Submitted to *J. Am. Soc. Mass Spectrom.*

Thermodynamics and Mechanisms of Protonated Diglycine Decomposition: A Computational Study

P. B. Armentrout and A. L. Heaton

Department of Chemistry, University of Utah, Salt Lake City, Utah 84112

Abstract

We present a full computational description of the fragmentation reactions of protonated diglycine (H^+GG). Relaxed potential energy surface scans performed at B3LYP/6-31G(d) or B3LYP/6-311+G(d,p) levels are used to map the reaction coordinate surfaces and identify the transition states (TSs) and intermediate reaction species for seven reactions observed experimentally in the succeeding companion paper. All structures are optimized at the B3LYP/6-311+G(d,p) level, with single point energies of the key optimized structures calculated at B3LYP and MP2(full) levels using a 6-311+G(2d,2p) basis set. These theoretical structures and energies are compared to extensive calculations in the literature. Although the pathways elucidated here are generally in agreement with those previously outlined, new details and, for some reactions, lower energy transition states are located. Further, the mechanism for the combined loss of carbon monoxide and ammonia is explored for the first time.

Key words: diglycine, protonated peptides, mechanism, mobile proton, theoretical calculations

Correspondence to: P. B. Armentrout; *e-mail*: armentrout@chem.utah.edu

UU IR Author Manuscript

Introduction

Detailed characterizations of gas-phase decompositions of peptides as accomplished by mass spectrometric means are important in understanding biologically relevant reactions and in using this technique to provide detailed sequence information. The gas phase fragmentation of protonated peptides occurs primarily at the amide bonds of peptide linkages, where the charge may be retained on either the N-terminal or C-terminal fragment, producing b or y ions, respectively. Additional fragmentations have been observed, including the loss of CH_2O_2, a process investigated in numerous amino acid and peptide systems in gas phase studies [1-4]. Significant progress in analyzing such processes computationally has been made over the past decade. For instance, multiple pathways for the loss of CH_2O_2 have been characterized [5] and the most thermodynamically and kinetically favored mechanism involves loss of H_2O + CO. In the specific case of the decomposition of protonated diglycine, H^+GG, computational studies include the work of Klassen and Kebarle [6] and Wesdemiotis and coworkers [7]. More comprehensive examinations of the main fragmentation pathways have been conducted by Paizs and coworkers [8,9] and Balta et al. [10].

In the following paper [11], guided ion beam tandem mass spectrometry is used to quantitatively characterize the energetics for H^+GG decomposition for the first time. Six ionic products are formed by loss of CO, loss of H_2O forming the b_2 ion, combined loss of H_2O + CO forming the a_2 ion, combined loss of CO + NH_3, and formation of the y_1 (H^+G) and a_1 ($CH_2NH_2^+$) ions. In the present paper, we explore the reaction coordinate surfaces for these processes using computational theory. This work elucidates the mechanisms for these reactions as well as providing structures, vibrational frequencies, and rotational constants needed for accurate analysis of the experimental data. Single point energies also yield energetic information for comparison to the resultant experimental threshold energies, which allows the mechanistic pathways found computationally to be validated. The theoretical mechanisms for fragmentation of H^+GG elucidated here generally match those previously reported [5,8-10], although alternate low energy pathways are found here and loss of ammonia is characterized for the first time.

Computational Section

Model structures, vibrational frequencies, and energetics for all reaction species, including all transition state (TS) and intermediate species, were calculated using Gaussian 03 [12]. Series of relaxed potential energy surface (PES) scans at the B3LYP/6-31G(d) or B3LYP/6-311+G(d,p) levels were performed in order to identify the elementary steps. All intermediate and TS structures occurring along the PESs were optimized at the B3LYP/6-311+G(d,p) level, where each TS was found to contain one imaginary frequency and each intermediate was vibrationally stable. Rotational constants were obtained from the optimized structures, and all vibrational frequencies were also calculated at this level. Zero-point vibrational energy (ZPE) corrections to the relative energies use vibrational frequencies scaled by 0.99 [13]. Single point energies were determined at the B3LYP and MP2(full) levels using the 6-311+G(2d,2p) basis set. In a recent study of protonated glycine [14], these levels of theory were found to provide accurate reproduction of several relevant experimental results. Other approaches, e.g., larger basis sets or MP2 geometry optimizations, were not found to provide any additional accuracy. Single point energies at the B3P86/6-311+G(2d,2p) level were also obtained in the course of the present work, but as also shown in the glycine work, these substantially overestimate experimental values and therefore are not reported.

As a means of identifying the various conformations of H^+GG, we use a nomenclature that specifies the site of protonation in brackets followed by a designation of the five dihedral angles going from the N terminus to the hydroxyl group (i.e., ∠NCCN, ∠CCNC, ∠CNCC, ∠NCCO, and ∠CCOH), where c (cis) stands for angles < 45°, g (gauche) for angles between 45° and 135°, and t (trans) for angles > 135°. Thus the GS is [N_1]-ttttt where the N_1 designation indicates the proton is on the nitrogen of the first (N-terminal) residue. For diglycine species in which the proton is on the oxygen of the first residue (O_1), it is also useful to designate the orientation of the proton (c or t) by indicating the ∠CCO_1H dihedral angle, either [O_{1c}] or [O_{1t}]. When the proton is located on the carbonyl oxygen of the second residue (O_2) forming a gem-diol structure, the ∠CCO_2H dihedral angle is also specified by a sixth letter. The position of the

N-terminal amino group (when not protonated or involved in a hydrogen bond) can also be rotated but is often such that the lone-pair is cis with respect to the CC bond. When the amino group is rotated to the trans position, the ∠NCCN dihedral is augmented by a subscript t. In a couple of cases, cis and gauche dihedral angles can have opposite signs leading to distinct conformations and in such cases a subscript + and ─ are used to distinguish them, although not all such possibilities are explicitly noted. Transition states are indicated by TS followed by the protonation site and backbone conformation. Transition states for proton transfer steps are named like TS[N_1-O_1]-ttttt and those for dihedral angle rotations as TS[N_1]-t(tc)ttt. Although a bit more complicated than simply numbering the various species, we believe this nomenclature allows better visualization of the species and can potentially be systematically extended to longer chains as well.

Results

Mechanisms for the decomposition of H^+GG have been elucidated previously using computational theory by Paizs and coworkers (Paizs, Csonka, Lendvay, and Suhai, PCLS [8] and Paizs and Suhai, PS [9]) at the B3LYP/6-31G(d) level and in the latter, at the B3LYP/6-31+G(d,p) level as well. These results were later updated by Balta, Aviyente, and Lifshitz (BAL) at the B3LYP/6-31+G(d,p) level [10]. Some minor revisions in these findings are suggested by results of the present, higher level B3LYP/6-311+G(d,p) calculations.

Protonation of GG most favorably occurs at the amide nitrogen with the structure further stabilized by several hydrogen bonds. Thus, the GS of H^+GG is [N_1]-ttttt (designated as **A1** by PCLS and **1** by BAL). The GS structure of H^+GG has previously been characterized computationally [6,8-10,15,16] and the results verified using IRMPD spectroscopy [15,16]. In the following sections, the various conformers available to protonated GG and its decomposition products are discussed along with the TSs that connect them. Relative energies of the former at several levels of theory and from the literature are listed in Table S1 with structures shown in Figure S1 of the Supplementary Material. Table S2 includes relative energies of the various TSs.

Proton Migration Before Decomposition

As previously noted by PCLS [8], there are a multitude of conformers for the protonated GG dipeptide, several of which have enantiomeric forms, e.g., $[N_1]$-ttg$_+$tt and $[N_1]$-ttg$_-$tt where the CNCC dihedral angle leads to a chiral center. Because such enantiomers have identical energies, such forms will generally not be independently described here. In the present work, we have located 32 additional conformers of H^+GG compared to the 28 found by PCLS and BAL, although it should be noted that neither those studies nor this one are designed to be comprehensive elucidations of all possible H^+GG structures. Descriptions of these species and comparisons with literature results can be found in the Supplementary Material.

Transition states connecting the various conformers of H^+GG having trans ∠CCNC peptide bonds vary considerably in their energies, Table S2, but all lie below the energies needed for decomposition, as previously concluded by PCLS as well. Dihedral angle rotations for $[N_1]$ and $[O_1]$ structures are relatively low in energy, never exceeding 100 kJ/mol above the GS, whereas those for the higher lying $[N_2]$ and $[O_2]$ structures lie between 80 and 130 kJ/mol, commensurate with the higher relative energies of these intermediates. Proton transfers between the $[N_1]$ and $[O_1]$ structures never exceed 35 kJ/mol, whereas TS $[N_1\text{-}N_2]$ structures fall in the 70 – 120 kJ/mol energy range, and other variations lie over 100 kJ/mol above the GS. Fifteen of the proton transfer TSs found here were also located by PCLS or BAL, whereas we identify another eleven. We also locate 30 additional dihedral angle rotations, compared to the 20 found by PCLS and BAL, some of which interconnect stable conformers previously identified.

y_1/a_1 Formation: Rate Limiting Steps

Formation of the y_1 (H^+G) product ion must involve transfer of the mobile proton to the N_2 amide nitrogen. Paizs and Suhai (PS) identify the "reactive configuration" as $[N_2]$-cggtt (called **P** in [9] and **D1** in [8]). The route identified by PCLS from GS $[N_1]$-ttttt (**A1**) to $[N_2]$-cggtt involves passing through $[N_1]$-ttgtt (**A2**) and $[N_1]$-gtgtt (**A3**). Because they involve only rotations of dihedral angles, the TSs between the three $[N_1]$ species are low in energy, 6 – 7 (9.5, PCLS) kJ/mol for TS$[N_1]$-tt(tg)tt (**A1_A2**) and 10 - 21 (18.5, PCLS) kJ/mol for TS$[N_1]$-(tg)tgtt

(**A2_A3**) relative to the GS. (Note that the energy of TS[N_1]-(tg)tgtt is 0.4 kJ/mol below to 0.3 kJ/mol above [N_1]-gtgtt, such that the latter structure very easily converts to [N_1]-ttgtt.) In contrast, the proton transfer step from [N_1]-gtgtt to [N_2]-cggtt requires passing over TS[N_1-N_2]-(gc)(tg)gtt (**A3_D1**), 71 – 86 (78.0 , PCLS) kJ/mol above the GS. However, even this barrier is much lower than the rate-limiting step, TS[N_2]-cggtt{$C_{\alpha1}$~OC~N_2} (called **TS_B0** by PS and **tsy1** by BAL), in which the $C_{\alpha1}$-CO and OC-N_2 bonds are synchronously broken (as indicated by the ~ symbol and shown in the inset of Figure 1). Like PS and BAL, we were unable to locate any TSs corresponding to sequential cleavage of these two bonds in this mechanism, and with one exception, this is true for the other TSs discussed below. Indeed, location of suitable TSs for cleavage of the peptide bond in [N_2] species was most conveniently performed by inducing cleavage of the $C_{\alpha1}$-CO bond instead. PS (BAL) find that this TS lies 159 – 161 (143 – 163) kJ/mol above the GS, comparable to results of the present calculations, 149 – 168 kJ/mol. Once over this barrier, a complex of $CH_2NH_2^+$, glycine, and CO is formed.

In addition to this path for formation of the dominant y_1 and a_1 products, we found an energetically comparable pathway, shown in Figure 1. The GS [N_1]-ttttt structure first rearranges by rotation of the N-terminal CC bond such that the protonated N-terminal amino group can transfer a proton to the imide group. This transfer passes over TS[N_1-N_2]-(tc)t(tg)tt and forms the imino protonated complex, [N_2]-ctgtt (**D2**). Note that this structure retains a hydrogen bond between the two nitrogen atoms (N_1 and N_2) as well as between N_2 and O_2. From [N_2]-ctgtt, rotation about the N-terminal CC bond forms an alternative conformer of this complex, [N_2]-ttgtt by passing over TS[N_2]-(ct)tgtt. From this intermediate, the system can pass over TS[N_2]-ttgtt{$C_{\alpha1}$~OC~N_2}, where again the $C_{\alpha1}$-CO and OC-N_2 bonds are synchronously broken to form a complex of $CH_2NH_2^+$, glycine, and CO. Compared to the pathway found by PS, the rate-limiting step for this pathway lies 1.2 kJ/mol lower in energy in the B3LYP calculations, but 0.7 kJ/mol higher at the MP2 level of theory. The main difference between these two pathways is that TS[N_2]-cggtt{$C_{\alpha1}$~OC~N_2} has a hydrogen bond between the NH of the incipient immonium ion and the carbonyl of glycine, such that it is somewhat more

constrained. Indeed, calculations of the 298 K free energies of these two species indicate that the ttgtt conformer is 1 – 3 kJ/mol lower than cggtt. Likewise, calculations of energy-dependent rate coefficients for each pathway reveal entropies of activation of 50 versus 45 J/K mol, respectively, i.e., ttgtt is a looser TS.

Two alternative pathways for formation of the y_1/a_1 products found here are described in the Supplementary Material. These routes are either higher in energy or entropically disfavored compared to those shown in Figure 1. PS and BAL also examined pathways yielding an aziridinone, i.e., 3-membered cyclic CH_2NHCO, but both found this pathway to exceed those outlined above by about 50 kJ/mol. Therefore, this route was not investigated here.

y_1/a_1 Formation: Products

Once over one of the rate limiting steps elucidated above, complexes of the $CH_2NH_2^+$ immonium ion to glycine and CO are formed. PS located three such complexes, and we find six others along with TSs connecting all nine, Tables S1 and S2. In agreement with PS, we find that intrinsic reaction coordinate (IRC) calculations from both low-lying TSs ($TS[N_2]$-cggtt$\{C_{\alpha1}$~OC~$N_2\}$ and $TS[N_2]$-ttgtt$\{C_{\alpha1}$~OC~$N_2\}$) appear to lead to the formation of $(OC_C)(CH_2NH_2^+)(G_{NH\bullet OC,CH\bullet N})$ (called **CO_PBD_B3** by PS), Figure 1, where our nomenclature indicates that the CO molecule is bound to the carbon atom of the immonium ion and the glycine forms two hydrogen bonds (•) with the immonium ion (whose components are on the left of the hydrogen bond designation). In all cases, the first hydrogen bond listed identifies the shared proton. This complex lies 80 – 83 (101.7, PS) kJ/mol above the global GS. A third pathway over $TS[N_2]$-cttct$\{C_{\alpha1}$~OC~$N_2\}$ (see Supplementary Material) initially forms $(OC_C)(CH_2NH_2^+)(G_{NH\bullet N})$, lying 99 – 104 kJ/mol above the GS. As noted by PS, these systems are floppy enough that the IRC calculations are not necessarily unique and can lead to two other complexes found by PS: $(OC_C)(CH_2NH_2^+)(G_{NH\bullet N,CH\bullet OC})$ (**CO_PBD_B1**) and $(OC_C)(CH_2NH_2^+)(G_{NH\bullet N,NH\bullet OC})$ (**CO_PBD_B2**). All these OC_C complexes are higher in energy than complexes in which the CO molecule rearranges to bind in the plane of the immonium ion. $(OC)(CH_2NH_2^+)(G)$ complexes having the CO molecule bound to HN in the immonium ion are

generally more stable than other configurations. The lowest of these is $(OC_{HN})(CH_2NH_2^+)(G_{NH•N,CH•OC})$, 62 – 67 kJ/mol above the H^+GG GS, but $(OC_{HN})(CH_2NH_2^+)(G_{NH•N,NH•OC})$ and $(OC_{HN})(CH_2NH_2^+)(G_{NH•OC,CH•N})$ were also located and lie 0.1 – 3.1 and 6 – 9 kJ/mol, respectively, higher in energy. The CO can also hydrogen bind to the HC end of the immonium ion, $(OC_{HC})(CH_2NH_2^+)(G_{NH•N,NH•OC})$, which is 7 – 9 kJ/mol higher in energy. Relative to $(OC_{HN})(CH_2NH_2^+)(G_{NH•N,CH•OC})$, TSs between the various complexes are low lying, only 18 – 20 kJ/mol higher in energy. As loss of CO from the lowest energy complex requires only 15 – 22 kJ/mol, dissociation of any of these complexes to form the proton bound dimer $(CH_2NH_2^+)(G)$ + CO can occur readily, especially because the rate-limiting TSs for formation of these complexes lie 86 – 104 kJ/mol above $(OC_{HN})(CH_2NH_2^+)(G_{NH•N,CH•OC})$, Figure 1.

Once the CO molecule has left, the $(CH_2NH_2^+)(G)$ complexes can also explore the same variations in conformations that occur for the CO complexes. The lowest energy conformer found is $CH_2NH_2^+(G_{NH•N,CH•OC})$ (called **PBD_B1** by PS) with variants differing only in the hydrogen bonding configurations (relative energies in kJ/mol): NH•N, NH•OC (0.4 – 0.6) (**PBD_B2**), NH•OC, NH•N (5.5 – 7.5), and NH•OC, CH•N (7.7 – 9.6) (**PBD_B3**). Variants in which the shared proton lies closer to the glycine are found for $CH_2NH(H^+G_{N•HN,CH•OC})$ (-2.7 – 2.6) (**PBD_B1a**) and for $CH_2NH(H^+G_{N•HN})$ (2.8 – 5.4) (**PBD_B2a**), where the NH•OC hydrogen bond is lost upon proton transfer. As also found by PS, the TSs for proton transfer between CH_2NH and G in both of these complexes are low-lying, actually lying below the NH•N forms once zero point energies are included (by 2 – 4 and 0.5 – 2.2 kJ/mol, respectively). This indicates that the barrier for proton motion lies below the zero point level such that the proton is shared nearly equally between the two molecules. Indeed, the N-H-N bond lengths are 1.308 + 1.297 and 1.326 + 1.277 Å, respectively, for these two TSs. In all cases, the various $CH_2NH_2^+(G)$ + CO species lie well below the energy of the rate-limiting $TS[N_2]\{C_{\alpha1}{\sim}OC{\sim}N_2\}$ (ttgtt or cggtt) transition states passed in order to produce them, Figure 1.

For any of these paths, the $(CH_2NH_2^+)(G)$ complexes can dissociate to form either the y_1 (H^+G) or a_1 ($CH_2NH_2^+$) ionic products at somewhat higher energies, Figure 1. Overall, the energy required for loss of CO is clearly limited by the two TS[N_2]{$C_{\alpha 1}$~OC~N_2} transition states (ttgtt and cggtt). TS[N_2]-cggtt{OC~N_2} (see Supplementary Material) has a similar energy, but is much tighter such that it is unlikely to contribute appreciably to this pathway. Subsequent formation of the y_1 and a_1 product ions is limited by their asymptotic product energies, i.e., they have loose TSs. Formation of y_1 + CH_2NH is calculated to lie only 21 – 31 kJ/mol higher in energy than the lowest rate-limiting TS, with formation of a_1 + G another 16 – 22 kJ/mol higher still. Because the proton is shared by the two molecules at their favored protonation sites, dissociation can yield both the y_1 and a_1 ions competitively.

