

RESEARCH ARTICLE | OCTOBER 16 2023

## Amide isomerization pathways: Electronic and structural background of protonation- and deprotonation-mediated *cis-trans* interconversions

Ádám A. Kelemen ; András Perczel  ; Dániel Horváth ; Imre Jákli  



*J. Chem. Phys.* 159, 154301 (2023)

<https://doi.org/10.1063/5.0165772>



CrossMark

### Articles You May Be Interested In

A functional integral formalism for quantum spin systems

*J. Math. Phys.* (July 2008)

Modes selection in polymer mixtures undergoing phase separation by photochemical reactions

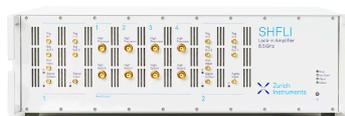
*Chaos* (June 1999)

Spreading of a surfactant monolayer on a thin liquid film: Onset and evolution of digitated structures

*Chaos* (March 1999)

500 kHz or 8.5 GHz?  
And all the ranges in between.

Lock-in Amplifiers for your periodic signal measurements



Find out more



# Amide isomerization pathways: Electronic and structural background of protonation- and deprotonation-mediated *cis-trans* interconversions

Cite as: J. Chem. Phys. 159, 154301 (2023); doi: 10.1063/5.0165772

Submitted: 30 June 2023 • Accepted: 28 September 2023 •

Published Online: 16 October 2023



View Online



Export Citation



CrossMark

Ádám A. Kelemen,<sup>1,2</sup>  András Perczel,<sup>1,3,a)</sup>  Dániel Horváth,<sup>1,3</sup>  and Imre Jákli<sup>1,3,a)</sup> 

## AFFILIATIONS

<sup>1</sup> HUN-REN-ELTE Protein Modelling Research Group, Pázmány Péter sétány 1/A, H-1117 Budapest, Hungary

<sup>2</sup> Hevesy György PhD School of Chemistry, Institute of Chemistry, Eötvös Loránd University, Pázmány Péter sétány 1/A, H-1117 Budapest, Hungary

<sup>3</sup> Laboratory of Structural Chemistry and Biology, Institute of Chemistry, Eötvös Loránd University, Pázmány Péter sétány 1/A, H-1117 Budapest, Hungary

<sup>a)</sup> Authors to whom correspondence should be addressed: [perczel.andras@ttk.elte.hu](mailto:perczel.andras@ttk.elte.hu) and [imre.jakli@ttk.elte.hu](mailto:imre.jakli@ttk.elte.hu)

## ABSTRACT

The *cis-trans* isomerization of amide bonds leads to wide range of structural and functional changes in proteins and can easily be the rate-limiting step in folding. The *trans* isomer is thermodynamically more stable than the *cis*, nevertheless the *cis* form can play a role in biopolymers' function. The molecular system of *N*-methylacetamide · 2H<sub>2</sub>O is complex enough to reveal energetics of the *cis-trans* isomerization at coupled cluster single-double and coupled cluster single-double and perturbative triple [CCSD(T)] levels of theory. The *cis-trans* isomerization cannot be oversimplified by a rotation along  $\omega$ , since this rotation is coupled with the *N*-atom pyramidal inversion, requesting the introduction of a second dihedral angle " $\alpha$ ." Full  $f(\omega, \alpha)$  potential energy surfaces of the different amide protonation states, critical points and isomerization reaction paths were determined, and the barriers of the neutral, *O*-protonated and *N*-deprotonated amides were found too high to allow *cis-trans* interconversion at room temperature: ~85, ~140, and ~110 kJ mol<sup>-1</sup>, respectively. For the *N*-protonated amide bond, the *cis* form ( $\omega = 0^\circ$ ) is a maximum rather than a minimum, and each  $\omega$  state is accessible for less than ~10 kJ mol<sup>-1</sup>. Here we outline a *cis-trans* isomerization pathway with a previously undescribed low energy transition state, which suggests that the proton is transferred from the more favorable *O*- to the *N*-protonation site with the aid of nearby water molecules, allowing the *trans* → *cis* transition to occur at an energy cost of ≤11.6 kJ mol<sup>-1</sup>. Our results help to explain why isomerase enzymes operate via protonated amide bonds and how *N*-protonation of the peptide bond occurs via *O*-protonation.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0165772>

## NOMENCLATURE

### Abbreviations

CPCM conductor-like polarizable continuum model  
IEFPCM integral equation formalism polarizable continuum model  
IRC intrinsic reaction coordinate  
PEC potential energy curve  
PES potential energy surface,  $\omega = [C^{\text{Me}}-C-N-C^{\text{NMe}}]$  ( $^\circ$ ) in the periodic range of  $-180^\circ < \omega < +180^\circ$ ,

$\alpha = -(180^\circ - [C'-C^{\text{NMe}}-N-H])$  ( $^\circ$ )  
RC reaction complex of the amide and the protonated or deprotonated form of water  
RT room temperature  
TS transition state  
Xxx-nonPro a dipeptide formed between amino acid Xxx and any residue but Pro  
Xxx-Pro a dipeptide formed between amino acid Xxx and Pro residues  
Xxx-Yyy a dipeptide formed by amino acid Xxx and Yyy residues

## INTRODUCTION

The extensive conjugation of the C=O double bond and the lone pair of the adjacent N-atom within an amide bond prevents *cis-trans* isomerization at room temperature. The activation energy required for isomerization between the *cis* and *trans* forms is  $\sim 80\text{--}100\text{ kJ mol}^{-1}$ . The hindered transition along  $\omega$ , the torsion angle formed by the  $(\text{Ca}^i\text{--C}^i\text{--N--Ca}^{i+1})$  atoms between the *cis* ( $\omega = 0^\circ$ ) and *trans* ( $\omega = \pm 180^\circ$ ) isomers gives rise to a tuned rigidity, an essential stabilizing element of the backbone structure of polypeptides and proteins. Nuclear magnetic resonance (NMR) data show that the *trans* isomer is more stable than the *cis* isomer for both *N*-methylacetamide ( $\Delta G_{\text{trans/cis}} \sim 8\text{ kJ mol}^{-1}$ ) and *N*-methylformamide ( $\Delta G_{\text{trans/cis}} \sim 6\text{ kJ mol}^{-1}$ ), leading to a prevalence of *trans* amide linkages at room temperature.<sup>1,2</sup>

The controlled isomerization of peptide bonds, and thus the accumulation of selected *cis* isomers, is related to protein folding and to misfolding.<sup>3</sup> Despite their rarer occurrence, *cis* peptide bonds of Xxx-nonPro amides can be found in functional proteins, unambiguously identified at or near their functionally important sites.<sup>4</sup> For example, in an  $\alpha$ -like neurotoxin of *Buthus martensii* Karsch (a scorpion found in East Asia), called BmK M7 (PDB: 1KV0, a dimer, consisting 66 residues), the Pro<sup>9</sup>-His<sup>10</sup> peptide bond can be either *cis*, or a mixture of *cis* and *trans* isomers.<sup>4</sup> Only the *cis* form of the Pro-His subunit, forming a type-II  $\beta$ -turn, can accommodate the H-bond donated by the C-terminal Ala<sup>66</sup> to form a compact 3D-fold. This 3D fold of the toxin, which contains a *cis* amide bond as a central element, may be evolutionarily conserved, as can be seen in the M1 structure of the neurotoxin BmK (PDB: 1SN1), too.<sup>4</sup> Concanavalin A (ConA, PDB: 1JBC) is a larger protein of 237 residues that binds a metal ion whose coordination is the driving force for the *trans*  $\rightarrow$  *cis* isomerization of the Ala<sup>207</sup>-Asp peptide bond, resulting in the locked state of ConA.<sup>5</sup> In wild-type ribonuclease T1, the Tyr-Pro<sup>39</sup> peptide bond forms a *cis* isomer and remains so even when Pro39 is replaced by Ala.<sup>6</sup> Non-proline *cis* amide bond can be found at the active site of *Staphylococcus aureus* PC1  $\beta$ -lactamase (PDB: 3BLM), between Glu<sup>166</sup>-Ile on its N-terminal  $\alpha$ -helix.<sup>7</sup>

As *cis-trans* isomerization of the Xxx-Pro and Xxx-nonPro bonds are often not spontaneous processes, aided by chaperones, they formation can be the rate-limiting steps of protein folding.<sup>3</sup> In the case of the peptidylprolyl *cis-trans* isomerase enzyme [PPIase (EC 5.2.1.8)], various reaction mechanisms have been proposed, with one of them suggesting an acid-base reaction.<sup>8</sup> Cyclophilin, a protein that also exhibits prolyl isomerase activity, is known to bind substrates in a Pro-turn (inverse  $\gamma$ -turn) conformation, in which an Arg residue is close to the targeted Pro of the substrate within the active site and can act as a proton donor during the catalysis of the isomerization reaction through an H-bond.<sup>9</sup> The *cis-trans* isomerization of nonPro peptide bonds can also be aided by isomerase enzymes, such as the non-prolyl *cis-trans* isomerase hsp70 chaperone DnaK.<sup>10</sup>

For the *cis-trans* isomerization of the amide bond the weakening of the conjugation is necessary, usually indicated by the non-planarity of the peptide bond. The pyramidalization of the N-atom can be enhanced by steric repulsion/restriction, conformational and electronic effects.<sup>11</sup> Several molecular switches based on amide *cis-trans* isomerization were developed. The acid mediated switching behavior of *N*-methyl-*N*-(2-pyridyl)benzamides was studied by x-ray

crystallography and <sup>1</sup>H NMR spectroscopy, where *cis*  $\rightarrow$  *trans* isomerization of the amides occurred upon addition of acid.<sup>12</sup> The *pH*-dependent reversible *cis-trans* isomerization propensity of various amide models were also studied<sup>13</sup> and experimental results can be found where nearby functional groups may fixate and/or protonate the amide oxygen to induce *cis-trans* isomerization.<sup>14-19</sup> Therefore, the interconversion of amide conformations can be modulated by acidic environment.<sup>14</sup>

In total, four electrons of a  $\pi$  bond and a lone pair of the N-atom are delocalized over the  $\sigma$ -bond of the O, C' and N-atoms. In a chemical point of view, *cis-trans* interconversion involves breaking this conjugated bond system, during the rotation of  $\omega$ . Rotation of  $\omega$  triggers significant structural changes, (i) as the hybridization state of the N-atom changes from  $sp^2$  to  $sp^3$ , (ii) the C'-N bond lengthens, while (iii) the C'-O bond length decreases due to the re-localization of the  $\pi$ -bond on the C'-O  $\sigma$ -bond. The dipole moment also varies strongly with  $\omega$  rotation, and a charge transfer along the C-N bond also changes the atomic monopoles, altering the electrostatic properties of the functional group. The energy barrier of *cis-trans* isomerization was first estimated by Ramachandran and Mitra to be  $\sim 85\text{ kJ mol}^{-1}$ .<sup>20</sup> As a system of high biological interest, the energetics of *cis-trans* isomerization has already been calculated using different models, such as formamide, *N*-methylformamide, *N,N*-dimethylformamide and *N*-formylglycinamide, at different levels of theory [e.g., HF/3-21G, HF/6-31G(d) and B3LYP/6-31G(d)].<sup>21</sup> The density functional theory (DFT) calculations [B3LYP/6-311++G(3df,3pd)] gave values between 81.7 and 109.1  $\text{kJ mol}^{-1}$ .<sup>22</sup> Using a higher level of theory (MP2/6-31G\* in vacuum), two distinct transition states (TSs), with barrier heights of 60.3 and 69.7  $\text{kJ mol}^{-1}$  were determined.<sup>23</sup> Similar  $\Delta G$  values were obtained from NMR measurements:  $62.8 < \Delta G < 83.7\text{ kJ mol}^{-1}$ .<sup>24</sup> Energy barriers of C'-N rotation for *N*-methylacetamide were measured by NMR at 333 K in water and found to be  $\Delta G = 79\text{ kJ mol}^{-1}$  for *cis*  $\rightarrow$  *trans* and  $\Delta G = 89\text{ kJ mol}^{-1}$  for *trans*  $\rightarrow$  *cis* interconversion.<sup>25</sup>

Based on the literature data, the energy barrier for *cis-trans* isomerization is high, but as this process takes part in protein folding and misfolding,<sup>3</sup> so there must be a lower energy route for this "reaction." Regarding the effect of protonation or deprotonation of the amide group, it is known from NMR spectroscopy that the NH exchange is highly *pH* sensitive and thus the isomerization rate is influenced by *pH*.<sup>26</sup> Thus, a coherent description of amide bond isomerization as a function of *pH*, explaining the details and energy costs of rotation, may resolve the problem of high energy barriers and reveal low energy pathways.

## AIMS

Our aim was to provide a quantitative description of the acid- and base-catalyzed isomerization reaction pathways of the amide bond and to discriminate between alternative reaction routes. The potential energy surfaces of the neutral, protonated and deprotonated *N*-methylacetamide model systems were determined, including the geometric and energetic properties of all their critical points and intrinsic reaction coordinates (IRCs). In our minimalistic model, both protonation and deprotonation will be modeled as water-assisted processes, the minimum number of explicit H<sub>2</sub>O

is determined in order to perform more realistic calculations and obtain a more reliable mechanism.

