UC Irvine
UC Irvine Previously Published Works

Title
Accurate Reproduction of Quantum Mechanical Many-Body Interactions in Peptide Main-Chain Hydrogen-Bonding Oligomers by the Polarizable Gaussian Multipole Model

Permalink
https://escholarship.org/uc/item/5ts318mh

Journal
Journal of Chemical Theory and Computation, 18(10)

ISSN
1549-9618

Authors
Zhao, Shiji
Wei, Haixin
Cieplak, Piotr
et al.

Publication Date
2022-10-11

DOI
10.1021/acs.jctc.2c00710

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

J Chem Theory Comput. Author manuscript; available in PMC 2023 May 02.

Published in final edited form as:

J Chem Theory Comput. 2022 October 11; 18(10): 6172–6188. doi:10.1021/acs.jctc.2c00710.

Accurate Reproduction of Quantum Mechanical Many-Body Interactions in Peptide Main-Chain Hydrogen-Bonding Oligomers by the Polarizable Gaussian Multipole Model

Shiji Zhao,

Departments of Molecular Biology and Biochemistry, Chemical and Biomolecular Engineering, Materials Science and Engineering, and Biomedical Engineering, University of California, Irvine, Irvine, California 92697, United States

Haixin Wei,

Departments of Molecular Biology and Biochemistry, Chemical and Biomolecular Engineering, Materials Science and Engineering, and Biomedical Engineering, University of California, Irvine, Irvine, California 92697, United States

Piotr Cieplak,

SBP Medical Discovery Institute, La Jolla, California 92037, United States

Yong Duan,

UC Davis Genome Center and Department of Biomedical Engineering, University of California, Davis, Davis, California 95616, United States

Ray Luo

Departments of Molecular Biology and Biochemistry, Chemical and Biomolecular Engineering, Materials Science and Engineering, and Biomedical Engineering, University of California, Irvine, Irvine, California 92697, United States

Corresponding Authors Yong Duan – *UC Davis Genome Center and Department of Biomedical Engineering, University of California, Davis, Davis, California 95616, United States;* duan@ucdavis.edu; **Ray Luo** – *Departments of Molecular Biology and Biochemistry, Chemical and Biomolecular Engineering, Materials Science and Engineering, and Biomedical Engineering, University of California, Irvine, Irvine, California 92697, United States;* rluo@uci.edu.

Author Contributions

S.Z. performed pGM parameterizations and molecular mechanics calculations. H.W. participated in preparing the parameter and topology files for the pGM models. Y.D. performed all QM energy calculations. S.Z. and Y.D. wrote the original manuscript. P.C., Y.D., and R.L. directed the project and supported manuscript revision. All authors have given approval to the final version of the manuscript.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.2c00710.

The formamide dimer hydrogen-bonding conformation (Figure S1); the Gly:Gly dimer hydrogen-bonding conformations (Figure S2); the $Gly:Gly_2$ trimer hydrogen-bonding conformations (Figure S3); the $Gly:Gly_3$ tetramer hydrogen-bonding conformations (Figure S4); the $Gly_2:Gly_2$ tetramer hydrogen-bonding conformations (Figure S5); the $Gly_3:Gly_3$ hexamer hydrogen-bonding conformations (Figure S6); the interaction energies of the two middle peptides $IE_{mid}(Gly_m:Gly_n)$ in the presence of the neighboring peptides of glycine dipeptide oligomers $Gly_m:Gly_n$ calculated by $\omega B97X-D$-D/aTZ and polarizable force fields (kcal/mol) (Table S1); the interaction energies $IE(Gly_m:Gly_n)$ and many-body interaction energies $ME(Gly_m:Gly_n)$ of glycine dipeptide oligomers $Gly_m:Gly_n$ calculated by the pGM-ind model with the alternative polarizabilities (kcal/mol) (Table S2); and the nonadditive contributions $ME_{NA}(Gly_m:Gly_n)$ and additive contributions $ME_A(Gly_m:Gly_n)$ to the many-body interaction energies of glycine dipeptide oligomers $Gly_m:Gly_n$ calculated by the pGM-ind model with the alternative polarizabilities (kcal/mol) (Table S2) (PDF)

Coordinates of glycine dipeptide oligomers used for QM and molecular mechanical force fields calculations (GLY_OLIG.xyz) (XYZ)

The authors declare no competing financial interest.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Abstract

A key advantage of polarizable force fields is their ability to model the atomic polarization effects that play key roles in the atomic many-body interactions. In this work, we assessed the accuracy of the recently developed polarizable Gaussian Multipole (pGM) models in reproducing quantum mechanical (QM) interaction energies, many-body interaction energies, as well as the nonadditive and additive contributions to the many-body interactions for peptide main-chain hydrogen-bonding conformers, using glycine dipeptide oligomers as the model systems. Two types of pGM models were considered, including that with (pGM-perm) and without (pGM-ind) permanent atomic dipoles. The performances of the pGM models were compared with several widely used force fields, including two polarizable (Amoeba13 and ff12pol) and three additive (ff19SB, ff15ipq, and ff03) force fields. Encouragingly, the pGM models outperform all other force fields in terms of reproducing QM interaction energies, many-body interaction energies, as well as the nonadditive and additive contributions to the many-body interactions, as measured by the root-mean-square errors (RMSEs) and mean absolute errors (MAEs). Furthermore, we tested the robustness of the pGM models against polarizability parameterization errors by employing alternative polarizabilities that are either scaled or obtained from other force fields. The results show that the pGM models with alternative polarizabilities exhibit improved accuracy in reproducing QM many-body interaction energies as well as the nonadditive and additive contributions compared with other polarizable force fields, suggesting that the pGM models are robust against the errors in polarizability parameterizations. This work shows that the pGM models are capable of accurately modeling polarization effects and have the potential to serve as templates for developing next-generation polarizable force fields for modeling various biological systems.

INTRODUCTION

Development of molecular mechanical force fields has been at the forefront of molecular modeling research due to the critical roles that force fields play in applications such as molecular dynamics (MD) simulations, Monte Carlo (MC) simulations, and protein structure prediction.[1-4] Force fields that have the ability to provide accurate energy calculations and are highly transferable to a wide range of molecular systems have become highly desirable. With graphical processing unit (GPU)-accelerated and specialized high-performance computational platforms,[5,6] it becomes increasingly feasible to conduct simulations at time scales of biological relevance. The extensively used point-charge additive force fields, such as Amber ff19SB,[7] CHARMM,[8] and OPLS,[9] share similar functional forms. In the additive Amber force fields, the following general functional form is used to calculate the potential energies of molecular systems:

$$E_{\text{total}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} + E_{\text{ele}} + E_{\text{vdW}} \quad (1)$$

The first three terms are short-range bonded terms, including the bond stretching terms E_{bond}, the angle bending terms E_{angle}, and the dihedral angle torsion terms E_{dihedral}, with the following formulas:

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$E_{\text{bond}} = \sum_{\text{bonds}} k_{\text{b}}(r - r_0)^2 \tag{2}$$

$$E_{\text{angle}} = \sum_{\text{angles}} k_\theta(\theta - \theta_0)^2 \tag{3}$$

$$E_{\text{dihedral}} = \sum_{\text{dihedrals}} V_{\text{n}}[1 + \cos(n\phi - \gamma)] \tag{4}$$

The last two terms are nonbonded terms between any two atoms i and j. The electrostatic term E_{ele}, usually modeled by the interactions between fixed atom-centered partial charges (Coulomb's law), is a long-range term; whereas the van der Waals term E_{vdW}, modeled by the 6–12 Lennard-Jones potential, is nominally also a long-range term, although it decays rather quickly with increasing distance. E_{ele} and E_{vdW} are formulated as

$$E_{\text{vdW}} = \sum_{i<j} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} \right) \tag{5}$$

$$E_{\text{ele}} = \sum_{i<j} \frac{q_i q_j}{\epsilon R_{ij}} \tag{6}$$

both of which are pairwise and additive. Therefore, in this framework, the interaction between any two atoms is not affected by the presence or absence of other nonbonded atoms.

While additive force fields will continue to play important roles, polarizable force fields are expected to extend our ability to study biomolecular systems more adequately due to their ability to model the atomic polarization effects, which are the redistribution of atomic electron density due to the electric field produced by nearby atoms.[10] Polarization effects are important in biological processes such as ligand–receptor interactions,[11-14] the interactions of ions with nucleic acids,[15,16] the dielectric environmental changes during protein folding,[17,18] and enzymatic mechanisms.[19] If more than two atoms are involved, polarization effects lead to nonadditivity, since when polarized by a third atom, any two atoms interact differently from the situation where the third atom is absent. Lacking proper representation of the polarization effects is considered a major shortcoming of the additive force fields. For over five decades, many attempts have been directed to properly incorporate polarization effects into polarizable force fields. A variety of methods have been explored, including the induced dipole models,[20-28] the fluctuating charge models,[29,30] the Drude oscillator models,[31,32] and the continuum dielectric models.[33,34]

The induced point dipole model is one of the most studied approaches with a long history since the 1970s.[35,36] In this approach, the induced dipole of atom i, subject to the external electric field $\boldsymbol{E}_i$, is

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\boldsymbol{\mu}_i = \alpha_i \left[\boldsymbol{E}_i - \sum_{j \neq i}^{n} \boldsymbol{T}_{ij} \boldsymbol{\mu}_j \right] \tag{7}$$

where α_i is the isotropic polarizability of atom i and $\boldsymbol{T}_{ij}$ is the dipole field tensor with the matrix form

$$\boldsymbol{T}_{ij} = \frac{f_e}{r_{ij}^3} \boldsymbol{I} - \frac{3 f_t}{r_{ij}^5} \begin{bmatrix} x^2 & xy & xz \\ xy & y^2 & yz \\ xz & yz & z^2 \end{bmatrix} \tag{8}$$

where $\boldsymbol{I}$ is the identity matrix; x, y, and z are the Cartesian components along the vector between atoms i and j at distance r_{ij}; and f_e and f_t are distance-dependent damping functions that modify $\boldsymbol{T}_{ij}$ to avoid the so-called "polarization catastrophe" problem, i.e., the phenomenon that induced dipole diverges due to the cooperative induction between induced dipoles at short distances.[10,37] Several damping schemes have been proposed by Thole using a smeared charge distribution $\rho(u)$, where $u = r_{ij} / (\alpha_i \alpha_j)^{1/6}$ is the effective distance.[38,39] Thole's damping schemes have been incorporated into several important polarizable force fields. For example, in the ff12pol force field,[22-25] the linear damping scheme is adopted

$$\rho(u) = \begin{cases} \frac{3}{\pi} \frac{(a-u)}{a^4}, & u < a \\ 0, & u \geq a \end{cases} \tag{9}$$

and the damping functions f_e and f_t have the form

$$\begin{aligned} v &= u / a \\ f_e &= \begin{cases} 4v^3 - 3v^4, & v < 1 \\ 1.0, & v \geq 1 \end{cases} \\ f_t &= \begin{cases} v^4, & v < 1 \\ 1.0, & v \geq 1 \end{cases} \end{aligned} \tag{10}$$

In the Amoeba polarizable force field,[26-28] an exponential damping scheme is used

$$\rho(u) = \frac{3a}{4\pi} \exp(-au^3) \tag{11}$$

and the damping functions f_e and f_t become

$$\begin{aligned} v &= au^3 \\ f_e &= 1 - \exp(-v) \\ f_t &= 1 - (v+1)\exp(-v) \end{aligned} \tag{12}$$

However, since Thole's schemes only screen the interactions between induced dipoles, leaving the polarization due to fixed charges and permanent multipoles unaffected, one caveat is the possibility of producing large atomic induced dipoles when other highly

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

charged species are nearby. About a decade ago, Elking et al. developed a scheme that models atomic electric multipoles using Gaussian electron densities,[40-42] which was originally proposed by Wheatley,[43,44] and this model was later named as the polarizable Gaussian Multipole (pGM) model.[45-47] The pGM model can overcome the potential problem of Thole's scheme by screening all short-range electrostatic interactions in a consistent manner, including the interactions of charge—charge, charge—dipole, charge—quadrupole, dipole—dipole, and so on, eliminating a potential source of singularity in the electrostatic term E_{ele}. Consequently, it has been shown that the pGM model notably improves the prediction of molecular polarizability anisotropy compared with that of Thole models.[45] In the pGM model, the nth order Gaussian multipole at distance r with atom i is defined as

$$\rho^{(n)}(r) = \mathbf{\Theta}^{(n)} \cdot \nabla^{(n)} \left(\frac{\beta_i}{\sqrt{\pi}} \right)^3 \exp(-\beta_i^2 r^2) \tag{13}$$

where $\mathbf{\Theta}^{(n)}$ is the nth rank momentum tensor, $\nabla^{(n)}$ is the nth rank gradient operator, and β_i is the Gaussian exponent controlling the "radius" of the distribution with the following formula:

$$\beta_i = s \left(\frac{2\alpha_i}{3\sqrt{2\pi}} \right)^{-1/3} \tag{14}$$

where α_i is the atomic polarizability and s is a constant screening factor. Although any order of multipoles can be modeled by the pGM model, only charges (zeroth order multipole, eq 15) and dipoles (first order multipole, eq 16) are considered in the current pGM model design

$$\rho^{(0)}(r) = q_i \left(\frac{\beta_i}{\sqrt{\pi}} \right)^3 \exp(-\beta_i^2 r^2) \tag{15}$$

$$\rho^{(1)}(r) = \mathbf{p}_i \cdot \nabla \left(\frac{\beta_i}{\sqrt{\pi}} \right)^3 \exp(-\beta_i^2 r^2) \tag{16}$$

where q_i is the permanent charge and $\boldsymbol{p}_i$ is the permanent dipole of atom i. Replacing $\boldsymbol{p}_i$ in eq 16 with $\boldsymbol{\mu}_i$ in eq 7 will give the pGM distribution of the induced dipole, which has the same form as that of the permanent dipole. For the pGM model, we have the following formula of damping functions f_e and f_t:

$$\begin{aligned} S_{ij} &= \frac{\beta_i \beta_j r_{ij}}{\sqrt{2(\beta_i^2 + \beta_j^2)}} \\ f_e &= \operatorname{erf}(S_{ij}) - \frac{2}{\sqrt{\pi}} S_{ij} \exp(-S_{ij}^2) \\ f_t &= \operatorname{erf}(S_{ij}) - \frac{2}{\sqrt{\pi}} S_{ij} \exp(-S_{ij}^2) \left(1 + \frac{2}{3} S_{ij}^2 \right) \end{aligned} \tag{17}$$

where erf(erf(S_{ij})) is the error function of S_{ij}.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

