Journal of Computational Chemistry

GIAO-PCM Calculations on Alanine Diamide Models Aimed at Predicting Protein Secondary Structures

Journal:	*Journal of Computational Chemistry*
Manuscript ID:	Draft
Wiley - Manuscript type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Hudáky, Ilona; Eötvös Loránd University, Institute of Chemistry, Laboratory of Structural Chemistry and Biology Jákli, Imre; Eotvos University, Department of Organic Chemistry Náray-Szabó, Gábor; Eötvös Loránd University, Institute of Chemistry, Laboratory of Structural Chemistry and Biology Perczel, András; Eotvos Lorand University,
Key Words:	protein secondary structure, relative chemical shift, PCM, ab initio, DFT

GIAO-PCM Calculations on Alanine Diamide Models Aimed at Predicting Protein Secondary Structures

Ilona Hudáky,[a] Imre Jákli,[b] Gábor Náray-Szabó[a] and András Perczel[a,b]*

In this paper we extend our theoretical studies dealing with the dependence of relative proton and carbon chemical shifts (CSs) of protein backbone atoms on their conformational position. In an earlier paper (A. Czajlik, I. Hudáky, A. Perczel, J Comp Chem 2011, 32, 3362) we reported on a fair agreement between calculated and observed backbone CSs as a function of backbone conformation. Applying the polarizable continuum model (PCM) in this work, we compare relative CSs of fully optimized alanine diamide conformers with gas phase calculations and experimental results. Along a path on the Ramachandran surface, we collated calculated relative CSs obtained with and without explicit water molecules, as well as with and without considering the PCM reaction field. Furthermore, we traced the energetically relevant reaction paths along the torsional angle ψ connecting the lowest energy minima (helical, extended, polyproline II and inverse γ-turn) on the Ramachandran plot, with the prospect to facilitate identifying them by their relative CSs. We found that consideration of the solvent effect of the environment around a diamide model improves the agreement with experimental findings on abundant conformers. This agreement is of the level achieved previously by a thorough gas phase investigation on considerably larger oligoalanine models. By relating $\Delta\delta C^{\alpha}$, $\Delta\delta H^{\alpha}$ and $\Delta\delta C^{\beta}$ values of polyproline II and inverse γ-turn to the experimentally well characterized helical and extended data, our calculations contribute to protein secondary structure prediction based on nuclear magnetic CSs.

Keywords: protein secondary structure, relative chemical shift, PCM, ab initio, DFT

[a] *I.Hudáky, G. Náray-Szabó, A. Perczel*
Laboratory of Structural Chemistry and Biology, Institute of Chemistry, Eötvös Loránd University, P.O. Box 32, 1518 Budapest 112, Hungary
E-mail: perczel@chem.elte.hu
[b] *I. Jákli, A. Perczel*
Protein Modelling Group, Hungarian Academy of Sciences-Eötvös Loránd University, 1117 Budapest, Pázmány Péter sétány 1/A, Hungary

Contract/grant sponsor: ICGEB; Contract/grant numbers: CRP/HUN08-03; Contract/grant sponsor: Hungarian Scientific Research Fund; Contract/grant numbers: OTKA K72973, NK67800 and NI-68466; Contract/grant sponsor: European Union and the European Social Fund; Contract/grant number: TAMOP 4.2.1./B-09/KMR-2010-0003.

Introduction

There are intensive attempts to use nuclear magnetic resonance (NMR) chemical shifts (CS) in the determination of protein 3D structures either exclusively, or additionally to other parameters (such as nuclear Overhauser effects, NOE, or residual dipolar couplings, RDC).[1] The first step in protein structure refinement by NMR is the assignment of each CS to a certain nucleus of an amino acid residue along the primary sequence. Beyond the identity of the nucleus and the molecular fold, CSs are influenced by several factors such as the proximity of charged sites, ring current effects, solvent, pH, temperature, *etc*.. Nevertheless those nuclei that either constitute the polypeptide backbone or are very close to it (Figure 1/A) show such differences from nucleus and amino acid specific reference values that depend on the backbone fold of the residues. Accordingly, these data can be used for the estimation of backbone conformations.[2-6] This is why we attempted earlier the identification of the nine typical backbone folds[7] of chiral amino acid residues on the basis of CS calculated for homoconformer polyalanine models in the gas phase.[8] That was a sequel of the description of the CS properties of a simple diamide model.[9,10]

The benefit of the inclusion of solvent effects through the PCM reaction field[11] into the quantum mechanical calculations on small model peptides aimed at predicting the geometry of amino acid residues in proteins was demonstrated in one of our previous studies.[12] As published there, the conformational search of the Ramachandran map using the HCO-L-Ala-NH_2 model peptide at either the RHF/PCM/6-31+G(d) or the B3LYP/PCM/6-31+G(d) level of theory readily delivered two low energy regions, one around conformer α_L (characteristic for the right-handed α-helix) and another around conformers β_L/ε_L (elongated backbone fold). Without restriction on ϕ or ψ torsional angles, conformers α_L and ε_L are not genuine energy minima in the gas phase,[7] however, these regions are well populated in experimentally determined protein structures. Good correlation was found between the relative energies, ΔE, of the representative minima at RHF/PCM/6-31+G(d), B3LYP/PCM/6-31+G(d), as well as B3LYP/PCM/6-311++G(d,p) levels of theory and the logarithm of experimental abundances of alanine residues retrieved from the Protein Data Bank. At the RHF/6-31+G(d) level more alanine residues were assigned around the calculated minima than at other levels. The PCM results on the model peptide HCO-L-Ala-NH_2 referring to an aqueous solution proved to match experimental data of internal alanine residues of proteins just as well as the data of surface alanines, even though the former ones are not surrounded by

solvent molecules. It was concluded that the most important effect of considering solvation within the model is that polarizability of the environment is respected.[12]

Without constrain on the backbone torsional angles ϕ and ψ (Figure 1/B), α_L or ε_L conformations of amino acid residues can be optimized in the gas phase only as parts of larger model peptides. In these, several amide groups are engaged in hydrogen bond formation. A single right-handed helix[13] and a collagen triple helix[14] consist of successive amino acid residues folded into α_L conformation in the former and into ε_L one in the latter case, respectively, and both of these homoconformers can be optimized in the gas phase. Intra- and intermolecular H-bonds within the single right-handed helix as well as the three polyproline II helices of the collagen model stabilize these structures. By reducing the model to a single diamide, however, the H-bonded partners disappear, thus conformers α_L and ε_L converge to other minima. Gas phase optimizations deliver conformer γ_L as the global minimum. This structure is an inverse γ-turn and has an H-bond within itself. In proteins, however, it is occurs seldom compared to helical and polyproline II structures.

As the application of the PCM reaction field[11] improves the performance of a diamide model for an amino acid residue as far as geometry and conformational abundance in proteins is concerned,[12] it seems to be plausible that PCM may help to reflect the dependence of certain CSs on the backbone fold, too. Applying the PCM reaction field, full and partial optimizations of diamide models and subsequent calculations of CSs were performed for the following specific aims:

i) to characterize all fully optimized alanine conformers by relative CSs at various *ab initio* and DFT levels;

ii) to compare relative CSs along a path on the Ramachandran surface calculated with different approaches such as with and without explicit water molecules as well as with and without the PCM reaction field;

iii) to trace the energetically relevant reaction paths along the torsional angle ψ connecting the lowest energy minima, i.e. α_L, β_L, γ_L and ε_L, and to characterize them by relative CSs.

