

# Governing Parameters of Long-Range Intramolecular S-to-N Acyl Transfers within (S)-Acyl Isopeptides

Jean-Christophe M. Monbaliu,<sup>#,‡</sup> Georges Dive,<sup>§</sup> Christian V. Stevens,<sup>‡</sup> and Alan R. Katritzky<sup>\*,#,||</sup>

<sup>#</sup>Center for Heterocyclic Compounds, Department of Chemistry, University of Florida, Gainesville, Florida 32611-7200, United States

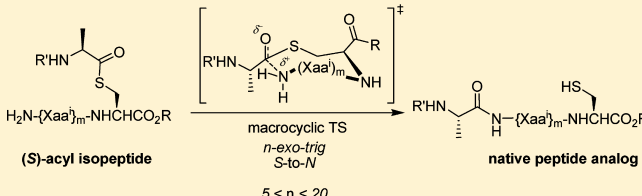
<sup>‡</sup>Department of Sustainable Organic Chemistry and Technology, Faculty of Bioscience Engineering, Ghent University, B-9000 Ghent, Belgium

<sup>§</sup>Center of Protein Engineering, Institut de Chimie, University of Liège, Bâtiment B.6, B-4000 Sart Tilman, Liège, Belgium

<sup>||</sup>Chemistry Department, King Abdulaziz University, Jeddah, 21589 Saudi Arabia

## S Supporting Information

**ABSTRACT:** The governing parameters for the long-range intramolecular S-to-N acyl transfer in (S)-acyl isopeptides are shown by computational and statistical methods (principal component analysis and cluster analysis) to be driven by enthalpic and geometric effects over the range  $n = 5–20$ . The results emphasize the dependency of  $\Delta G^\ddagger$  on the geometrical parameters governing the approach of the reactive termini and the importance of stabilizing intramolecular hydrogen bonds in the transition states (TSs), rather than the effects of TS ring-size. The competition between the intra- (uni-) and inter- (bi)molecular acyl transfers were studied for representative examples.



(S)-acyl isopeptide  $\xrightarrow[\text{macrocyclic TS } n\text{-exo-trig S-to-N}]{}$  native peptide analog

$5 < n < 20$

## 1. INTRODUCTION

Since the first report of glycylglycine in 1901 by Fisher,<sup>1</sup> a remarkable evolution of synthetic techniques<sup>2–6</sup> has enabled chemists to reach complex targets including not only peptides (<100 amino acid residues) but also small proteins (>100 amino acid residues).<sup>7–14</sup> Notably, the development of polymer-supported techniques (solid phase peptide synthesis, SPPS) and the automation of peptide synthesis have enabled large scale industrial and pharmaceutical applications.<sup>15,16</sup> Despite the arsenal of coupling reagents and protecting groups developed over a century of research, peptide synthesis remains challenging: increasing complexity of the targets, aggregation/solubility issues, regulatory constraints, and synthetic cost are among the main hurdles that continuously stimulate the development of new techniques.<sup>17</sup>

Within the context of convergent synthesis, chemoselective and fast transformations under smooth conditions are preferred for the ligation of two peptide fragments, preferably eliminating the need for protecting groups and minimizing epimerization. The particular reactivity of thioesters offered a solid opportunity for the development of such techniques.<sup>18–23</sup> Two main synthetic<sup>24</sup> techniques based on thioester reactivity have emerged during the last two decades, namely native chemical ligation (NCL)<sup>8,12,25–29</sup> and (S)-acyl isopeptide strategies.<sup>30–36</sup> Although conceptually different, both involve two common features (i) an initial capture event forming a (S)-isopeptide and (ii) an isomerization to the native peptide analog proceeding through an entropically favored S-exo-trig S-to-N acyl transfer (Figure 1a).<sup>37</sup> In these methods, a favorable positioning of the

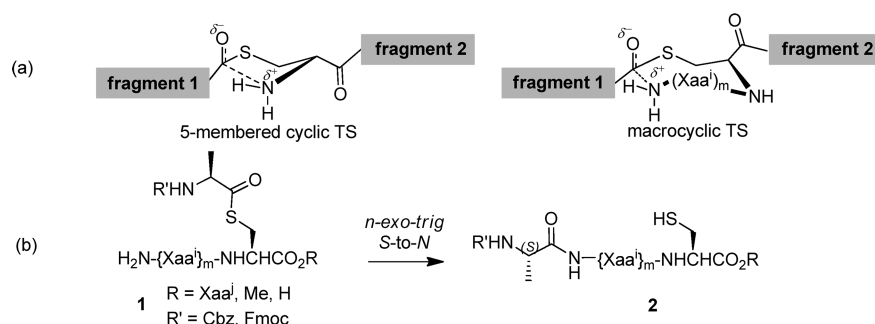
reactive moieties in close proximity can compensate<sup>38</sup> the modest enthalpic activation of thioesters (Figure 1a). Traditionally, these techniques rely on the nucleophilicity of a Cys  $\beta$ -sulfhydryl side chain, although NCL have now been extended to a variety of ligation sites using thiolated analogs of amino acids.<sup>8,12,13</sup>

This paradigm was challenged by Haase and Seitz<sup>29</sup> who introduced the concept of internal Cys ligation in 2009, extending the proof-of-concept already demonstrated in Sugar Assisted Ligation (SAL)<sup>42–46</sup> and auxiliary-assisted NCL.<sup>8,12,47</sup> Haase and Seitz studied whether internal cysteine ligation—i.e. proceeding through a macrocyclic transition state (TS)—would also accelerate the rates of thioester-based segment condensation. They found that S-to-N acyl transfers through 17–20 macrocyclic TSs were optimal and that, by contrast, 8-, 11-, and 14-membered macrocycles were difficult to form (Figure 1a).

In a series of recent papers, we reported intramolecular S-to-N acyl transfers of enantiopure and stable (S)-acyl isopeptides **1** through a palette of macrocyclic TSs.<sup>30–33</sup> As in Seitz's internal cysteine-accelerated NCL,<sup>29</sup> variations on the Xaa<sup>i</sup> sequence (Figure 1b) were studied to access the corresponding native oligopeptide analogs **2**. The rates and yields of the ligations were found to be highly variable and somehow dependent on the size of the macrocyclic transition state<sup>30,32,33</sup> but less on the substitution.<sup>31</sup> S-to-N Acyl transfer through a 5-membered cyclic TS proceeded in excellent yield at room temperature,<sup>32,33</sup> while

Received: September 25, 2012

Published: November 30, 2012



**Figure 1.** (a) 5-Exo-trig S-to-N acyl transfer as commonly encountered in classical NCL and (S)-isopeptide rearrangements (left) and *n-exo-trig* S-to-N acyl transfer in internal Cys NCL and extended isopeptide rearrangements (right). (b) This work: isomerization of (S)-acyl isopeptides **1** to native peptide analogs **2**. If  $m = 0$ ,  $n = 5$ .

larger macrocyclic TSs required prolonged MW heating to afford the desired ligated peptides in moderate to good yields.<sup>30–33</sup> Interestingly, an attempt to isomerize a S-acyl isopeptide through an 8-membered TS failed.<sup>32,33</sup> A preliminary computational study suggested the probable implication of preorganizational factors to govern these intramolecular long-range acyl transfers.<sup>34</sup>

The contamination of the desired ligation product by a variety of side products including hydrolysis,<sup>29</sup> direct aminolysis,<sup>29</sup> transacylation, and disulfide formation<sup>29,30,32,33</sup> complicates the interpretation of the results. To suppress such experimental interferences, we now report on a computational study of the long-range S-to-N acyl transfer in (S)-acyl isopeptides **1** proceeding via *n*-(macro)cyclic TSs ranging from  $n = 5$ –20. Systematic calculation of the activation parameters and determination of the geometrical strain involved at the TS can rationalize long-range S-to-N intramolecular acyl transfer in (S)-acyl isopeptides **1** in terms of the governing parameters.