Further Decomposition of H^+G

In previous work [14], we elucidated the mechanism for decomposition of protonated glycine, with results in good agreement with previous computational studies [5,17,18]. Experimentally, the dominant product observed is loss of CO + H_2O to yield $CH_2NH_2^+$ (a_1), with smaller amounts of a $CH_2NH_2^+(H_2O)$ complex observed at slightly lower energies. This latter product is not observed in the fragmentation of H^+GG, which is consistent with the cross section magnitude observed in this previous work. The rate-limiting step for formation of both products is TS(H^+G[N-O_2]-cc), proton transfer from the protonated amino group to the hydroxyl group. The energy of this TS (accompanied by CH_2NH + CO products) relative to the H^+GG GS is 319 – 324 kJ/mol, Table S2.

b_2 Formation: Oxazolone – Rate Limiting Steps

A low energy pathway for loss of H_2O from H^+GG to form the b_2 ion has previously been elucidated by BAL [10] and the energetics detailed there are reproduced here within 11 kJ/mol. The mechanism elucidated there is described in Supplementary Material and leads to [O_{1t}]-ctggt (**6**, **O6**) as the critical precursor for water loss. Lower energy pathways to this critical species are shown in Figure 2 and start with [O_{1c}]-ttttt (**3**, **O1**): either [O_{1c}]-ttttt → [O_{1c}]-ctttt (**O5**) → [O_{1t}]-ctttt → [O_{1t}]-ctctt (**O2**), which has a rate-limiting step of TS[$O_{1c\text{-}t}$]-ctttt at 60 – 64 kJ/mol above

the GS; or [O_{1c}]-ttttt → [O_{1c}]-ttgtt (**O3**) → [O_{1c}]-ctgtt → [O_{1t}]-ctctt (**O2**), which has a rate-limiting step of TS[O_{1c}]-(tc)tgtt at 62 – 66 kJ/mol. From [O_{1t}]-ctctt (**O2**), [O_{1t}]-ctggt (**6, O6**) is formed by passing over TS[O_{1t}]-ct(cg)(tg)t, only 50 kJ/mol above the GS. Thus, these two pathways to formation of [O_{1t}]-ctggt (**6, O6**) (neither of which was characterized by BAL or PCLS) are 36 – 41 kJ/mol lower than the pathway located by BAL.

From [O_{1t}]-ctggt (**6**, **O6**), transfer of the proton to the hydroxyl group over TS[O_1-O_3]-ctggt (**6_comp4**), shown in the inset of Figure 3, leads directly to a complex of water with 2-aminomethyl-5-oxazolone protonated on the ring nitrogen, (H^+AMOx[N_2]-c)(H_2O_R) (called **comp4** by BAL and **B1** by PCLS). The oxazolone has a cis ∠NCCN dihedral angle such that the amine group hydrogen bonds with the imine group in the ring. The water binds through its oxygen atom to the oxazolone ring, as indicated by the R subscript. For this pathway, TS[O_1-O_3]-ctggt, lying 141 – 151 (157, BAL) kJ/mol above the H^+GG reactant GS, is rate-limiting in the formation of the oxazolone b_2 ion. BAL also elucidated a parallel pathway, differing only in that the N-terminus has a cis orientation relative to the amide oxygen, which is described in the Supplementary Material along with two high energy pathways for b_2 ion formation previously elucidated by PCLS [8]. Another pathway more comparable in energy was also located by BAL and is shown in Figure 3. Here the [N_1]-ttttt (**1, A1**) GS reorganizes by rotation around the CN bond of the second residue. TS[N_1]-tt(tg)(tc)t (**1_2**) lies 28 – 29 (31) kJ/mol above the GS and forms [N_1]-ttgct (**2, A5**), lying 19 – 21 (18, 22) kJ/mol above the GS. (This TS is the same as **A2e_A5b** found by PCLS at 31 kJ/mol. We believe it is more appropriately designated as **A1_A5** on the basis of a relaxed potential energy surface scan, which identifies these as the appropriate intermediates it connects. An IRC calculation was indeterminate.) A similar energy but slightly more complicated pathway converts [N_1]-ttttt GS (**1, A1**) → [N_1]-ttgtt (**A2**) → [N_1]-tttct (**A4**) → [N_1]-ttgct (2, **A5**), where TS[N_1]-tt(gt)(tc)t is rate limiting and lies 27 – 29 kJ/mol above the GS, 0.3 kJ/mol below TS[N_1]-tt(tg)(tc)t. (This TS is the same as **A2e_A5a** found by PCLS at 31 kJ/mol. We believe it is more appropriately identified as **A2_A4** on the basis of a relaxed potential energy surface scan. Again an IRC calculation was indeterminate.) From [N_1]-

ttgct, one of the protons on the N-terminal amino group is transferred to the hydroxyl group, while the existing NH•OC interaction is retained. As this transfer occurs, the system passes over TS[N_1-O_3]-ttgct (**2_comp2**), Figure 3, 134 – 155 (165) kJ/mol above the GS, in which the C-OH_2 bond is broken and the five-membered oxazolone ring is formed synchronously. This pathway is calculated to lie 4 – 5 (8) kJ/mol above TS[O_1-O_3]-ctggt (**6_comp4**) at the DFT levels, but 7 kJ/mol below at the MP2 level of theory. When the relative free energies of these two TSs are compared, TS[O_1-O_3]-ctggt becomes another 4 kJ/mol more favorable, a result of the more constrained head-to-tail geometry of TS[N_1-O_3]-ttgct.

This TS leads to a complex of water with AMOx protonated on the ring nitrogen with its terminal amine group approximately perpendicular to the ring (i.e., ∠NCCN = 108°), H^+AMOx[N_2]-g. Water binds to the oxazolone ring and hydrogen bonds to the terminal amine group, (H^+AMOx[N_2]-g)($H_2O_{R,N1•HO}$) (**comp2**), which lies 58 – 64 (61) kJ/mol below the TS. Rotation about the side-chain CC bond over TS(H^+AMOx[N_2]-(cg))(H_2O_R), which requires only 2.5 – 5.3 kJ/mol, brings the amine group into position to hydrogen bond with the imine group in the ring (cis ∠NCCN dihedral), forming the lower energy intermediate (H^+AMOx[N_2]-c)(H_2O_R) (**comp4**, **B1**), which lies 26 – 29 (43) kJ/mol below (H^+AMOx[N_2]-g)($H_2O_{R,N1•HO}$).

The water can also migrate to other stable positions, namely either to hydrogen bond at the terminal amino group yielding (H^+AMOx[N_2]-c)(H_2O_{HN1}), which is 2 – 10 kJ/mol less stable than (H^+AMOx[N_2]-c)(H_2O_R), or at the protonation site forming (H^+AMOx[N_2]-c)(H_2O_{HN2}) (**comp3**), Figure 3, which is 6 – 12 (17) kJ/mol more stable and the most stable conformation of this complex found. Transition states to these two conformations are low in energy, requiring only 4 – 9 and 2 – 10 kJ/mol, respectively. Finally, supported by the hydration, the proton can transfer from the ring (N_2) in (H^+AMOx[N_2]-c)(H_2O_{HN1}) to the terminal amino group yielding (H^+AMOx[N_1]-c)(H_2O_{HN1}), which lies 7 – 17 kJ/mol lower in energy. The TS for this transformation, TS(H^+AMOx[N_1-N_2]-c)(H_2O_{HN1}), lies 18 – 25 kJ/mol above (H^+AMOx[N_2]-c)(H_2O_{HN1}). Because the rate-limiting TSs for formation of H^+AMOx, TS[O_1-O_3]-ctggt at

B3LYP and TS[N_1-O_3]-ttgct at MP2, lie 91 – 105 (113 – 120, BAL) kJ/mol above (H^+AMOx[N_2]-c)(H_2O_{HN2}), any of these transformations can occur readily.

Any of the (H^+AMOx)(H_2O) complexes can lose the water molecule to form the simplest b_2 ion, protonated 2-aminomethyl-5-oxazolone, (H^+AMOx[N_2]-c), Figure 3. This product asymptote lies considerably below TS[O_1-O_3]-ctggt and TS[N_1-O_3]-ttgct, by 35 – 57 kJ/mol, such that one of these TSs is rate-limiting for H_2O loss from H^+GG to form the H^+AMOx b_2 ion. An alternate form of the b_2 ion is protonated at the terminal amino site, H^+AMOx[N_1]-c and lies 13 – 24 kJ/mol higher in energy with a connecting TS(H^+AMOx[N_1-N_2]-c) lying 29 – 39 kJ/mol above H^+AMOx[N_2]-c. Also the amino methyl side-chain can have a gauche configuration, H^+AMOx[N_2]-g, which lies 31 – 36 kJ/mol above H^+AMOx[N_2]-c. These alternate conformations lie 20 – 33 kJ/mol and 0 – 26 kJ/mol, respectively, below the rate-limiting TSs.

a_2 Formation: Oxazolone – Further Decomposition

At higher energies, any of the (H^+AMOx)(H_2O) complexes can rearrange by ring opening (OC~O bond cleavage) to form hydrated acylium ions, Figure 4. The TSs for the ring-opening step, TS(H^+AMOx){OC~O}(H_2O), have similar energies for all conformers of the (H^+AMOx)(H_2O) complexes, lying within 12 kJ/mol of one another and 180 – 201 kJ/mol above the H^+GG GS. In all cases, these TSs lie 36 – 64 kJ/mol above the energy required to lose CO. Three different ($C_3H_7N_2O^+$)(H_2O) complexes can be formed: ($C_3H_7N_2O^+$[N_1]-ct)(H_2O_{HN1}), ($C_3H_7N_2O^+$[N_2]-ct)(H_2O_{HN2}), and ($C_3H_7N_2O^+$[N_2]-ct)(H_2O_{HN1}), lying 126 – 133, 142 – 149, and 153 – 160 kJ/mol, respectively, above the H^+GG GS. These species require modest energies (62 – 68, 45 – 54, 34 – 39 kJ/mol, respectively) to also lose water forming the a_2 product ion in its $C_3H_7N_2O^+$[N]-ct conformation, but these asymptotes lie only 0 – 9 above, 1 – 16 above (shown in Figure 4), and 7 below to 4 kJ/mol above their respective TSs. The $C_3H_7N_2O^+$ product initially has an iminium ion structure, $CH_2NHCOCH_2NH_2^+$, with a hydrogen bond between the two nitrogen atoms, where protonation on N_2 is 1 – 3 kJ/mol more favorable at the B3LYP level and 1 kJ/mol less favorable than N_1 protonation at the MP2 level. The TS between the N_1 and N_2 forms, TS$C_3H_7N_2O^+$[N_1-N_2]-ct, lies 13 – 15 kJ/mol above the lower energy form.

As recently demonstrated by IRMPD studies [19] for the a_2 ion derived from protonated triglycine, the $C_3H_7N_2O^+$ product ion may cyclize to form N_1-protonated 4-imidazolidinone, c-$C_3H_7N_2O^+[N_1]$. In the present system, this transformation can potentially occur before or after losing water, as shown by the parallel pathways in Figure 4 (black and blue surfaces, respectively). Cyclic c-$C_3H_7N_2O^+$ is more stable than the acyclic form by 40 – 55 (40 at the B3LYP/6-311++G(d,p) level in Verkerk et al. [19]) kJ/mol, and the hydrated versions have a stability difference of 48 – 60 kJ/mol. Cyclization requires two steps with the former being rate-limiting both with and without water present: a trans-cis rotation about the C-N_2 bond over TS($C_3H_7N_2O^+[N_2]$-c(tc))(H_2O_{HN2}), followed by $C_{\alpha 2}$-N_1 coupling over TS(c-$C_3H_7N_2O^+[N_1]${$C_{\alpha 2}$~N_1})(H_2O). Note that in the latter step, the nomenclature for the protonation site changes from N_2 to N_1 without the need for a proton migration because of the formation of a covalent C-N bond. Also the ($C_3H_7N_2O^+[N_2]$-cc)(H_2O) intermediate has an alternate form in which the water hydrogen bonds to the N_1 terminus, but this intermediate is 28 – 30 kJ/mol higher in energy. Transition states leading to this [N_1] intermediate are also higher in energy by 18 – 22 and 30 – 32 kJ/mol, respectively. For the [N_2] intermediate, the cyclic complex formed is (c-$C_3H_7N_2O^+[N_1]$)(H_2O_{HN2}), whereas (c-$C_3H_7N_2O^+[N_1]$)(H_2O_{HN1}) lies lower in energy by 22 – 23 kJ/mol. (This latter species has two variants differing in energy by only 1 – 2 kJ/mol. The five-membered ring is puckered with the protonated N_1 nitrogen lying out of the plane defined by the remaining heavy atoms. The lower energy structure has the water bound to the NH bond pointing perpendicular to the plane of the ring (the "up" position as designated by u in Table S1). In the higher energy "down" position, the water hydrogen bonds to the NH bond nearly parallel to the ring plane. The TS between these two forms is low-lying, 3 – 7 kJ/mol above the GS form.) The TS for motion of the water between the N_1 and N_2 positions lies 33 – 34 kJ/mol above the N_1 GS form (11 kJ/mol above the N_2 form). Overall, if the cyclization occurs when the water is retained, the TSs for cyclization lie below the TS(H^+AMOx{OC~O})(H_2O) for CO loss by 4 – 19 kJ/mol and are 19 – 29 kJ/mol below the energy needed for subsequent H_2O loss, Figure 4.

We also considered further decomposition of the oxazolone product ion after loss of H_2O, i.e., loss of CO from H^+AMOx to form the a_2 ion. This is the pathway examined computationally by Verkerk et al. [19], who considered four linear and three cyclic isomers of a_2 as well as the TSs connecting them. (The calculations of Verkerk et al. find comparable energies to the present results for this part of the reaction coordinate surface; however, they did not locate the $C_3H_7N_2O^+[N_2]$-cc intermediate connecting $C_3H_7N_2O^+[N_2]$-ct (**L1**, Verkerk) with c-$C_3H_7N_2O^+[N_1]$ (**C1**). We note that the IR spectrum predicted for the cc species is very similar to that for the more stable ct conformer, such that independent identification of such a species using IRMPD spectroscopy is unlikely.) The present calculations, Figure 4, indicate that CO loss from H^+AMOx$[N_2]$-c passes over TS(H^+AMOx$[N_2]$-c){OC~O}, 36 – 45 kJ/mol higher than TS(H^+AMOx){OC~O}(H_2O). This is followed by steps parallel to those described above, with the key feature being that the rate-limiting step for cyclization, TS($C_3H_7N_2O^+[N_2]$-c(tc)) for B3LYP and TS(c-$C_3H_7N_2O^+[N_1]${$C_{\alpha 2}$~N_1}) for MP2, lies above TS(H^+AMOx$[N_2]${OC~O}) by 8 – 14 kJ/mol, Figure 4. Thus, once the water has left, cyclization is less likely to occur because it requires more energy and is entropically hindered. This conclusion potentially disagrees with the findings of Verkerk et al. who use IRMPD spectroscopy to identify the cyclic species as the dominant a_2 isomer. In their system, the a_2 ion is also formed by decarbonylation of the H^+AMOx b_2 ion, which is a primary fragment of H^+GGG; however, in their work, the a_2 product is formed by in-source fragmentation of H^+GGG, such that the cyclization could be assisted by association with the other fragments, G and CO, which are then dissociated before spectral interrogation.

Overall, the lowest energy pathway for formation of the a_2 ion from H^+GG is limited by TS(H^+AMOx){OC~O}(H_2O) at 180 – 194 kJ/mol, presuming that the $C_3H_7N_2O^+$ product cyclizes. Formation of the open form of $C_3H_7N_2O^+$ + CO + H_2O requires an additional 1 – 16 kJ/mol. If water is lost from the b_2 oxazolone product ion first, then TS(H^+AMOx$[N_2]${OC~O}) becomes the rate-limiting step at 225 – 230 kJ/mol, and cyclization requires an additional 8 – 14 kJ/mol.

U of U IR Author Manuscript

We also located an alternative pathway for loss of CO + H_2O to form the a_2 ion, in which the b_2 ion is not an intermediate. Again the mechanism starts with $H^+GG[O_{1t}]$-ctggt (**6, O6**), but upon $[O_1\text{-}O_3]$ proton transfer, a higher energy TS is overcome such that the C-CO bond is broken. $TS[O_1\text{-}O_3]\{C\sim CO\}$-ctggt, shown in the inset in Figure 4, lies 179 – 187 kJ/mol above the H^+GG GS, 35 – 38 kJ/mol above $TS[O_1\text{-}O_3]$-ctggt that leads to the b_2 oxazolone product ion, and 1 – 8 kJ/mol below $TS(H^+AMOx[N_2]\text{-}c\{OC\sim O\})(H_2O_{HN2})$ leading to CO loss in Figure 4. IRC calculations connect $TS[O_1\text{-}O_3]\{C\sim CO\}$-ctggt with $H^+GG[O_{1t}]$-ctggt and lead to a complex of the acylic a_2 ion with water and CO, $(C_3H_7N_2O^+[N_2]\text{-}ct)(H_2O_{HC\alpha 2}\text{-}c)(OC_{C\alpha 2})$, in which the CO is bound to the terminal carbon atom in a acylium-like structure and the water is hydrogen bound to the CH_2 terminus in a cis position relative to the CO. The CO and H_2O ligands can migrate along the $C_3H_7N_2O^+$ ion, but loss of CO requires only another 7 – 15 kJ/mol and lies 23 – 30 kJ/mol below the rate-limiting TS. (Note that this pathway for CO loss lies 11 – 34 kJ/mol higher than $TS[N_2]\{C_{\alpha 1}\sim OC\sim N_2\}$ elucidated above for CO loss.) Subsequent migration of the water to more stable $HC_{\alpha 2}$-t, HN_1, and HN_2 positions also requires little energy, only 1 – 3 kJ/mol more, and this allows formation of the $(C_3H_7N_2O^+[N_2]\text{-}ct)(H_2O_{HN2})$ species, Figure 4. Overall, this mechanism for loss of both CO and H_2O has a rate-limiting TS that is slightly lower in energy (1 – 8 kJ/mol) than the pathway passing through the b_2 oxazolone product, however, because the oxazolone pathway requires less energy (by 35 – 38 kJ/mol) and originates from the same intermediate, $H^+GG[O_{1t}]$-ctggt (**6, O6**), it seems unlikely that this alternative pathway is competitive.

b_2 Formation: Diketopiperazine

The other possible structure for the b_2 ion is protonated diketopiperazine (H^+DKP), a six-membered ring heterocycle. In order to definitively identify the structure of the b_2 product ion formed in the decomposition of H^+GG, the pathway to formation of this alternate species was also investigated, with results comparable to those of BAL [10]. Formation of H^+DKP requires that the OC-NH peptide bond be in a cis configuration, rather than its normal trans orientation. Several such species are identified and included in Figure S1. (No $[N_2]$ complexes having such a

cis configuration were explored as they do not lead to H^+DKP. Likewise, cis conformations of the CCOH dihedral undoubtedly exist for many of these complexes but were not examined.) Details of the literature pathway to formation of H^+DKP + H_2O (as well as several alternative lower energy pathways determined in the present work) are provided in Supplementary Material. For instance, we find that the lowest energy pathway to [N_1]-gcgtt, Figure S2, has a rate-limiting step of TS[N_1]-g(ct)gtt at 59 – 73 kJ/mol above the GS, or 26 – 38 kJ/mol below the analogous TS located by BAL, TS[N_1]-t(tc)ttt (**1_cis1**). TS[N_1]-g(ct)gtt is lower in energy because hydrogen bonds between the protonated N-terminus and both carbonyls are maintained throughout. This result further underscores the observation of BAL that the trans-cis rotation of the peptide bond is *not* the limiting step in formation of H^+DKP. Overall, the lowest energy pathways to formation of H^+DKP + H_2O involve TS[N_1–$O_{2,3}$]-ccctgc, Figure S3, which are 88 – 97 kJ/mol above those for generating H^+AMOx + H_2O. Thus, even though the final product is lower in energy by 7 – 12 (12, BAL) kJ/mol, formation of the H^+DKP b_2 ion cannot compete with formation of the H^+AMOx b_2 ion.