In a series of recent works, the functional form and parameterization schemes for the pGM model have been designed and implemented. First, a set of isotropic atomic polarizabilities and radii for the pGM model were obtained by fitting to molecular polarizability tensors of 1405 molecules or dimers calculated at the B3LYP/aug-cc-pVTZ//B3LYP/aug-cc-pVTZ level of theory using an optimization method based on the genetic algorithm (GA).[45] Second, a local frame for the pGM permanent dipoles formed by covalent basis vectors (CBVs), which are unit vectors along the direction of covalent bonds or virtual bonds, has been proposed based on the observation that atomic permanent dipole moments mainly result from covalent bonding interactions.[46] Third, the analytical formula of the electrostatic term of the pGM models has been derived,[46] which is the sum of a permanent electrostatic term $E_{\text{ele-perm}}$ and an induced electrostatic term $E_{\text{ele-ind}}$

$$E_{\text{ele}} = E_{\text{ele-perm}} + E_{\text{ele-ind}} \tag{18}$$

with the following formula

$$E_{\text{ele-perm}} = \sum_{i<j} (q_i + \boldsymbol{p}_i \cdot \nabla_i)(q_j + \boldsymbol{p}_j \cdot \nabla_j)\frac{\text{erf}(S_{ij})}{r_{ij}} \tag{19}$$

$$E_{\text{ele-ind}} = \sum_{i<j} \boldsymbol{\mu}_i (q_j + \boldsymbol{p}_j \cdot \nabla_j)\nabla_i \frac{\text{erf}(S_{ij})}{r_{ij}} \tag{20}$$

Therefore, in the functional form of the pGM models, the electrostatic term in eq 6 is replaced by eqs 18-20, and the rest of the terms remain unchanged (eqs 2-5). In addition, the pGM electrostatic term has been interfaced with the particle mesh Ewald (PME) method for molecular simulations under periodic boundary conditions.[46,48-51] Fourth, the pGM internal stress tensor expression for constant pressure MD simulations of both the flexible and rigid-body molecular system has been derived.[47] Finally, the *PyRESP* program enabling parameterizations for the pGM models with and without atomic permanent dipoles by reproducing quantum mechanical (QM) electrostatic potential (ESP) around molecules has been implemented.[52] All of the components mentioned above, including the pGM polarizabilities and radii, the *sander* program enabling MD simulations for the pGM models, and the *PyRESP* parametrization program, are available in the AmberTools22 program suite that can be downloaded from http://ambermd.org/.[53]

In this work, we assessed the ability of the pGM models to reproduce QM many-body interaction energies in peptide oligomers, specifically the influences of neighboring peptides upon a pair of interacting peptide monomers. For polarizable force fields, the many-body interaction energies can be decomposed into nonadditive and additive contributions. The detailed definitions of the many-body interaction energy, as well as its nonadditive and additive contributions, will be presented in the Theory section. Glycine dipeptide oligomers arranged in three main-chain hydrogen-bonding conformations were used as the model peptide systems because glycine has the minimalist side chain so that we can focus on main-chain hydrogen-bonding interactions. Two types of pGM models were considered, including pGM-perm, in which the atomic dipoles are represented by a combination of both

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

induced and permanent dipoles, and pGM-ind, in which the atomic dipoles are represented by the induced dipoles only. We compared the performances of the pGM-perm and pGM-ind models with several other widely used force fields in terms of reproducing QM interaction energies and many-body interaction energies, including four Amber force fields: ff12pol,[22-25] ff19SB,[7] ff15ipq,[54] and ff03,[55] as well as the 2013 version of the Amoeba protein force field (Amoeba13).[28] Among the seven force fields tested, pGM-perm, pGM-ind, Amoeba13, and ff12pol are polarizable force fields, while ff19SB, ff15ipq, and ff03 are classical point-charge additive force fields. The results show that the pGM models perform significantly better than all other force fields in terms of reproducing QM interaction energies, many-body interaction energies, and the nonadditive and additive contributions to the many-body interactions. In addition, we tested the robustness of the pGM models against parameterization errors by employing alternative atomic polarizabilities, including the pGM polarizabilities scaled by a factor of 0.9,[45] the Amoeba polarizabilities,[26] and the ff12pol polarizabilities.[22] The results show that the pGM models are highly robust and perform well even with those “wrong” polarizabilities.

THEORY

In this work, each oligomer is arranged in a general form of m glycine dipeptides interacting with n glycine dipeptides, named Gly_m:Gly_n, where m and n are ranged from 1 to 3. Each “Gly” in this work represents a glycine dipeptide (ACE-GLY-NME) capped with an *N*-acetyl (ACE) group at the N-terminal and an *N*-methylamide (NME) group at the C-terminal. For example, Figure 1A shows the Gly_2:Gly_2 oligomer.

The interaction energy $\mathrm{IE}(\mathrm{Gly}_m{:}\mathrm{Gly}_n)$ between Gly_m and Gly_n of the Gly_m:Gly_n oligomer can be calculated by the following equation:

$$\mathrm{IE}(\mathrm{Gly}_m{:}\,\mathrm{Gly}_n) = E(\mathrm{Gly}_m{:}\,\mathrm{Gly}_n) - E(\mathrm{Gly}_m) - E(\mathrm{Gly}_n) \tag{21}$$

where $E(\mathrm{Gly}_m{:}\mathrm{Gly}_n)$ is the potential energy of the entire Gly_m:Gly_n oligomer and $E(\mathrm{Gly}_m)$, $E(\mathrm{Gly}_n)$ are the potential energies of isolated Gly_m and Gly_n, respectively.

More importantly, we intend to study the many-body effects in the Gly_m:Gly_n oligomer, specifically, the influence of the neighboring glycine dipeptides Gly_{m-1} and Gly_{n-1} upon the interaction between the two middle glycine dipeptides Gly:Gly in the Gly_m:Gly_n oligomer. Here, we define the many-body interaction energy $\mathrm{ME}(\mathrm{Gly}_m{:}\mathrm{Gly}_n)$ as the difference between $\mathrm{IE}(\mathrm{Gly}_m{:}\mathrm{Gly}_n)$ and IE(Gly:Gly). That is,

$$\mathrm{ME}(\mathrm{Gly}_m{:}\,\mathrm{Gly}_n) = \mathrm{IE}(\mathrm{Gly}_m{:}\,\mathrm{Gly}_n) - \mathrm{IE}(\mathrm{Gly}{:}\mathrm{Gly}) \tag{22}$$

Taking the Gly_2:Gly_2 oligomer in Figure 1A as an example, the difference between $\mathrm{IE}(\mathrm{Gly}_2{:}\mathrm{Gly}_2)$ and IE(Gly:Gly) of the two middle peptides (displayed in brown) is the many-body interaction energy $\mathrm{ME}(\mathrm{Gly}_2{:}\mathrm{Gly}_2)$ caused by the presence of the two neighboring peptides (displayed in cyan).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The many-body interaction energy $ME(Gly_m:Gly_n)$ can be decomposed into the nonadditive contribution $ME_{NA}(Gly_m:Gly_n)$ and the additive contribution $ME_A(Gly_m:Gly_n)$. Before showing their formulas, we first need to define the interaction energies of the two middle peptides $IE_{mid}(Gly_m:Gly_n)$ in the presence of the neighboring peptides Gly_{m-1} and Gly_{n-1}.

$$\begin{aligned} IE_{mid}(Gly_m:Gly_n) = IE(Gly_m:Gly_n) - IE(Gly_m:XGly_{n-1}) \\ - IE(Gly_{m-1}X:Gly_n) \\ + IE(Gly_{m-1}X:XGly_{n-1}) \end{aligned} \quad (23)$$

where X indicates the absence of either one of the two middle peptides. $IE(Gly_m:XGly_{n-1})$ is the interaction energy between Gly_m and the neighboring peptides Gly_{n-1}; $IE(Gly_{m-1}X:Gly_n)$ is the interaction energy between the neighboring peptides Gly_{m-1} and Gly_n; and $IE(Gly_{m-1}X:XGly_{n-1})$ is the interaction energy between the neighboring peptides Gly_{m-1} and Gly_{n-1} on both sides. Then, we have the formulas of $ME_{NA}(Gly_m:Gly_n)$ and $ME_A(Gly_m:Gly_n)$, the proof of which can be found in the Appendix.

$$ME_{NA}(Gly_m:Gly_n) = IE_{mid}(Gly_m:Gly_n) - IE(Gly:Gly) \quad (24)$$

$$ME_A(Gly_m:Gly_n) = ME(Gly_m:Gly_n) - ME_{NA}(Gly_m:Gly_n) \quad (25)$$

For the $Gly_2:Gly_2$ oligomer example, Figure 1B shows the oligomer $Gly_2:XGly$ with the interaction energy $IE(Gly_2:XGly)$; Figure 1C shows the oligomer $GlyX:Gly_2$ with the interaction energy $IE(GlyX:Gly_2)$; and Figure 1D shows the oligomer GlyX:XGly with the interaction energy IE(GlyX:XGly). The interaction energy of the two middle peptides in the presence of the neighboring peptides is $IE_{mid}(Gly_2:Gly_2) = IE(Gly_2:Gly_2) - IE(Gly_2:XGly) - IE(GlyX:Gly_2) + IE(GlyX:XGly)$. The nonadditive and additive contributions to the many-body interaction energy $ME(Gly_2:Gly_2)$ are $ME_{NA}(Gly_2:Gly_2) = IE_{mid}(Gly_2:Gly_2) - IE(Gly:Gly)$ and $ME_A(Gly_2:Gly_2) = ME(Gly_2:Gly_2) - ME_{NA}(Gly_2:Gly_2)$, respectively.

For additive force fields, $ME_{NA}(Gly_m:Gly_n)$ is guaranteed to be zero, so that $ME(Gly_m:Gly_n)$ is equivalent to $ME_A(Gly_m:Gly_n)$, while for polarizable force fields, $ME_{NA}(Gly_m:Gly_n)$ is nonzero, so that $ME(Gly_m:Gly_n)$ has both additive and nonadditive contributions.

COMPUTATIONAL DETAILS

Geometry Preparations.

The formamide dimer and three glycine dipeptide dimers were used to select density functional theory (DFT) methods for subsequent QM energy calculations. The formamide dimer was first arranged into hydrogen-bonding conformation, and the geometry was optimized at the B3LYP/6-311++G(d,p) level of theory. A total of 15 glycine dipeptide oligomers were constructed. First, three glycine dipeptide dimers were configured and arranged into α-helix, anti-parallel β-sheet, and parallel β-sheet hydrogen-bonding

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

conformations observed in proteins. Then, the geometries were optimized at the B3LYP/6-311++G(d,p) level of theory with the main-chain torsion angles fixed at (ϕ, ψ) = (−57, −47°), (−140, 135°) and (−119, 113°), corresponding to the α-helix, anti-parallel β-sheet, and parallel β-sheet conformations, respectively. Higher-order oligomers were constructed from these three optimized dimers by rigid-body translations and rotations. For example, to produce a Gly_3:Gly_3 dipeptide hexamer (in the conformation of an interacting pair of trimers) while maintaining the central dimer in the optimized conformation, both dipeptides of the Gly:Gly dimer are rotated and moved toward both sides along the hydrogen bond direction. The structures of formamide dimer and glycine dipeptide oligomers are presented in Figures S1-S6, and the detailed coordinates of glycine dipeptide oligomers are available in the Supporting Information.

Quantum Mechanical Calculations.

Three DFT methods were tested to calculate the QM interaction energies of the formamide dimer and the glycine dipeptide dimers, including ωB97X − D,[56] M062X,[57] and B3LYP,[58,59] all with the aug-cc-pVTZ (aTZ) basis set. The basis set superposition errors (BSSEs) were corrected through the counterpoise corrections.[60] To select the most suitable DFT method for our systems, the CCSD(T)/CBS interaction energies $IE_{CCSD(T)/CBS}$ were calculated as the reference energies using Helgaker's extrapolation method.[61,62] First, the HF and MP2 interaction energies were calculated with aug-cc-pVTZ (aTZ) and aug-cc-pVQZ (aQZ) basis sets, and the correlation (CORR) energies IE_{CORR} were defined as the difference between the MP2 and HF energies $IE_{CORR} = IE_{MP2} - IE_{HF}$. Next, $IE_{HF/CBS}$ and $IE_{CORR/CBS}$ were calculated using the following equations:

$$IE_{HF/CBS} = \frac{IE_{HF/aTZ}\exp(-1.63\times 4) - IE_{HF/aQZ}\exp(-1.63\times 3)}{\exp(-1.63\times 4) - \exp(-1.63\times 3)} \tag{26}$$

$$IE_{CORR/CBS} = \frac{IE_{CORR/aTZ}\times 3^3 - IE_{CORR/aQZ}\times 4^3}{3^3 - 4^3} \tag{27}$$

and $IE_{MP2/CBS}$ can be calculated as

$$IE_{MP2/CBS} = IE_{HF/CBS} + IE_{CORR/CBS} \tag{28}$$

Note that the average of $IE_{MP2/CBS}$ with and without counterpoise correction was used as the final $IE_{MP2/CBS}$. Finally, $IE_{CCSD(T)/CBS}$ were calculated by adding a CCSD(T) correction calculated at a small basis set to the averaged $IE_{MP2/CBS}$

$$IE_{CCSD(T)/CBS} = IE_{MP2/CBS} + (IE_{CCSD(T)} - IE_{MP2})_{\text{small basis set}} \tag{29}$$

For formamide dimers, aug-cc-pVTZ was used as the small basis set; for glycine dipeptide dimers, cc-pVTZ was used as the small basis set.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Following the strategy that has been successfully used in Amber force field development in which the partial charges were fit to QM electrostatic potentials (ESPs), the QM ESPs were calculated at the MP2/aug-cc-pVTZ level of theory for a set of points in the solvent-accessible region around each glycine dipeptide molecule in the α-helix, anti-parallel β-sheet, and parallel β-sheet conformations. The points were generated using the method developed by Singh et al. on molecular surfaces (with a density of 6 points/Å^2) at each of 1.4, 1.6, 1.8, and 2.0 times the van der Waals radii.[63,64] All QM calculations were performed using the Gaussian 16 software.[65]

pGM Parameterizations.