Methods

Calculations were carried out using the Gaussian 09[15] (and occasionally the Gaussian 03[16]) program package (Table 1). The PCM reaction field[11] was applied with water as solvent. The model peptide was HCO-L-Ala-NH_2 (Figure 1/A) in most cases, and CH_3CO-L-Ala-$NHCH_3$ (Figure 1/B) in the few other instances. The HCO-L-Ala-NH_2 model peptide was surrounded by four water molecules (Figure 1/C) in the case of a path on the Ramachandran surface. During optimizations, the GDIIS algorithm[17] was opted for. The PCM reaction field may not be applied with the basis set 3-21G according to an earlier observation indicating that the RHF/PCM/3-21G level reflects gas phase calculations rather than models with explicit solvent molecules.[12] Chemical shielding tensors were calculated with the GIAO method. Isotropic chemical shielding values (σ) were collected for the selected nuclei (amide nitrogen and hydrogen, α- and β- carbon, α-hydrogen, as well as carbonyl carbon, see Figure 1/A). CSs (δ) were not referenced to molecules such as TMS or NH_3. Instead, relative CSs ($\Delta\delta$) were deduced directly from chemical shielding values:

$$\Delta\delta = \sigma_{ref} - \sigma \qquad (1)$$

The preference of relative values over absolute ones for correlating experimental chemical shifts in peptide analogues to computed *ab initio* or DFT results was suggested by Pulay and coworkers.[18] In the present study, the reference (σ_{ref}) is the right-handed helical α_L conformer[8] except when otherwise stated.

Detection and characterization of fully optimized alanine conformers

A thorough conformational search[12] of the HCO-L-Ala-NH_2 model peptide using the PCM reaction field as implemented in Gaussian 03 revealed the existence of minima in seven out of the nine conformational regions of the Ramachandran map. Multiplication of minima was typical. In the present study, optimization of all nine ideal amino acid conformers (α_L, β_L, γ_L, δ_L, ε_L, α_D, γ_D, δ_D and ε_D, Figure 2)[7] was attempted by four different levels of theory using the polarizable continuum model (RB3LYP/PCM/6-311+G(d,p), RB3LYP/PCM/6-311++G(2d,2p), RHF/PCM/6-311++G(d,p) and RHF/PCM/6-311++G(2d,2p)). The performances of program packages Gaussian 03 and Gaussian 09 were compared at the level RB3LYP/PCM/6-311++G(2d,2p). Difference caused by taking CH_3COAlaNHCH$_3$ instead of

HCOAlaNH$_2$ as the model peptide was tested at levels RB3LYP/PCM/6-311+G(d,p) and RHF/PCM/6-311++G(d,p).

Chemical shielding values (σ) were calculated usually at the theoretical level identical to that of the optimization. At levels RB3LYP/6-311++G(2d,2p) and RHF/6-311++G(2d,2p), however, the effect of omission of PCM from CS calculations was investigated. Relative CSs ($\Delta\delta$) at each inspected level of theory were referenced to the respective CS in the right-handed helical α_L conformer.

Comparison of reaction paths calculated with different approaches

Greatly differing conformers accommodate surrounding water molecules dissimilarly.[19] In a thorough conformational search, the solvent molecules should be placed around the solute according to a preceding molecular mechanical investigation. However, the H-bond pattern of CH_3CO-L-Ala-$NHCH_3$ and surrounding four water molecules is very similar in the cases of conformers β_L and ε_L (β'_2 and P_{II} by the original notation).[19] Namely, water molecules form two pairs, and each pair connects an amide H to the carbonyl O of the other peptide group (Figure 1/C). It was therefore expected and later positively confirmed that four water molecules placed in such a way around HCO-L-Ala-NH_2 would keep the H-bond pattern along the whole reaction path between conformational regions β_L and ε_L (Figure S-1). Hence a β_L - ε_L path allows the same polarized moieties to remain engaged in H-bonds with explicit water molecules during the whole scan. It is thus ideal for the comparison of approaches with and without explicit water molecules.

The β_L - ε_L path was traced at five values of torsional angle ϕ: 180°, 210°, 240°, 270° and 300° (*i.e.* -180°, -150°, -120°, -90° and -60°). Five different RHF approaches (methods I to M in Table 1) were used. Method J is the only one where the level of optimization (RHF/3-21G) differs from that of the CS calculation (RHF/6-311++g(d,p)). Methods J, K and L involve 4 explicit water molecules in the model. In methods I and L the PCM reaction field is applied. Method M refers to a pure gas phase calculation. This is the only one where more restriction had to be applied than that of the torsional angle ϕ. Because the conformation at ϕ=270° without any constrain on ψ converged to the γ_L region, ψ was set equal to 144°. Relative CSs ($\Delta\delta$) at each inspected level of theory were referenced to the respective chemical shielding values of the conformation with ϕ= 270° (*i.e.* -90°).

The reaction path from conformer α_L through γ_L reaching either β_L or ε_L

Reaction path along the torsional angle ψ was traced by method I (RHF/PCM/6-311++G(d,p) level of theory) because it performed among the best methods (F, H and I) in reflecting experimental protein data (see section bellow). As far as initial conformations are concerned, the tracing of two paths was attempted: δ_D-δ_L-β_L and α_L-γ_L-β_L. At the grid points, torsional angle ψ was fixed at -90°, -60°, -30°, 0°, 30°, 60°, 90°, 120°, 150° and 180°, and all other parameters were let relax. Relative CSs ($\Delta\delta$) were referenced to the respective chemical shielding values of the right-handed helical α_L conformer.

Results and Discussion

Detection and characterization of fully optimized alanine conformers

Optimizations of all the nine ideal conformers were attempted by five different methods using the polarizable continuum model (Table 2, Figures 2 and 3/A). Only seven of them were found to be genuine minima, while conformations δ_D and δ_L converged to other ones. Additionally, conformer γ_L disappeared on optimization with Gaussian 03 (method A: at PCM/RB3LYP/6-311++G(2d,2p) level of theory). This result is in accordance with the previously published[12] sets of conformers obtained with Gaussian 03 at levels RHF/PCM/6-31+G(d), B3LYP/PCM/6-31+G(d) and B3LYP/PCM/6-311++G(d,p). On the contrary, conformer γ_L proved to be a genuine minimum with Gaussian 09 and has low relative energy (between 1 and 1.5 kJ.mol^{-1} by RB3LYP methods and between 4 and 4.4 kJ.mol^{-1} by RHF methods, see Table 2).

The position of a conformer relative to α_L on the 2D relative chemical shielding plots ($\Delta\delta H^{\alpha}$-$\Delta\delta C^{\alpha}$, $\Delta\delta C^{\alpha}$-$\Delta\delta C^{\beta}$ and $\Delta\delta H^{\alpha}$-$\Delta\delta C^{\beta}$) is qualitatively independent of the applied method (Figures 3/B-C). Numerical differences, however, occur (Table S-I). When PCM is not included in the CS calculation (methods C and G) $\Delta\delta H^{\alpha}$ values shift upfield for all conformers without exception (Figures 3/B and C). This is because the absolute chemical shielding (σ) decreases for α_L while increases for all other conformers. When compared to *ab initio* (methods F to G), DFT results (methods A to E) give downfield shifted $\Delta\delta C^{\alpha}$ for conformer γ_D, and upfield shifted $\Delta\delta C^{\beta}$ for conformers α_D, ε_D and γ_L. With all methods, the great shifts of the rare conformers (α_D and γ_D and ε_D) are striking (downfield δC^{α}, upfield δC^{β} and $\Delta\delta H^{\alpha}$), however, not observed in proteins.

