First-Principle Computational Study on the Full Conformational Space of L-Threonine Diamide, the Energetic Stability of Cis and Trans Isomers

Michelle A. Sahai,*,† Szilard N. Fejer,‡ Bela Viskolcz,§ Emil F. Pai,†,|| and Imre G. Csizmadia*,§,⊥

Department of Medical Biophysics, University of Toronto, Toronto Medical Discovery Tower, 101 College Street, Room 5-359, Toronto, Ontario, Canada M5G 1L7, University Chemical Laboratories, Lensfield Road, Cambridge, United Kingdom CB2 1EW, Department of Chemistry and Chemical Informatics, Faculty of Education, University of Szeged, Boldogasszony sgt. 6, Szeged, Hungary, 6725, Department of Biochemistry and Molecular and Medical Genetics, University of Toronto, Toronto, Ontario, Canada M5S 1A8, and Department of Chemistry, University of Toronto, 80 St. George St., Toronto, Ontario, Canada M5S 3H6

Received: April 7, 2006

First-principle computations were carried out on the conformational space of trans and cis peptide bond isomers of HCO−Thr−NH_2. Using the concept of multidimensional conformational analysis (MDCA), geometry optimizations were performed at the B3LYP/6-31G(d) level of theory, and single-point energies as well as thermodynamic functions were calculated at the G3MP2B3 level of theory for the corresponding optimized structures. Two backbone Ramachandran-type potential energy surfaces (PESs) were computed, one each for the cis and trans isomers, keeping the side chain at the fully extended orientation ($\chi^1 = \chi^2 =$ anti). Similarly, two side chain PESs for the cis and trans isomers were generated for the ($\phi = \psi =$ anti) orientation corresponding to approximately the β_L backbone conformation. Besides correlating the relative Gibbs free energy of the various stable conformations with the number of stabilizing hydrogen bonds, the process of trans → cis isomerization is discussed in terms of intrinsic stabilities as measured by the computed thermodynamic functions.

Introduction

Biological Considerations of the First-Principle Study. First-principle studies describing the full conformational space of different amino acids,[1−6] the building units of peptides and proteins, have become a key factor in chemistry, biochemistry and pharmaceutical chemistry, where they are undoubtedly important in correlating and understanding conformation and molecular function.

Threonine is an essential amino acid that cannot be produced by metabolism and therefore must be taken up in the diet.[7] Found in high concentrations in the heart, skeletal muscles and central nervous system,[8,9] threonine is also required to maintain proper protein balance in the body, as well as assist in the formation of collagen and elastin in the skin.[8]

The side chain of threonine (Thr) (Figure 1) contains a secondary alcohol functional group, explaining its role in many proteins of the body. Threonine is often found at active sites of enzymes with especially important catalytic functions, proteases,[10,11] components of the proteosome[12] or members of the glycosylasparaginase as well as γ-glutamyltransferase families. Threonine proteases act very similarly to their serine counterparts. In both cases, surrounding residues activate the hydroxyl groups of the respective amino acid side chains to attack the peptide bond of the substrate.[10−13]

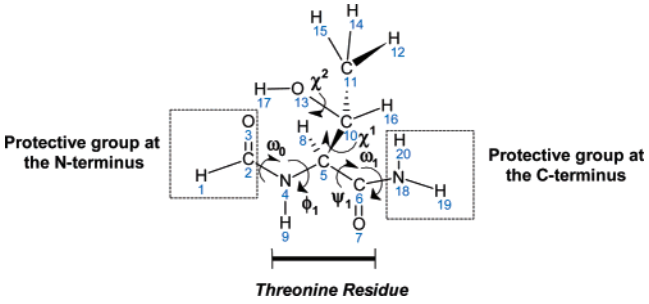


Figure 1. Numbering system employed for HCO−Thr−NH_2 with definitions of backbone and side chain dihedral angles.

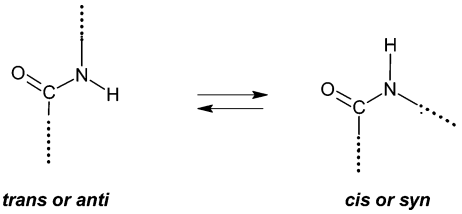
Mature proteins containing carbohydrate side chains of varying lengths are termed glycoproteins. These sugar residues not only considerably influence the conformational and physicochemical properties of the proteins but also are important for biological recognition processes such as cellular adhesion (binding), and in immunological defense by providing coating protection against proteases or antibodies.[14−18] A characteristic feature of core carbohydrates is the covalent linking of the protein chain with the sugar chain via either an *N*-{via asparagine (Asn) or glutamine (Gln)} or *O*-glycosidic bond {via Ser, Thr, tyrosine (Tyr), or hydroxylysine}.[17]

It is hoped that the present in-depth theoretical study will enhance the knowledge that has already accumulated regarding this amino acid by providing a more secure physiochemical basis.[7,19,20]

* Corresponding authors. M.A.S.: e-mail, michelle.sahai@utoronto.ca; tel, (01) 416 581 7548; fax, (01) 416 581 7560. I.G.C.: e-mail, imre.csizmadia@utoronto.ca; tel, (36) 62 544 720; fax, (36) 62 420 953.
† Department of Medical Biophysics, University of Toronto.
‡ University Chemical Laboratories.
§ University of Szeged.
|| Department of Biochemistry and Molecular and Medical Genetics, University of Toronto.
⊥ Department of Chemistry, University of Toronto.

10.1021/jp0680488 CCC: $33.50
Published on Web 09/19/2006

SCHEME 1



Structural Considerations of the First-Principle Study. In general, trans peptide bonds are considered to be more stable than cis peptide bonds, because most peptide bonds in proteins are observed to be in their trans form (Scheme 1). However, that empirical observation does not prove that the trans orientation is intrinsically more stable than the cis form in all cases, and this has been shown to be true in a number of previous publications.[21−27]

In fact, the trans form is not truly a trans isomer because the C−N bond is only slightly stronger than the genuine single bond. This marginally increased strength is due to the conjugative stabilization that is operative in the O═C−NH moiety. The above two forms may differ by less than 5 kcal·mol^{-1}, and the barrier to interconversion is in the vicinity of 20 kcal·mol^{-1}.[28] This barrier height is much less than the value associated with the torsional activation energy of a genuine double bond, which could be as high as 60 kcal·mol^{-1}. Thus, the syn and anti designations may be more appropriate than cis and trans isomers.

In addition to the intrinsic stability of a peptide bond associated with an amino acid residue, intramolecular hydrogen bonds could further stabilize the cis or syn form with respect to the trans or anti form. Therefore, we have chosen to study the threonine diamide to examine the intrinsic stabilities of the two forms of a peptide bond.

It has been previously shown that free amino acid models are not appropriate for modeling peptide folding, because the free amino acid model allows for special stabilizing and destabilizing forces to be included that are not valid for peptides.[29] Therefore, to study the conformational behavior of Thr, a simple peptide model, such as the N- and C-terminally protected amino acid diamide HCO−Thr−NH$_2$, would be more appropriate to model the inductive (through bond) and field (through space) interactions for energetic and electronic density contributions of neighboring peptide residues. This was previously shown to be an acceptable practice for the analysis of peptide models, as only minor differences in geometry and stability are observed.[2]

A previous ab initio study on the trans peptide bond of MeCO−Thr−NH−Me diamide resulted in 39 out of the possible 81 conformations at the RHF/3-21G level of theory.[30] This undoubtedly describes the conformational diversity of Thr and underscores the importance of a conformational study of this amino acid residue. As such, the present study aims to fully understand the conformational space of Thr in the (*S*) absolute configuration of the backbone and in the (*R*) absolute configuration of the side chain.

Geometrical Considerations of the First-Principle Study. For any given peptide with two optimizable dihedral angles about the peptide bond, ϕ and ψ, the laws of multidimensional conformational analysis (MDCA)[3−6,31−34] predict nine possible backbone (BB) conformations (β_L, γ_L, γ_D, δ_L, δ_D, α_L, α_D, ϵ_L, ϵ_D). These conformations are often depicted on a Ramachandran map (Figure 2). In Thr, there are two additional dihedral angles of interest in the side chain (SC): χ^1 and χ^2 (Figure 1). Like the BB dihedrals, each SC dihedral can assume three possible conformations, gauche$^+$ (g$^+$), anti (a) and gauche$^-$ (g$^-$), leading to $3 \times 3 = 9$ possible side chain conformations (Figure 2). As a result, $3(\phi) \times 3(\psi) = 9$ for the BB, and $3(\chi^1) \times 3(\chi^2) = 9$ for the SC, leading to $9 \times 9 = 81$ possible optimizable conformations.

