

Critical Role of Anisotropy for the Dimerization Energies of Two Protein–Protein Recognition Motifs: *cis*-*N*-Methylacetamide versus a β -Sheet Conformer of Alanine Dipeptide. A Joint *ab Initio*, Density Functional Theory, and Molecular Mechanics Investigation

Nohad Gresh,^{*,†} Hong Guo,^{‡,§,⊥} Dennis R. Salahub,^{‡,§} Bernard P. Roques,^{*,†} and Sherif A. Kafafi^{||}

Contribution from the Département de Pharmacochimie Moléculaire et Structurale, INSERM U266, CNRS UMR 8600, UFR des Sciences Pharmaceutiques & Biologiques, 4, Avenue de l'Observatoire, 75270 Paris Cedex 06, France, Centre de Recherche en Calcul Appliqué, Bureau 400, 5160, Boulevard Décarie, Montréal, QUEBEC H3X 2H9 Canada, Département de Chimie, Université de Montréal, C.P. 6128, succursale A Montréal, Québec, Canada, and Center for Advanced Research in Biotechnology, Maryland Biotechnology Institute, 9600 Gudelsky Drive, Rockville, Maryland 20851

Received December 16, 1997. Revised Manuscript Received June 18, 1999

Abstract: The dimerization energies of formamide and *cis*-*N*-methylacetamide (*cis*-NMA) are compared with those of the Ala and Gly dipeptides in their canonical β -sheet conformations using *ab initio* (SCF and MP2), density functional theory (DFT), and the SIBFA molecular mechanics procedure. Consistent with the gas-phase *ab initio* and DFT results, the SIBFA procedure is able to account for the larger dimerization energies of formamide and *cis*-NMA than of the dipeptides. In contrast, the majority of “conventional” force fields produced an inversion of the relative dimerization energies, giving rise to a more favorable dimerization of the Ala dipeptide than of *cis*-NMA (Beachy, M. D.; et al. *J. Am. Chem. Soc.* **1997**, *119*, 5908). Energy decomposition analysis on the dimers of formamide and the Gly dipeptide shows the Coulombic energy contribution to be the most important term favoring the formamide dimer. The analysis based on the SIBFA procedure similarly shows the multipolar energy term (E_{MTP}) to be the most important contributor to this difference. This is due to its monopole–dipole and monopole–quadrupole components. The issue of the transferability of the multipolar expansion is discussed in the context of simulations of oligopeptides.

Introduction

Development of molecular mechanics models at different levels of sophistication is of considerable interest. The simplicity of the conventional potential energy functions allows us to study the structure, dynamics, and thermodynamics of macromolecules containing tens of thousands of atoms for time scales of nanoseconds or longer.¹ But in some cases, the accuracy of the potential energy functions may be in question, presumably due to large inhomogeneities of the condensed phase and other effects that have not been taken into account in the development of such potential functions.² Although the improvement of the performances of conventional potential energy functions remains

an important objective, the parallel development of more elaborate molecular mechanics procedures^{3,4} could be of crucial importance when more quantitative results are required. One of the models, which goes beyond the use of conventional, atom-centered partial charges for electrostatics and was developed for biomolecular modeling, is the sum of interactions between fragments *ab initio* (SIBFA)-computed procedure.³ It incorporates a distributed multipole description of electrostatics and polarization and a representation of lone-pair directionality, which, in the *ab initio* computations, is due to the overlap-dependent components of the binding energy, short-range exchange, and charge transfer. The intermolecular interaction energy is computed as a sum of five separate contributions: electrostatic and short-range repulsion in first order, and polarization, charge transfer, and dispersion in second order.

* To whom correspondence should be addressed. Tel.: (33) 1-53.73.95.74. Fax: (33) 1-43.26.69.18. E-mail: gresh@bisance.citi2.fr.

† UFR des Sciences Pharmaceutiques & Biologiques.

‡ Centre de Recherche en Calcul Appliqué.

§ Université de Montréal.

⊥ Present address: Departments of Chemistry and Chemical Biology, Harvard University, 12 Oxford St., Cambridge, MA 02139.

|| Maryland Biotechnology Institute.

(1) (a) Mackerell, A. D., Jr.; Bashford, D.; Bellot, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Ha, H.; Gao, J.; Guo, H.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Nguyen, D. T.; Ngo, T.; Prodhom, B.; Reither, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiokiewicz-Kuczera, Yin, J. D.; Karplus, M. *J. Phys. Chem.* **1998**, *B102*, 3586. (b) Duan, Y.; Wang, L.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9897.

(2) Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 5908.

(3) (a) Gresh, N.; Claverie, P.; Pullman, A. *Int. J. Quantum Chem.* **1986**, *29*, 101–118. (b) Gresh, N.; Leboeuf, M.; Salahub, D. R. In *Modeling the Hydrogen Bond*; Smith, D. A., Ed.; American Chemical Society Symposium Series 569; ACS: Washington, DC, 1994; p 82. (c) Gresh, N. *J. Comput. Chem.* **1995**, *16*, 856. (d) Gresh, N.; Garmer, D. R. *J. Comput. Chem.* **1996**, *17*, 1481. (e) Gresh, N. *J. Chim.-Phys.* **1997**, *94*, 1365.

(4) See, e.g.: (a) Dinur, U. *J. Phys. Chem.* **1990**, *94*, 5669. (b) Millot, C.; Stone, A. J. *Mol. Phys.* **1992**, *77*, 439. (c) Mitchell, J. B. O.; Nandi, C. L.; Thornton, J. M.; Price, S. L.; Singh, J.; Snarey, M. *J. Chem. Soc., Faraday Trans.* **1993**, *89*, 2619. (d) Astrand, P. O.; Wallqvist, A.; Karlstrom, G. *J. Chem. Phys.* **1994**, *100*, 1262. (e) Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, *101*, 6141. (f) Dudek, M. J.; Ponder, J. W. *J. Comput. Chem.* **1995**, *16*, 791. (g) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Basch, H.; Cohen, D. *J. Chem. Phys.* **1996**, *105*, 1968.

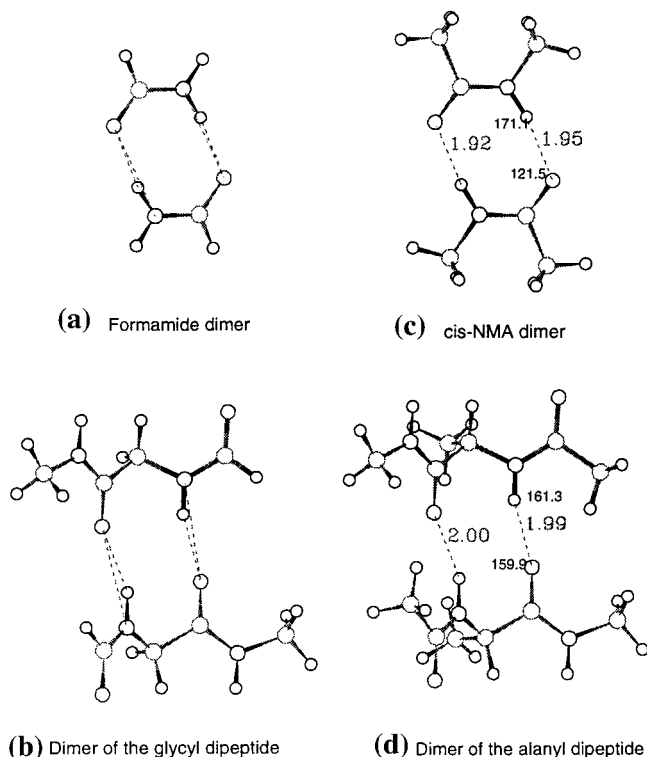


Figure 1. Representation of the dimers of (a) formamide, (b) glycyl dipeptide, (c) *N*-methylacetamide, and (d) alanyl dipeptide.

This procedure was recently refined^{3b,c} and shown to reproduce reliably the results of high-quality ab initio supermolecule computations on polycoordinated complexes of divalent cations,^{3c,d,5} multiply hydrogen-bonded complexes,⁶ and cation- π interactions (Gresh, N., manuscript in preparation).

This paper is the first in a series of papers evaluating the performance of SIBFA for determining the gas-phase and condensed-phase properties of amides and peptides. We report the results of the SIBFA calculations on the gas-phase dimerization energies of formamide and *cis*-*N*-methylacetamide (*cis*-NMA) along with the dimerization of the Ala and Gly dipeptides in their canonical β -sheet conformation. Such model dimers are important motifs of protein-protein recognition.⁷ We compare these energies to those obtained from ab initio and DFT calculations and analyze the different components of the interactions. Recently, Beachy et al. have studied the dimerization of *cis*-NMA and the alanine dipeptide (AD) (see Figure 1).² They found that the ab initio computations resulted in a larger dimerization energy for *cis*-NMA than for AD. In contrast, all the conventional force fields tested, except one, gave rise to a stronger dimerization energy for AD than for *cis*-NMA, by amounts of up to 3.8 kcal/mol. Understanding the origin of this discrepancy, which Beachy et al. ascribe to fundamental limitations of force field models that rely on single atom-centered partial atomic charges, is of clear importance.