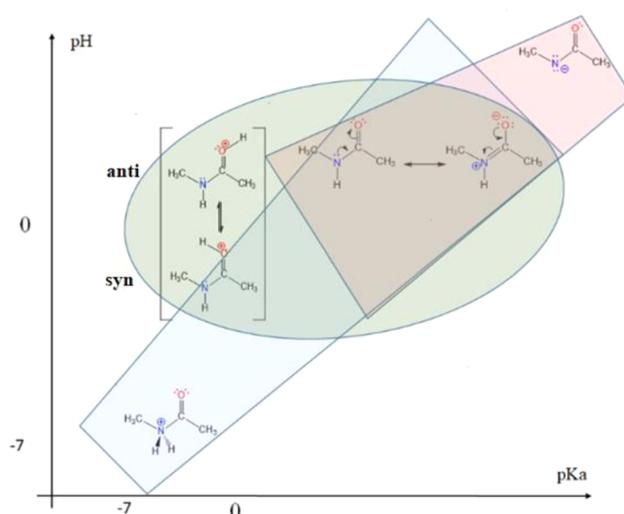
## METHODS

Both geometry optimizations, vibrational frequencies and energy calculations were performed at two different levels of theory (B3LYP<sup>27</sup> and M06-2X<sup>28</sup>) within the Gaussian 09 revision E.01<sup>29</sup> program, using the 6-311++G(d,p) basis set.<sup>30</sup> Following the optimizations, single point energy calculations were performed at the coupled cluster single-double (CCSD)<sup>31</sup> and coupled cluster single-double and perturbative triple [CCSD(T)]<sup>32</sup> levels of theory on structures obtained at the B3LYP/6-311++G(d,p) level, using the def2-TZVP<sup>33,34</sup> basis set, within the Orca 4.0.1 program.<sup>35</sup> Input files were created in GaussView 6.1<sup>36</sup> or Avogadro<sup>37</sup> programs. Calculations were performed using the water-solvent model implemented in the integral equation formalism polarizable continuum model (IEFPCM) method<sup>38</sup> for DFT calculations. For CCSD and CCSD(T) calculations the conductor-like polarizable continuum model (CPCM)<sup>39</sup> was used as the water-solvent model with  $\epsilon = 78.355$  300 and refractive index set to 1.777 849.

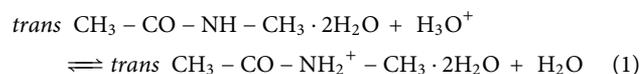
To study the *cis-trans* isomerization of the peptide bond, *N*-methylacetamide was used as the peptide bond model. As a first step, a relaxed scan was performed along the amide  $\omega$  ( $\text{Ca}^i\text{-C}^j\text{-N-Ca}^{i+1}$ ) torsion angle in 36 steps from  $\omega = -180^\circ$  to  $\omega = 180^\circ$  to assign the approximate positions of the critical points. Then energy optimization was performed on the selected geometries to determine the exact geometric and energetic properties of the critical points. In order to characterize the acid-base reaction steps that may facilitate isomerization, four different protonation states of *N*-methylacetamide were studied, in addition to the neutral, *N*-protonated, *N*-deprotonated and *anti*- and *syn*-*O*-protonated forms (Fig. 1, STables 2–27, SFigs. 2–27).

The PES  $f(\omega, \alpha)$  of neutral and *O*-protonated *N*-methylacetamide were determined by calculating a relaxed scan of dihedral angle  $\alpha$  [for definition see Fig. 2(a)] between  $-80^\circ$  and  $80^\circ$  and dihedral angle  $\omega$  between  $-180^\circ$  and  $180^\circ$  in  $5^\circ$  steps. The potential energy curve (PEC) of *N*-protonated and *N*-deprotonated *N*-methylacetamide were calculated in  $1^\circ$  steps along the dihedral angle  $\omega$ . To study the coordinates of the isomerization reaction, IRC paths<sup>41</sup> were calculated using the program Gaussian 09 Revision E.01, starting from the calculated geometries of the transition states (TSs) found on the PESs of neutral and *O*-protonated *N*-methylacetamide. Selected critical points on the PES of neutral *N*-methylacetamide were reoptimized at the CCSD/def2-TZVP level of theory.

In order to keep the number of H-bonds constant and to eliminate the effect of internal H-bonds to the total energy occurring within the model system during protonation or deprotonation reactions, two explicit water molecules in the desired protonated state were placed to all critical donor-acceptor interaction and proton transfer sites of *N*-methylacetamide (STables 43–81, SFigs. 28–64). Protonation and deprotonation processes are represented as a formal isodesmic reaction between the neutral *trans* *N*-methylacetamide  $\cdot 2\text{H}_2\text{O}$  complex + water molecule in the appropriate protonated form and the *N*-methylacetamide  $\cdot 2\text{H}_2\text{O}$  complex in the investigated protonation state and structural form + water molecule, e.g.,



**FIG. 1.** The one-to-one or 50%–50% ratio of the adjacent protonation states of *N*-methylacetamide occurs when  $\text{pH}$  equals  $\text{pKa}$ , based on the equation  $\text{pKa} = \text{pH} + \lg([\text{HA}]/[\text{A}^-])$ , as if  $[\text{HA}] \sim [\text{A}^-]$ , then  $\lg([\text{HA}]/[\text{A}^-]) = 0$ . Thus, at a  $\text{pH}$  of  $-7$ , the *N*-protonated ( $\text{pKa} = -7$ ) and neutral forms are in equilibrium (species within the blue trapezoid), whereas at  $\text{pH} = 0$ , the *O*-protonated forms (*syn* and *anti*;  $\text{pKa} = 0$ ) are in equilibrium with the neutral form of *N*-methylacetamide (molecular structures within the green oval). The *N*-deprotonation occurs in a very basic molecular environment, as the  $\text{pKa}$  is 18 (molecular structures within the red trapezoid).<sup>40</sup>

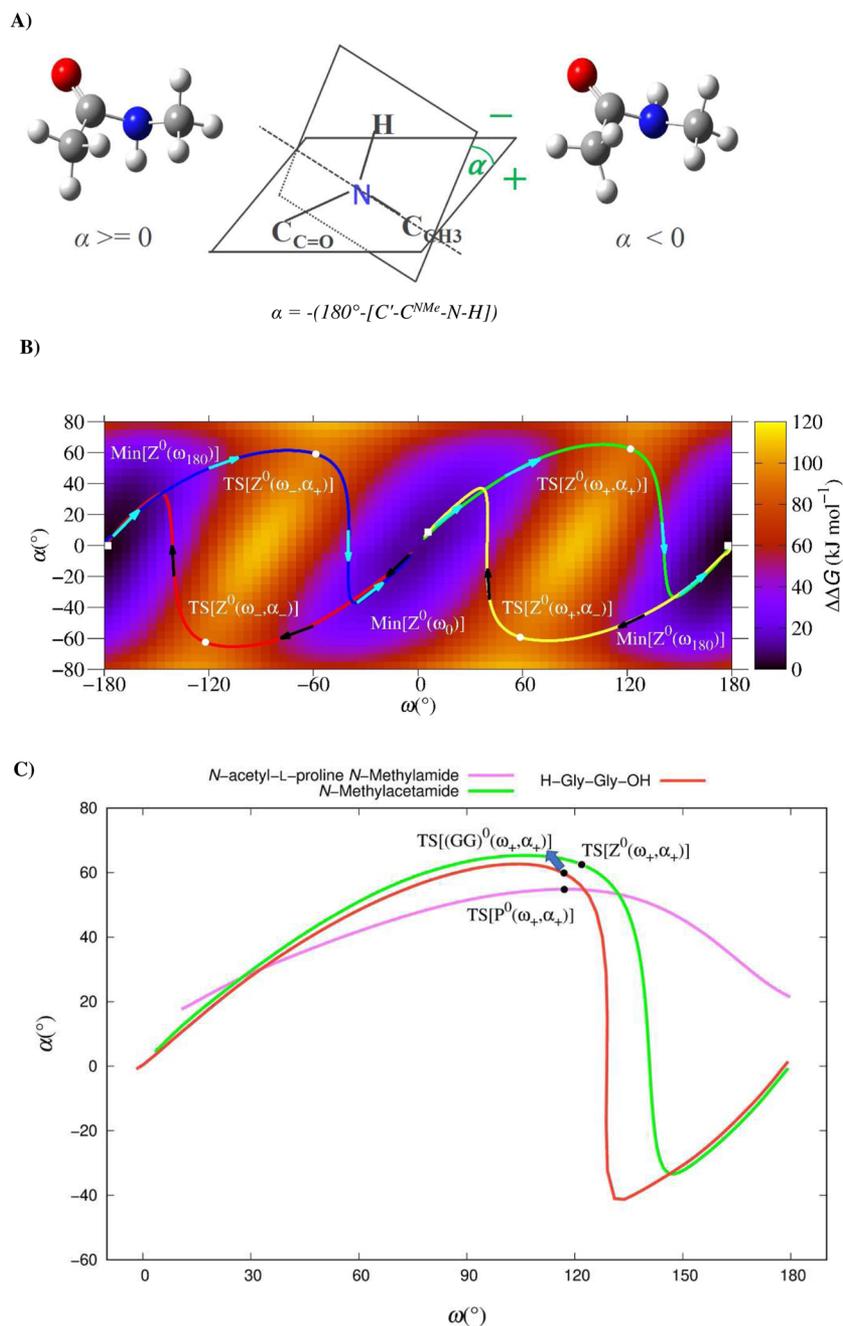


Critical points of the PECs and PESs were reoptimized and vibrational frequencies and energetic properties were recalculated at the B3LYP using the 6-311++G(d,p) TZP type basis set, followed by single point energy calculations at the CCSD/def2-TZVP (STables 42 and 99) and CCSD(T)/def2-TZVP levels of theory (Tables I–IV).

To compare the non-proline isomerization reaction with the *cis-trans* isomerization of proline containing peptide bonds, *N*-acetyl-L-proline *N*-methylamide was optimized in the *trans*, *cis* and in the isomerization transition states at the theoretical level B3LYP/6-311++G(d,p) in the IEFPCM water solvent model. (STables 109–111, SFigs. 73–75).

In order to be able to distinguish clearly between the different protonation states and the critical points, the molecular structures have been labelled with the following abbreviations and spellings (Scheme 1):

- (i) The leftmost label **X** indicates whether the molecular structure is a minimum (Min), a transition state (TS), a reaction complex (RC) (a transition state with low local energy barrier, however a key element in the reaction path) or a maximum (Max).
- (ii) Within the square bracket the protonation state is indicated:  $\text{Z}^0$  means zero-charge or neutral amide,  $\text{O}^{+s}$  stands for the *syn*-*O*-protonated, while  $\text{O}^{+a}$  stands for the *anti*-*O*-protonated *N*-methylacetamide.  $\text{N}^+$  stands for the *N*-protonated while  $\text{N}^-$  stands for the *N*-deprotonated



**FIG. 2.** (a) The  $\alpha = -(180^\circ - [C'-C^{NMe}-N-H])$  dihedral angle as defined to depict the pyramidal N-atom inversion. The extremes of this dihedral angle are around  $-60^\circ$  and  $60^\circ$  according to the IRC calculations. (b) The PES  $\sim f(\omega, \alpha)$  potential energy surface of *N*-methylacetamide [at theoretical level B3LYP/6-311++G(d,p), using the IEFPCM water solvent model] with the oriented IRC paths, depicting the route of peptide bond isomerization under neutral conditions. Critical points depicted are as follows: (i) the *trans* minimum;  $\text{Min}[Z^0(\omega_{180})] := (\pm 180.0^\circ, 0.0^\circ)$  and the *cis* minimum;  $\text{Min}[Z^0(\omega_0)] := (5.8^\circ, 8.7^\circ)$ , with the (ii) transition states:  $\text{TS}[Z^0(\omega_+, \alpha_+)] := (121.9^\circ, 62.5^\circ)$ ,  $\text{TS}[Z^0(\omega_-, \alpha_+)] := (-58.6^\circ, 59.2^\circ)$ ,  $\text{TS}[Z^0(\omega_-, \alpha_-)] := (-121.9^\circ, -62.5^\circ)$  and  $\text{TS}[Z^0(\omega_+, \alpha_-)] := (58.6^\circ, -59.2^\circ)$ . The “upper” IRC path connects the critical points  $\text{Min}[Z^0(\omega_{180})]$ ,  $\text{TS}[Z^0(\omega_-, \alpha_+)]$ ,  $\text{Min}[Z^0(\omega_0)]$ ,  $\text{TS}[Z^0(\omega_+, \alpha_+)]$  and  $\text{Min}[Z^0(\omega_{180})]$  denoted as “forward” isomerization route (follow the light blue arrows), while the “lower” IRC path, called as “backward” isomerization route connects  $\text{Min}[Z^0(\omega_{180})]$ ,  $\text{TS}[Z^0(\omega_+, \alpha_-)]$ ,  $\text{Min}[Z^0(\omega_0)]$ ,  $\text{TS}[Z^0(\omega_-, \alpha_-)]$  and  $\text{Min}[Z^0(\omega_{180})]$  (follow the black arrows). (c) The IRC paths ( $0^\circ \leq \omega \leq +180^\circ$ ), calculated from the TSs of the *cis-trans* isomerization of *N*-methylacetamide (green), H-Gly-Gly-OH (red) and *N*-acetyl-L-proline *N*-methylamide (magenta) models obtained at the B3LYP/6-311++G(d,p) level of theory. Note the pyramidal inversion of the N-atom in the case of *N*-methylacetamide and H-Gly-Gly-OH and its absence for the proline derivative.

**TABLE I.** Selected structural ( $\omega$  and  $\alpha$ ) and energetic properties ( $\Delta\Delta E$  in  $\text{kJ mol}^{-1}$ ) of the optimized minima (*cis* and *trans*), transition states and maxima optimized at CCSD(T)/def2-TZVP//B3LYP/6-311++G(d,p) level of theory (using CPCM water solvent model). The neutral ( $Z^0$ ) and the O-protonated forms ( $O^{+a}$  and  $O^{+s}$ ).