Loss of CO and NH_3

Neither PCLS nor BAL examined the loss of ammonia from H^+GG. Pingitore et al. [7] have suggested that the sequential loss of CO and ammonia from H^+GG occurs by nucleophilic displacement (S_N2) of NH_3 by CH_2NH in the $(CH_2NH)(H^+G)$ complex formed after CO expulsion, but performed no explicit calculations of this pathway. This S_N2 pathway was explored here and two rate-limiting TSs were located, one lying 230 – 248 kJ/mol above the GS and another having a cis CCOH dihedral that lies 34 – 41 kJ/mol higher. Further details of such a reaction was not explored as much lower energy pathways for deamidation following initial decarbonylation were located.

The lowest energy pathway found is shown in Figure 5. This route starts with the $(CH_2NH_2^+)(G_{NH\bullet OC,CH\bullet N})$ complex (also shown in Figure 1), but could begin with any of its variants. Rearrangement of this complex over a fairly low energy barrier, TS($C_3H_9N_2O_2^+$[N_2]-tgtt{C~N}) at 16 – 18 kJ/mol above the complex, forms a bond between the immonium ion

carbon and the nitrogen atom of glycine. This yields the $C_3H_9N_2O_2^+[N_2]$-tgtt molecule that is 5 – 22 kJ/mol more stable than the starting complex (73 – 82 kJ/mol above GS). Proton transfer from N_2 to N_1 over TS($C_3H_9N_2O_2^+[N_2\text{-}N_1]$-($g_+$t)gtt) requires 161 – 175 kJ/mol and is the rate-limiting TS in the subsequent ammonia expulsion. This forms $C_3H_9N_2O_2^+[N_1]$-g_+gtt, 54 – 62 kJ/mol above the GS, which can rearrange to form $C_3H_9N_2O_2^+[N_1]$-gttt, 98 – 102 kJ/mol above the GS, Figure 5. Not all variants of these $C_3H_9N_2O_2^+$ species were explored. Slightly different pathways were also located and have rate-limiting TSs lying at energies of 169 – 175 kJ/mol for TS($C_3H_9N_2O_2^+[N_2\text{-}N_1]$-g(gt)tc) as shown in the inset of Figure 5, 173 – 183 kJ/mol for TS($C_3H_9N_2O_2^+[N_2\text{-}N_1]$-tgtt), and 179 – 186 kJ/mol for TS($C_3H_9N_2O_2^+[N_2\text{-}N_1]$-gttt). The two lower energy pathways, Figure 5, are facilitated by having the proton pass near the carbonyl oxygen as it transfers. These two pathways are similar to the lower energy mechanism found by Bythell et al. for loss of CO and NH_3 from H^+AGG [[20]], although the protonated carbonyl is a stable intermediate in that system. Ammonia can probably be lost from any of the $C_3H_9N_2O_2^+$ complexes formed after the rate-limiting TSs, but $C_3H_9N_2O_2^+[N_1]$-gttt generates $C_3H_6NO_2^+$[N]-ttt, Figure 5, the lowest energy form of this molecule because it is stabilized by a NH•OC hydrogen bond. These products lie 148 – 157 kJ/mol above the GS and 4 – 37 kJ/mol below the rate-limiting steps. Alternate conformers of the $C_3H_6NO_2^+$ product include [N]-gtt and [N]-gct, 8 – 9 and 16 – 18 kJ/mol, respectively, above [N]-ttt. There are also cyclic versions of this molecule, oxazolone protonated on the N, carbonyl, and ring oxygen (H^+Ox[N], H^+Ox[OC], and H^+Ox[O_R]), but they are 47 – 59, 82 – 85, and 139 – 155 kJ/mol, respectively, above the acyclic [N]-ttt conformer. Overall, the lowest energy pathway found for loss of NH_3 from the ($CH_2NH_2^+$)(G) decarbonylation products involves either TS($C_3H_9N_2O_2^+[N_2\text{-}N_1]$-($g_+$t)gtt or -g(gt)tc), which are 5 – 27 kJ/mol lower in energy than formation of the H^+G + CH_2NH + CO products.

We also looked for mechanisms involving initial loss of NH_3 from H^+GG. Conceivably any of the [N_1] conformers can lose ammonia by cleaving the C-N_1 bond, potentially assisted by backside attack from other parts of the molecule. Although an exhaustive search for such

pathways was not conducted, several routes having rate-limiting TSs above 220 kJ/mol were located, making them too high in energy for further consideration. The only low energy process found starts with $H^+GG[N_1]$-tcgtt and breaks the C-N_1 bond by backside attack of O_2, thereby forming ammonia bound to a six-membered ring, c-$OC(=O)CH_2NHC(=O)CH_2$, protonated at the first carbonyl, protonated diketo-oxazine, H^+DKOx. The TS for this process lies 162 – 181 kJ/mol above the H^+GG GS. The ammonia can presumably move around to different sites around the H^+DKOx molecule, but these transformations were not explored because loss of ammonia lies only 5 – 12 kJ/mol above the TS (169 – 191 kJ/mol above the H^+GG GS). No efficient pathways for CO loss from H^+DKOx could be located. Note that this pathway for NH_3 loss lies only 7 – 12 kJ/mol below (B3LYP) to 12 – 20 kJ/mol above (MP2) that for combined CO + NH_3 loss involving the $C_3H_9N_2O_2^+[N_2\text{-}N_1]$ TSs, and well above the TSs for initial loss of H_2O (by 18 – 56 kJ/mol) or CO (by 16 – 25 kJ/mol). Thus, it is not surprising that ionic products corresponding to initial ammonia loss are not observed here or in previous work.

Discussion

On the basis of the reaction coordinate surfaces calculated here, loss of CO and H_2O both occur as primary dissociations from H^+GG. These pathways are limited by the tight transition states, $TS[N_2]\{C_{\alpha1}\sim OC\sim N_2\}$ (ttgtt or cggtt) or $TS[N_1\text{-}O_3]$-ttgct or $TS[O_1\text{-}O_3]$-ctggt, respectively, rather than by the asymptotic energies of the products, Figures 1 and 3, respectively. These TSs have similar energies near 150 kJ/mol at the B3LYP level, but MP2 suggests the TSs for H_2O loss lie 10 – 15 kJ/mol lower and those for CO loss ~18 kJ/mol higher. Loss of H_2O to form the b_2 product ion having the oxazolone structure, H^+AMOx, passes over TSs lying 89 – 97 kJ/mol lower than those leading to diketopiperazine, H^+DKP. H^+AMOx can further dissociate to form a_2 by losing CO. If the water is lost before decomposition of the oxazolone ring, then the reactions are sequential with formation of the a_2 ion limited by the tight transition state, $TS(H^+AMOx[N_2]\text{-}c\{OC\sim O\})$, Figure 4. However, if the oxazolone ring decomposes before the water is lost, the system passes over $TS(H^+AMOx[N_2]\text{-}c\{OC\sim O\})(H_2O_{HN2})$ and the loss of CO +

H_2O is limited by the energy of the product asymptote, i.e., a loose TS, Figure 4. The ionic product formed by CO loss, $(CH_2NH_2^+)(G)$, can also dissociate further by loss of NH_3, CH_2NH, or G. The former pathway passes over a tight TS, $TS(C_3H_9N_2O_2^+[N_2\text{-}N_1]\text{-}(g_+t)gtt$ or -g(gt)tc), Figure 5, and the latter two channels involve loose TSs, Figure 1. Finally, the y_1 (H^+G) product can dissociate further to form the a_1 product ion in a sequential process.

Acknowledgement

This work is supported by the National Science Foundation, Grant CHE-1049580. A grant of computer time from the Center for High Performance Computing at the University of Utah is gratefully acknowledged.

Appendix A

Supplementary Material

Supplementary material associated with this article may be found in the online versions at doi:

References

1. Tsang, C. W.; Harrison, A. G.: Chemical Ionization of Amino Acids. *J. Am. Chem. Soc.* **98**, 1301-1308 (1976).
2. El Aribi, H.; Rodriquez, C. F.; Almeida, D. R. P.; Ling, Y.; Mak, W. W.-N.; Hopkinson, A. C.; Siu, K. W. M.: Elucidation of Fragmentation Mechanisms of Protonated Peptide Ions and Their Products: A Case Study on Glycylglycylglycine Using Density Functional Theory and Threshold Collision-Induced Dissociation. *J. Am. Chem. Soc.* **125**, 9229-9236 (2003).
3. Abirami, S.; Xing, Y. M.; Tsang, C. W.; Ma, N. L.: Theoretical Study of α/β-Alanine and Their Protonated/Alkali Metal Cationized Complexes. *J. Phys. Chem. A* **109**, 500-506 (2005).
4. Laskin, J.; Denisov, E.; Futrell, J.: A Comparative Study of Collision-Induced and Surface-Induced Dissociation. 1. Fragmentation of Protonated Dialanine. *J. Am. Chem. Soc.* **122**, 9703-9714 (2000).
5. O'Hair, R. A. J.; Broughton, P. S.; Styles, M. L.; Frink, B. T.; Hadad, C. M.: The Fragmentation Pathways of Protonated Glycine: A Computational Study. *J. Am. Soc. Mass Spectrom.* **11**, 687-696 (2000).
6. Klassen, J. S.; Kebarle, P.: Collision-Induced Dissociation Threshold Energies of Protonated Glycine, Glycinamide, and Some Related Small Peptides and Peptide Amino Amides. *J. Am. Chem. Soc.* **119**, 6552-6563 (1997).

7. Pingitore, F.; Polce, M. J.; Wang, P.; Wesdemiotis, C.; Paizs, B.: Intramolecular Condensation Reactions in Protonated Dipeptides: Carbon Monoxide, Water, and Ammonia Losses in Competition. *J. Am. Soc. Mass Spectrom.* **15**, 1025-1038 (2004).
8. Paizs, B.; Csonka, I.; Lendvay, G.; Suhai, S.: Proton mobility in protonated glycylglycine and N-formylglycylglycinamide: a combined quantum chemical and RKKM study. *Rapid Commun. Mass Spectrom.* **15**, 637-650 (2001).
9. Paizs, B.; Suhai, S.: Theoretical study of the main fragmentation pathways for protonated glycylglycine. *Rapid Commun. Mass Spectrom.* **15**, 651-663 (2001).
10. Balta, B.; Aviyente, V.; Lifshitz, C.: Elimination of Water from the Carboxyl Group of $GlyGlyH^+$ *J. Am. Soc. Mass Spectrom.* **14**, 1192-1203 (2003).
11. Armentrout, P. B.; Heaton, A. L.: Thermodynamics and Mechanisms of Protonated Diglycine Decomposition: A Guided Ion Beam Study. *J. Am. Soc. Mass Spectrom.*, following article (2011).
12. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A.: Gaussian 03, revision B.02; Gaussian, Inc.: Pittsburgh, PA, 2003.
13. Montgomery, J. A., Jr.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A.: A complete basis set model chemistry. VI. Use of density functional geometries and frequencies. *J. Chem. Phys.* **110**, 2822-2827 (1999).
14. Armentrout, P. B.; Heaton, A. L.; Ye, S. J.: Thermodynamics and Mechanisms for Decomposition of Protonated Glycine and Its Protonated Dimer *J. Phys. Chem. A*, in press (2011).
15. Wu, R.; McMahon, T. B.: Protonation Sites and Conformations of Peptides of Glycine ($Gly_{1-5}H^+$) by IRMPD Spectroscopy. *J. Phys. Chem. B* **113**, 113, 8767–8775 (2009).
16. Wu, R.; McMahon, T. B.: Infrared multiple-photon dissociation mechanisms of peptides of glycine. *Chem. Eur. J.* **14**, 7765-7770 (2008).
17. Balta, B.; Basma, M.; Aviyente, V.; Zhub, C.; Lifshitz, C.: Structures and reactivity of gaseous glycine and its derivatives. *Int. J. Mass Spectrom.* **201**, 69-85 (2000).
18. Uggerud, E.: The unimolecular chemistry of protonated glycine and the proton affinity of glycine: a computational model. *Theor. Chem. Acc.* **97**, 313-316 (1997).
19. Verkerk, U. H.; Siu, C.-K.; Steill, J. D.; Aribi, H. E.; Zhao, J.; Rodriquez, C. F.; Oomens, J.; Hopkinson, A. C.; Siu, K. W. M.: a_2 Ion Derived from Triglycine: An N_1-Protonated 4-Imidazolidinone. *J. Phys. Chem. Lett.* **1**, 868–872 (2010).
20. Bythell, B. J.; Barofsky, D. F.; Pingitore, F.; Polce, M. J.; Wang, P.; Wesdemiotis, C.; Paizsc, B.: Backbone Cleavages and Sequential Loss of Carbon Monoxide and Ammonia

from Protonated AGG: A Combined Tandem Mass Spectrometry, Isotope Labeling, and Theoretical Study. *J. Am. Soc. Mass Spectrom.* **18**, 1291-1303 (2007).

U of U IR Author Manuscript

Figure Captions

Figure 1. Reaction coordinate surface for CO loss from H^+GG and subsequent y_1 and a_1 formation. Geometry optimizations and single point energies of each elementary step are determined at the B3LYP/6-311+G(d,p) level of theory and corrected for ZPE. Short dashed lines indicate bonds that are breaking or forming for transition states and hydrogen bonds for intermediates and products. The inset and additional horizontal bar indicate the structure and energy of an alternative rate-limiting transition state.

Figure 2. Reaction coordinate surface for two (red and black lines) low-energy paths to form the $H^+GG[O_{1t}]$-ctggt intermediate leading to H_2O loss over TS($H^+GG[O_1\text{-}O_3]$-ctggt). Geometry optimizations and single point energies of each elementary step are determined at the B3LYP/6-311+G(d,p) level of theory and corrected for ZPE.

Figure 3. Reaction coordinate surface for H_2O loss from H^+GG. Geometry optimizations and single point energies of each elementary step are determined at the B3LYP/6-311+G(d,p) level of theory and corrected for ZPE. Short dashed lines indicate bonds that are breaking or forming. The inset and additional horizontal bar indicate the structure and energy of an alternative rate-limiting transition state.

Figure 4. Reaction coordinate surface for CO loss from H^+AMOx (b_2) with H_2O present (black line) and absent (blue line). Geometry optimizations and single point energies of each elementary step are determined at the B3LYP/6-311+G(d,p) level of theory and corrected for ZPE. The inset and additional horizontal bar (in red) indicate the structure and energy of an alternative pathway for loss of CO + H_2O to form the a_2 ion, in which H^+AMOx ion is not an intermediate (see text).

UU IR Author Manuscript

Figure 5. Reaction coordinate surface for NH_3 loss from the $(CH_2NH_2^+)$G intermediate formed in Figure 1. Geometry optimizations and single point energies of each elementary step are determined at the B3LYP/6-311+G(d,p) level of theory and corrected for ZPE. Short dashed lines indicate bonds that are breaking or forming for transition states and hydrogen bonds for the initial reactant. The inset and additional horizontal bar indicate the structure and energy of an alternative rate-limiting transition state.