Furthermore, it was pointed out by Mackerell et al.¹ and Beachy et al.² that the results from the gas-phase ab initio calculations and conventional force fields may not be comparable, since the latter were generally developed in relation to condensed-phase properties. Thus, in this phase, the existence

of ligands bound to the free C=O and N-H groups of the dimer of AD, such as solvent molecules or protein groups, can induce a resonance-assisted cooperative effect,⁸ increasing the dimerization energy for AD in the condensed phase as compared to that in the gas phase. For the *cis*-NMA dimer, on the other hand, such a resonance-assisted stabilization is built-in in the electronic structure of the molecule, since the N-H and C=O groups are in spatial proximity due to the *cis* conformation. But there will be no additional C=O and N-H groups to enhance this effect upon passing from the gas phase to the condensed phase.

The caveats of comparing gas-phase ab initio and conventional force field results should vanish if one uses polarizable molecular mechanics functions, which should be able to reproduce *both* gas-phase and condensed-phase intermolecular interactions.² However, such comparisons have been virtually limited to studies of bulk liquids⁹ and water clusters.^{6,10} This constitutes an additional incentive to test the SIBFA procedure and establish the importance and relative weights of first- and second-order effects. Here, we accomplish this by comparing the results to those from ab initio (SCF/MP2) and density functional calculations. The analysis of the separate components of the interaction energies should offer an explanation for the shortcomings of the atom-centered point charge models in accounting for gas-phase dimerization energies.

In addition to discussing the need for anisotropy and separability features in a force field, we also address another important issue: the transferability of the multipolar expansion.¹¹ Specifically, we assess the extent to which the multipolar expansion derived for the peptide elementary fragments can be used for the energy computations of relatively large-sized Ala oligopeptides in specific conformations. This evaluation is made by comparing the ab initio/DFT results to the SIBFA results that utilize different representations of the peptide (see below).

The Results and Discussion section is divided into two parts. The first deals with the dimerization of formamide and of the glycyl dipeptide (GD) in its canonical β -sheet conformation. Formamide and GD are the unmethylated analogues of *cis*-NMA and AD, respectively, and the relatively small sizes of these compounds allow us to perform single-point high-level ab initio supermolecule SCF/MP2 computations with energy decomposition. In the second part, we compare the SIBFA dimerization energies of *cis*-NMA and AD to the results of ab initio and DFT computations as well as to those reported by Beachy et al.² In a forthcoming paper, we will evaluate the ability of SIBFA to describe condensed-phase cooperative effects in peptides.

(8) (a) Guo, H.; Sirois, S.; Proynov, E. I.; Salahub, D. R. In *Theoretical Treatments of Hydrogen Bonding*; Hadzi, D., Ed.; 1997. (b) Guo, H.; Karplus, M. *J. Phys. Chem.* **1994**, *98*, 7104. (c) Guo, H.; Salahub, D. R. *Angew. Chem., Int. Ed.* **1998**, *37*, 2985. (d) St. Amant, A.; Cornell, W.; Kollman, P. A.; Halgren, T. A. *J. Comput. Chem.* **1995**, *16*, 1483.

(9) (a) Stillinger, F. H.; David, C. W. *J. Chem. Phys.* **1978**, *69*, 1473. (b) Belford, D.; Campbell, E. *J. Chem. Phys.* **1987**, *86*, 7013. (c) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269. (d) Sprik, M.; Klein, M. L. *J. Chem. Phys.* **1988**, *89*, 7556. (e) Ahlstrom, P.; Wallqvist, A.; Engstrom, S.; Jonsson, B. *Mol. Phys.* **1989**, *68*, 563. (f) Dykstra, C. E. *J. Chem. Phys.* **1989**, *91*, 6472. (g) Niesar, U.; Corongiu, G.; Huang, M.-J.; Dupuis, M.; Clementi, E. *Int. J. Quantum Chem.* **1989**, *23*, 421. (h) Cieplak, P.; Lybrand, T. A.; Kollman, P. A. *J. Chem. Phys.* **1990**, *92*, 6755. (i) Wallqvist, A.; Ahlstrom, P.; Karlstrom, G. *J. Phys. Chem.* **1990**, *94*, 1649. (j) Guillot, B. *J. Chem. Phys.* **1991**, *95*, 1543. (k) Sprik, M. *J. Chem. Phys.* **1991**, *95*, 6762. (l) Halley, J. W.; Rustad, J. R.; Rahman, A. *J. Chem. Phys.* **1993**, *98*, 4110. (m) Bernardo, D. N.; Ding, Y.; Krogh-Jespersen, K.; Levy, R. M. *J. Phys. Chem.* **1994**, *98*, 4180.

(10) (a) Masella, M.; Flament, J.-P. *J. Chem. Phys.* **1997**, *107*, 9105. (b) Hodges, M. P.; Stone, A. J.; Xantheas, S. X. *J. Phys. Chem. A* **1997**, *101*, 9163. (c) Merrill, G.; Gordon, M. S. *J. Phys. Chem. A* **1998**, *102*, 2650.

(11) Faerman, C. H.; Price, S. L. *J. Am. Chem. Soc.* **1990**, *112*, 4915.

(5) Garmer, D. R.; Gresh, N.; Roques, B. P. *Proteins: Struct., Funct., Genet.* **1998**, *31*, 42.

(6) (a) Gresh, N. *J. Phys. Chem. A* **1997**, *A101*, 8680. (b) Masella, M.; Gresh, N.; Flament, J.-P. *J. Chem. Soc., Faraday Trans.* **1998**, *94*, 2745.

(7) *Protein Engineering*; Oxender, D. L., Fox, C. F., Eds.; A. R. Liss, Inc.: New York, 1988.

Procedure

Ab Initio Computations. The ab initio computations are performed with the Gamess package.¹² The basis set used is the coreless effective potential (CEP/4-31G+(2d)) basis set derived by Stevens et al.,¹³ supplemented on the heavy atoms by two uncontracted 3d orbitals, the exponents of which were given in ref 14. This basis set was previously used by us in several studies of cation–ligand^{3c,d,5} and hydrogen-bonded complexes.^{3b,6} The energy decompositions are done using the Kitaura–Morokuma procedure,¹⁵ and for the formamide dimer, the reduced variational space analysis (RVS) procedure developed by Stevens and Fink¹⁶ was used as well. These analyses deconvolute the total ab initio SCF interaction energy, $\Delta E(\text{SCF})$, into its individual components, E_1 , E_{rep} , and E_{ct} . The basis set superposition error is computed using the Boys procedure.¹⁷ For the formamide dimer, the BSSE correction with the virtual space¹⁸ is included as well, as it is computed during the RVS process implemented in the Gamess package.¹² The RVS analysis could not be applied for the GD and AD dimers because the amount of disk space needed to store the integrals was too large on the available computers. The correlation energy is computed using the Moller–Plesset 2 (MP2) procedure.¹⁹

Density Functional Theory Computations. The DFT²⁰ computations are performed with the deMon-KS program²¹ using the generalized gradient approximation (GGA). The functionals of Becke²² and Perdew–Wang²³ are used for the exchange, and the Perdew functional²⁴ is used for correlation. The resulting schemes are designated as Becke–Perdew (BP) and Perdew–Perdew (PP), respectively. The basis sets used are DZVP and 6-311++G**. The DZVP basis set used here is a Gaussian basis set that is comparable to 6-31G**, but optimized for DFT calculations.^{21b} Previous tests on a variety of hydrogen-bonded systems showed that this basis set produces hydrogen bond energies and geometries that are close to those obtained with much more extensive basis sets with a considerable saving in computer time.^{8a} For instance, for the HF–HF dimer, the hydrogen bond energy obtained from the BP/DZVP calculations with BSSE correction is 4.4 kcal/mol, compared to 4.1 kcal/mol from MP2/aug-cc-pVTZ and 4.2 kcal/mol from B3LYP/aug-cc-pVTZ. Furthermore, for neutral complexes containing a peptide linkage, namely systems closely related to those in the present study, the BP/DZVP hydrogen bond energies (without BSSE correction) were found to be only 0.2–0.6 kcal/mol higher than those from the BP/6-311++G** calculations, and the hydrogen-bond distances were equal to within 0.01 Å. In this study, 6-311++G** is only applied to the dimerizations of formamide and GD. The DFT calculations (BP and PP) were performed at the SIBFA-optimized geometries for all the systems studied. In addition, for the dimers of *cis*-NMA and AD, we performed DFT geometry optimizations at the BP/DZVP level of the six intermolecular variables that define the approach of the second monomer with respect to the first.

These DFT results are consistent with previous validation studies of the generalized gradient approximation (GGA) functionals (e.g., ref 8a) that indicate typical errors for hydrogen bond energies of around 0.5 kcal/mol. While some cases involving larger GGA errors have been identified (for example, the energy difference between the two lowest conformers of glycine,^{8a} where the error is about 1.5 kcal/mol), efforts are ongoing to develop more accurate functionals. The comparisons between DFT and MP2 results which were carried out in ref 3b as well as in ref 8a on several model hydrogen-bonded complexes indicate that the present GGA results should be of roughly equivalent quality to those obtained with MP2, and we view this as adequate for the present purposes.