These 81 possible conformers correspond only to a trans isomer of the peptide bond within the backbone. However, the backbone of any amino acid residue in a peptide or in a protein molecule is a triple rotor. The three dihedral angles ϕ, ψ, and ω measure the rotation about the N−C, the C−CO and the OC−NH (i.e., the peptide) bonds respectively, leading to a potential energy hypersurface: $E = f(\phi,\psi,\omega)$. ω can potentially assume values in the vicinity of 0° or 180°. As a result, it is customary to consider a potential energy surface (PES) cross-section of the full potential energy hypersurface (PEHS) such as $E = f(\phi,\psi)$, where ω could be optimized at about 0° or 180°. Most Ramachandran maps of the type $E = f(\phi,\psi)$ found in the literature are associated with the trans peptide bond where ω is ∼180°. However, in keeping with a full conformational study, the present study also aims to investigate the cis peptide bond where ω is ∼0°. Therefore, a total of $81 \times 2 = 162$ possible optimizations will be carried out by examining both the cis and trans isomers of the peptide bond for HCO−Thr−NH$_2$.

Altogether, the present study aims to explore, characterize and present the geometric preferences of HCO−Thr−NH$_2$ developed for both cis and trans isomers. Preliminary work at the RHF/3-21G level of theory and subsequent refinement at the DFT/B3LYP/6-31G(d) level of theory was employed to optimize all possible (162) conformers. The G3MP2B3 method was further used to yield more reliable energies and thermodynamic functions to describe the process of peptide folding in terms of entropy change and backbone−backbone (BB−BB)

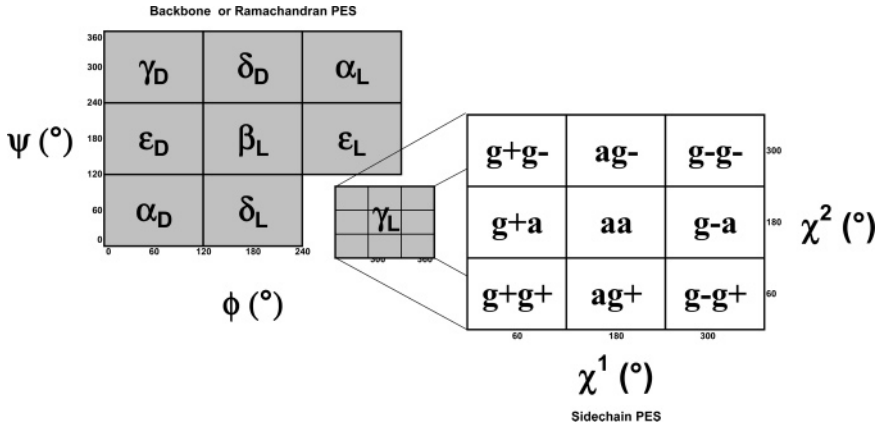


Figure 2. 2D topology of a Ramachandran PEHS $E = f(\phi,\psi)$ of an amino acid residue in a peptide with conformers designated by traditional conventions. An extrapolation of conformations described by the PEHS $E = f(\chi^1,\chi^2)$ is also shown by IUPAC conventions.

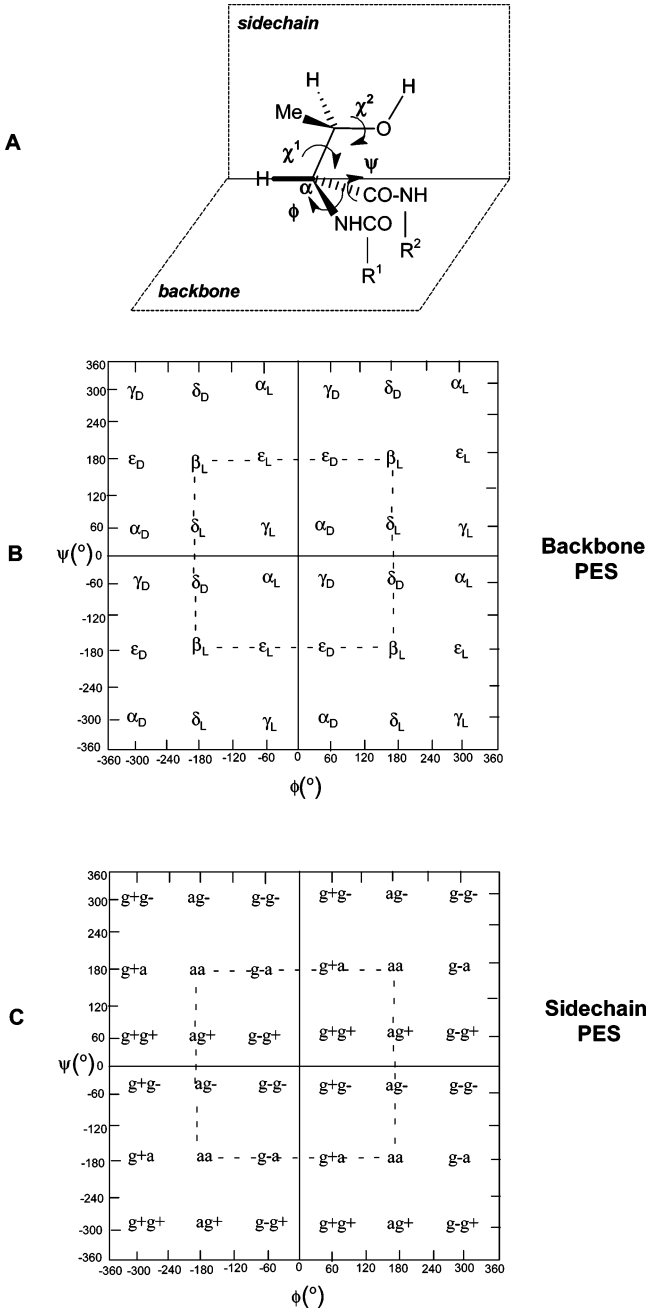


Figure 3. (A) Partitioning the four independent variables (ϕ, ψ, χ^1, χ^2) to backbone (ϕ and ψ) and side chain (χ^1 and χ^2) domains. Unless otherwise stated, $R^1 = R^2 = H$ in the present study. (B) Topology of the Ramachandran backbone PES. (C) Topology of the side chain PES. For both, $\chi^1 = \chi^2 = 180°$.

as well as backbone−side chain (BB−SC) hydrogen bonding interactions. Although this is only a gas phase study, this should establish a primary standard against which future solvation studies may be compared.

Scope. Three topics will be addressed in the present paper: peptide folding as a prelude to protein folding, the role of hydrogen bonding in peptide folding, and the role of trans → cis isomerization in peptide folding.

(1) Peptide folding is a prelude to protein folding. Thus, an in-depth analysis of peptide conformation is an appropriate first step. Peptide chemists traditionally divide the problem of the full conformational space into the backbone and side chain domains, as illustrated schematically for the case of the threonine diamide (Figure 3A). It is clear even from this schematic illustration that the partitioning of the full conformational space is arbitrary, because three rotors (ϕ, ψ, and χ^1) are connected to the α-carbon and must be extensively coupled.

(2) The side chain−backbone interaction via hydrogen bonding is expected to play a significant role in protein folding.

(3) Finally, the role of trans → cis isomerization (Scheme 1) has been postulated to play some role in protein folding either during or after the synthesis of the polypeptide chain. For this reason we shall compute thermodynamic functions (ΔH, ΔG, and ΔS) to characterize the intrinsic stability of cis and trans forms of the peptide bond.

Method

Conformational and Configurational Specifications. Numeric definitions of the relative spatial orientation of all constituent atoms of HCO−Thr−NH_2 follow an established standard,[3,4] shown explicitly in Figure 1. As a result, the Thr amino acid residue as well as the protecting end groups were exclusively defined using the *z*-matrix internal coordinate system to characterize molecular structure, stereochemistry, and geometry. Both cis and trans forms of the first peptide bond (HCO−NH) were investigated.

Molecular Computations of Structures and Energies. All computations were carried out using the Gaussian 03 program package (G03).[35] Each structure was initially optimized using the ab initio[36] restricted Hartree−Fock (RHF)[37] method with the split valence 3-21G basis set.[38−40] Multidimentional conformation analysis (MDCA)[6] was used to define the topologically possible set of conformers represented by a grid-defined set of catchment regions (Figure 2). Presently, it is possible to accurately characterize the topologically probable set of stable conformers emerging from the larger set of topologically possible conformers.[41]

The RHF/3-21G geometry optimized structural parameters were then used as the input in a subsequent theoretical refinement step with the inclusion of electron correlation effects at the B3LYP/6-31G(d) level of theory to obtain more reliable geometry and stability data. Here, B3LYP[42] denotes the combination of Becke's three-parameter exchange functional with the Lee−Yang−Parr (LYP)[43] correlation functional and also employs the mathematically more complete 6-31G(d) basis set. To yield more accurate energies, all stationary points were further refined using the G3-based quantum chemistry method G3MP2B3.[44−46] Energies of this type are labeled as $E^{\text{uncorrected}}$. Total energies are given in hartrees, and the relative energies are given in kilocalories per mole (with the conversion factor: 1 hartree = 627.5095 kcal·mol^{-1}).