		TS[ $\omega_+, \alpha_+$ ]	TS[ $\omega_-, \alpha_+$ ]	TS[ $\omega_-, \alpha_-$ ]	TS[ $\omega_+, \alpha_-$ ]	$\omega_{180}$ ( <i>trans</i> )	$\omega_0$ ( <i>cis</i> ) <sup>a</sup>
$Z^0$	$\omega$ (deg)	+121.9	-58.6	-121.9	+58.6	-180.0	+5.8
	$\alpha$ (deg)	+62.5	+59.2	-62.5	-59.2	0.0	+8.7
	$\Delta\Delta E_{\text{trans}}$ <sup>b</sup>	80.8	83.5	80.8	83.5	0.0	11.1
$O^{+a}$	$\omega$ (deg)	+119.2	-60.2	-119.2	+60.2	+180.0	+0.1
	$\alpha$ (deg)	+57.5	+56.7	-57.5	-56.7	0.0	+0.3
	$\Delta\Delta E_{\text{trans}}$	155.1	155.9	155.0	155.9	0.0	9.3
$O^{+s}$	$\omega$ (deg)	+116.3	-61.1	-116.3	+61.1	-180.0	0.8
	$\alpha$ (deg)	+51.4	+56.8	-51.4	-56.8	0.0	-1.8
	$\Delta\Delta E_{\text{trans}}$	152.1	138.8	152.0	138.8	0.0	4.2

<sup>a</sup>Min[ $Z^0(\omega_0)$ ] is the *cis* isomer of the neutral *N*-methylacetamide, while Min[ $O^{+a}(\omega_0)$ ] and Min[ $O^{+s}(\omega_0)$ ] stand for the *cis* isomers of the O-protonated forms.

<sup>b</sup>Single point energies  $\Delta\Delta E$ s were calculated, using CPCM water solvent model ( $\epsilon = 78.355\ 300$  and refractive index 1.777 849) with respect to the more stable *trans* isomer.

**TABLE II.** Selected structural ( $\omega$  and  $\alpha$ ) and energetic properties ( $\Delta\Delta E$  in  $\text{kJ mol}^{-1}$ ) of the optimized minima (*cis* and *trans*), transition states and maxima optimized at CCSD(T)/def2-TZVP//B3LYP/6-311++G(d,p) level of theory (using CPCM water solvent model). The *N*-protonated form ( $N^+$ ).

		$\omega_+$	$\omega_-$	$\omega_{180}$ ( <i>trans</i> )	$\omega_0$ ( <i>cis</i> ) <sup>a</sup>
$N^+$	$\omega$ (deg)	+61.4	-62.7	-173.9	0.0
	$\alpha$ (deg)	-56.4	+56.3	+57.2	+57.6
	$\Delta\Delta E_{\text{trans}}$ <sup>b</sup>	5.2	5.2	0.0	8.9

<sup>a</sup>Min[ $N^+(\omega_0)$ ] is the *cis* isomer of the *N*-protonated *N*-methylacetamide.

<sup>b</sup>Single point energies  $\Delta\Delta E$ s were calculated, using CPCM water solvent model ( $\epsilon = 78.355\ 300$  and refractive index 1.777 849) with respect to the more stable *trans* isomer.

**TABLE III.** Selected structural ( $\omega$  and  $\alpha$ ) and energetic properties ( $\Delta\Delta E$  in  $\text{kJ mol}^{-1}$ ) of the optimized minima (*cis* and *trans*), transition states and maxima optimized at CCSD(T)/def2-TZVP//B3LYP/6-311++G(d,p) level of theory (using CPCM water solvent model). The *N*-deprotonated form ( $N^-$ ).

		$\omega_+$	$\omega_-$	$\omega_{180}$ ( <i>trans</i> )	$\omega_0$ ( <i>cis</i> ) <sup>a</sup>
$N^-$	$\omega$ (deg)	+94.5	-94.5	-180.0	0.0
	$\Delta\Delta E_{\text{trans}}$ <sup>b</sup>	120.9	120.9	0.0	15.2

<sup>a</sup>Min[ $N^-(\omega_0)$ ] is the *cis* isomer of the *N*-deprotonated *N*-methylacetamide.

<sup>b</sup>Single point energies  $\Delta\Delta E$ s were calculated, using CPCM water solvent model ( $\epsilon = 78.355\ 300$  and refractive index 1.777 849) with respect to the more stable *trans* isomer.

amide model system. Furthermore, (**GG**)<sup>0</sup> stands for the neutral H-Gly-Gly-OH dipeptide and **P**<sup>0</sup> stands for neutral *N*-acetyl-L-proline *N*-methylamide.

- (iii) If an H<sub>2</sub>O or H<sub>3</sub>O(+) is H-bonded to the parent amide model, it is labelled and distinguished by the lower index, whose formalism is as follows: **aq** indicates a neutral water molecule, **aq+** for a hydronium or hydroxonium ion, H<sub>3</sub>O(+), and **aq-** for the hydroxide ion, OH(-). Next to this index in brackets is the atom of the amide to which the solvent is H-bonded.

- (iv) A *trans* amide bond is abbreviated as  $\omega_{180}$ , while the *cis* bond is abbreviated as  $\omega_0$ . Furthermore,  $\omega_+$  denotes a positive  $\omega$  dihedral angle, while  $\omega_-$  denotes a negative  $\omega$  dihedral angle. Similarly,  $\alpha_+$  indicates positive  $\alpha$  dihedral angles, while  $\alpha_-$  indicates negative  $\alpha$  dihedral angles. In the case of *N*-protonated *N*-methylacetamide there are stable gauche states,  $\omega \approx \pm 60^\circ$ , labelled  $\omega_+$  and  $\omega_-$  respectively. As an example, the fully optimized molecular structure (Min) of *syn*-O-protonated *N*-methylacetamide ( $O^{+s}$ ) in its *trans* orientation ( $\omega_{180}$ ), complexed with a protonated H<sub>2</sub>O (aq+) bonded to the C=O group and a "normal" water molecule (aq) H-bonded to the NH of the amide, is abbreviated as: **Min**[ $O^{+s}_{\text{aq}+(O),\text{aq}(NH)}(\omega_{180})$ ]. For further examples, see STable 102.

Assuming a Boltzmann distribution, the *cis/trans* isomer ratios of the optimized conformers of *N*-methylacetamide were calculated at different levels of theory (Table V). To validate these calculations, the *cis/trans* isomer ratios were also determined by NMR spectroscopy. NMR measurements were performed on a sample containing 10% (v/v) *N*-methylacetamide in 10% (v/v) D<sub>2</sub>O/H<sub>2</sub>O. DSS was added as an internal proton reference to set the 0.0 ppm value and NaN<sub>3</sub> for sample stability. <sup>1</sup>H experiments were performed at 298 K on a Bruker Avance III 700 MHz spectrometer operating at 700.13 MHz for <sup>1</sup>H, equipped with z-gradient and a 5-mm Prodigy CryoProbe head. 1D-<sup>1</sup>H (*zgesgp*) and 2D-spectra, <sup>1</sup>H-<sup>1</sup>H NOESY (*noesygpph19*), were recorded, using 64 and 32 scans respectively. The NOESY mixing time (*d8*) was set to 500 ms. All spectra were processed using TopSpin 4.0.7. Protons were assigned based on NOE cross-peaks and the *cis/trans* isomer ratio was determined by using the integrals of the <sup>1</sup>H resonances.

To determine the *cis/trans* isomer ratio of non-proline peptide bonds (Xxx-Yyy, where Yyy: non-Pro) among stable proteins thus far studied experimentally, 3760 chains containing a total of 146 375 residues were extracted from the PDB.<sup>42</sup> In this pre-filtered PISCES dataset of 2020,<sup>43</sup> all protein structures have a resolution better than 1.6 Å with R factors better than 0.25 to ensure that the correct isomer forms were assigned, and primary sequence homology less than

**TABLE IV.** The  $\omega$  and  $\alpha$  dihedral angles of the *trans*, *cis*, TSs and RCs of *N*-methylacetamide optimized with two water molecules. DFT optimized molecular complexes optimized at B3LYP/6-311++G(d,p) were used for  $\Delta\Delta E(\text{CCSD(T)/def2-TZVP})$  single point energy calculations. The  $\Delta\Delta E$  ( $\text{kJ mol}^{-1}$ ) values of minima and TSs of the (a) neutral, (b) *anti*-*O*-protonated, (c) *syn*-*O*-protonated, (d) *N*-protonated, (e) selected reaction complexes and (f) *N*-deprotonated forms were determined.

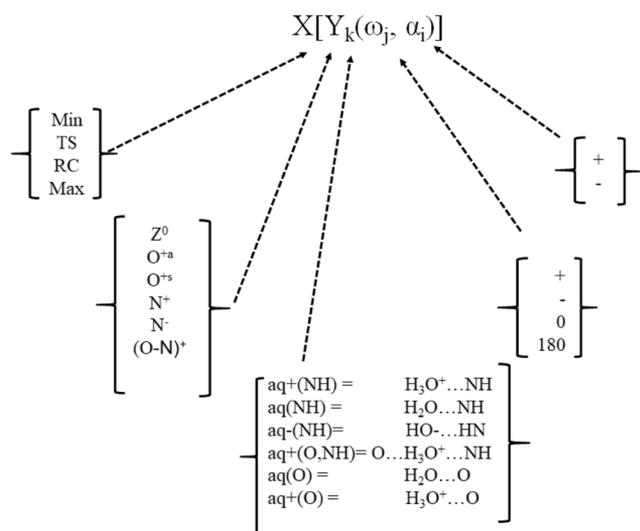
N-methylacetamide neutral forms	$\omega$ (deg)	(a)		$\Delta\Delta E(\text{CCSD(T)/def2-TZVP//B3LYP/6-311++G(d,p)})$
		$\alpha$ (deg)		
Min[ $Z^0_{\text{aq(O),aq(NH)}}(\omega_{180})$ ]	+179.9	-0.4		0.0
Min[ $Z^0_{\text{aq(O),aq(NH)}}(\omega_0)$ ]	-1.1	-1.2		10.4 <sup>a</sup>
TS[ $Z^0_{\text{aq(O),aq(NH)}}(\omega_+, \alpha_+)$ ]	+122.3	+61.6		76.5
TS[ $Z^0_{\text{aq(O),aq(NH)}}(\omega_-, \alpha_+)$ ]	-60.3	+58.9		91.3
TS[ $Z^0_{\text{aq(O),aq(NH)}}(\omega_-, \alpha_-)$ ]	-122.3	-61.6		76.6
TS[ $Z^0_{\text{aq(O),aq(NH)}}(\omega_+, \alpha_-)$ ]	+60.2	-59.0		91.3
N-methylacetamide <i>anti</i> - <i>O</i> -protonated forms	$\omega$ (deg)	(b)		$\Delta\Delta E(\text{CCSD(T)/def2.TZVP//B3LYP/6-311++G(d,p)})$
		$\alpha$ (deg)		
Min[ $O^{+a}_{\text{aq(O),aq(NH)}}(\omega_{180})$ ]	+179.8	-0.1		-95.9 <sup>a</sup>
Min[ $O^{+a}_{\text{aq(O),aq(NH)}}(\omega_0)$ ]	-0.9	-0.6		-84.2
TS[ $O^{+a}_{\text{aq(O),aq(NH)}}(\omega_+, \alpha_+)$ ]	+119.1	+59.2		50.1
TS[ $O^{+a}_{\text{aq(O),aq(NH)}}(\omega_-, \alpha_+)$ ]	-61.3	+58.1		53.1
TS[ $O^{+a}_{\text{aq(O),aq(NH)}}(\omega_-, \alpha_-)$ ]	-119.1	-59.2		50.0
TS[ $O^{+a}_{\text{aq(O),aq(NH)}}(\omega_+, \alpha_-)$ ]	+61.3	-58.1		51.5
N-methylacetamide <i>syn</i> - <i>O</i> -protonated forms	$\omega$ (deg)	(c)		$\Delta\Delta E(\text{CCSD(T)/def2.TZVP//B3LYP/6-311++G(d,p)})$
		$\alpha$ (deg)		
Min[ $O^{+s}_{\text{aq(O),aq(NH)}}(\omega_{180})$ ]	+179.4	-0.5		-83.1 <sup>a</sup>
Min[ $O^{+s}_{\text{aq(O),aq(NH)}}(\omega_0)$ ]	-0.7	-1.3		-85.5
TS[ $O^{+s}_{\text{aq(O),aq(NH)}}(\omega_+, \alpha_+)$ ]	+119.8	+55.0		52.3
TS[ $O^{+s}_{\text{aq(O),aq(NH)}}(\omega_-, \alpha_+)$ ]	-62.1	+58.2		43.2
TS[ $O^{+s}_{\text{aq(O),aq(NH)}}(\omega_-, \alpha_-)$ ]	-119.8	-55.0		52.1
TS[ $O^{+s}_{\text{aq(O),aq(NH)}}(\omega_+, \alpha_-)$ ]	+62.1	-58.2		44.0
N-methylacetamide <i>N</i> -protonated forms	$\omega$ (deg)	(d)		$\Delta\Delta E(\text{CCSD(T)/def2.TZVP//B3LYP/6-311++G(d,p)})$
		$\alpha$ (deg)		
Min[ $N^+_{\text{aq(O),aq(NH)}}(\omega_{180})$ ]	+179.7	+58.1		-56.6 <sup>a</sup>
Min[ $N^+_{\text{aq(O),aq(NH)}}(\omega_-)$ ]	-42.1	+57.8		-47.4
Min[ $N^+_{\text{aq(O),aq(NH)}}(\omega_+)$ ]	+42.8	+58.1		-48.2
Max[ $N^+_{\text{aq(O),aq(NH)}}(\omega_0)$ ]	+2.4	+57.2		-45.8
N-methylacetamide reaction complexes	$\omega$ (deg)	(e)		$\Delta\Delta E(\text{CCSD(T)/def2.TZVP//B3LYP/6-311++G(d,p)})$
		$\alpha$ (deg)		
RC[ $Z^0_{\text{aq+(O),aq(NH)}}(\omega_{180})$ ]	+180.0	-0.3		-3.5 <sup>a</sup>
RC[ $O^+_{\text{aq+(O),aq(NH)}}(\omega_0)$ ]	+0.1	+0.1		9.1
RC[ $Z^0_{\text{aq(O),aq-(NH)}}(\omega_{180})$ ]	-179.7	+1.0		-3.7
RC[ $N^-_{\text{aq(O),aq-(NH)}}(\omega_0)$ ]	+0.0	+0.2		12.9
RC[ $(N-O)^+_{\text{aq(O),aq(NH)}}$ ]	+47.1	-50.6		161.4
RC[ $O^{+s}_{\text{aq+(O),aq(NH),aq(NH)}}(\omega_{180})$ ]	-139.6	+41.5		11.6
RC[ $N^+_{\text{aq+(O),aq(NH),aq(NH)}}(\omega_0)$ ]	+1.3	+46.2		9.4
N-methylacetamide <i>N</i> -deprotonated forms	$\omega$ (deg)	(f)		$\Delta\Delta E(\text{CCSD(T)/def2.TZVP//B3LYP/6-311++G(d,p)})$
		$\alpha$ (deg)		
Min[ $Z^0_{\text{aq(O),aq-(NH)}}(\omega_{180})$ ]	+179.6	-0.3		0.0 <sup>b</sup>
Min[ $Z^0_{\text{aq(O),aq-(NH)}}(\omega_0)$ ]	+0.1	0.0		16.3
TS[ $Z^0_{\text{aq(O),aq-(NH)}}(\omega_+, \alpha_+)$ ]	+117.9	+61.1		64.7

TABLE IV. (Continued.)