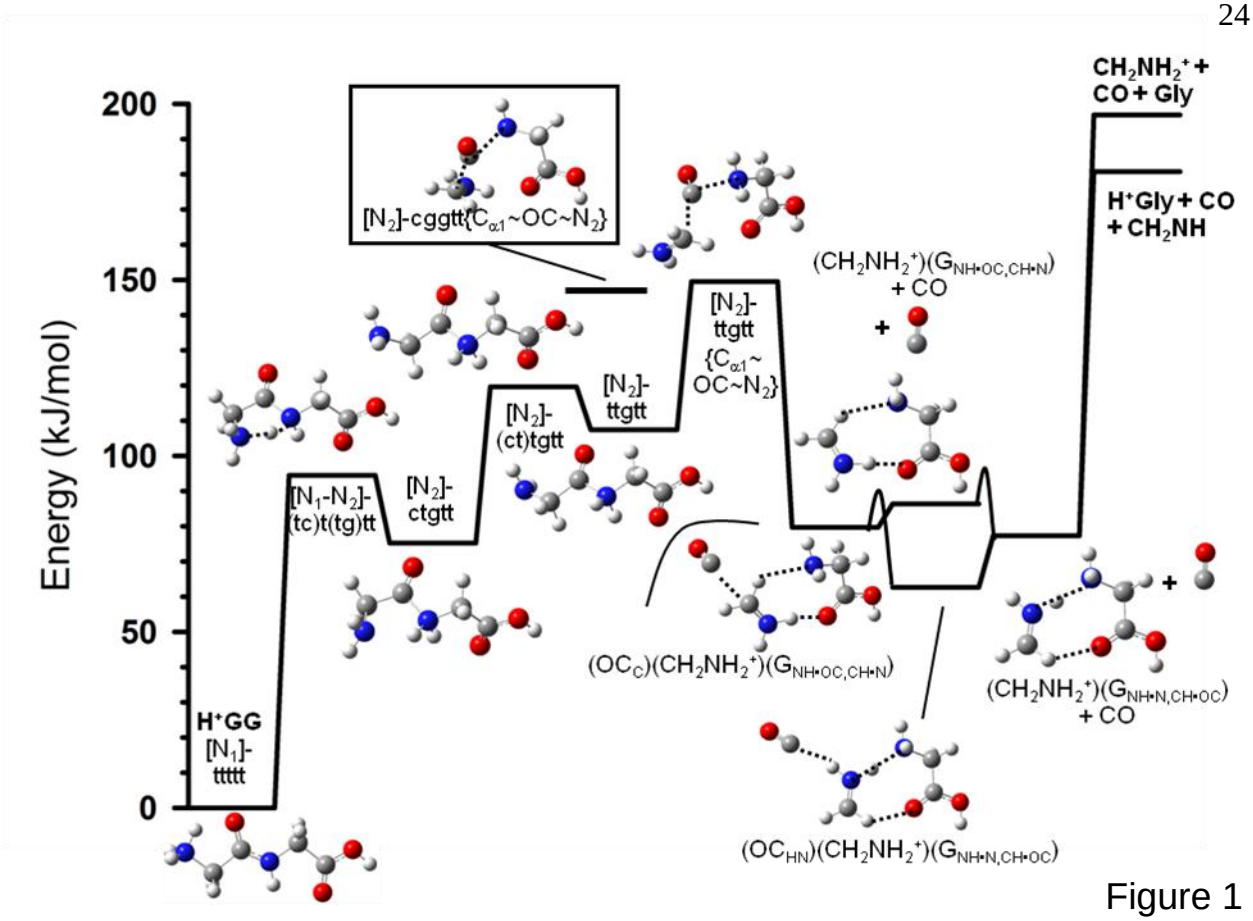


Figure 1

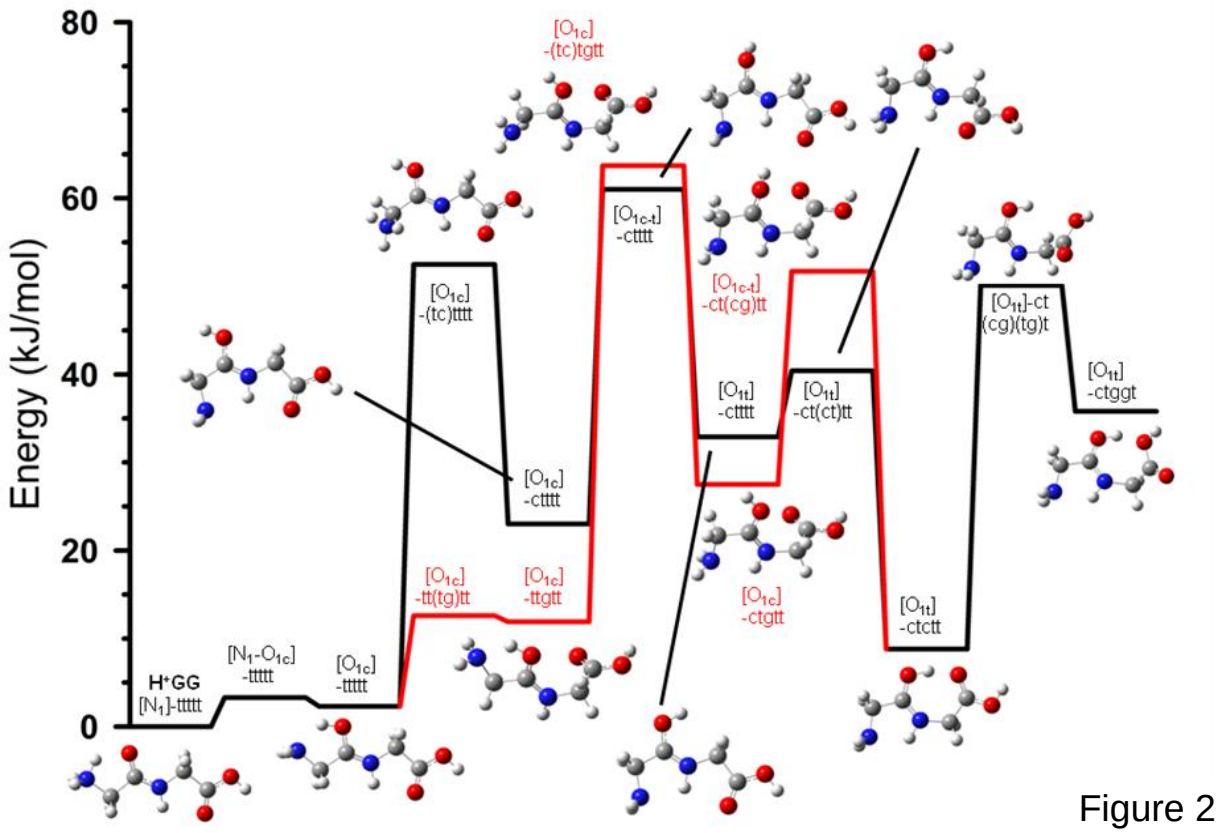


Figure 2

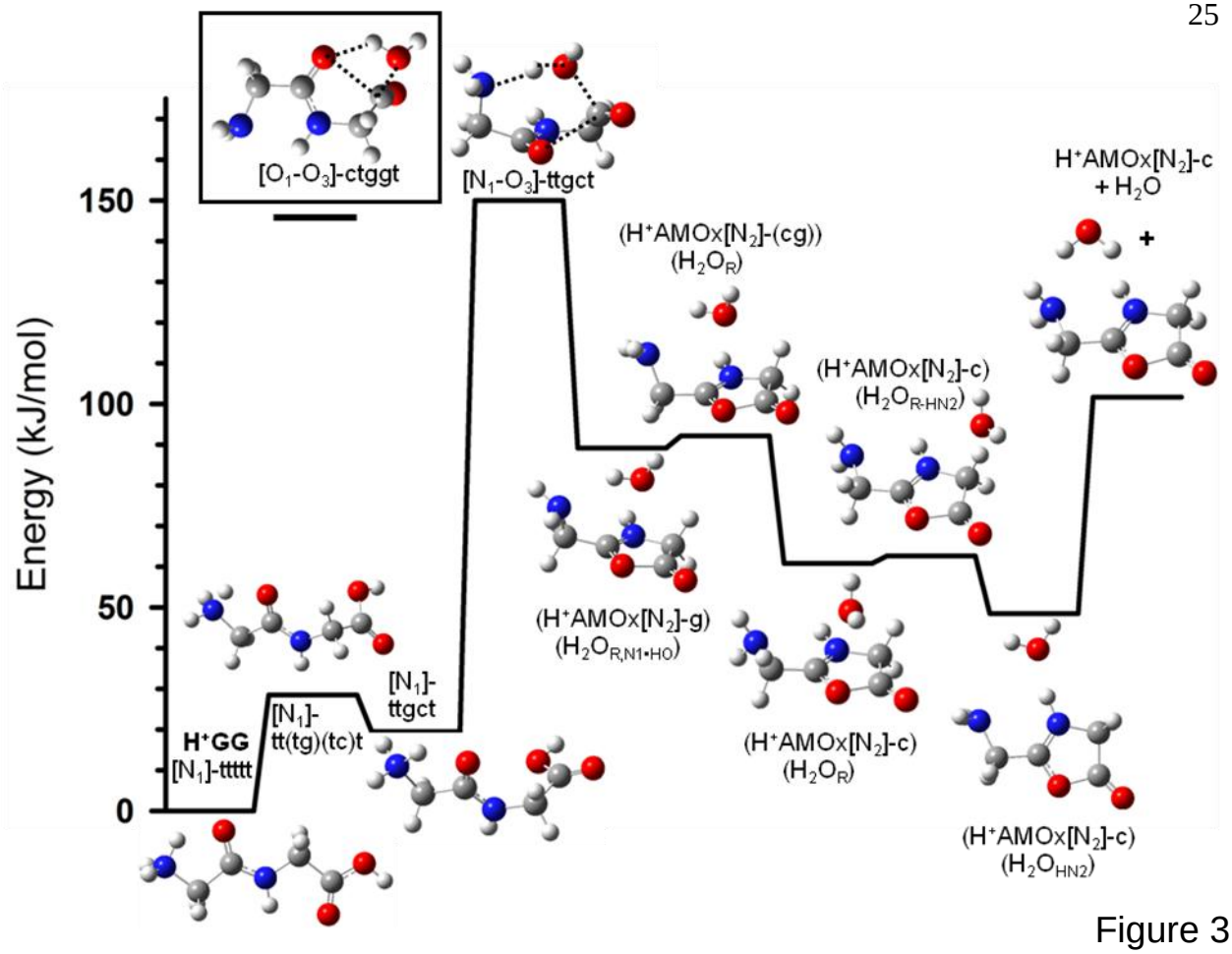


Figure 3

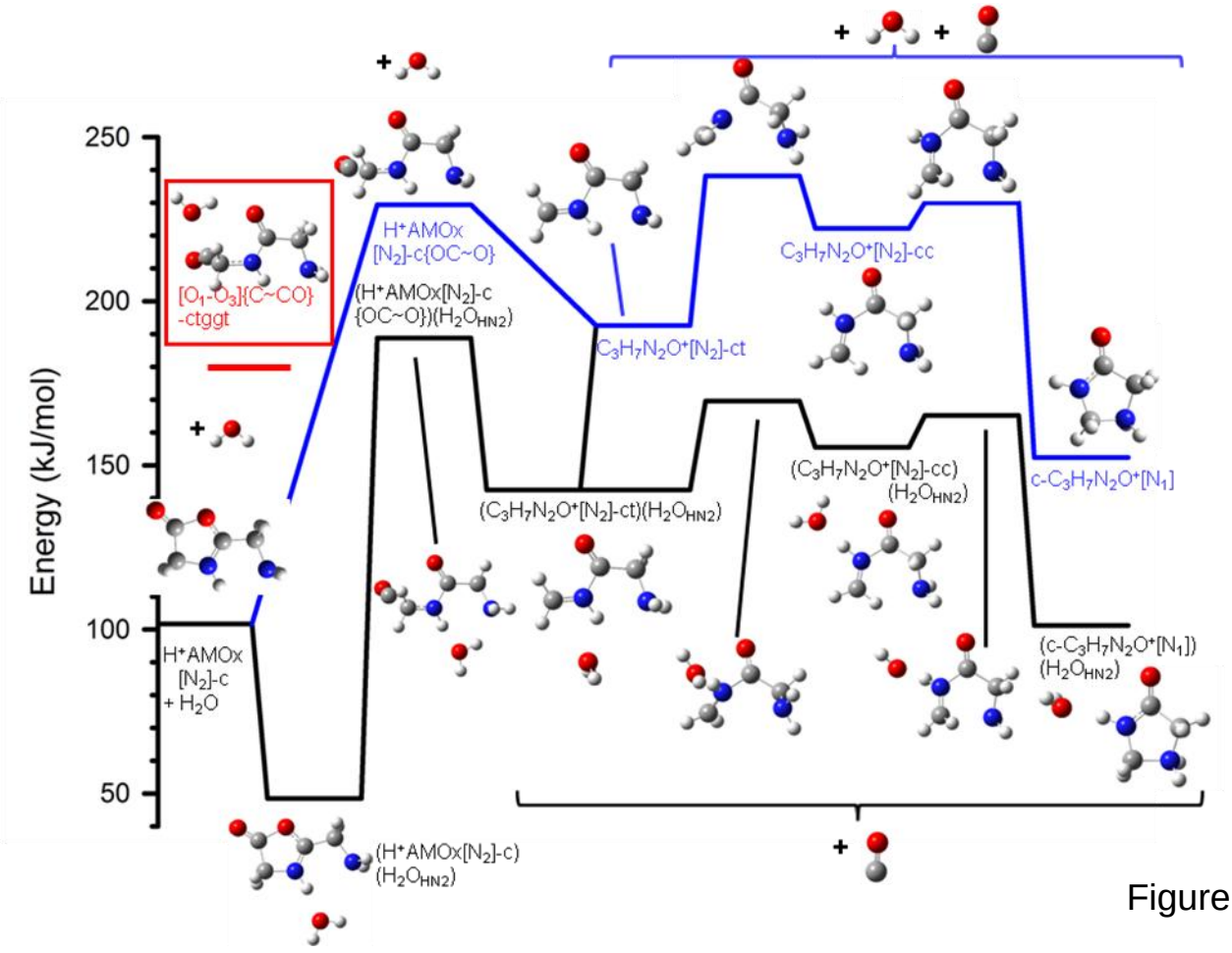


Figure 4

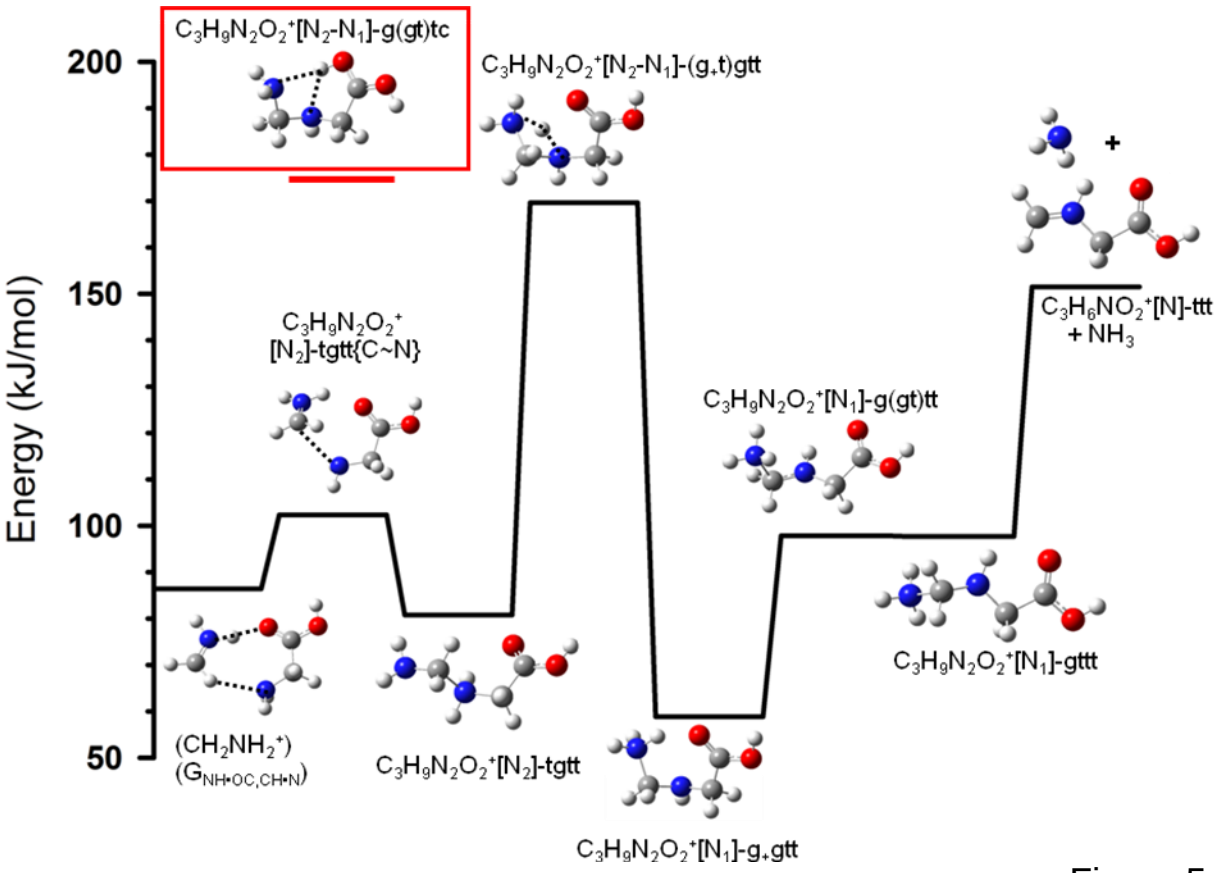

Energy (kJ/mol)
200
150
100
50
C3H9N2O2+[N2-N1]-g(gt)tc
C3H9N2O2+[N2-N1]-(g+t)gtt
C3H9N2O2+[N2]-tgtt{C~N}
C3H9N2O2+[N1]-g(gt)tt
C3H6NO2+[N]-ttt
+ NH3
(CH2NH2+)
(GNH•OC,CH•N)
C3H9N2O2+[N2]-tgtt
C3H9N2O2+[N1]-gttt
C3H9N2O2+[N1]-g+gtt

Figure 5

Supplementary Material for

Thermodynamics and Mechanisms of Protonated Diglycine Decomposition:

A Computational Study

P. B. Armentrout and A. L. Heaton

Proton Migration Before Decomposition

Among the additional structures located here that were not elucidated by PCLS and BAL, there are four additional [N_1] structures that all have the terminal hydroxyl group rotated (∠CCOH = c) compared to the five [N_1] structures listed by PCLS (∠CCOH = t), Figure S1. In general, this removes the hydrogen bond between the carbonyl and the hydroxyl groups (designated by CO•HO), such that these cis structures lie about 29 kJ/mol higher than the trans analogues, e.g., compare [N_1]-ttttt versus [N_1]-ttttc and [N_1]-ttgtt versus [N_1]-ttgtc, Figure S1. The [N_1]-gtgtt and [N_1]-gtgtc pair are closer in energy (22 kJ/mol) because loss of the CO•HO hydrogen bond allows a stronger N_1H•O_2C interaction, as indicated by a shorter hydrogen bond (1.93 versus 2.00 Å). Interestingly, when the CCOH dihedral angle of the [N_1]-tttct and [N_1]-ttgct conformers are rotated in attempts to form [N_1]-tttcc and [N_1]-ttgcc, respectively, both transform to [N_1]-ttggc, lying about 38 kJ/mol higher in energy. We also located six [N_1] conformers having a cis peptide bond (∠CCNC = c), Figure S1, which are included as they are possible intermediates for formation of a diketopiperazine b_2 product ion. Three of these were previously identified by BAL for the same purpose. In only one of these cases was a conformer with a cis ∠CCOH dihedral orientation located, [N_1]-gcgtc, and it lies 25 kJ/mol above its [N_1]-gcgtt counterpart. Other potential cis ∠CCOH conformers were left unexplored.

PCLS located six structures having the proton at the first oxygen, and BAL located another three. (One of these, [O_{1c}]-ttgct, could only be geometry optimized at the B3LYP/6-31+G(d,p) level. At the B3LYP/6-311+G(d,p) level used here, this structure collapses to [O_{1c}]-tttct as verified by a very careful examination of the reaction coordinate connecting these structures.) The present work located seven more [O_1] structures with trans peptide bonds

(∠CCNC) and nine with cis peptide bonds, Figure S1. None of the 25 [O_1] structures has a cis ∠CCOH dihedral angle, although such conformers should be stable for many of the [O_1] conformers.

As noted by PCLS, key species in the decomposition of H^+GG are the [N_2] conformers (which they designate as D), as cleavage of the central CN peptide bond can lead to y_1/a_1 product formation, as detailed in the main paper. The five lowest energy structures of this type were located by PCLS and lie 70 - 100 kJ/mol above the [N_1]-ttttt GS. We found another three structures of this type lying only slightly higher in energy, [N_2]-cgctt, [N_2]-$t_t$tgtt, and [N_2]-ttgtt. Two other N_2 structures are higher in energy because they have cis ∠CCOH termini, [N_2]-$t_t$tgtc and [N_2]-ttgtc. Comparing these last four conformations, the $t_t$tgtx orientation of the amine group is 4 kJ/mol lower in energy than ttgtx because the amine hydrogens are directed towards the carbonyl. No exploration of cis peptide bonds or other cis ∠CCOH conformers was attempted as such species are unlikely to be involved in the decomposition pathways observed.

Finally, PCLS found five [O_2] conformers (which they call G) as these are potentially key intermediates in the loss of water for the b_2/a_2 decomposition pathway. We reproduced their results for the G1 – G4 structures and found one additional structure, [O_2]-$t_t$tcgtg. PCLS also found a G5 structure with a ∠NCCN dihedral angle of 179.8°, but the closest species that we could geometry optimize was [O_2]-$g_t$tcgtt, with an angle of 137.5°, Figure S1, thereby allowing a stabilizing NH•OC interaction. In addition, we located four [O_2] structures having cis ∠CCNC peptide bonds, one of which had previously been optimized by BAL (cis4). These cis structures lie only slightly higher in energy than several of the trans conformers and are potential routes to formation of a b_2 product having a diketopiperazine structure, as discussed in the main paper.