## 2. METHODOLOGY

Quantum chemical calculations were performed using the Gaussian 09 package of programs.<sup>48</sup> HF method employing the 6-31G\* basis set with 5 pure d functions and gradient techniques using internal coordinates with very tight optimization convergence criteria (each component of the first energy derivative below  $2.0 \times 10^{-6}$  Hartree/Bohr or radian) were used for both geometry optimization and computation of vibrational properties. The transition states were localized using the New-Raphson algorithm, and the nature of the stationary points was determined by analysis of the Hessian. The impact of both the method and the basis set was assessed for representative examples using HF or B3LYP methods and 6-31G\* or 6-31+G\* basis sets with 5 pure d functions. The activation and reaction energies were calculated from the thermochemical output computed for the reagents, transition states, and products, using standard thermochemistry as implemented in Gaussian 09. Thermochemistry was computed at 298.150 K under a pressure of 1 atm. A detailed analysis of the atomic movements in the intramolecular S-to-N acyl transfer process was further obtained by intrinsic reaction coordinate (IRC) calculations. IRC calculations were used in the gas phase to localize the nearest local minima on the reactant and product sides of the reaction coordinate. Statistical analysis was performed on the TS structures using SAS software.<sup>49</sup>

## 3. RESULTS

**3.1. Transition States and Reaction Path.** Diverse (S)-acyl isopeptide sequences **1** systematically covered *n-exo-trig* S-

to-N acyl transfers ranging from  $n = 5$  to  $n = 20$  (See Table 1 and Figure 1b; see Table S1 in the Supporting Information for

**Table 1.** (S)-Acyl Isopeptide Sequences Studied and Corresponding TS Ring Size for *n-exo-trig* S-to-N Acyl Migration<sup>a</sup>

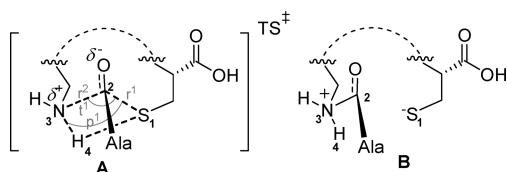
ring size	TS	{Xaa <sup>i</sup> } <sub>m</sub>	ring size	TS	{Xaa <sup>i</sup> } <sub>m</sub>
5	lig <sub>5</sub>		12	lig <sub>12b</sub>	β-Ala-Gly
6	lig <sub>6a</sub>		13	lig <sub>13a</sub>	Gly-GABA
6	lig <sub>6b</sub>		13	lig <sub>13b</sub>	GABA-Gly
7	lig <sub>7a</sub>		14	lig <sub>14a</sub>	{Gly} <sub>3</sub>
7	lig <sub>7b</sub>		14	lig <sub>14b</sub>	β-Ala-GABA
8	lig <sub>8</sub>	Gly	15	lig <sub>15</sub>	β-Ala-{Gly} <sub>2</sub>
9	lig <sub>9</sub>	β-Ala	16	lig <sub>16a</sub>	GABA-{Gly} <sub>2</sub>
10	lig <sub>10</sub>	GABA	16	lig <sub>16b</sub>	β-Ala-Gly-β-Ala
11	lig <sub>11</sub>	{Gly} <sub>2</sub>	17	lig <sub>17</sub>	{Gly} <sub>4</sub>
12	lig <sub>12a</sub>	Gly-β-Ala	20	lig <sub>20</sub>	{Gly} <sub>5</sub>

<sup>a</sup>See also Figure 1b for disambiguation and full details in the Supporting Information.

detailed sequence). The different ring sizes were reached by incrementing the {Xaa<sup>i</sup>}<sub>m</sub> sequence, including α-unsubstituted natural and unnatural amino acids: glycine (Gly), β-alanine (β-Ala), and γ-aminobutyric acid (GABA). Substituted amino acids were not considered to limit the resource-consuming nature of the computations. The sequences presented in this study were selected on the basis of their availability in published experimental reports.<sup>29–33</sup> Isomeric macrocyclic TSs were obtained for representative examples by considering the most obvious sequence alternative. For the 6- and 7-membered ring S-to-N acyl transfer ( $m = 0$ ), unnatural analogs of Cys were used to reach the desired ring size. The corresponding TSs were labeled with respect to ring size, i.e. lig<sub>n</sub>. The indices a and b were used to distinguish sequence variations proceeding through isomeric *n-exo-trig* TS (e.g., lig<sub>12a</sub> and lig<sub>12b</sub>). In addition to the basic sequential variations presented in Table 1, the impact of the substitution (R and R'; see Figure 1b) was studied on the sequences associated with lig<sub>9</sub> and lig<sub>13a</sub>: lig<sub>9</sub><sup>e</sup> (R = Me), lig<sub>9</sub><sup>ec</sup> (R = Me, R' = CO<sub>2</sub>Me), lig<sub>13a</sub><sup>e</sup> (R = Me), lig<sub>13a</sub><sup>c</sup> (R = CO<sub>2</sub>Me), and lig<sub>13a</sub><sup>ec</sup> (R = Me, R' = CO<sub>2</sub>Me). The conformation of the S-acylating Ala residue was also taken into account for some representative examples: lig<sub>9</sub><sup>rot1</sup>, lig<sub>9</sub><sup>rot2</sup>, and lig<sub>14b</sub><sup>rot1</sup> (rotation of the amino group). The impact of the cis/trans configuration of the peptide backbone was assessed by comparing lig<sub>16a</sub> (featuring all amide bonds in trans) and lig<sub>16a</sub><sup>oc</sup> (featuring one amide bond in cis). One example including explicitly a molecule of water in the mechanism was computed (lig<sub>9</sub><sup>w</sup>).

Two additional data points were considered to measure the impact of the method level and basis set expansion on the TS structure: **lig**<sub>9</sub><sup>Hd+</sup> (HF/6-31+G\*) **lig**<sub>9</sub><sup>Bd</sup> (B3LYP/6-31G\*). Notably, the structures of **lig**<sub>9</sub>, **lig**<sub>9</sub><sup>Hd+</sup>, and **lig**<sub>9</sub><sup>Bd</sup> were perfectly superimposable (99% molecular overlay between **lig**<sub>9</sub> and **lig**<sub>9</sub><sup>Hd+</sup> and 98% molecular overlay between **lig**<sub>9</sub><sup>Hd+</sup> and **lig**<sub>9</sub><sup>Bd</sup>)<sup>50</sup> demonstrating the minimal impact of the method level and basis set expansion on TS structure in this study due to the compactness of the geometry.