Additionally, each stable conformer was subjected to frequency calculations at the B3LYP/6-31G(d) level of theory to confirm their identity as being true minima. The results also provided zero-point energy (ZPE) values, which were scaled using a correction factor and added to the total energy of each conformer to provide more accurate energetic characterization of the conformers as well as the vibrational frequency of each of the normal modes. Corrected energies for these geometries are labeled as $E^{\text{corrected}}$.

Applied frozen core coupled cluster computations performed at the CCSD/6-311++G(d,p) level of theory were also used to test the accuracy of the results obtained for the cis and trans isomers of the $\alpha_D[-+]$ conformer at the G3MP2B3 level of theory.

Potential Energy Surfaces. Potential energy surface scans were made at the RHF/3-21G level of theory by varying the (ϕ, ψ) and (χ^1, χ^2) dihedrals in 10° increments under tight

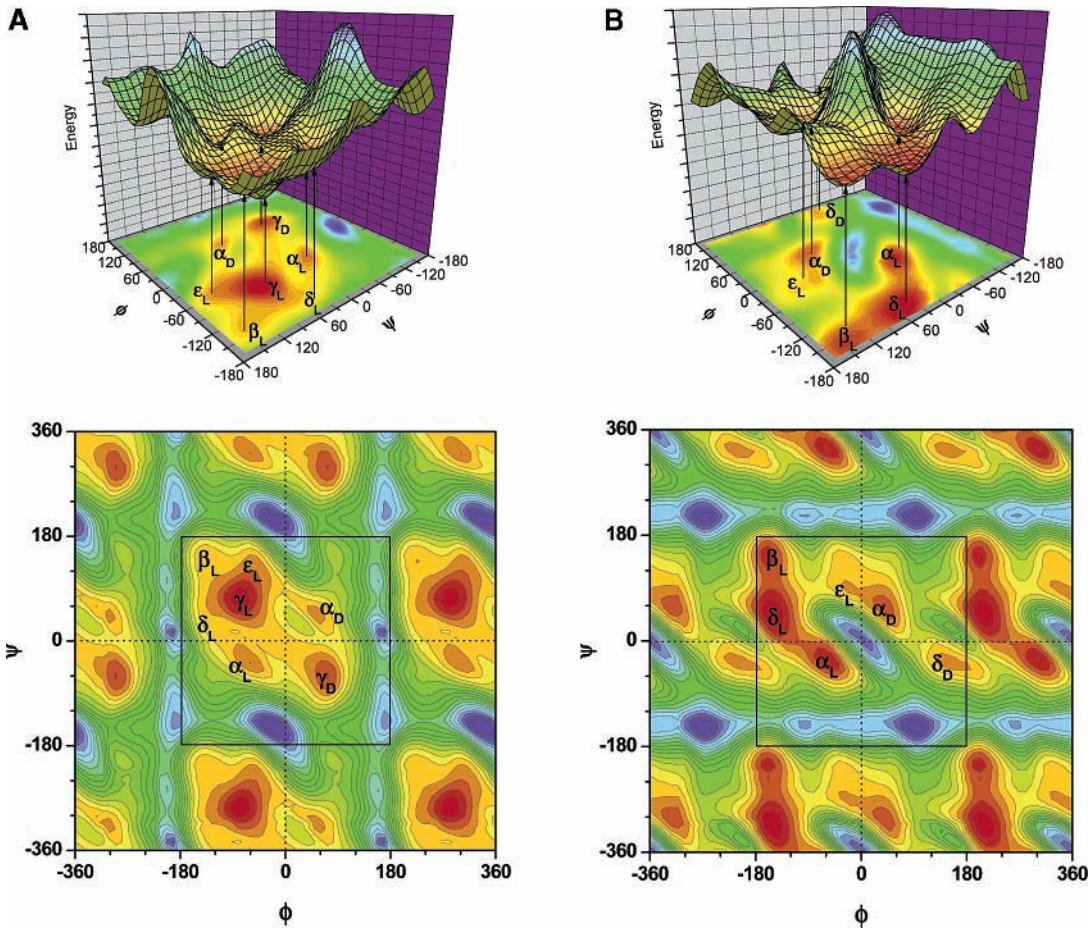


Figure 4. Backbone (BB) PES Ramachandran topology each with two independent variables (ϕ, ψ): (A) trans HCO−Thr−NH_2; (B) cis HCO−Thr−NH_2. A total of 12 × 12 = 144 grid points were computed at the RHF/3-21G level of theory using tight optimization conditions. Note: dihedral angles (ϕ, ψ) are in degrees, and energy is in kcal·mol^{-1}.

optimization options. The Ramachandran type backbone PES, $E = f(\phi,\psi)$, and side chain PES, $E = f(\chi^1,\chi^2)$, are presented as contour diagrams in the range −360° to +360° in Figure 3B,C, respectively. Pseudo-3D representations of these PESs are presented in the IUPAC recommended range of −180° to +180°.[47]

Thermodynamic Functions (ΔH, ΔG and ΔS) and Hydrogen Bonding Interactions. Besides the relative energy ($\Delta E^{\text{uncorrected}}$ and $\Delta E^{\text{corrected}}$) values computed by the G3MP2B3 method, the thermodynamic functional changes of enthalpy (ΔH), Gibbs free energy (ΔG), and entropy (ΔS) were also calculated for the stable conformations. The values of total enthalpy (H) and Gibbs free energy (G) given in Hartrees were converted to their respective relative values, ΔH and ΔG, in kcal·mol^{-1} as described above, whereas the values for entropy (S) and entropy change (ΔS) were given in cal·mol^{-1}·K^{-1}.

Within the G3MP2B3 method, refinement of energy was employed using the perturbation theory with the Møller−Plesset second order (MP2) method in combination with the 6-31G(d) basis set. This was applied to determine the intrinsic stability of HCO−Thr−NH_2 including the energetic contributions from the side chain via intramolecular interactions involving the secondary alcohol functional group with the backbone of the residue.

Results and Discussion

Molecular Conformations. Geometry optimizations were initially carried out at the RHF/3-21G level of theory on the expected 81 trans and 81 cis isomeric forms resulting in only 46 and 36 stable conformers, respectively. Subsequent refinement of these stable conformers at the B3LYP/6-31G(d) level of theory resulted in 40 trans and 30 cis structures. Figures A and B and Tables A and B of the Supporting Information report these geometries, as well as thermodynamic functions computed at the G3MP2B3 level of theory. Additional attempts were made to optimize the missing conformers, but the geometries shifted as recorded in Table C of the Supporting Information. A pair of PES cross-sections of the PEHS for the trans and cis peptide bond containing isomers is shown in Figures 4 and 5. In Figure 4, the PES gives the backbone-fold with variation along ϕ and ψ (Ramachandran potential energy surface). In Figure 5, the other pair of PESs gives the side chain fold involving variation along χ^1 and χ^2. The central points of all surfaces are of identical energies, because $\phi = \psi = \chi^1 = \chi^2 = 180°$.

Only a few mimina are shown on these 2D cross-sections. For example, only 7 minima appear for each of the trans forms for the PES of $E = f(\phi, \psi)$ with $\chi^1 = \chi^2 = 180°$, and $E = f(\chi^1,\chi^2)$ with β_L ($\phi = \psi = 180°$) backbone, for a total of only 14 minima. In the case of 3D cross-sections of the trans form, the PEHSs of $E = f(\phi,\psi,\chi^1)$ with $\chi^2 = 180°$ and $E = f(\phi,\psi,\chi^2)$ with $\chi^1 = 180°$, a total of 34 minima may be found (Figure 6). This was compared to the 40 minima (Table A of the Supporting Information) located on the 4D PEHS. The situation is analogous to the case of the cis form.

Molecular Stability. The relative energies (ΔE) and associated relative thermodynamic functions (ΔH, ΔG, and ΔS) for all conformers of the trans and cis isomers are tabulated in Tables A and B of the Supporting Information, respectively.