As far as the analysis of energy components of the DFT calculations is concerned, it should be remembered that the concepts of Kohn–Sham DFT are different from those of the more traditional Hartree–Fock plus configuration mixing approaches, sometimes in ways that are more subtle than commonly assumed. All effects of exchange and correlation, including the difference in kinetic energy between the Kohn–Sham reference system of noninteracting electrons and the real system, are included in the exchange correlation functional. In particular, the question of “dispersion in DFT” arises often.^{8d,25} Obviously, the exact (unknown) functional would yield exact energy surfaces, including the asymptotic regions. The widely used functionals (GGA, B3LYP, etc.) have not been designed with dispersion in mind; however, they can provide reasonable values of the C_6/R^6 , C_8/R^8 , etc. terms when used in a time-dependent response formalism.²⁶

For traditional configurational mixing methods, difficulties, and the possibility for confusion of terminology, arise when one enters the region where overlap becomes non-negligible. For example, the analysis of excitations in an MP2 calculation that give rise to the “instantaneous dipole–induced dipole” language becomes imprecise when the two systems are not disjoint, since the excitation of a single electron on each molecule is not taking place, because of the overlap. From the Kohn–Sham DFT perspective, this small overlap region also presents a distinct challenge. The focus is on the exchange correlation functional rather than on the wave function. To understand this region, it might be helpful to think of correlation on various length scales for electron–electron interactions. The overall great success of the widely used functionals (GGA, B3LYP, etc.) indicates that correlation is reasonably handled on the scale of a covalent bond of about 1.5 Å length and also, as the above-mentioned validation results suggest, on the length scale of a hydrogen bond of about 3 Å. There are also indications that some of the most modern GGA functionals that incorporate more constraints derived from the physics of electron–electron interactions yield quite reasonable results for rare gas dimers.²⁷ Hence, even though dispersion interaction are not included explicitly, statements to the effect that “DFT does not contain dispersion” are probably excessive.

In this connection, we will incorporate DFT results using a functional recently developed by Kafafi and co-workers²⁸ and denoted by K2-BVWN, which was designed to handle the long-range interactions taking place in model van der Waals systems, namely noble gas dimers but hydrogen-bonded systems as well. In this methodology, the total exchange correlation energy functional, E_{xc} , is approximated by a sum of two terms, an exchange component, E_{x} , and a correlation term, E_{corr} . The E_{x} term consists of a hybrid mixture of 37.5% exact exchange and the appropriate local spin density exchange using the adiabatic connection formula.^{28a,b} E_{c} is a linear combination of the VWN correlation energy functional of the free electron gas and of a GGA term containing one adjustable parameter.^{28a–c}

The K2-BVWN scheme predicted the binding energies of nine noble gas dimers (helium through xenon) and a variety of charge-transfer complexes accurately.^{28a,b} This was shown to be due to the $\sim 1/R^6$ long-

(12) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K.A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A., Jr. *J. Comput. Chem.* **1993**, *14*, 1347.

(13) Stevens, W. J.; Basch, H.; Krauss, M. *J. Chem. Phys.* **1984**, *81*, 6026.

(14) (a) Garmer, D. R.; Gresh, N. *J. Am. Chem. Soc.* **1994**, *116*, 3556.

(b) Gresh, N.; Stevens, W. J.; Krauss, M. *J. Comput. Chem.* **1995**, *16*, 843.

(15) Kitaura, K.; Morokuma, K. *Int. J. Quantum Chem.* **1976**, *10*, 325.

(16) Stevens, W. J.; Fink, W. H. *Chem. Phys. Lett.* **1987**, *139*, 15.

(17) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.

(18) Cammi, R.; Hofmann, H.-J.; Tomasi, J. *Theor. Chim. Acta* **1989**, *76*, 297.

(19) Pople, J. A.; Binkley, J. S.; Seeger, R. *Int. J. Quantum Chem.* **1976**, *10*, 1.

(20) Parr, R. G.; Yang, W. *Density Functional Theory of Atoms and Molecules*; Oxford University Press: Oxford, 1989.

(21) (a) Andzelm, J. W.; Casida, M. E.; Koester, A.; Proynov, E.; St-Amant, A.; Salahub, D. R.; Godbout, N.; Guan, J.; Jamorski, C.; Leboeuf, M.; Malkin, V.; Malkina, O.; Vela, A. deMon Software, University of Montreal, 1995. (b) Godbout, N.; Salahub, D. R.; Andzelm, J. W.; Wimmer, E. *Can. J. Chem.* **1992**, *70*, 560.

(22) Becke, A. D. *J. Chem. Phys.* **1988**, *88*, 1053.

(23) Perdew, J. P.; Wang, Y. *Phys. Rev.* **1986**, *B33*, 8800.

(24) Perdew, J. P. *Phys. Rev.* **1986**, *B33*, 8822; *Phys. Rev.* **1986**, *B34*, 7406E.

(25) Hobza, P.; Sponer, J.; Reschel, T. *J. Comput. Chem.* **1995**, *16*, 1315.

(26) Osinga, V. P.; van Gisbergen, S. J. A.; Baerends, E. J. *J. Chem. Phys.* **1997**, *106*, 5091.

(27) (a) Zhang, Y.; Pei, W.; Yang, M. *J. Chem. Phys.* **1997**, *107*, 7921.

(b) Patton, D. C.; Pederson, M. R. *Phys. Rev. A* **1997**, *56*, R2495.

(28) (a) Kafafi, S. A.; El-Gharkawy, E. R. H. *J. Phys. Chem. A* **1998**, *102*, 3202. (b) Kafafi, S. A. *J. Phys. Chem. A* **1998**, *102*, 5639. (c) Kafafi, S. A. *J. Phys. Chem. A* **1998**, *102*, 10404. (d) Kafafi, S. A.; Krauss, M. *Int. J. Quantum Chem.*, in press. (e) Kafafi, S. A.; Gregurick, S. K. *J. Carbohydr. Res.*, in press.

range behavior of the correlation potential, which mimics the van der Waals interaction. An additional advantage in using this functional is related to the BSSE corrections, as these were shown to be less than 0.2 kcal/mol.^{28a,b}

In all of the computations reported in this work using the K2-BVWN scheme, full geometry optimizations with analytical gradients using 6-311G** basis sets were performed. These were followed by single-point computations at the K2-BVWN/6-311+g(2df,p) level of theory. Geometry optimizations were terminated when the largest component of the gradient was smaller than 0.0005 Hartree/Bohr. The minimum energy geometries of all of the species of interest obtained at the K2-BVWN/6-311G** level of theory were similar to the corresponding SIBFA-optimized structures.

Computation of the SIBFA Intermolecular Interaction Energies. The intermolecular interaction term, ΔE , is computed as a sum of five separate contributions:

$$\Delta E = E_{\text{MTP}} + E_{\text{rep}} + E_{\text{pol}} + E_{\text{ct}} + E_{\text{disp}}$$

A more complete presentation of the SIBFA procedure is given in recent papers,^{3c,d,6} including a review article.^{3c} The electrostatic (multipolar) energy contribution, E_{MTP} , is computed with distributed multipoles derived from the ab initio SCF wave function of the constitutive fragments. The multipoles (up to quadrupoles) are distributed on the atoms and bond barycenters using the procedure developed by Vigné-Maeder and Claverie.²⁹ E_{rep} , the short-range repulsion energy, is computed as a sum of bond–bond, bond–lone pair, and lone pair–lone pair interactions.

E_{pol} is the polarization energy contribution, calculated with distributed, anisotropic polarizabilities on the individual molecules. The polarizabilities are distributed on the centroids of the localized orbitals (heteroatom lone pairs and bond barycenters) using the procedure of Garmer and Stevens.³⁰ A Gaussian screening of the polarizing field is used. The field polarizing each molecule is determined with the permanent multipoles and the induced dipoles of all the other molecules in an iterative fashion. Two values of the polarization energy will be given in the Results and Discussion section. The first value is denoted as $E_{\text{pol}}^{\text{a}}$ (SIBFA), where the polarizable field exerted on a center of a given fragment is from the permanent multipoles of the other fragments. The second is denoted as $E_{\text{pol}}^{\text{b}}$ (SIBFA), where the polarizing field is due to the permanent multipoles + induced dipole and is computed in an iterative fashion. E_{pol} can be written as

$$E_{\text{pol}}(\text{P}) = -0.5 \sum_i E_0(i) \sum_j \alpha_{\text{P}(i,j)} E(j)$$

Here, E_0 and E designate the polarizing fields due to the permanent multipoles and the permanent multipoles + induced dipoles, respectively, and $\alpha_{\text{P}(i,j)}$ represents the polarizability tensor of components i and j of centroid P. The ab initio SCF computations with the basis set of Stevens et al.¹³ were performed on the individual molecules for the derivation of the distributed multipoles and polarizabilities.

E_{ct} is the charge-transfer contribution. An expression for E_{ct} was derived in our previous papers^{3a,c} starting from the formula due to Murrell et al.³¹ Refinements to SIBFA led to the introduction of a coupling with the polarization. This is done principally at the level of the denominator of E_{ct} , which takes into account the difference between the ionization potential, $I_{L\alpha}$, of the electron donor and the electron affinity, A_{β^*} , of the acceptor. $I_{L\alpha}$ is increased by the predominantly positive electrostatic potential exerted on this atom by all the other molecules in the complex, whereas A_{β^*} is reduced by the predominantly negative electrostatic potential due to its surrounding ligands. These potentials are those due to the permanent multipoles and the induced dipoles of the interacting molecules. More details are given in refs 3a,c and 6.