In all, a total of 32 TSs were isolated. In the vast majority of the cases, the imaginary frequency was associated with the concomitant transfer of (i) the Ala residue from the C-terminal cysteine to the N-terminus of the isopeptide **1** and of (ii) a proton from the N-terminus to the cysteinyl thiolate, defining a pseudo 4-membered ring (concerted-like TS **A** in Figure 2). The most relevant geometrical parameters governing the approach of the two reactive termini are summarized in Table 2.



**Figure 2.** Pseudo 4-membered core associated with the vast majority of intramolecular *n*-*exo*-*trig* S-to-N acyl migrations in oligopeptides (concerted-like TS, **A**) and reactive intermediate **B** for the stepwise-like path.

**Table 2.** Connectivity Matrix for the TSs **lig**<sub>n</sub><sup>a</sup>

atom no.		geometrical parameters						
1	S							
2	C	1	$r^1$					
3	N	2	$r^2$	1	$t^1$			
4	H	3	$r^3$	2	$t^2$	1	$p^1$	

<sup>a</sup>Bond distances  $r^1$  and  $r^2$  and valence and dihedral angles  $t^1$  and  $p^1$  define the approach of the reactive termini (See also Figure 2).

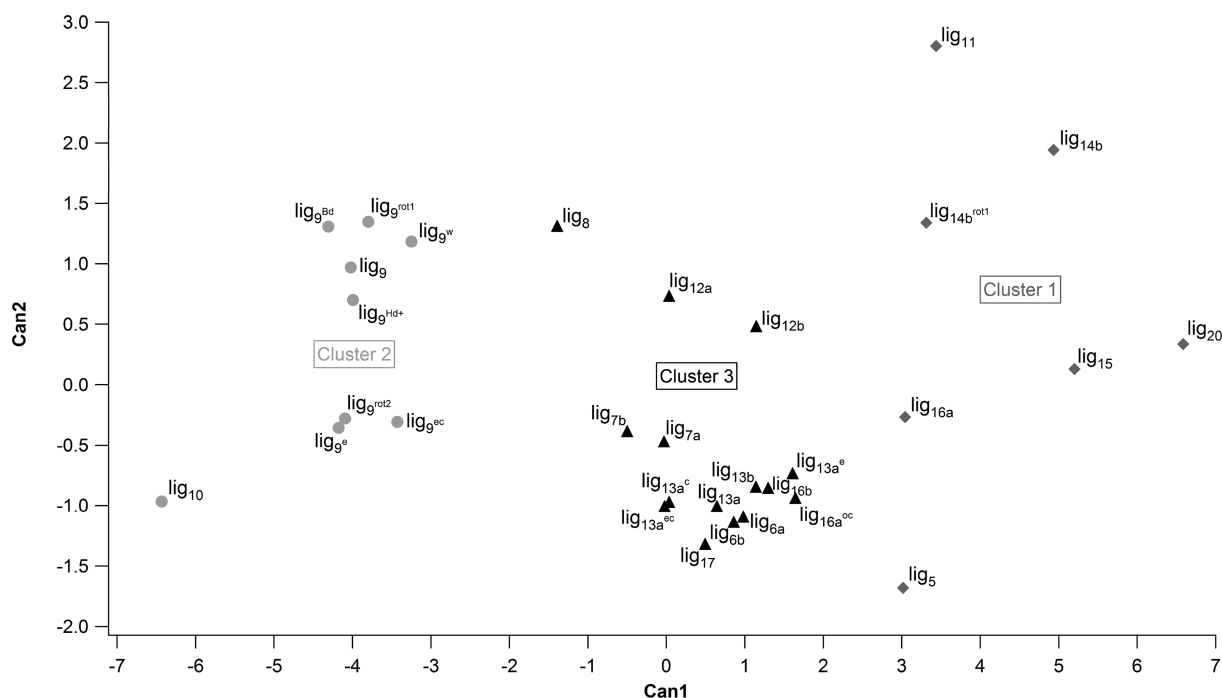
These TSs are late—in the Hammond postulate sense, i.e. product-like with a net transfer of the alanine residue from the S-cysteine terminus to the N-terminus (average  $r^1 = 2.43$  Å). In some cases (**lig**<sub>11</sub>, **lig**<sub>14b</sub>, **lig**<sub>15</sub>, and **lig**<sub>16a,b</sub>), the TSs were solely associated with the Ala transfer (average  $r^1 = 2.49$  Å); following the right branch of the IRC led to an intermediate with net charge separation (see intermediate **B** in Figure 2 for the stepwise-like path). This mechanistic feature is, however, only anecdotal since, in both cases (concerted or stepwise), the transfer of the Ala residue remains the rate-determining step. TS **lig**<sub>14a</sub> appeared as marginal with a particularly long  $r^1$  distance ( $r^1 = 2.93$  Å). Activation and reaction parameters are summarized in Table S2 in the Supporting Information (SI) and discussed in section 4.

**3.2. Statistical Analysis.** The generation of a data set of observations belonging to the same family of reactions can be subjected to statistical analysis in order to simplify, extract the most relevant information, and draw trends in highlighting similarities and differences. While performing diagonalization of the Hessian matrices associated to the 32 TSs localized, we found that the main components of the eigenvector associated to the negative eigenvalue were associated with the geometrical parameters dictating the approach of the nucleophilic amine onto the electrophilic carbonyl, namely  $r^1$ ,  $r^2$ ,  $t^1$ , and  $p^1$  (Figure 1 and Table 2). Among the 32 computed TS structures, **lig**<sub>14a</sub> was

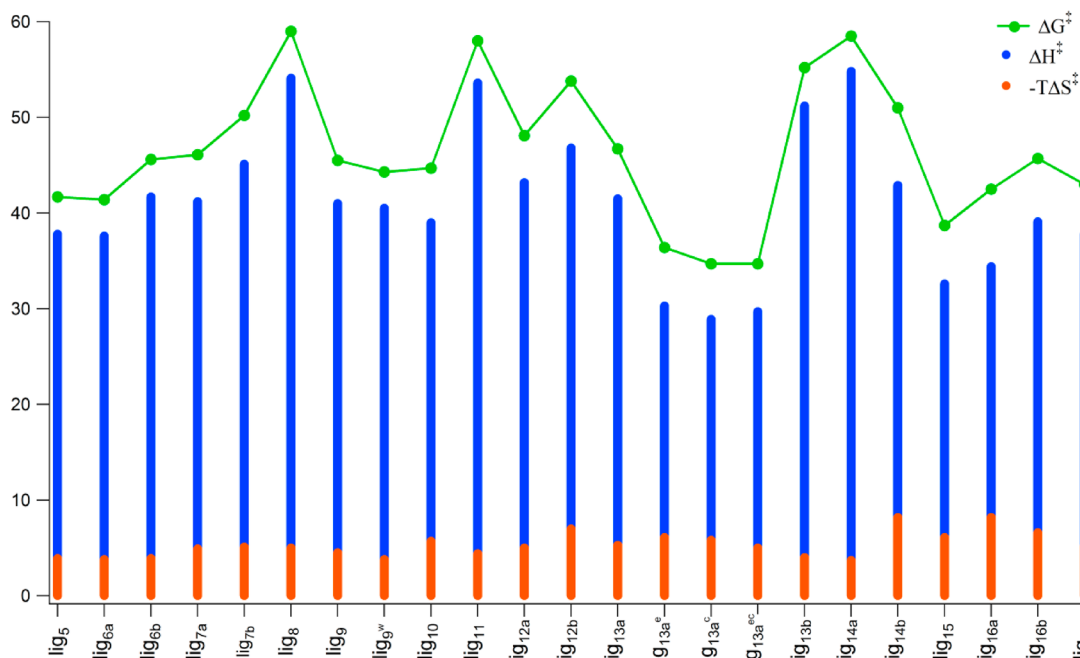
considered as an outlier due to an unusually long  $r^1$  distance (2.93 Å) and was discarded from the data set. For the 31 remaining TS structures, the correlation matrix generated by the four variables  $r^1$ ,  $r^2$ ,  $t^1$ , and  $p^1$  revealed a high relation between the variations of  $r^2$  and those of  $p^1$  (See SI Table S3).

