

基于键偶极模型的多肽-水体系多体极化作用强度高效计算新方法

祝佳怡, 姜笑楠, 郑笑函, 郝强, 王长生

(辽宁师范大学化学化工学院, 大连 110629)

摘要 为解决传统非极化力场无法准确描述三体及更高阶多体作用的问题, 并精确模拟液态水环境下蛋白质的结构和功能, 提出了一种可以快速准确计算多肽-水体系中的多体极化作用强度的新方法. 该方法将多肽分子中的N—H和C=O极性化学键及水分子的O—H键视为键偶极, 周围环境对化学键的极化作用使得该化学键上产生了诱导键偶极, 使用键偶极之间的相互作用来描述体系中的多体极化作用. 通过拟合模型体系中三体作用能随分子间距离变化的势能曲线确定所需参数. 将该方法应用于多肽-水体系中的三体作用强度计算, 并与高精度的MP2方法和AMOEBABIO18可极化力场方法进行了系统比较. 结果表明, 本文方法计算的6个多肽水团簇(共包含92290个三体作用)的总三体作用强度与MP2方法的计算结果具有极高的线性相关性(线性相关系数为0.9965, 均方根误差为7.29 kJ/mol), 准确度略优于AMOEBABIO18可极化力场方法(其与MP2方法的线性相关系数为0.9950, 均方根误差为10.74 kJ/mol). 此外, 本文方法的计算效率与AMOEBABIO18可极化力场方法相比也有显著提高, 在涉及超过20000次三体相互作用的模拟中, 计算时间缩短了约50%. 本文提出的高效方法为大规模蛋白质-水体系的多体作用模拟提供了新思路, 在相关领域中具有重要应用潜力.

关键词 多体极化作用; 化学键偶极; 诱导键偶极; 多肽-水体系; 三体作用能
中图分类号 O641.3 **文献标志码** A **doi:** 10.7503/cjcu20240518

A Novel Efficient Method Based on the Bond Dipole Model for Computing Many-body Polarization Strength in Peptide-Water Systems

ZHU Jiayi, JIANG Xiaonan, ZHENG Xiaohan, HAO Qiang*, WANG Changsheng*
(School of Chemistry and Chemical Engineering, Liaoning Normal University, Dalian 116029, China)

Abstract To address the limitations of traditional non-polarizable force fields in describing three-body and higher-order many-body interactions, and to accurately simulate protein structure and function in liquid water environments, we present a novel method for the efficient and accurate calculation of many-body polarization strengths in peptide-water systems. The N—H and C=O polar bonds in peptide molecules and the O—H bonds in water were treated as bond dipoles. The polarization effect of the environment on the chemical bond causes an induced bond dipole, and interactions of bond dipoles were employed to describe the many-body polarization interaction. The required parameters were determined by fitting the three-body interaction energy curves of the model molecules with different interaction distance. The method and parameters were applied to calculate the three-body polarization strength in peptide-water systems and the results were compared with those of high-precision MP2 method and the

收稿日期: 2024-11-25. 网络首发日期: 2024-12-31.

联系人简介: 郝强, 男, 博士, 讲师, 主要从事理论与计算化学研究. E-mail: qh@luun.edu.cn

王长生, 男, 博士, 教授, 主要从事理论与计算化学研究. E-mail: chwanges@luun.edu.cn

基金项目: 国家自然科学基金(批准号: 21773102)和辽宁省教育厅基础科研项目(批准号: LJKMZ20221411)资助.

Supported by the National Natural Science Foundation of China(No.21773102) and the General Research Program of the Department of Education of Liaoning Province, China(No.LJKMZ20221411).

AMOEBABIO18 polarizable force field method. It is showed that the results of our method have a high linear correlation with those of MP2 method (correlation coefficient=0.9965, RMSE=7.29 kJ/mol) in the calculation of three-body interaction strengths in six peptide-water clusters (with a total of 92290 three-body interactions), which is superior to AMOEBABIO18 polarizable force field method (correlation coefficient=0.9950, RMSE=10.74 kJ/mol with the MP2 method). Moreover, the computational efficiency is significantly improved, with calculation time reduced by approximately 50% compared to AMOEBABIO18 polarizable force field method in simulations involving clusters with over 20000 three-body interactions. This efficient method offers a new approach for large-scale simulations of many-body interactions in protein-water systems, demonstrating significant potential for applications in related fields.

Keywords Many-body polarization interaction; Chemical bond dipole; Induced bond dipole; Peptide-water system; Three-body interaction energy

蛋白质与水分子间的多体极化作用在生命活动中发挥着关键作用^[1-3]. Guo等^[4]和Wu等^[5]研究发现,多肽和蛋白质中广泛存在的氢键网络和协同效应对稳定蛋白质的二级结构起重要作用,而这种氢键协同效应的主要来源就是多体极化作用. Ejtehadi等^[6]采用分子动力学模拟研究了多体作用对蛋白质折叠速率和机制的影响,结果表明,考虑蛋白质与溶剂分子间的三体作用可显著增加模拟结果与实验值的相关性. Ghosh等^[7]通过分子动力学模拟研究了水溶液中的多肽,发现水分子与多肽之间的相互作用能引起多肽中酰胺键振动频率的波动. Li等^[8]使用中子散射方法和分子动力学模拟方法研究了蛋白质水体系,发现多体极化效应使不同蛋白质上的水分子具有普遍的亚扩散性. Yamasaki等^[9]发现,三聚体中多体效应的贡献极大地稳定了蛋白质结构,并将其归因于诱导极化效应. 上述研究表明,准确计算蛋白质、多肽与水分子间的三体作用强度对模拟蛋白质的物理和化学性质十分重要.