Nuclei C^{α}, H^{α} and C^{β} were previously selected as those that best reflect the secondary structure of amino acid residues by their relative CSs.[8] Relevant experimental results published there are now quoted and compared to $\Delta\delta C^{\alpha}$, $\Delta\delta H^{\alpha}$ and $\Delta\delta C^{\beta}$ of the seven optimized conformers (methods B to I, Table S-II). The best performance is obtained by methods F, H and I, where R^2 is 0.42-0.43 for $\Delta\delta C^{\alpha}$, 0.69-0.71 for $\Delta\delta H^{\alpha}$ and 0.69-0.75 for $\Delta\delta C^{\beta}$. Correlation coefficients are highly similar to relevant data (0.42, 0.80, and 0.73, respectively) obtained by a very extensive gas phase investigation carried out on oligoalanine models which involved even double-stranded β-sheets.[8] Without differentiating between

inward and outward looking H^{α} nuclei, correlation coefficients for the nine conformers of single stranded oligoalanines would lower to 0.37, 0.65, and 0.42, respectively.

There are analogous gas phase data on the alanine diamide model available in the literature.[9] Collating them to the above quoted experimental results, one theoretical level proves to be adequate (R^2 is 0.41 for $\Delta\delta C^{\alpha}$, 0.69 for $\Delta\delta H^{\alpha}$ and 0.56 for $\Delta\delta C^{\beta}$ at GIAO-RHF/TZ2P//B3LYP/6-311++G** when conformer α_L is represented by a model with α-helix-like backbone constrains). This level predicts, however, that the δH^{α} difference between β_L and α_L, is as small as 0.04 ppm.[9] This is in poor accordance with the experimental data of 0.57 ppm.[8] The same value ranges between 0.15 and 0.49 when the CSs are calculated with PCM (methods A-B, D-F and H-I, Table S-II) and becomes negative when PCM is not involved in the CS calculation (methods C and G). All in all, the very important advantage of the application of the PCM model lies in its simplicity coupled with fair accuracy of the obtained results even on a model as small as a diamide.

Correlations between the applied theoretical levels are tabulated in the Supporting Information (Table S-III). As mentioned above, conformer γ_L disappears when treated with G03 instead of G09. Correlation between the two program packages (methods A and B), R^2 is the worst (i.e. 0.29) for $\Delta\delta H^{NH}$, 0.93 for ΔE and $\Delta\delta N^{NH}$, and 0.98-1.00 for all other relative CSs and the torsional angles. No significant difference is detected when the model is changed from HCO-L-Ala-NH_2 to CH_3CO-L-Ala-$NHCH_3$ (methods E vs. D, and I vs. H) as R^2 is 1.000 for ϕ and ψ, 1.00 for ΔE, 0.99 for $\Delta\delta C^{\alpha}$, $\Delta\delta H^{\alpha}$ and $\Delta\delta C^{\beta}$, and not less than 0.95 for other CSs. *Ab initio* and DFT results can be scaled excellently to each other (methods B vs. F, D vs. H) with R^2 as large as 0.96, except for ΔE, $\Delta\delta C^{\beta}$ and $\Delta\delta C'$ (R^2=0.8-0.9). When PCM is neglected in the single point calculation of energies and CSs (methods B vs. D, F vs. G), correlation is poor for ΔE and $\Delta\delta C'$ (R^2=0.4-0.6), but excellent for aliphatic carbons (R^2=0.99).

Comparison of a reaction path calculated with different approaches

Along the β_L - ε_L path (Table S-IV), the unrestricted torsional angle ψ remained within the interval 130°-165° (Figure 3/A). The only exception is the grid point at ϕ=270° (i.e. -90°) by pure gas phase calculation (method M), where ψ had to be constrained, and thus set to 144°, in order to avoid convergence to the γ_L region.

The local minimum with four explicit water molecules in the gas phase along the path is ε_L as obtained with the basis sets 3-21G and 6-311++G(d,p) (ϕ = –92° and –89°,

respectively (see methods J and K in Table S-IV and Figure 4/A). It is interesting to note that the RHF/6-311++g(d,p)//RHF/3-21G single point energy minimum is closer to $\phi = -120^o$ than to -90^o. The reason for this may be the difference in optimal values of parameters (bond lengths and angles) obtained by these two basis sets. Though the ΔE curves by methods J and K show only light increase up to $\phi = -150^o$ (ΔE is not greater than 3.5 kJ.mol^{-1}), the β_L minimum could not be optimized. A greater effort on optimally placing solvent molecules around the solute might have rendered a β_L minimum, as reported for CH_3CO-L-Ala-$NHCH_3$.[19] In that case, conformer β_L ($\phi = -151^o$) has an energy of 7.89 kJ.mol^{-1} higher than that of the global minimum ε_L ($\phi = -94^o$). This energy difference is the double of the "single point" ΔE obtained by methods J and K.

Retaining the explicit solvent molecules and applying the PCM reaction field (method L) gives very similar results (ε_L minimum at $\phi = -82^o$). On the contrary, the local minimum without explicit water molecules is β_L (methods I and M). In the gas phase (method M), neither conformer ε_L, nor the grid point at ϕ=270^o (i.e. -90^o) without constraining ψ could be optimized, instead both converge to the γ_L region. Applying the PCM (method I), however, ε_L is a low energy minimum (0.48 kJ.mol^{-1} above conformer β_L, method I).

Clearly, the PCM β_L - ε_L path does not reflect exactly the ΔE curve obtained with four explicit water molecules. This should not be considered as a deficiency, however, when the goal is to interpret experimental protein Ramachandran plots, because β_L and ε_L appear there as two "conformational attractors"[20] of fusing spots.[12]

Contrary to energetic differences, relative CS, $\Delta\delta C^\alpha$, $\Delta\delta C^\beta$ and $\Delta\delta H^\alpha$ appear very similar as obtained by the five inspected methods (Figures 4/B-D). $\Delta\delta C^\alpha$ gives Λ-shaped, while $\Delta\delta H^\alpha$ shows V-shaped curves with their respective maximum and minimum in the vicinity of ϕ=240^o (i.e. -120^o). $\Delta\delta C^\beta$ curves seem to have a minimum at a somewhat different value of ϕ, between 210^o and 240^o, where the C'_{formyl}–N^{NH}–C^α– C^β torsion is almost 90^o. The various methods give very similar $\Delta\delta C^\beta$ data, except for method J. After fitting a line on points (either chemical shielding, or absolute or relative chemical shielding, σ, δ, $\Delta\delta$) of either C^α or H^α with ϕ greater than 240^o and another line on those with ϕ not less than 240^o, these lines intercept at 241^o<ϕ<250^o. By method I, the intercept for $\Delta\delta C^\alpha$ and $\Delta\delta H^\alpha$ is equally at 246^o (-114^o). When $\phi \approx$ -114^o, H^α lies in the plane of the first amide group (C'_{formyl}–N^{NH}–C^α–H^α torsion is about zero). At lower values, H^α is on the same side of the first amide plane as C', while at greater values, H^α moves to the side of C^β. For values -180^o < ϕ < -114^o (given in

degree), σC^{α} is approximately (0.05698ϕ+159.4) ppm, while σH^{α} is (-0.008155ϕ+27.14) ppm by method I. For values $-114^{o} < \phi < -60^{o}$, σC^{α} is about (-0.07446ϕ+144.4) ppm, while σH^{α} is (-0.01520ϕ+29.80) ppm. As a consequence, a conformation in the β_L region and another in ε_L may be equally far away from $\phi = 114^{o}$, and therefore both their δC^{α} and δH^{α} values are identical. In such case, their assignation to either of these regions has to rely on δC^{β}.