Two different types of statistical analysis have been carried out in order to visualize the data points within this four-dimensional space (see the SI for full procedure details).<sup>51</sup> First, a principal component analysis (PCA) was performed to express the data according to their most variations. The diagonalization of the metric matrix generates four eigenvectors as linear orthogonal combinations of the four original variables ( $r^1$ ,  $r^2$ ,  $t^1$ , and  $p^1$ ). From the corresponding four eigenvalues, the two least were ignored as of lesser significance, while the two first eigenvalues (2.033 and 1.107) combined 78.5% of the total variance and were considered further. This minimal loss of information allowed us to reduce the dimensionality of the data set. The two selected eigenvectors define the factor pattern which mainly combines  $r^2$  and  $p^1$  in the first principal component and  $t^1$  and  $r^1$  in the second one. Performing orthogonal transformation of this matrix using the Varimax procedure, the variance was redistributed on the two selected components in order to maximize the correlation between the variations of the original data and the retained components (principal components PC1 & PC2; see Table S4 in the SI). At this point, it was possible to re-express the TS data set in terms of PC1 & PC2. The projection of the coordinates of the 31 molecules (factor scores) into this two-dimensional space has clearly showed a structuring into the data. For instance, the unsubstituted and substituted 9- and 13-macrocyclic TS structures appeared as homogeneous groups. The main feature of this PCA analysis is related to the dispersion of the molecules following the PC1 (42.7% of variance) axis due to the variations of  $p^1$  and  $r^2$  (see Figure S1 in the Supporting Information). These results suggested a structuring of the macrocyclic TS structures into well-defined groups.

The FASTCLUS procedure was then used to perform a clustering out of a data set free of any a priori known groups (31 TS structures; see Supporting Information for full procedure details). Starting from four geometrical variables ( $r^1$ ,  $r^2$ ,  $t^1$ , and  $p^1$ ), it was reasonable to split the data into three groups ("clusters"). A set of three points was thus selected as first guesses (cluster seeds); each TS structure was then assigned to the nearest seed to form temporary clusters, the means of which are then iteratively determined by the *k*-means algorithm. Three clusters respectively containing 7, 9, and 15 macrocyclic TS structures were obtained accordingly (Figure 3). The criterion to define the clusters was then tested by means of a multivariate analysis of variance (MANOVA). The variable which significantly differs between clusters appeared to be  $p^1$  with a mean value of 73.287, −8.710, and 36.886 in clusters 1, 2, and 3, respectively. ANOVA results have shown that  $p^1$  and  $r^2$  generated an  $F(2,28) = 138.96$  and 15.25, respectively (significant at  $p < 0.0001$ ). The Wilks'  $\lambda$  of the MANOVA taking into account the incidence of the four variables was found equal to 0.06767 corresponding to  $F(8,50) = 17.77$ , which is also highly significant. Using the CANDISC procedure, a Rao's canonical analysis was performed to find linear combinations of the variables that provided maximal separation between the clusters. The first eigenvalue combined 96.95% of the information, while the second eigenvalue, 3.05%. A plot of the two canonical variables (Can1 and Can2) shows that Can1 splits the data set into three clusters without overlapping (Figure 3).



**Figure 3.** Plot of the two canonical variables Can1 and Can2.



**Figure 4.** Evolution of the activation parameters  $\Delta G^\ddagger$ ,  $\Delta H^\ddagger$ , and  $-T\Delta S^\ddagger$  (kcal·mol<sup>-1</sup>) as a function of the ring size and substitution for the intramolecular acyl transfer (representative examples).

## 4. DISCUSSION

Intramolecular S-to-N acyl transfers were documented on 5-membered cyclic systems by kinetic studies in the late 1950s and 1960s,<sup>52,53</sup> but since then, no further mechanistic nor computational study on larger intramolecular S-to-N acyl transfers have been reported. Illuminati and Mandolini pioneered the study and interpretation of the governing parameters for intramolecular reactions in their seminal report on macrolactonization.<sup>54–59</sup> Reinterpretation of their results by Bruice emphasized the differential contribution of enthalpy and entropy versus ring size;

for small to medium ring sizes ( $n = 3-8$ ), the enthalpy ( $\Delta H^\ddagger$ , i.e. the ring strain) is the major contributor to the observed rates (related to  $\Delta G^\ddagger$ ), while for larger ring sizes ( $n > 8$ ), the entropic contribution ( $T\Delta S^\ddagger$ , i.e. the probability of encounter by the reactive termini) gains importance.<sup>55</sup> Classically, macrocyclization proceeding through 5- and 6-membered rings is known to proceed with low activation barriers. An increase in ring size is first characterized by a significant increase in activation energy (maximum for  $n = 8$ ) and then a progressive leveling-off for higher ring sizes ( $n > 13$ ).<sup>55</sup> Thus, there is a general trend for  $\Delta H^\ddagger$  to decrease as the chain length increases. For small and



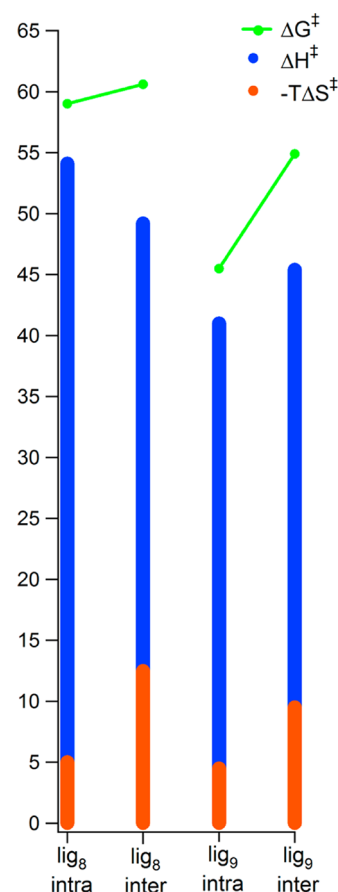
medium ring size ( $3 < n < 16$ ), large differences usually appear between the intramolecular and intermolecular  $\Delta H^\ddagger$ , i.e. the ring strain weighs a lot. The highest strains were reported for  $n = 3$  and  $n = 8$ . It has been shown that for the formation of large rings ( $n = 18\text{--}23$ ), the intramolecular reaction has a similar  $\Delta H^\ddagger$  as its intermolecular counterpart, i.e. the reaction is almost strainless.