量子力学方法和分子力学方法均可通过结合多体展开法计算体系的三体极化作用能. Cisneros等^[10]使用MP2/aug-cc-pVTZ(BSSE)方法计算了约500个由水分子构成的三聚体的三体作用能,发现采用MP2/aug-cc-pVTZ(BSSE)方法计算得到的三体极化作用能与采用CCSD(T)/aug-cc-pVTZ(BSSE)方法计算结果的均方根误差(RMSE)仅为0.08 kJ/mol. 然而,高精度量子力学方法计算时间较长,模拟的空间和时间尺度有限,不适用于蛋白质等生物大分子体系. 分子力学方法中的固定电荷力场方法将原子电荷作为恒定参数处理,导致分子中各个原子上的电荷不能随外界环境的变化而变化,所以固定电荷力场方法无法描述分子体系中存在的极化作用,因而不能用于计算三体极化作用能^[10]. 极化力场方法包含了对极化作用的处理,可以较准确地描述环境变化所带来的极化效应,是预测三体极化作用能的重要工具^[11]. 极化力场中常用的极化处理包括诱导点偶极法、Drude Oscillator法和浮动电荷法^[12]. 以AMOEBABIO18可极化力场为代表的诱导点偶极法需要通过迭代求解矩阵方程获得原子的诱导偶极,较为耗时^[13]. Drude Oscillator法通过加入一个与原子核相连的Drude粒子来描述极化作用,粒子的偏离程度反映极化作用的强弱^[14,15]. 浮动电荷模型将原子点电荷视为动态变量,根据电负性重新分配,分子上的总电荷保持不变^[16,17].

近年来,本课题组^[18-22]建立了基于化学键偶极的可极化偶极-偶极作用模型并将之应用于计算包含氢键作用、堆积作用、T-shape和X—H \cdots π 作用的分子间相互作用强度,结果与高精度量子力学方法相当. 在此基础上,本文将该模型扩展应用于多肽-水体系中的多体极化作用强度的快速计算. 蛋白分子与环境水分子间的多体极化作用对蛋白分子的结构和性质有重要影响,肽键是蛋白分子的基本结构单元. 本文选取含有肽键的酰胺、甘氨酸和天门冬酰胺与水分子组成的体系为研究对象,将分子中的N—H和C=O极性化学键以及水分子的O—H化学键视为键偶极,键偶极矩的大小可随分子环境的变化而改变,通过构建键偶极和诱导键偶极之间的相互作用模型,提出了一种新的多体极化作用计算方法. 通过拟合训练集分子的三体作用能随分子间距离变化的势能曲线确定了该方法所需参数. 将本文方法和参数应用于快速计算二肽-水三聚体和多肽-水团簇的三体作用能,并与高精度量子力学方法的计算结果进行了比较,以检验本文方法和参数的合理性和可转移性,并进一步探讨其在更大规模的多肽-水体系中的应用潜力.

1 理论模型

对于多肽水体系, 本文将多肽分子中极性化学键以及水分子中O—H化学键视为键偶极, 键偶极矩的大小可随分子环境的变化而改变. 当孤立存在的多肽分子处于最稳定状态时, 其极性化学键的偶极矩被视为固有偶极矩; 然而, 当其周围环境发生变化时, 这些化学键会瞬时产生诱导偶极矩. 为描述多肽与水分子之间的极化相互作用, 本文利用片段间的固有偶极矩与诱导偶极矩以及诱导偶极矩与诱导偶极矩的相互作用之和描述多肽与水分子之间的极化相互作用, 具体如下式所示:

$$E_{\text{pol}}(i, j) = \sum_{\substack{l \in i \\ m \in j}} \frac{\mu_l \delta\mu_m + \delta\mu_l \mu_m + \delta\mu_l \delta\mu_m}{R_{ml}^3} (2\cos\theta \cos\theta' + \sin\theta \sin\theta' \cos\zeta) \quad (1)$$

式中: $E_{\text{pol}}(i, j)$ (kJ/mol) 为第 i 和第 j 个分子之间的极化作用能; μ_l 和 $\delta\mu_l$ (C·m) 分别为第 i 个分子中的第 l 个化学键的固有偶极矩和诱导偶极矩; μ_m 和 $\delta\mu_m$ (C·m) 分别为第 j 个分子中的第 m 个化学键的固有偶极矩和诱导偶极矩; θ (°), θ' (°), ζ (°) 和 R_{ml} (nm) 是与两个键偶极相对位置有关的几何参量. 以水分子O—H化学键偶极与多肽分子N—H化学键偶极间的相互作用来说明本文模型中对 θ , θ' , ζ 和 R_{ml} 几何参量的具体定义. 如图1所示, O—H化学键偶极矩位于O—H化学键上, 其键偶极矩的正方向为由氧原子指向氢原子方向; N—H化学键偶极矩位于N—H化学键上, 其键偶极矩正方向为由氮原子指向氢原子方向; A点为O—H化学键的中点, B点为N—H化学键的中点, R_{ml} 为两个化学键中点A和B间的距离, θ 和 θ' 分别是O—H和N—H化学键偶极正方向与线段AB间的夹角, ζ 是O—H键与线段AB所在平面和N—H键与线段AB所在平面间的二面角.