Reaction path from conformer α_L through γ_L reaching either β_L or ε_L

According to the previous section, a path of low energy barrier exists between conformers β_L and ε_L. In contrast, the search along the torsional angle ψ (Figures 5 and S-3) does not indicate whether both conformers β_L and ε_L are directly connected to γ_L, or only one of them. Torsional angles ϕ and ψ may differ from the ideal values obtained by diamide optimizations. It is therefore very important to know, what overlaps are to be expected among the most populated conformers. According to Figures 5/C-E, characteristic C^{α}, C^{β} and H^{α} CSs of conformers ε_L and γ_L can be positioned between typical helical and extended values:

$$\Delta\delta C^{\alpha}(\alpha_L) >> \Delta\delta C^{\alpha}(\varepsilon_L) > \Delta\delta C^{\alpha}(\beta_L) > \Delta\delta C^{\alpha}(\gamma_L) \quad (2)$$

$$\Delta\delta C^{\beta}(\beta_L) > \Delta\delta C^{\beta}(\alpha_L) \approx \Delta\delta C^{\beta}(\varepsilon_L) > \Delta\delta C^{\beta}(\gamma_L) \quad (3)$$

$$\Delta\delta H^{\alpha}(\beta_L) \approx \Delta\delta H^{\alpha}(\gamma_L) > \Delta\delta H^{\alpha}(\varepsilon_L) \approx \Delta\delta H^{\alpha}(\alpha_L) \quad (4)$$

Thus both ε_L and γ_L give rather extended-like δC^{α} (upfield), and helical-like δC^{β} (upfield), while δH^{α} is extended-like for γ_L (downfield) and helical-like for ε_L (upfield). These tendencies may explain several cases of secondary structure prediction where it is contradicting between the nuclei C^{α}, C^{β} or H^{α}. When confronting the above inequalities with experimental results of reference [8], only the upfield shift of $\Delta\delta C^{\beta}(\gamma_L)$ in (3) cannot be traced on the protein alanine 2D CS plots.

Conclusion

We present a simple computational scheme for the estimation of the secondary structure dependence of backbone chemical shifts (CS) in proteins. The application of the polarizable continuum model (PCM) to CS calculations seems plausible as this reaction field proved to improve the performance of a diamide model for an amino acid residue as far as geometry and conformational abundance in proteins is concerned.[12] Relative energies calculated on diamides in gas phase do not succeed in reflecting experimental propensities of protein amino acid backbone folds as conformations α_L and ε_L (α-helical and polyproline II) converge to other minima, and the global energy minimum is γ_L (inverse γ-turn). This problem arises from unsatisfied polarized groups, and can be settled either by longer models, or adding explicit solvent molecules, but the computationally simplest solution is treating the surroundings as a polarizable continuum by the PCM reaction field.

Prediction of relative chemical shifts also improves on applying PCM, as the experimentally well characterized δH^{α} difference between β_L and α_L conformers (extended and α-helical) is now approximated, while it is negligible in gas phase results of diamides. Protein relative CSs are best reflected by GIAO-RHF/PCM methods at 6-311++G(2d,2p) or 6-311++G(d,p) level of theory calculated on the seven conformers optimized at identical levels, where R^2 is 0.42-0.43 for $\Delta\delta C^{\alpha}$, 0.69-0.71 for $\Delta\delta H^{\alpha}$ and 0.69-0.75 for $\Delta\delta C^{\beta}$. Correlation coefficients are highly similar to relevant previously published data (0.42, 0.80, and 0.73, respectively)[8] obtained by a very extensive gas phase investigation carried out on several residue long oligoalanine models involving even double-stranded β-sheets. Inclusion of explicit solvent molecules appears unnecessary when PCM is applied.

Through path calculation connecting the lowest energy minima, i.e., β_L, γ_L and ε_L, the well known and widely used distinction between helical (α_L) and extended (β_L) residues on the ground of relative CSs is extended to the less frequent γ_L and ε_L conformers: Thus both ε_L and γ_L give rather extended-like δC^{α} (upfield), and helical-like δC^{β} (upfield), while δH^{α} is extended-like for γ_L (downfield) and helical-like for ε_L (upfield). This explains several cases of secondary structure prediction where it is contradicting between the nuclei C^{α}, C^{β} or H^{α}.

Figure captions

Figure 1. The investigated model peptides. **A)** Model peptide $HCOAlaNH_2$ with the labels of the most important nuclei. **B)** Model peptide $CH_3COAlaNHCH_3$ with the definition of torsional angles ϕ and ψ. **C)** Model composed of the peptide $HCOAlaNH_2$ and four explicit water molecules.

Figure 2. The Ramachandran map defined by torsional angles ϕ and ψ. The nine regions are named according to Perczel et al.[7] The seven minima of the $HCOAlaNH_2$ model optimized with PCM is presented to characterize the relevant conformational regions.

Figure 3. Geometry and relative chemical shielding of optimized alnine diamide conformers: α_D (black +), α_L (cyan x), β_L (black triangle), γ_D (red hollow square), γ_L (red square), ε_D (green hollow diamond), ε_L (green diamond). For the applied levels of theory see **Table I.** Several data points are signed by the letter of the applied method. **A)** Ramachandran plot, **B)** $\Delta\delta H^{\alpha}$-$\Delta\delta C^{\alpha}$ plot, **C)** $\Delta\delta C^{\alpha}$-$\Delta\delta C^{\beta}$ plot, **D)** $\Delta\delta H^{\alpha}$-$\Delta\delta C^{\beta}$ plot. The origin of every relative chemical shift scale is set at the right-handed helical α_L conformer (see **Table S-I**).

Figure 4. Relative energy and chemical shifts of alanine diamide conformations optimized along the β_L - ε_L path. Data points at ϕ=180°, 210°, 240°, 270° and 300° (i.e. at -180°, -150°, -120°, -90° and -60° refer to structures of constrained ϕ. All other structures are fully optimized β_L or ε_L conformers. **A)** Relative energy over the (lower) minimum of the path. **B)** $\Delta\delta C^{\alpha}$, **C)** $\Delta\delta C^{\beta}$, **D)** $\Delta\delta H^{\alpha}$. Relative chemical shifts are referenced to the data point at 270° (i.e. -90°).

Figure 5. Geometry, energy and relative chemical shifts of alanine diamide conformations optimized along the α_L - γ_L- ε_L- β_L paths. Data points at ψ=-90°, -60°, -30°, 0°, 30°, 60°, 90°, 120°, 150° and 180° refer to structures of constrained ψ. All other structures are fully optimized α_L, γ_L, ε_L or β_L conformers. **A)** Ramachandran plot. **B)** Relative energy over that of the α_L conformer. **C)** $\Delta\delta C^{\alpha}$, **D)** $\Delta\delta C^{\beta}$, **E)** $\Delta\delta H^{\alpha}$. Relative chemical shifts are referenced to the right-handed helical α_L conformer.