With regard to location of transition states, PCLS identify two pathways between [N_1]-ttgtt (called **A2** by PCLS) and [N_1]-ttgct (**A5**), which we believe are better assigned to TS[N_1]-tt(gt)(tc)t (**A2_A4**) and TS[N_1]-tt(tg)(tc)t (**A1_A5**) on the basis of relaxed potential energy surfaces scans of these motions (IRC calculations were inconclusive). In agreement with our assignment, BAL identify TS[N_1]-tt(tg)(tc)t as **1_2** (the equivalent of **A1_A5**). In part, our

assignment relies on finding TS[N_1]-tt(tg)ct (**A4_A5**), which was not identified by PCLS and allows facile interchange of the [N_1]-tttct (**A4**) and [N_1]-ttgct (**A5**) structures. Indeed, this TS is 0.2 kJ/mol below (B3LYP) to 0.5 kJ/mol above (MP2) the ttgct conformer once zero point energies are included.

y_1/a_1 Formation: Rate Limiting Steps

In addition to the pathways for formation of the y_1/a_1 products elucidated in the main paper, a third route converts [N_2]-ctgtt (**D2**) to [N_2]-cttct (**D3**) followed by synchronous cleavage of the $C_{\alpha 1}$-CO and OC-N_2 bonds over TS[N_2]-cttct{$C_{\alpha 1}$~OC~N_2}, which lies 187 – 214 kJ/mol above the GS (or 38 – 46 kJ/mol above the two lower energy pathways). We did find one pathway in which the two bonds do not break synchronously that is a variant of the pathway elucidated by PS. From [N_2]-cggtt (**D1**), cleavage of the peptide bond along a path that forms a NH•N hydrogen bond goes over TS[N_2]-cggtt{OC~N_2} to yield an acylium ion ($H_2NCH_2CO^+$) stabilized by binding to glycine. The ($H_2NCH_2CO^+$)(G) complex lies only 3 – 6 kJ/mol lower in energy than the barrier, and its decomposition into a (CO)($CH_2NH_2^+$)(G) complex requires only -0.5 – 2.6 kJ/mol, such that cleavage of the $C_{\alpha 1}$~CO bond occurs readily and the initial C-N bond cleavage is rate-limiting. This TS lies between 7 – 9 kJ/mol above (B3LYP) to 8 kJ/mol below (MP2) the energies of TS[N_2]-tggtt{$C_{\alpha 1}$~OC~N_2} and TS[N_2]-cggtt{$C_{\alpha 1}$~OC~N_2}; however, the 298 K free energy of this TS increases by 5 – 7 kJ/mol relative to these TSs and its entropy of activation is 17 J/K mol. Thus, it is much tighter than the other two TSs and therefore is unlikely to contribute appreciably to the observed product formation.

y_1/a_1 Formation: Products

In all of the (OC_C)($CH_2NH_2^+$)(G) complexes, the OC_C molecule can bind on the opposite side of the plane defined by the immonium ion. Table S1 includes two such variations, distinguished by u (up) or d (down). Because the glycine is not symmetric above and below this plane, these complexes have slightly different energies, although within 0.4 kJ/mol according to the present results.

UU IR Author Manuscript

In addition to the $(OC)(CH_2NH_2^+)(G)$ complexes, there also exist $(OC)(CH_2NH)(H^+G)$ complexes in which the NH•N shared proton shifts to N•HN. These were investigated only for the two lowest energy complexes, those which have OC_{HN} connectivity. In both cases, shifting the proton to the glycine increases the energy by only 4 – 10 kJ/mol; however, with zero point energies included, the TSs between these forms lie 4 – 8 kJ/mol *below* the respective $(OC)(CH_2NH)(H^+G)$ complexes (and within 3 kJ/mol of the respective $(OC)(CH_2NH_2^+)(G)$ complexes). Thus, the N•HN species collapse to their NH•N forms, which seems likely in other conformations as well.

b_2 Formation: Oxazolone – Rate Limiting Steps

A low energy pathway for loss of H_2O from H^+GG to form the oxazolone b_2 ion has previously been elucidated by BAL [10] and the energetics detailed there are reproduced here within 11 kJ/mol. This mechanism starts with the $[N_1]$-ttttt GS (called **1** by BAL, **A1** by PCLS) and transfers the proton to the amide oxygen via TS$[N_1$-$O_1]$-ttttt (**1_3, A1_O1**) to form $[O_{1c}]$-ttttt (**3**, **O1**). Rotation of the C-terminal CC bond passes over TS$[O_{1c}]$-ttt(tc)t (**3_4**), which leads to $[O_{1c}]$-ttgct (**4**) according to BAL. This species is not stable at the B3LYP/6-311+G(d,p) level of theory used here and collapses to $[O_{1c}]$-tttct (**O4**), 4 – 5 kJ/mol lower in energy. $[O_{1c}]$-ttgct can be optimized at the B3LYP/6-31+G(d,p) level used by BAL, however even at this level, the barrier to forming $[O_{1c}]$-tttct is only 0.5 kJ/mol before including zero point energies and disappears once zero point energies are included. An alternative pathway to the $[O_{1c}]$-tttct intermediate starts with the GS, forms $[N_1]$-ttgtt (**A2**) followed by $[N_1]$-tttct (**A4**), which then transfers the proton to O_1 over TS$[N_1$-$O_1]$-tttct (**A4_O4**), 20 – 25 kJ/mol above the GS. We find the highest energy step of this route at the B3LYP/6-311+G(d,p) level to be 5 kJ/mol lower in energy than that for the pathway elucidated by BAL. In either case, the transformations needed to form $[O_{1c}]$-tttct are much lower in energy than the next step, 99 – 103 (102, BAL) kJ/mol, in which the proton in $[O_{1c}]$-tttct is rotated from cis to trans over TS$[O_{1c\text{-}t}]$-(tg)t(tg)(cg)t (**4_5**). This forms $[O_{1t}]$-gtggt (**5**) in which the proton is now pointed at the hydroxyl oxygen. Next, rotation of the CC bond in the N-terminal peptide over a very small barrier (0.7 kJ/mol) at TS$[O_{1t}]$-

UU IR Author Manuscript

(gc)tggt (**5_6**) leads to [O_{1t}]-ctggt (**6**, **O6**), Figure S1, which is the critical precursor for water loss.

BAL also elucidated a parallel pathway to formation of the oxazolone, differing only in that the N-terminus has a cis orientation relative to the amide oxygen. Because there is no longer a N_1•HN_2 hydrogen bond, [O_{1t}]-ttggt (**7**) lies 38 – 42 (40, BAL) kJ/mol above [O_{1t}]-ctggt (**6**, **O6**). The rate-limiting TS[O_1-O_3]-ttggt (**7_comp5**) lies 28 – 31 (30) kJ/mol above TS[O_1-O_3]-ctggt (**6_comp4**). The initial product formed, (H^+AMOx[N_2]-g)(H_2O_R) (**comp5**), lies 30 – 34 (30) kJ/mol above (H^+AMOx[N_2]-c)(H_2O_R) and differs only in the orientation of the aminomethyl side-chain. The gauche orientation (∠NCCN = 127°) no longer allows a hydrogen bond between the amine and imine groups, hence the higher energy.

There are two high energy pathways for b_2 ion formation that were previously elucidated by PCLS [8]. In the first, the [N_1]-ttttt (**A1**) GS is converted to [N_2]-ctgct (**D5**), which can occur by several multistep pathways. The simplest converts [N_1]-ttttt (**A1**) → [N_1]-ttgtt (**A2**) → [N_1]-ttgct (**A5**) → [N_2]-ctgct (**D5**), where the proton transfer step, TS[N_1-N_2]-(tc)tgct (**A5_D5**) is rate-limiting, calculated to be 111 – 122 (115, PCLS) kJ/mol above the GS. Alternatively, the sequence [N_1]-ttttt (**A1**) → [N_1]-ttgtt (**A2**) → [N_1]-gtgtt (**A3**) → [N_2]-cggtt (**D1**) → [N_2]-ctgtt (**D2**) → [N_2]-cttct (**D3**) → [N_2]-ctgct (**D5**) can be followed. The steps interconverting the various [N_2] conformers all lie relatively high in energy, but TS[N_2]-ct(gt)(tc)t (**D2_D3**) is the highest at 105 – 112 (109, PCLS) kJ/mol. As noted by PCLS, several of the conversions can occur by dihedral angle rotations in both directions such that there are two TSs having similar energies connecting several of these intermediates. In all cases, these various transformations require much less energy than the conversion of [N_2]-ctgct (**D5**) to the oxazolone product, (H^+AMOx[N_2]-c)(H_2O_R) (**B1**, **comp4**), over TS[N_2-O_3]-ctgct (**B1_D5, 8_comp4**), which PCLS (BAL) calculate requires 187 (185) kJ/mol, in agreement with the range determined here of 174 – 185 kJ/mol. Note this is 32 – 34 (28, BAL) kJ/mol above TS[O_1-O_3]-ctggt (**6_comp4**) and 29 – 39 (20, BAL) kJ/mol above TS[N_1-O_3]-ttgct (**2_comp2**), the pathways elucidated in the main paper.

UU IR Author Manuscript

The second route to the b_2 ion found by PCLS notes that [N_2]-ctgtt (**D2**) can transfer the proton from N_2 to the terminal carbonyl forming [O_2]-ctcgtg (**G1**). Transfer of a proton from one hydroxyl group to another passes through TS[O_2-O_3]-ctcgtg (**G1_B1, 9_comp4**), 246 – 252 (250, 248) kJ/mol above the GS or 94 – 111 (84 and 92, BAL) kJ/mol higher than the two low energy pathways identified above.

b_2 Formation: Diketopiperazine

BAL located a pathway for formation of H^+DKP that starts with the [N_1]-ttttt GS (1) and rotates the peptide bond over TS[N_1]-t(tc)ttt (1_cis1), yielding [N_1]-tcttt (cis1), only 17 – 18 (18) kJ/mol higher in energy. As BAL point out, the TS for this rotation is not that high in energy, 97 – 100 (102) kJ/mol above the GS, such that this transformation can occur at the higher energies required for fragmentation. From [N_1]-tcttt, the system passes through [N_1]-tcgtt (**cis2**) and [N_1]-gcgtt (**cis3**), but these transformations are all low in energy (<25 kJ/mol above the GS). Although not explored by BAL, the peptide bond can also be rotated from other conformations. Examination of alternate pathways finds that the lowest energy pathway to [N_1]-gcgtt, Figure S2, has a rate-limiting step of TS[N_1]-g(ct)gtt at 59 – 73 kJ/mol above the GS, or 26 – 38 kJ/mol below TS[N_1]-t(tc)ttt (**1_cis1**). This TS is lower in energy because hydrogen bonds between the protonated N-terminus and both carbonyls are maintained throughout. This result further underscores the observation of BAL that the trans-cis rotation of the peptide bond is not the limiting step in formation of H^+DKP.

As shown in Figure S3, from [N_1]-gcgtt (**cis3**), the proton is transferred from N_1 to O_2 via TS[N_1-O_2]-(gc)c(gc)t(tg)c (**cis3_cis4**), which lies 132 – 135 (139) kJ/mol above the GS. As the proton is transferred, a C-N bond is formed yielding the six-membered ring complex, [O_2]-ccgtgc (an intermediate conformer not located by BAL), which easily rearranges over TS[O_2]-cc(cg)tgc, only 0.6 – 1.0 kJ/mol higher in energy, to form [O_2]-ccctgc (**cis4**). From [O_2]-ccctgc, which lies 79 – 106 (110) kJ/mol above the GS, one of the two hydrogen atoms on N_1 can be transferred to either hydroxyl group. BAL located TS[N_1-O_2]-ccctgc (**cis4_comp1**) (inset in Figure S3), which lies 222 – 247 (244) kJ/mol above the GS. Of nearly identical energy is

TS[N_1-O_3]-ccctgc, Figure S3, which lies at 223 – 246 kJ/mol, 0.5 – 0.7 kJ/mol lower in energy at B3LYP and 0.9 kJ/mol higher at MP2. Either of these steps is the rate-limiting step in the formation of H^+DKP and leads directly to (H^+DKP[O_c])(H_2O_{HO}) (**comp1**), the lowest energy (H^+DKP)(H_2O) complex. Here, water is hydrogen bound to DKP protonated on one of the oxygens, and the proton on the oxygen is in a cis orientation (away from the peptide bond, where the nomenclature used follows the same path along the backbone as used for H^+GG complexes). This complex lies only 5 – 9 (6, BAL) kJ/mol above the [N_1]-ttttt GS and 35 – 43 (38) kJ/mol below the lowest energy (H^+AMOx)(H_2O) complex. Other variants of this complex have the proton in a trans orientation, (H^+DKP[O_t])(H_2O_{HO}), or with the water bound to the ring, (H^+DKP[O_c])(H_2O_R). These complexes lie 2 – 3 and 34 – 44 kJ/mol, respectively, higher in energy than (H^+DKP[O_c])(H_2O_{HO}). The latter complex requires only 77 – 84 kJ/mol to form H^+DKP[O_c] + H_2O, well below the rate-limiting TS, Figure S3.

Another route to DKP elucidated here involves [O_{1c}]-ccgtt as the key precursor, Figure S1. This complex can be formed through several routes. Starting with the [N_1]-ttttt GS (**1**), the most direct route starts with rotation of the peptide bond (which requires 97 – 100 (102) kJ/mol) to form [N_1]-tcttt (**cis1**), lying 17 – 18 (18) kJ/mol above the GS, followed by rotation about the NC bond of the second residue, 23 – 25 (28) kJ/mol above GS, leading to [N_1]-tcgtt (**cis2**), 10 – 13 (13) kJ/mol above GS. A lower energy but more complicated pathway to this complex follows [N_1]-ttttt $\rightarrow$ [N_1]-ttgtt $\rightarrow$ [N_1]-gtgtt, then rotates the peptide bond through TS[N_1]-g(ct)gtt at 59 – 73 kJ/mol. This forms [N_1]-gcgtt (**cis3**) at 5 – 15 (18) kJ/mol above GS, which rotates about the NCCN bond, TS[N_1]-(gt)cgtt (**cis2_cis3**) at 10 – 17 (21) kJ/mol, to form [N_1]-tcgtt (**cis2**). From this complex, proton transfer from N to O over TS[N_1-O_1]-tcgtt, only 17 – 18 kJ/mol above the GS, forms [O_{1c}]-tcgtt. However, this complex lies 0.5 – 2.9 kJ/mol lower than the TS once zero point energies are included, such that it collapses back to the N_1 form. From [O_{1c}]-tcgtt, rotation about the NCCN bond, which requires 69 – 71 kJ/mol, leads to [O_{1c}]-ccgtt, 58 – 59 kJ/mol above the GS.

UU IR Author Manuscript

Alternatively, one can follow $[N_1]$-ttttt → $[N_1]$-ttgct → $[N_1]$-tttct and rotate the peptide bond (TS$[N_1]$-t(tc)tct at 110 – 114 kJ/mol) to form $[N_1]$-tctct at 31 – 32 kJ/mol above the GS. Transfer of the proton to O_1 through TS$[N_1$-$O_1]$-tctct at 32 – 36 kJ/mol leads to $[O_{1c}]$-tctct at 30 – 36 kJ/mol above the GS. Rotation about the NCCN bond over TS$[O_{1c}]$-(tg)ctct (80 – 85 kJ/mol) forms $[O_{1c}]$-gctct (72 – 77 kJ/mol), which then passes over TS$[O_{1c}]$-(cg)c(gt)(tc)t (80 – 84 kJ/mol) to again form $[O_{1c}]$-ccgtt. Overall, the three pathways to formation of this complex have rate-limiting TSs below ~100 kJ/mol with the lowest energy pathway, $[N_1]$-ttttt → $[N_1]$-ttgtt → $[N_1]$-gtgtt → $[N_1]$-gcgtt → $[N_1]$-tcgtt → $[O_{1c}]$-tcgtt → $[O_{1c}]$-ccgtt, requiring only 71 – 73 kJ/mol.

The $[O_{1c}]$-ccgtt conformer is oriented such that the N terminal nitrogen atom can form a bond with the carbon of the carboxylic acid. The TS for formation of the six-membered ring synchronously transfers a hydrogen atom from the N to the C terminal carbonyl. The $[N_1$-$O_2]$-cc(gc)(tg)t transition state lies 244 – 260 kJ/mol above the GS (12 – 22 kJ/mol above TS$[N_1$-$O_{2,3}]$-ccctgc) and forms $[O_1,O_2,$-$N_1]$-cccgtg, Figure S1, which lies 80 – 105 kJ/mol above the GS. The awkward nomenclature indicates that all three oxygen atoms are now hydroxyl groups with the N-terminus having only one hydrogen rather than two. From this complex, loss of water can proceed by transfer of a hydrogen atom to one of the gem-diol hydroxyl groups. This can occur either from the hydroxyl group on the same carbon (TS$[O_3$-$O_2]$-cccgtg) or the one across the ring (TS$[O_1$-$O_2]$-cccgt). These routes are comparable in energy lying 250 – 252 and 244 – 264 kJ/mol above the GS, respectively. These TSs lie 4 – 28 and 17 – 22 kJ/mol, respectively, higher than the TS$[N_1$-$O_{2,3}]$-ccctgc transition states of Figure S3. In either case, a complex of DKP protonated on a carbonyl oxygen with a water molecule bound to the ring, H^+DKP$[O_1](H_2O_R)$, is formed. This complex lies only 43 – 51 kJ/mol above the GS, about the same energy as the most stable $(H^+$AMOx$)(H_2O)$ complex, but 34 – 44 kJ/mol above $(H^+$DKP$[O_c])(H_2O_{HO})$. Loss of water from this complex requires only 35 – 49 kJ/mol and lies well below the rate-limiting TSs.