To assess the robustness of the pGM models against errors in polarizability parameterization, four sets of atomic polarizabilities were employed to parameterize the pGM models, including the pGM polarizabilities,[45] the pGM polarizabilities scaled by a factor of 0.9, the Amoeba polarizabilities,[26] and the ff12pol polarizabilities,[22] for a combined total of two pGM models and six variants. The recently developed *PyRESP* program was used to parameterize the point charges and permanent point dipoles of the glycine dipeptide molecule for the pGM-perm and pGM-ind models, and a two-stage parameterization procedure was adopted.[52] In the first stage, all charges and permanent dipoles were set free to change, and a weak restraining strength of 0.0005 was applied. In the second stage, intra-molecular equivalencing was enforced on all charges and permanent dipoles that share an identical chemical environment with others, such as those of methyl and methylene hydrogens. A stronger restraining strength of 0.001 was applied, and all other fitting centers were set frozen to keep the values obtained from the first stage. In both stages, the restraints were only applied to non-hydrogen-heavy atoms. Only the total charge constraint was enforced in the parameterization process, and no additional intra-molecular charge constraint was applied. Inter-molecular equivalencing was enforced in both the first and the second stages for the three conformations of glycine dipeptides. For the parameterizations of both the pGM-perm and pGM-ind models, both 1–2 and 1–3 polarization interactions were included for reasons elucidated before.[45,66]

The parameters of bonded terms (bond stretching terms, angle bending terms, dihedral angle torsion terms) and the van der Waals terms for both the pGM-perm and pGM-ind models were obtained from the ff12pol force field without any change.[25]

Molecular Mechanics Calculations.

Seven force fields were explored for calculating the molecular mechanics energies of glycine dipeptide oligomers, including four polarizable force fields: pGM-perm, pGM-ind, ff12pol,[22-25] and Amoeba13,[28] and three additive force fields: ff19SB,[7] ff15ipq,[54] and ff03.[55] The parameter and topology files for the Amber force fields (pGM-perm, pGM-ind, ff12pol, ff19SB, ff15ipq, and ff03) were generated using the *tleap* program from the AmberTools22 program suite.[53] The coordinate files for the Amber force fields and the .xyz files for the Amoeba13 force field were generated from the geometries optimized by Gaussian 16 software.[65] The single-point energies of the Amber force fields were calculated by the *sander* program with extensions to accommodate the pGM models.[46,53] The *dynamic* program from the Tinker 8.6.1 software package was used to calculate the single-point

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

energies of the Amoeba13 force field.[67] All nonbonded interactions were calculated in gas phase without distance cutoff.

The performance of each force field in each energy calculation task was evaluated by the root-mean-squared error (RMSE) and mean absolute error (MAE) given by

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (E_i^{\mathrm{QM}} - E_i)^2}{N}} \tag{30}$$

$$\mathrm{MAE} = \frac{\sum_{i=1}^{N} | E_i^{\mathrm{QM}} - E_i |}{N} \tag{31}$$

where E_i^{QM} is the energy given by QM calculations and E_i is the energy calculated by molecular mechanics force fields.

RESULTS AND DISCUSSION

ωB97X – D without Counterpoise Correction Most Accurately Reproduces CCSD(T)/CBS Interaction Energies.

There have been numerous works documenting the performances of various DFT methods in their ability to model the dispersion effect. Among these DFT methods, ωB97X – D[56] and M062X[57] exhibit great trade-offs between computation speed and accuracy.[68] One observation is that the accuracy of DFT methods depends on the particular molecular systems being studied. To determine which one of these DFT methods is the most suitable to our systems, we compared the interaction energies of the formamide dimer and the glycine dipeptide dimers obtained from three DFT methods, including ωB97X – D,[56] M062X,[57] and B3LYP,[58,59] with those calculated at the CCSD(T)/CBS level of theory, which have been considered as the “gold standard” of computational chemistry. Table 1 shows the interaction energies calculated with these DFT methods with and without counterpoise BSSE corrections. We can see that the interaction energies by ωB97X – D without counterpoise correction are the closet to the CCSD(T)/CBS results (eq 29), with an RMSE of 0.17 kcal/mol and an MAE of 0.12 kcal/mol. Not surprisingly, B3LYP interaction energies consistently exhibit the highest deviations from the CCSD(T)/CBS results, since B3LYP lacks proper consideration of dispersion contributions. Without counterpoise corrections, the RMSEs of M062X and B3LYP are 0.32 and 3.37 kcal/mol, respectively, and the MAEs are 0.30 and 3.12 kcal/mol, respectively. With counterpoise corrections, the RMSEs of ωB97X – D, M062X, and B3LYP are 0.31, 0.71, and 3.70 kcal/mol, respectively, and the MAEs are 0.29, 0.67, and 3.42 kcal/mol, respectively. Therefore, for the formamide dimer and the glycine dipeptide dimers, ωB97X – D without counterpoise BSSE correction best reproduces the CCSD(T)/CBS interaction energies. For this reason, ωB97X – D without counterpoise correction was chosen as the QM reference method to evaluate various molecular mechanical force fields in the following discussions.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

pGM Models Show the Best Performances in Interaction Energy Calculations.

Listed in Table 2 are the interaction energies calculated at the ωB97X – D/aug-cc-pVTZ level of theory without counterpoise corrections and seven molecular mechanical force fields. The interaction energies IE(Gly_m:Gly_n) between Gly_m and Gly_n of the Gly_m:Gly_n oligomers were calculated following eq 21 without consideration of the deformation energies to avoid complications in energy calculations that may arise from structural changes. The interaction energies calculated by the ωB97X – D method exhibit an increasing trend in the order of Gly:Gly, Gly:Gly_2, Gly:Gly_3, Gly_2:Gly_2, and Gly_3:Gly_3 for all three conformations. It is notable that IE(Gly_3:Gly_3) is 11.00 kcal/mol stronger than IE(Gly:Gly) in the α-helix conformation. In comparison, for the anti-parallel β-sheet and parallel β-sheet conformations, IE(Gly_3:Gly_3) are only 3.50 and 4.11 kcal/mol stronger than IE(Gly:Gly), respectively. The larger difference of the α-helix conformation is attributable to the strong polarization effect caused by the alignments of the main-chain peptide hydrogen bonds. Another observation is that although Gly:Gly_3 and Gly_2:Gly_2 are both tetramers, Gly_2:Gly_2 consistently shows stronger interaction energies than Gly:Gly_3 in all three conformations. This shows that the inner parts of the peptide secondary structures are expected to have stronger main-chain hydrogen bonding than the outer parts.

The interaction energies calculated by the pGM-perm and pGM-ind models stand out as the closest to the DFT results, with RMSEs of 1.35 and 1.37 kcal/mol, respectively. The similarity between the performances of the pGM-perm and pGM-ind models indicates that, with the pGM damping schemes, the induced dipoles are sufficient for calculating the interaction energies of glycine dipeptides. The next best performance is given by the ff15ipq force field with an RMSE of 1.87 kcal/mol, which is an additive force field whose charges were fit to the ESP of peptides in the presence of explicit solvent water.[54] The polarizable force field ff12pol, the additive force field ff19SB, and the polarizable force field Amoeba13 are ranked fourth to sixth, with RMSE of 2.28, 2.67, and 2.91 kcal/mol, respectively. The observation that Amoeba13 performs worse than ff12pol in this test set is somewhat surprising, given that Amoeba13 is such an elaborate force field that includes atomic permanent dipoles and quadrupoles, in addition to the polarizable induced dipoles.[26,28] In contrast, the ff12pol force field is a minimalist polarizable induced dipole force field with neither permanent dipoles nor quadrupoles.[22-25]

Another interesting observation is that, compared with the ωB97X – D results, the pGM-perm and pGM-ind models systematically overestimate the interaction energies IE(Gly_m:Gly_n). On average, pGM-perm and pGM-ind overestimate the interaction energies by 1.32 and 1.34 kcal/mol, respectively. Also interesting is the consistency of the deviations between the interaction energies of ωB97X – D and the pGM models across different conformations. Since the main-chain hydrogen bonds contribute to peptide secondary structure formations, the balance across different conformations can influence the peptide secondary structure preference and the capability of modeling the relative strength of different hydrogen-bonding systems and in peptide main-chain secondary structures. In this regard, both pGM models show good balance and their differences in the interaction energies are about the same magnitude across the three conformations, with pGM-perm

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

exhibiting slightly better consistency than pGM-ind. For example, the Gly:Gly interaction energies are overestimated by 1.48–1.88 kcal/mol by pGM-perm and overestimated by 1.36–1.92 kcal/mol by pGM-ind. In contrast, all other force fields show nonuniform deviations across different conformations. Taking ff12pol as an example, the largest deviation of interaction energies among the three conformations consistently comes from the oligomers in the α-helix conformation. For Amoeba13, the deviations of α-helix and parallel β-sheet conformers are comparable, whereas the deviations of anti-parallel β-sheet conformer are notably smaller. Since uniform deviations across different conformers naturally avoid introducing conformational bias, it is much more preferable than nonuniform deviations. Furthermore, when taking all Gly_m:Gly_n oligomers into account, the deviations of interaction energies range between −0.96 and −1.88 kcal/mol for pGM-perm, and between −0.90 and −1.92 kcal/mol for pGM-ind, which are more consistent than other force fields. On the other hand, all other force fields (except ff15ipq) exhibit a tendency of growing deviations with increasing size of oligomers, suggesting that they underestimate the many-body interactions when there are multiple peptides in the oligomers. Therefore, it is encouraging that the pGM models outperform all other five force fields in terms of interaction energy calculations across oligomers with different conformations and with different sizes.

pGM Models Most Accurately Reproduce QM Many-Body Interaction Energies.

Subtracting corresponding rows of the Gly:Gly dimers from other rows in Table 2 gives Table 3, which lists the many-body interaction energies ME(Gly_m:Gly_n) calculated by the seven force fields compared with those calculated by the ωB97X – D method. The many-body interaction energies defined in eq 22 describe the overall (additive and nonadditive) contributions from the neighboring glycine dipeptides (Gly_{m-1} and Gly_{n-1}, excluding the Gly:Gly at the interface) to the dimerization energy at the interface of Gly_m:Gly_n. As shown in the ωB97X – D results, the many-body interaction energies again increase in the order of Gly:Gly_2, Gly:Gly_3, Gly_2:Gly_2, and Gly_3:Gly_3 for all three conformations. Comparing oligomers in the anti-parallel and parallel β-sheet conformations, we can see that the addition of outer peptides does not significantly increase the many-body interactions. Thus, the cross-strand effects in the β-sheet conformations are mainly limited to those in close contact and diminish rather quickly with distance. For the α-helix conformation, notably, a stronger many-body interaction is observed because the hydrogen bonds are aligned in the same directions. An interesting observation is that the many-body interaction energy of tetramer Gly:Gly_3 is only marginally stronger than that of trimer Gly:Gly_2 by 1.16 kcal/mol, which is a much smaller increase compared to the 4.21 kcal/mol increase of tetramer Gly_2:Gly_2. By adding one more peptide at each side, the many-body interaction energy of hexamer Gly_3:Gly_3 becomes stronger by 3.62 kcal/mol than that of tetramer Gly_2:Gly_2. Therefore, for the α-helix conformation, in contrast to the marginal effect of adding peptides to one side of the interface, symmetric addition makes the interaction at the interface significantly stronger, so that a much stronger many-body interaction is expected in the inner part of α-helices. Therefore, the outer and inner parts of α-helices could have considerably different stabilities. This effect could be significant in nonpolar environments such as transmembrane proteins.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Among the seven force fields tested, pGM-perm and pGM-ind again show the lowest RMSEs (0.40 and 0.38 kcal/mol, respectively) and the lowest MAEs (0.37 and 0.35 kcal/mol, respectively), making them the best force fields in terms of many-body interaction energy calculations. It is encouraging that the RMSEs and MAEs of both pGM models are lower than the thermal fluctuation energy of 0.60 kcal/mol at 300 K. Similar to the case of interaction energy calculations, both pGM models give similar performances in terms of many-body interaction energy calculations. This indicates a potential advantage of the ESP fitting strategy employed for pGM parameterizations. Both the interaction energies and many-body interaction energies between molecules are largely dependent on the electrostatic interactions, and the ESP surrounding molecules are one of the most important electrostatic properties. Since both the pGM-ind and pGM-perm models are able to reproduce QM ESPs with low errors,[52] it is expected that both models can accurately reproduce QM interaction energies and many-body interaction energies, therefore giving similar performances.

The next best-performing force field is the Amoeba13 force field, which exhibits the largest improvement compared with interaction energy calculations in Table 2, with the RMSE reduced from 2.91 to 1.16 kcal/mol. The significant improvement of Amoeba13 shows that the short-range interactions are the main cause of the large errors observed in the interaction energies. The ff12pol force field is ranked the third best-performing force field, with an RMSE of 1.62 kcal/mol. It is remarkable that all polarizable force fields perform better than all additive force fields. In fact, all additive force fields notably underestimate the many-body interactions by more than 2.00 kcal/mol. Among the additive force fields tested, ff15ipq once again shows the best performance, with an RMSE of 2.04 kcal/mol. However, this RMSE is slightly higher than that of interaction energies (1.87 kcal/mol) given by ff15ipq.

