

Use of a Modified Romberg Formalism for the Extrapolation of Molecular Properties from Oligomers to Polymers. Polyalanine Diamide in Its $(\gamma_L)_n$ or $(C_7)_n$ Conformation

Christoph-Maria Liegener,[†] Gabor Endredi,[†] Michael A. McAllister,[‡] András Perczel,^{‡,§} Janos Ladik,^{*†} and Imre G. Csizmadia^{†,‡,§}

Contribution from the Institute for Theoretical Chemistry and Laboratory of the National Foundation for Cancer Research, Friedrich Alexander University Erlangen-Nürnberg, Egerlandstrasse 3, D-8520 Erlangen, Germany, Department of Chemistry, University of Toronto, Toronto, Ontario, Canada M5S 1A1, and Department of Organic Chemistry, Eötvös University, Budapest, Hungary

Received December 23, 1992

Abstract: Computed torsional angles, HOMO/LUMO energies, total energy per unit, polymerization energy, dipole moment, and selected net charges are reported for the For-(Ala)₄-NH₂ tetrapeptide in its $(\gamma_L)_4$ or $(C_7^{qu})_4$ conformation. These molecular properties are compared with those of shorter members of the oligomer series, and an extrapolation of the computed numerical values for infinite backbone length was attempted using a modified Romberg algorithm.

Introduction

A protein, containing n amino acid residues, has N_0 legitimate minimum-energy conformations:¹⁻³

$$N_0 = 9^n = 10^{n \log 9} \quad (1)$$

The numerical comparison shown in Table I clearly indicates that for peptides and proteins N_0 is not only astronomically large but an order or magnitude increase in n will create 10 orders of magnitude more conformations. This appears to be an unprecedented chemical amplification typical to macromolecules.

Needless to say, it is impossible at this time even to contemplate that one might compute, using an empirical force field method, all the conformations of an intermediate-size peptide for example for $n = 10$. Even if molecular mechanics could optimize the geometry of these structures as fast as 1 s/conformation, for $n = 10$ the total computer time requirement for all the conformations would be roughly 110 calendar years. Furthermore, it is useless to plan today an ab initio SCF molecular computation for a single 3D conformational structure of a peptide with $n = 10$ even if a modest 3-21G basis set is used. In the case of this particular basis set, for polyalanine diamide (HCO-(NH-CHMe-CO) _{n} -NH₂) the current technical limitation may be in the vicinity of $n = 6$ or 7 but definitely $n < 10$.

For this reason, methods other than those currently in use might be required to gain some further insight into the physical properties of protein models. If periodically repeating units are involved, then a modified Romberg's formalism⁴ might be used

Table I. Numerical Illustration of the Increase of the Number of Legitimate Minimum-Energy Conformations (N_0) with the Increase of the Degree of Polymerization (n)

n	N_0	n	N_0
1	$9 = 10^{0.9542425}$	100	$9^{100} = 10^{95.42425}$
10	$9^{10} = 10^{9.542425}$	1000	$9^{1000} = 10^{954.2425}$

to extrapolate computed molecular properties (energy per unit, geometry, net charges, etc.) from oligomers, $n = 1, 2, 3,$ and 4, to the infinite polymer. The method has been successfully applied to a small chemical system previously.⁵

Methods

The modified Romberg algorithm^{4,6} starts from a given sequence $\{q_{00}, q_{10}, \dots, q_{n0}\}$ of $n + 1$ elements (e.g. the oligomer energies or other properties) where the first four elements are calculated directly. The elements q_{ik} of a triangular matrix ($1 \leq k \leq i \leq n$) are evaluated recursively by the expression

$$q_{i,k}^{(\alpha)} = \frac{h_{i-k}^{(\alpha)} q_{i,k-1} - h_i^{(\alpha)} q_{i-1,k-1}^{(\alpha)}}{h_{i-k}^{(\alpha)} - h_i^{(\alpha)}} \quad (2)$$

where $h_i^{(\alpha)} = (i + 1)^{-\alpha}$. This is done for any of a set of α values (normally a grid between 0 and 10). The array $q_{ik}^{(\alpha)}$ for which $|q_{n,n-1}^{(\alpha)} - q_{n-1,n-1}^{(\alpha)}|$ becomes minimal is taken as the optimal one, and q_{nn} is then the extrapolated values for the above sequence.