<i>N</i> -methylacetamide <i>N</i> -deprotonated forms	$\omega$ (deg)	$\alpha$ (deg)	$\Delta\Delta E(\text{CCSD(T)/def2.TZVP//B3LYP/6-311++G(d,p)})$
$\text{TS}[Z^0_{\text{aq(O),aq-(NH)}}(\omega_-, \alpha_+)]$	-66.6	+59.6	98.1
$\text{TS}[Z^0_{\text{aq(O),aq-(NH)}}(\omega_-, \alpha_-)]$	-114.9	-60.7	100.4
$\text{TS}[Z^0_{\text{aq(O),aq-(NH)}}(\omega_+, \alpha_-)]$	+66.6	+59.6	100.2
$\text{Min}[N^-_{\text{aq(O),aq(NH)}}(\omega_{180})]$	+179.8	n.a.	-29.9
$\text{Min}[N^-_{\text{aq(O),aq(NH)}}(\omega_0)]$	0.0	n.a.	-23.3
$\text{Max}[N^-_{\text{aq(O),aq(NH)}}(\omega_+)]$	+99.3	n.a.	80.5
$\text{Max}[N^-_{\text{aq(O),aq(NH)}}(\omega_-)]$	-90.5	n.a.	88.4

<sup>a</sup>All  $\Delta\Delta E$  values are with respect to the  $\text{Min}[Z^0_{\text{aq(O),aq(NH)}}(\omega_{180})]$ , taking the isodesmic reaction into consideration.

<sup>b</sup>All  $\Delta\Delta E$  values are with respect to the  $\text{Min}[Z^0_{\text{aq(O),aq(NH)}}(\omega_{180})]$ , taking the isodesmic reaction into consideration.



**SCHEME 1.** Coding scheme of the minima (Min), transition states (TS), reaction coordinates (RC) and maxima (Max) of the neutral, differently protonated and deprotonated *N*-methylacetamide structures (see text for more details).

**TABLE V.** The  $\Delta\Delta G$  differences between the *cis* and *trans* isomers<sup>a</sup> of the optimized *N*-methylacetamide as calculated at different theoretical levels.

Method <sup>b</sup>	$\Delta\Delta G^{\text{cis/trans}}$ (kJ mol <sup>-1</sup> )	<i>cis/trans</i> ratio <sup>c</sup>
M06-2X/6-311++G(d,p)	3.7	1:4
B3LYP/6-311++G(d,p)	7.9	1:24
CCSD/def2-TZVP	10.9	1:81
NMR-spectroscopy	10.5	1:70

<sup>a</sup>The fully optimized two conformers are  $\text{Min}[Z^0(\omega_{180})]$  and  $\text{Min}[Z^0(\omega_0)]$ .

<sup>b</sup>The IEFFCM implicit water solvent model was applied at all theoretical levels.

<sup>c</sup>The derived *cis/trans* ratios assuming a Boltzmann-distribution at 298 K.

20%. To compare these statistical results with the *cis/trans* ratio derived from QM calculated energies, the DFT energies of H-Gly-Gly-OH dipeptide were also calculated at B3LYP/6-311++G(d,p) level of theory (STables 103–108, S Figs 67–72).

## RESULTS

### Statistical analysis of *cis-trans* isomerization of non-proline peptides

The residue-independent average *cis:trans* ratio determined from the PDB database is  $\sim 0.52\%$ , a value that shows large variations as a function of the chemical and topological properties of the side chain. This ratio is low for Ile, but 2–3 times higher than the calculated average for Gly and Trp [S Fig. 1(a)]. For non-Pro peptide bonds, the 1.18% found for Yyy=Gly is the highest *cis:trans* ratio, while the lowest one is that of Ile: 0.15%. Understandably, we found that Yyy residues with a large hydrophobic (Ile, Leu, Val), branched (Thr, Ile) or long (Lys, Arg) sidechain lower the *cis:trans* isomer ratio, as a *cis* amide form is sterically more compact.

In proteins, as expected, the highest number of *cis* amide bonds precede the Pro residue [S Fig. 1(b)], the average value of which is about one hundred times higher than that of Yyy=nonPro. There are significantly more Xxx-Pro *cis* amide bonds than those observed for Pro-Yyy dipeptides [S Fig. 1(c)]. However, it is worth noting that for nine residues we could not identify any *cis* amide bonds within the Pro-Yyy framework, whereas many of their palindromic forms (Xxx-Pro) embed several *cis* amide bonds. For example, the Asn-Pro amide bond has a *cis/trans* ratio of 53/1000, whereas not a single Pro-Asn *cis* amide bond was found in this highly filtered database.

The tabulation of the *cis:trans* ratio for all 20 proteinogenic amino acid pairs revealed that the most flexible Gly-Gly dipeptides contain the highest number of *cis* amide bonds between residues: a total of 24 *cis* bonds were assigned among the 4347 Gly-Gly pairs observed in proteins, corresponding to 5.6% (STable 1). Although a comparable number of Gly-Ala, Ala-Gly and Ala-Ala dipeptides were also found (4471, 4929, and 6323), surprisingly low numbers of *cis* amide bonds were assigned to the latter residue pairs, namely 0.9%, 0.6%, and 0.6%, respectively. This indicates

that even the shortest methyl side chain group of Ala introduces steric hindrance and thus, greatly reduces the occurrence of the *cis* form.

### The *cis-trans* isomerization of the peptide models *N*-methylacetamide, H-Gly-Gly-OH and *N*-acetyl-L-proline *N*-methylamide

In order to determine and characterize all critical points of the *cis-trans* isomerization potential energy curve, *N*-methylacetamide was first mapped using an implicit water-solvent model (IEFPCM). A relaxed scan calculation was performed over  $\omega = [C^{Me}-C'-N-C^{NMe}]$ ;  $-180^\circ \leq \omega \leq +180^\circ$  incremented by  $5^\circ$ . However, as the PEC  $\sim f(\omega)$  is not a continuous function due to the pyramidal inversion of the amide N-atom, a dihedral angle  $\alpha$  had to be introduced, defined as  $-(180^\circ - [C'-C^{NMe}-N-H])$ . Values of  $\alpha$  are shifted into the range  $[-80^\circ, +80^\circ]$  due to periodicity. Although alternative definitions exist to describe the pyramidal-ity of the N-atom,<sup>21,44</sup> they are often coupled to the dihedral angle  $\omega$ , whereas the one used here [Fig. 2(a)] is based solely on the spatial arrangement of the C', C<sup>NMe</sup>, and H-atoms around the pyramidal N-atom.

The N-atom pyramidal inversion was observed during the *cis-trans* isomerization of the more complex H-Gly-Gly-OH dipeptide model [Fig. 2(c)]. However, in the case of the *N*-acetyl-L-proline *N*-methylamide (SFIGs. 73–75) this inversion is absent due to the pyrrolidine ring of Pro. Moreover, a non-zero pyramidal-ity is observed for both the *trans* and *cis* proline peptide bonds:  $\alpha_{trans} = 16.2^\circ$  and  $\alpha_{cis} = 17.0^\circ$  at B3LYP/6-311++G(d,p), respectively. This pyramidal-ity of an amide system introduces a weaker amide bond conjugation, which may favor an easier isomerization for Pro.

The neutral PES of *N*-methylacetamide was calculated in  $5^\circ$  steps  $[-180^\circ \leq \omega \leq +180^\circ, -80^\circ \leq \alpha \leq +80^\circ]$  [Fig. 2(b)]. Topological analysis of the PES  $\sim f(\omega, \alpha)$  shows that there are 4 TSs/period within the studied interval, complementing previous studies that only mention 2 TSs[16]. The PES has a  $C_2$  mirror image symmetry, with four transition states labelled as  $TS[Z^0(\omega_+, \alpha_+)]$ ,  $TS[Z^0(\omega_-, \alpha_+)]$ ,  $TS[Z^0(\omega_+, \alpha_-)]$  and  $TS[Z^0(\omega_-, \alpha_-)]$  which are distinguished by the sign of the dihedral angle  $\omega$  and  $\alpha$ . Pairs of mirror image TSs,  $TS[Z^0(\omega_+, \alpha_+)] \leftrightarrow TS[Z^0(\omega_-, \alpha_-)]$ , and  $TS[Z^0(\omega_-, \alpha_+)] \leftrightarrow TS[Z^0(\omega_+, \alpha_-)]$ , have the same relative Gibbs free energies ( $\Delta\Delta G$ ): 85.1 and 87.8 kJ mol<sup>-1</sup>, respectively using the B3LYP functional and 79.5 and 82.4 kJ mol<sup>-1</sup>, respectively using the M06-2X functional using basis set 6-311++G(d,p). (STables 29 and 35, respectively) These stability data show that both the *cis* and *trans* minima are connected by four approximately equally high isomerization TSs, resulting that spontaneous isomerization cannot occur at room temperature.

The  $\Delta\Delta G$  differences between the *cis* and *trans* isomers were determined at the different theoretical levels (DFT and CCSD), and the relative ratios were calculated assuming a Boltzmann distribution at 298 K (Table V).

The currently calculated energy barriers can be compared with those measured by NMR and published previously.<sup>22,24</sup> While the DFT stabilities are slightly underestimated, the CCSD/def2-TZVP data agree with the values derived from our NMR based results. We found that the  $\Delta\Delta G$  values calculated at the CCSD/def2-TZVP

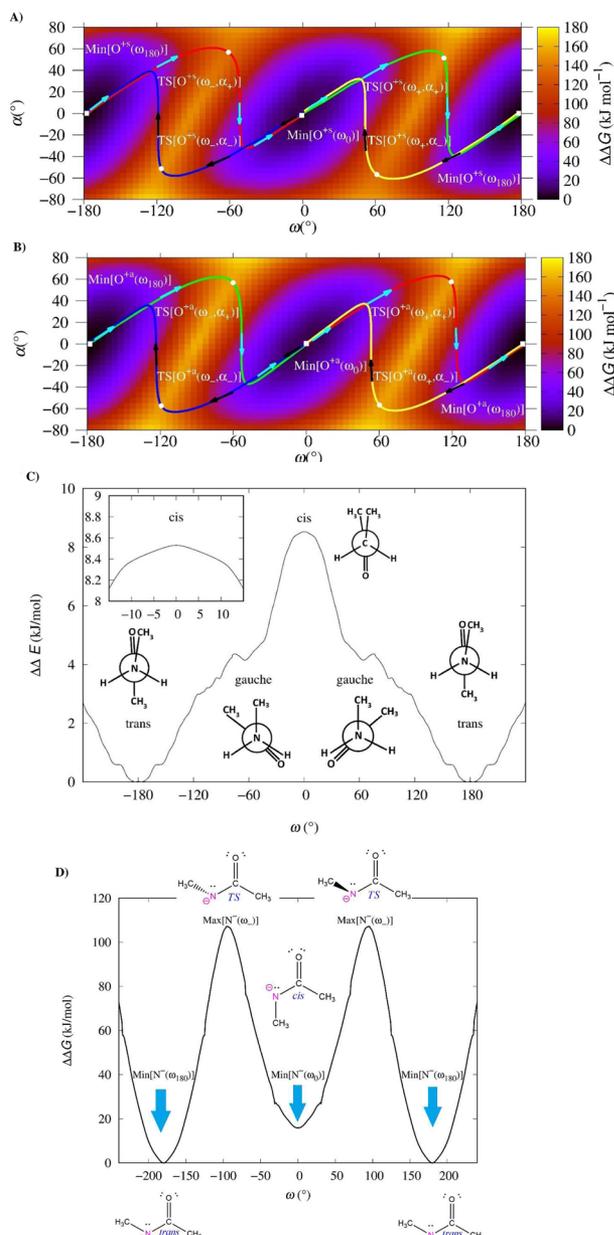
and B3LYP/6-311++G(d,p) levels of theory show a significant correlation, as the Pearson correlation coefficient is  $\approx 0.999$ . The same Pearson's  $r$  values were calculated between  $\Delta\Delta G$  values at theoretical levels B3LYP/6-311++G(d,p) and M06-2X/6-311++G(d,p). Furthermore, the root-mean-square deviations (RMSDs) between the fully optimized B3LYP and CCSD structures are small: e.g., 0.046 Å for  $\text{Min}[Z^0(\omega_{180})]$ , 0.06 Å for  $\text{Min}[Z^0(\omega_0)]$  and 0.02 Å for  $\text{TS}[Z^0(\omega_+, \alpha_+)]$  respectively, indicating that the structural properties of the conformer hardly change. In the case of  $\text{Min}[Z^0(\omega_{180})]$ , the slightly higher RMSD value (STable 41) is due to the different orientation of the methyl groups, which is irrelevant in our view as the amide bonds are very similar. The RMSD values between fully optimized B3LYP and M06-2X structures of *N*-methylacetamide are in the narrower range [0.01, 0.05 Å] (STable 40). In conclusion, the low RMSD values between the optimized structures show that geometries obtained at B3LYP level are relevant and can be used for later higher-level calculations. Moreover, the significant correlation between the stability data show that even B3LYP relative energy differences are reliable, and, with the use of a suitable offset value, the data are in agreement with each other.

