Computational study of interstellar glycine formation occurring at radical surfaces of water-ice dust particles. *The Astrophysical Journal* **2012**, *754*, 24.

(21) Ogata, Y.; Kawasaki, A. Kinetics of the condensation of some aliphatic aldehydes with ammonia. *Tetrahedron* **1964**, *20*, 1573–1578.

(22) Ogata, Y.; Kawasaki, A. Equilibrium additions to carbonyl compounds. *The Carbonyl Group (1970)* **1970**, *2*, 1–69.

(23) Feraud, P.; Le Henaff, P. Study of some condensation reactions of formol with amino acids. *Bulletin de la Societe chimique de France* **1968**, *5*, 1968–1978.

(24) Sander, E. G.; Jencks, W. P. Equilibria for additions to the carbonyl group. *Journal of the American Chemical Society* **1968**, *90*, 6154–6162.

(25) Pérez-Villa, A.; Pietrucci, F.; Saitta, A. M. Prebiotic chemistry and origins of life research with atomistic computer simulations. *Physics of life reviews* **2018**, *34-35*, 105–135.

(26) Branduardi, D.; Gervasio, F. L.; Parrinello, M. From A to B in free energy space. *The Journal of chemical physics* **2007**, *126*, 054103.

(27) Pietrucci, F.; Saitta, A. M. Formamide reaction network in gas phase and solution via a unified theoretical approach: Toward a reconciliation of different prebiotic scenarios. *Proceedings of the National Academy of Sciences* **2015**, *112*, 15030–15035.

(28) Pietrucci, F.; Aponte, J. C.; Starr, R.; Pérez-Villa, A.; Elsila, J. E.; Dworkin, J. P.; Saitta, A. M. Hydrothermal decomposition of amino acids and origins of prebiotic meteoritic organic compounds. *ACS Earth and Space Chemistry* **2018**, *2*, 588–598.

(29) Pérez-Villa, A.; Saitta, A. M.; Georgelin, T.; Lambert, J.-F.; Guyot, F.; Maurel, M.-C.; Pietrucci, F. Synthesis of RNA nucleotides in plausible prebiotic conditions from ab initio computer simulations. *The journal of physical chemistry letters* **2018**, *9*, 4981–4987.

(30) Cassone, G.; Sponer, J.; Sponer, J. E.; Pietrucci, F.; Saitta, A. M.; Saija, F. Synthesis of (d)-erythrose from glycolaldehyde aqueous solutions under electric field. *Chemical Communications* **2018**, *54*, 3211–3214.

(31) O'Connor, C. Acidic and basic amide hydrolysis. *Quarterly Reviews, Chemical Society* **1970**, *24*, 553–564.

(32) Zavitsas, A. A.; Coffiner, M.; Wiseman, T.; Zavitsas, L. R. Reversible hydration of formaldehyde. Thermodynamic parameters. *The Journal of Physical Chemistry* **1970**, *74*, 2746–2750, doi : `10.1021/j100708a003`.

(33) Riley, K. E.; Op't Holt, B. T.; Merz, K. M. Critical assessment of the performance of density functional methods for several atomic and molecular properties. *Journal of chemical theory and computation* **2007**, *3*, 407–433.

(34) Kua, J.; Rodriguez, A. A.; Marucci, L. A.; Galloway, M. M.; De Haan, D. O. Free energy map for the co-oligomerization of formaldehyde and ammonia. *The Journal of Physical Chemistry A* **2015**, *119*, 2122–2131.

(35) Ogata, Y.; Kawasaki, A. The kinetics of the reaction of formaldehyde with ammonia. *Bulletin of the Chemical Society of Japan* **1964**, *37*, 514–519.

(36) Bossa, J.; Theule, P.; Duvernay, F.; Chiavassa, T. NH2CH2OH thermal formation in interstellar ices contribution to the 5-8 μm region toward embedded protostars. *The Astrophysical Journal* **2009**, *707*, 1524.

(37) Koch, D. M.; Toubin, C.; Peslherbe, G. H.; Hynes, J. T. A theoretical study of the formation of the aminoacetonitrile precursor of glycine on icy grain mantles in the interstellar medium. *The Journal of Physical Chemistry C* **2008**, *112*, 2972–2980.

(38) Tewari, Y. B.; Goldberg, R. N. Thermodynamics of the hydrolysis reactions of nitriles. *The Journal of Chemical Thermodynamics* **2005**, *37*, 720–728.