The other possibility would be to use the Hartree-Fock (HF) crystal orbital formalism applied to solids,⁶ but in this case the states belonging to the chain ends have to be computed with the aid of a Green's matrix formalism.^{7,8}

The molecular computations on the peptide oligomers were carried out on a CRAY X-MP/28 (Toronto) and on a CRAY Y-MP4/464 (München) supercomputer using a GAUSSIAN 90 program at a HF/3-21G level of theory with full geometry optimization. The extrapolations have been performed on a CDC Cyber 995 E. Computer at the University of Erlangen-Nürnberg.

(5) Liegener, C.-M.; Beleznyay, F.; Ladik, J. *Phys. Lett. A* 1987, 123, 399.

(6) Del Re, G.; Ladik, J.; Biczo, G. *Phys. Rev.* 1967, 155, 997. André, J.-M.; Gouverneur, L.; Leroy, G. *Int. J. Quantum Chem.* 1967, 1, 427, 451.

(7) Ladik, J.; Seel, M. *Phys. Rev.* 1976, B13, 5388.

(8) Del Re, G.; Ladik, J. *Chem. Phys.* 1980, 49, 32.

[†] Friedrich Alexander University Erlangen-Nürnberg.

[‡] University of Toronto.

[§] Eötvös University.

(1) Perczel, A.; Angyán, J. G.; Kajtár, M.; Viviani, W.; Rivail, J.-L.; Marcocchia, J.-F.; Csizmadia, I. G. *J. Am. Chem. Soc.* 1991, 113, 6256.

(2) Perczel, A.; Kajtár, M.; Marcocchia, J.-F.; Csizmadia, I. G. *THEOCHEM* 1991, 232, 291.

(3) Perczel, A.; Viviani, W.; Csizmadia, I. G. Peptide Conformational Energy Surfaces and their Relevance to Protein Folding. In *Molecular Aspects of Biotechnology: Computational Models and Theories*; Bertran, J., Ed.; Kluwer Academic Publishers: Hingham, MA, 1992; pp 39-82.

(4) (a) Romberg, W. *Det. K. Nor. Vidensk. Selsk. Forh.* 1955, 28, 7. (b) Bauer, F. L.; Rutishauser, H.; Stiefel, E. *Porc. Symp. Appl. Math.* 1963, 15, 199. (c) Stoer, J.; Bulirsch, R. *Introduction to Numerical Analysis*; Springer: Berlin, 1980.

Table II. Computed and Extrapolated Molecular Characteristics for the $(\gamma_L)_n$ Conformation of $\text{HCO}-(\text{NH}-\text{CHMe}-\text{CO})_n-\text{NH}_2$

n	1	2	3	4	∞
max force (a.u.)	3.0×10^{-6}	1.0×10^{-5}	2.3×10^{-5}	1.62×10^{-4}	
RMS force (a.u.)	1.0×10^{-6}	3.0×10^{-6}	4.0×10^{-6}	4.3×10^{-5}	
ϕ_{n-3}				-83.64270	-83.50350 ^a
ψ_{n-3}				+66.80091	+66.74948
ϕ_{n-2}			-83.78156	-84.3196	
ψ_{n-2}			+66.85134	+65.49751	
ϕ_{n-1}		-84.21039	-84.63500	-84.85120	
ψ_{n-1}		+67.01872	+65.68652	+65.52944	
ϕ_n	-84.47953	-84.96256	-85.22487	-85.15418	-85.41777 ^b
ψ_n	+67.25868	+66.39361	+66.25060	+65.96468	+66.19617 ^c
ϵ_{LUMO}	+0.17743	+0.16402	+0.15689	+0.15235	+0.12147
ϵ_{HOMO}	-0.38194	-0.36970	-0.36320	-0.35939	-0.34427
E_n	-412.474779964	-656.963680615	-901.453597680	-1145.943652420	
$\Delta E_n = E_n - E_{n-1}$		-244.488900651	-244.489917065	-244.490054740	-244.490132361
ΔE_n^{ID} (kcal/mol)		+0.614	-0.024	-0.110	-0.148
$\mu(\text{D})$	3.2648	6.6960	8.1530	11.3012	
$\Delta\mu(\text{D})$		2.4312	1.4570	3.1482	not monotonic
$g(\text{N}_1)$	-0.963399	-0.917228	-0.918229	-0.918298	not monotonic
$g(\text{H}_1)$	+0.394722	+0.411550	+0.414234	+0.415988	+0.418527 ^c
$g(\text{O}_1)$	-0.634862	-0.629329	-0.630471	-0.630197	not monotonic
$g(\text{N}_n)$	-0.963399	-0.966053	-0.966295	-0.966793	-0.966720 ^c
$g(\text{H}_n)$	+0.394722	+0.394334	+0.395772	+0.395966	not monotonic
$g(\text{O}_n)$	-0.634862	-0.676313	-0.677337	-0.680132	-0.680961 ^c