Another observation in Table 3 is that all force fields consistently underestimate the many-body interaction energies $ME(Gly_m:Gly_n)$ when compared with the ωB97X – D results. For the pGM models, this is in sharp contrast to the systematic overestimations in the interaction energies $IE(Gly_m:Gly_n)$, as shown in Table 2, indicating that the long-range terms are still under-represented in the pGM models. Moreover, this suggests that the overestimations of $IE(Gly_m:Gly_n)$ of the pGM models are primarily due to the short-range van der Waals terms. Because of this, we anticipate that the present van der Waals parameters, which were taken directly from ff12pol without optimization, need to be tuned to make the short-range terms less attractive. For Amoeba13 and ff12pol, however, since their $IE(Gly_m:Gly_n)$ and $ME(Gly_m:Gly_n)$ are systematically weaker than the ωB97X – D results and the errors in $IE(Gly_m:Gly_n)$ are larger than $ME(Gly_m:Gly_n)$, it appears that these two force fields could be improved by strengthening both their short-range and polarization terms.

Similar to the case of interaction energies, the underestimations of the pGM models compared to ωB97X – D across different conformations and across different oligomers are consistent, which range between 0.17 and 0.61 kcal/mol for pGM-perm and between 0.15 and 0.59 kcal/mol for pGM-ind. In contrast, all other force fields once again show nonuniform deviations across different conformations, and the deviations increase with the size of oligomers. It is notable that the additive ff15ipq force field, which exhibits relatively

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

consistent deviations in terms of interaction energies, also shows gradually increasing deviations with the oligomer size in terms of many-body interaction energies. Based on the above observations, we conclude that the polarization effects play critical roles in the many-body interactions, and the additive force fields are, in general, incapable of modeling them accurately.

pGM Models Perform the Best in Reproducing QM Nonadditive and Additive Contributions to the Many-Body Interactions.

The many-body interaction energies in Table 3 depend on the nonbonded terms in the functional form of each molecular mechanical force field. For additive force fields such as Amber ff19SB, ff15ipq, and ff03, the many-body interactions only have contributions from the additive electrostatic and van der Waals terms. In polarizable force fields ff12pol and pGM-ind, the nonadditive induced dipole polarization energy is also involved. The pGM-perm model has additional energy contributions from atomic permanent dipoles, and the Amoeba13 force field also has contributions from atomic permanent quadrupoles, which are both additive terms. It is difficult to decipher which of these terms plays a more important role if we just look at the total many-body interaction energies shown in Table 3. Therefore, we decompose the many-body interaction energies into nonadditive and additive contributions to gain insight into these force fields. For each Gly_m:Gly_n oligomer, the formulas of the nonadditive contributions $ME_{NA}(Gly_m:Gly_n)$ and the additive contributions $ME_A(Gly_m:Gly_n)$ are given in eqs 24 and 25, respectively. The proof of the decomposition is shown in the Appendix. Note that the functional forms of additive force fields only have additive terms, so that the nonadditive contribution $ME_{NA}(Gly_m:Gly_n)$ of any additive force field is guaranteed to be zero. For this reason, we will only compare the performances of polarizable force fields pGM-perm, pGM-ind, Amoeba13, and ff12pol in this section.

The interaction energies of the two peptides at the interface, $IE_{mid}(Gly_m:Gly_n)$, in the presence of the neighboring peptides Gly_{m-1} and Gly_{n-1} defined in eq 23, calculated by the four polarizable force fields and by ωB97X – D, are shown in Table S1. For additive force fields, the values will be identical to those of IE(Gly:Gly) in the absence of neighboring peptides Gly_{m-1} and Gly_{n-1}, as shown in Table 2, if listed. For polarizable force fields, a trend similar to that in Table 2 is observed. First, the pGM-perm and pGM-ind models outperform the other two polarizable force fields, with RMSEs of 1.39 and 1.43 kcal/mol, respectively. The RMSEs of the Amoeba13 force field and the ff12pol force field are 2.54 and 1.69 kcal/mol, respectively. Second, compared with the ωB97X – D results, the pGM-perm and pGM-ind models systematically overestimate the interaction energies (by 1.37 and 1.41 kcal/mol, respectively), whereas both Amoeba13 and ff12pol underestimate (by 2.31 and 1.18 kcal/mol, respectively). Third, the deviations between the interaction energies given by ωB97X – D and the pGM models across different conformations and different oligomers are highly consistent. For pGM-perm, the largest spread (0.41 kcal/mol) comes from $Gly:Gly_3$ between the anti-parallel β-sheet (−1.25 kcal/mol) and parallel β-sheet (−1.66 kcal/mol) conformers. For pGM-ind, the largest spread (0.56 kcal/mol) is between the α-helix (−1.36 kcal/mol) and parallel β-sheet (−1.92 kcal/mol) conformers of Gly:Gly. Overall, the deviations range between −1.02 and −1.88 kcal/mol for pGM-perm, and between −1.15 and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

−1.92 kcal/mol for pGM-ind, for all five oligomers and all three conformations. However, Amoeba13 and ff12pol show nonuniform deviations across different conformations, and these deviations tend to increase with the size of oligomers.

Table 4 shows the nonadditive contributions $ME_{NA}(Gly_m:Gly_n)$ of ωB97X – D and the four polarizable force fields obtained by subtracting corresponding rows of the Gly:Gly dimers from other rows in Table S1, and Table 5 shows the additive contributions $ME_{A}(Gly_m:Gly_n)$ obtained by subtracting the corresponding nonadditive contributions $ME_{NA}(Gly_m:Gly_n)$ from the many-body interaction energies $ME(Gly_m:Gly_n)$ in Table 3. For nonadditive contributions $ME_{NA}(Gly_m:Gly_n)$, pGM-ind produces the lowest RMSE (0.30 kcal/mol) among all four polarizable force fields, and pGM-perm is the second best, with an RMSE of 0.33 kcal/mol. However, for additive contributions $ME_{A}(Gly_m:Gly_n)$, pGM-perm performs the best with an RMSE of 0.09 kcal/mol, and pGM-ind is the second-best, with an RMSE of 0.10 kcal/mol. The significantly lower RMSEs of the additive contributions of the pGM models than other polarizable force fields show the robustness of the ESP fitting scheme of *PyRESP* since the additive contribution is mainly due to the interactions involving fixed point charges and permanent dipoles.[52] For both nonadditive and additive contributions, the ff12pol force field gives the worst performance, with RMSEs of 0.95 and 0.72 kcal/mol for the nonadditive and additive contributions, respectively. Amoeba13 yields RMSEs of 0.59 kcal/mol for both the nonadditive and additive contributions, which are better than ff12pol in terms of both contributions, but still notably worse than the pGM models.

Interestingly, compared with the ωB97X – D results, all four polarizable force fields underestimate both the nonadditive and additive contributions to varying degrees. As measured by MAE, the nonadditive contributions are underestimated by 0.30, 0.27, 0.54, and 0.85 kcal/mol by pGM-perm, pGM-ind, Amoeba13, and ff12pol, respectively, and the additive contributions are underestimated by 0.06, 0.08, 0.42, and 0.46 kcal/mol by pGM-perm, pGM-ind, Amoeba13, and ff12pol, respectively. Therefore, MAEs show the same performance trend as RMSEs, where pGM-ind performs the best for calculating the nonadditive contributions, and pGM-perm performs the best for calculating the additive contributions. The consistent underestimations could potentially come from two sources: inadequate short-range damping and smaller-than-needed polarizabilities, both of which may be improved by further parameterizations. Encouragingly, the pGM models again exhibit consistent deviations from the ωB97X – D results for both contributions across different conformations and different sized oligomers. For the pGM-perm model, the deviations range between 0.12 and 0.52 kcal/mol for the nonadditive contributions, and between 0.01 and 0.22 kcal/mol for the additive contributions; For the pGM-ind model, the deviations range between 0.09 and 0.49 kcal/mol for the nonadditive contributions, and between 0.02 and 0.25 kcal/mol for the additive contributions. In contrast, Amoeba13 and ff12pol show significant variations in the deviations from the ωB97X – D results for both nonadditive and additive contributions across different conformations. Overall, Tables 4 and 5 show that compared with the other two polarizable force fields tested here, the pGM models (with or without permanent atomic dipoles) perform the best in terms of reproducing QM nonadditive and additive contributions to the many-body interaction energies.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

pGM Models Are Robust with Altered Atomic Polarizabilities.

Because of the inherent approximations to either experimental observations or QM calculations, all mechanical force fields are subject to errors that can come from both the functional forms and parameterization processes. In the development of Amber force fields, a consistent electrostatic parameterization approach is to fit the QM calculated ESPs of small molecules or fragments of large molecules to obtain atomic charges and multipoles. Technically, this approach is rather straightforward and allows the development of consistent parameters across a wide variety of chemistry. In the cases of polarizable force fields, another advantage is that the errors in the initial fitting of polarizabilities can be partially compensated at the stage when charges and permanent multipoles are calculated, yielding a more robust force field. This feature is potentially advantageous because of the nonlinear nature of the polarization energy.

In our previous work, the pGM atomic polarizabilities and radii were obtained by fitting QM molecular polarizability tensors of 1405 molecules or dimers.[45] In this work, we further evaluated the robustness of the pGM models by re-parameterizing the glycine dipeptide charges and permanent dipoles using the recently developed *PyRESP* program,[52] with alternative polarizabilities, including the pGM polarizabilities scaled by a factor of 0.9, the Amoeba13 polarizabilities,[26] and the Amber ff12pol polarizabilities.[22] These alternative polarizability sets are either scaled or taken from different sources and have been developed for different polarization schemes. Therefore, we expect that the energies related to polarization calculated with these three polarizability sets be less accurate than those produced by the pGM models with original pGM polarizabilities shown in Tables 2-5. Our objective is to see whether these “wrong” polarizabilities would lead to intolerable errors in energy calculations.

The interaction energies IE(Gly_m:Gly_n) as well as the many-body interaction energies ME(Gly_m:Gly_n) of each Gly_m:Gly_n oligomer of the pGM-perm and pGM-ind models calculated with the alternative polarizabilities are shown in Tables 6 and S2, respectively. Interestingly, for both pGM models, the “wrong” polarizabilities produce interaction energies IE(Gly_m:Gly_n) with lower RMSEs compared with those obtained by the “correct” pGM polarizabilities shown in Table 2. For the pGM-perm model, the overall RMSE of interaction energies decreased from 1.35 kcal/mol to 0.80, 0.82, and 0.62 kcal/mol for the scaled pGM, Amoeba13, and ff12pol polarizabilities, respectively. For the pGM-ind model, the overall RMSE of interaction energies decreased from 1.37 kcal/mol to 0.81, 0.97, and 0.57 kcal/mol for the scaled pGM, Amoeba13, and ff12pol polarizabilities, respectively. The higher RMSEs associated with the “correct” pGM polarizabilities compared with that of the “wrong” polarizabilities might be explained by the fact that the van der Waals parameters for the pGM models were taken directly from the ff12pol force field without any optimization. As shown in Table 2, with the original polarizabilities, both pGM-perm and pGM-ind overestimate the interaction energies. Since the interaction energies can be decomposed to the electrostatic and van der Waals contributions, the overestimation in the interaction energies can be explained by the overestimation of the van der Waals term in the current pGM models. Specifically, the dispersion effect of the van der Waals term might be too attractive. As shown in Tables 6 and S2, the amount of overestimation in the interaction

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

energies is reduced with the alternative polarizabilities, indicating weaker electrostatic attractions. Consequently, the overestimation of the van der Waals term is compensated by the underestimation of the electrostatic term with the alternative polarizabilities, leading to lower overall RMSEs. Therefore, there is a need to reparameterize the van der Waals terms, and we anticipate the pGM models with re-parameterized van der Waals terms will give interaction energies with better agreement with QM results than those with the "wrong" polarizabilities.

In contrast to the interaction energies IE(Gly_m:Gly_n), as expected, the many-body interaction energies ME(Gly_m:Gly_n) calculated by the pGM models with the "wrong" polarizabilities are consistently worse than those with the "correct" pGM polarizabilities shown in Table 3. For the pGM-perm model, the overall RMSE of the many-body interaction energies increased from 0.40 kcal/mol to 0.70, 0.69, and 1.11 kcal/mol for the scaled pGM, Amoeba13, and ff12pol polarizabilities, respectively. Similarly, for the pGM-ind model, the overall RMSE increased from 0.38 kcal/mol to 0.69, 0.68, and 1.10 kcal/mol for the scaled pGM, Amoeba13, and ff12pol polarizabilities, respectively. Remarkably, the many-body interaction energies produced by the pGM models with alternative polarizabilities consistently outperform the Amoeba13 and ff12pol force fields, with their respective native polarizabilities shown in Table 3. With the Amoeba13 polarizabilities, the RMSEs of the Amoeba13 force field, the pGM-perm, and pGM-ind models are 1.16, 0.69, and 0.68 kcal/mol, respectively; with the ff12pol polarizabilities, the RMSEs of the ff12pol force field, the pGM-perm, and pGM-ind models are 1.62, 1.11, and 1.10 kcal/mol, respectively. Because short-range terms contribute much less to ME(Gly_m:Gly_n) than to IE(Gly_m:Gly_n), these improvements are likely attributable to the differences in the treatment of long-range terms, including the electrostatic and polarization terms. This shows that the pGM models are highly robust in terms of modeling the many-body interactions of peptide main-chain hydrogen-bonding structures. The improvement is remarkable, given the substantial differences among the different force fields in their functional forms of the electrostatic and polarization terms. Both the Amoeba13 and ff12pol force fields are based on the Thole screening schemes, in which only the cross induction between the induced dipoles is screened to avoid polarization catastrophe. In the pGM models, all electrostatic terms are represented as Gaussian densities. Consequently, all electrostatic interactions are screened, including charge—charge, charge—dipole, and dipole—dipole interactions. The improvement observed in this comparison is likely attributable to the inherent consistency of the treatment of electrostatic terms in the pGM models.