E_{disp} is the dispersion energy contribution, developed using the formulation of Creuzet et al.³² and expressed as a sum of $1/R^6$, $1/R^8$, and $1/R^{10}$ terms. These terms are reduced at short distances by means of an exponential damping term. An explicit exchange–dispersion term is introduced for the mutual interactions between polyatomic molecules. Directionality effects are accounted for by the explicit introduction of fictitious atoms with reduced van der Waals radii to represent the lone pairs.

Standard bond lengths and valence angles are used throughout. The C–N, C=O, and C–C bond lengths are 1.32, 1.25, and 1.53 Å, respectively. The C–N–C', N–C'=O, N–C'–C, C'–C $_{\alpha}$ –N, and C–C $_{\alpha}$ –C $_{\beta}$ angles are 123.0°, 125.0°, 114.0°, 109.5°, and 109.5° respectively. For the *cis*-NMA and the AD dimers, energy minimization was performed on the six intermolecular variables that define the position of the second molecule with respect to the first one; the “Merlin” minimizer³³ was used. These energy-minimized conformations were then used to perform single-point computations on the dimers of the unmethylated analogues, as well as the ab initio computations of the corresponding dimerization energies.

The evaluation of the transferability of the multipolar expansion is undertaken by the comparison of the ab initio/DFT results with those obtained with SIBFA for three different representations of the peptide:

(a) A multipolar expansion is derived for the whole backbone of the Ala dipeptide in a frozen β -sheet conformation as defined by the ϕ and ψ of -139° and 135° , respectively.

(b) The multipolar expansions on the peptides are based on those of their basic building blocks, *N*-methylformamide (NMF) for the backbone and methane for the Ala side chains; NMF is split into a pseudo-formamide and a pseudo-methane fragment in such a way as to preserve the original multipolar distribution at the two junctional bonds NH and CH. This will allow the rotations along the ϕ torsional angle.

(c) The same expansion as in (b) is adopted, but the redistribution of the multipoles occurring within each dipeptide is accounted for by computing the mutual intermolecular polarization energy between the constitutive fragments. Thus, each of the two C–H bonds at the junction between the successive NMF fragments is collapsed by carrying back its hydrogen and its barycenter onto its carbon atom, namely, the methyl carbon for the NMF fragment on the N-terminal side and the carbonyl carbon for the NMF fragment on the C-terminal side. This is necessary to avoid the overlaps between these two fragments while preserving their original net charge of zero. The intermolecular interaction energy is then computed as the difference between the total energy of the two dimers and those of each dimer in the absence of the other.

Representation b is the one used in applications of SIBFA to conformational studies of oligopeptides,³⁴ and representation c is presently used for the purpose of taking into account long-range polarization effects.

The 1999 version of the SIBFA Fortran code, sample input data, the geometries of the complexes, and the library of fragments used in this study are available upon request from the authors.

Results and Discussion

1. Dimerization of Formamide and GD. The model GD dimer was held in the same geometry as the energy-minimized AD dimer. This was necessary in order to prevent the two GD monomers from approaching each other too closely upon dimer formation due to the absence of a methyl group on the N-terminal end. The structures of the formamide and GD dimers are reported in Figure 1, along with those of the *cis*-NMA and AD dimers. The results of the dimerization energies from the ab initio, DFT, and SIBFA calculations are given in Table 1.

(32) Creuzet, S.; Langlet, J.; Gresh, N. *J. Chim.-Phys. (Paris)* **1991**, *88*, 2399.

(33) Evangelakis, G.; Rizos, J.; Lagaris, I.; Demetropoulos, G. N. *Comput. Phys. Commun.* **1987**, *43*, 401.

(34) (a) Demetropoulos, G. N.; Gresh, N. *J. Comput.-Aided Mol. Design* **1991**, *5*, 81. (b) Gresh, N.; Tiraboschi, G.; Salahub, D. R. *Biopolymers* **1998**, *45*, 405.

(29) Vigné-Maeder, F.; Claverie, P. *J. Chem. Phys.* **1988**, *88*, 4934.

(30) Garmer, D. R.; Stevens, W. J. *J. Phys. Chem.* **1989**, *93*, 8263.

(31) Murrell, J. N.; Randic, M.; Williams, D. R. *Proc. R. Soc. Ser. A* **1966**, *284*, 566.

Table 1. Dimerization Energies of Formamide and GD and Contributions from Different Components and Values of Their Difference, δ (kcal/mol)

(a) Ab Initio SCF/MP2 and DFT/DZVP Computations ^a							
	formamide		GD		δ		
Ab Initio Computations							
$\Delta E(\text{SCF})$	-13.9		-9.3		4.6		
$\Delta E(\text{SCF/BSSE})$	-12.6		-7.9		4.7		
$\Delta E(\text{SCF/BSSE}^*)$	-13.3						
$\Delta E(\text{MP2})$	-18.1		-15.0		3.1		
$\Delta E(\text{MP2/BSSE})$	-13.8		-9.7		4.1		
DFT Computations							
$\Delta E(\text{DFT/BP})$	-14.7 (-13.5)		-8.0 (-7.4)		6.7 (6.1)		
$\Delta E(\text{BP/BSSE})$	-13.9 (-13.5)		-6.8 (-6.9)		7.1 (6.6)		
$\Delta E(\text{DFT/PP})$	-17.4 (-16.0)		-11.2 (-10.5)		6.2 (5.5)		
$\Delta E(\text{PP/BSSE})$	-16.4 (-15.9)		-9.9 (-10.1)		6.5 (5.8)		
Kitaura–Morokuma (KM) and RVS Analysis of SCF Energies							
E_{Coulomb}	-22.1		-14.3		7.8		
E_{exchange}	16.5		10.2		-6.3		
E_1	-5.6		-4.1		1.5		
$E_{\text{pol}}(\text{RVS})$	-4.5						
$E_{\text{pol}}(\text{KM})$	-5.7		-3.0		2.7		
$E_{\text{ct}}(\text{RVS})$	-2.4						
$E_{\text{ct}}(\text{KM})$	-2.6		-2.2		0.4		
E_2	-8.3		-5.2		3.1		
(b) SIBFA Computations ^b							
	formamide	GD					
	energy	(a)		(b)		(c)	
	energy	energy	δ	energy	δ	energy	δ
$E_{\text{MTP}}^{\#}$	-20.9	-13.2	7.7	-9.3	11.6	-9.3	11.6
E_{rep}	13.7	6.5	-7.2	6.6	-7.1	6.9	-6.8
$E_{\text{rep}}^{\#}$	14.4	6.8	-7.6				
E_1	-7.3	-6.7	0.6	-2.7	4.6	-2.4	4.9
$E_1^{\#}$	-6.6	-6.4	0.2				
E_{pol}	-4.4	-2.6	1.8	-1.9	2.5	-3.9	0.5
E_{pol}^*	-3.6	-2.2	1.4	-1.6	2.0	-3.4	0.2
E_{ct}	-1.7	-1.4	0.3	-1.3	0.4	-1.0	0.7
E_2	-6.1	-3.9	2.2	-3.2	2.9	-4.9	1.2
ΔE_0	-13.3	-10.6	2.7	-5.9	7.4	-7.5	5.8
$\Delta E_0^{\#}$	-12.6	-10.3	2.3				
E_{disp}	-4.9	-5.2	-0.3	-5.1	-0.3	-5.2	-0.3
$E_{\text{disp}}^{\#}$	-4.1	-4.3	-0.2				
$\Delta E(\text{SIBFA})$	-18.3	-15.8	2.5	-11.1	7.2	-12.6	5.7
$\Delta E^{\#}(\text{SIBFA})$	-16.7	-14.6	2.1				

^a Values in parentheses are those obtained using the 6-311+G** basis set. ^b $E_{\text{rep}}^{\#}$, $\Delta E_0^{\#}(\text{SIBFA})$, $E_{\text{disp}}^{\#}$, and $\Delta E^{\#}(\text{SIBFA})$ denote the values obtained upon rescaling E_{rep} and E_{disp} by the factors 1.05 and 0.81, respectively.

Ab Initio Results. The ab initio energies are decomposed into a first-order term ($E_1 = E_{\text{Coulomb}} + E_{\text{exchange}}$) and a second-order term ($E_2 = E_{\text{pol}} + E_{\text{ct}}$). The values of these components are obtained using the RVS and the Kitaura–Morokuma (KM) procedures. To avoid the ambiguities inherent in the E_{mix} term in the KM treatment and for the purpose of comparison, we have defined $E_{\text{ct}}(\text{KM})$ as the difference between the total binding energy at the SCF level, $\Delta E(\text{SCF})$, and the sum of E_1 and $E_{\text{pol}}(\text{KM})$. Also listed are the values of the BSSE corrections using the standard Boys procedure¹⁷ as well as the version using only the virtual orbitals (for the formamide dimer only, and denoted as BSSE*). The corresponding $\Delta E(\text{SCF})$ values are denoted as $\Delta E(\text{SCF/BSSE})$ and $\Delta E(\text{SCF/BSSE}^*)$, respectively. Table 1a also reports the correlation energies (E_{corr}) and the intermolecular interaction energies at the MP2 level, $\Delta E(\text{MP2})$; the BSSE corrections at the MP2 level, BSSE(MP2), and the corresponding interaction energies with the BSSE corrections, $\Delta E(\text{MP2/BSSE})$, are also given.