^a For the series of four ϕ_1 , the extrapolation yielded an exponent of $\alpha = 2.5$. ^b The original data did not form a monotonic set of values. The extrapolation of $\{\phi\}_n$ to ϕ_∞ was carried out with the use of $\alpha = 2.5$ (the exponent of footnote a). ^c Four-element extrapolation did not yield an optimized exponent α . Exponents used in the four-element extrapolation were in the different rows $\alpha = 2.7, 2.2, 3.2$, and 5 (optimum for three elements).

Results and Discussion

Several papers^{1,9,10} reported the geometry-optimized structure and energy for the γ_L , or C_7^{equ} , conformation of $\text{HCO}-\text{NH}-\text{CHMe}-\text{CO}-\text{NH}_2$. High-power computations^{10c} clearly indicated that torsional angle pairs do change with basis set size and with the inclusion of electron correlation. Nevertheless, we have reason to believe that the method used will lead to a reliability that may be classified as semiquantitative. The $\gamma_L\gamma_L$ conformation of $\text{HCO}-(\text{NH}-\text{CHMe}-\text{CO})_2-\text{NH}_2$ has been reported earlier,¹¹⁻¹³ and the trimer $\text{HCO}-(\text{NH}-\text{CHMe}-\text{CO})_3-\text{NH}_2$ in its $\gamma_L\gamma_L\gamma_L$ conformation has also been studied.¹⁴ We now wish to report the fully optimized $\gamma_L\gamma_L\gamma_L\gamma_L$ conformation of the tetramer $\text{HCO}-(\text{NH}-\text{CHMe}-\text{CO})_4-\text{NH}_2$ computed at the HF/3-21G level of theory together with the results of the modified Romberg extrapolation.

Figure 1 illustrates schematically the $(\gamma_L)_n$, or $(C_7^{\text{equ}})_n$, structures of the alanine oligomer diamides $\text{HCO}-(\text{NH}-\text{CHMe}-\text{CO})_n-\text{NH}_2$ for $n = 1, 2, 3$, and 4. Table II summarizes the computed and extrapolated results, and Figure 2 illustrates graphically the convergence of selected molecular properties.

Inspection of the data clearly shows that energy, which is a global property, always changes in a smooth monotonic fashion with n . Therefore it is easy to carry out the extrapolation. However, some other local properties may not vary smoothly. If the data are monotonic but the variation is not smooth, then a three-point fit with a fixed exponent, α , may be attempted (cf. footnote of Table II). If, however, the points are not even monotonic, no fit is possible.

The HOMO-LUMO gap, $\Delta\epsilon$, is decreasing with the increase of the degree of polymerization, and it converges to 292.25 kcal/mol (12.67 eV). This is a typical value, at the HF level, for an

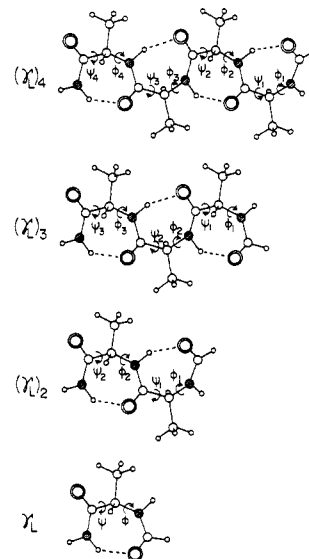
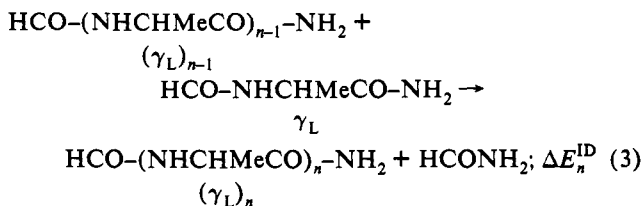


Figure 1. Schematic illustration of the $(\gamma_L)_n$ oligomer conformations of $\text{HCO}-(\text{NH}-\text{CHMe}-\text{CO})_n-\text{NH}_2$ for $n = 1, 2, 3$, and 4.

insulator. Figure 2a clearly indicates that the HOMO converges faster than the LUMO.

The energy of polymerization is measured by the change in energy associated with the following isodesmic reaction



This energy value, ΔE_n^{ID} is changing from a slightly endothermic to a slightly exothermic value, as shown by Figure 2b.