In sharp contrast with these prerequisite data, the computed  $\Delta G^\ddagger$  for *n-exo-trig* S-to-N acyl transfer displayed a high dependency on  $\Delta H^\ddagger$ , while the entropy contribution appeared to be rather invariant and of minimal contribution even for  $n > 8$  (Figure 4). It is worth noting that, in isomeric TSs, the presence of unnatural amino acids in the sequence came with a higher  $\Delta S^\ddagger$  contribution, likely arising from an increased chain flexibility (e.g., **lig**<sub>14a</sub> and **lig**<sub>14b</sub>). A rather similar  $\Delta G^\ddagger$  profile than the one reported for lactone macrocyclization was obtained for medium macrocyclic TS formation ( $5 < n < 10$ ). At first, a progressive increase of  $\Delta G^\ddagger$  for  $n = 5, 6, 7$ , and 8 was noticed, with a maximum in  $\Delta G^\ddagger$  associated with **lig**<sub>8</sub>, and then followed by a significant drop for  $n = 9, 10$ . For  $n > 10$ , an unexpected trend featuring a succession of extrema and the absence of the expected leveling-off of  $\Delta G^\ddagger$  at  $n > 13$  was observed (Figure 4).

Thus, the contribution of the chain strain remains high through the entire series, by contrast with lactones and cycloalkanes. From these observations, it came out that (S)-acyl isopeptides bearing at least one non- $\alpha$ -amino acid in the {Xaa<sup>i</sup>}<sub>m</sub> sequence (see Figure 1 and Table 1) are usually characterized by a lower cyclic strain (lower Pitzer strain contribution), while (S)-acyl isopeptide sequences made of  $\alpha$ -amino acids are usually characterized by a higher cyclic strain. This phenomenon probably arises from an increase in torsional freedom, and therefore a decrease in the Pitzer strain contribution to the total cyclic strain, in (S)-acyl  $\beta$ - and  $\gamma$ -isopeptides, which is reflected in the evolution of the activation barrier as well. Notably, the three absolute maxima in Figure 4 are **lig**<sub>8</sub>, **lig**<sub>11</sub>, and **lig**<sub>14a</sub>, the sequences of which contain exclusively  $\alpha$ -amino acids. The succession of maxima for **lig**<sub>8</sub>, **lig**<sub>11</sub>, and **lig**<sub>14</sub> (Xaa<sup>i</sup> = Gly and  $m = 1, 2$ , and 3 for {Xaa<sup>i</sup>}<sub>m</sub>, respectively) is consistent with the experimental observations reported by Haase and Seitz in their internal Cys ligation experiments.<sup>29</sup>

This phenomenon can also be correlated with our early experimental reports on the long-range acyl transfer within (S)-acyl isopeptides:<sup>30–33</sup> for instance, the intramolecular acyl transfer product was not observed through eight-membered macrocyclic TSs, while the main reaction product was identified as resulting from an intermolecular (bimolecular) acyl transfer. To further document this matter, the intermolecular (bimolecular) acyl transfers of **lig**<sub>8</sub> and **lig**<sub>9</sub> were computed and compared to their intramolecular (unimolecular) counterparts (Figure 5). The intermolecular acyl transfer for **lig**<sub>8</sub> requires a comparable activation barrier than for the intramolecular acyl transfer, whereas the intermolecular acyl transfer for **lig**<sub>9</sub> has a significantly higher activation barrier (ca. 10 kcal·mol<sup>−1</sup>) than its intramolecular counterpart. Interestingly, the enthalpy of activation for the intramolecular acyl transfer for **lig**<sub>8</sub> is higher than the intermolecular transfer, while an opposite trend is observed for **lig**<sub>9</sub>. This difference emphasizes the higher internal strain imposed to **lig**<sub>8</sub> than to **lig**<sub>9</sub> for reaching the intramolecular transition state for the acyl transfer. Clearly, the competition between the intra- and intermolecular acyl transfers is driven by the parameters that govern the approach of the reactive termini, i.e. the ring strain.<sup>60</sup>

The explicit inclusion of water in the reaction mechanism did not affect the activation barriers significantly on **lig**<sub>9</sub>,<sup>w</sup> compared

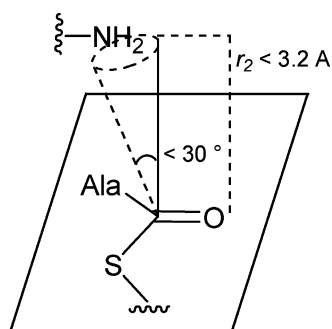


**Figure 5.** Comparison of the activation parameters for the intra- (uni-) and inter- (bi-) molecular acyl transfer for **lig**<sub>8</sub> and **lig**<sub>9</sub>.

to **lig**<sub>9</sub>. The substitution of R and R' (Figure 1b) was also studied: if R = Me or/and R' = CO<sub>2</sub>Me (**lig**<sub>13a</sub><sup>e</sup> R = Me, R' = H; **lig**<sub>13a</sub><sup>c</sup> R = H, R' = CO<sub>2</sub>Me; **lig**<sub>13a</sub><sup>ec</sup> R = Me, R' = CO<sub>2</sub>Me), a significant decrease in activation barrier (ca. 10 kcal·mol<sup>−1</sup>) was noticed by comparison with **lig**<sub>13a</sub> (R = R' = H). However, their belonging to the same geometrical cluster (Figure 3) led us to conclude that this decrease in  $\Delta G^\ddagger$  arose from a relative destabilization of the corresponding reactive (S)-acyl isopeptide conformers, thus minimizing the global effect of the substitution at R and R'.<sup>31–33</sup>

Figure 4 also emphasizes the impact of the amino acid nature and sequence ({Xaa<sup>i</sup>}<sub>m</sub>) through the isomeric TSs **lig**<sub>12a</sub>, **lig**<sub>12b</sub>, **lig**<sub>13a</sub>, **lig**<sub>13b</sub>, **lig**<sub>14a</sub>, **lig**<sub>14b</sub>, **lig**<sub>16a</sub>, and **lig**<sub>16b</sub>, which led to drastically different activation barriers. Also, a cis-configured amide bond in the sequence increased the activation barrier of 7.3 kcal·mol<sup>−1</sup> (**lig**<sub>16a</sub> vs **lig**<sub>16a</sub><sup>oc</sup>; see the SI). Clearly, the size of the macrocycle is not the only determinant factor for the activation barrier, a fact supported by the stackable nature of the TS according to three distinct clusters regardless of the ring size (Figure 3).<sup>61</sup>

These observations prompted us to analyze the nature of the reactive (S)-acyl isopeptide conformers. Several authors tried to rationalize intramolecular reactivity by decoupling reactive conformer and TS effects. Among others, Menger,<sup>62</sup> Houk,<sup>63</sup> and Bruice<sup>64</sup> have considered the proximity of the reactive termini in the reactive conformer as a crucial parameter for intramolecular reactivity. In particular, Bruice introduced the concept of propinquity,<sup>65</sup> later extended to the concept of Near Attack Conformation (NAC), i.e. the conformation of a genuine reactive conformer featuring the required conformation to juxtapose reactive termini to enter the TS (Figure 6).<sup>61</sup>



**Figure 6.** NAC concept extended to intramolecular S-to-N acyl transfer (adapted from ref 61).

According to these studies, mostly dedicated to the intramolecular formation of small-size cyclic anhydrides, a high molarity in NAC is correlated with high intramolecular rates.<sup>61</sup> Recently, Karaman revisited this concept in a series of contributions and emphasized the importance of torsion and proximity for accessing intramolecular TSs.<sup>66–68</sup> However, in our case, none of the conformers isolated corresponded to the NAC definition (Figure 6), with termini separations ranging from  $r^2 = 3.82$  Å (**lig**<sub>5</sub>) to 10.12 Å (**lig**<sub>14b</sub>). The conformation of the reactive (S)-isopeptide could not be solely responsible for variations in  $\Delta G^\ddagger$  between the different macrocyclic TSs.