The nonadditive and additive contributions to the many-body interactions calculated by the pGM-perm and pGM-ind models with the alternative polarizabilities are shown in Tables 7 and S3, respectively. With the scaled pGM polarizabilities, the RMSEs of the nonadditive contributions of the pGM-perm and pGM-ind models increase from 0.33 and 0.30 kcal/mol to 0.60 and 0.58 kcal/mol, respectively. Despite the fact that the functional forms of the polarization terms of the pGM models are different from those of the Amoeba13 and ff12pol force fields, the nonadditive contributions of the pGM models with the Amoeba13 and ff12pol polarizabilities are remarkably similar to Amoeba13 and ff12pol with their respective native polarizabilities. With the Amoeba13 polarizabilities, the RMSEs of the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Amoeba13 force field, the pGM-perm, and pGM-ind models are 0.59, 0.59, and 0.57 kcal/mol, respectively; with the ff12pol polarizabilities, the RMSEs of the ff12pol force field, the pGM-perm, and pGM-ind models are 0.95, 0.96, and 0.95 kcal/mol, respectively. We therefore conclude that the changes in the functional forms in the calculations of induction energies from their respective native forms did not lead to an intolerable level of error.

For the additive contribution calculations, the RMSEs of the pGM-perm and pGM-ind models with the scaled pGM polarizabilities increase from 0.09 and 0.10 kcal/mol to 0.12 and 0.13 kcal/mol, respectively, which are essentially unchanged. Remarkably, both pGM models with the Amoeba13 and ff12pol polarizabilities notably outperform their respective native counterparts (Amoeba13 and ff12pol force fields) in the calculation of the additive contributions. With the Amoeba13 polarizabilities, the RMSEs of the Amoeba13 force field, the pGM-perm, and pGM-ind models are 0.59, 0.12, and 0.13 kcal/mol, respectively; with the ff12pol polarizabilities, the RMSEs of the ff12pol force field, the pGM-perm, and pGM-ind models are 0.72, 0.17, and 0.18 kcal/mol, respectively. In fact, for the additive contribution calculations, the RMSEs of both pGM models with the alternative polarizabilities are comparable to those with the original polarizabilities. The notably better performance of the pGM models with the Amoeba13 and ff12pol polarizabilities than the Amoeba13 and ff12pol force fields in the additive contribution calculations suggests that the *PyRESP* scheme of fitting charges and permanent multipoles from QM calculated ESPs is a reliable approach in the development of molecular mechanical force fields and has the ability to compensate the errors in the initial parameterization of polarizabilities.

CONCLUSIONS

In this work, we assessed the capabilities of the recently developed pGM models[46,47] in modeling the many-body interactions of glycine dipeptides main-chain hydrogen-bonding conformers. Two types of pGM models were considered, including that with (pGM-perm) and without (pGM-ind) permanent atomic dipoles. The performances of the pGM models were compared with several other widely used force fields, including Amoeba13,[28] ff12pol,[22-25] ff19SB,[7] ff15ipq,[54] and ff03.[55] The glycine dipeptide oligomers were selected as the model systems since glycine has the minimalist side chain so that we can focus on main-chain hydrogen-bonding interactions.

We first identified ωB97X – D/aug-cc-pVTZ without counterpoise BSSE correction as the most suitable DFT method for our molecular systems. Compared with other DFT methods tested (M062X and B3LYP) with and without counterpoise corrections, ωB97X – D without counterpoise correction produced the interaction energies of the formamide dimer and the glycine dipeptide dimers with the best agreement to those calculated at the CCSD(T)/CBS level of theory.

Next, we compared the interaction energies IE(Gly_m:Gly_n) and many-body interaction energies ME(Gly_m:Gly_n) calculated at the ωB97X – D/aug-cc-pVTZ level of theory and those calculated by the seven molecular mechanical force fields. The overall RMSEs of the interaction energies and many-body interaction energies of the seven force fields are shown in Figure 2. Encouragingly, the overall RMSEs of the interaction energies IE(Gly_m:Gly_n) calculated

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

by the pGM-perm and pGM-ind models are 1.35 and 1.37 kcal/mol, respectively, which significantly outperform other polarizable (Amoeba13, 2.91 kcal/mol; ff12pol, 2.28 kcal/mol) and additive (ff19SB, 2.67 kcal/mol; ff15ipq, 1.87 kcal/mol; ff03, 3.78 kcal/mol) force fields. For the many-body interaction energies $ME(Gly_m:Gly_n)$, the overall RMSEs of the pGM-perm and pGM-ind models are 0.40 and 0.38 kcal/mol, respectively. In comparison, the RMSEs of other polarizable (Amoeba13, 1.16 kcal/mol; ff12pol, 1.62 kcal/mol) and additive (ff19SB, 2.45 kcal/mol; ff15ipq, 2.04 kcal/mol; ff03, 2.58 kcal/mol) force fields are notably higher than those of the pGM models. In addition, for both interaction energies and many-body interaction energies, the deviations between the ωB97X – D results and the pGM models results across different conformations and oligomers with different sizes are highly consistent, while all other force fields exhibit nonuniform deviations across different conformations, and these deviations increase with the size of oligomers. Therefore, our data show that the pGM models perform the best among all seven tested force fields in terms of calculating interaction energy and many-body interaction energy.

For polarizable force fields, the many-body interaction energy can be decomposed into the nonadditive contribution $ME_{NA}(Gly_m:Gly_n)$ and the additive contribution $ME_A(Gly_m:Gly_n)$, so that we compared both contributions calculated by the four polarizable force fields with those of ωB97X-D calculations. Figure 2 shows the overall RMSEs of the nonadditive and additive contributions to the many-body interaction energies of the four polarizable force fields. Encouragingly, the pGM models result in the lowest RMSEs for both nonadditive (pGM-perm, 0.33 kcal/mol; pGM-ind, 0.30 kcal/mol) and additive (pGM-perm, 0.09 kcal/mol; pGM-ind, 0.10 kcal/mol) contributions. In comparison, the Amoeba13 force field gives RMSEs of 0.59 kcal/mol for both the nonadditive and additive contributions. The ff12pol force field gives RMSEs of 0.95 kcal/mol for the nonadditive contribution and 0.72 kcal/mol for the additive contribution. Therefore, the pGM models perform the best among all tested polarizable force fields in terms of modeling both the nonadditive and additive contributions to the many-body interactions.

Finally, we tested the robustness of the pGM models against parameterization errors by employing alternative polarizabilities. Interestingly, the pGM models with the alternative polarizabilities produce interaction energies with lower RMSEs compared with those produced by the original pGM polarizabilities. This might be explained by the fact that the current pGM models share identical van der Waals parameters as the ff12pol force field, and the overestimation of the van der Waals term is compensated by the underestimation of the electrostatic term with the alternative polarizabilities. In future works, the van der Waals parameters of the pGM models will be re-parameterized using similar ways as we did for parameterizing the ff12pol force field.[25] On the other hand, the pGM models with the alternative polarizabilities produce many-body interaction energies as well as the nonadditive and additive contributions to the many-body interactions with higher RMSEs compared with those with the original pGM polarizabilities. Even so, both pGM models with the alternative polarizabilities still give better or similar performances compared with the Amoeba13 and ff12pol force fields. Our data show that the pGM models are robust against polarizability errors and perform well even with those “wrong” polarizabilities.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

In summary, this work validates that the pGM models have the capabilities to accurately model the interaction energies, many-body interaction energies, as well as the nonadditive and additive contributions to the many-body interactions of peptide main-chain hydrogen-bonding structures. We expect that the pGM models have the potential to serve as templates for developing the next-generation polarizable force fields for modeling various polarization-sensitive biological processes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the research support from NIH (GM79383 to Y.D. and GM130367 to R.L.).

APPENDIX: PROOF OF MANY-BODY INTERACTION ENERGIES DECOMPOSITION

In the Theory section, we claim that the many-body interaction energies $ME(Gly_m:Gly_n)$ (eq 22) can be decomposed into the nonadditive contributions $ME_{NA}(Gly_m:Gly_n)$ and the additive contributions $ME_A(Gly_m:Gly_n)$, whose formulas are given in eqs 24 and 25, respectively. We first prove the formula of the nonadditive contributions in Appendix A1 and then prove the formula of the additive contributions in Appendix A2.

A1 Nonadditive Contributions

The nonadditive effect refers to that for a molecular system with more than two atoms involved; any two atoms will interact differently compared with the situation where other atoms were not present.[10] For additive force fields, the nonadditive effect does not exist, i.e., $ME_{NA}(Gly_m:Gly_n)$ defined in eq 24 is always zero. Therefore, the interaction energy of the two middle peptides $IE_{mid}(Gly_m:Gly_n)$ in the presence of the neighbor peptides Gly_{m-1} and Gly_{n-1} defined in eq 23 should be the same as the interaction energy of the two middle peptides IE(Gly:Gly) in the absence of the neighbor peptides. The key to prove this is that, for additive force fields, the interaction energy IE(A,B:C,D) of a four-body system A,B:C,D can be decomposed into

$$\begin{aligned} IE(A, B: C, D) = {} & IE(A: C) + IE(B: C) + IE(A: D) \\ & + IE(B: D) \end{aligned} \tag{A1}$$

Therefore, the first three terms in eq 23 can be decomposed to

$$\begin{aligned} IE(Gly_m: Gly_n) = {} & IE(Gly_{m-1}X: XGly_{n-1}) \\ & + IE(Gly_{m-1}X: Gly) + IE(Gly: XGly_{n-1}) \\ & + IE(Gly: Gly) \end{aligned} \tag{A2}$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\mathrm{IE}(\mathrm{Gly}_m\mathrm{:\ XGly}_{n-1}) = \mathrm{IE}(\mathrm{Gly}_{m-1}\mathrm{X:\ XGly}_{n-1}) + \mathrm{IE}(\mathrm{Gly:\ XGly}_{n-1}) \tag{A3}$$

$$\mathrm{IE}(\mathrm{Gly}_{m-1}\mathrm{X:\ Gly}_n) = \mathrm{IE}(\mathrm{Gly}_{m-1}\mathrm{X:\ XGly}_{n-1}) + \mathrm{IE}(\mathrm{Gly}_{m-1}\mathrm{X:\ Gly}) \tag{A4}$$

Substituting eqs A2-A4 into eq 23 gives

$$\mathrm{IE}_{\mathrm{mid}}(\mathrm{Gly}_m\mathrm{:\ Gly}_n) = \mathrm{IE}(\mathrm{Gly:\ Gly}) \tag{A5}$$

Therefore, for additive force fields, the nonadditive contribution $\mathrm{ME}_{\mathrm{NA}}(\mathrm{Gly}_m\mathrm{:Gly}_n)$ in eq 24 becomes zero. For polarizable force fields, the difference between $\mathrm{IE}_{\mathrm{mid}}(\mathrm{Gly}_m\mathrm{:Gly}_n)$ and IE(Gly:Gly) is naturally the nonadditive contribution $\mathrm{ME}_{\mathrm{NA}}(\mathrm{Gly}_m\mathrm{:Gly}_n)$, which is a nonzero value.

A2 Additive Contributions

For either additive or polarizable force fields, the additive contribution can be expressed in the following alternative formula by substituting eqs 22-24 into eq 25

$$\mathrm{ME}_{\mathrm{A}}(\mathrm{Gly}_m\mathrm{:\ Gly}_n) = \mathrm{IE}(\mathrm{Gly}_m\mathrm{:\ XGly}_{n-1}) + \mathrm{IE}(\mathrm{Gly}_{m-1}\mathrm{X:\ Gly}_n) - \mathrm{IE}(\mathrm{Gly}_{m-1}\mathrm{X:\ XGly}_{n-1}) \tag{A6}$$

For additive force fields, we need to show that the many-body interaction energies $\mathrm{ME}(\mathrm{Gly}_m\mathrm{:Gly}_n)$ only have the additive contribution $\mathrm{ME}_{\mathrm{A}}(\mathrm{Gly}_m\mathrm{:Gly}_n)$. This can be done by substituting eq A2 into the formula of the many-body interaction energy in eq 22 and substituting eq A3-A4 into the formula of the additive contribution in eq A6. Therefore, we have proved that for additive force fields, we have

$$\mathrm{ME}(\mathrm{Gly}_m\mathrm{:\ Gly}_n) = \mathrm{ME}_{\mathrm{A}}(\mathrm{Gly}_m\mathrm{:\ Gly}_n) \tag{A7}$$

For polarizable force fields, the difference between the many-body interaction energy $\mathrm{ME}(\mathrm{Gly}_m\mathrm{:Gly}_n)$ and the nonzero nonadditive contribution $\mathrm{ME}_{\mathrm{NA}}(\mathrm{Gly}_m\mathrm{:Gly}_n)$ naturally gives the additive contribution $\mathrm{ME}_{\mathrm{A}}(\mathrm{Gly}_m\mathrm{:Gly}_n)$.

REFERENCES

(1). Leach AR Molecular Modelling: Principles and Applications, 2nd ed.; Pearson Education, 2001.
(2). Monticelli L; Tieleman DPForce Fields for Classical Molecular Dynamics. In Biomolecular Simulations, Methods in Molecular Biology; Humana Press, 2013; pp 197–213.
(3). Vitalis A; Pappu RV Methods for Monte Carlo simulations of biomacromolecules. Annu. Rep. Comput. Chem 2009, 5, 49–76. [PubMed: 20428473]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(4). Jumper J; Evans R; Pritzel A; Green T; Figurnov M; Ronneberger O; Tunyasuvunakool K; Bates R; Žídek A; Potapenko A; et al. Highly accurate protein structure prediction with AlphaFold. Nature 2021, 596, 583–589. [PubMed: 34265844]

(5). Le Grand S; Götz AW; Walker RC SPFP: Speed without compromise—A mixed precision model for GPU accelerated molecular dynamics simulations. Comput. Phys. Commun 2013, 184, 374–380.

(6). Lee T-S; Cerutti DS; Mermelstein D; Lin C; LeGrand S; Giese TJ; Roitberg A; Case DA; Walker RC; York DM GPU-accelerated molecular dynamics and free energy methods in Amber18: performance enhancements and new features. J. Chem. Inf. Model 2018, 58, 2043–2050. [PubMed: 30199633]

(7). Tian C; Kasavajhala K; Belfon KA; Raguette L; Huang H; Migues AN; Bickel J; Wang Y; Pincay J; Wu Q; Simmerling C ff19SB: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. J. Chem. Theory Comput 2020, 16, 528–552. [PubMed: 31714766]

(8). Brooks BR; Brooks CL III; Mackerell AD Jr.; Nilsson L; Petrella RJ; Roux B; Won Y; Archontis G; Bartels C; Boresch S; et al. CHARMM: the biomolecular simulation program. J. Comput. Chem 2009, 30, 1545–1614. [PubMed: 19444816]

(9). Jorgensen WL; Maxwell DS; Tirado-Rives J Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J. Am. Chem. Soc 1996, 118, 11225–11236.