(39) Guthrie, J. P.; Yim, J. C.-H.; Wang, Q. Hydration of nitriles: An examination in terms of no barrier theory. *Journal of Physical Organic Chemistry* **2014**, *27*, 27–37.

(40) Zavoianu, D.; Cavadia, E. Structure and reactivity of nitriles and thioamides-alkanine hydrolysis of some niriles and of corresponding thioamides. *Bulletin de la Societe chimique de France* **1971**, *9*, 3340.

(41) Markovitch, O.; Agmon, N. Structure and energetics of the hydronium hydration shells. *The Journal of Physical Chemistry A* **2007**, *111*, 2253–2256, doi : `10.1021/jp068960g`.

(42) Linstrom, P. J.; Mallard, W. G. The NIST Chemistry WebBook: A chemical data resource on the internet. *Journal of Chemical & Engineering Data* **2001**, *46*, 1059–1063, doi: `10.18434/T4D303`.

(43) Johnson III, R. D. Computational chemistry comparison and benchmark database. *NIST standard reference database* **2013**, *69*.

(44) Islam, S.; Bučar, D.-K.; Powner, M. W. Prebiotic selection and assembly of proteinogenic amino acids and natural nucleotides from complex mixtures. *Nature Chemistry* **2017**, *9*, 584.

(45) Wołos, A.; Roszak, R.; Żadło-Dobrowolska, A.; Beker, W.; Mikulak-Klucznik, B.; Spólnik, G.; Dygas, M.; Szymkuć, S.; Grzybowski, B. A. Synthetic connectivity, emergence, and self-regeneration in the network of prebiotic chemistry. *Science* **2020**, *369*.

(46) Moutou, G.; Taillades, J.; Bénefice-Malouet, S.; Commeyras, A.; Messina, G.; Mansani, R. Equilibrium of α-aminoacetonitrile formation from formaldehyde, hydrogen cyanide and ammonia in aqueous solution: Industrial and prebiotic significance. *Journal of Physical Organic Chemistry* **1995**, *8*, 721–730, doi : `10.1002/poc.610081105`.

(47) Smirnov, V.; Badelin, V. The enthalpies of solution and solvation of glycine in mixed water-formamide solvents at 298.15 K. *Russian journal of physical chemistry* **2006**, *80*, 357–360, doi : `10.1134/S0036024406030095`.

(48) Svec, H. J.; Clyde, D. D. Vapor Pressures of Some α-Amino Acids. *Journal of Chemical and Engineering Data* **1965**, *10*, 151–152, doi :`10.1021/je60025a024`.

(49) Takagi, S.; Chihara, H.; Seki, S. Vapor pressure of molecular crystals. XIII. Vapor pressure of α-glycine crystal. The energy of proton transfer. *Bulletin of the Chemical Society of Japan* **1959**, *32*, 84–88, doi :`10.1246/bcsj.32.84`.

(50) Locke, M. J.; McIver Jr, R. T. Effect of solvation on the acid/base properties of glycine. *Journal of the American Chemical Society* **1983**, *105*, 4226–4232, doi : `10.1021/ja00351a017`.

(51) Haberfield, P. What is the energy difference between H2NCH2CO2H and+ H3NCH2CO2-? *Journal of Chemical Education* **1980**, *57*, 346, doi : `10.1021/ed057p346`.

(52) Belloche, A.; Menten, K.; Comito, C.; Müller, H.; Schilke, P.; Ott, J.; Thorwirth, S.; Hieret, C. Detection of amino acetonitrile in Sgr B2 (N). *Astronomy & Astrophysics* **2008**, *492*, 769–773.

(53) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Physical review letters* **1996**, *77*, 3865.

(54) Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *Journal of computational chemistry* **2006**, *27*, 1787–1799.

(55) Troullier, N.; Martins, J. L. Efficient pseudopotentials for plane-wave calculations. *Physical review B* **1991**, *43*, 1993.

(56) Hutter, J.; Curioni, A. Car–Parrinello molecular dynamics on massively parallel computers. *ChemPhysChem* **2005**, *6*, 1788–1793.

(57) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A., *et al.* PLUMED: A portable plugin

22

for free-energy calculations with molecular dynamics. *Computer Physics Communications* **2009**, *180*, 1961–1972.

(58) Grossfield, A. *WHAM: the weighted histogram analysis method* **version 2.0.9.**,

(59) Goings, J. *WHAM, https://github.com/jjgoings/wham* **2018**,