UU IR Author Manuscript

Table S1. Relative theoretical energies at 0 K of H^+GG intermediates (kJ/mol)

species	B3LYP[a]	B3LYP// B3LYP[a]	MP2(full) //B3LYP[a]	Literature[b]
$H^+GG[N_1]$-ttttt	0.0	0.0	0.0	0.0 (A1), *0.0 (1)*
$H^+GG[N_1]$-ttgtt	6.2	7.8	3.7	8.9 (A2)
$H^+GG[N_1]$-tcgtt	12.6	13.0	9.9	*13.0 (cis2)*
$H^+GG[N_1]$-gcgtt	13.0	15.4	5.3	*17.6 (cis3)*
$H^+GG[N_1]$-tttct	16.9	18.8	17.6	19.0 (A4)
$H^+GG[N_1]$-tcttt	18.2	17.9	16.8	*18.4 (cis1)*
$H^+GG[N_1]$-gtgtt	18.6	21.3	9.8	17.7 (A3)
$H^+GG[N_1]$-ttgct	19.5	21.0	19.0	22.4 (A5), *17.6 (2)*
$H^+GG[N_1]$-tcgct	28.8	30.8	25.5	
$H^+GG[N_1]$-ttttc	29.4	26.2	26.8	
$H^+GG[N_1]$-tctct	30.9	32.2	30.8	
$H^+GG[N_1]$-ttgtc	35.3	33.5	29.9	
$H^+GG[N_1]$-gcgtc	38.3	38.2	28.2	
$H^+GG[N_1]$-gtgtc	40.9	41.1	29.7	
$H^+GG[N_1]$-ttggc	56.0	55.0	51.2	
$H^+GG[O_{1c}]$-ttttt	2.3	2.6	8.0	5.0 (O1), *2.5 (3)*
$H^+GG[O_{1t}]$-ctctt	8.8	8.3	14.3	10.8 (O2)
$H^+GG[O_{1c}]$-ttgtt (collapses to $[N_1]$)	11.9	13.1	15.3	17.4 (O3)
$H^+GG[O_{1c}]$-tcgtt (collapses to $[N_1]$)	17.8	18.6	20.8	
$H^+GG[O_{1c}]$-tttct	19.9	22.4	26.6	25.0 (O4)
$H^+GG[O_{1c}]$-ctttt	23.0	22.1	25.0	30.4 (O5)
$H^+GG[O_{1c}]$-ttgct (collapses to tttct)[c]	24.5	26.7	30.5	*20.1 (4)*
$H^+GG[O_{1c}]$-ctgtt	27.5	27.5	27.5	
$H^+GG[O_{1c}]$-tctct	30.2	31.9	36.0	

$H^+GG[O_{1c}]$-tcgct	32.3	34.3	35.5	
$H^+GG[O_{1t}]$-ctttt	32.9	32.0	34.7	
$H^+GG[O_{1t}]$-ctggt	35.8	39.5	39.8	40.8 (O6), *38.5 (6)*
$H^+GG[O_{1t}]$-gtgtt	48.6	45.8	52.9	
$H^+GG[O_{1t}]$-$g_t$tgtt	49.3	46.2	55.8	
$H^+GG[O_{1c}]$-ccttt	56.0	54.2	58.5	
$H^+GG[O_{1c}]$-ccgtt	58.4	58.1	58.0	
$H^+GG[O_{1t}]$-gtttt	63.4	60.7	66.1	
$H^+GG[O_{1t}]$-$g_t$tttt	67.4	63.9	71.0	
$H^+GG[O_{1t}]$-ccgtt	69.2	68.9	69.9	
$H^+GG[O_{1c}]$-gctct	72.3	73.0	77.1	
$H^+GG[O_{1t}]$-gtggt	73.0	74.7	76.6	*74.5 (5)*
$H^+GG[O_{1c}]$-ccgc$_-$t	74.3	75.0	76.5	
$H^+GG[O_{1c}]$-ccgc$_+$t	74.7	75.0	80.0	
$H^+GG[O_{1t}]$-$t_t$tggt	76.2	77.6	81.6	*78.2 (7)*
$H^+GG[O_{1t}]$-ggctt	148.1	146.2	135.8	
$H^+GG[N_2]$-cggtt	73.6	76.9	62.8	70.5|76.6 (D1)
$H^+GG[N_2]$-ctgtt	75.2	77.3	72.6	73.8 (D2)
$H^+GG[N_2]$-cttct	95.1	100.2	94.3	96.1 (D3)
$H^+GG[N_2]$-ctg$_-$ct	96.7	101.7	88.5	96.1 (D4)
$H^+GG[N_2]$-ctg$_+$ct	97.8	102.7	93.0	98.1 (D5), *96.2 (8)*
$H^+GG[N_2]$-cgctt	102.8	105.6	92.5	
$H^+GG[N_2]$-$t_t$tgtt	107.4	106.1	104.5	
$H^+GG[N_2]$-ttgtt	111.4	109.9	107.3	
$H^+GG[N_2]$-$t_t$tgtc	138.5	133.9	132.8	
$H^+GG[N_2]$-ttgtc	142.4	137.5	135.7	

$(OCCH_2NH_2^+)(G_{NH•N})$	153.5	157.3	154.1	
$H^+GG[O_2]$-ctcgtg	71.4	69.4	64.1	72.8 (G1)
$H^+GG[O_2]$-ctcgtt	72.3	69.9	63.8	74.5 (G2)
$H^+GG[O_2]$-gtcgtg	98.3	95.0	91.0	102.4 (G3)
$H^+GG[O_2]$-$g_t$tcgtg	102.5	98.5	95.8	105.9 (G4)
$H^+GG[O_2]$-$t_t$tcgtg	102.9	99.1	96.8	
$H^+GG[O_2]$-$g_t$tcgtt	103.1	98.9	95.7	107.4 (G5)
$H^+GG[O_2]$-ccctgc	104.0	105.5	78.7	*109.6 (cis4)*
$H^+GG[O_2]$-ccgtgc	106.2	108.0	83.4	
$H^+GG[O_2]$-ccgttc	110.0	111.5	85.2	
$H^+GG[O_1,O_2,\text{-}N_1]$-cccgtg	104.9	102.1	80.0	
$(OC_{HN})(CH_2NH_2^+)(G_{NH•N,CH•OC})$	62.6 (0.0)	66.5 (0.0)	64.2 (0.0)	
$(OC_{HN})(CH_2NH_2^+)(G_{NH•N,NH•OC})$	65.7 (3.1)	69.2 (2.7)	64.3 (0.1)	
$(OC_{HC})(CH_2NH_2^+)(G_{NH•N,NH•OC})$	69.9 (7.3)	73.8 (7.3)	73.5 (9.3)	
$(OC_{HN})(CH_2NH)(H^+G_{N•HN,CH•OC})$	70.4 (7.8)	73.1 (6.6)	68.3 (4.1)	
$(OC_{HN})(CH_2NH_2^+)(G_{NH•OC,CH•N})$	70.6 (8.0)	73.1 (6.6)	72.7 (8.5)	
$(OC_{Cd})(CH_2NH_2^+)(G_{NH•N,\ CH•OC})$	70.9 (8.3)	74.9 (8.4)	73.3 (9.1)	
$(OC_{Cu})(CH_2NH_2^+)(G_{NH•N,CH•OC})$	71.3 (8.7)	75.3 (8.8)	73.7 (9.5)	89.1|82.0 (CO_PBD_B1)
$(OC_C)(CH_2NH_2^+)(G_{NH•N,NH•OC})$	71.3 (8.7)	75.2 (8.7)	74.0 (9.8)	89.1|82.0 (CO_PBD_B2)
$(OC_{HN})(CH_2NH)(H^+G_{N•HN})$	74.0 (11.4)	75.7 (9.2)	74.7 (10.5)	
$(OC_{Cu})(CH_2NH_2^+)(G_{NH•OC,CH•N})$	79.7 (17.1)	82.3 (15.8)	82.5 (18.3)	101.7|91.6 (CO_PBD_B3)
$(OC_{Cd})(CH_2NH_2^+)(G_{NH•OC,CH•N})$	79.8 (17.2)	82.4 (15.9)	82.7 (18.5)	
$(OC_C)(CH_2NH_2^+)(G_{NH•N})$	99.1 (36.5)	103.4 (36.9)	104.1 (39.9)	
$(CH_2NH_2^+)(G_{NH•N,CH•OC})$ + CO	77.3 (0.0)	81.5 (0.0)	85.7 (0.0)	97.9|87.9 (PBD_B1)
$(CH_2NH_2^+)(G_{NH•N,NH•OC})$ + CO	77.9 (0.6)	81.9 (0.4)	86.3 (0.6)	98.3|88.3 (PBD_B2)
$(CH_2NH)(H^+G_{N•HN,CH•OC})$ + CO	79.9 (2.6)	82.5 (1.0)	83.0 (-2.7)	99.2|90.0 (PBD_B1a)

$(CH_2NH)(H^+G_{N\bullet HN})$ + CO	82.7 (5.4)	84.3 (2.8)	88.9 (3.2)	102.1|93.3 (PBD_B2a)
$(CH_2NH_2^+)(G_{NH\bullet OC,NH\bullet N})$ + CO	84.5 (7.2)	86.9 (5.5)	93.2 (7.5)	
$(CH_2NH_2^+)(G_{NH\bullet OC,CH\bullet N})$ + CO	86.4 (9.1)	89.2 (7.7)	95.3 (9.6)	111.3|97.9 (PBD_B3)
$(CH_2NH_2^+)(G_{ZW,C\bullet OC})$ + CO	92.1	93.2	95.7	
$(CH_2NH)(H^+G_{N\bullet HN,CH\bullet OH})$ + CO	98.6	102.9	104.3	
$(CH_2NH_2^+)(G_{ZW,C\bullet OH})$ + CO	111.5	115.9	114.8	
$(CH_2NH)(H^+G_{N\bullet HO})$ + CO	118.1	115.4	121.6	
H^+G + CH_2NH + CO	180.7 (103.4)	179.5 (98.0)	189.0 (103.3)	207.5|193.3, *193.3*
$CH_2NH_2^+$ + G + CO	196.8 (119.5)	197.6 (116.1)	210.8 (125.1)	243.9|215.5
$CH_2NH_2^+$ + H_2O + CH_2NH + 2 CO	318.4	316.2	326.5	
$(H^+AMOx[N_2]\text{-}c)(H_2O_{HN2})$	48.5 (0.0)	50.0 (0.0)	43.8 (0.0)	*44.4 (comp3)*
$(H^+AMOx[N_1]\text{-}c)(H_2O_{HN1})$	55.0 (6.5)	57.4 (7.4)	43.7 (-0.1)	
$(H^+AMOx[N_2]\text{-}c)(H_2O_R)$	60.8 (12.3)	62.2 (12.2)	50.0 (6.2)	81.3 (B1), *61.1 (comp4)*
$(H^+AMOx[N_2]\text{-}c)(H_2O_{HN1})$	63.6 (15.1)	64.5 (14.5)	60.4 (16.6)	
$(H^+AMOx[N_2]\text{-}g)(H_2O_{R,N1\bullet HO})$	89.2 (40.7)	90.8 (40.8)	76.3 (32.5)	*104.2 (comp2)*
$(H^+AMOx[N_2]\text{-}g)(H_2O_R)$	92.7 (44.2)	92.4 (52.4)	83.7 (39.9)	*90.8 (comp5)*
$H^+AMOx[N_2]\text{-}c$ + H_2O	101.7 (53.2)	98.0 (48.0)	99.1 (55.3)	139.3, *105.4 (b2b+H_2O)*
$H^+AMOx[N_1]\text{-}c$ + H_2O	125.7 (77.2)	122.2 (72.2)	112.1 (68.3)	
$H^+AMOx[N_2]\text{-}g$ + H_2O	135.3 (86.8)	129.6 (79.6)	135.0 (91.2)	
$(c\text{-}C_3H_7N_2O^+[N_1])(H_2O_{HN1u})$ + CO	78.4 (0.0)	85.4 (0.0)	66.8 (0.0)	
$(c\text{-}C_3H_7N_2O^+[N_1])(H_2O_{HN1d})$ + CO	79.2 (0.7)	86.3 (0.9)	68.8 (2.0)	
$(c\text{-}C_3H_7N_2O^+[N_1])(H_2O_{HN2})$ + CO	101.2 (22.8)	107.2 (21.8)	89.7 (22.9)	
$(C_3H_7N_2O^+[N_1]\text{-}ct)(H_2O_{HN1})(OC_{C\alpha 2})$	122.7 (44.3)	129.3 (43.9)	118.5 (51.7)	
$(C_3H_7N_2O^+[N_1]\text{-}ct)(H_2O_{HN1})$ + CO	126.6 (48.2)	133.2 (47.8)	126.9 (60.1)	
$(C_3H_7N_2O^+[N_2]\text{-}ct)(H_2O_{HN2})$ + CO	142.5 (64.1)	148.8 (63.4)	142.1 (75.3)	
$(C_3H_7N_2O^+[N_2]\text{-}ct)(H_2O_{HC\alpha 2}\text{-}c)(OC_{C\alpha 2})$	143.6 (65.2)	150.2 (64.8)	141.2 (74.4)	

$(C_3H_7N_2O^+[N_2]\text{-ct})(H_2O_{HC\alpha 2}\text{-t})$ + CO	147.4 (69.0)	153.8 (68.4)	152.8 (86.0)	
$(C_3H_7N_2O^+[N_2]\text{-ct}(H_2O_{HC\alpha 2}\text{-c})$ + CO	151.0 (72.6)	157.7 (72.3)	156.4 (89.6)	
$(C_3H_7N_2O^+[N_2]\text{-ct})(H_2O_{HN1})$ + CO	153.5 (75.1)	159.6 (74.2)	156.5 (89.7)	
$(C_3H_7N_2O^+[N_2]\text{-cc})(H_2O_{HN2})$ + CO	155.5 (77.1)	161.8 (76.4)	156.9 (90.1)	
$(C_3H_7N_2O^+[N_1]\text{-ct})(H_2O_{HC\alpha 2})$ + CO	166.8 (88.4)	172.3 (86.9)	168.6 (101.8)	
$(C_3H_7N_2O^+[N_2]\text{-cc})(H_2O_{HC\alpha 2})$ + CO	181.4 (102.9)	186.1 (100.7)	179.1 (112.3)	
$(C_3H_7N_2O^+[N_2]\text{-cc})(H_2O_{HN1})$ + CO	184.5 (106.1)	189.8 (104.4)	186.9 (120.1)	
c-$C_3H_7N_2O^+[N_2]$ + H_2O + CO	152.4 (0.0)	153.4 (0.0)	141.0 (0.0)	
$C_3H_7N_2O^+[N_2]$-ct + H_2O + CO	192.6 (40.2)	194.0 (40.7)	195.8 (54.9)	
$C_3H_7N_2O^+[N_1]$-ct + H_2O + CO	194.4 (42.0)	195.4 (42.0)	194.5 (53.6)	
$C_3H_7N_2O^+[N_1]$-cc + H_2O + CO	222.2 (69.8)	223.0 (69.6)	224.6 (83.6)	
$(H^+DKP[O_c])(H_2O_{HO})$	5.7	8.5	9.0	*5.9 (comp1)*
$(H^+DKP[O_t])(H_2O_{HO})$	7.8	11.6	10.9	
$(H^+DKP[O_c])(H_2O_R)$	49.6	50.6	43.3	
$H^+DKP[O_c]$ + H_2O	89.9	86.1	91.8	*93.3 (b2a+H_2O)*
$C_3H_9N_2O_2^+[N_1]$-g_+gtt + CO	58.8	62.4	54.5	
$C_3H_9N_2O_2^+[N_2]$-tgtt + CO	80.8	82.2	73.5	
$C_3H_9N_2O_2^+[N_1]$-g_-gtt + CO	98.5	101.8	93.8	
$C_3H_9N_2O_2^+[N_1]$-gttt + CO	97.7	101.2	101.5	
$C_3H_6NO_2^+[N]$-gtc($NH_{3,HO}$) + CO	109.4	108.8	117.1	
$C_3H_6NO_2^+[N]$-gtt($NH_{3,HC}$) + CO	114.3	116.2	121.1	
$C_3H_6NO_2^+[N]$-ttt + NH_3 + CO	151.5	148.6	157.0	
$C_3H_6NO_2^+[N]$-gtt + NH_3 + CO	159.5	157.3	166.2	
$C_3H_6NO_2^+[N]$-gct + NH_3 + CO	167.6	166.7	174.2	
H^+DKOx[CO] + NH_3	174.4	169.0	190.7	
H^+Ox[N] + NH_3 + CO	210.1	205.8	203.9	

$H^+Ox[OC] + NH_3 + CO$	236.3	230.3	239.2
$H^+Ox[O_R] + NH_3 + CO$	303.3	303.1	295.8

[a] Values from the present study calculated at the B3LYP/6-311+G(d,p), B3LYP/6-311+G(2d,2p)//B3LYP/ 6-311+G(d,p), and MP2(full)/6-311+G(2d,2p)//B3LYP/6-311+G(d,p) levels of theory, including zero point energy corrections. Values in parentheses refer to the energy relative to a structurally similar species. ZW indicates the glycine has a zwitterionic $NH_3^+CH_2CO_2^-$ structure.

[b] Values from PS [9], B3LYP/6-31G(d) | B3LYP/6-31+G(d,p); PCLS [8], B3LYP/6-31G(d); and in italics from BAL [10], B3LYP/6-31+G(d,p). Designations in parentheses are the names used by these authors.

[c] Could only be optimized at the B3LYP/6-31+G(d,p) level.

UU IR Author Manuscript

Table S2. Relative theoretical energies at 0 K of H^+GG transition states (kJ/mol) and their imaginary frequencies (cm^{-1})

transition state[a]	B3LYP[b]	B3LYP// B3LYP[b]	MP2(full) //B3LYP[b]	imag freq (cm^{-1})	Literature[c]
$[N_1]$-tt(tg)tt	6.8	7.3	6.4	42	9.5 (A1_A2)
$[O_1]$-tt(tg)tt	12.6	13.5	17.8	45	(O1_O3)
$[N_1]$-(tg)cgtt	15.6	16.6	9.9	48	*21.3 (cis2_cis3)*
$[N_1]$-(tg)tgtt	18.8	20.9	10.1	62	18.5 (A2_A3)
$[N_1]$-tt(tg)ct	19.3	20.8	19.5	25	(A4_A5)
$[N_1]$-tc(gt)tt	24.7	24.6	23.1	45	*28.0 (cis1_cis2)*
$[N_1]$-tt(gt)(tc)t	28.2	29.0	27.6	52	31.0 (A2e_A5a)
$[N_1]$-tt(tg)(tc)t	28.5	29.3	27.9	49	31.4 (A2e_A5b) *31.4 (1_2)*
$[N_1]$-tc(gt)(tc)t	31.1	32.3	31.0	17	
$[O_{1c}]$-tt(gt)(tc)t	32.6	33.7	38.1	53	(O3_O4)
$[O_{1c}]$-ttt(tc)t	33.2	34.3	38.5	52	*36.0 (3_4)*
$[O_{1c}]$-tc(tg)(ct)t	33.6	34.3	37.1	41	
$[N_1]$-tc(tg)(tc)t	34.5	35.2	31.2	48	
$[O_{1t}]$-ct(ct)tt	40.4	39.9	42.1	43	
$[O_{1c}]$-tt(gt-tc)t	49.4	50.2	51.5	84	
$[O_{1t}]$-ct(cg)(tg)t	50.1	50.2	50.0	74	
$[O_{1c\text{-}t}]$-ct(cg)tt	51.7	52.1	52.9	543	
$[N_1]$-tttt(tc)	52.2	50.4	51.4	592	
$[O_{1c}]$-(tc)tttt	52.5	50.5	56.4	120	(O1_O5)
$[O_{1c\text{-}t}]$-ctttt	61.0	59.8	63.8	591	
$[O_{1c}]$-(tc)tgtt	63.7	62.5	65.5	126	
$[O_{1c}]$-cc(gt)tt	66.2	65.2	67.9	52	
$[O_{1c}]$-(tc)cgtt	70.7	69.1	71.3	125	

[N_1]-g(ct)gtt	70.9	73.2	58.8	134	
[O_{1t}]-(cg)tggt	73.7	75.6	77.5	81	*79.1 (5_6)*
[O_{1c}]-ccgc$_{(+/-)}$t	75.3	75.4	79.9	26	
[O_{1t}]-g$_{(+/-)}$tggt	78.3	79.0	83.6	56	
[O_{1c}]-ccg(c$_+$t)t	79.2	78.6	81.9	50	
[O_{1c}]-(cg)c(gt)(tc)t	80.5	80.0	83.7	54	
[O_{1c}]-(tg)ctct	80.9	80.7	85.0	159	
[O_{1c}]-cc(gt)(c–t)t	82.9	82.3	85.2	55	
[O_{1c-t}]-(tg)tttt	84.2	81.9	88.7	600	
[O_{1c}]-(tc)cgct	84.8	84.1	86.2	148	
[O_{1t}]-gt(tg)(tg)t	86.7	85.0	88.1	69	
[N_2]-c(gt)gtt	87.2	89.8	84.3	38	86.3 (D1_D2)
[O_{1t}]-(gt)tggt	87.7	88.2	92.8	303	
[O_{1c-t}]-(tg)$_+$tttt	94.2	91.2	97.9	605	
[N_2]-ct(tg)ct	98.0	102.9	96.1	40	99.3 (D3_D5)
[O_2]-(cg)tcgtg	98.4	95.2	91.4	48	(G1_G3)
[N_1]-t(tc)ttt	99.6	99.3	97.0	189	*102.1 (1_cis1)*
[O_{1c-t}]-(tg)t(tg)(cg)t	99.6	100.7	103.2	595	*101.7 (4_5)*
[N_2]-cg(gc)tt	102.5	105.5	94.9	51	
[N_2]-cg(gt)(cg)tt	106.1	108.2	94.2	44	
[O_2]-cc(cg)tgc	106.9	108.6	84.4	78	
[N_2]-ct(gt)(tc)_t	107.6	110.9	105.4	72	109.5 (D2_D3a)
[N_2]-ct(gt)(tc)$_+$t	108.5	112.0	106.1	69	108.8 (D2_D3b)
[O_2]-ccgttc$_{(+/-)}$	108.8	110.4	85.0	134	
[N_1]-t(tc)tct	112.5	113.8	110.1	215	
[N_2]-ct(gt)(tc)_t	118.1	122.3	108.2	29	(D2_D3)