The Mullikan net charges, the last six lines in Table II, sometimes vary monotonically, sometimes not, but they do not change dramatically with n . Also there is no significant charge separation between the N-terminus and C-terminus at $n = 4$. Yet

(9) McAllister, M. A.; Császár, P.; Viviani, W.; Rivail, J.-L.; Csizmadia, I. G. *THEOCHEM*, in press.

(10) (a) Head-Gordon, T.; Head-Gordon, M.; Frish, M. J.; Brooks, C. L., III; Pople, J. C. *J. Am. Chem. Soc.* **1991**, *113*, 5989. (b) Böhm, H.-J.; Brode, S. *J. Am. Chem. Soc.* **1991**, *113*, 7129. (c) Frey, R. F.; Coffin, J.; Newton, S. Q.; Ramek, M.; Cheng, V. K. W.; Momany, F. A.; Schäfer, L. *J. Am. Chem. Soc.* **1992**, *114*, 5369.

(11) McAllister, M. A.; Perczel, A.; Császár, P.; Csizmadia, I. G. *THEOCHEM*, in press.

(12) Perczel, A.; McAllister, M. A.; Császár, P.; Csizmadia, I. G. *J. Am. Chem. Soc.*, in press.

(13) Perczel, A.; McAllister, M. A.; Császár, P.; Csizmadia, I. G. *J. Am. Chem. Soc.*, in press.

(14) Cheung, M.; McGovern, M. E.; Danchuan, T. J.; Zhao, C.; McAllister, M. A.; Perczel, A.; Császár, P.; Csizmadia, I. G. *THEOCHEM*, in preparation.

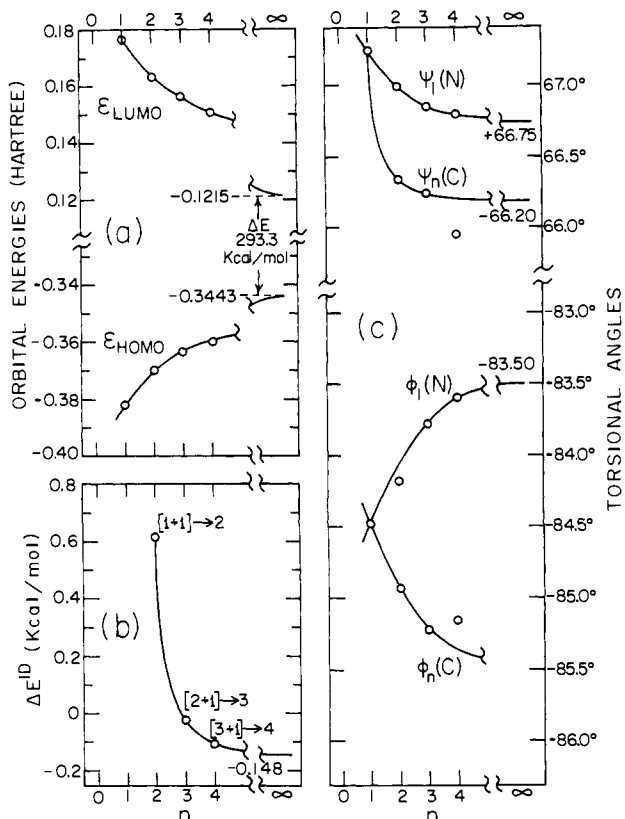


Figure 2. Graphical illustration of the convergence of selected molecular properties of $\text{HCO}-(\text{NH}-\text{CHMe0CO})_n-\text{NH}_2$ in their $(\gamma_L)_n$ conformations.

the increase of dipole moment is very pronounced with increasing n . Clearly, the length of the molecule is responsible for this, as no significant charge separation was found between the two ends of the chain. Of course one can always envisage this dramatic increase in μ as a result of the vectorial sum of the dipoles associated with the individual peptide bonds.

Perhaps the most interesting data are the variation of the (ϕ, ψ) angle pairs at the amino end, $\phi_1(\text{N})$ and $\psi_1(\text{N})$, and at the carboxyl end, $\phi_n(\text{C})$ and $\psi_n(\text{C})$. It is clear from Figure 2c that $\psi_1(\text{N})$ and $\psi_n(\text{C})$ change in the same direction; the variation of $\psi_n(\text{C})$ with n is, however, larger than that of $\psi_1(\text{N})$. The two extrapolated values differ from each other by only 0.5° . In contrast to that, $\phi_1(\text{N})$ and $\phi_n(\text{C})$ change in opposite directions with n . In the limit, $\phi_1(\text{N})$ reaches a value of -83.5° while $\phi_n(\text{C})$ converges toward -85.5° , creating a difference of $\sim 2^\circ$ between the two ϕ values. This of course is only an approximate value because $\phi_n(\text{C})$ could not be extrapolated, as the points did not form a monotonic set.