References

[1] D. S. Wishart, Prog Nucleic Magn Reson Spectrosc 2011, 58, 62, and references therein.

[2] Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, J Biomol NMR 2009;44:213 and references therein.

[3] A. Cavalli, X. Salvatella, C. M. Dobson, M. Vendruscolo, Proc Natl Acad Sci 2007;104:9615

[4] Y. Shen, P. N. Bryan, Y. He, J. Orban, D. Baker, A. Bax, Protein Science 2010, 19, 349, and references therein.

[5] D. S. Wishart, D. Arndt, M. Berjanskii, P. Tang, J. Zhou, G. Lin, Nucleic Acids Res 2008, 36, W496.

[6] M. Berjanskii, P. Tang, J. Liang, J. A. Cruz, J. Zhou, Y. Zhou, E. Bassett, C. MacDonell, P. Lu, D. S. Wishart, Nucleic Acids Res 2009, 37, W670.

[7] A. Perczel, J. G. Ángyán, M. Kajtár, W. Viviani, L. J. Rivail, F. J. Marcoccia, G. I. Csizmadia, J Am Chem Soc, 1991, 113, 6256

[8] A. Czajlik, I. Hudáky, A. Perczel, J Comp Chem 2011, 32, 3362.

[9] A. Perczel, A. G. Császár, J Comp Chem 2000, 21, 882.

[10] A. Perczel, A. G. Császár, Chem Eur J 2001, 7, 1069.

[11] J. Tomasi, B. Mennucci, R. Cammi, Chem Rev 2005, 105, 2999.

[12] I. Hudáky, P. Hudáky, A. Perczel, J Comp Chem 2004, 25, 1522.

[13] I. A. Topol, S. K. Burt, E. Deretey, T. H. Tang, A. Perczel, A. Rashin, I. G. Csizmadia, J Am Chem Soc 2001, 123, 6054.

[14] V. K. Pálfi, A. Perczel, J Comp Chem 2008, 29, 1374.

[15] M. J. Frisch, et al. Gaussian 09, Revision B.01, Gaussian, Inc., Wallingford CT, 2010 (for complete reference see Supporting Information).

[16] M. J. Frisch, et al. Gaussian 03, Revision E.01, Gaussian, Inc., Wallingford CT, 2004. (for complete reference see Supporting Information).

[17] Ö. Farkas, H. B. Schlegel, Phys Chem Chem Phys Lett 2002;4:11-15.

[18] B. Wang, J. F. Hinton, P. Pulay, J Comp Chem 2002, 23, 493.

[19] W. G. Han, K. J. Jalkanen, M. Elstner, S. Suhai, J Phys Chem B 1998, 102, 2587.

[20] D. Walther, F. E. Cohen, Acta Crystallogr D 1999, 55, 506.

Table 1. Methods for chemical shielding calculations applying different levels of theory.

Method	Gaussian program package	Model peptide	Level of theory
A	G03	$HCOAlaNH_2$	GIAO-RB3LYP/PCM/6-311++G(2d,2p)//RB3LYP/PCM/6-311++G(2d,2p)
B	G09	$HCOAlaNH_2$	GIAO-RB3LYP/PCM/6-311++G(2d,2p)//RB3LYP/PCM/6-311++G(2d,2p)
C	G09	$HCOAlaNH_2$	GIAO-RB3LYP/6-311++G(2d,2p)//RB3LYP/PCM/6-311++G(2d,2p)
D	G09	$CH_3COAlaNHCH_3$	GIAO-RB3LYP/PCM/6-311+G(d,p)//RB3LYP/PCM/6-311+G(d,p)
E	G09	$HCOAlaNH_2$	GIAO-RB3LYP/PCM/6-311+G(d,p)//RB3LYP/PCM/6-311+G(d,p)
F	G09	$HCOAlaNH_2$	GIAO-RHF/PCM/6-311++G(2d,2p)//RHF/PCM/6-311++G(2d,2p)
G	G09	$HCOAlaNH_2$	GIAO-RHF/6-311++G(2d,2p)//RHF/PCM/6-311++G(2d,2p)
H	G09	$CH_3COAlaNHCH_3$	GIAO-RHF/PCM/6-311++G(d,p)//RHF/PCM/6-311++G(d,p)
I	G09	$HCOAlaNH_2$	GIAO-RHF/PCM/6-311++G(d,p)//RHF/PCM/6-311++G(d,p)
J	G09	$HCOAlaNH_2 + 4\ H_2O$	GIAO-RHF/6-311++G(d,p)//RHF/3-21G
K	G09	$HCOAlaNH_2 + 4\ H_2O$	GIAO-RHF/6-311++G(d,p)//RHF/6-311++g(d,p)
L	G09	$HCOAlaNH_2 + 4\ H_2O$	GIAO-RHF/PCM/6-311++G(d,p)//RHF/PCM/6-311++g(d,p)
M	G09	$HCOAlaNH_2$	GIAO-RHF/6-311++G(d,p)//RHF/6-311++g(d,p)

Table 2. Chemical shielding and other data of optimized alnine diamide conformers.

Conf	ΔE	ϕ	ψ	σN^{NH}	σH^{NH}	σC^{α}	σH^{α}	σC^{β}	$\sigma C'$
Method:	**A**								
α_D	8.71	61.8	37.1	107.2	24.26	127.4	27.61	168.2	1.7
α_L	**0.00**	**-86.4**	**-17.8**	102.4	24.72	129.1	27.10	164.1	-0.2
β_L I	0.34	-131.8	143.1	104.7	24.53	133.3	26.60	161.2	0.8
ε_D	15.12	59.5	-144.1	107.4	24.38	127.3	27.73	167.1	2.1
ε_L	0.64	-89.7	143.7	102.3	24.77	132.2	26.89	163.2	-0.5
γ_D	11.49	72.5	-49.3	106.8	24.42	119.8	27.84	165.2	-1.9
Method:	**B**								
α_D	10.13	63.7	34.4	110.7	25.48	126.9	27.92	168.4	3.3
α_L	0.80	-85.6	-15.5	104.7	25.95	128.7	27.41	164.4	1.7
β_L	**0.00**	**-153.6**	**158.6**	109.0	25.00	131.3	27.12	161.5	2.6
ε_D	14.40	58.8	-141.6	110.1	25.56	127.0	28.08	167.1	3.6
ε_L	3.02	-75.2	146.9	104.6	26.00	130.6	27.43	164.3	1.2
γ_D	8.73	72.1	-49.0	109.6	25.65	119.3	28.13	165.2	-0.2
γ_L	1.51	-84.1	69.5	100.7	26.06	132.0	27.09	168.4	2.1
Method:	**C**								
α_D	24.41	63.7	34.4	116.1	26.45	126.4	28.31	167.5	8.0
α_L	14.85	-85.6	-15.5	108.2	26.96	129.5	27.19	163.7	5.9
β_L	3.62	-153.6	158.6	114.7	25.45	131.4	27.25	161.3	4.4
ε_D	28.00	58.8	-141.6	116.7	26.42	127.4	28.41	166.3	8.3
ε_L	15.27	-75.2	146.9	109.8	26.80	131.4	27.60	164.0	4.9
γ_D	9.08	72.1	-49.0	112.2	26.40	119.0	28.27	165.0	4.0
γ_L	**0.00**	**-84.1**	**69.5**	102.7	26.82	132.4	27.18	167.7	5.6
Method:	**D**								
α_D	10.28	64.5	33.3	113.8	25.85	126.0	28.32	168.6	3.8
α_L	**0.00**	-89.6	-12.2	107.6	26.40	127.4	27.80	164.1	2.6
β_L	0.88	-151.9	154.8	111.0	25.44	129.7	27.64	161.6	2.9
ε_D	14.96	58.5	-139.3	111.6	25.78	126.3	28.47	167.7	4.1
ε_L	3.08	-74.5	142.5	105.8	26.24	128.3	27.94	164.3	1.2
γ_D	9.18	73.4	-53.2	111.9	25.89	118.5	28.47	165.6	0.7
γ_L	1.04	**-85.1**	**71.0**	101.7	26.36	130.5	27.56	168.1	2.8
Method:	**E**								
α_D	9.29	64.2	33.7	110.5	25.94	126.7	28.17	168.1	3.7
α_L	0.31	-91.4	-10.6	104.5	26.42	129.4	27.58	163.9	2.2
β_L	**0.00**	-152.9	158.4	108.8	25.50	131.5	27.41	161.1	3.2
ε_D	13.86	59.3	-142.4	109.7	25.98	127.0	28.32	166.9	4.2
ε_L	2.70	-75.5	147.2	104.5	26.39	130.5	27.70	164.1	1.6
γ_D	8.46	72.6	-52.3	109.5	26.04	119.0	28.37	165.0	0.3
γ_L	1.22	**-84.4**	**70.1**	100.5	26.42	132.0	27.41	168.1	2.8