As is evident from Table 1a, formamide is found to have a much larger dimerization energy than GD from the ab initio and DFT calculations. The energy difference amounts to 4.6–4.7 kcal/mol at the SCF level with or without the BSSE correction. This difference is reduced to 3.1 kcal/mol at the MP2 level without the BSSE correction and to 4.1 kcal/mol with the correction. The largest differences are found with the DFT methods, reaching as much as 6–7 kcal/mol. Correction of the BSSE or increase of the size of the basis set from DZVP to 6-311++G** has only a small effect on these differences. The dimerization energies of formamide reported here are consistent with those of previous quantum mechanical studies.^{35,36} Thus, the values of $\Delta E(\text{SCF})$ and $\Delta E(\text{MP2})$ of -13.9 and -18.1 kcal/mol are close to the corresponding -13.4 and -17.4 kcal/mol values obtained by Florian and Johnson using a 6-31G(d,p) basis set.³⁶ The SCF values are also close to that of -12.3 kcal/mol obtained in a pioneering study by Dreyfus and Pullman using a (7s,3p/3s) Gaussian basis set.³⁵ The BSSE correction obtained with the CEP 4-31G+(2d) basis set amounts to 1.4 kcal/mol, somewhat smaller than the 2.2 kcal/mol value given by Florian and Johnson.³⁶ An even smaller value (BSSE* = 0.6 kcal/mol) is obtained with the RVS approach, which takes into account the virtual orbitals only. On the other hand, the magnitude of the BSSE correction at the MP2 level of 4.3 kcal/mol, including the 1.4 kcal/mol SCF contribution, reduces $\Delta E(\text{MP2/BSSE})$ to -13.8 kcal/mol, so that the BSSE-corrected correlation energy would be of only -1.2 kcal/mol. A BSSE value of 5.0 kcal/mol at the correlated level was given in ref 36, having an SCF contribution of 2.2 kcal/mol, similarly reducing the gain in correlation energy from -4.0 to -1.3 kcal/mol. Despite the fact that the cyclic formamide dimer is stabilized by two hydrogen bonds, such reduced E_{corr} values are now smaller than those provided for the correlation contribution to the binding energy of the water dimer, which amounts to -1.5 kcal/mol using an aug-cc-pVTZ basis set,³⁷ an indication that the BSSE at the MP2 level may be overestimating the magnitude of the correction. In their original study of the interaction energies between two hydrogen fluoride molecules, Schwenke and Truhlar showed that correcting for the BSSE does not guarantee the improvement of the dimerization energies.³⁸ However, it could be pointed out that the caveats they offered concern actual repulsive close-contact geometries, and that for equilibrium geometries the BSSE does lead to an improved convergence of the Hartree–Fock binding energies upon increasing the size of the basis set. The dangers of using the BSSE correction in attempts to improve the MP2 intermolecular interaction hypersurfaces of hydrogen fluoride trimers were recently pointed out,³⁹ but this concern applies to a large, diffuse-function-augmented, correlation-consistent basis set which is not presently affordable for the complexes considered here. The need for the BSSE correction at the HF as well as the MP2 levels was emphasized by Feller³⁷ and by Paizs and Suhai.⁴⁰ This raises some concern as to the calibration of the dispersion contribution in the SIBFA procedure. In light of the data provided by the above-mentioned authors, it has led us to rescale E_{disp} and provide an alternative set of SIBFA results, but this will be shown not to affect the relative binding energies of formamide versus GD dimers (see below).

Inspection of the components of the interaction energies indicates that the dominant contributor to the 3–4 kcal/mol

(35) Dreyfus, M.; Pullman, A. *Theor. Chim. Acta* **1970**, *19*, 20.

(36) Florian, J.; Johnson, B. G. *J. Phys. Chem.* **1995**, *99*, 5899.

(37) Feller, D. *J. Chem. Phys.* **1992**, *96*, 6104.

(38) Schwenke, D. W.; Truhlar, D. W. *J. Chem. Phys.* **1985**, *82*, 2418.

(39) Liedl, K. R. *J. Chem. Phys.* **1998**, *108*, 3199.

(40) Paizs, B.; Suhai, S. *J. Comput. Chem.* **1998**, *19*, 575.

energy difference favoring the formamide dimer over the GD dimer is the Coulombic energy term, E_c , which contributes 7.7 kcal/mol. This is strongly opposed within E_1 by the exchange term E_e , which disfavors this dimer by 6.3 kcal/mol. This reduces the preference stemming from E_1 to 1.5 kcal/mol, whereas the corresponding preference by E_2 , dominated by a large differential E_{pol} (2.7 kcal/mol), amounts to 3.1 kcal/mol, i.e., twice as much.

DFT Results. The values given in parentheses are those computed using a 6-311++G** basis set. Although the BP dimerization energies without the BSSE correction (−13.5 and −14.7 kcal/mol) are 3.5–4.5 kcal/mol smaller than the MP2 value of −18.1 kcal/mol, the energies with the BSSE correction (−13.5 and −13.9 kcal/mol) are close to $\Delta E(\text{MP2}/\text{BSSE})$, namely −13.8 kcal/mol, due to the much smaller BSSE corrections with the DFT methods (0–0.8 kcal/mol). The PP dimerization energies are 2.5 kcal/mol larger than the corresponding values from the BP calculations, and are in the −16 to −17.4 kcal/mol range. As with the BP functional, the BSSE is smaller than 1 kcal/mol.

For GD, no previous results are available for comparison. The BP energies are about 1–2 kcal/mol smaller than the corresponding SCF values and much smaller (3–7 kcal/mol) than the MP2 values. The PP values are in the −10.5 to −11.2 kcal/mol range, and the BSSE correction is very small (0.4 kcal/mol).

SIBFA Results. The SIBFA results are given in Table 1b. We first use representation a, using the whole backbone of the dipeptide, and compare the dimerization energies of formamide and GD. Without E_{disp} , $\Delta E_0(\text{SIBFA})$ for the formamide dimer amounts to −13.3 kcal/mol, close to the $\Delta E(\text{SCF})$ and $\Delta E(\text{SCF}/\text{BSSE})$ values of −13.9 and −12.6 kcal/mol, respectively. For the GD dimer, $\Delta E_0(\text{SIBFA})$ amounting to −10.6 kcal/mol is somewhat larger than the $\Delta E(\text{SCF})$ and $\Delta E(\text{SCF}/\text{BSSE})$ values of −9.3 and −7.9 kcal/mol, respectively.

An important issue we need to address is that of the real magnitude of the BSSE correction to be brought to the correlation energy, $E_{\text{corr}}(\text{MP2})$, the energy contribution which $E_{\text{disp}}(\text{SIBFA})$ is designed to reproduce. Recent computations on multiply H-bonded complexes, such as four distinct cyclic water tetramers, showed SIBFA to reproduce well not only the results of $\Delta E(\text{MP2})$ using the CEP 4-31G+(2d) basis set, but also those using a more extended basis set, namely 6-311+G(2d,2p) in the absence of BSSE.^{6b} In the present case, the inclusion of E_{disp} leads to the dimerization energies of −18.3 and −15.8 kcal/mol for the formamide and GD dimers, respectively. These values are very close to those of $\Delta E(\text{MP2})$ computed in the absence of the BSSE correction, which amount to −18.1 and −15.0 kcal/mol, respectively. Such numerical agreements are consistent with those found in the water oligomers.⁶ However, these two $\Delta E(\text{SIBFA})$ values are much larger than the corresponding BSSE-corrected values of −13.8 and −9.7 kcal/mol, as well as those from the DFT calculations. In view of the concerns raised above as to the “real” magnitude of the correlation energy, this has led us to reconsider the calibration of E_{disp} . This contribution was recently calibrated^{6a} by referring to symmetry-adapted perturbation theory computations⁴¹ which do not suffer from basis set extension effects. Thus, for the linear water dimer at equilibrium distance ($d_{\text{O-O}} = 2.90 \text{ \AA}$), the value of E_{disp} computed by Langlet et al. is −1.84 kcal/mol,

(41) (a) Hess, O.; Caffarel, M.; Langlet, J.; Caillet, J.; Huiszoon, C.; Claverie, P. In *Modelling of molecular structures and properties*; Rivail, J.-L., Ed.; Studies in Physical and Theoretical Chemistry 71; Elsevier: Amsterdam, 1990; p 323. (b) Hess, O.; Caffarel, M.; Claverie, P. *J. Chem. Phys.* **1990**, 92, 6049.