The QM calculated and NMR derived *cis/trans* ratios (1:81 and 1:70) of *N*-methylacetamide (Table V) are about 24–28 times larger than those we found and calculated for proteins using PDB data: 1:1930 or  $\sim 0.518\%$  (STable 1). This large difference could be due to the fact that the *N*-methylacetamide model cannot take into account the steric effect of the amino acid side chains. However, if we consider only the *cis/trans* ratio of H-Gly-Gly-OH, then the calculated  $\Delta\Delta G_{cis/trans}$  value is 12.5 kJ mol<sup>-1</sup> at B3LYP/6-311++G(d,p) level of theory, which converts to an isomer ratio of 1:180, almost perfectly matching the ratio obtained from the PDB data: *cis:trans*<sub>Gly-Gly</sub> = 1:181 = 5.6‰ (STable 1). Furthermore, the *cis-trans* isomerization TSs were optimized at B3LYP/6-311++G(d,p) for H-Gly-Gly-OH, giving the following values: 87.9 kJ mol<sup>-1</sup> ( $\text{TS}[(\text{GG})^0(\omega_-, \alpha_+)]$ , SFIG. 71), 79.0 kJ mol<sup>-1</sup> ( $\text{TS}[(\text{GG})^0(\omega_+, \alpha_+)]$ , SFIG. 69), 83.1 kJ mol<sup>-1</sup> ( $\text{TS}[(\text{GG})^0(\omega_+, \alpha_-)]$ , SFIG. 70) and 75.7 kJ mol<sup>-1</sup> ( $\text{TS}[(\text{GG})^0(\omega_-, \alpha_-)]$ ), respectively. (SFIG. 72) (STable 112) These DFT calculated values are similar to those obtained in the literature for *N*-methylacetamide, 62.8–109.1 kJ mol<sup>-1</sup>,<sup>22,24</sup> suggesting that *N*-methylacetamide is indeed a relevant model to study further details of the *cis-trans* isomerization, as completed below.

### The *cis-trans* isomerization of protonated and deprotonated *N*-methylacetamide

One might wonder how such a high activation energy of isomerization (81.7–109.1 kJ mol<sup>-1</sup>) could be reduced when the process occurs in nature. Isomerization enzymes are thought to act via acid-base catalysis, the probable mechanism of which has not been elucidated in detail as of yet.<sup>8,9</sup> As the *O*- and *N*-protonation of the peptide bond takes place at  $pK_a = 0$  and  $-7$ ,<sup>40</sup> respectively; thus neither direct protonation nor deprotonation can be easily achieved under normal cellular conditions ( $pH \sim 6 \pm 2$ ).

To gain an understanding of the structural and energetic consequences of acid-base catalysis, all the components of isomerization were brought onto a common molecular platform that allowed us to compare the  $\Delta\Delta G$  values of the protonated, deprotonated and neutral amide systems. We aimed to find and identify the most



**FIG. 3.** (a) The PES  $\sim f(\omega, \alpha)$  potential energy surface of *syn-O*-protonated *N*-methylacetamide. The top “forward” oriented IRC path (light blue arrows) connects critical points as:  $\text{Min}[\text{O}^{+s}(\omega_{180})]$ ,  $\text{TS}[\text{O}^{+s}(\omega_-, \alpha_+)]$ ,  $\text{Min}[\text{O}^{+s}(\omega_0)]$ ,  $\text{TS}[\text{O}^{+s}(\omega_+, \alpha_+)]$  and  $\text{Min}[\text{O}^{+s}(\omega_{180})]$ , while the bottom, “backward” IRC path (black arrows) connects  $\text{Min}[\text{O}^{+s}(\omega_{180})]$ ,  $\text{TS}[\text{O}^{+s}(\omega_+, \alpha_-)]$ ,  $\text{Min}[\text{O}^{+s}(\omega_0)]$  and  $\text{TS}[\text{O}^{+s}(\omega_-, \alpha_-)]$ , respectively. (b) The PES  $\sim f(\omega, \alpha)$  potential energy surface of *anti-O*-protonated *N*-methylacetamide. The top “forward” oriented IRC path (light blue arrows) connects critical points as:  $\text{Min}[\text{O}^{+a}(\omega_{180})]$ ,  $\text{TS}[\text{O}^{+a}(\omega_-, \alpha_+)]$ ,  $\text{Min}[\text{O}^{+a}(\omega_0)]$ ,  $\text{TS}[\text{O}^{+a}(\omega_+, \alpha_+)]$  and  $\text{Min}[\text{O}^{+a}(\omega_{180})]$ , while the lower, “backward” IRC path (black arrows) connects  $\text{Min}[\text{O}^{+a}(\omega_{180})]$ ,  $\text{TS}[\text{O}^{+a}(\omega_+, \alpha_-)]$ ,  $\text{Min}[\text{O}^{+a}(\omega_0)]$  and  $\text{TS}[\text{O}^{+a}(\omega_-, \alpha_-)]$ , respectively. (c) The PEC of the *N*-protonated *N*-methylacetamide given by relaxed “scan” calculations, with a “low energy” maximum determined at dihedral angle  $\omega = 0^\circ$ . (d) The potential energy curve of the *N*-deprotonated *N*-methylacetamide. [Calculations were completed at B3LYP/6-311++G(d,p) level of theory, using the IEFPCM water solvent model.]

favorable transition pathways with the lowest energy TSs. In line with the above, the PESs of both the *anti*- and the *syn-O*-protonated *N*-methylacetamide were determined at the DFT level of theory. We found that these PESs also exhibit  $C_2$  symmetry, as found for the neutral PES, presenting 4 plus 4 TSs [Figs. 3(a) and 3(b)], again grouped in pairs. For the *anti-O*-protonated form, the  $\Delta\Delta G$  at B3LYP/6-311++G(d,p) of the TS pairs are similar (152.7 and 153.8  $\text{kJ mol}^{-1}$ ) and have higher energy (STable 30), while the  $\Delta\Delta G$  of the *syn-O*-protonated TS pairs are less energetically similar (147.7 and 134.6  $\text{kJ mol}^{-1}$ ) and are lower compared to those of the *anti-O* forms (STable 31). However, as all these protonated TSs have significantly higher  $\Delta\Delta G$  values compared to those of the neutral molecule (85.1 and 87.8  $\text{kJ mol}^{-1}$ , respectively), they signal that paths with these TSs are less likely to occur and thus this way *O*-protonation may not be the key to solution. *N*-protonation only occurs under extremely acidic conditions, as its  $pK_a = -7$ .<sup>40</sup> However, if the peptide is *N*-protonated, then the N-atom is trapped in a  $sp^3$  hybrid state and therefore no N-atom inversion could occur. In this case, the variable  $\alpha$  becomes unnecessary, the 2D PES is simplified to a 1D PEC and  $\omega$  truly monitors changes in  $\Delta\Delta G$ . The analysis of this PEC  $\sim f(\omega)$  revealed that (i) it has three minima, (ii) the *cis* amide bond ( $\omega = 0^\circ$ ) is a maximum, rather than a minimum [Fig. 3(c)], and (iii) the *trans* isomer,  $\text{Min}[\text{N}^+(\omega_{180})]$ , remains the global minimum of this PEC. The two very shallow local minima were assigned at  $\omega \approx \pm 60^\circ$  correspond to the *gauche*(+) and *gauche*(-) states, coded as  $\text{Min}[\text{N}^+(\omega_-)]$  and  $\text{Min}[\text{N}^+(\omega_+)]$ , respectively. The  $\sim 12.4 \text{ kJ mol}^{-1}$  relative energy of  $\text{Max}[\text{N}^+(\omega_0)]$  at theoretical level B3LYP/6-311++G(d,p), which corresponds to the *cis* isomer, is so low that at room temperature the interconversion between the above 3 minima would occur spontaneously via the maximum, in case of the –albeit very unlikely – event of *N*-protonation [Fig. 3(c)].

*N*-methylacetamide can only be *N*-deprotonated at  $pK_a = 18$ .<sup>40</sup> In this extremely basic molecular environment hydrolysis of the polypeptide chain would also occur, making this mode of isomerization unlikely. For the *N*-deprotonated *N*-methylacetamide, two TSs were found at  $\omega \approx -94^\circ$  and  $\omega \approx 94^\circ$  at B3LYP/6-311++G(d,p) level of theory, forming a mirror image pair [Fig. 3(d)]. The energy barrier of this isomerization pathway is  $\sim 107.1 \text{ kJ mol}^{-1}$  at B3LYP/6-311++G(d,p), higher than the TSs of the neutral and *N*-protonated forms. In conclusion, *N*-deprotonation and *O*-protonation cannot increase the isomerization rate in a continuous solvent model.

### Explicit solvent assisted acidic-base reactions of the *N*-methylacetamide

Amide structures (minima, TSs, and maxima) complexed with a critical number of correctly positioned explicit  $\text{H}_2\text{O}$ s could be involved in a catalyzed *cis-trans* isomerization route. Recognizing the limitations of an implicit hydrate model system, all minima, TSs and maxima of *N*-methylacetamide have been reoptimized so that the amide model is now complexed with two explicit  $\text{H}_2\text{O}$ s.

In general, the reoptimization with the two H-bonded waters has only a marginal effect on the geometrical properties of *N*-methylacetamide (STables 100 and 101). Since the intramolecular H-bonds lower the total energy of the explicitly hydrated molecular complexes, the two water molecules in the desired protonation state

(H<sub>2</sub>O, H<sub>3</sub>O<sup>+</sup> or OH<sup>-</sup>) were added in such a way, that the whole protonation/deprotonation and isomerization reaction could be treated and evaluated within a common molecular framework.

The addition of two H-bonded water molecules to the neutral amide system and the formation of H-bonded complexes does not qualitatively change the isomerization PES: its C<sub>2</sub> symmetry is preserved and the relative energy of TSs remain high enough to practically disallow isomerization via this path at room temperature [Table IV(a)]:  $\Delta\Delta E^{\text{CCSD(T)/def2-TZVP}}$  activation energies are 76.5 and 91.3 kJ mol<sup>-1</sup>, respectively. Thus, the TSs with explicit H<sub>2</sub>O remain approximately the same as those calculated using the implicit (using only CPCM) approach (80.8 and 83.5 kJ mol<sup>-1</sup>, respectively) (Table I). Nevertheless, the lowest energy pathway selected in the explicitly hydrated amide model (76.6 kJ mol<sup>-1</sup>) is about 7 kJ mol<sup>-1</sup> lower than its implicitly hydrated counterpart (83.5 kJ mol<sup>-1</sup>).

N-deprotonation of the amide bond does not provide a better alternative either, even with the inclusion of the two explicit water molecules. The energy gap between the *cis* (Min[N<sup>-</sup><sub>aq(O),aq(NH)</sub>( $\omega_0$ )] (SFig. 51) and *trans* (Min[N<sup>-</sup><sub>aq(O),aq(NH)</sub>( $\omega_{180}$ )] (SFig. 48) isomers shrinks to 6.6 kJ mol<sup>-1</sup> ( $\Delta E^{\text{Min}[N_{aq(O),aq(NH)}^-](\omega_0)} - \Delta E^{\text{Min}[N_{aq(O),aq(NH)}^-](\omega_{180})}$ ), [Table IV(f)] as compared to that the 15.2 kJ mol<sup>-1</sup> calculated applying the implicit water solvent model ( $\Delta E^{\text{Min}[N^-](\omega_0)} - \Delta E^{\text{Min}[N^-](\omega_{180})}$ ) (Table III). More importantly, the maximum separating the *trans* from the *cis* isomer of this form remains also too high to overcome at room temperature:  $\Delta E^{\text{Max}[N_{aq(O),aq(NH)}^-](\omega_-)} \Delta E^{\text{Min}[N_{aq(O),aq(NH)}^-](\omega_{180})} = 118.3$  kJ mol<sup>-1</sup> [Table IV(f)].

The examples of both cyclophilin and dihydrofolate reductase (DHFR) enzyme substrate complexes<sup>9</sup> show that the *cis-trans* isomerization of Xxx-Pro amides is associated with an *N*-protonation step, which terminates amide conjugation and facilitates  $\omega$  rotation. While direct *N*-protonation occurs only at very acidic pH (e.g., *N*-methylacetamide *pKa* = -7), *O*-protonation could be completed under less acidic conditions at pH ~ 0. Here we propose that *O*-protonation could occur first, followed by a still unclear proton “jump” from the *O*-atom of the carbonyl group to the nitrogen of the amide linkage. Therefore, the associated PESs of both the *syn*- and *anti-O*-protonated *N*-methylacetamide were analyzed focusing especially on this aspect. The topologies of the PESs change only slightly when explicit H<sub>2</sub>O are applied, compared to those obtained using the implicit water model [Figs. 3(a) and 3(b)]. In case of *O*-protonation of the amide bond, proton transfer from H<sub>3</sub>O(+) to *N*-methylacetamide · 2H<sub>2</sub>O lowers the relative energy of the complex to -83.1 (*syn*) (SFig. 39) and -95.9 kJ mol<sup>-1</sup> (*anti*) (SFig. 34), respectively. The lowest energy TS for the *syn-O*-protonated form is +43.2 kJ mol<sup>-1</sup> (SFig. 42), while for the *anti* configuration this value is +50.0 kJ mol<sup>-1</sup> (SFig. 37), with respect to the neutral amide · 2H<sub>2</sub>O plus H<sub>3</sub>O(+), which means significantly smaller net energy cost for the *cis-trans* isomerization, compared to that of the neutral form (see above):  $\Delta\Delta E^{\text{TS}[Z_{aq(O),aq(NH)}^0](\omega_+, \alpha_+)}} = +76.5$  kJ mol<sup>-1</sup> [Tables IV(a)–IV(c)]. The conformational shift of the *O*-protonated *N*-methylacetamide · 2H<sub>2</sub>O from Min[O<sup>+</sup><sub>aq(O),aq(NH)</sub>( $\omega_{180}$ )] (SFig. 39) to TS[O<sup>+</sup><sub>aq(O),aq(NH)</sub>( $\omega_-, \alpha_+$ )] (SFig. 42) is similar in extent to that seen along the IRC path, calculated using the implicit water model [Fig. 5(a)].