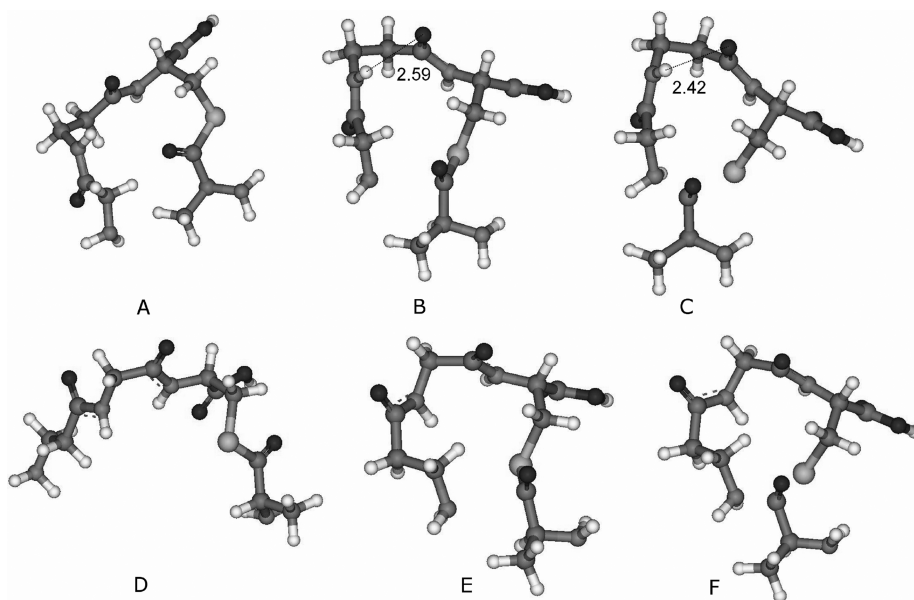
Intramolecular complexes (IC) featuring the characteristics of a NAC ( $\bar{r}^2 = 2.59$  Å) were obtained by following IRC branches from TSs down to reactants.<sup>64,69</sup> As seen in the statistical analysis, the two geometrical parameters governing the formation of the TSs are  $p^1$  (related to the torsion of the nucleophilic N-terminus) and  $r^2$  (termini separation). Therefore, the amount of energy needed to bring the structurally closest reactive conformer to its corresponding IC appeared as another determinant parameter, requiring an average of 14 kcal·mol<sup>−1</sup> (as a rough estimation, one-third of the energy required to access the TS is related to structural reorganization to the IC). For instance, the reorganization of the (S)-acyl isopeptides **1** to the ICs leading

to **lig**<sub>11</sub> and **lig**<sub>14</sub> required a maximum of ca. 25 kcal·mol<sup>−1</sup>, while this reorganization for **lig**<sub>9</sub> required less than 10 kcal·mol<sup>−1</sup> (Figure 7).

The emergence of intramolecular electrostatic interactions is of fundamental importance in peptide chains: promiscuity of donors and acceptors through a succession of amide bonds, van der Waals interactions between side-chains, and other weak noncovalent interactions. In many cases, these effects are so dramatic that they can trigger aggregation of the peptide.<sup>70–72</sup> The impact of the electrostatic interactions does not stop at the reactive conformers and may have a dramatic impact on the development of TSs.<sup>66,73–75</sup> The results for  $n > 9$  were reexamined to identify the intervention of any intramolecular hydrogen bonds (IHBs), and their impact on  $\Delta G^\ddagger$ , by comparing the structures of reactive (S)-acyl isopeptide conformers, ICs, and TSs (See Figure 7 for representative examples). Hydrogen bonds involved in the folding of the oligopeptide sequence and also involving the reactive termini were considered. Strikingly, the complex dependency of  $\Delta G^\ddagger$  vs sequence/ring size for  $n > 9$  could be correlated to the emergence and the strengthening of intramolecular hydrogen bonds from the ICs to the TSs. In **lig**<sub>11</sub> and **lig**<sub>14a</sub>, the geometric constraints imposed on the system to enter the TS and to maximize stereoelectronic effects<sup>19,22,76</sup> led to a weakening of the pre-existing hydrogen bonds. Comparing **lig**<sub>12a</sub> and **lig**<sub>12b</sub> revealed the absence of hydrogen bonds in **lig**<sub>12b</sub>, while the hydrogen bond folding the structure is strengthened in **lig**<sub>12a</sub> (Figure 7). The same line was verified on **lig**<sub>13a</sub>/**lig**<sub>13b</sub>, **lig**<sub>14a</sub>/**lig**<sub>14b</sub>, **lig**<sub>16a</sub>/**lig**<sub>16a</sub>. The sequence dependency of  $\Delta G^\ddagger$  appeared clearly to be caused by structural features enabling or disabling the emergence of stabilizing hydrogen bonding in the TS. The emergence of four strengthening IHBs in **lig**<sub>17</sub> vs 3 IHBs in **lig**<sub>20</sub> afforded a rationalization of the final jump in  $\Delta G^\ddagger$  from  $n = 17$  to  $n = 20$ .

## CONCLUSION

In conclusion, this computational approach affords the first systematic theoretical background for *n*-*exo-trig* intramolecular S-to-N acyl transfer, a phenomenon encountered in recent



**Figure 7.** Reactive conformer (A = **1**<sub>12a</sub>, D = **1**<sub>12b</sub>), IC (B = **IC**<sub>12a</sub>, E = **IC**<sub>12b</sub>), and TS (C = **lig**<sub>12a</sub>, F = **lig**<sub>12b</sub>) illustrated for S-to-N acyl transfer proceeding through 12-membered macrocyclic TSs. Distances for hydrogen bonding present in B and C = 2.59 and 2.42 Å, respectively.

examples of internal Cys ligation and isomerization of (S)-acyl isopeptides. The reaction is driven by enthalpic effects for the range  $n = 5-20$ . For medium ring sizes ( $n = 5-10$ ), a classical evolution of  $\Delta G^\ddagger$  vs ring size was found, while for larger macrocycles ( $n > 10$ ), a succession of extrema complicated the evolution of  $\Delta G^\ddagger$ . By using a statistical treatment, this work emphasizes the importance of the geometrical parameters  $p^1$  (the torsion imposed to the reactive termini by the geometrical strains) and  $r^2$  (the distance between the nucleophilic amine and the electrophilic alanyl carbonyl) which govern the approach of the reactive termini at the TS. This treatment also emphasized that the substitution at R and R' (Figure 1b) has a little impact on the nature of the TS. The preorganization of the structures, and, in particular, the emergence of stabilizing hydrogen bonds through ICs to TSs appeared as a major governing parameter correlating the variations of  $\Delta G^\ddagger$  to sequence and ring-size. In isomeric structures, the presence of an internal non-natural amino acid ( $\beta$ -Ala or GABA) directly after the Cys residue favored the emergence of stabilizing hydrogen bonds (5.7–8.4 kcal·mol<sup>-1</sup>) which tighten and stabilize the TS. Building on these results, it is anticipated that the incorporation of turn inducers to fold the (S)-acyl isopeptide sequence (e.g., inclusion of proline,<sup>77</sup> pseudoproline residues,<sup>78,79</sup> or  $\alpha,\alpha$ -disubstituted amino acids<sup>80,81</sup>) might have a beneficial impact on the activation barriers for intramolecular long-range S-to-N acyl transfer by promoting internal hydrogen bonding. The competition between the intra- and intermolecular acyl transfers is driven by the parameters that govern the approach of the reactive termini, i.e. the ring strain.