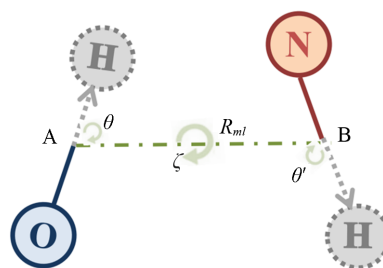


Fig. 1 Interaction schematic diagram of O—H and N—H dipoles

式(1)中化学键的诱导偶极矩 $\delta\mu$ 通过下式估算:

$$\delta\mu = c \cdot \frac{q - q_0}{q_0} \cdot \mu \quad (2)$$

式中: 对于O—H化学键偶极, q (e) 为O—H键上H的原子电荷, q_0 (e) 为参考电荷; 对N—H化学键偶极, q (e) 为N—H键上H的原子电荷, q_0 (e) 为参考电荷; 对C=O化学键偶极, q (e) 为C=O键上O的原子电荷, q_0 (e) 为参考电荷; c 是依赖于原子电荷计算方法的校正系数; μ (C·m) 为化学键的固有偶极矩. 本文使用AM1方法计算原子电荷 q , 几何参量 θ , θ' , ζ 和 R_{ml} 均很容易从体系的笛卡尔坐标计算得到. 使用式(1)计算极化作用能 E_{pol} , 需已知化学键的固有偶极矩 μ 、参考电荷 q_0 以及校正系数 c . 本文将化学键的固有偶极矩 μ 、参考电荷 q_0 以及校正系数 c 作为参数处理.

包含 N 个分子的团簇的总三体作用能 ($\text{IE}_{3\text{B}}$, kJ/mol) 为团簇中每个三聚体的三体作用能之和^[23,24].

$$\text{IE}_{3\text{B}} = \sum_{i \neq j \neq k} V_{3\text{B}}(i, j, k) \quad (3)$$

$$V_{3\text{B}}(i, j, k) = E(i, j, k) - [E(i, j) + E(i, k) + E(j, k)] + E(i) + E(j) + E(k) \quad (4)$$

式(3)中, $V_{3\text{B}}(i, j, k)$ (kJ/mol) 为由第 i, j, k 个分子所形成三聚体的三体作用能. 包含 N 个分子的团簇需计算 C_N^3 个三体作用, 每个三体作用能均可通过式(4)计算获得. 式(4)中, $E(i, j, k)$ (kJ/mol) 为由第 i, j, k 个分子构成的三聚体的总能量; $E(i, j)$ (kJ/mol) 为由第 i, j 个分子构成的二聚体的总能量; $E(i)$ (kJ/mol) 为第 i 个分子的总能量. 使用高精度量子力学方法可以计算得到三聚体能量 $E(i, j, k)$ 、二聚体能量 $E(i, j)$ 和单体能量 $E(i)$, 进而通过式(4)得到三聚体的三体作用能.

使用本文方法或可极化力场方法可以非常容易地计算得到三聚体、二聚体以及单体的极化作用能 E_{pol} , 进而通过下式获得三聚体的三体作用能.

$$V_{3\text{B}}(i, j, k) = E_{\text{pol}}(i, j, k) - [E_{\text{pol}}(i, j) + E_{\text{pol}}(i, k) + E_{\text{pol}}(j, k)] + E_{\text{pol}}(i) + E_{\text{pol}}(j) + E_{\text{pol}}(k) \quad (5)$$

式中: $E_{\text{pol}}(i, j, k)$ (kJ/mol) 为第 i, j, k 个分子所形成三聚体的极化作用能; $E_{\text{pol}}(i, j)$ 为第 i, j 个分子所形成二聚体的极化作用能; $E_{\text{pol}}(i)$ (kJ/mol) 为第 i 个分子的极化作用能.

本文方法通过计算化学键的固有偶极与诱导偶极之间的相互作用, 以及诱导偶极与诱导偶极之间的相互作用, 来获得极化作用能. 以甘氨酸二肽与两个水分子形成的三聚体为例, 水分子有 2 个 O—H 化学键, 甘氨酸二肽分子有 C=O, N—H, C—N, C—H 和 C—C 共计 18 个化学键. C=O, N—H, C—N, C—H 和 C—C 键的平均键偶极矩实验值分别为 8.01×10^{-30} , 4.37×10^{-30} , 1.50×10^{-30} , 1.00×10^{-30} 和 $0 \text{ C} \cdot \text{m}$ ^[25]. 使用本文方法计算其三体作用能时, 如果将所有化学键均视为键偶极, 则需计算 152 个化学键偶极-偶极相互作用. 综合考虑计算精度和计算效率, 本文不考虑酰胺和多肽分子中弱极性 C—H, C—N 键以及非极性 C—C 键的偶极矩, 仅考虑 C=O 和 N—H 键的偶极矩. 采用上述方法处理后, 本文方法计算该体系的三体作用能时, 只需计算 40 个化学键偶极-偶极相互作用, 这大大减少了计算量.

为了检验式(2)预测化学键诱导偶极矩的准确性, 分别使用 B3LYP/aug-cc-pVTZ^[26] 和本文方法计算了 441 个不同结构水分子单体的分子偶极矩. 本文方法通过将分子上极性化学键的固有偶极矩和诱导偶极矩的矢量加和得到水分子单体的分子偶极矩(图 2). 由图 2(C)可知, 使用本文方法计算 441 个水分子单体分子偶极矩与 B3LYP/aug-cc-pVTZ 方法的最大误差不超过 $0.032 \times 10^{-30} \text{ C} \cdot \text{m}$, 最大相对误差仅为 0.52%. 这一结果表明, 式(2)能够准确预测化学键的诱导偶极矩.