### Proton transfer between the *O*- and *N*-protonated states of *N*-methylacetamide

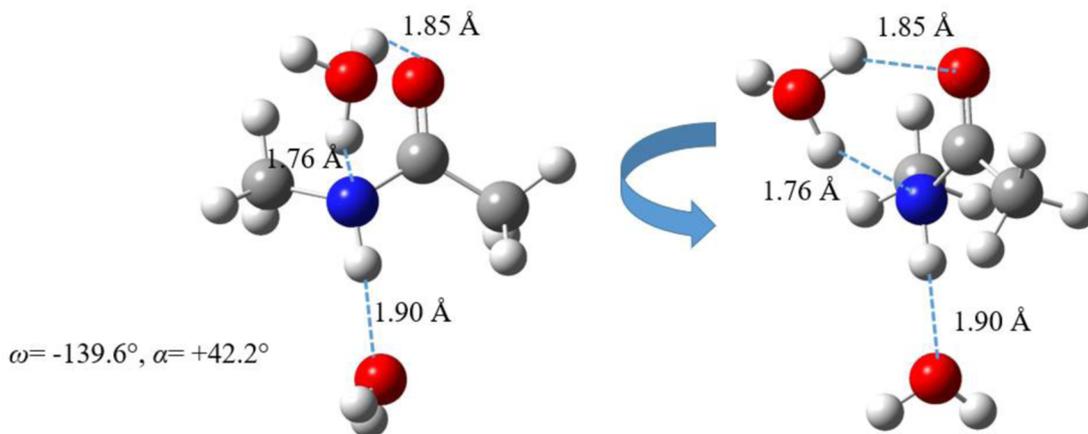
The energy barrier of tautomerization between *O*- and *N*-protonated states (RC[(*O*-N)+aq(O),aq(NH)]) (SFig. 62) is even higher ( $\Delta\Delta E^{\text{CCSD(T)/def2-TZVP}} = 161.4$  kJ mol<sup>-1</sup>) [Table IV(e)] than the energy barrier of *cis-trans* isomerization of the neutral *N*-methylacetamide. Thus, it is not a likely route for proton transfer between the two protonated states. However, at ( $\omega = -139.6^\circ$ ,  $\alpha = +41.5^\circ$ ) the transition state RC[O<sup>+</sup><sub>aq+(O,NH),aq(NH)</sub>( $\omega_{180}$ )] (SFig. 63) ( $\Delta\Delta E^{\text{CCSD(T)/def2-TZVP}} = +11.6$  kJ mol<sup>-1</sup>) points to a possible gateway leading from the *O*- to the *N*-protonated state. This TS-optimized feature marks a particular point on the *O* → *N* proton transition pathway, along which the rotation of  $\omega$  progresses from -180° to 0° [Fig. 5(a)]. The molecular structure of this key junction of the two paths (RC[O<sup>+</sup><sub>aq+(O,NH),aq(NH)</sub>( $\omega_{180}$ )] is also peculiar: that six atoms (N-, C', O-, H<sub>aq+</sub>, O<sub>aq+</sub> and H<sub>aq+</sub>) form a pseudoring containing two H-bonds. The O–H bond of the hydroxonium ion coordinated by both heteroatoms of the amide is slightly weakened [Fig. 4(a)]. There is an overlap between the lone electron pair of the nitrogen and the anti-bonding orbital of the H<sub>3</sub>O<sup>+</sup> ion (SFig. 66) thus the partial donation of the hydrogen toward the N can be seen.

The critical points of the *N*- and *syn-O*-protonated amide potential curve and surface were obtained by IRC calculations started from the RC[O<sup>+</sup><sub>aq+(O,NH),aq(NH)</sub>( $\omega_{180}$ )] (SFig. 63) and RC[N<sup>+</sup><sub>aq+(O,NH),aq(NH)</sub>( $\omega_0$ )] (SFig. 64) at ( $\omega = -1.3^\circ$ ,  $\alpha = +46.2^\circ$ ) transition state pair. Assuming a Boltzmann distribution, the ratio of *trans syn-O*-protonated to *N*-protonated *N*-methylacetamide is 1:10<sup>-7</sup>, which result is in agreement with previous NMR data,<sup>45</sup> where this ratio was measured to be 1:10<sup>-6</sup>. This also further supports our thesis that *N*-protonation of amides can only occur via *O*-protonation with the help of the aqueous environment through this special RC[O<sup>+</sup><sub>aq+(O,NH),aq(NH)</sub>( $\omega_{180}$ )] transition state, however this route does proceed with surprisingly low energy barrier. The *trans* to *cis* isomerization pathway follows the *N*-protonated PEC, as reported above for the implicitly hydrated amide system [Figs. 3(c) and 5(a)], leading via the *N*-protonated minimum, termed Min[N<sup>+</sup><sub>aq(O),aq(NH)</sub>( $\omega_-$ )] (SFig. 47) ( $\omega = -42.1^\circ$ ,  $\alpha = 57.8^\circ$ ) with a relative energy of -47.4 kJ mol<sup>-1</sup> [Table IV(d)], to the local *cis* maximum labelled Max[N<sup>+</sup><sub>aq(O),aq(NH)</sub>( $\omega_0$ )] (SFig. 46) ( $\omega = +2.4^\circ$ ,  $\alpha = +57.2^\circ$ ).

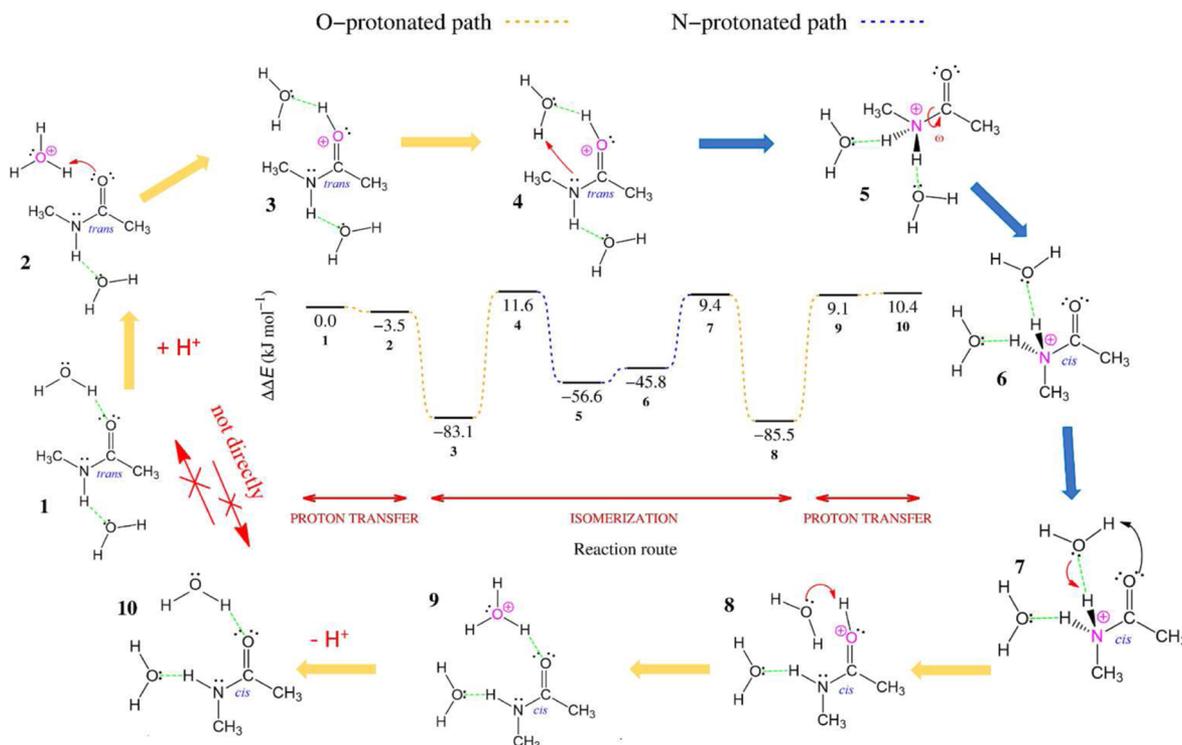
As mentioned above, for the implicitly hydrated model, the conformer Min[N<sup>+</sup><sub>aq(O),aq(NH)</sub>( $\omega_-$ )] (SFig. 47) is a fully optimized minimum with no negative second derivatives, while the critical point Max[N<sup>+</sup><sub>aq(O),aq(NH)</sub>( $\omega_0$ )] (SFig. 46) is a structure with one imaginary vibrational frequency (-75.62 cm<sup>-1</sup>), representing mainly the torsional vibrational mode operating along the C'–N  $\sigma$  bond. As expected also for the *N*-protonated explicitly hydrated amide model, the relative CCSD(T) energy range associated with a complete  $\omega$  period (180° ≤  $\omega$  ≤ +180°) is in the range 0–10.8 kJ mol<sup>-1</sup> [Table IV(d)]. This relatively small energy difference between the *cis* and *trans* isomers means that any configuration along  $\omega$  is possible, meaning that even the *cis* isomer is readily accessible at room temperature.

A similar transition state has been found for ( $\omega = -1.3^\circ$ ,  $\alpha = +46.2^\circ$ ) (RC[N<sup>+</sup><sub>aq+(O,NH),aq(NH)</sub>( $\omega_0$ )] (SFig. 64) with  $\Delta\Delta E^{\text{CCSD(T)/def2-TZVP}} = 9.4$  kJ mol<sup>-1</sup> compared to the neutral

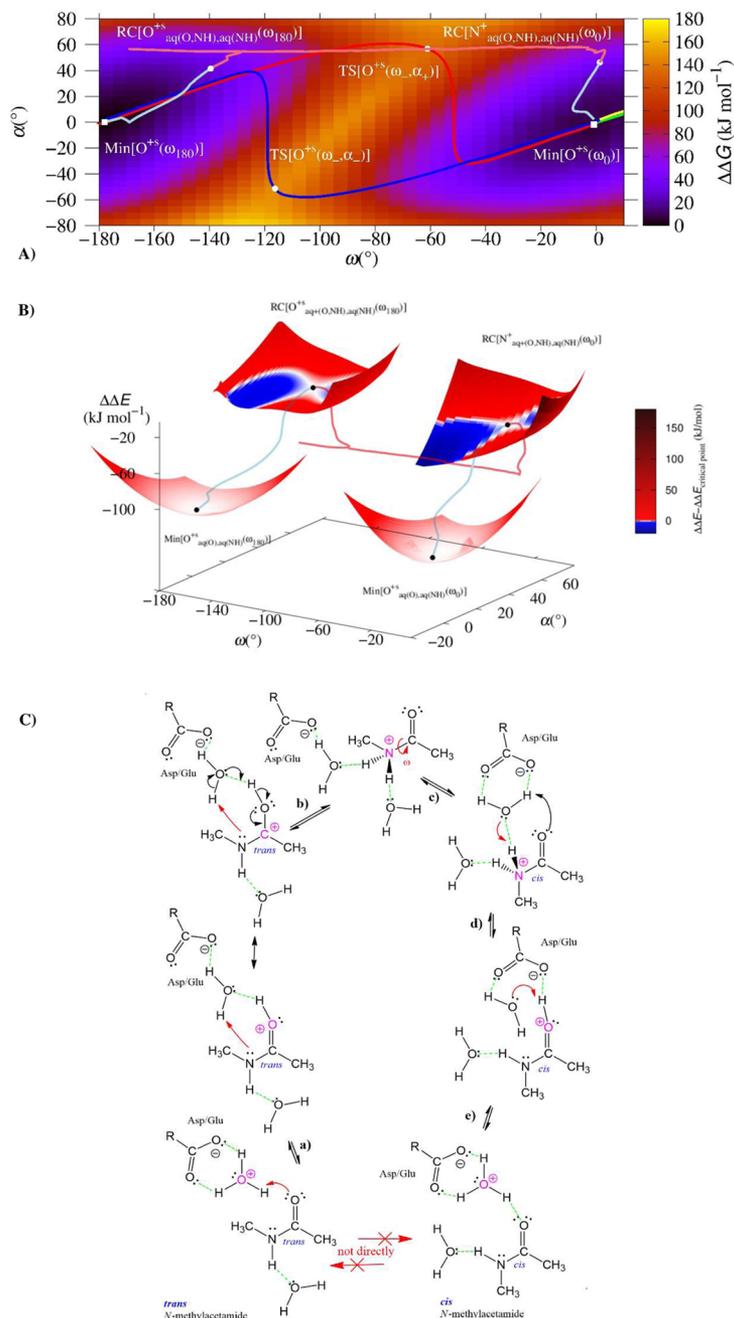
A)



B)