## ■ ASSOCIATED CONTENT

### ■ Supporting Information

Full sequence details for the (S)-acyl isopeptides **1**, Cartesian coordinates for all transition states, (S)-acyl isopeptides **1** and native peptide analogs **2**, summary of the computed thermochemistry, and details of the statistical treatment. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [katritzky@chem.ufl.edu](mailto:katritzky@chem.ufl.edu).

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was supported by the Research Foundation-Flanders (FWO-Vlaanderen; fellowship to J.-C.M.M.) and the Funds for Scientific Research-Belgium (F.R.S-FNRS; G.D. is research associate of F.R.S-FNRS). J.-C.M.M. acknowledge the ICT Department of Ghent University for access to HPC facilities. G.D. thanks the F.R.S-FNRS for access to HPC facilities installed in Liège (Belgium) and Louvain-la-Neuve (Belgium). The authors acknowledge the Kenan Foundation (University of Florida) and the King Abdulaziz University (Saudi Arabia) for their financial support.

## ■ REFERENCES

- (1) Fischer, E.; Fourneau, E. *Ber. Dtsch. Chem. Ges.* **1901**, *34*, 2868–2879.
- (2) Pedersen, S. L.; Tofteng, A. P.; Malik, L.; Jensen, K. J. *Chem. Soc. Rev.* **2012**, *41*, 1826–1844.
- (3) Mende, F.; Seitz, O. *Angew. Chem., Int. Ed.* **2011**, *50*, 1232–1240.
- (4) El-Faham, A.; Albericio, F. *Chem. Rev.* **2011**, *111*, 6557–6602.
- (5) Joullie, M. M.; Lassen, K. M. *ARKIVOC* **2010**, *viii*, 189–250.
- (6) Valeur, E.; Bradley, M. *Chem. Soc. Rev.* **2009**, *38*, 606–631.
- (7) Kumar, K. S. A.; Bavikar, S. N.; Spasser, L.; Moyal, T.; Ohayon, S.; Brik, A. *Angew. Chem., Int. Ed.* **2011**, *50*, 6137–6141.
- (8) Payne, R. J.; Wong, C.-H. *Chem. Commun.* **2010**, *46*, 21–43.
- (9) Tam, J. P.; Yu, Q.; Miao, Z. *Biopolymers (Pept. Sci.)* **1999**, *51*, 311–332.
- (10) Coin, I. J. *Pept. Sci.* **2010**, *16*, 223–230.
- (11) Kent, S. B. H. *Chem. Soc. Rev.* **2009**, *38*, 338–351.
- (12) Hackenberger, C. P. R.; Schwarzer, D. *Angew. Chem., Int. Ed.* **2008**, *47*, 10030–10074.
- (13) Dirksen, A.; Dawson, P. E. *Curr. Opin. Chem. Biol.* **2008**, *12*, 760–766.
- (14) Coltart, D. M. *Tetrahedron* **2000**, *56*, 3449–3491.
- (15) Pedersen, S. L.; Sørensen, K. K.; Jensen, K. J. *Biopolymers (Pept. Sci.)* **2010**, *94*, 206–212.
- (16) Verlander, M. *Int. J. Pept. Res. Ther.* **2007**, *13*, 75–82.
- (17) Vlieghe, P.; Lisowski, V.; Martinez, J.; Khrestchatsky, M. *Drug Discovery Today* **2010**, *15*, 40–56.
- (18) Deakyne, C. A.; Ludden, A. K.; Roux, M. V.; Notario, R.; Demchenko, A. V.; Chickos, J. S.; Liebman, J. F. *J. Phys. Chem. B* **2010**, *114*, 16253–16262.
- (19) Yang, W.; Drueckhammer, D. G. *J. Am. Chem. Soc.* **2001**, *123*, 11004–11009.
- (20) Castro, E. A. *Pure Appl. Chem.* **2009**, *81*, 685–696.
- (21) Yang, W.; Drueckhammer, D. G. *Org. Lett.* **2000**, *2*, 4133–4136.
- (22) Erben, M. F.; Boese, R.; Della Védova, C. O.; Oberhammer, H.; Willner, H. *J. Org. Chem.* **2006**, *71*, 616–622.
- (23) Castro, E. A. *Chem. Rev.* **1999**, *99*, 3505–3524.
- (24) Monbaliu, J.-C. M.; Katritzky, A. R. *Chem. Commun.* **2012**, *48*, 11601–11622.
- (25) Ollivier, N.; Vicogne, J.; Vallin, A.; Drobecq, H.; Desmet, R.; El Mahdi, O.; Leclercq, B.; Goormachtigh, G.; Fafeur, V.; Melnyk, O. *Angew. Chem., Int. Ed.* **2012**, *51*, 209–213.
- (26) Boll, E.; Dheur, J.; Drobecq, H.; Melnyk, O. *Org. Lett.* **2012**, *14*, 2222–2225.
- (27) Yang, R.; Hou, W.; Zhang, X.; Liu, C.-F. *Org. Lett.* **2012**, *14*, 374–377.
- (28) Liu, S.; Pentelute, B. L.; Kent, S. B. H. *Angew. Chem., Int. Ed.* **2012**, *51*, 993–999.
- (29) Haase, C.; Seitz, O. *Eur. J. Org. Chem.* **2009**, 2096–2101.
- (30) Ha, K.; Chahar, M.; Monbaliu, J.-C. M.; Todadze, E.; Hansen, F. K.; Oliferenko, A. A.; Ocampo, C. E.; Leino, D.; Lillicotch, A.; Stevens, C. V.; Katritzky, A. R. *J. Org. Chem.* **2012**, *77*, 2637–2648.
- (31) Hansen, F. K.; Ha, K.; Todadze, E.; Lillicotch, A.; Frey, A.; Katritzky, A. R. *Org. Biomol. Chem.* **2011**, *9*, 7162–7167.
- (32) Katritzky, A. R.; Tala, S. R.; Abo-dya, N. E.; Ibrahim, T. S.; El-feky, S. A.; Gyanda, K.; Pandya, K. M. *J. Org. Chem.* **2011**, *76*, 85–96.
- (33) Katritzky, A. R.; Abo-dya, N. E.; Tala, S. R.; Abdel-Samii, Z. K. *Org. Biomol. Chem.* **2010**, *8*, 2316–2319.
- (34) Oliferenko, A. A.; Katritzky, A. R. *Org. Biomol. Chem.* **2011**, *9*, 4756–4759.
- (35) Yoshiya, T.; Hasegawa, Y.; Kawamura, W.; Kawashima, H.; Sohma, Y.; Kimura, T.; Kiso, Y. *Biopolymers (Pept. Sci.)* **2010**, *96*, 228–239.
- (36) Yoshiya, T.; Ito, N.; Kimura, T.; Kiso, Y. *J. Pept. Sci.* **2008**, *14*, 1203–1208.
- (37) Wang, C.; Guo, Q.-X.; Fu, Y. *Chem. Asian J.* **2011**, *6*, 1241–1251.
- (38) The spatial proximity between the nucleophilic  $\alpha$ -amine and the thioester in the isopeptide intermediate which leads to a facile S-to-N acyl transfer is often referred to as “entropic activation”, although this notion is quite ambiguous and poorly defined.<sup>39–41</sup>
- (39) Pascal, R. J. *Phys. Org. Chem.* **2002**, *15*, 566–569.
- (40) Villa, J.; Strajbl, M.; Glennon, T. M.; Sham, Y. Y.; Chu, Z. T.; Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 11899–11904.
- (41) Gennari, C.; Molinari, F.; Piarulli, U.; Bartoletti, M. *Tetrahedron* **1990**, *46*, 7289–7300.