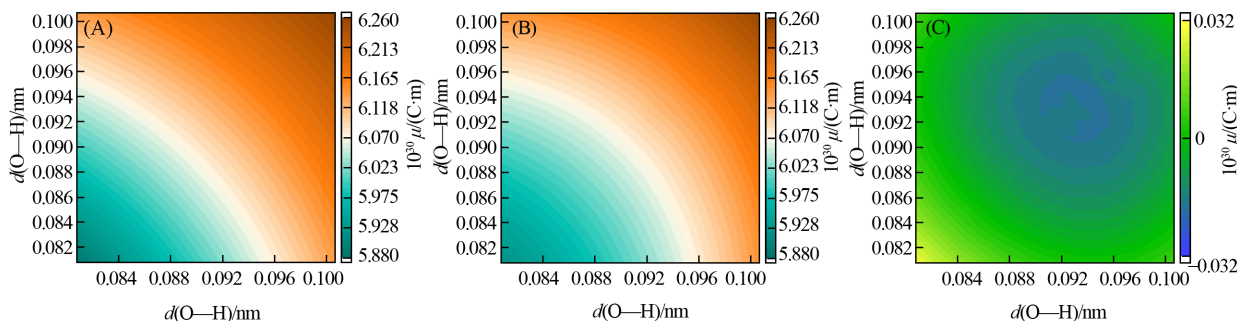


Fig. 2 Molecular dipole moments of water monomer calculated by using B3LYP/aug-cc-pVTZ method(A) and our method(B) and the difference between the results of the two methods

综上所述, 需要确定酰胺和多肽分子中 C=O 键和 N—H 键的 μ 、 q_0 和 c 值. 以实验报道的 C=O 和 N—H 键偶极矩为初始值^[25], 通过拟合酰胺分子的实验分子偶极矩, 得到了 C=O 键与 N—H 键的 μ 值. 表 1 列出了本文确定的 μ 值, 其中, 水分子中 O—H 键的固有偶极矩为文献[25]中的实验值. 进而使用表 1 中参数计算得到甲酰胺、*N*-甲基甲酰胺、乙酰胺和 *N*-甲基乙酰胺的分子偶极矩分别为 11.84×10^{-30} , 13.84×10^{-30} , 11.71×10^{-30} 和 $13.88 \times 10^{-30} \text{ C} \cdot \text{m}$, 与实验值 12.44×10^{-30} , 12.78×10^{-30} , 12.28×10^{-30} 和 $14.34 \times 10^{-30} \text{ C} \cdot \text{m}$ ^[27] 非常接近, 说明本文确定的固有偶极矩参数是合理的.

Table 1 Permanent bond dipole moment parameters determined by our method

Chemical bond	$10^{30} \mu / (\text{C} \cdot \text{m})$
C=O(in —CONH— group)	8.85
N—H[in —CONHX group, (X≠H)]	5.04
N—H(in —CONH ₂ group)	4.37
O—H(in H ₂ O)	5.04 ^[25]

使用 MP2/6-31+G(d, p) 方法优化得到了甲酰胺、*N*-甲基甲酰胺、乙酰胺、*N*-甲基乙酰胺和甘氨酸二肽的稳定构象, 再使用 AM1 方法计算了上述分子稳定构象的原子电荷, 从而确定了 q_0 . 在此基础上, 选取酰胺三聚体和甘氨酸二肽-水三聚体(图 3), 固定两个分子不动, 改变第 3 个分子与前两个分子间的距离 $R(\text{O} \cdots \text{H})$, 将 $R(\text{O} \cdots \text{H})$ 值分别取为平衡氢键键长的 0.90, 1.00, 1.10, 1.25 和 1.50 倍以及 0.5, 1.0 和 1.5 nm, 得到多个三体结构. 使用高精度 MP2/aug-cc-pVTZ(BSSE) 方法计算得到三体作用能随距离的变化曲线. 通过调整 C=O 键和 N—H 键的 c 值, 使用本文方法计算得到的三体作用能随距离变化曲线与高精度 MP2/aug-cc-pVTZ(BSSE) 方法的结果基本一致, 从而确定了 c 值(表 2).

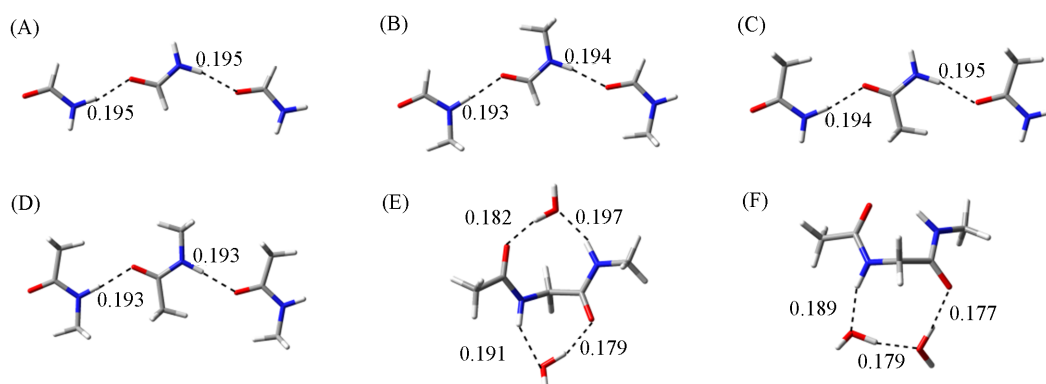


Fig. 3 Model molecules in training set

(A) Formamide trimer; (B) *N*-methylformamide trimer; (C) acetamide trimer; (D) *N*-methylacetamide trimer; (E) water-gly-water-line; (F) gly-water-water-cyc. White: hydrogen; red: oxygen; grey: carbon, blue: nitrogen. The hydrogen-bond distances are in nm.