**FIG. 4.** (a) Two quasi orthogonal views of the molecular structure of  $\text{RC}[\text{O}^{+\text{s}}_{\text{aq}+(\text{O},\text{NH}),\text{aq}(\text{NH})}(\omega_{180})]$ , the key transition state comprising the unique six-membered pseudo-ring of N-, C', O-, H-, O- and H-atoms containing two H-bonds. (b) Critical points and relative energies ( $\Delta\Delta E$ s/kJ mol<sup>-1</sup>) determined at the CCSD(T)/def2-TZVP//B3LYP/6-311++G(d,p) level of theory (IEFPCM) of the "superway" obtained by judicious selection among the elements reported in Table IV for the *cis-trans* isomerization of *N*-methylacetamide compared with those obtained for the neutral *trans* molecule  $\text{Min}[\text{Z}^0_{\text{aq}(\text{O}),\text{aq}(\text{NH})}(\omega_{180})]$ , with the following labeling: 1 is  $\text{Min}[\text{Z}^0_{\text{aq}(\text{O}),\text{aq}(\text{NH})}(\omega_{180})]$ , 2 is  $\text{RC}[\text{Z}^0_{\text{aq}+(\text{O}),\text{aq}(\text{NH})}(\omega_{180})]$ , 3 is  $\text{Min}[\text{O}^{+\text{s}}_{\text{aq}(\text{O}),\text{aq}(\text{NH})}(\omega_{180})]$ , 4 is  $\text{RC}[\text{O}^{+\text{s}}_{\text{aq}+(\text{O},\text{NH}),\text{aq}(\text{NH})}(\omega_{180})]$ , 5 is  $\text{Min}[\text{N}^{+\text{s}}_{\text{aq}(\text{O}),\text{aq}(\text{NH})}(\omega_{180})]$ , 6 is  $\text{Max}[\text{N}^{+\text{s}}_{\text{aq}(\text{O}),\text{aq}(\text{NH})}(\omega_0)]$ , 7 is  $\text{RC}[\text{N}^{+\text{s}}_{\text{aq}+(\text{O},\text{NH}),\text{aq}(\text{NH})}(\omega_0)]$ , 8 is  $\text{Min}[\text{O}^{+\text{s}}_{\text{aq}(\text{O}),\text{aq}(\text{NH})}(\omega_0)]$ , 9 is  $\text{RC}[\text{O}^{+\text{s}}_{\text{aq}+(\text{O}),\text{aq}(\text{NH})}(\omega_0)]$  and 10 is  $\text{Min}[\text{Z}^0_{\text{aq}(\text{O}),\text{aq}(\text{NH})}(\omega_0)]$ . All  $\Delta\Delta E$  values are with respect to the  $\text{Min}[\text{Z}^0_{\text{aq}(\text{O}),\text{aq}(\text{NH})}(\omega_{180})]$ , taking the isodesmic reaction into consideration.



**FIG. 5.** (a) The IRC path (red) connecting  $\text{Min}[\text{O}^{+s}(\omega_{180})]$  ( $\omega = -180.0^\circ$ ,  $\alpha = +0.0^\circ$ ) to  $\text{TS}[\text{O}^{+s}(\omega_{-},\alpha_{+})]$  ( $\omega = -61.1^\circ$ ,  $\alpha = +56.8^\circ$ ) and the lowest energy *trans* to *cis* isomerization path ("superway") (orange) connecting  $\text{Min}[\text{O}^{+s}_{\text{aq}(\text{O},\text{NH}),\text{aq}(\text{NH})}(\omega_{180})]$  to  $\text{Max}[\text{N}^{+}_{\text{aq}(\text{O},\text{NH}),\text{aq}(\text{NH})}(\omega_0)]$  via the O- to N- proton transfer step ( $\text{RC}[\text{O}^{+s}_{\text{aq}(\text{O},\text{NH}),\text{aq}(\text{NH})}(\omega_{180})]$ ), depicted by the PEC obtained from a "scan" calculation in 1° steps in the range ( $-180^\circ \leq \omega \leq 0^\circ$ ) from  $\text{Min}[\text{N}^{+}_{\text{aq}(\text{O},\text{NH}),\text{aq}(\text{NH})}(\omega_{180})]$  and IRC path calculations started from  $\text{RC}[\text{O}^{+s}_{\text{aq}(\text{O},\text{NH}),\text{aq}(\text{NH})}(\omega_{180})]$  and  $\text{RC}[\text{N}^{+}_{\text{aq}(\text{O},\text{NH}),\text{aq}(\text{NH})}(\omega_0)]$  at the B3LYP/6-311++G(d,p) level of theory (b) The IRC paths and the potential energy curve for the N-protonated *N*-methylacetamide  $\cdot 2\text{H}_2\text{O}$  complex calculated from the proton transfer steps in both the *trans* ( $\text{RC}[\text{O}^{+s}_{\text{aq}(\text{O},\text{NH}),\text{aq}(\text{NH})}(\omega_{180})]$ ) and *cis* ( $\text{RC}[\text{N}^{+}_{\text{aq}(\text{O},\text{NH}),\text{aq}(\text{NH})}(\omega_0)]$ ) structures with the local potential energy surfaces  $[\Delta\Delta E(\omega,\alpha)]$  for both of them and for the *cis* ( $\text{Min}[\text{O}^{+s}_{\text{aq}(\text{O},\text{NH}),\text{aq}(\text{NH})}(\omega_{180})]$ ) and *trans* ( $\text{Min}[\text{O}^{+s}_{\text{aq}(\text{O},\text{NH}),\text{aq}(\text{NH})}(\omega_{180})]$ ) *syn*-O-protonated *N*-methylacetamide  $\cdot 2\text{H}_2\text{O}$  complexes obtained at B3LYP/6-311++G(d,p). The potential energy surfaces are calculated relative to the given critical point (black)  $\text{RC}[\text{O}^{+s}_{\text{aq}(\text{O},\text{NH}),\text{aq}(\text{NH})}(\omega_{180})]$ ,  $\text{RC}[\text{N}^{+}_{\text{aq}(\text{O},\text{NH}),\text{aq}(\text{NH})}(\omega_0)]$ ,  $\text{Min}[\text{O}^{+s}_{\text{aq}(\text{O},\text{NH}),\text{aq}(\text{NH})}(\omega_{180})]$  and  $\text{Min}[\text{O}^{+s}_{\text{aq}(\text{O},\text{NH}),\text{aq}(\text{NH})}(\omega_{180})]$ . The blue curve represents the O-protonated part of the IRC curve and the red curve represents the N-protonated part. (c) Based on our data, the critical steps of a hypothetical enzyme-catalyzed *cis-trans* isomerization, focusing on the substrate, explicit  $\text{H}_2\text{O}$  and acid-catalyzed sidechain (e.g., Asp or Glu) rearrangements at the active site.

*trans* structure. This low energy barrier allows the proton to be transferred back from the N-atom to the O-atom of the amide, giving a *cis* O-protonated form with  $\Delta\Delta E^{\text{CCSD(T)/def2-TZVP}} = -85.5 \text{ kJ mol}^{-1}$  compared to the neutral *trans* structure. This O-protonated *cis* state is even in a lower energy state than the O-protonated *trans* N-methylacetamide due to the extra hydrogen bond between the two surrounding water molecules. If back-protonation occurs when  $\omega$  is close to or exactly  $0^\circ$  (close to the *cis* isomer), the neutral *cis* amide form is obtained, as the highest barrier it had to overcome during the entire isomerization pathway was only  $+11.6 \text{ kJ mol}^{-1}$ . Note that this relative energy is only one seventh of the lowest TS ( $\text{TS}[Z^0_{\text{aq(O),aq(NH)}}(\omega_+, \alpha_+)]: \Delta\Delta E^{\text{CCSD(T)/def2-TZVP}} = 76.5 \text{ kJ mol}^{-1}$ ) for the neutral IRCs.

### A hypothetical reaction route for the *cis-trans* isomerization of non-proline amides

Figure 4(b) summarizes all the critical points of the latter “superway,” starting with the proton transfer step leading to the rare, but more abundant O-protonated state, which conformer shifts to a critical TS where the proton is transferred from the O-atom of the carbonyl to the N-atom, catalyzed by a well-positioned water. Once the  $n-\pi$  conjugation is terminated, as the N-atom becomes protonated and acquires its  $sp^3$  hybrid state, the rotation around  $\omega$  becomes nearly barrier-free and thus, the *cis* isomer is reachable. At certain points of continuous rotation of the  $C'-N$  bond the proton can be back-transferred to the O-atom with the assistance of a water molecule, which can also take up the proton from the amide. The six critical steps of the protonated N-methylacetamide·2H<sub>2</sub>O complex offer a possibility to propose a hypothetical enzyme-catalyzed *cis-trans* isomerization pathway [Fig. 5(c)]: (i) at  $pH = 6$ , the carboxyl sidechain of Asp or Glu protonates the key water molecule, producing a COO<sup>-</sup> group and a H<sub>3</sub>O<sup>+</sup> ion. (ii) In the second step, the proton is transferred from H<sub>3</sub>O<sup>+</sup> to the C=O group of the amide bond. (iii) During the third step, the O → N proton shift occurs with the assistance of the H<sub>2</sub>O molecule. (iv) Now the N-atom of the protonated amide becomes positively charged ( $sp^3$  hybrid state), the  $C'-N$  rotation becomes almost barrier-free, thus the  $\omega$  dihedral angle rotates to  $\sim 0^\circ$  (*cis* form). (v) In the fifth step, the proton is transferred back to the oxygen, resulting in an O-protonated *cis* form. (vi) In the final, sixth step, the amide is deprotonated. The key water molecule (H<sub>3</sub>O<sup>+</sup>) held by the sidechain of Asp/Glu.

### CONCLUSION

Statistical analysis of proteins retrieved from the PDB confirms that the *cis* isomer of the amide bond exists for hundreds of protein-embedded Xxx-Yyy dipeptides. Undoubtedly, most of the *cis* amide bonds are associated with the Xxx-Pro units, but even when Pro is excluded from the analysis, dozens of *cis* examples were identified for almost every amino acid residue (STable 1). Since the thermodynamically more stable *trans* isomer is separated from the *cis* isomer by a high kinetic energy barrier ( $\sim 85 \text{ kJ mol}^{-1}$ ), it ensures that there is no spontaneous transition between *cis-trans* isomers under physiological conditions, raising the question of what kind of molecular machinery could nevertheless ensure the emergence of such an equilibrium in living cells? In a cellular environment, isomerase enzymes

facilitate this otherwise too high energy barrier to maintain equilibrium and “repair” “faulty” amide bonds. We have shown that the energetics of the isomerization of N-methylacetamide is quite similar to dipeptide segments thus is it a suitable model – especially with two explicit H<sub>2</sub>Os – for its study. Reaction pathways of both the neutral system and protonated and deprotonated, variants were considered. In the case of O-protonation, the isomerization energy barrier remains too high ( $\sim 150 \text{ kJ mol}^{-1}$ ) regardless of whether it is in *syn*- or *anti*-configuration. However, we also found that the N-protonated amide bond can be rotated around  $\omega$  at a low energy cost ( $0 \leq \Delta\Delta E \leq +10 \text{ kJ mol}^{-1}$ ). Direct -spontaneous - N-protonation of the amide can only occur under at extremely acidic conditions of  $-5 \leq pH \leq -9$ , as its  $pKa$  is  $-7$ .<sup>31</sup> Interestingly, O-protonation has a lower  $pKa$  of 0 and can therefore occur under milder acidic conditions or be initiated by an ideally positioned donor trapped in a proximal position – as in the case of an activated water or a protonated side chain within a protein matrix. In view of the above, we started a systematic search for possible transitions between O- and N-protonated isomerization pathways of amides. We found that any tautomerization would have too high energy cost, but if two H<sub>2</sub>O molecules, both H-bonded to the amide, are explicitly included in the model, the connecting TSs could be identified and optimized ( $\text{RC}[O^+_{\text{aq+(O,NH),aq(NH)}}(\omega_{180})], \text{RC}[N^+_{\text{aq+(O,NH),aq(NH)}}(\omega_0)]$ ) (Fig. 4). N-protonation in aqueous solution can occur via O-protonation through these TSs.

Two reversible low energy pathways were also found for proton transfer between the amide O- and N-atoms with the active assistance of a water molecule. Interestingly, there is no stable *cis* ( $\omega = 0^\circ$ ) state for the N-protonated amides since the *cis* critical point is a maximum. However, as the torsion angle along  $\omega$ , requires only low energy to rotate ( $0 \leq \Delta\Delta E \leq +10 \text{ kJ mol}^{-1}$ ), the *cis* orientation also becomes available; this amide form could be stabilized via a proton back transfer. In other words,  $\omega$  will be either *trans* or *cis*, depending on the rotational state of the torsional angle  $\omega$ , if back-protonation stabilizes one of the two isomers. The probability of O-protonation is proportional to the acidity of the medium, but if the amide is locally protonated in an aqueous environment, the *cis* amide bond can be spontaneously formed and “fixed” by deprotonation. This explains why such a large variety and small but significant number of *cis* amide bonds can be identified in oligo- and polypeptides, but also in globular proteins (STable 1).

Here we have studied the detailed mechanism of the amide isomerization reaction supported by a pair of water molecules in the context of an acid-base reaction, revealing the elementary steps of a catalysis. We have identified six key steps, determined the associated molecular complexes, and outlined a possible low-energy pathway for *trans* to *cis* isomerization. The series of minima, TSs and maxima now revealed can be regarded as a model of the defining steps of such a biocatalyzed reaction. Without the explicit knowledge of a protein scaffold, we have outlined the essence of such an isomerization with calculating the CCSD(T) energy levels of the critical points of the process. We present here, for the first time, a coherent mechanism of the acid/water catalyzed, *cis-trans* amide isomerization pathway, along which the highest TS is only  $11.6 \text{ kJ mol}^{-1}$  [Fig. 4(b)], creating the possibility for the two isomers to be in contact with each other at room temperature in an aqueous medium.