$[N_2]$-(ct)tgtt	119.6	118.6	117.7	57	
$[N_1\text{-}O_1]$-ttttt	3.3	3.5	6.3	1015	4.1 (A1_O1) *2.9 (1_3)*
$[N_1\text{-}O_1]$-ttgtt	11.4	12.5	12.3	994	15.3 (A2_O3)
$[N_1\text{-}O_1]$-tcgtt	17.3	18.1	17.9	988	
$[N_1\text{-}O_1]$-tttct	20.6	22.8	24.4	1011	23.8 (A4_O4)
$[N_1\text{-}O_1]$-ttgct	24.2	25.9	27.1	1006	
$[N_1\text{-}O_1]$-tctct	32.4	34.0	35.5	1017	
$[N_1\text{-}O_1]$-tcgct	32.5	34.5	32.8	1008	
$[N_1\text{-}N_2]$-(gc)(tg)gtt	81.8	86.4	71.5	1145	78.0 (A3_D1)
$[N_1\text{-}N_2]$-(tc)t(tg)tt	94.5	98.1	92.7	1189	92.2 (A1_D2)
$[N_2\text{-}O_2]$-ct(gc)(tg)t	102.2	103.9	101.8	1074	104.4 (G1_D2)
$[N_1\text{-}N_2]$-(tc)tg_ct	109.0	115.5	100.2	1141	107.4 (A5_D4) *108.8 (2_8)*
$[N_1\text{-}N_2]$-(tc)tg_+ct	115.8	121.6	111.6	1193	115.2 (A5_D5)
$[N_1\text{-}O_2]$-(gc)c(gc)t(tg)c	133.3	135.0	132.2	148	*139.3 (cis3_cis4)*
$[O_1\text{-}O_3]$-ctggt $\rightarrow$ H^+AMOx	**146.2**	**150.6**	**141.5**	**143**	***156.9 (6_comp4)***
$[N_1\text{-}O_3]$-ttgct $\rightarrow$ H^+AMOx	**150.0**	**155.2**	**134.6**	**148**	***164.8 (2_comp2)***
$[O_1\text{-}O_3]$-ttggt $\rightarrow$ H^+AMOx	176.6	179.2	171.6	124	*187.0 (7_comp5)*
$[N_2\text{-}O_3]$-ctgct $\rightarrow$ H^+AMOx	178.7	184.4	173.9	572	187.3 (B1_D5) *184.9 (8_comp4)*
$[O_1\text{-}O_3]${C~CO}-ctggt $\rightarrow$ a_2	**180.9**	**187.1**	**179.1**	**269**	
$[N_1\text{-}O_3]$-ccctgc $\rightarrow$ H^+DKP	**243.1**	**245.9**	**223.5**	**1497**	
$[N_1\text{-}O_2]$-ccctgc $\rightarrow$ H^+DKP	**243.6**	**246.6**	**222.6**	**1466**	***243.5 (cis4_comp1)***
$[O_2\text{-}O_3]$-ctcgtg $\rightarrow$ H^+AMOx	251.6	249.4	245.9	1772	249.7 (G1_B1) *248.5 (9_comp4)*
$[O_3\text{-}O_2]$-cccgtg $\rightarrow$ H^+DKP	**252.3**	**250.6**	**250.5**	**1602**	
$[O_{1t}\text{-}N_2]$-c(ct)gtt	259.9	261.1	254.6	1743	
$[N_1\text{-}O_2]$-cc(gc)(tg)t	260.1	258.8	244.6	1760	
$[O_1\text{-}O_2]$-cccgt $\rightarrow$ H^+DKP	263.2	263.7	244.0	1145	

$[N_1\text{-}O_1,O_2]$-cc(cg)(gt)t(cg)	316.6	315.4	284.1	1652	
$[N_2]$-ttgtt{$C_{\alpha1}$~OC~N_2} → (-CO)	**149.5**	**152.9**	**168.4**	**176**	
$[N_2]$-cggtt{$C_{\alpha1}$~OC~N_2} → (-CO)	**150.7**	**154.1**	**167.7**	**237**	**159.0|160.7 (TS_B0)** ***154.0 (tsy1)***
$[N_2]$-cggtt{OC~N_2} → acylium	**158.9**	**161.7**	**160.0**	**83**	
$[N_1]$-tcgtt{C~N_1} → H^+DKOx	162.6	163.6	181.3	377	
$[N_2]$-cttct{$C_{\alpha1}$~OC~N_2}→ (-CO)	187.3	191.0	214.2	127	
$(OC_{HN})(CH_2NH)H^+(G_{N\bullet H\bullet N,CH\bullet OC})$	64.6	68.0	64.5	619	
$(OC_{HN})(CH_2NH)H^+(G_{N\bullet H\bullet N,NH\bullet OC})$	68.2	71.4	66.9	596	
$(OC_{HC\text{-}HN})(CH_2NH_2^+)(G_{NH\bullet N,NH\bullet OC})$	69.8	73.7	73.2	10	
$(OC_{C\text{-}HC})(CH_2NH_2^+)(G_{NH\bullet N,\ NH\bullet OC})$	71.3	75.2	75.0	22	
$(OC_{HN})(CH_2NH_2^+)(G_{NH\bullet N,NH\text{-}CH\bullet OC})$	73.0	76.0	72.7	32	
$(OC_{HN})(CH_2NH_2^+)(G_{NH\bullet OC\text{-}N})$	73.4	76.0	72.7	134	
$(CH_2NH)H^+(G_{N\bullet H\bullet N,CH\bullet OC})$ + CO	74.9	78.7	81.7	797	94.1|84.5 (TS_B1)
$(CH_2NH)H^+(G_{N\bullet H\bullet N,NH\sim OC})$ + CO	77.4	81.1	84.1	744	97.1|86.6 (TS_B2)
$(OC_{C\text{-}HN})(CH_2NH_2^+)(G_{NH\bullet OC,CH\bullet N})$	79.6	82.3	83.7	10	
$(OC_C)(CH_2NH_2^+)(G_{NH\bullet OC\text{-}N})$	82.1	84.9	83.0	168	
$(CH_2NH_2^+)(G_{NH\bullet N\text{-}OC,NH\bullet OC\text{-}N})$ + CO	84.8	87.7	94.3	45	
$(CH_2NH_2^+)(G_{NH\bullet N\text{-}OC,CH\bullet OC\text{-}NH\bullet N})$ + CO	84.9	87.7	92.6	99	
$(CH_2NH_2^+)(G_{NH\bullet N,CH\text{-}NH\bullet OC})$ + CO	87.3	90.6	94.2	87	
$(CH_2NH_2^+)(G_{NH\bullet N\text{-}OC,NH\bullet OC\text{-}CH\bullet N})$ + CO	89.6	92.6	96.4	173	
$(CH_2NH_2^+)(G_{NH\bullet OC,CH\text{-}NH\bullet N})$ + CO	92.0	94.2	101.4	72	
$(OC\sim CH_2NH_2^+)(G_{NH\bullet N})$	153.1	156.8	156.3	290	
H^+G(TS[N-O_2]-ct) + CH_2NH + CO	319.2	323.6	319.2	97	
H^+AMOx$[N_2]$-c($H_2O_{R\text{-}HN2}$)	62.7	63.9	59.5	20	
H^+AMOx$[N_2]$-c($H_2O_{R\text{-}HN1}$)	65.8	66.2	58.5	25	
H^+AMOx$[N_2]$-c($H_2O_{HN2\text{-}HN1}$)	68.0	68.0	61.6	53	

$H^+AMOx[N_1\text{-}N_2]\text{-}c(H_2O_{HN1})$	86.9	89.8	78.3	1249
$H^+AMOx[N_2]\text{-}(cg)(H_2O_R)$	92.2	93.3	81.6	94
$H^+AMOx[N_1\text{-}N_2]\text{-}c + H_2O$	138.9	136.6	128.6	1269
$H^+AMOx[N_2]\text{-}c\{OC\sim O\}(H_2O_{HN2})$	**188.8**	**193.5**	**180.1**	**349**
$H^+AMOx[N_1]\text{-}c\{OC\sim O\}(H_2O_{HN1})$	**190.7**	**195.4**	**185.7**	**396**
$H^+AMOx[N_2]\text{-}c\{OC\sim O\}(H_2O_{HN1})$	**196.5**	**201.2**	**192.5**	**348**
$H^+AMOx[N_2]\text{-}c\{OC\sim O\} + H_2O$	**229.4**	**229.7**	**225.5**	**338**
$C_3H_9N_2O_2^+[N_2]\text{-}tgtt\{C\sim N\}$	102.3	105.8	113.3	155
$C_3H_9N_2O_2^+[N_1]\text{-}g(gt)tt + CO$	102.8	105.9	101.6	52
$C_3H_9N_2O_2^+[N_2\text{-}N_1]\text{-}(g_+t)gtt + CO$	**169.6**	**174.8**	**161.6**	**1638**
$C_3H_9N_2O_2^+[N_2\text{-}N_1]\text{-}g(gt)tc + CO$	**174.9**	**173.8**	**169.4**	**170**
$C_3H_9N_2O_2^+[N_2\text{-}N_1]\text{-}tgtt + CO$	**178.5**	**183.3**	**172.9**	**1627**
$C_3H_9N_2O_2^+[N_2\text{-}N_1]\text{-}gttt + CO$	**181.2**	**185.6**	**178.9**	**1626**
$c\text{-}C_3H_7N_2O^+[N_1](H_2O_{HN1u\text{-}d}) + CO$	81.8	88.1	73.4	92
$c\text{-}C_3H_7N_2O^+[N_1](H_2O_{HN2\text{-}HN1}) + CO$	112.6	118.3	101.1	48
$C_3H_7N_2O^+[N_1\text{-}N_2]\text{-}ct(H_2O_{HN1}) + CO$	155.2	162.8	159.9	1040
$c\text{-}C_3H_7N_2O^+[N_1]\{C_{\alpha2}\sim N_1\}(H_2O_{HN2}) +$ CO	165.2	172.3	175.5	190
$C_3H_7N_2O^+[N_2]\text{-}c(tc)(H_2O_{HN2}) + CO$	169.6	174.5	167.3	46
$c\text{-}C_3H_7N_2O^+[N_1]\{C_{\alpha2}\sim N_1\}(H_2O_{HN1}) +$ CO	187.4	192.8	194.2	150
$C_3H_7N_2O^+[N_2]\text{-}c(tc)(H_2O_{HN1}) + CO$	201.4	205.4	199.0	136
$C_3H_7N_2O^+[N_1\text{-}N_2]\text{-}ct + H_2O + CO$	205.7	208.4	209.8	1157
$c\text{-}C_3H_7N_2O^+[N_1]\{C_{\alpha2}\sim N_1\} + H_2O +$ CO	229.9	231.1	239.4	182
$C_3H_7N_2O^+[N_2]\text{-}c(tc) + H_2O + CO$	238.2	237.5	237.2	114

[a] Entries including arrows indicate the product ion formed by this TS. Bold entries indicate rate-limiting transition states for key processes observed experimentally.

[b] Values from the present study at the B3LYP/6-311+G(d,p), B3LYP/6-311+G(2d,2p)// B3LYP/6-311+G(d,p), and MP2(full)/6-311+G(2d,2p)//B3LYP/6-311+G(d,p) levels of theory, including zero point energy corrections.

[c] Values from PS [9], B3LYP/6-31G(d) | B3LYP/6-31+G(d,p); PCLS [8], B3LYP/6-31G(d); and in italics from BAL [10], B3LYP/6-31+G(d,p). Designations in parentheses are the names used by these authors.

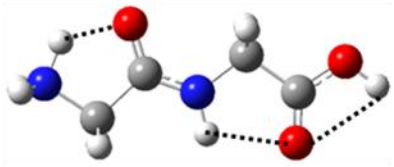
UU IR Author Manuscript

Figure Captions

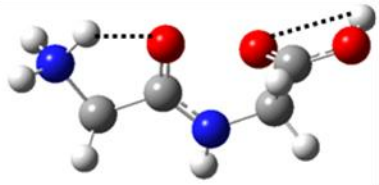
Figure S1. Structures for H^+GG with relative energies in kJ/mol calculated at the B3LYP/6-311+G(d,p) level. Hydrogen bonds are shown by dashed lines. Names used by PCLS, PS, and BAL are in parentheses.

Figure S2. Reaction coordinate surface for the lowest energy cis-trans peptide bond isomerization of H^+GG. Geometry optimizations and single point energies of each elementary step are determined at the B3LYP/6-311+G(d,p) level of theory and corrected for ZPE.

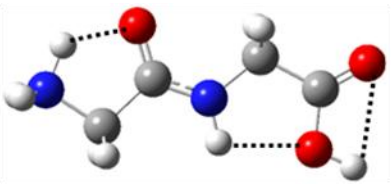
Figure S3. Reaction coordinate surface for H_2O loss from $H^+GG[N_1]$-gcgtt to form H^+DKP. Geometry optimizations and single point energies of each elementary step are determined at the B3LYP/6-311+G(d,p) level of theory and corrected for ZPE. The inset and additional horizontal bar indicate the structure and energy of an alternative rate-limiting transition state. Short dashed lines indicate bonds that are breaking or forming for transition states and hydrogen bonds for intermediates.