Table 2 Reference charge and corrected factor for induced bond dipole

Atom	q_0/e	c
O[in the X—CONH—Y group, (X, Y≠H, Me)]	-0.3732	1.00
O[in the MeCONHX group, (X≠H)]	-0.3699	1.00
O[in the HCONHX group, (X≠H)]	-0.3645	0.90
O(in the —CONH ₂ group)	-0.3643	1.20
H[in the X—CONH—Y group, (X, Y≠H, Me)]	+0.2627	1.00
H(in the N—H of —CONHMe group)	+0.2247	1.00
H(in the —CONH ₂ group)	+0.2262	1.20
H(in water)	+0.1926	1.08

2 模型的应用

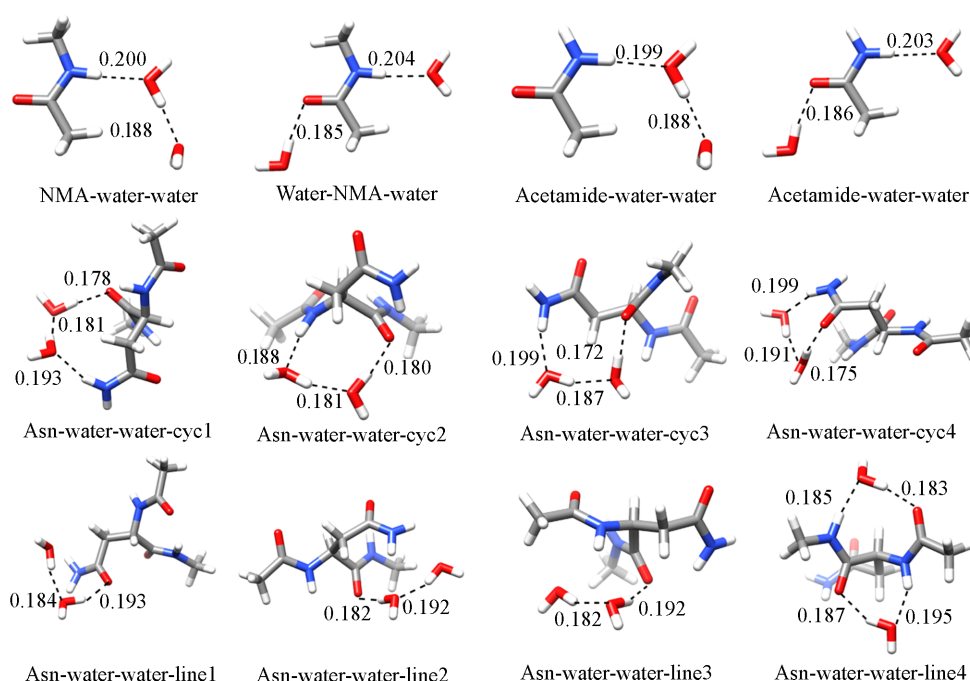
2.1 酰胺-水三聚体和二肽-水三聚体三体作用能的计算

2.1.1 稳定三聚体结构三体作用能的计算 极化作用有利于水环境中蛋白质的溶剂化现象^[28]. 蛋白质肽键上的C=O和N—H极性化学键易与周围的溶剂水分子上的O—H键形成氢键作用. 因此, 本文选择酰胺-水三聚体和二肽-水三聚体为研究对象, 采用MP2/6-31+G(*d,p*)方法优化得到了*N*-甲基乙酰胺-水稳定三聚体、乙酰胺-水稳定三聚体以及天门冬酰氨二肽-水稳定三聚体总计12个三聚体结构(图4, 坐标见本文支持信息表S1).

酰胺-水三聚体分别包含酰胺中C=O键与水分子形成氢键作用以及N—H键与水分子形成氢键作用的构象. 天门冬酰氨二肽-水稳定三聚体中主链的C=O和N—H以及侧链的C=O和N—H均与水分子上的O—H形成氢键作用, 这些结构的不单一性使得三体作用能的范围更加广泛, 可以充分验证上述极性化学键的相关参数.

分别采用MP2/aug-cc-pVTZ(BSSE)方法、M06-2X/aug-cc-pVTZ(BSSE)法和本文方法计算上述12个稳定三聚体的三体作用能. 由表3可知, 采用MP2/aug-cc-pVTZ(BSSE)方法计算的12个稳定三聚体结构中, 三体作用最强的是三聚体Asn-Water-Water-cyc1, 三体作用能为-13.77 kJ/mol; 三体作用最弱的是Water-NMA-Water, 三体作用能为-1.72 kJ/mol. Asn-Water-Water-cyc1中天门冬酰氨二肽侧链的N—H键和主链的C=O键分别与不同水分子形成氢键作用, 水分子之间也形成氢键作用, 每个分子均同时作为氢键供体和受体, 本文将此类三聚体结构称为环状三聚体. 三聚体Water-NMA-Water中*N*-甲基乙酰胺分子的极性化学键C=O键和N—H键分别与两个水分子形成分子间氢键作用, 水分子之间距离较远, 未形成氢键作用, *N*-甲基乙酰胺分子同时作为氢键供体和受体, 本文将此类三聚体结构称为折线型三聚体结构.