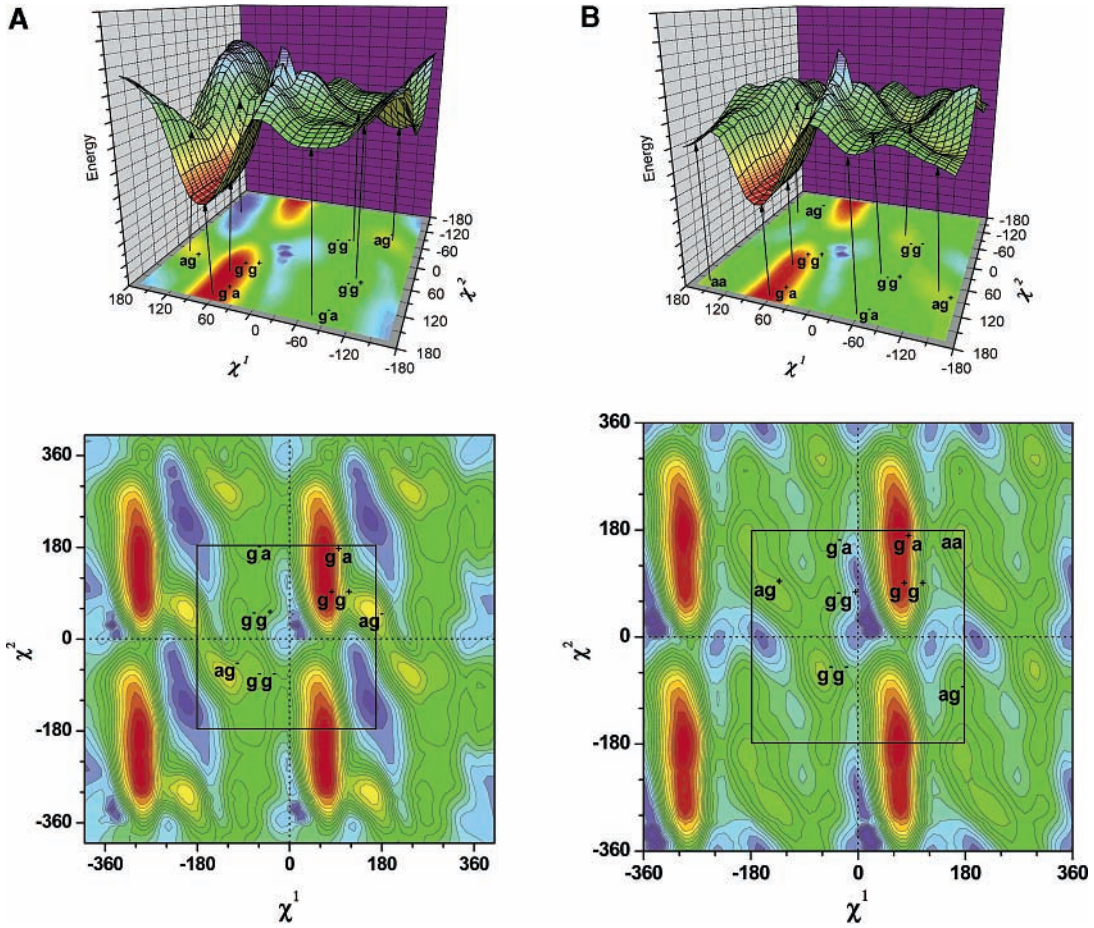


Figure 5. Side chain (SC) PES topology each with two independent variables (χ^1, χ^2). (A) trans HCO−Thr−NH_2; (B) cis HCO−Thr−NH_2. For both panels, $\phi = \psi = 180°$ in an approximate β_L conformation. A total of $12 \times 12 = 144$ grid points were computed at the RHF/3-21G level of theory using tight optimization conditions. Note: dihedral angles (χ^1, χ^2) are in degrees, and energy is in kcal·mol^{-1}.

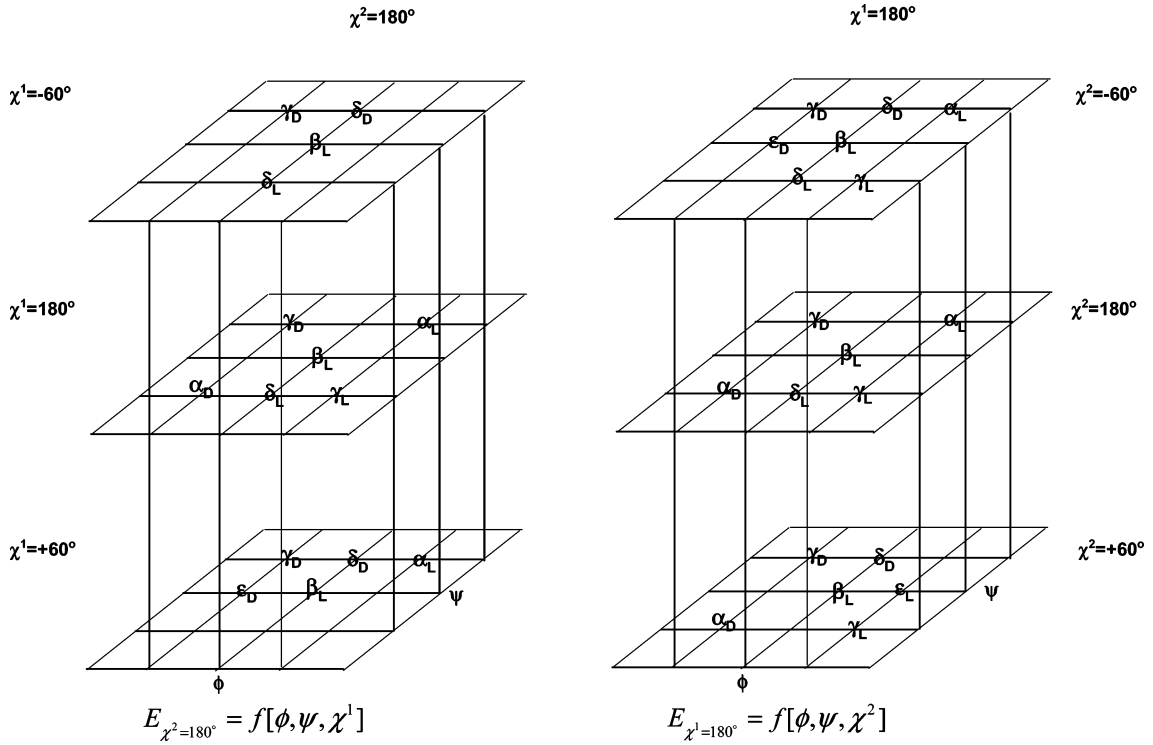


Figure 6. Topology of three independent variable-hypersurface cross-sections of the four independent variable Ramachandran type hypersurface: $E = f(\phi,\psi,\chi^1,\chi^2)$.

The trans global minimum was previously[30] reported to be $\gamma_L[+-]$ at the HF/3-21G level of theory. However, the trans global minimum $\gamma_L[-+]$ reported in this work (Table A of the Supporting Information) was used to calculate relative values for the conformers. It is possible that the discrepancy between these two reported global minima with different side chain conformations is the result of the B3LYP/6-31G(d) level of theory providing theoretical refinement with the inclusion of electron correlation effects lacking in the HF/3-21G level of theory, thereby obtaining a more reliable geometry. Figure 7 provides the schematic comparisons for the various stabilities as well as the relative energies of all conformers calculated with respect to the trans global minimum $\{\gamma_L[-+]\}$ (Table A of the Supporting Information). Thus, the relative energies for all the cis isomers are also calculated with respect to the global minimum of the trans form. The data calculated in this fashion are useful in assessing the relative energetics of trans → cis isomerization.

Hydrogen Bonding. The traditional types of hydrogen bonds involved in backbone/backbone (BB/BB) interactions (types 1A, 1B and 1C) and side chain/backbone (SC/BB) interactions (types 2A to 2D) are shown in Figure 8. The actual optimized structures specifying hydrogen bond distances are shown for the trans and cis isomers in Figures A and B of the Supporting Information, respectively. The hydrogen bond distances are also summarized in the Tables A and B of the Supporting Information, respectively. It is interesting to note that weak hydrogen bonds between the α-proton (i.e., $C^{\alpha}-H$) and the nearest carbonyl oxygen have been observed in some of the conformers. This type of BB/BB hydrogen bond is labeled as Type 1C.