(to be continued)

(**Table 2**. continued)

Method:	**F**								
α_D	9.72	62.5	37.9	147.8	26.70	147.3	28.90	179.7	13.7
α_L	**0.00**	**-80.8**	**-21.0**	143.5	27.08	147.6	28.46	177.7	11.9
β_L	0.54	-150.7	153.8	146.5	26.32	150.5	28.13	174.8	13.6
ε_D	15.20	58.6	-142.7	146.7	26.62	147.3	29.01	179.0	13.9
ε_L	1.24	-71.8	148.6	143.1	27.07	149.4	28.49	177.8	12.5
γ_D	13.36	75.7	-45.1	146.8	26.88	141.4	28.99	178.3	10.6
γ_L	3.96	-88.7	76.2	140.8	27.20	152.0	28.11	179.7	13.7
Method:	**G**								
α_D	22.89	62.5	37.9	152.4	27.74	146.8	29.36	179.0	18.4
α_L	13.91	-80.8	-21.0	147.0	28.15	148.2	28.21	177.0	16.1
β_L	0.88	-150.7	153.8	151.0	26.74	150.5	28.30	174.6	15.6
ε_D	25.91	58.6	-142.7	152.0	27.50	147.4	29.36	178.5	18.2
ε_L	11.12	-71.8	148.6	147.4	27.89	149.7	28.73	177.4	16.2
γ_D	11.20	75.7	-45.1	148.9	27.69	141.2	29.14	178.0	14.5
γ_L	**0.00**	**-88.7**	**76.2**	142.2	27.97	152.2	28.21	179.1	16.9
Method:	**H**								
α_D	10.19	62.9	37.6	148.7	27.17	147.1	29.25	180.0	14.3
α_L	**0.00**	-80.6	-21.6	143.9	27.55	146.6	28.81	177.9	12.4
β_L	1.87	-151.7	153.0	146.9	26.70	149.4	28.56	175.1	13.9
ε_D	16.14	58.4	-140.7	146.3	26.96	146.8	29.32	179.5	14.5
ε_L	1.75	-70.2	145.9	142.5	27.39	148.1	28.93	178.0	12.5
γ_D	14.89	76.8	-51.7	147.4	27.19	140.8	29.25	178.8	11.7
γ_L	4.40	**-89.5**	**77.9**	140.1	27.56	150.9	28.52	179.8	14.4
Method:	**I**								
α_D	9.32	62.7	37.4	147.8	27.10	147.6	29.11	179.9	14.4
α_L	**0.00**	**-81.3**	**-21.2**	143.4	27.45	147.9	28.68	177.9	12.2
β_L	1.11	-149.9	153.9	146.6	26.72	150.9	28.37	174.9	14.2
ε_D	15.18	59.2	-144.0	146.6	26.99	147.6	29.18	179.1	14.7
ε_L	1.59	-71.7	149.2	143.1	27.39	149.7	28.73	178.1	13.0
γ_D	13.45	76.0	-50.4	147.1	27.19	141.3	29.16	178.6	11.6
γ_L	4.15	-89.2	76.0	140.7	27.52	152.2	28.39	179.9	14.5

Method: A-I: as defined in Table I. **Conf:** conformer defined according to Perczel et al.[7]

Global minima are set in bold (**A**: **α_L**, E_h= -417.399952, **B**: **β_L**, E_h= -417.387894, **C**: **γ_L**, E_h= -417.372090, **D**: **α_L**, E_h= -496.026422, **E**: **β_L**, E_h= -417.375569, **F**: **α_L**, E_h= -414.943855, **G**: **γ_L**, E_h= -414.924922, **H**: **α_L**, E_h= -493.018634, **I**: **α_L**, E_h= -414.928398, in hartree) **ΔE**: relative energy above the global minimum in kJ.mol^{-1}. At **methods C** and **G**, relative energies deduced from the relevant single point calculations are given.

Backbone torsional angles ϕ and ψ are given in degree, chemical shielding in ppm.

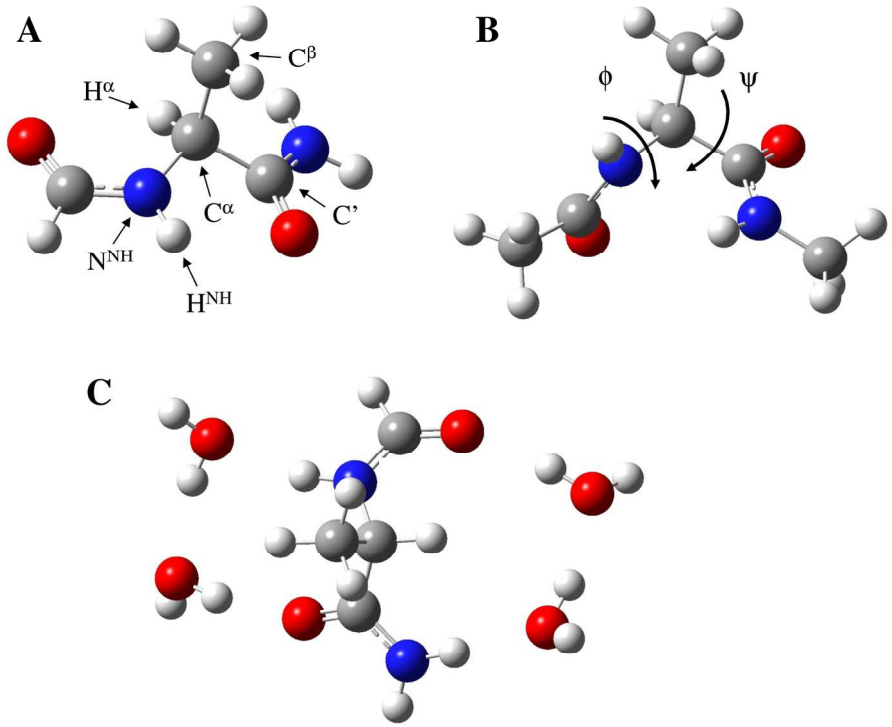


Figure 1. The investigated model peptides. A) Model peptide HCO-Ala-NH_2 with the labels of the most important nuclei. B) Model peptide CH_3CO-Ala-$NHCH_3$ with the definition of torsional angles φ and ψ. C) Model composed of the peptide HCO-Ala-NH_2 and four explicit water molecules.
185x148mm (300 x 300 DPI)