(10). Cieplak P; Dupradeau F-Y; Duan Y; Wang J Polarization effects in molecular mechanical force fields. J. Phys.: Condens. Matter 2009, 21, No. 333102. [PubMed: 21828594]

(11). Friesner RAModeling Polarization in Proteins and Protein–ligand Complexes: Methods and Preliminary Results. In Peptide Solvation and HBonds, Advances in Protein Chemistry; Elsevier Inc., 2005; Vol. 72, pp 79–104.

(12). Gresh N; Cisneros GA; Darden TA; Piquemal J-P Anisotropic, polarizable molecular mechanics studies of inter-and intramolecular interactions and ligand–macromolecule complexes. A bottom-up strategy. J. Chem. Theory Comput 2007, 3, 1960–1986. [PubMed: 18978934]

(13). Zhao S; Schaub AJ; Tsai S-C; Luo R Development of a Pantetheine Force Field Library for Molecular Modeling. J. Chem. Inf. Model 2021, 61, 856–868. [PubMed: 33534558]

(14). King E; Qi R; Li H; Luo R; Aitchison E Estimating the roles of protonation and electronic polarization in absolute binding affinity simulations. J. Chem. Theory Comput 2021, 17, 2541–2555. [PubMed: 33764050]

(15). Draper DE; Grilley D; Soto AM Ions and RNA folding. Annu. Rev. Biophys. Biomol. Struct 2005, 34, 221–243. [PubMed: 15869389]

(16). Lipfert J; Doniach S; Das R; Herschlag D Understanding nucleic acid–ion interactions. Annu. Rev. Biochem 2014, 83, 813–841. [PubMed: 24606136]

(17). Fitch CA; Karp DA; Lee KK; Stites WE; Lattman EE; García-Moreno EB Experimental pKa values of buried residues: analysis with continuum methods and role of water penetration. Biophys. J 2002, 82, 3289–3304. [PubMed: 12023252]

(18). Dill KA; Bromberg S; Yue K; Chan HS; Fiebig KM; Yee DP; Thomas PD Principles of protein folding—a perspective from simple exact models. Protein Sci. 2008, 4, 561–602.

(19). Greatbanks SP; Gready JE; Limaye AC; Rendell AP Enzyme polarization of substrates of dihydrofolate reductase by different theoretical methods. Proteins: Struct., Funct., Bioinf 1999, 37, 157–165.

(20). Cieplak P; Caldwell J; Kollman P Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: aqueous solution free energies of methanol and N-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases. J. Comput. Chem 2001, 22, 1048–1057.

(21). Wang ZX; Zhang W; Wu C; Lei H; Cieplak P; Duan Y Strike a balance: optimization of backbone torsion parameters of AMBER polarizable force field for simulations of proteins and peptides. J. Comput. Chem 2006, 27, 781–790. [PubMed: 16526038]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(22). Wang J; Cieplak P; Li J; Hou T; Luo R; Duan Y Development of polarizable models for molecular mechanical calculations I: parameterization of atomic polarizability. J. Phys. Chem. B 2011, 115, 3091–3099. [PubMed: 21391553]

(23). Wang J; Cieplak P; Li J; Wang J; Cai Q; Hsieh M; Lei H; Luo R; Duan Y Development of polarizable models for molecular mechanical calculations II: induced dipole models significantly improve accuracy of intermolecular interaction energies. J. Phys. Chem. B 2011, 115, 3100–3111. [PubMed: 21391583]

(24). Wang J; Cieplak P; Cai Q; Hsieh M-J; Wang J; Duan Y; Luo R Development of polarizable models for molecular mechanical calculations. 3. Polarizable water models conforming to Thole polarization screening schemes. J. Phys. Chem. B 2012, 116, 7999–8008. [PubMed: 22712654]

(25). Wang J; Cieplak P; Li J; Cai Q; Hsieh M-J; Luo R; Duan Y Development of polarizable models for molecular mechanical calculations. 4. van der Waals parametrization. J. Phys. Chem. B 2012, 116, 7088–7101. [PubMed: 22612331]

(26). Ren P; Ponder JW Consistent treatment of inter-and intramolecular polarization in molecular mechanics calculations. J. Comput. Chem 2002, 23, 1497–1506. [PubMed: 12395419]

(27). Ren P; Ponder JW Polarizable atomic multipole water model for molecular mechanics simulation. J. Phys. Chem. B 2003, 107, 5933–5947.

(28). Shi Y; Xia Z; Zhang J; Best R; Wu C; Ponder JW; Ren P Polarizable atomic multipole-based AMOEBA force field for proteins. J. Chem. Theory Comput 2013, 9, 4046–4063. [PubMed: 24163642]

(29). Banks JL; Kaminski GA; Zhou R; Mainz DT; Berne B; Friesner RA Parametrizing a polarizable force field from ab initio data. I. The fluctuating point charge model. J. Chem. Phys 1999, 110, 741–754.

(30). Patel S; Brooks CL III CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. J. Comput. Chem 2004, 25, 1–16. [PubMed: 14634989]

(31). Lamoureux G; Harder E; Vorobyov IV; Roux B; MacKerell AD Jr. A polarizable model of water for molecular dynamics simulations of biomolecules. Chem. Phys. Lett 2006, 418, 245–249.

(32). Lopes PEM; Lamoureux G; Roux B; MacKerell AD Polarizable empirical force field for aromatic compounds based on the classical drude oscillator. J. Phys. Chem. B 2007, 111, 2873–2885. [PubMed: 17388420]

(33). Tan Y-H; Luo R Continuum treatment of electronic polarization effect. J. Chem. Phys 2007, 126, No. 094103. [PubMed: 17362100]

(34). Tan Y-H; Tan C; Wang J; Luo R Continuum polarizable force field within the Poisson-Boltzmann framework. J. Phys. Chem. B 2008, 112, 7675–7688. [PubMed: 18507452]

(35). Warshel A; Levitt M Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. J. Mol. Biol 1976, 103, 227–249. [PubMed: 985660]

(36). Vesely FJ N-particle dynamics of polarizable Stockmayer-type molecules. J. Comput. Phys 1977, 24, 361–371.

(37). Applequist J; Carl JR; Fung K-K Atom dipole interaction model for molecular polarizability. Application to polyatomic molecules and determination of atom polarizabilities. J. Am. Chem. Soc 1972, 94, 2952–2960.

(38). Thole BT Molecular polarizabilities calculated with a modified dipole interaction. Chem. Phys 1981, 59, 341–350.

(39). Van Duijnen PT; Swart M Molecular and atomic polarizabilities: Thole's model revisited. J. Phys. Chem. A 1998, 102, 2399–2407.

(40). Elking D; Darden T; Woods RJ Gaussian induced dipole polarization model. J. Comput. Chem 2007, 28, 1261–1274. [PubMed: 17299773]

(41). Elking DM; Cisneros GA; Piquemal J-P; Darden TA; Pedersen LG Gaussian multipole model (GMM). J. Chem. Theory Comput 2010, 6, 190–202. [PubMed: 20209077]

(42). Elking DM; Perera L; Duke R; Darden T; Pedersen LG Atomic forces for geometry-dependent point multipole and Gaussian multipole models. J. Comput. Chem 2010, 31, 2702–2713. [PubMed: 20839297]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(43). Wheatley RJ Gaussian multipole functions for describing molecular charge distributions. Mol. Phys 1993, 79, 597–610.

(44). Wheatley RJ; Mitchell JB Gaussian multipoles in practice: Electrostatic energies for intermolecular potentials. J. Comput. Chem 1994, 15, 1187–1198.

(45). Wang J; Cieplak P; Luo R; Duan Y Development of polarizable Gaussian model for molecular mechanical calculations I: Atomic polarizability parameterization to reproduce ab initio anisotropy. J. Chem. Theory Comput 2019, 15, 1146–1158. [PubMed: 30645118]

(46). Wei H; Qi R; Wang J; Cieplak P; Duan Y; Luo R Efficient formulation of polarizable Gaussian multipole electrostatics for biomolecular simulations. J. Chem. Phys 2020, 153, No. 114116. [PubMed: 32962395]

(47). Wei H; Cieplak P; Duan Y; Luo R Stress tensor and constant pressure simulation for polarizable Gaussian multipole model. J. Chem. Phys 2022, 156, No. 114114. [PubMed: 35317572]

(48). Darden T; York D; Pedersen L Particle mesh Ewald: An N·log (N) method for Ewald sums in large systems. J. Chem. Phys 1993, 98, 10089–10092.

(49). Essmann U; Perera L; Berkowitz ML; Darden T; Lee H; Pedersen LG A smooth particle mesh Ewald method. J. Chem. Phys 1995, 103, 8577–8593.

(50). Crowley M; Darden T; Cheatham T; Deerfield D Adventures in improving the scaling and accuracy of a parallel molecular dynamics program. J. Supercomput 1997, 11, 255–278.

(51). Duke RE; Cisneros GA Ewald-based methods for Gaussian integral evaluation: application to a new parameterization of GEM*. J. Mol. Model 2019, 25, No. 307. [PubMed: 31501946]

(52). Zhao S; Wei H; Cieplak P; Duan Y; Luo R PyRESP: A Program for Electrostatic Parameterizations of Additive and Induced Dipole Polarizable Force Fields. J. Chem. Theory Comput 2022, 18, 3654–3670. [PubMed: 35537209]

(53). Case DA; Aktulga HM; Belfon K; Ben-Shalom I; Berryman JT; Brozell SR; Cerutti DS; Cheatham TE III; Cisneros GA; Cruzeiro VWD; Darden TA; Duke RE; Giambasu G; Gilson MK; Gohlke H; Goetz AW; Harris R; Izadi S; Izmailov SA; Kasavajhala K; Kaymak MC; King E; Kovalenko A; Kurtzman T; Lee T; LeGrand S; Li P; Lin C; Liu J; Luchko T; Luo R; Machado M; Man V; Manathunga M; Merz KM; Miao Y; Mikhailovskii O; Monard G; Nguyen H; O'Hearn KA; Onufriev A; Pan F; Pantano S; Qi R; Rahnamoun A; Roe DR; Roitberg A; Sagui C; Schott-Verdugo S; Shajan A; Shen J; Simmerling CL; Skrynnikov NR; Smith J; Swails J; Walker RC; Wang J; Wang J; Wei H; Wolf RM; Wu X; Xiong Y; Xue Y; York DM; Zhao S; Kollman PA Amber 2022; University of California: San Francisco, 2022.

(54). Debiec KT; Cerutti DS; Baker LR; Gronenborn AM; Case DA; Chong LT Further along the road less traveled: AMBER ff15ipq, an original protein force field built on a self-consistent physical model. J. Chem. Theory Comput 2016, 12, 3926–3947. [PubMed: 27399642]

(55). Duan Y; Wu C; Chowdhury S; Lee MC; Xiong G; Zhang W; Yang R; Cieplak P; Luo R; Lee T; et al. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J. Comput. Chem 2003, 24, 1999–2012. [PubMed: 14531054]

(56). Chai J-D; Head-Gordon M Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. Phys. Chem. Chem. Phys 2008, 10, 6615–6620. [PubMed: 18989472]

(57). Zhao Y; Truhlar DG The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, non-covalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. Theor. Chem. Acc 2008, 120, 215–241.

(58). Lee C; Yang W; Parr RG Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Phys. Rev. B: Condens. Matter Mater. Phys 1988, 37, 785.

(59). Becke AD Density-functional thermochemistry. III. The role of exact exchange. J. Chem. Phys 1993, 98, 5648–5652.

(60). Boys SF; Bernardi F The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. Mol. Phys 1970, 19, 553–566.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(61). Halkier A; Helgaker T; Jørgensen P; Klopper W; Koch H; Olsen J; Wilson AK Basis-set convergence in correlated calculations on Ne, N2, and H2O. Chem. Phys. Lett 1998, 286, 243–252.

(62). Halkier A; Helgaker T; Jørgensen P; Klopper W; Olsen J Basis-set convergence of the energy in molecular Hartree–Fock calculations. Chem. Phys. Lett 1999, 302, 437–446.

(63). Connolly ML Analytical molecular surface calculation. J. Appl. Crystallogr 1983, 16, 548–558.

(64). Singh UC; Kollman PA An approach to computing electrostatic charges for molecules. J. Comput. Chem 1984, 5, 129–145.

(65). Frisch MJ; Trucks GW; Schlegel HB; Scuseria GE; Robb MA; Cheeseman JR; Scalmani G; Barone V; Petersson GA; Nakatsuji H; Li X; Caricato M; Marenich AV; Bloino J; Janesko BG; Gomperts R; Mennucci B; Hratchian HP; Ortiz JV; Izmaylov AF; Sonnenberg JL; Williams-Young D; Ding F; Lipparini F; Egidi F; Goings J; Peng B; Petrone A; Henderson T; Ranasinghe D; Zakrzewski VG; Gao J; Rega N; Zheng G; Liang W; Hada M; Ehara M; Toyota K; Fukuda R; Hasegawa J; Ishida M; Nakajima T; Honda Y; Kitao O; Nakai H; Vreven T; Throssell K; Montgomery JA Jr.; Peralta JE; Ogliaro F; Bearpark MJ; Heyd JJ; Brothers EN; Kudin KN; Staroverov VN; Keith TA; Kobayashi R; Normand J; Raghavachari K; Rendell AP; Burant JC; Iyengar SS; Tomasi J; Cossi M; Millam JM; Klene M; Adamo C; Cammi R; Ochterski JW; Martin RL; Morokuma K; Farkas O; Foresman JB; Fox DJGaussian 16, revision A.03; Gaussian Inc.: Wallingford, CT, 2016.

(66). Xie W; Pu J; Gao J A coupled polarization-matrix inversion and iteration approach for accelerating the dipole convergence in a polarizable potential function. J. Phys. Chem. A 2009, 113, 2109–2116. [PubMed: 19123850]

(67). Ponder JWTINKER: Software Tools For Molecular Design, Washington University School of Medicine: Saint Louis, MO, 2004; 3.