Two additional types of interactions (Types 2E and 2F) involving the amide nitrogen as a proton acceptor are also represented in Figure 8 and reported in Tables A and B. We were interested in the sp^2 bonding characteristic of the amide nitrogen in this model; an attempt was made to determine whether the interaction can serve as a type of side chain/backbone interaction, although previously published works[48−50] did not classify them as true hydrogen bonds. In particular, it was found from analyzing the hydrogen bonding patterns that the cis local minima $\gamma_D[a+]$ and $\delta_D[a+]$ have relatively higher conformational energies of 17.12 and 14.61 kcal·mol^{-1} at the B3LYP/3-21G(d) level of theory, respectively, despite having a shorter N···H−O distance of 2.31 Å as reported in Table B. This may indicate that the type 2E and type 2F interactions proposed in this work cannot be classified as true hydrogen bonds.

The relative Gibbs free energy (ΔG) values (from Figure 7) were plotted against the number of hydrogen bonds present in the various conformers (Figure 9). There are six points to note in connection with this figure:

(i) With an increase in the number of hydrogen bonds, ΔG decreases.

(ii) The cis forms (open circles) are clustered primarily where there are fewer hydrogen bonds, and the trans forms (solid circles) appear to have more hydrogen bonds.

(iii) For hydrogen bond numbers 0 and 1, the lowest points are of the cis form: $\delta_L[aa]$ and $\delta_L[-a]$.

(iv) During trans → cis isomerization, the number of hydrogen bonds present in the trans form is reduced when converted to the cis form. This is illustrated in Figure 9 for $\delta_L[-a]$ and $\beta_L[+a]$.

(v) A "tilted funnel" may be drawn (solid lines) in which we join the highest free energy points and the lowest ΔG points. This funnel converges to the global minimum $\gamma_L[-+]$.

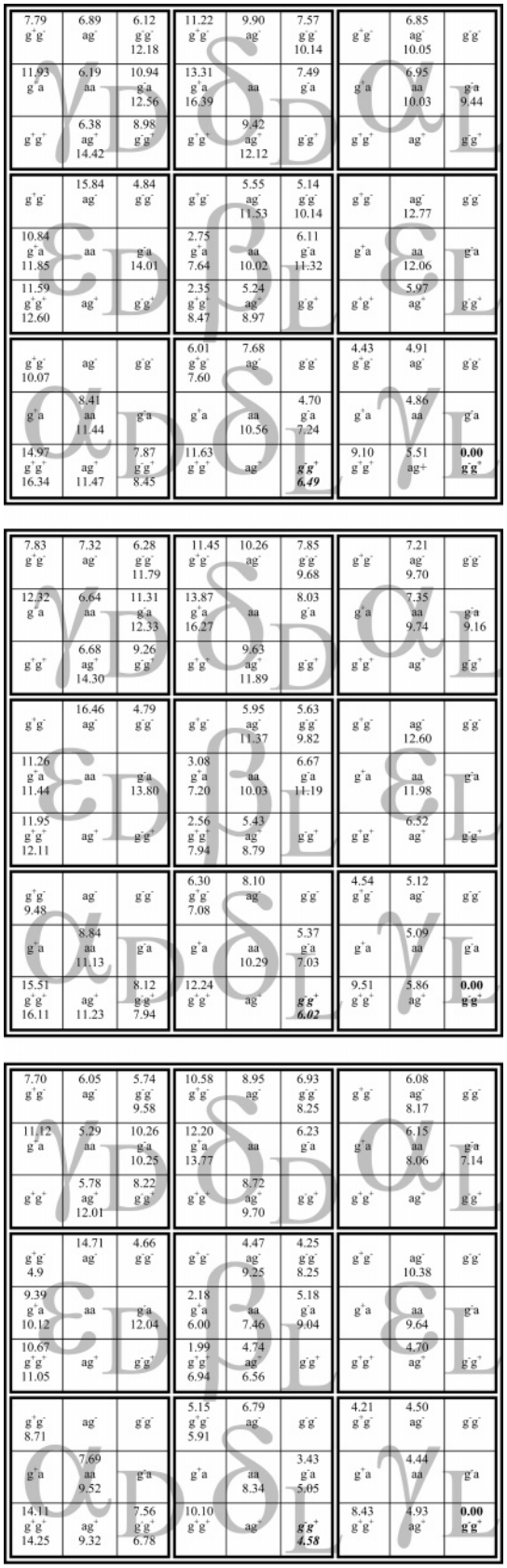


Figure 7. (A) relative energy, (B) relative enthalpy and (C) relative Gibbs free energy represented in 4D Ramachandran topological matrices, for trans HCO−Thr−NH_2 (upper value) and cis HCO−Thr−NH_2 (lower value). The conformational assignments (middle denote) are for the χ^1 and χ^2 orientations, respectively. The difference for the cis energetic values are calculated with respect to the trans global minimum $\gamma_L[-+]$ (bold). Alternatively, the cis global minimum is $\delta_L[-+]$ (bold/italic).

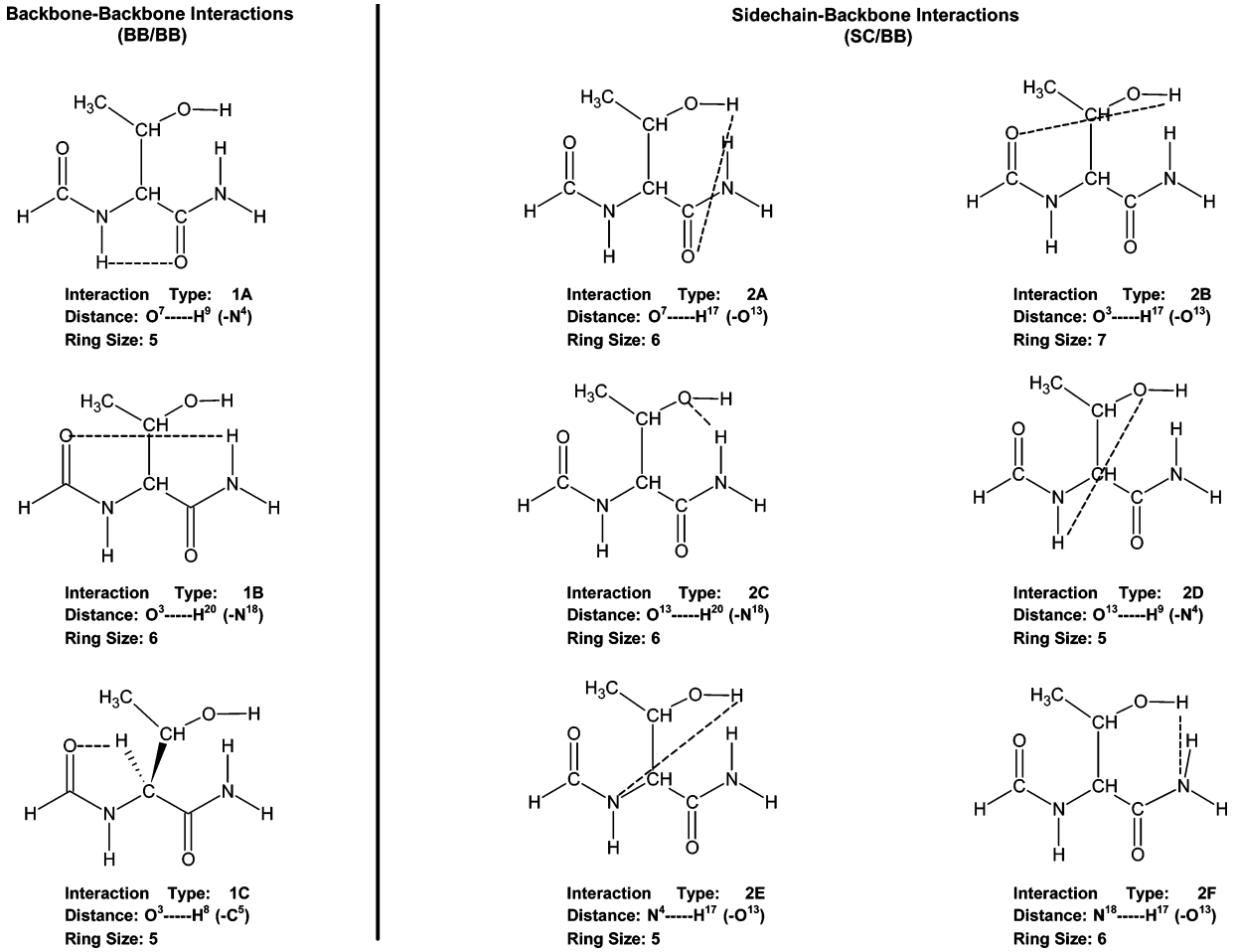


Figure 8. Classification of the types of internal hydrogen bonding for HCO−Thr−NH_2. Corresponding ring sizes indicative of the number of hydrogen bonds are also represented.