## DEDICATION

This work is dedicated to the memory of Professor Imre G. Csizmadia: a teacher, a mentor and a friend of fruitful decades, who initiated this research.

## SUPPLEMENTARY MATERIAL

The detailed statistical data from the PDB for the *cis:trans* ratio of peptide bonds can be found in SFig. 1 and STable 1. Geometrical and energetical data for neutral, protonated and deprotonated *N*-methylacetamide calculated at the theoretical levels B3LYP/6-311++G(d,p), M06-2X/6-311++G(d,p) and CCSD/def2-TZVP//B3LYP/6-311++G(d,p) in the implicit solvent model can be found in STables 2–42 and SFIGs. 2–27. Geometric and energetic data for neutral, protonated and deprotonated *N*-methylacetamide, calculated at the theoretical levels B3LYP/6-311++G(d,p) and CCSD/def2-TZVP//B3LYP/6-311++G(d,p) in explicit solvent model can be found in STables 43–101 and SFIGs. 28–64. Some examples of the structural and conformational notation developed in the article for neutral, protonated and deprotonated molecular structures can be found in STable 102. Critical points and relative energies compared to  $\text{Min}[Z^0_{\text{aq(O),aq(NH)}}(\omega_{180})]$  calculated at CCSD(T)/def2-TZVP//B3LYP/6-311++G(d,p) for *N*-deprotonated, *N*-protonated and *O*-protonated structures are shown in SFig. 65. Second order perturbation energies from natural bond orbital calculations at B3LYP/6-311++G(d,p) of the  $\text{H}_2\text{O} \cdots \text{NH}$  and  $\text{H}_2\text{O} \cdots \text{O}=\text{C}'$  interactions at points of the IRC path calculated from  $\text{RC}[\text{O}^{\text{+s}}_{\text{aq+(O,NH),aq(NH)}}(\omega_{180})]$  and  $\text{RC}[\text{N}^+_{\text{aq+(O,NH),aq(NH)}}(\omega_0)]$  can be found at SFig. 66. Geometric and energetic data for the neutral H-Gly-Gly-OH and *N*-acetyl-L-proline *N*-methylamide can be found in STables 103–113 and SFIGs. 67–75.

## ACKNOWLEDGMENTS

This work was completed in the ELTE Thematic Excellence Programme supported by the Hungarian Ministry for Innovation and Technology (SzintPlus). Project No. 2018-1.2.1-NKP-2018-00005 has been implemented with the support provided from the National Research, Development, and Innovation Fund of Hungary, financed under the Grant No. 2018-1.2-1-NKP funding scheme (HunProtExc). Project No. RRF-2.3.1-21-2022-00015 is implemented with the support of the European Union's Recovery and Resilience Instrument (PharmaLab).

## AUTHOR DECLARATIONS

## Conflict of Interest

The authors have no conflicts to disclose.

## Author Contributions

**Ádám A. Kelemen:** Data curation (lead); Investigation (lead); Visualization (lead); Writing – original draft (lead). **András Perczel:** Conceptualization (equal); Supervision (equal); Writing – review & editing (equal). **Dániel Horváth:** Project administration

(supporting). **Imre Jáklí:** Conceptualization (equal); Software (lead); Supervision (equal); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (equal).

## DATA AVAILABILITY

The data that support the findings of this study are available within the article and its supplementary material.

## REFERENCES

- 1 A. E. Tonelli, "Stability of *cis* and *trans* amide bond conformations in polypeptides," *J. Am. Chem. Soc.* **93**(26), 7153–7155 (1971).
- 2 L. A. LaPlanche and M. T. Rogers, "*cis* and *trans* configurations of the peptide bond in *N*-monosubstituted amides by nuclear magnetic resonance," *J. Am. Chem. Soc.* **86**(3), 337–341 (1964).
- 3 C. Odefey, L. M. Mayr, and F. X. Schmid, "Non-prolyl *cis-trans* peptide bond isomerization as a rate-determining step in protein unfolding and refolding," *J. Mol. Biol.* **245**(1), 69–78 (1995).
- 4 R.-J. Guan, Y. Xiang, X.-L. He, C.-G. Wang, M. Wang, Y. Zhang, E. J. Sundberg, and D.-C. Wang, "Structural mechanism governing *cis* and *trans* isomeric states and an intramolecular switch for *cis/trans* isomerization of a non-proline peptide bond observed in crystal structures of scorpion toxins," *J. Mol. Biol.* **341**(5), 1189–1204 (2004).
- 5 J. Bouckaert, Y. Dewallef, F. Poortmans, L. Wyns, and R. Loris, "The structural features of concanavalin A governing non-proline peptide isomerization," *J. Biol. Chem.* **275**(26), 19778–19787 (2000).
- 6 L. M. Mayr, D. Willbold, P. Rösch, and F. X. Schmid, "Generation of a non-prolyl *cis* peptide bond in ribonuclease T<sub>1</sub>," *J. Mol. Biol.* **240**(4), 288–293 (1994).
- 7 S. Banerjee, N. Shigematsu, L. K. Pannell, S. Ruvinov, J. Orban, F. Schwarz, and O. Herzberg, "Probing the non-proline *cis* peptide bond in  $\beta$ -lactamase from *Staphylococcus aureus* PC1 by the replacement Asn136  $\rightarrow$  Ala," *Biochemistry* **36**(36), 10857–10866 (1997).
- 8 G. Fischer, "Chemical aspects of peptide bond isomerisation," *Chem. Soc. Rev.* **29**(2), 119–127 (2000).
- 9 C. Cox and T. Lectka, "Synthetic catalysis of amide isomerization," *Acc. Chem. Res.* **33**(12), 849–858 (2000).
- 10 C. Schiene-Fischer, J. Habazettl, F. X. Schmid, and G. Fischer, "The hsp70 chaperone DnaK is a secondary amide peptide bond *cis-trans* isomerase," *Nat. Struct. Biol.* **9**(6), 419–424 (2002).
- 11 R. Szostak, J. Aubé, and M. Szostak, "An efficient computational model to predict protonation at the amide nitrogen and reactivity along the C–N rotational pathway," *Chem. Commun.* **51**(29), 6395–6398 (2015).
- 12 I. Okamoto, M. Nabeta, T. Minami, A. Nakashima, N. Morita, T. Takeya, H. Masu, I. Azumaya, and O. Tamura, "Acid-induced conformational switching of aromatic *N*-methyl-*N*-(2-pyridyl)amides," *Tetrahedron Lett.* **48**(4), 573–577 (2007).
- 13 R. Thurston, V. Zantop, K. S. Park, H. Maid, A. Seitz, and M. R. Heinrich, "pH-dependent conformational switching of amide bonds from full *trans* to full *cis* and vice versa," *Org. Lett.* **24**(19), 3488–3492 (2022).
- 14 A. L. Bartuschat, K. Wicht, and M. R. Heinrich, "Switching and conformational fixation of amides through proximate positive charges," *Angew. Chem., Int. Ed.* **54**(35), 10294–10298 (2015).
- 15 O. Roy, C. Caumes, Y. Esvan, C. Didierjean, S. Faure, and C. Taillefumier, "The *tert*-butyl side chain: A powerful means to lock peptoid amide bonds in the *cis* conformation," *Org. Lett.* **15**(9), 2246–2249 (2013).
- 16 B. C. Gorske, B. L. Bastian, G. D. Geske, and H. E. Blackwell, "Local and tunable  $n \rightarrow \pi^*$  interactions regulate amide isomerism in the peptoid backbone," *J. Am. Chem. Soc.* **129**(29), 8928–8929 (2007).
- 17 B. C. Gorske, J. R. Stringer, B. L. Bastian, S. A. Fowler, and H. E. Blackwell, "New strategies for the design of folded peptoids revealed by a survey of noncovalent interactions in model systems," *J. Am. Chem. Soc.* **131**(45), 16555–16567 (2009).

- <sup>18</sup>C. Caumes, O. Roy, S. Faure, and C. Taillefumier, "The click triazolium peptoid side chain: A strong *cis*-amide inducer enabling chemical diversity," *J. Am. Chem. Soc.* **134**(23), 9553–9556 (2012).
- <sup>19</sup>R. Yamasaki, A. Tanatani, I. Azumaya, S. Saito, K. Yamaguchi, and H. Kagechika, "Amide conformational switching induced by protonation of aromatic substituent," *Org. Lett.* **5**(8), 1265–1267 (2003).
- <sup>20</sup>G. N. Ramachandran and A. K. Mitra, "An explanation for the rare occurrence of *cis* peptide units in proteins and polypeptides," *J. Mol. Biol.* **107**(1), 85–92 (1976).
- <sup>21</sup>H. A. Baldoni, G. N. Zamarbide, R. D. Enriz, E. A. Jauregui, Ö. Farkas, A. Perczel, S. J. Salpietro, and I. G. Csizmadia, "Peptide models XXIX. *cis*-*trans* isomerism of peptide bonds: Ab initio study on small peptide model compound; the 3D-Ramachandran map of formylglycinamide," *J. Mol. Struct.: THEOCHEM* **500**(1–3), 97–111 (2000).
- <sup>22</sup>B. Thakkar, J. Svendsen, and R. Engh, "Density functional studies on secondary amides: Role of steric factors in *cis*/*trans* isomerization," *Molecules* **23**(10), 2455 (2018).
- <sup>23</sup>F. J. Luque and M. Orozco, "Theoretical study of *N*-methylacetamide in vacuum and aqueous solution: Implications for the peptide bond isomerization," *J. Org. Chem.* **58**(23), 6397–6405 (1993).
- <sup>24</sup>W. E. Stewart and T. H. Siddall, "Nuclear magnetic resonance studies of amides," *Chem. Rev.* **70**(5), 517–551 (1970).
- <sup>25</sup>G. Scherer, M. L. Kramer, M. Schutkowski, U. Reimer, and G. Fischer, "Barriers to rotation of secondary amide peptide bonds," *J. Am. Chem. Soc.* **120**(22), 5568–5574 (1998).
- <sup>26</sup>T. Schleich, R. Gentzler, and P. H. Von Hippel, "Proton exchange of *N*-methylacetamide in concentrated aqueous electrolyte solutions. I. Acid catalysis," *J. Am. Chem. Soc.* **90**(22), 5954–5960 (1968).
- <sup>27</sup>A. D. Becke, "Density-functional thermochemistry. III. The role of exact exchange," *J. Chem. Phys.* **98**(7), 5648–5652 (1993).
- <sup>28</sup>Y. Zhao and D. G. Truhlar, "The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals," *Theor. Chem. Acc.* **120**(1–3), 215–241 (2008).
- <sup>29</sup>M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, R. Gomperts, B. Menucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, Ö. Farkas, J. B. Foresman, and D. J. Fox, *Gaussian 09*, 2016.
- <sup>30</sup>A. D. McLean and G. S. Chandler, "Contracted Gaussian basis sets for molecular calculations. I. Second row atoms,  $Z = 11$ –18," *J. Chem. Phys.* **72**(10), 5639–5648 (1980).
- <sup>31</sup>J. Čížek, "On the correlation problem in atomic and molecular systems. Calculation of wavefunction components in Ursell-type expansion using quantum-field theoretical methods," *J. Chem. Phys.* **45**(11), 4256–4266 (1966).
- <sup>32</sup>Y. S. Lee, S. A. Kucharski, and R. J. Bartlett, "A coupled cluster approach with triple excitations," *J. Chem. Phys.* **81**(12), 5906–5912 (1984).
- <sup>33</sup>A. Schäfer, C. Huber, and R. Ahlrichs, "Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr," *J. Chem. Phys.* **100**(8), 5829–5835 (1994).
- <sup>34</sup>J. Zheng, X. Xu, and D. G. Truhlar, "Minimally augmented Karlsruhe basis sets," *Theor. Chem. Acc.* **128**(3), 295–305 (2011).
- <sup>35</sup>F. Neese, "Software update: The ORCA program system, version 4.0," *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **8**(1), e1327 (2018).
- <sup>36</sup>R. Dennington, T. A. Keith, and J. M. Millam, *GaussView*, 2016.
- <sup>37</sup>M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, and G. R. Hutchison, "Avogadro: An advanced semantic chemical editor, visualization, and analysis platform," *J. Cheminf.* **4**(1), 17 (2012).
- <sup>38</sup>E. Cancès, B. Mennucci, and J. Tomasi, "A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics," *J. Chem. Phys.* **107**(8), 3032–3041 (1997).
- <sup>39</sup>M. Cossi, N. Rega, G. Scalmani, and V. Barone, "Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model," *J. Comput. Chem.* **24**(6), 669–681 (2003).
- <sup>40</sup>M. A. Eriksson, T. Härd, and L. Nilsson, "On the pH dependence of amide proton exchange rates in proteins," *Biophys. J.* **69**(2), 329–339 (1995).
- <sup>41</sup>K. Fukui, "The path of chemical reactions - the IRC approach," *Acc. Chem. Res.* **14**(12), 363–368 (1981).
- <sup>42</sup>H. M. Berman, "The protein data bank," *Nucleic Acids Res.* **28**(1), 235–242 (2000).
- <sup>43</sup>M. Cyterski, C. Barber, M. Galvin, R. Parmar, J. M. Johnston, D. Smith, A. Ignatius, L. Prieto, and K. Wolfe, "PiSCES: Pi(scine) stream community estimation system," *Environ. Modell. Software* **127**, 104703 (2020).
- <sup>44</sup>B. S. Thakkar, J.-S. M. Svendsen, and R. A. Engh, "Cis/*trans* isomerization in secondary amides: Reaction paths, nitrogen inversion, and relevance to peptidic systems," *J. Phys. Chem. A* **121**(36), 6830–6837 (2017).
- <sup>45</sup>R. B. Martin, "O-protonation of amides in dilute acids," *J. Chem. Soc., Chem. Commun.* **13**, 793 (1972).