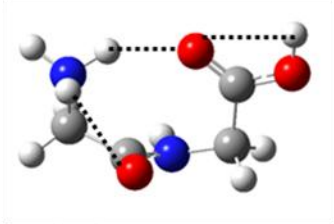
[N_1]-ttttt (A1, 1), 0.0



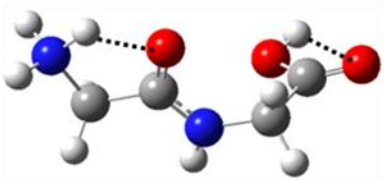
[N_1]-ttg_tt (A2), 6.2



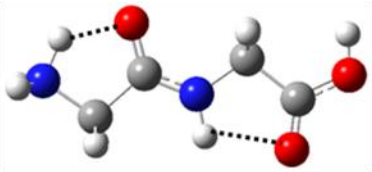
[N_1]-tttct (A4), 16.9



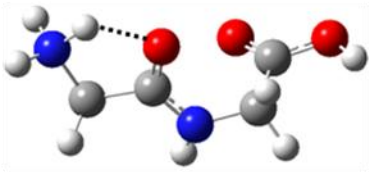
[N_1]-gtgtt (A3), 18.6



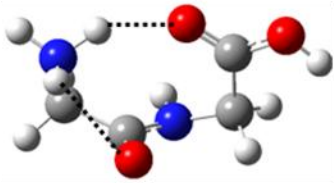
[N_1]-ttgct (A5, 2),19.5



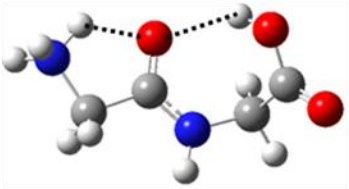
[N_1]-ttttc, 29.4



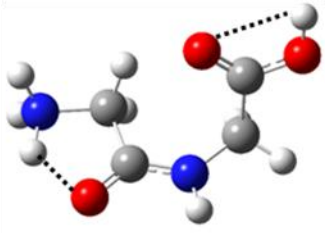
[N_1]-ttgtc, 35.3



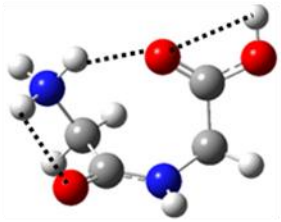
[N_1]-gtgtc, 40.9



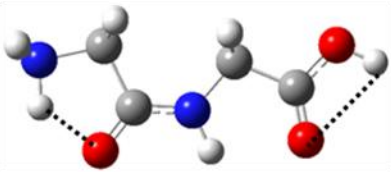
[N_1]-ttggc, 56.0



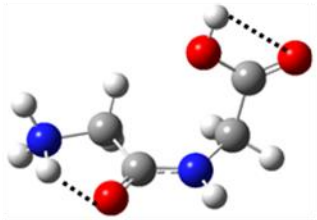
[N_1]-tcgtt (cis2), 12.6



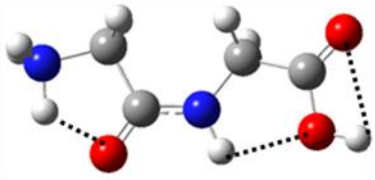
[N_1]-gcgtt (cis3), 13.0



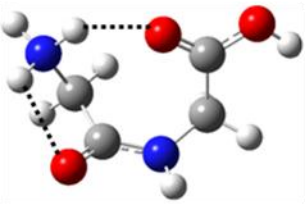
[N_1]-tcttt (cis1), 18.2



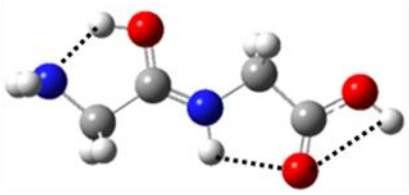
[N_1]-tcgct, 28.8



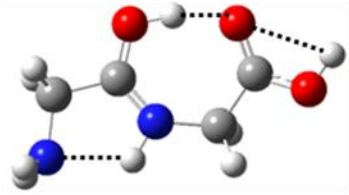
[N_1]-tctct, 30.9



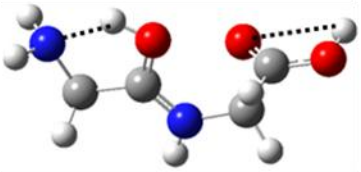
[N_1]-gcgtc, 38.3



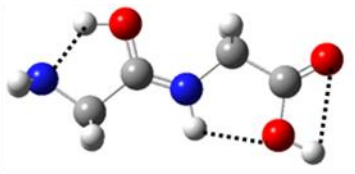
[O_{1c}]-ttttt (O1, 3), 2.3



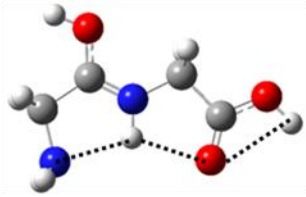
[O_{1t}]-ctctt (O2), 8.8



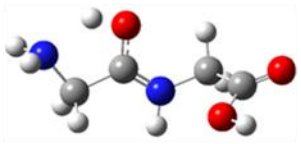
[O_{1c}]-ttgtt (O3), 11.9



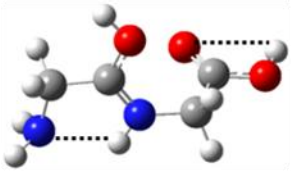
[O_{1c}]-tttct (O4), 19.9



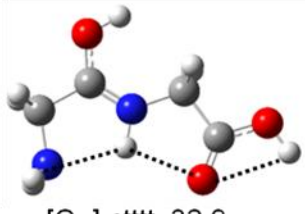
[O_{1c}]-ctttt (O5), 23.0



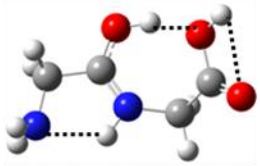
[O_{1c}]-ttgct (4), 24.5
DZ opt



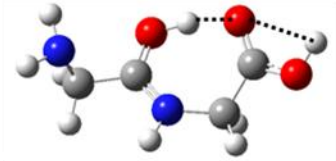
[O_{1c}]-ctgtt, 27.5



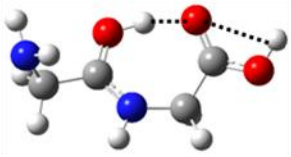
[O_{1t}]-ctttt, 32.9



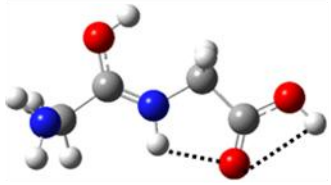
[O_{1t}]-ctggt (O6, 6), 35.8



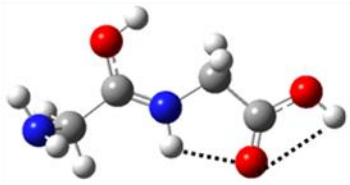
[O_{1t}]-gtgtt, 48.6



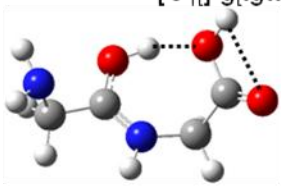
[O_{1t}]-$g_t$tgtt, 49.3



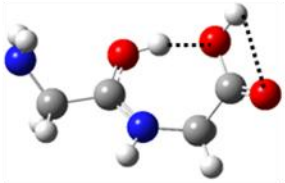
[O_{1t}]-gtttt, 63.4



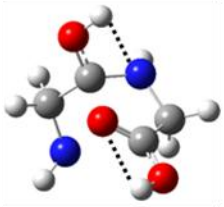
[O_{1t}]-$g_t$tttt, 67.4



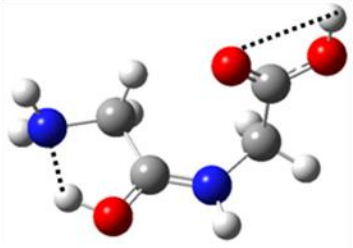
[O_{1t}]-gtggt (5), 73.0



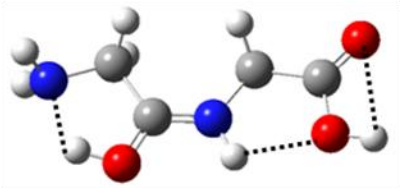
[O_{1t}]-$t_t$tggt (7), 76.2



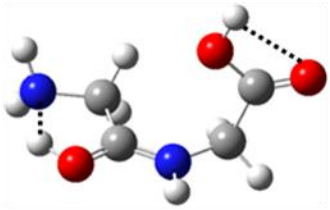
[O_{1t}]-ggctt, 148.1



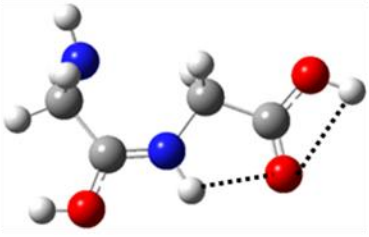
[O_{1c}]-tcgtt, 17.8



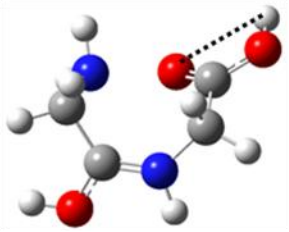
[O_{1c}]-tctct, 30.2



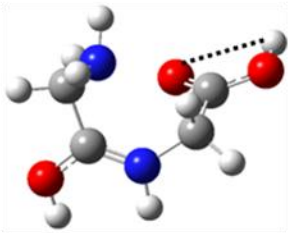
[O_{1c}]-tcgct, 32.3



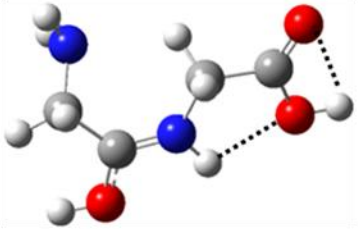
[O_{1c}]-ccttt, 56.0



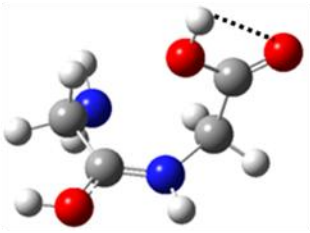
[O_{1c}]-ccgtt, 58.4



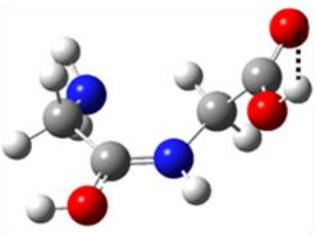
[O_{1t}]-ccgtt, 69.2



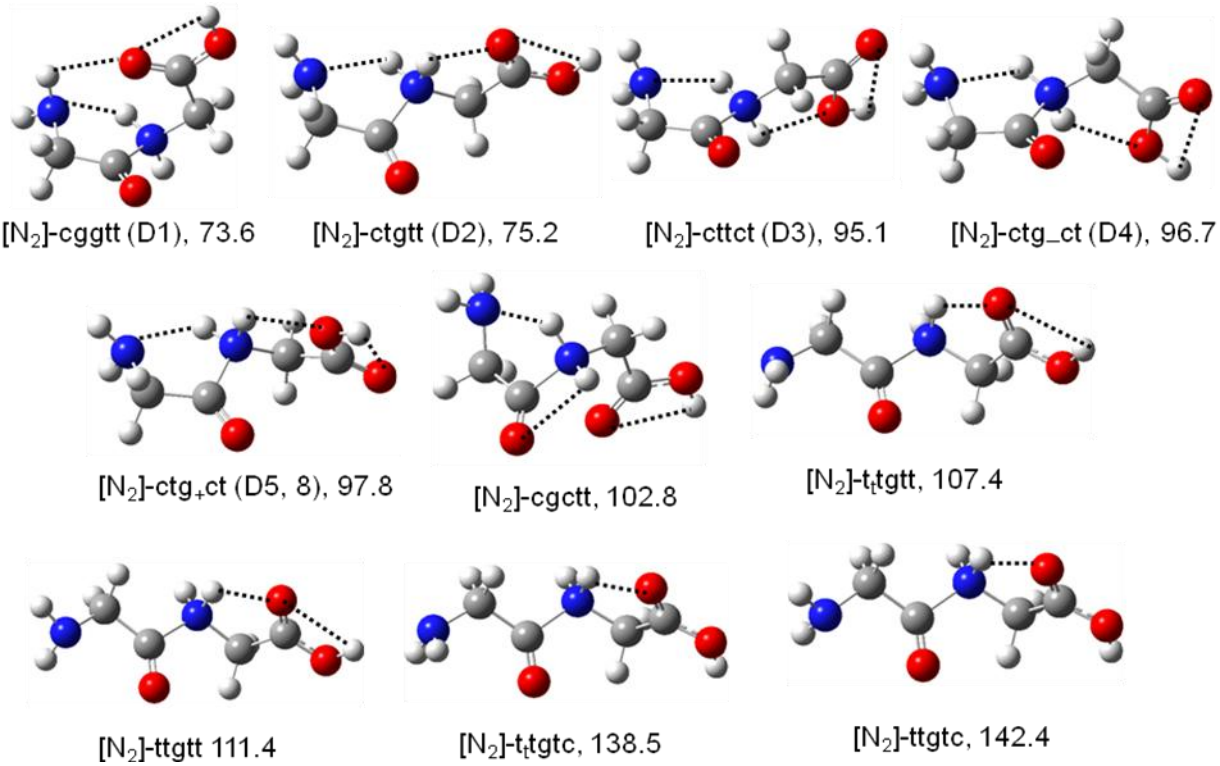
[O_{1c}]-gctct, 72.3



[O_{1c}]-ccgc_t, 74.3



[O_{1c}]-ccgc$_+$t, 74.7


[N2]-cggtt (D1), 73.6
[N2]-ctgtt (D2), 75.2
[N2]-cttct (D3), 95.1
[N2]-ctg_ct (D4), 96.7
[N2]-ctg+ct (D5, 8), 97.8
[N2]-cgctt, 102.8
[N2]-ttttgtt, 107.4
[N2]-ttgtt 111.4
[N2]-ttttgtc, 138.5
[N2]-ttgtc, 142.4

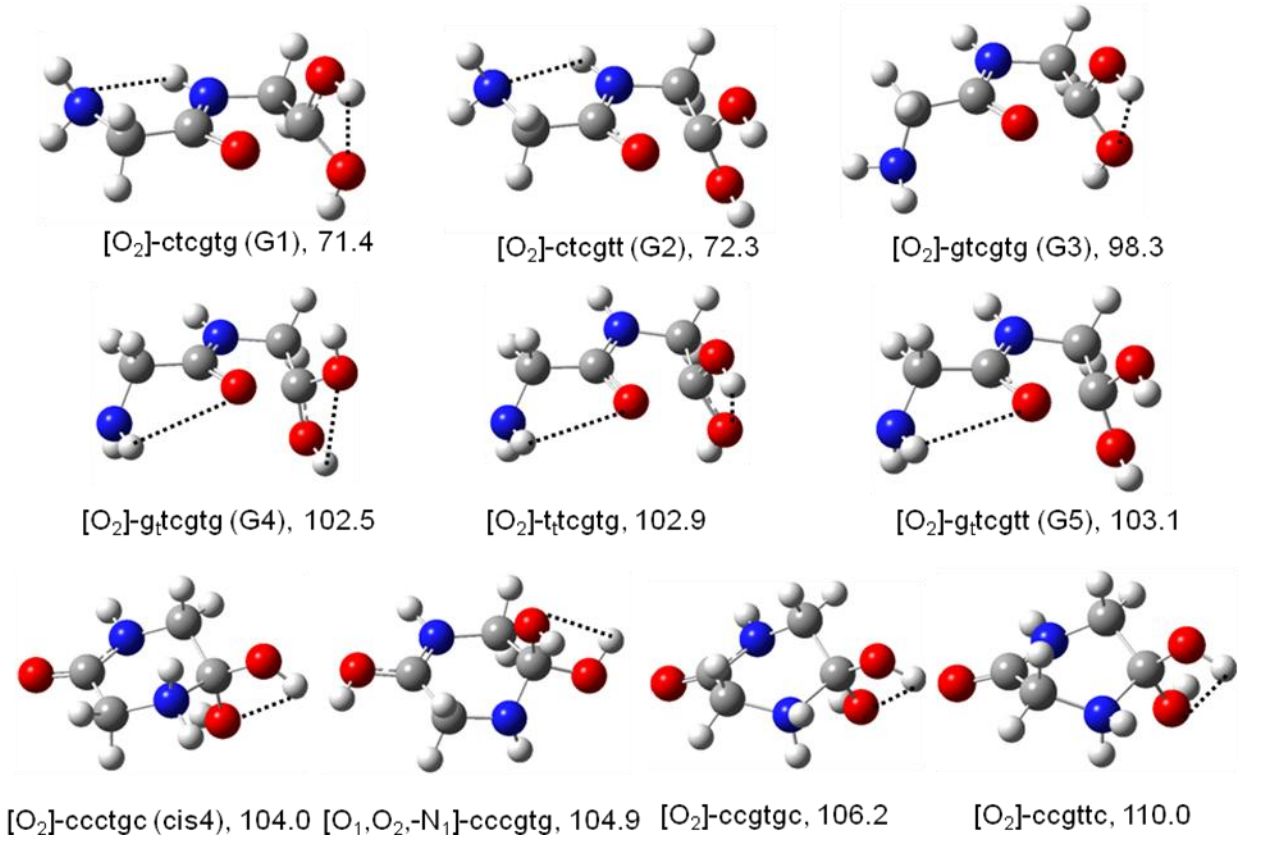

[O2]-ctcgtg (G1), 71.4
[O2]-ctcgtt (G2), 72.3
[O2]-gtcgtg (G3), 98.3
[O2]-gttcgtg (G4), 102.5
[O2]-ttttcgtg, 102.9
[O2]-gttcgtt (G5), 103.1
[O2]-ccctgc (cis4), 104.0
[O1,O2,-N1]-cccgtg, 104.9
[O2]-ccgtgc, 106.2
[O2]-ccgttc, 110.0

Figure S1

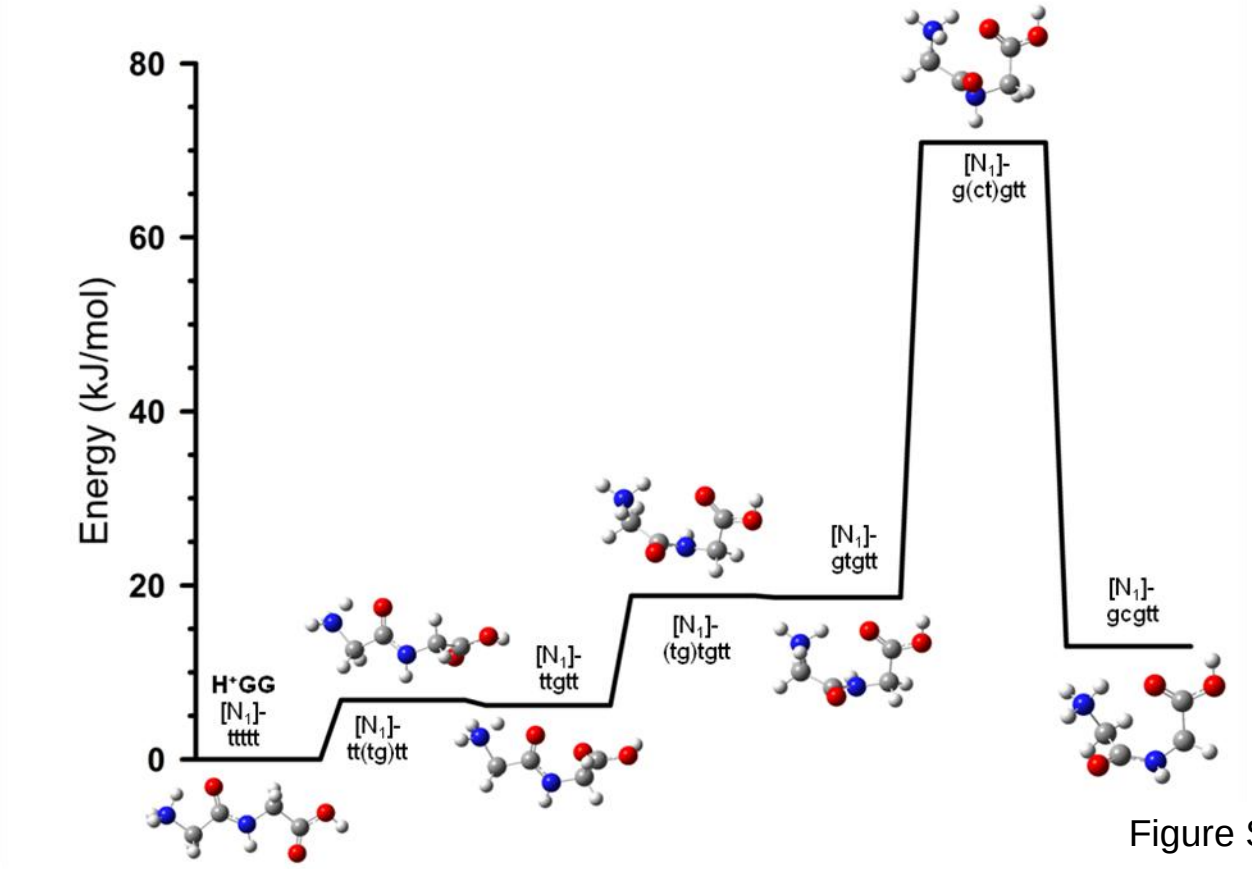


Figure S2

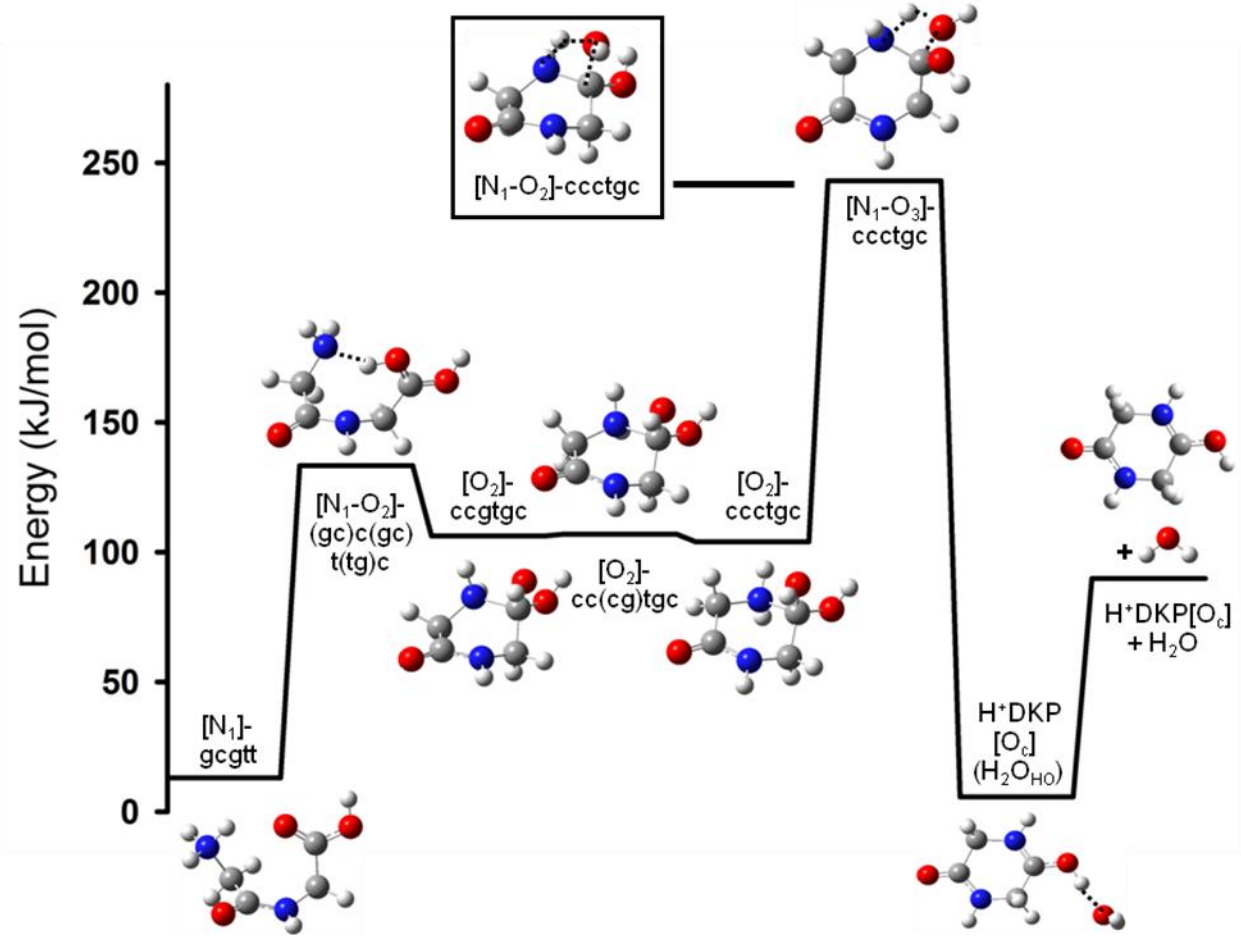


Figure S3