N-甲基乙酰胺分子和乙酰胺分子结构仅相差一个甲基基团, 与两个水分子形成稳定三聚体的方式


Fig. 4 Total 12 equilibrium structures of amide, peptide-water trimers

The hydrogen bond distance are in nm.

Table 3 Three-body interaction energies(V_{3B}) calculated by different methods for total 12 amide-water trimers and peptide-water trimers

Trimer	MP2/aug-cc-pVTZ(BSSE) method	M06-2X/aug-cc-pVTZ(BSSE) method		This work	
	$V_{3B}/(\text{kJ}\cdot\text{mol}^{-1})$	$V_{3B}/(\text{kJ}\cdot\text{mol}^{-1})$	$\Delta V_{3B}/(\text{kJ}\cdot\text{mol}^{-1})$	$V_{3B}/(\text{kJ}\cdot\text{mol}^{-1})$	$\Delta V_{3B}/(\text{kJ}\cdot\text{mol}^{-1})$
NMA-water-water	-4.02	-4.31	-0.29	-3.43	0.59
Water-NMA-water	-1.72	-1.72	0.00	-2.22	-0.50
Acetamide-water-water	-3.97	-4.27	-0.29	-3.81	0.17
Water-acetamide-water	-1.76	-1.80	-0.04	-2.64	-0.88
Asn-water-water-cyc1	-13.77	-13.85	-0.08	-12.34	1.42
Asn-water-water-cyc2	-13.43	-14.18	-0.75	-12.18	1.26
Asn-water-water-cyc3	-9.37	-10.38	-1.00	-9.37	0.00
Asn-water-water-cyc4	-7.91	-8.28	-0.38	-9.25	-1.34
Asn-water-water-line1	-6.78	-8.12	-1.34	-7.61	-0.84
Asn-water-water-line2	-5.52	-6.32	-0.79	-4.27	1.26
Asn-water-water-line3	-4.52	-5.44	-0.92	-5.31	-0.79
Asn-water-water-line4	-3.97	-3.93	0.04	-4.60	-0.63
Maximum absolute error			1.34		1.42
Root-mean square error			0.66		0.92
Mean absolute error			0.50		0.81

及分子间氢键也十分相似. 当 *N*-甲基乙酰胺分子和乙酰胺分子同时作为氢键供体和受体时, 其形成的稳定三聚体结构的氢键键长之差不超过 0.002 nm, 三体作用能之差也仅为 0.04 kJ/mol. 同理, NMA-water-water 和 Acetamide-water-water 中水分子同时作为氢键供体和受体时, 氢键键长也基本一致, 三体作用能之差仅为 0.05 kJ/mol. 上述结果说明, 忽略 C—H 化学键偶极是合理的.

天门冬酰氨二肽与两个水分子形成稳定三聚体的三体作用强度比 *N*-甲基乙酰胺分子和乙酰胺分子的更强, 这是由于天门冬酰氨二肽包含的极性化学键更多, 主链和侧链的极性化学键能够同时与水分子形成氢键作用, 分子之间的极化作用更强. 当多个氨基酸形成多肽或蛋白质时, 准确计算其与水分子之间的三体作用强度十分重要. 采用本文方法和 MP2/aug-cc-pVTZ(BSSE) 方法预测环状三聚体的三体作用强度较折线型三聚体更强. 以 MP2/aug-cc-pVTZ(BSSE) 方法的计算结果为基准, 采用本文方法计算 12 个稳定结构三体作用能的最大误差为 1.42 kJ/mol, 均方根误差为 0.92 kJ/mol, 平均绝对误差

为 0.81 kJ/mol, 线性相关系数为 0.95. 上述结果表明, 本文方法计算酰胺-水和二肽-水体系稳定三聚体的三体作用能时表现较好.

2.1.2 三体作用能随分子间距离的变化 在蛋白质的液态分子动力学模拟中, 根据牛顿运动方程可知, 蛋白质与溶剂水分子的相对位置是不断变化的. 因此, 预测蛋白质与水分子之间的三体作用能随分子间距离变化的势能曲线十分必要. 本文选择 *N*-甲基乙酰胺、乙酰胺、天门冬酰胺二肽分别与水分子形成的稳定三聚体为研究对象, 采用 MP2/aug-cc-pVTZ (BSSE) 方法和本文方法预测体系的三体作用能随分子间距离变化的势能.