(68). Li A; Muddana HS; Gilson MK Quantum mechanical calculation of non-covalent interactions: a large-scale evaluation of PMx, DFT, and SAPT approaches. J. Chem. Theory Comput 2014, 10, 1563–1575. [PubMed: 24803867]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

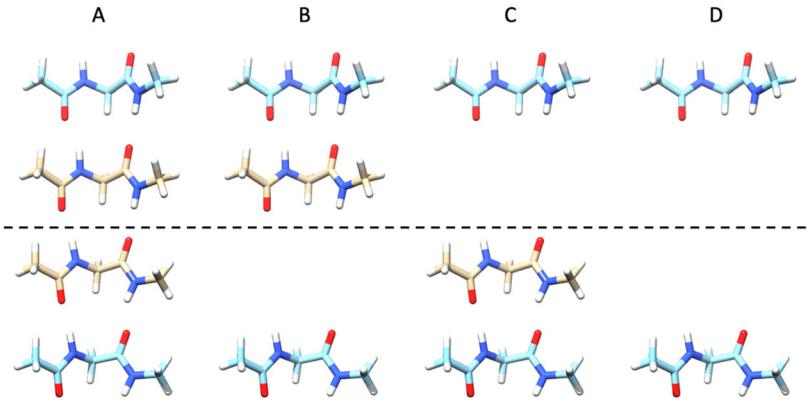

Figure 1.

Glycine dipeptide oligomers (in the parallel β-sheet conformation) used to calculate the interaction energy, many-body interaction energy, and the nonadditive and additive contributions to the many-body interaction of the Gly_2:Gly_2 oligomer. For each oligomer, the interaction energies between glycine dipeptides above and below the dashed line are calculated using eq 21. (A) Gly_2:Gly_2, (B) Gly_2:XGly, (C) GlyX:Gly_2, and (D) GlyX:XGly. Refer to the Theory section for detailed descriptions.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

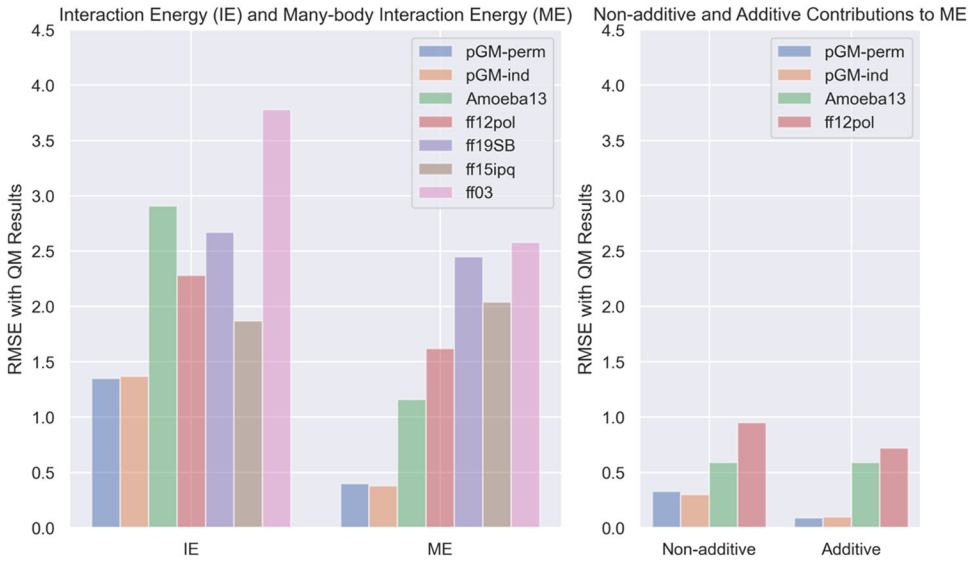

Figure 2.

Overall RMSEs of the interaction energies IE(Gly_m:Gly_n), many-body interaction energies ME(Gly_m:Gly_n) (left), as well as the nonadditive contribution ME_{NA}(Gly_m:Gly_n) and the additive contribution ME_A(Gly_m:Gly_n) to the many-body interaction energies (right) of the tested force fields with the ωB97X – D/aug-cc-pVTZ calculated results.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1.

Quantum Mechanical Interaction Energies of the Formamide Dimer and the Glycine Dipeptide Dimers Calculated by the CCSD(T)/CBS and Density Functional Theory Methods (kcal/mol)

		ωB97X – D/aTZ		M062X/aTZ		B3LYP/aTZ	
	CCSD(T)/CBS[a]	**noCp**[b]	**CP**[b]	**noCp**[b]	**CP**[b]	**noCP**[b]	**CP**[b]
	Formamide Dimer						
	−7.04	−6.96 (0.08)[d]	−6.83 (0.21)	−6.91 (0.13)	−6.77 (0.27)	−6.15 (0.89)	−6.05 (0.99)
	Glycine Dipeptide Dimer (Gly:Gly)						
αR[c]	−14.40	−14.73 (−0.33)	−14.29 (0.11)	−14.00 (0.40)	−13.57 (0.83)	−10.37 (4.03)	−10.01 (4.39)
aβ[c]	−14.49	−14.53 (−0.04)	−14.10 (0.39)	−14.22 (0.27)	−13.77 (0.72)	−10.54 (3.95)	−10.17 (4.32)
β[c]	−17.34	−17.36 (−0.02)	−16.91 (0.43)	−16.93 (0.41)	−16.47 (0.87)	−13.75 (3.59)	−13.37 (3.97)
RMSE		0.17	0.31	0.32	0.71	3.37	3.70
MAE		0.12	0.29	0.30	0.67	3.12	3.42

[a] The formula for the CCSD(T)/CBS interaction energies is shown in eq 29.

[b] noCP means no counterpoise BSSE corrections; CP means with counterpoise BSSE corrections.

[c] αR, aβ, and β represent α-helix, anti-parallel β-sheet, and parallel β-sheet conformations, respectively.

[d] The values in parentheses are differences between the values calculated by DFT methods and corresponding values calculated by the CCSD(T)/CBS method.

J Chem Theory Comput. Author manuscript; available in PMC 2023 May 02.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

J Chem Theory Comput. Author manuscript; available in PMC 2023 May 02.

Table 2.

Interaction Energies **IE**(**Gly$_m$:Gly$_n$**) of Glycine Dipeptide Oligomers **Gly$_m$:Gly$_n$** Calculated by **ωB97X – D**/aTZ and Molecular Mechanical Force Fields (kcal/mol)[a]

	ωB97X – D/aTZ		pGM-perm	pGM-ind	Amoeba13	ff12pol	ff19SB	ff15ipq	ff03
				Gly:Gly					
αR	*b*	−14.73	−16.21 (−1.48)[d]	−16.09 (−1.36)	−12.98 (1.75)	−12.91 (1.81)	−16.04 (−1.31)	−19.08 (−4.35)	−15.35 (−0.62)
a*β*	*b*	−14.53	−16.01 (−1.48)	−16.13 (−1.60)	−13.82 (0.71)	−14.67 (−0.15)	−13.60 (0.92)	−16.31 (−1.78)	−11.83 (2.70)
β	*b*	−17.36	−19.24 (−1.88)	−19.28 (−1.92)	−14.16 (3.20)	−17.82 (−0.46)	−15.69 (1.67)	−18.85 (−1.48)	−14.78 (2.58)
				Gly:Gly$_2$					
αR	*b*	−17.90	−19.21 (−1.31)	−19.11 (−1.22)	−15.38 (2.52)	−14.94 (2.96)	−17.90 (−0.01)	−21.30 (−3.40)	−17.09 (0.80)
a*β*	*b*	−16.00	−17.26 (−1.26)	−17.38 (−1.38)	−14.89 (1.11)	−15.68 (0.32)	−14.06 (1.94)	−16.83 (−0.83)	−12.28 (3.72)
β	*b*	−18.99	−20.64 (−1.65)	−20.67 (−1.69)	−15.42 (3.57)	−18.97 (0.02)	−16.19 (2.80)	−19.45 (−0.46)	−15.26 (3.73)
				Gly:Gly$_3$					
αR	*b*	−19.06	−20.32 (−1.26)	−20.23 (−1.18)	−16.26 (2.80)	−15.68 (3.37)	−18.42 (0.63)	−21.93 (−2.87)	−17.57 (1.48)
a*β*	*b*	−16.12	−17.36 (−1.24)	−17.47 (−1.35)	−14.95 (1.17)	−15.74 (0.38)	−14.09 (2.03)	−16.86 (−0.74)	−12.31 (3.81)
β	*b*	−19.16	−20.79 (−1.63)	−20.83 (−1.67)	−15.56 (3.61)	−19.08 (0.08)	−16.25 (2.91)	−19.52 (−0.36)	−15.31 (3.85)
				Gly$_2$:Gly$_2$					
αR	*b*	−22.11	−23.20 (−1.09)	−23.12 (−1.01)	−18.54 (3.57)	−17.71 (4.40)	−20.30 (1.81)	−24.16 (−2.04)	−19.36 (2.75)
a*β*	*b*	−17.72	−18.75 (−1.03)	−18.85 (−1.13)	−16.21 (1.51)	−16.88 (0.84)	−14.61 (3.11)	−17.49 (0.23)	−12.80 (4.92)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

	ωB97X − D/aTZ	pGM-perm	pGM-ind	Amoeba13	ff12pol	ff19SB	ff15ipq	ff03
β [b]	−21.02	−22.34 (−1.33)	−22.40 (−1.38)	−16.98 (4.03)	−20.27 (0.75)	−16.77 (4.25)	−20.14 (0.87)	−15.80 (5.22)
			Gly_3:Gly_3					
αR [b]	−25.73	−26.68 (−0.96)	−26.63 (−0.90)	−21.27 (4.45)	−20.09 (5.64)	−21.91 (3.82)	−26.11 (−0.39)	−20.86 (4.87)
aβ [b]	−18.03	−18.99 (−0.97)	−19.09 (−1.07)	−16.37 (1.66)	−17.04 (0.98)	−14.69 (3.33)	−17.57 (0.46)	−12.90 (5.12)
β [b]	−21.47	−22.74 (−1.27)	−22.80 (−1.32)	−17.33 (4.14)	−20.56 (0.91)	−16.92 (4.55)	−20.33 (1.14)	−15.94 (5.53)
RMSE		1.35	1.37	2.91	2.28	2.67	1.87	3.78
MAE		1.32	1.34	2.65	1.54	2.34	1.43	3.45
max dev[c]		−0.96	−0.90	4.45	5.64	4.55	1.14	5.53
min dev[c]		−1.88	−1.92	0.71	−0.46	−1.31	−4.35	−0.62

[a] The formula for the interaction energies $IE(Gly_m{:}Gly_n)$ is shown in eq 21.

[b] αR, aβ, and β represent α-helix, anti-parallel β-sheet, and parallel β-sheet conformations, respectively.

[c] The maximum and minimum deviations with ωB97X − D/aTZ results.

[d] The values in parentheses are differences between the values calculated by molecular mechanical force fields and corresponding values calculated by the ωB97X − D/aTZ method.

J Chem Theory Comput. Author manuscript; available in PMC 2023 May 02.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 3.

Many-Body Interaction Energies **ME**(**Gly$_m$:Gly$_n$**) of Glycine Dipeptide Oligomers **Gly$_m$:Gly$_n$** Calculated by **ωB97X – D**/aTZ and Molecular Mechanical Force Fields (kcal/mol)[a,b]

ωB97X – D/aTZ		pGM-perm	pGM-ind	Amoeba13	ff12pol	ff19SB	ff15ipq	ff03
			Gly:Gly$_2$					
αR	−3.17	−3.00 (0.17)	−3.02 (0.15)	−2.40 (0.77)	−2.03 (1.14)	−1.87 (1.30)	−2.22 (0.95)	−1.75 (1.42)
aβ	−1.47	−1.25 (0.22)	−1.24 (0.23)	−1.07 (0.40)	−1.00 (0.47)	−0.46 (1.02)	−0.53 (0.95)	−0.44 (1.03)
β	−1.63	−1.39 (0.23)	−1.39 (0.23)	−1.25 (0.37)	−1.14 (0.48)	−0.50 (1.12)	−0.60 (1.02)	−0.48 (1.15)
			Gly:Gly$_3$					
αR	−4.33	−4.11 (0.22)	−4.14 (0.19)	−3.28 (1.05)	−2.77 (1.56)	−2.39 (1.94)	−2.85 (1.48)	−2.23 (2.10)
aβ	−1.59	−1.35 (0.24)	−1.34 (0.25)	−1.14 (0.46)	−1.07 (0.52)	−0.49 (1.11)	−0.55 (1.04)	−0.48 (1.11)
β	−1.80	−1.55 (0.25)	−1.55 (0.25)	−1.39 (0.41)	−1.26 (0.54)	−0.56 (1.24)	−0.67 (1.13)	−0.53 (1.27)
			Gly$_2$:Gly$_2$					
αR	−7.38	−6.99 (0.39)	−7.03 (0.35)	−5.57 (1.82)	−4.80 (2.58)	−4.27 (3.12)	−5.08 (2.31)	−4.01 (3.37)
aβ	−3.19	−2.74 (0.45)	−2.72 (0.47)	−2.39 (0.80)	−2.20 (0.99)	−1.01 (2.19)	−1.19 (2.01)	−0.97 (2.22)
β	−3.66	−3.10 (0.55)	−3.12 (0.54)	−2.82 (0.84)	−2.45 (1.21)	−1.08 (2.57)	−1.30 (2.36)	−1.02 (2.64)
			Gly$_3$:Gly$_3$					
αR	−11.00	−10.47 (0.53)	−10.53 (0.47)	−8.29 (2.71)	−7.18 (3.82)	−5.88 (5.12)	−7.03 (3.97)	−5.51 (5.49)
aβ	−3.50	−2.98 (0.51)	−2.96 (0.53)	−2.55 (0.95)	−2.37 (1.13)	−1.09 (2.41)	−1.26 (2.24)	−1.07 (2.43)
β	−4.11	−3.50 (0.61)	−3.52 (0.59)	−3.17 (0.94)	−2.74 (1.37)	−1.23 (2.87)	−1.48 (2.63)	−1.16 (2.95)
RMSE		0.40	0.38	1.16	1.62	2.45	2.04	2.58
MAE		0.37	0.35	0.96	1.32	2.17	1.84	2.26
max dev		0.61	0.59	2.71	3.82	5.12	3.97	5.49
min dev		0.17	0.15	0.37	0.47	1.02	0.95	1.03

[a]The formula for the many-body interaction energies ME(Gly$_m$:Gly$_n$) is shown in eq 22.