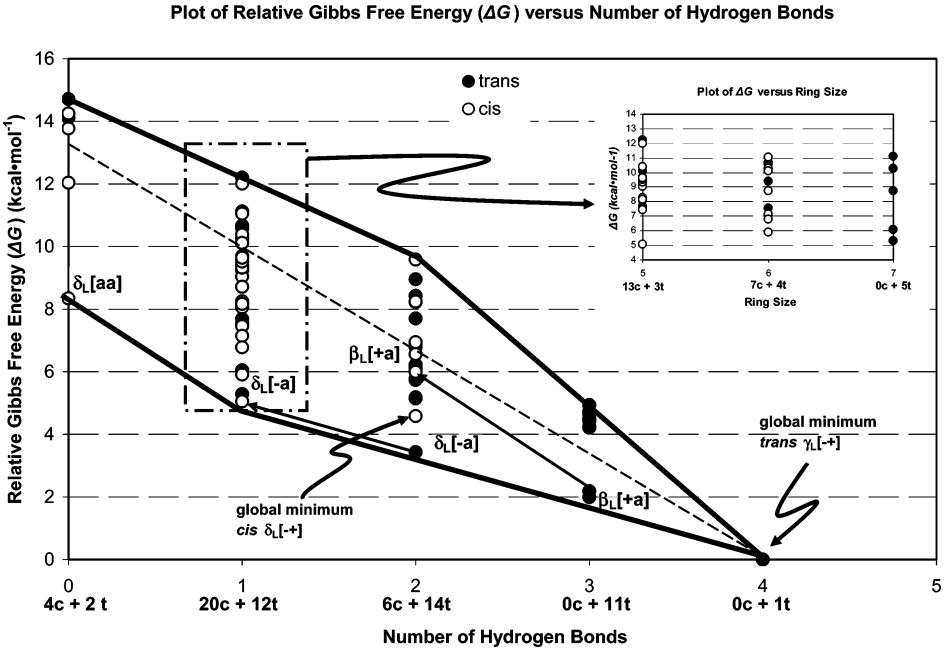


Figure 9. Relative stability (ΔG) versus the number of hydrogen bonds for all possible conformers of cis (open circles) and trans (solid circles) HCO−Thr−NH_2. The difference for the cis energetic values were calculated with respect to the trans global minimum $\gamma_L[-+]$ (bold) according to Figure 7C. The subplot of relative stability (ΔG) versus ring size uses hydrogen bonding schemes from Figure 8 for the trans and cis isomers with one hydrogen bond. In general, the number of stable conformers changes with the number of bonds.

(vi) From the insert at the upper right-hand side of Figure 9, it is clear that a larger ring size favors trans isomers and a smaller ring size favors cis isomers.

Trans → Cis Isomerization. Gibbs free energy (ΔG) values for trans → cis isomerization were computed for those structures in the trans and the cis isomeric forms with the same backbone

TABLE 1: Change of Thermodynamic Functions (All in kcal/mol) for Trans → Cis Isomerization of the Peptide Bond in HCO−Thr−NH_2 at the G3MP2B3 Level of Theory

conformer	$\Delta E^{\text{uncorrected}}$	$\Delta E^{\text{corrected}}$	ΔH	ΔG
β_L[++]	4.985	6.121	5.384	4.952
β_L[+a]	3.865	4.895	4.117	3.817
β_L[a+]	4.556	3.723	3.360	1.825
β_L[a−]	6.206	5.990	5.421	4.774
β_L[−a]	3.982	5.214	4.522	3.862
β_L[−]	3.509	4.999	4.198	3.999
γ_D[−a]	0.634	1.621	1.020	−0.013
δ_L[+−]	0.556	1.589	0.774	0.763
δ_L[−a]	1.347	2.535	1.666	1.622
δ_D[+a]	1.898	3.088	2.408	1.570
δ_D[a+]	3.972	2.694	2.254	0.978
δ_D[−]	1.404	2.570	1.835	1.319
α_L[aa]	1.648	3.080	2.387	1.902
α_L[a−]	1.627	3.198	2.491	2.095
α_D[++]	−0.026	1.370	0.601	0.137
α_D[aa]	1.441	3.030	2.295	1.825
α_D[−+]	−0.971	0.579	−0.186	−0.777
ϵ_D[++]	−0.213	1.013	0.163	0.377
ϵ_D[+a]	0.338	1.007	0.187	0.727

conformation. The ΔE, ΔH, and ΔG values for the isomerization reactions are shown in Table 1.

$\Delta E^{\text{uncorrected}}$ values that were originally negative became positive after ZPE correction ($\Delta E^{\text{corrected}} > 0$). In addition, the calculated ΔG values were systematically smaller for these $\Delta E^{\text{corrected}}$ values. The overall change from the negative $\Delta E^{\text{uncorrected}}$ to ΔG was an increase from the negative toward the positive, suggesting that the trans conformers are more stable than the cis conformers. However, there are two exceptions: the α_D[−+] and γ_D[−a] conformers assumed a slightly negative ΔG values, indicating that the cis form of HCO−Thr−NH_2 was intrinsically more stable and less probable than the trans form, at least for these conformers.

To test the accuracy of the level of theory, applied frozen core coupled cluster computations were performed at the CCSD/6-311++G(d,p) level of theory on the HCO−Thr−NH_2 model peptide. The results indicated that a slightly negative ΔE turned to a virtually thermoneutral ΔG, shown in Table 2.

Figure 10 was obtained by taking the ΔG values from Figure 7C and plotting them against S (Table A of the Supporting Information). There are four points to observe with respect to the data presented in this figure:

(i) The trans forms are clustered primarily around lower entropy values, and the cis forms are clustered around higher entropy values.

(ii) In general, trans → cis isomerizations (shown by solid arrows) have a positive slope; i.e., they are going from trans to cis with increasing ΔG and increasing S.

(iii) There are only two cases of trans → cis isomerization that proceed with increasing ΔG and with entropy reduction. Here, the lines (dashed arrows) have negative slopes for ϵ_D[++] and ϵ_D[+a].

(iv) The nearly thermoneutral isomerization of the γ_D[−a] conformation has an almost horizontal line with the vector pointing from left to right (dotted line). Similarly for the α_D[−+] conformation, the dotted line vector has a negative slope pointing from left to right.

Without examining the relative ΔG values of conformers, as was done in Figure 10, some important points emerge concerning ΔG [trans → cis] isomerization, as shown in Figure 11. It is interesting to observe the following points:

(i) The points may be classified into three domains in terms of their ΔG values (boxes in dashed lines):

$$-1.0 \le \Delta G \le +1.0$$
$$+1.0 \le \Delta G \le +3.0$$
$$+3.5 \le \Delta G \le +5.0$$

(ii) The cis forms found to be computationally stable are at the bottom cluster within the −1.0 to +1.0 kcal·mol^{-1} range.

(iii) The various conformers form four different clusters: β_L on the top and extreme right, ϵ_D at the extreme left, δ_L and δ_D horizontal central and α_L and α_D vertical central.

Conclusions

The exploration of the full conformational space of L-threonine diamide included both cis and trans isomeric forms, each considering 9 backbone (along ϕ and ψ) and 9 side chain (along χ^1 and χ^2) geometries totaling $9 \times 9 = 81$ conformations. Geometry optimizations initially carried out at the RHF/3-21G level of theory were then subsequently refined at the B3LYP/6-31G(d) level of theory, resulting in 40 trans conformers and 30 cis conformers. Single-point energies and thermodynamic functions were further computed for these conformers at the G3MP2B3 level of theory. For the trans form, all 9 backbone conformers exhibited energy minimum conformers (i.e., α_L, α_D, β_L, γ_L, γ_D, δ_L, δ_D, ϵ_L, ϵ_D) whereas only 8 backbone conformers were obtained for the cis form as the γ_L backbone conformation was annihilated at all side chain orientations.

Of the 40 conformers for the trans form and 30 conformers for the cis form, only 19 discrete conformers were found to exist in both cis and trans forms. For these 19 conformers trans → cis isomerization was studied in terms of computed thermodynamic functions (ΔH, ΔS, and ΔG). The ΔG values for the trans → cis isomerization fell within the range −1.0 to +5.0 kcal·mol^{-1}; however, only two conformers, α_D[−+] and γ_D[−a], exhibited negative ΔG values. In these two conformers, the cis isomers appeared to be intrinsically slightly more stable than the trans forms. The cis α_D[−+] structure was in fact 0.777 kcal·mol^{-1} more stable than the trans α_D[−+] form. Consequently, this pair of structures was subjected to a high level of computation, specifically the CCSD/6-311++G(d,p) level of theory, where the isomerization became virtually thermoneutral but the trans form turned out to be 0.039 kcal·mol^{-1} more stable than the cis form. This suggests that the trans → cis isomerization Gibbs free energy computed is sensitive to the method used.