如图 5 所示, *N*-甲基乙酰胺与水分子形成的稳定三聚体为折线型结构, 中心水分子同时作为氢键

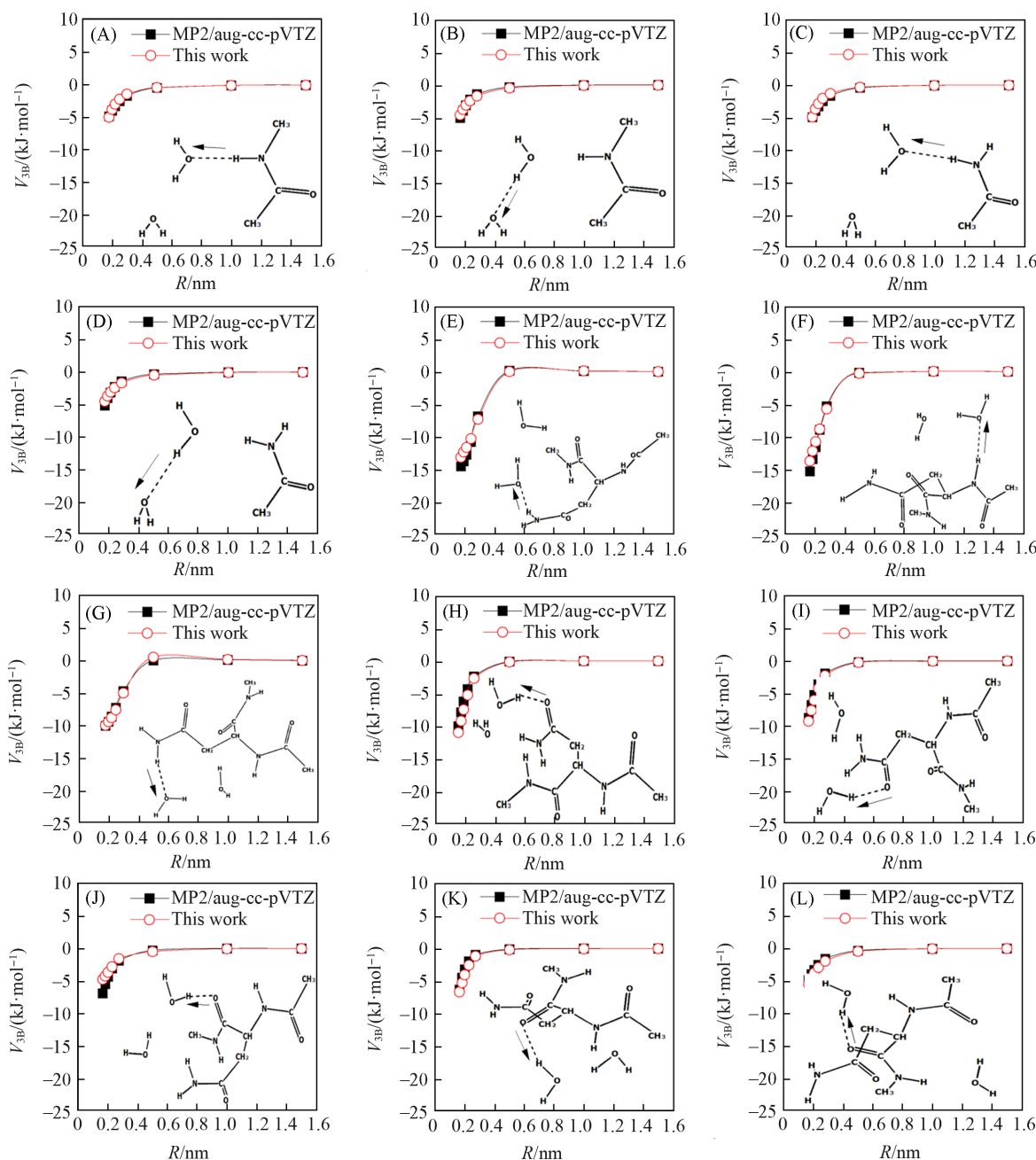


Fig. 5 Three-body interaction energies vs. the intermolecular distances

(A, B) NMA-water trimers, (C, D) acetamide-water trimers, (E—L) asparagine-water trimers. The planar structure schematic in (A—L) illustrates the spatial configuration of molecules through a two-dimensional representation, which provides a more straightforward visualisation of the intermolecular hydrogen bond of equilibrium trimers. The arrow indicates the direction in which the molecules have shifted.

供体和受体. 固定中心水分子的位置不变, 分别改变 *N*-甲基乙酰胺和另一个水分子的位置, 采用本文方法与 MP2/aug-cc-pVTZ(BSSE) 方法计算对应结构的三体作用能, 得到了两条三体作用能随分子间距离变化的势能曲线. 结果表明, 改变三聚体分子中 *N*-甲基乙酰胺或水分子的位置, 本文方法预测的势能曲线与 MP2/aug-cc-pVTZ(BSSE) 方法结果趋势一致且数值基本吻合. 乙酰胺计算结果与 *N*-甲基乙酰胺一致. 上述结果表明, 采用本文方法计算酰胺-水三聚体的三体作用能是合理的.

天门冬酰氨二肽与水分子形成 8 个稳定三聚体, 改变稳定三聚体中水分子的位置, 分别采用 MP2/aug-cc-pVTZ(BSSE) 方法和本文方法计算对应结构的三体作用能. 计算结果表明, 本文方法计算天门冬酰氨二肽-水三聚体的三体作用能随分子间距离变化的势能曲线与 MP2/aug-cc-pVTZ(BSSE) 方法的计算结果一致. 考虑量子力学方法的计算速度和计算成本, 分别采用 MP2/aug-cc-pVTZ(BSSE) 和 MP2/aug-cc-pVDZ(BSSE) 方法计算了上述 8 个天门冬酰氨二肽-水三聚体的三体作用能随分子间距离变化的势能曲线. 表 S2 (见本文支持信息) 列出了 MP2/aug-cc-pVTZ(BSSE) 和 MP2/aug-cc-pVDZ(BSSE) 方法计算天门冬酰氨二肽-水三聚体的三体作用能随分子间距离变化的结果. 由表 S2 可知, MP2/aug-cc-pVDZ(BSSE) 方法与 MP2/aug-cc-pVTZ(BSSE) 方法的误差在 -0.04~0.12 kJ/mol 之间. 因此, 本文采用 MP2/aug-cc-pVDZ(BSSE) 方法代替 MP2/aug-cc-pVTZ(BSSE) 方法计算多肽-水团簇的总三体作用能.