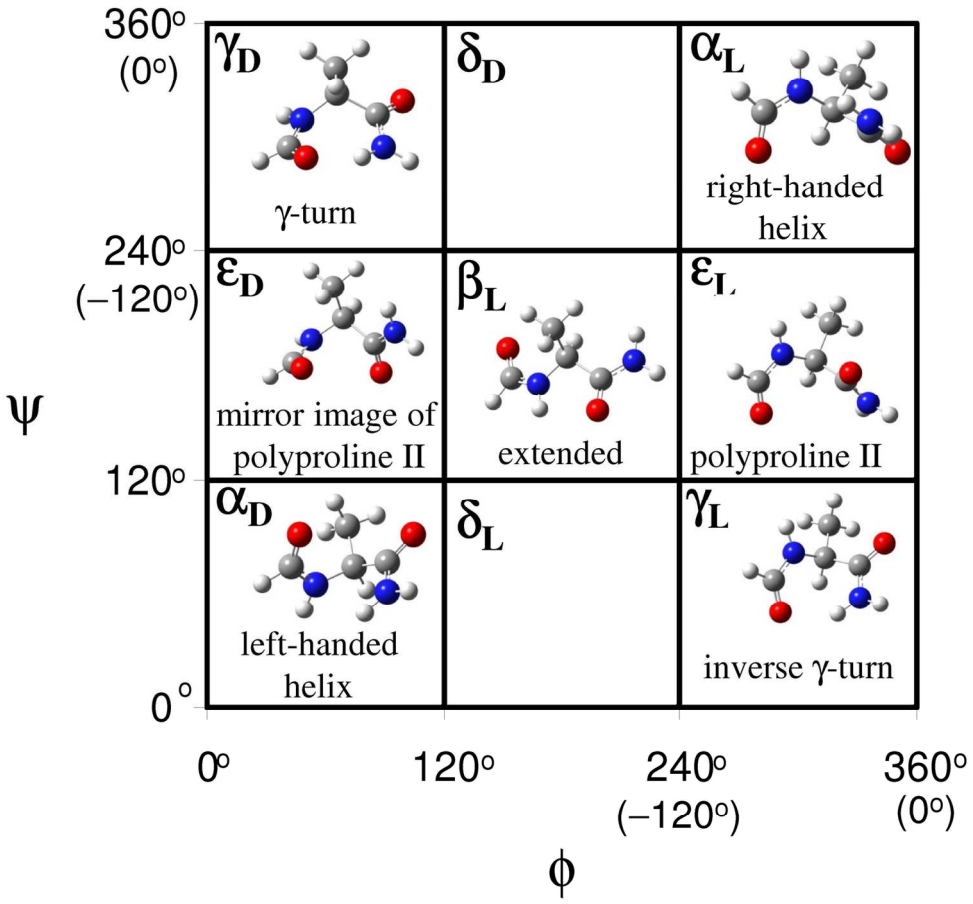


Figure 2. The Ramachandran map defined by torsional angles φ and ψ. The nine regions are named according to Perczel et al.[7] The seven minima of the HCO-Ala-NH_2 model optimized with PCM is presented to characterize the relevant conformational regions.
153x144mm (300 x 300 DPI)

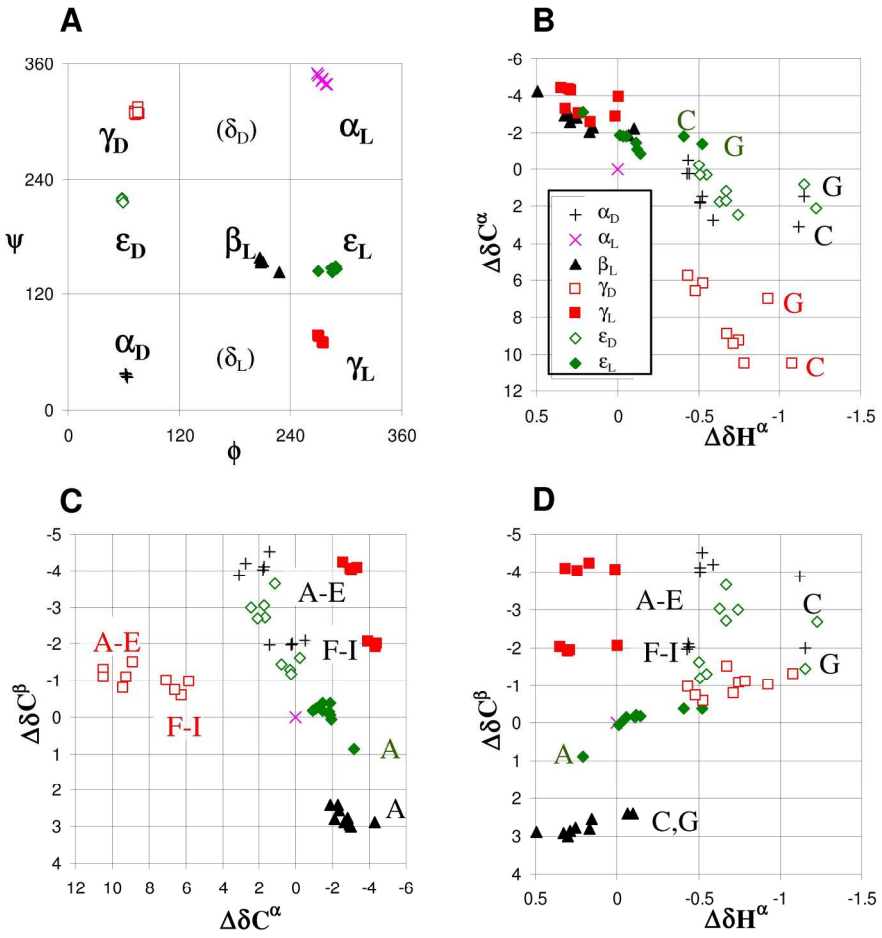


Figure 3. Geometry and relative chemical shielding of optimized alnine diamide conformers: α_D (black +), α_L (cyan x), β_L (black triangle), γ_D (red hollow square), γ_L (red square), ε_D (green hollow diamond), ε_L (green diamond). For the applied levels of theory see Table I. Several data points are signed by the letter of the applied method. A) Ramachandran plot, B) $\Delta\delta H^{\alpha}$-$\Delta\delta C^{\alpha}$ plot, C) $\Delta\delta C^{\alpha}$-$\Delta\delta C^{\beta}$ plot, D) $\Delta\delta H^{\alpha}$-$\Delta\delta C^{\beta}$ plot. The origin of every relative chemical shift scale is set at the right-handed helical α_L conformer (see Table S-I).
202x209mm (300 x 300 DPI)