Table 2. Components of the Electrostatic Energy, E_{MTP} , in the Formamide and GD Dimers and Their Differences, δ (kcal/mol)

	formamide energy	GD			
		representation a		representation b	
		energy	δ	energy	δ
E_{MTP}	−20.9	−13.2	7.7	−9.3	11.6
E_{mm}	−16.2	−20.0	−3.8	−14.0	2.2
E_{md}	−1.6	5.8	7.4	2.2	3.8
E_{mq}	−3.1	1.8	4.9	3.1	6.2
E_{dd}	0.5	−0.6	−1.1	1.1	0.6
E_{dq}	0.1	−1.0	−1.1	−2.5	−2.6
E_{qq}	−0.5	0.8	1.3	0.9	1.4

encompassing an exchange contribution $E_{\text{disp-exch}}$ of 0.48 kcal/mol.⁴² The corresponding values in SIBFA are −1.83 and 0.52 kcal/mol. The value of $\Delta E_0(\text{SIBFA})$, −3.9 kcal/mol, leads to a total $\Delta E(\text{SIBFA})$ value of −5.7 kcal/mol for the linear water dimer. $\Delta E_0(\text{SIBFA})$ compares well to the estimate of the HF limit interaction energy of −3.6 kcal/mol computed by Feller,³⁷ but $\Delta E(\text{SIBFA})$ is larger than the value of -5.0 ± 0.1 kcal/mol recently derived by Feyereisen et al.⁴³ and recommended as a reference value for molecular mechanics potentials. For the water dimer, such a target value can be obtained by rescaling the values of E_{rep} and E_{disp} by factors of 1.05 and 0.81, yielding values of $\Delta E_0(\text{SIBFA})$ and $\Delta E(\text{SIBFA})$ of −3.6 and −5.1 kcal/mol, respectively. We accordingly report in Tables 1 and 2 the rescaled SIBFA values along with the ones using the standard calibration.

The computed 2.5 kcal/mol difference in the $\Delta E(\text{SIBFA})$ dimerization energies favoring formamide over GD is close to the corresponding difference at the uncorrected MP2 level of 3.1 kcal/mol. The rescaled values, $\Delta E^\#(\text{SIBFA})$, of −16.7 and −14.6 kcal/mol reduce this difference by only 0.4 kcal/mol.^{44,45}

Examination of the individual energy components of $\Delta E(\text{SIBFA})$ shows the electrostatic term, E_{MTP} , to provide the major contribution to the relatively larger dimerization energy of formamide, a result consistent with the ab initio computations. The value of δ amounting to 7.7 kcal/mol is virtually identical to the corresponding δ value of the $E_c(\text{SCF})$ component of 7.8 kcal/mol. E_{MTP} is counteracted by E_{rep} , leading to an E_1 preference for formamide by only 0.6 kcal/mol. This value is smaller than the corresponding 1.5 kcal/mol value at the SCF level. E_2 makes the major contribution, 2.2 kcal/mol, to the overall δ value, also consistent with the ab initio results. E_{disp} favors the GD dimer by only a small amount (−0.3 kcal/mol).

Representation a cannot be used to construct larger-sized oligopeptides. This is because the two NMF moieties making up the GD backbone have different electronic distributions due to their different local environments. In the β -sheet conformation, the NH and CO groups around the central C_α carbon (denoted as NH_i and CO_i , respectively, in Figure 2) are mutually close, whereas the C=O and NH ones flanking them on the N and C ends, respectively (denoted as CO_{i-1} and NH_{i+1}), are away from polar neighbors. Extension of the β -sheet backbone on the N-terminal side is done by the replacement of the CH_3

(42) Langlet, J.; Caillet, J.; Caffarel, M. *J. Chem. Phys.* **1995**, 103, 8043.

(43) Feyereisen, M. W.; Feller, D.; Dixon, D. A. *J. Phys. Chem.* **1996**, 100, 2993.

(44) An instructive estimate of the possible impact of rescaling E_{rep} and E_{disp} can be provided by the case of another doubly hydrogen-bonded dimer, that of formic acid. A recent computation by Liedl et al.⁴⁵ at the MP2/aug-cc-pVTZ level gives dimerization energies of −16.8 and −15.4 kcal/mol without and with the BSSE correction. The values of $\Delta E(\text{SIBFA})$ are −17.4 and −15.6 kcal/mol without and with rescaling, respectively.

(45) Liedl, K. R.; Sekusak, S.; Mayer, E. *J. Am. Chem. Soc.* **1997**, 119, 3782.

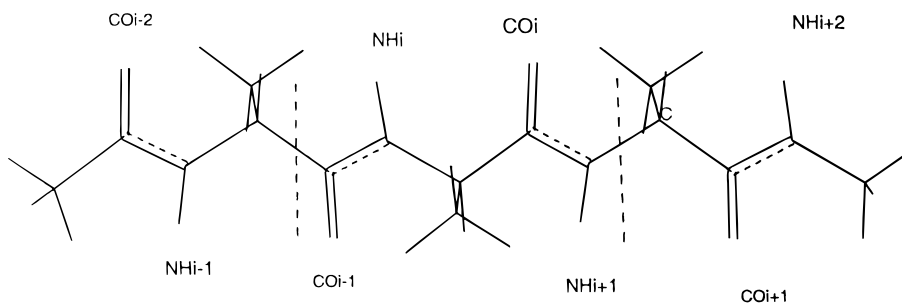


Figure 2. Extension of the oligopeptide backbone from both termini of the central "dipeptide" unit.

group by a *N*-methylacetamide moiety. It is done on the C-terminal side by the incorporation of an NMF group along the CH bond of the terminal NMF. This results in the CO_{i-1} and NH_{i+1} groups now being in turn flanked by the polar amino and carbonyl groups NH_{i-1} and CO_{i+1} , respectively, whereas the new terminal groups CO_{i-2} and NH_{i+2} are now those devoid of polar neighbors. Each extension of GD and AD thus results in a new alteration of the electronic distribution of the backbone due to the addition of polar neighbors to the terminal ends, as well as the propagation of conjugation effects. Furthermore, the intensities of the multipoles are likely to vary as a function of the conformation. The need for transferable multipole models for the electrostatic interactions of peptides and amides was first pointed out by Faerman and Price.¹¹ GD or AD could not be used as peptide building blocks, owing to the imbalance in the electronic distributions of their constitutive $\text{CO}_{i-1}\text{NH}_i$ and $\text{CO}_i\text{NH}_{i+1}$ moieties, whereas these should be more balanced in actual oligopeptides. The simplest candidate as a peptide building block is NMF itself. It was used in representation b to build GD from two successive NMF molecules, each of which is split into a pseudo-formamide and a pseudo-methane fragment, and the β -sheet conformation is obtained by the appropriate rotations around the N–C and C–C bonds at the junctions.³⁴ Each of these two moieties interacts with the two of the other dimer, but not with its congener in the same peptide. However, Table 2 shows that this representation significantly underestimates E_{MTP} , which decreases from -13.2 to -9.3 kcal/mol upon going from representation a to b. E_{pol} also decreases, but only by 0.7 kcal/mol. As a result, $\Delta E_0(\text{SIBFA})$ and $\Delta E(\text{SIBFA})$ are reduced to -5.9 and -11.1 kcal/mol, respectively, and the dimerization energy of GD becomes 7.2 kcal/mol smaller than that of formamide.

Representation b uses a multipolar expansion for NMF that is unperturbed by the proximity of its congener in the actual dipeptide. A possible way to account for the alteration of the electronic distribution due to its integration into a larger molecular structure is to include the mutual NMF–NMF polarization energy which occurs in the dipeptide simultaneously with that taking place between the two peptides. This is done in representation c. We note that such a representation is in line with the statement by Faerman and Price, that "some model for polarization effect shall be required in a peptide intermolecular force field for it to be able to model molecular recognition processes with quantitative accuracy".¹¹ The total $\Delta E(\text{SIBFA})$ is then computed as the difference between the peptide–peptide intermolecular interaction and that occurring within each individual peptide. This is equivalent to considering the GD dimer as a complex formed between four separate NMF molecules and subtracting from the complexation energy the "dimerization" energies of the pairs of NMFs within each individual peptide. This occurs at the cost of making the necessary approximation that, within each peptide, the C–H bonds at the junctions of successive NMFs are collapsed onto

their C atoms. This alone can avoid overflows due to overlaps between these units while preserving their net charge. Table 1b shows the value of E_1 in representation c for GD dimerization to remain close to that obtained for representation b, namely -2.4 versus -2.7 kcal/mol. This indicates that collapsing the junctional CH bonds onto their C atoms did not downgrade the representation of E_{MTP} and E_{rep} for computing intermolecular interaction energies. Representation c is seen to lead to an improvement in the value of $\Delta E(\text{SIBFA})$ for GD dimerization, which increases from -11.1 kcal/mol in representation b to -12.6 kcal/mol in representation c, as compared to -15.8 kcal/mol in representation a. This is due to a gain of 2 kcal/mol in the polarization energy. Such a gain appears encouraging, because it implies that electronic redistribution effects taking place in an actual glycol dipeptide molecule, accounted for in representation a but absent in representation b, can be recovered to some extent by incorporating the polarization energy between the NMF molecules. The gain is smaller than if one were dealing with pure "intermolecular" interaction energies, implying that further improvements have to be sought to handle the interactions occurring at the junction between two successive NMF fragments. This is the incentive of an ongoing investigation of intramolecular polarization effects in alanine tetrapeptides and their impact on intermolecular interaction as well as conformational energies.