The findings from this peptide model study show that in certain conformations, the cis and trans forms may have nearly identical stabilities, or the cis form may be more stable than

TABLE 2: Change of Thermodynamic Functions for Trans → Cis Isomerization of the Peptide Bond in HCO−Thr−NH_2 at the CCSD/6-311++G(d,p) (Frozen Core) Level of Theory

	energy (uncorrected)			Gibbs free energy		
	total		relative	total		relative
conformer	trans	cis	trans → cis	trans	cis	trans → cis
α_D[−+]	−530.552861	−530.553108	−0.155	−529.857488	−529.8574257	0.039

[a] Note: total energies are given in hartrees and relative energies are given in kcal·mol^{-1}.

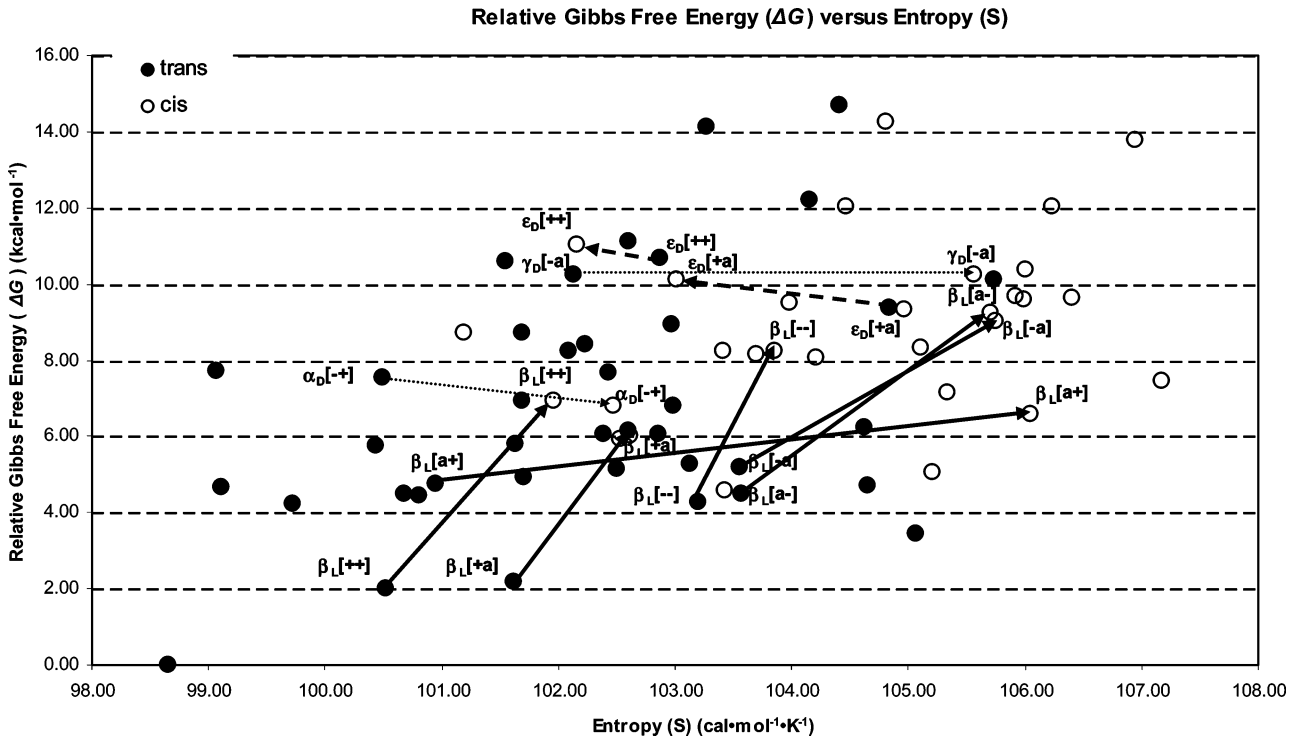


Figure 10. Relative stability (ΔG) versus entropy (S) for all possible conformers of cis (open circles) and trans (solid circles) HCO−Thr−NH_2. The difference for the cis energetic values were calculated with respect to the trans global minimum $\gamma_L[-+]$ (bold) according to Figure 7C.

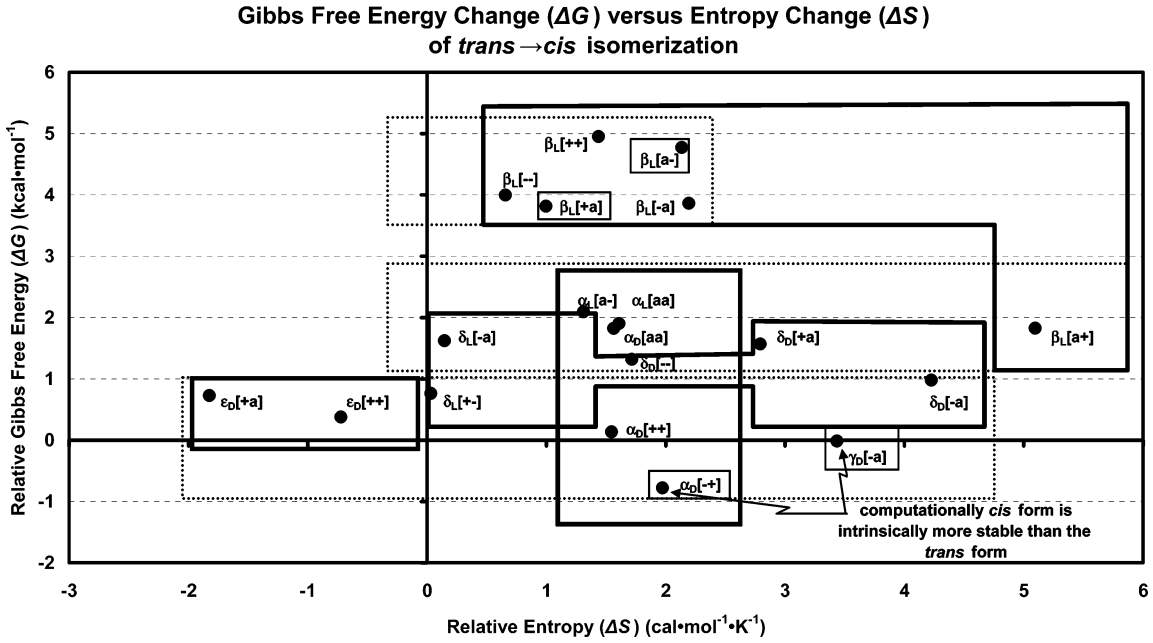


Figure 11. Variation of Gibbs free energy change (ΔG) with entropy change (ΔS) associated with the trans → cis isomerization of HCO−Thr−NH_2.

the trans form. However, there is still a notable preference for the trans peptide bond over the cis form, which agrees with the traditional observation that the majority of all peptide bonds in proteins occur in the trans conformation. As long as the forces that govern protein folding and stability are not completely understood, there can be no exact explanation for the observed rare occurrence of these cis conformations. Some plausible explanations could include the restriction of conformational space available in the presence of a cis peptide bond, 3-D protein structures preferentially stabilizing the trans peptide bond as a consequence of function, kinetics during the protein folding process, or that some cis peptide bonds might have gone unreported in the course of structure determination at lower resolution. However, what is clear is that the occurrence of non-Pro cis peptide bonds in proteins has been associated with steric strain. Their peculiar location near the active sites or binding pockets or at dimerization interfaces could imply that there is a natural tendency to conserve non-Pro cis peptide bonds, emphasizing their importance in protein structure and function.

Taking this into account, the threonine peptide model presented in this paper lays the foundation for future studies whereby a better understanding of the conformational preference and rarity of non-Pro cis peptide bonds can be determined. It is anticipated that theoretically calculated ΔG values of the most

probable cis conformers can be compared to the most probable trans conformers in this gas phase study while subsequently comparing the results with data extracted from the Protein Data Bank (PDB) and observed results from X-ray or NMR experiments.

Acknowledgment. I.G.C. thanks the Ministry of Education of Hungary for the award of a Szent-Györgyi Visiting Professorship. We are grateful to the Hungarian Scientific Research Fund for financial support (OTKA T046861 and F037648). E.F.P. gratefully acknowledges support from the Canada Research Chairs Program, the Natural Sciences and Engineering Research Council of Canada and the Canadian Institutes of Health Research. We gratefully thank Professor Ödön Farkas, Suzanne K. Lau, Jian Payandeh, Konstantin Popovic, Katalin Fejér, András Bakk-Vitális, and Zsuzsanna Jenei for their kind help while preparing this manuscript.