- (42) Ficht, S.; Payne, R. J.; Brik, A.; Wong, C.-H. *Angew. Chem., Int. Ed.* **2007**, *46*, 5975–5979.
- (43) Yang, Y.-Y.; Ficht, S.; Brik, A.; Wong, C.-H. *J. Am. Chem. Soc.* **2007**, *129*, 7690–7701.
- (44) Payne, R. J.; Ficht, S.; Tang, S.; Brik, A.; Yang, Y.-Y.; Case, D. A.; Wong, C.-H. *J. Am. Chem. Soc.* **2007**, *129*, 13527–13536.
- (45) Brik, A.; Ficht, S.; Yang, Y.-Y.; Bennett, C. S.; Wong, C.-H. *J. Am. Chem. Soc.* **2006**, *128*, 15026–15033.
- (46) Brik, A.; Yang, Y.-Y.; Ficht, S.; Wong, C.-H. *J. Am. Chem. Soc.* **2006**, *128*, 5626–5627.
- (47) Lutsky, M.-Y.; Nepomniaschii, N.; Brik, A. *Chem. Commun.* **2008**, 1229–1231.
- (48) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, revision B.1; Gaussian, Inc.: Wallingford CT, 2009.
- (49) SAS software, version 9.1; SAS Institute: Cary, NC, 2002.
- (50) *Discovery Studio Modeling Environment*, version 3.1; Accelrys: San Diego, CA, 2011.
- (51) Afifi, A.; Clark, V. A.; May, S. *Computer-aided multivariate analysis*; CRC Press: Boca Raton, FL, 2004.
- (52) Martin, R. B.; Hedrick, R. I. *J. Am. Chem. Soc.* **1962**, *84*, 106–110.
- (53) Barnett, R. E.; Jencks, W. P. *J. Am. Chem. Soc.* **1969**, *91*, 2358–2369.
- (54) Galli, C.; Mandolini, L. *Eur. J. Org. Chem.* **2000**, 3117–3125.
- (55) Lightstone, F. C.; Bruce, T. C. *Bioorg. Chem.* **1998**, *26*, 193–199.
- (56) Illuminati, G.; Mandolini, L. *Acc. Chem. Res.* **1981**, *14*, 95–102.
- (57) Cort, A. D.; Illuminati, G.; Mandolini, L.; Masci, B. *J. Chem. Soc., Perkin Trans. 2* **1980**, 1774–1777.
- (58) Galli, C.; Illuminati, G.; Mandolini, L. *J. Org. Chem.* **1980**, *45*, 311–315.
- (59) Galli, C.; Illuminati, G.; Mandolini, L.; Tamborra, P. *J. Am. Chem. Soc.* **1977**, *99*, 2591–2597.
- (60) The relative activation barriers between the intramolecular and the intermolecular process might partially explain why the intermolecular product becomes the main component of the reaction mixture, although we also believe that the experimental conditions (concentration, pH) have also a dramatic impact on the ratio between inter/intramolecular products (see for example ref 30).
- (61) Bruce, T. C.; Lightstone, F. C. *Acc. Chem. Res.* **1999**, *32*, 127–136.
- (62) Menger, F. M. *Acc. Chem. Res.* **1985**, *18*, 128–134.
- (63) Houk, K. N.; Tucker, J. A.; Dorigo, A. E. *Acc. Chem. Res.* **1990**, *23*, 107–113.
- (64) Lightstone, F. C.; Bruce, T. C. *J. Am. Chem. Soc.* **1997**, *119*, 9103–9113.
- (65) Bruce, T. C. *Annu. Rev. Biochem.* **1976**, *45*, 331.
- (66) Karaman, R. *Comp. Theor. Chem.* **2011**, *966*, 311–321.
- (67) Karaman, R. *Tetrahedron Lett.* **2010**, *51*, 5185–5190.
- (68) Karaman, R. *Tetrahedron Lett.* **2009**, *50*, 452–456.
- (69) Lightstone, F. C.; Bruce, T. C. *J. Am. Chem. Soc.* **1996**, *118*, 2595–2605.
- (70) Sohma, Y.; Hirayama, Y.; Taniguchi, A.; Mukai, H.; Kiso, Y. *Bioorg. Med. Chem.* **2011**, *19*, 1729–1733.
- (71) Cao, P.; Raleigh, D. P. *J. Am. Chem. Soc.* **2010**, *132*, 4052–4053.
- (72) Katritzky, A. R.; Haase, D. N.; Johnson, J. V.; Chung, A. *J. Org. Chem.* **2009**, *74*, 2028–2032.
- (73) Schiøtt, B.; Iversen, B. B.; Madsen, G. K. H.; Larsen, F. K.; Bruce, T. C. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12799–12802.
- (74) Cleland, W. W.; Frey, P. A.; Gerlt, J. A. *J. Biol. Chem.* **1998**, *273*, 25529–25532.
- (75) Chen, J.; Mcallister, M. A.; Lee, J. K.; Houk, K. N. *J. Org. Chem.* **1998**, *63*, 4611–4619.
- (76) Kurteva, V. B.; Lyapova, M. J.; Pojarlieff, I. G. *ARKIVOC* **2006**, ii, 91–100.
- (77) Monbaliu, J.-C. M.; Hansen, F. K.; Beagle, L. K.; Panzner, M. J.; Steel, P. J.; Todadze, E.; Stevens, C. V.; Katritzky, A. R. *Chem.—Eur. J.* **2012**, *18*, 2632–2638.
- (78) Cremer, G.-A.; Tariq, H.; Delmas, A. F. *J. Pept. Sci.* **2006**, *12*, 437–442.
- (79) Wöhr, T.; Wahl, F.; Nefzi, A.; Rohwedder, B.; Sato, T.; Sun, X.; Mutter, M. *J. Am. Chem. Soc.* **1996**, *118*, 9218–9227.
- (80) Kaneti, J.; Kirby, A. J.; Koedjikov, A. H.; Pojarlieff, I. G. *Org. Biomol. Chem.* **2004**, *2*, 1098–1103.
- (81) Toniolo, C.; Crisma, M.; Formaggio, F.; Peggion, C. *Biopolymers (Pept. Sci.)* **2001**, *60*, 396–419.