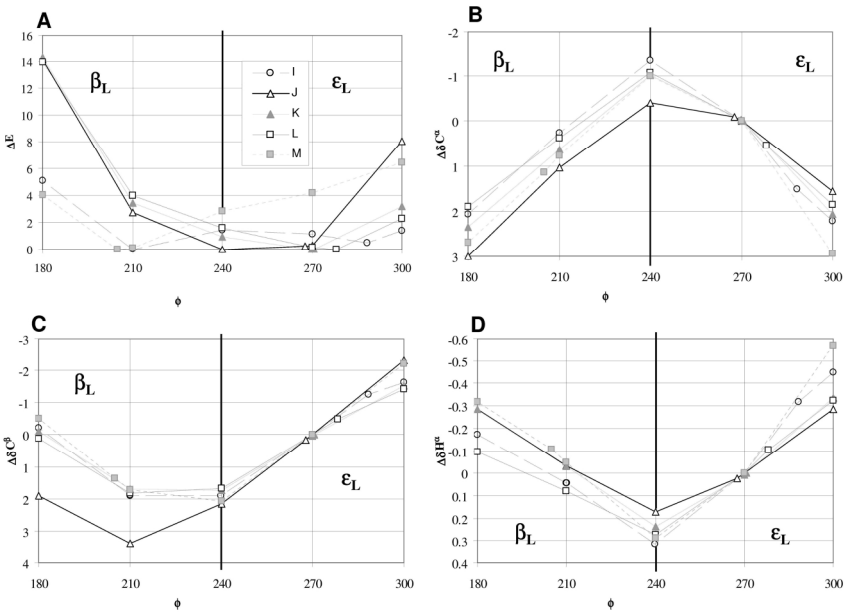


Figure 4. Relative energy and chemical shifts of alanine diamide conformations optimized along the β_L - ε_L path. Data points at φ=180°, 210°, 240°, 270° and 300° (i.e. at 180°, 150°, 120°, 90° and 60° refer to structures of constrained φ. All other structures are fully optimized β_L or ε_L conformers. A) Relative energy over the (lower) minimum of the path. B) $\Delta\delta C^{\alpha}$, C) $\Delta\delta C^{\beta}$, D) $\Delta\delta H^{\alpha}$. Relative chemical shifts are referenced to the data point at 270° (i.e. 90°).
209x148mm (300 x 300 DPI)

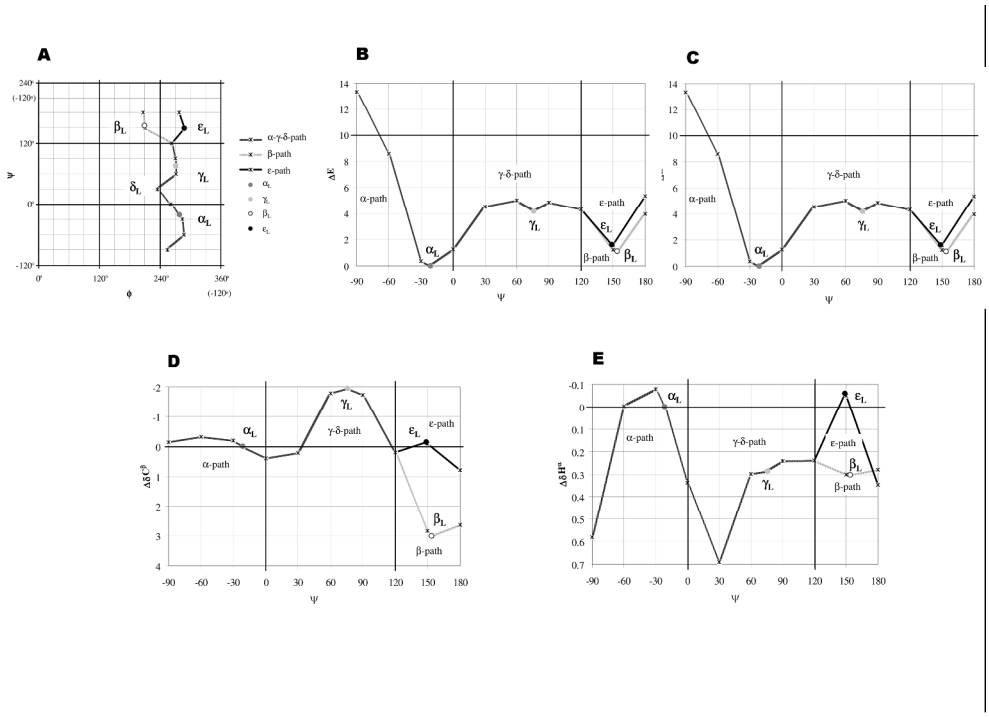
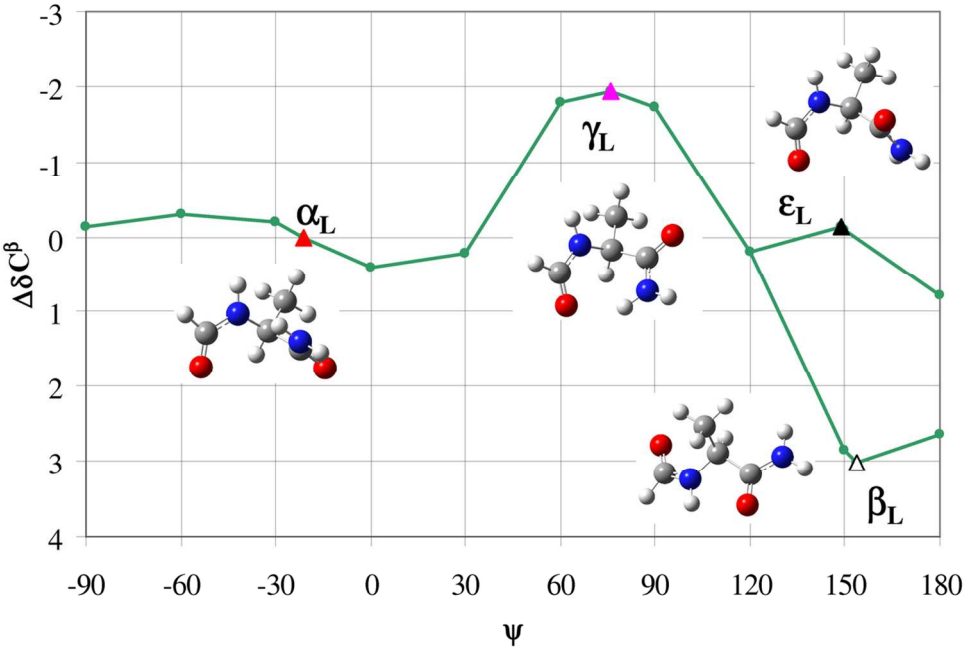


Figure 5. Geometry, energy and relative chemical shifts of alanine diamide conformations optimized along the α_L - γ_L- ε_L- β_L paths. Data points at ψ= -90°, -60°, -30°, 0°, 30°, 60°, 90°, 120°, 150° and 180° refer to structures of constrained ψ. All other structures are fully optimized α_L, γ_L, ε_L or β_L conformers. A) Ramachandran plot. B) Relative energy over that of the α_L conformer. C) $\Delta\delta C^{\alpha}$, D) $\Delta\delta C^{\beta}$, E) $\Delta\delta H^{\alpha}$. Relative chemical shifts are referenced to the right-handed helical α_L conformer.
297x209mm (300 x 300 DPI)



Graphical abstract
115x79mm (300 x 300 DPI)