[b]See Table 2 and text for notation.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

J Chem Theory Comput. Author manuscript; available in PMC 2023 May 02.

Table 4.

Nonadditive Contributions $\mathbf{ME_{NA}(Gly_m:Gly_n)}$ to the Many-Body Interaction Energies of Glycine Dipeptide Oligomers $\mathbf{Gly_m:Gly_n}$ Calculated by **ωB97X – D**/aTZ and Polarizable Force Fields (kcal/mol)[a,b]

ωB97X – D/**aTZ**		**pGM-perm**	**pGM-ind**	**Amoeba13**	**ff12pol**
Gly:Gly_2					
αR	−1.70	−1.57 (0.12)	−1.61 (0.09)	−1.30 (0.39)	−0.98 (0.71)
a*β*	−1.04	−0.83 (0.21)	−0.83 (0.21)	−0.76 (0.28)	−0.64 (0.40)
β	−1.12	−0.91 (0.21)	−0.92 (0.20)	−0.86 (0.26)	−0.73 (0.40)
Gly:Gly_3					
αR	−2.17	−2.02 (0.15)	−2.07 (0.10)	−1.65 (0.52)	−1.27 (0.90)
a*β*	−1.12	−0.89 (0.23)	−0.89 (0.22)	−0.80 (0.31)	−0.68 (0.44)
β	−1.21	−0.99 (0.22)	−1.00 (0.21)	−0.95 (0.26)	−0.79 (0.42)
Gly_2:Gly_2					
αR	−3.52	−3.27 (0.26)	−3.34 (0.18)	−2.68 (0.84)	−2.10 (1.43)
a*β*	−2.13	−1.71 (0.42)	−1.72 (0.42)	−1.56 (0.57)	−1.32 (0.81)
β	−2.48	−1.99 (0.49)	−2.02 (0.47)	−1.90 (0.58)	−1.51 (0.97)
Gly_3:Gly_3					
αR	−4.62	−4.32 (0.31)	−4.41 (0.21)	−3.48 (1.14)	−2.79 (1.83)
a*β*	2.29	−1.83 (0.46)	−1.83 (0.45)	−1.65 (0.64)	−1.40 (0.89)
β	−2.68	−2.16 (0.52)	−2.19 (0.49)	−2.05 (0.63)	−1.62 (1.06)
RMSE		0.33	0.30	0.59	0.95
MAE		0.30	0.27	0.54	0.85
max dev		0.52	0.49	1.14	1.83
min dev		0.12	0.09	0.26	0.40

[a] The formula for the nonadditive contributions $ME_{NA}(Gly_m:Gly_n)$ to the many-body interaction energies is shown in eq 24.

[b] See Table 2 and text for notation.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 5.

Additive Contributions $\mathbf{ME_A(Gly_m:Gly_n)}$ to the Many-Body Interaction Energies of Glycine Dipeptide Oligomers $\mathbf{Gly_m:Gly_n}$ Calculated by **ωB97X – D**/aTZ and Polarizable Force Fields (kcal/mol)[a,b]

	ωB97X – D/**aTZ**	**pGM-perm**	**pGM-ind**	**Amoeba13**	**ff12pol**
	Gly:Gly_2				
αR	−1.47	−1.43 (0.05)	−1.41 (0.06)	−1.10 (0.38)	−1.04 (0.43)
aβ	−0.43	−0.42 (0.01)	−0.41 (0.02)	−0.31 (0.12)	−0.36 (0.07)
β	−0.50	−0.48 (0.02)	−0.47 (0.03)	−0.39 (0.11)	−0.42 (0.08)
	Gly:Gly_3				
αR	−2.16	−2.09 (0.07)	−2.08 (0.08)	−1.63 (0.53)	−1.50 (0.66)
aβ	−0.48	−0.46 (0.02)	−0.45 (0.03)	−0.33 (0.15)	−0.39 (0.09)
β	−0.59	−0.56 (0.03)	−0.55 (0.04)	−0.44 (0.14)	−0.47 (0.11)
	Gly_2:Gly_2				
αR	−3.86	−3.73 (0.14)	−3.69 (0.17)	−2.88 (0.98)	−2.70 (1.16)
aβ	−1.06	−1.03 (0.03)	−1.01 (0.05)	−0.83 (0.23)	−0.88 (0.17)
β	−1.17	−1.11 (0.06)	−1.10 (0.07)	−0.92 (0.25)	−0.94 (0.23)
	Gly_3:Gly_3				
αR	−6.37	−6.16 (0.22)	−6.12 (0.25)	−4.81 (1.57)	−4.38 (1.99)
aβ	−1.21	−1.15 (0.06)	−1.13 (0.08)	−0.90 (0.31)	−0.97 (0.24)
β	−1.43	−1.34 (0.09)	−1.33 (0.10)	−1.11 (0.32)	−1.12 (0.31)
RMSE		0.09	0.10	0.59	0.72
MAE		0.06	0.08	0.42	0.46
max dev		0.22	0.25	1.57	1.99
min dev		0.01	0.02	0.11	0.07

[a] The formula for the additive contributions $ME_A(Gly_m:Gly_n)$ to the many-body interaction energies is shown in eq 25.

[b] See Table 2 and text for notation.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 6.

Interaction Energies **IE**(**Gly$_m$:Gly$_n$**) and Many-Body Interaction Energies **ME**(**Gly$_m$:Gly$_n$**) of Glycin Dipeptide Oligomers **Gly$_m$:Gly$_n$** Calculated by the pGM-perm Model with the Alternative Polarizabilities (kcal/mol)[a,b]

	pGM-perm/scaled[c]		pGM-perm/Amoeba[c]		pGM-perm/ff12pol[c]	
	IE(Gly$_m$:Gly$_n$)	ME(Gly$_m$:Gly$_n$)	IE(Gly$_m$:Gly$_n$)	ME(Gly$_m$:Gly$_n$)	IE(Gly$_m$:Gly$_n$)	ME(Gly$_m$:Gly$_n$)
			Gly:Gly			
αR	−16.06 (−1.33)		−16.13 (−1.40)		−15.86 (−1.13)	
aβ	−15.58 (−1.06)		−15.60 (−1.07)		−15.01 (−0.48)	
β	−18.64 (−1.28)		−18.62 (−1.26)		−17.75 (−0.39)	
			Gly:Gly$_2$			
αR	−18.89 (−0.99)	−2.83 (0.34)	−18.93 (−1.04)	−2.80 (0.36)	−18.44 (−0.54)	−2.58 (0.59)
aβ	−16.72 (−0.72)	−1.13 (0.34)	−16.75 (−0.75)	−1.14 (0.33)	−15.99 (0.01)	−0.98 (0.49)
β	−19.90 (−0.91)	−1.26 (0.37)	−19.89 (−0.90)	−1.27 (0.36)	−18.84 (0.15)	−1.09 (0.54)
			Gly:Gly$_3$			
αR	−19.91 (−0.85)	−3.85 (0.48)	−19.96 (−0.91)	−3.83 (0.50)	−19.35 (−0.30)	−3.49 (0.84)
aβ	−16.80 (−0.68)	−1.21 (0.38)	−16.83 (−0.71)	−1.23 (0.36)	−16.06 (0.06)	−1.05 (0.54)
β	−20.04 (−0.88)	−1.40 (0.40)	−20.03 (−0.87)	−1.41 (0.39)	−18.96 (0.20)	−1.21 (0.59)
			Gly$_2$:Gly$_2$			
αR	−22.64 (−0.53)	−6.58 (0.80)	−22.67 (−0.56)	−6.54 (0.84)	−21.86 (0.25)	−6.00 (1.38)
aβ	−18.06 (−0.34)	−2.47 (0.72)	−18.10 (−0.38)	−2.50 (0.69)	−17.15 (0.57)	−2.14 (1.05)
β	−21.43 (−0.42)	−2.79 (0.86)	−21.43 (−0.41)	−2.81 (0.85)	−20.15 (0.87)	−2.40 (1.25)
			Gly$_3$:Gly$_3$			
αR	−25.85 (−0.12)	−9.79 (1.21)	−25.90 (−0.17)	−9.77 (1.23)	−24.72 (1.01)	−8.86 (2.14)
aβ	−18.27 (−0.25)	−2.69 (0.81)	−18.33 (−0.30)	−2.73 (0.77)	−17.33 (0.69)	−2.32 (1.17)
β	−21.78 (−0.31)	−3.14 (0.97)	−21.80 (−0.33)	−3.18 (0.93)	−20.45 (1.02)	−2.71 (1.40)
RMSE	0.80	0.70	0.82	0.69	0.62	1.11
MAE	0.71	0.64	0.74	0.63	0.51	1.00
max dev	−0.12	1.21	−0.17	1.23	1.02	2.14

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

	pGM-perm/scaled[c]		pGM-perm/Amoeba[c]		pGM-perm/ff12pol[c]	
	IE(Gly_m:Gly_n)	ME(Gly_m:Gly_n)	IE(Gly_m:Gly_n)	ME(Gly_m:Gly_n)	IE(Gly_m:Gly_n)	ME(Gly_m:Gly_n)
min dev	−1.33	0.34	−1.40	0.33	−1.13	0.49

[a] See Table 2 and text for notation.

[b] The values in parentheses are differences between the values calculated by the pGM-perm model with the alternative polarizabilities and corresponding values calculated by the ωB97X − D/aTZ method, which can be found in Tables 2 and 3, respectively.

[c] pGM-perm/scaled, pGM-perm/Amoeba, and pGM-perm/ff12pol represent the pGM-perm models with the pGM polarizabilities scaled by a factor of 0.9, the Amoeba13 polarizabilities, and the ff12pol polarizabilities, respectively.

J Chem Theory Comput. Author manuscript; available in PMC 2023 May 02.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 7.

Nonadditive Contributions $\mathbf{ME_{NA}(Gly_m:Gly_n)}$ and Additive Contributions $\mathbf{ME_{A}(Gly_m:Gly_n)}$ to the Many-Body Interaction Energies of Glycine Dipeptide Oligomers $\mathbf{Gly_m:Gly_n}$ Calculated by the pGM-perm Model with the Alternative Polarizabilities (kcal/mol)[a,b]

	pGM-perm/scaled[c]		pGM-perm/Amoeba[c]		pGM-perm/ff12pol[c]	
	$ME_{NA}(Gly_m:Gly_n)$	$ME_{A}(Gly_m:Gly_n)$	$ME_{NA}(Gly_m:Gly_n)$	$ME_{A}(Gly_m:Gly_n)$	$ME_{NA}(Gly_m:Gly_n)$	$ME_{A}(Gly_m:Gly_n)$
			Gly:Gly_2			
αR	−1.38 (0.31)	−1.44 (0.03)	−1.37 (0.33)	−1.44 (0.04)	−1.12 (0.58)	−1.46 (0.01)
aβ	−0.71 (0.33)	−0.42 (0.01)	−0.73 (0.32)	−0.42 (0.01)	−0.56 (0.48)	−0.42 (0.01)
β	−0.78 (0.35)	−0.48 (0.02)	−0.79 (0.33)	−0.48 (0.02)	−0.61 (0.51)	−0.48 (0.02)
			Gly:Gly_3			
αR	−1.77 (0.40)	−2.08 (0.08)	−1.75 (0.42)	−2.08 (0.08)	−1.42 (0.75)	−2.07 (0.09)
aβ	−0.76 (0.36)	−0.46 (0.02)	−0.78 (0.34)	−0.45 (0.02)	−0.60 (0.52)	−0.45 (0.03)
β	−0.84 (0.37)	−0.55 (0.03)	−0.86 (0.35)	−0.55 (0.03)	−0.66 (0.55)	−0.55 (0.04)
			Gly_2:Gly_2			
αR	−2.87 (0.65)	−3.71 (0.15)	−2.84 (0.68)	−3.70 (0.16)	−2.31 (1.21)	−3.69 (0.17)
aβ	−1.46 (0.67)	−1.02 (0.04)	−1.49 (0.64)	−1.01 (0.05)	−1.15 (0.99)	−1.00 (0.06)
β	−1.69 (0.79)	−1.10 (0.07)	−1.71 (0.77)	−1.10 (0.07)	−1.32 (1.17)	−1.08 (0.09)
			Gly_3:Gly_3			
αR	−3.76 (0.86)	−6.03 (0.35)	−3.74 (0.89)	−6.03 (0.34)	−3.00 (1.62)	−5.86 (0.52)
aβ	−1.55 (0.73)	−1.13 (0.08)	−1.59 (0.69)	−1.13 (0.08)	−1.22 (1.07)	−1.10 (0.11)
β	−1.82 (0.85)	−1.32 (0.11)	−1.86 (0.82)	−1.32 (0.11)	−1.42 (1.26)	−1.29 (0.14)
RMSE	0.60	0.12	0.59	0.12	0.96	0.17
MAE	0.56	0.08	0.55	0.09	0.89	0.11
max dev	0.86	0.35	0.89	0.34	1.62	0.52
min dev	0.31	0.01	0.32	0.01	0.48	0.01

[a] See Table 6 and text for notation.

[b] The values in parentheses are differences between the values calculated by the pGM-perm model with the alternative polarizabilities and corresponding values calculated by the ωB97X − D/aTZ method, which can be found in Tables 4 and 5, respectively.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[c] pGM-perm/scaled, pGM-perm/Amoeba, and pGM-perm/ff12pol represent the pGM-perm models with the pGM polarizabilities scaled by a factor of 0.9, the Amoeba13 polarizabilities, and the ff12pol polarizabilities, respectively.