At this stage we cannot offer a definite explanation of the observation that some of the points do not fall on the curves. In the case of $n = 2$, $\phi_1(\text{N})$ deviates from the curve by 0.2° . Also the residual forces on $\phi_n(\text{C})$ and $\psi_n(\text{C})$ in the $(\gamma_L)_4$ structure are smaller than the usual threshold values (max force, 4.5×10^{-4}

Table III. Numerical Test for the Reliability of $\{\phi_n(\text{C}), \psi_n(\text{C})\}$ for the Fourth Angle Pair in the $(\gamma_L)_4$ Conformation of $\text{For}-(\text{Ala})_4-\text{NH}_2$

	original (all variables optimized)	constrained (fixed ^a $\{\phi_n(\text{C}), \psi_n(\text{C})\}$ at point 4, all other variables optimized)	final (all variables optimized)
$\phi_n(\text{C})$ at point 4	-85.154°	-85.322°	-85.140
$\psi_n(\text{C})$ at point 4	$+65.965^\circ$	$+66.222^\circ$	$+65.996$
max force (a.u.)	1.62×10^{-4}	2.67×10^{-4}	1.15×10^{-4}
RMS force (a.u.)	4.3×10^{-5}	6.4×10^{-5}	2.5×10^{-5}

^a Using $\alpha = 2.5$ in the Romberg expression (eq 2) for $\phi_n(\text{C})$ and $\psi_n(\text{C})$ at point 4 of $(\gamma_L)_4$.

a.u.; RMS force, 3.0×10^{-4} a.u.). Thus, on one hand, we may associate some chemical significance to this finding, since we have reason to believe that the deviations from monotony are real and therefore they must be diagnostic to some local interactions.

On the other hand, the fact that the variations of the ϕ and ψ torsional angles are monotonic may only have mathematical significance. Even at this highly fixed backbone conformation type [at the $(\gamma_L)_n$ location], the shape of the potential energy hypersurface could be shallow.

To be certain, we have carried out the following numerical test. We have determined in the cases of ϕ_n and ψ_n , respectively, the α value in the Romberg formalism from the first three points. In both cases $\alpha = 2.5$ has been obtained; using this value for the quantities $h_i^{(\alpha)}$, with the help of eq 2 we have calculated again $\phi_n(\text{C})$ and $\psi_n(\text{C})$ for the fourth point of $(\gamma_L)_4$. Keeping the obtained pair of values fixed, we have performed for $(\gamma_L)_4$ a new geometry optimization with the help of the gradient method for all the other geometrical variables. After convergence we have performed with their obtained values a further geometry optimization for all the variables including also the $(\phi_n(\text{C}), \psi_n(\text{C}))$ pair of angles at point four for $(\gamma_L)_4$ (see Table III).

As can be seen, the final pair of angles hardly differs from the ones obtained in the original optimization. This proves that the fourth points and the $\phi_n(\text{C})$ and $\psi_n(\text{C})$ curves of Figure 2c are not artifacts of the method but they really deviate from the smooth Figure 2c curves. This probably caused some collective effects.

In contrast, the change of the $\Delta E_n = E_n - E_{n-1}$ (Figure 2b) and $\Delta \epsilon = \epsilon_{\text{HOMO}} - \epsilon_{\text{LUMO}}$ (Figure 2a) of the monomer shows a monotonically converging tendency.

The monotonic decrease of the ΔE , measuring the energy of polymerization, as a function of n (see Figure 2b) also shows unambiguously that, after a specific value of n , the increase of the chain length (i.e. the addition of a residue to the $(i+n)$ th peptide) will have no influence on the residue(s) located toward the N-terminus of the polypeptide chain at position(s) $< i$. In the contrary the "last" residue perturbs the last n amino acid residues containing peptide fragment. In other words the effect of the long-range interaction through backbone atoms is limited. The limiting value may be somewhere in the vicinity of $4 \leq n \leq 6$.

Acknowledgment. I.G.C. would like to express his gratitude to the Deutsche Forschungsgemeinschaft for a Visiting Professorship which made it possible to do this work in Erlangen.