Supporting Information Available: Figure A: Optimized conformations of the 40 trans HCO−Thr−NH_2 showing the relevant hydrogen bond distances computed at the G3MP2B3 level of theory. Figure B: Optimized conformations of the 30 cis HCO−Thr−NH_2 showing the relevant hydrogen bond distances computed at the G3MP2B3 level of theory. Table A: Selected parameters of trans HCO−Thr−NH_2 conformers obtained at the B3LYP/6-31G and G3MP2B3 levels of theory. Table B: Selected parameters of cis HCO−Thr−NH_2 conformers obtained at the B3LYP/6-31G(d) and G3MP2B3 levels of theory. Table C: Summary of conformations obtained for HCO−Thr−NH_2 at the RHF/3-21G level of theory but migrated to a different conformation when re-optimized at the B3LYP/6-31G(d) level of theory. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Lang, A.; Csizmadia, I. G.; Perczel, A. *Proteins* **2005**, *58*, 571.
(2) Sahai, M. A.; Kehoe, T. A. K.; Koo, J. C. P.; Setiadi, D. H.; Chass, G. A.; Viskolcz, B.; Penke, B.; Pai, E. F.; Csizmadia, I. G. *J. Phys. Chem. A* **2005**, *109*, 2660.
(3) Chass, G. A.; Sahai, M. A.; Law, J. M. S.; Lovas, S.; Farkes, Ö.; Perczel, A.; Rivail, J. L.; Csizmadia, I. G. *Int. J. Quantum Chem.* **2002**, *90*, 933.
(4) Sahai, M. A.; Lovas, S.; Chass, G. A.; Penke, B.; Csizmadia, I. G. *J. Mol. Struct. (THEOCHEM)* **2003**, *666*−*667*, 169.
(5) Chass, G. A.; Rodríguez, A. M.; Mak, M. L.; Deretey, E.; Perczel, A.; Sosa, C. P.; Enriz, R. D.; Csizmadia, I. G. *J. Mol. Struct. (THEOCHEM)* **2001**, *537*, 319.
(6) Perczel, A.; Angyan, J. G.; Kajtar, M.; Viviani, W.; Rivail, J. L.; Marcoccia, J. F.; Csizmadia, I. G. *J. Am. Chem. Soc.* **1991**, *113*, 6256.
(7) Lodish, H.; Berk, A.; Zipursky, S. L.; Matsudaira, P.; Baltimore, D.; Darnell, J. E. *Molecular Cell Biology*, 4th ed.; W. H. Freeman & Co.: New York, 1999; p 183.
(8) Damian, M. S.; Koch, M. C.; Bachmann, G.; Schilling, G.; Fach, B.; Stoppler, S.; Trittmacher, S.; Dorndorf, W. *Nervenarzt* **1995**, *66*, 438.
(9) Bottner, M.; Unsicker, K.; Suter-Crazzolara, C. *Neuroreport* **1996**, *7*, 2903.
(10) Hooper, N. M. *Proteins in Biology and Medicine*; Portland Press: London, 2002.
(11) Puente, X. S.; Lopez-Otin C. *Genome Res.* **2004**, *14*, 609.
(12) Seemüller, E.; Lupus, A.; Stock, D.; Lowe, J.; Huber R.; Baumeister W. *Science* **1995**, *268*, 579.
(13) Kraut, J. *Annu. Rev. Biochem.* **1977**, *46*, 331.
(14) Herzner, H.; Reipen, T.; Schultz, M.; Kunz H. *Chem. Rev.* **2000**, *100*, 4495.
(15) Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683.
(16) Hounsell, E. F.; Davies, M. J.; Renouf, D. V. *Glycoconjugate J.* **1996**, *13*, 19.
(17) Vliegenthart, J. F.; Casset F. *Curr. Opin. Struct. Biol.* **1998**, *8*, 565.
(18) Csonka, G. I.; Schubert, G. A.; Perczel, A.; Sosa, C. P.; Csizmadia I. G. *Chem. Eur. J.* **2002**, *8*, 4718.
(19) Lakard, B. *J. Mol. Struct. (THEOCHEM)* **2004**, *681*, 183.
(20) Pawlukojc, A.; Leciejewicz, J.; Tomkinson, J.; Parker S. F. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2001**, *57*, 2513.
(21) Lorenzen, S.; Peters, B.; Goede, A.; Preissner, R.; Frommel C. *Proteins* **2005**, *58*, 589.
(22) Pappenberger, G.; Aygun, H.; Engels, J. W.; Reimer, U.; Fischer, G.; Kiefhaber, T. *Nature Struct. Biol.* **2001**, *8*, 452.
(23) Pal, D.; Chakrabarti P. *J. Mol. Biol.* **1999**, *294*, 271.
(24) Jabs, A.; Weiss, M. S.; Hilgenfeld, R. *J. Mol. Biol.* **1999**, *266*, 291.
(25) Bates, P. A.; Dokurno, P.; Freemont, P. S.; Sternberg, M. J. E. *J. Mol. Biol.* **1998**, *284*, 549.
(26) Odefey, C.; Mayr, L. M.; Schmid, F. X. *J. Mol. Biol.* **1995**, *245*, 69.
(27) Meng, H. Y.; Thomas, K. M.; Lee, A. E.; Zondlo, N. J. *Biopolymers* **2005**, *84*, 192.
(28) Baldoni, H. A.; Zamarbide, G. N.; Enriz, R. D.; Jauregui, E. A.; Farkas, Ö.; Perczel, A.; Salpietro, S. J.; Csizmadia, I. G. *J. Mol. Struct. THEOCHEM (Millennium Volume)* **2000**, *500*, 97.
(29) Chass, G. A.; Lovas, S.; Murphy, R. F.; Csizmadia, I. G. *Eur. Phys. J. D.* **2002**, *20*, 481.
(30) Sahai, M. A.; Motiwala, S. S.; Chass, G. A.; Penke, B.; Pai, E. F.; Csizmadia, I. G. *J. Mol. Struct. (THEOCHEM)* **2003**, *666*−*667*, 251.
(31) Peterson, M. R.; Csizmadia, I. G. *J. Am. Chem. Soc.* **1978**, *100*, 6911.
(32) Peterson, M. R.; Csizmadia, I. G. *Prog. Theor. Org. Chem.* **1982**, *3*, 190.
(33) Kehoe, T. A. K.; Peterson, M. R.; Chass, G. A.; Viskolcz, B.; Stacho, L.; Csizmadia, I. G. *J. Mol. Struct. (THEOCHEM)* **2003**, *666*, 79.
(34) Borics, A.; Chass, G. A.; Csizmadia, I. G.; Murphy, R. F.; Lovas, S. J. *J. Mol. Struct. (THEOCHEM)* **2003**, *666*, 355.
(35) 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.; Bakken, V.; 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.01; Gaussian, Inc.: Wallingford, CT, 2004.
(36) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Theory*; John Wiley & Sons: New York, 1986.
(37) Roothaan, C. C. *J. Rev. Mod. Phys.* **1951**, *23*, 69.
(38) Ditchfie, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724.
(39) Hehre, W. J.; Ditchfie, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257.
(40) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213.
(41) Berg, M. A.; Chasse, G. A.; Deretey, E.; Füzéry, A. K.; Fung, B. M.; Fung, D. Y. K.; Henry-Riyad, H.; Lin, A. C.; Mak, M. L.; Mantas, A.; Patel, M.; Repyakh, I. V.; Staikova, M.; Salpietro, S. J.; Tang, T. H.; Vank, J. C.; Perczel, A.; Farkas, Ö.; Torday, L. L.; Székely, Z.; Csizmadia, I. G. *J. Mol. Struct. THEOCHEM* **2000**, *500*, 5.
(42) Becke, A. D. *J. Chem. Phys.* **1996,** *104*, 1040.
(43) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(44) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 7764.
(45) Baboul, A. G.; Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **1999**, *110*, 7650.
(46) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1999**, *110*, 4703.
(47) IUPAC-IUB Commission on Biochemical Nomenclature. *Biochemistry* **1970**, *9*, 3471.
(48) Baker, E. N.; Hubbard, R. E. *Prog. Biophys. Mol. Biol.* **1984**, *44*, 97.
(49) Stickle, D. F.; Presta, L. G.; Dill, K. A.; Rose, G. D. *J. Mol. Biol.* **1992**, *226*, 1143.
(50) McDonald, I. K.; Thornton, J. M. *J. Mol. Biol.* **1994**, *238*, 777.