Owing to the important role of E_{MTP} for the differential dimerization energy, we have listed in Table 2 the values of each of the six components in both dimers, from monopole–monopole (E_{mm}) to quadrupole–quadrupole (E_{qq}) in representations a and b. E_{mm} is seen to provide the dominant contribution to E_{MTP} in all cases. In representation a, it is actually larger by 3.8 kcal/mol in the GD dimer than in the formamide dimer. Limiting E_{MTP} to this term only would result in a dramatic inversion in the ordering of dimerization energies of formamide and GD in this representation, with a difference of 9.0 kcal/mol now favoring the former. E_{md} and E_{mq} , the monopole–dipole and monopole–quadrupole terms, have a reduced amplitude, but both are attractive for the formamide dimer and sum up to -4.7 kcal/mol, whereas they are both repulsive in the GD dimer and sum up to 7.6 kcal/mol. They thus contribute 12.3 kcal/mol to favoring the formamide dimer, an amount that is 3 times as large as the opposite amount due to E_{mm} . The dipole–dipole, dipole–quadrupole, and quadrupole–quadrupole terms, E_{dd} , E_{dq} , and E_{qq} , are mutually compensatory. They are virtually canceled out in the formamide dimer and sum up to a small attractive contribution (-0.8 kcal/mol) in the GD dimer. In representation b, E_{mm} is only 2.2 kcal/mol smaller in the GD dimer than in the formamide dimer. Limiting E_{MTP} to only this term would again result in an inversion of the relative formamide versus GD dimerization energies, now amounting to 2.2 kcal/mol. Similar to representation a, E_{md} and E_{mq} are both repulsive

for the GD dimer; they sum up to 5.3 kcal/mol, whereas E_{dd} , E_{dq} , and E_{qq} sum up to a weak attractive contribution of -0.6 kcal/mol.

This analysis shows that, within the framework of our distributed multipole expansion, limiting E_{MTP} to its monopole–monopole term, as is done in all conventional force fields, would result in an erroneous energy preference favoring GD over formamide dimerization. In the present case, the anisotropy brought about by the inclusion of explicit polarization and charge-transfer terms in SIBFA, as well as of lone pair directionality, would not be sufficient, in the absence of higher-order multipoles, to prevent the inversion of relative dimerization energies.

2. Dimerization of *cis*-NMA and AD. For the *cis*-NMA dimer, the O–H hydrogen bond distances optimized by DFT at the BP/DZVP level and by SIBFA are 1.78 and 1.93 Å, respectively. The C=O–H and O–H–N angles are close to 120° and 180° in both procedures. For the AD dimer, the DFT distances are 1.97 and 2.00 Å, virtually identical to the 1.99 and 2.00 Å from SIBFA. The C=O–H angles have average values of 155° (DFT) and 160° (SIBFA). The O–H–N angles have average values of 166° (DFT) and 161° (SIBFA).

The dimerization energies from the ab initio (without considering the BSSE correction at this stage) and DFT computations are given in Table 3a. At the SCF level, AD has a smaller dimerization energy than GD (-7.6 kcal/mol as compared to -9.3 kcal/mol), whereas *cis*-NMA has a slightly improved (0.4 kcal/mol) one compared to formamide. On the other hand, at the MP2 level, AD now has a better dimerization energy than GD, namely -17.5 kcal/mol as compared to -15.0 kcal/mol. This could partly reflect the stabilization brought by the “dispersion” interactions contributed by the mutual interactions of the methyl groups. The BP dimerization energies of *cis*-NMA obtained without the BSSE correction, in the range -14.8 to -17.0 kcal/mol for *cis*-NMA, are close to the MP2/aug-cc-pVDZ//HF/DZP value of -16.4 kcal/mol reported in a previous study.⁴⁶ The BSSE-corrected value of -16.2 kcal/mol is close to that of -15.8 kcal/mol obtained using a scaled BSSE correction. The dimerization energies from the PP functionals are greater than those from the BP functionals, as observed before. The BP energies appear to be small for the AD dimerization as compared to the LMP2 values obtained by Beachy et al.² Thus, the AD dimerization energy reported by these authors at the LMP2/cc-pVTZ(-f) level with the HF CP correction amounts to -10.7 kcal/mol, namely about 2 kcal/mol larger than in the BP calculations. Although the PP geometry optimization is not performed, it appears, on the other hand, that the PP dimerization energies for AD, in the -11.9 to -13.6 kcal/mol range, are larger than the LMP2/cc-pVTZ(-f) values. As observed earlier, the differential dimerization energies from the DFT calculations are substantially larger than the 2.85 kcal/mol value derived by Beachy et al. at the HF/6-31G* level.^{2,47} This is due to the larger dimerization energies for *cis*-NMA and the smaller dimerization energies for AD from the DFT calculations. The exact reason for this behavior of the PP functional upon dealing with the AD dimer is not clear, and it seems that it is not due to a basis set effect,

(46) Dixon, D.; Dobbs, K. D.; Valentini, J. J. *J. Phys. Chem.* **1994**, *98*, 13435.

(47) The Beachy paper reported a smaller difference of 1.35 kcal/mol, but this value has been found to result from the inadvertent use of an incompletely optimized geometry for the alanine dipeptide monomer. The correct HF/6-31G* stabilization energy for the AD dimer is -10.30 kcal/mol, not the -11.80 kcal/mol reported by Beachy (Halgren, T. A., personal communication).

Table 3. Values of the Dimerization Energies of *cis*-NMA and AD and of Their Differences, δ (kcal/mol)^a

(a) Ab Initio and DFT Results							
	<i>cis</i> -NMA energy		AD energy		δ		
Ab Initio Results ^a							
SCF	-14.3		-7.6		6.7		
MP2	-20.5		-17.5		3.0		
DFT Results, ΔE (DFT)							
BP functional							
<i>a</i>	-15.4		-7.6		7.8		
<i>b</i>	-14.8		-7.8		7.0		
<i>c</i>	-17.0		-9.1		7.9		
<i>d</i>	-16.2		-8.4		7.8		
PP functional							
<i>a</i>	-18.1		-11.9		6.2		
<i>b</i>	-18.5		-13.6		4.9		
K2 functional							
<i>b</i>	-14.9		-14.0		0.9		
(b) SIBFA Computations							
		AD					
	<i>cis</i> -NMA	(a)		(b)		(c)	
	energy	energy	δ	energy	δ	energy	δ
E_{MTP}	-20.6	-14.1	6.5	-10.2	10.4	-10.2	10.4
E_{rep}	13.7	9.7	-4.0	9.8	-4.1	10.0	-3.7
$E_{rep}^{\#}$	14.4	10.2	-4.2				
E_1	-6.9	-4.4	2.5	-0.4	6.5	-0.4	6.5
$E_1^{\#}$	-6.2	-3.9	2.3				
E_{pol}	-4.7	-3.5	1.2	-2.7	2.0	-5.5	-0.8
E_{pol}^*	-3.9	-2.9	1.0	-2.3	1.6	4.4	-0.5
E_{ct}	-1.7	-1.5	0.2	-1.4	0.3	-1.0	0.7
E_2	-6.4	-4.9	1.3	-4.1	2.3	-6.5	-0.1
ΔE^0 (SIBFA)	-13.3	-9.3	4.0	-4.5	8.8	-6.7	6.6
ΔE^0 [#] (SIBFA)	-12.6	-8.8	3.8				
E_{disp}	-5.4	-7.8	-2.4	-7.8	-2.4	7.9	-2.4
$E_{disp}^{\#}$	-4.4	-6.3	-2.1				
ΔE (SIBFA)	-18.7	-17.1	1.6	-12.3	6.4	-14.6	4.1
$\Delta E^{\#}$ (SIBFA)	-17.0	-15.1	1.9				

^a Based on the SIBFA-optimized geometries for the dimer along with the standard and frozen SIBFA monomer geometries. ^b The DFT-optimized geometries for the dimer and the monomers independently. ^c The DFT-optimized geometries for the dimer and the monomer geometries based on those found in the dimer. ^d Same as in c, but taking into account the Boys–Bernardi BSSE counterpoise correction. ^e $E_{rep}^{\#}$, ΔE^0 [#](SIBFA), $E_{disp}^{\#}$, and $\Delta E^{\#}$ (SIBFA) denote the values obtained upon rescaling E_{rep} and E_{disp} by the factors 1.05 and 0.81, respectively. The SIBFA-optimized dimer geometries use the same frozen monomer geometries.

as the results on formamide and GD show such an effect to be small. Further investigation, including studies with the LAP correlation functionals,⁴⁸ is underway in our laboratories. Interestingly, the K2 functional, which was designed to handle model van der Waals systems, is the one that yields the smallest *cis*-NMA versus AD dimerization energy difference (0.9 kcal/mol). Thus, the MP2 energy difference of 3 kcal/mol is bracketed between the two sets of DFT computations. For the time being, we simply note that, although MP2 incorporates important correlation effects and provides convenient comparisons, it should not be taken as an absolute standard.

The SIBFA dimerization energies are given in Table 3b. The SIBFA dimerization energy of *cis*-NMA is slightly (0.4 kcal/mol) larger than that of formamide because of the dispersion energy term. Consistent with the SCF computations, ΔE^0 (SIBFA) is less favorable for AD than for GD dimerization. This is due to the E_1 term, which is 2.3 kcal/mol smaller than the former, owing to E_{rep} , whereas E_2 has an inverse but weaker preference.

(48) Sirois, S.; Proynov, E. I.; Nguyen T.; Salahub, D. R. *J. Chem. Phys.* **1997**, *107*, 6770 and references therein.

Table 4. Components of the Electrostatic Energy, E_{MTP} , in the *cis*-NMA and AD Dimers (kcal/mol)^a

	<i>cis</i> -NMA energy	AD			
		representation a		representation b	
		energy	δ	energy	δ
E_{MTP}	-20.6	-14.1	6.5	-10.2	10.4
E_{mm}	-15.5	-21.4	-5.9	-15.1	0.4
E_{md}	-2.1	6.7	8.8	2.8	4.9
E_{mq}	-3.0	1.3	4.3	2.7	5.7
E_{dd}	0.5	-0.9	-1.6	0.9	0.4
E_{dq}	0.1	-0.8	-0.9	-2.2	-2.3
E_{qq}	-0.5	0.8	1.3	0.8	1.3

^a δ denotes the difference in interaction energies of *cis*-NMA and AD.

ΔE_0 (SIBFA) amounts to -9.3, -4.5, and -6.7 kcal/mol in representations a–c, respectively. It is noteworthy that the values found in (a) and (c) bracket the value of -8.2 kcal/mol computed by Beachy et al. at the Hartree–Fock level, using a correlation-consistent polarized valence triple- ζ basis set without f functions and with the BSSE correction.² The decreased values of ΔE_0 (SIBFA) for AD compared to those for GD result in an increase of the *cis*-NMA versus AD energy difference, which amounts to 4.0, 8.8, and 6.6 kcal/mol in representations a–c. On the other hand, E_{disp} has a 2.6 kcal/mol larger value in the AD dimer than in the GD dimer, reflecting the effect of two additional methyl groups on each monomer, on the N-terminal end and as the Ala side chain. As with the MP2 calculations, this results in, overall, more favorable values for AD than GD dimerization, the difference amounting to 1.6 kcal/mol in representation a. The ΔE (SIBFA) values are -17.1, -12.3, and -14.6 kcal/mol in representations a, b, and c, respectively. This is reflected in differences in *cis*-NMA versus AD dimerization energies of 1.6, 6.4, and 4.1 kcal/mol with (a), (b), and (c), respectively, that are smaller than those with the unmethylated analogues. The values of ΔE (SIBFA) are closer to those of ΔE (DFT) obtained using the PP functional than to those obtained using the BP one, similar to the trends observed for the unmethylated analogues. The rescaled values $\Delta E^{\#}$ (SIBFA) amount to -17.0 and -15.1 kcal/mol for *cis*-NMA and AD dimers, whence a modest (0.3 kcal/mol) increase of δ .

Analysis of the factors leading to the more favorable *cis*-NMA dimerization energy, reported in Table 4, shows the same features as with the unmethylated analogues regarding the decisive role of E_{MTP} , and, within it, of the E_{md} and E_{mq} components and the additional contribution of E_2 to the overall δ . We also note the improved agreement of (c) with (a) relative to that obtained with (b).

The attainment of a more favorable dimerization energy for *cis*-NMA than for AD represents a significant improvement with respect to the conclusions from conventional molecular mechanics force fields. Among the 10 force fields tested by Beachy et al., nine were found to favor AD rather than *cis*-NMA dimerization by energy differences in a 0.9–3.8 kcal/mol range. The AMBER* force field (a modified version of AMBER implemented in the BatchMin 5.5 module of the MacroModel program suite) alone favored *cis*-NMA rather than AD dimerization, but by only a small (0.5 kcal/mol) margin.

The value of the actual difference in dimerization energies between *cis*-NMA and AD is an important issue. In the ab initio computations by Beachy et al.,² such a difference amounts to 2.9 kcal/mol. In the absence of E_{disp} , representation a gives an energy difference of 4.0 kcal/mol. Inclusion of E_{disp} reduces this difference to 1.6 and 1.9 kcal/mol without and with rescaling, respectively. Our large-basis-set DFT computations,

on the other hand, give larger energy differences that are in the 4.9–7.9 kcal/mol energy range. For the SIBFA computations, representation c, yielding a value of 4.0 kcal/mol, is the only one of the representations we have considered that is tractable for the study of large and flexible molecules and is being currently used for conformational studies of oligopeptides. It has also been recently applied to compute the mono- and bidentate complexation energies of the Zn^{2+} cation to flexible ligands having two ligating groups, such as glycine (Gilard et al., manuscript in preparation) or mercaptan α - and β -carboxamides, which are the Zn^{2+} binding units of a class of metalloprotease inhibitors (Tiraboschi et al., manuscript in preparation). In these cases, construction of the ligands using their separate constituent fragments and allowing for their mutual interactions in the presence of Zn^{2+} enabled a good reproduction of the ab initio binding energies for a range of representative conformations. It could thus be used as an alternative to deriving a new set of multipoles and polarizabilities for every novel and frozen conformation. Improvements on this approach are being sought in order to get a better reproduction of the available ab initio or DFT results.

Conclusions

We have compared the dimerization energies of two essential peptide and protein recognition motifs: formamide and *cis*-N-methylacetamide (*cis*-NMA) on one hand, and the Gly and Ala dipeptides in a β -sheet conformation on the other hand. This was done by a combination of ab initio SCF/MP2, DFT, and molecular mechanics computations. Energy decomposition of the ab initio supermolecule interaction energies performed for the formamide and GD dimers showed the Coulombic contribution to be the essential factor leading to a stronger dimerization energy for formamide than for GD. The correct representation of the Coulombic interaction is critical for accurate molecular mechanics. Within the framework of the SIBFA molecular mechanics procedure, this is achieved by the use of multipoles (monopoles, dipoles, and quadrupoles) distributed on the atoms and bond barycenters, which are derived from an ab initio computation on the isolated molecule or molecular fragment. Analysis of the components of E_{MTP} showed that, even though they had smaller numerical values than the monopole–monopole term, the monopole–dipole and monopole–quadrupole terms were decisive, giving rise to a better formamide dimerization as compared to GD. Truncating E_{MTP} to leave only the monopole–monopole term leads to an inversion of the order of formamide versus GD as well as *cis*-NMA versus AD dimerization energies, similar to that observed for the conventional force-fields. However, E_{MTP} was opposed within E_1 by the short-range energy E_{rep} , necessitating the intervention of E_{pol} and E_{ct} within E_2 to correctly reproduce the dimerization energy differences. Thus, in addition to the need for anisotropy, the need to account for the behavior of each individual component of the total energy is underlined. This is further highlighted upon comparing the formamide versus GD dimerization energy to the corresponding *cis*-NMA versus AD one. E_{disp} was found to have a marked preference now favoring AD over *cis*-NMA, counteracting E_2 , and reducing the latter dimerization energy difference with respect to that found with the unmethylated analogues.

The DFT computations favored *cis*-NMA over AD dimerization, consistent with the ab initio results of Beachy et al., but amplified the *cis*-NMA versus AD dimerization energy difference, an enhancement more pronounced with the BP functional (7.9 kcal/mol) than with PP (4.9 kcal/mol).

Because of the concern that the value of $E_{\text{corr}}(\text{MP2})$ could be overestimated on account of BSSE effects, and in light of recently available results on the linear water dimer investigated using an aug-cc-pVTZ basis set,^{37,43} we have rescaled, in an alternative set of computations, the values of E_{rep} and E_{disp} , using factors of 1.04 and 0.84, respectively. This resulted in small reductions of the $\Delta E(\text{SIBFA})$ values, passing from -18.7 and -17.1 kcal/mol for *cis*-NMA and AD, respectively, to -17.0 and -15.1 kcal/mol, thereby affecting by only 0.3 kcal/mol the value of the dimerization energy difference. Further refinements to our procedure are the inclusion of penetration effects in the first-order energy term, E_1 (Gresh, N.; Giessner-Prettre, work in progress) and accounting for the effects of correlation on the multipoles and polarizabilities.

An important issue which arose from the present investigation relates to the transferability of the multipolar expansion, as addressed first by Faerman and Price.¹¹ For GD and AD, the ideal procedure, denoted as representation a, consists of computing the multipolar expansion of the dipeptide backbone frozen in the specific conformation of interest, in this instance the canonical β -sheet. This is, however, untractable upon extending the peptide or performing a conformational change around the ϕ and ψ angles. On the other hand, using the multipolar expansion of the isolated constituent of the peptide backbone,

namely NMF, was seen to incur a significant underestimation of the intermolecular dimerization energies of GD and AD (representation b). As an alternative, we took into account the NMF–NMF interaction within each peptide simultaneously with the interpeptide interaction energy (representation c). This was done in order to tentatively account for the electronic redistributions occurring in the *ab initio* computations due to both intra- and intermolecular effects through the polarization energy component computed by molecular mechanics. Such a representation should enable computations on larger oligopeptides. It was found to produce an improved agreement with the results of representation a. To reduce the remaining energy difference between (a) and (c), further improvements have still to be sought along these lines, particularly for the treatment of intramolecular polarization effects at the junction between two successive NMF fragments.

Acknowledgment. Support from Astra Research Center, Montréal, for part of this work is gratefully acknowledged. The *ab initio* computations were performed on the computers of the Institut du Développement et des Ressources en Informatique Scientifique (IDRIS), 91405 Orsay, France. We wish to thank the reviewers for constructive comments.

JA9742489