2.2 大团簇体系三体作用能的计算

在蛋白质的液态模拟中, 蛋白质的极性化学键可与溶剂水分子形成氢键网络, 在这些氢键网络中存在着大量的三体作用. 本文采用 Tinker 程序包^[29], 在 AMBER99SB 力场下, 使用 NPT(298 K, 1×10^5 Pa) 系综分别对甘氨酸二肽和水分子形成的体系以及天门冬酰氨二肽和水分子形成的体系进行了 2 ns 的分子动力学模拟, 步长设置为 1 fs, 分别保存 1 和 2 ns 时的结构. 选取上述结构中甘氨酸二肽和距离其最近的 30 个水分子形成 $[\text{NME-Gly-ACE} \cdot (\text{H}_2\text{O})_{30}]$ 团簇以及天门冬酰氨二肽和距离其最近的 50 个水分子形成 $[\text{NME-Asn-ACE} \cdot (\text{H}_2\text{O})_{50}]$ 团簇. 本文也将甘氨酸和天门冬酰氨进行了组合, 得到 NME-Gly-Asn-ACE 和 NME-Gly-Asn-Gly-ACE 分子, 再经过上述步骤分别进行动力学模拟, 得到了两个多肽与距离其最近的 50 个水分子形成的团簇. 上述 6 个由动力学模拟采样的随机结构坐标见本文支持信息表 S3.

图 6 给出了 4 个随机二肽-水团簇的结构, 1 个随机三肽-水团簇的结构和 1 个随机四肽-水团簇的结构, $[\text{NME-Gly-ACE} \cdot (\text{H}_2\text{O})_{30}]-1$ 为本文选取的第一个甘氨酸二肽和距离其最近的 30 个水分子形成的团簇, 其它团簇以此类推. 分别使用 MP2/aug-cc-pVDZ(BSSE) 方法、本文方法和 AMOEBA18 力场方

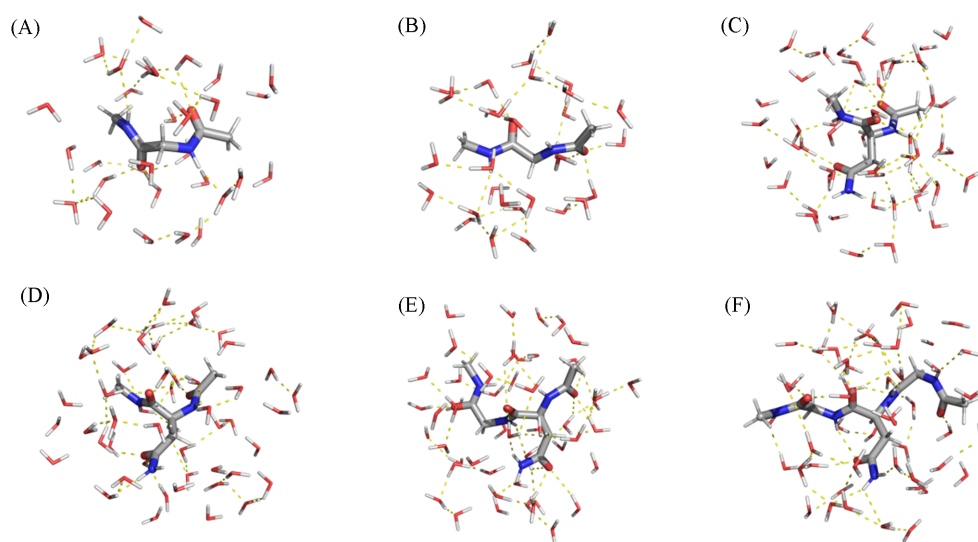


Fig. 6 Geometries of six random peptide-water clusters

(A) $[\text{NME-Gly-ACE} \cdot (\text{H}_2\text{O})_{30}]-1$; (B) $[\text{NME-Gly-ACE} \cdot (\text{H}_2\text{O})_{30}]-2$; (C) $[\text{NME-Asn-ACE} \cdot (\text{H}_2\text{O})_{50}]-1$;
(D) $[\text{NME-Asn-ACE} \cdot (\text{H}_2\text{O})_{50}]-2$; (E) $\text{NME-Gly-Asn-ACE} \cdot (\text{H}_2\text{O})_{50}$; (F) $\text{NME-Gly-Asn-Gly-ACE} \cdot (\text{H}_2\text{O})_{50}$.
White: hydrogen; red: oxygen; grey: carbon; blue: nitrogen.

法^[30,31]计算各个团簇的总三体作用能, 计算结果列于表4. 表4中还给出了各个团簇中的总三体作用数目.

Table 4 Three-body interaction energies(IE_{3B}) calculated by different methods for 6 peptide-water clusters

Cluster	Number of V_{3B} units	MP2/aug-cc-pVDZ(BSSE) method	This work		AMOEBABIO18 method	
		$IE_{3B}/(\text{kJ}\cdot\text{mol}^{-1})$	$IE_{3B}/(\text{kJ}\cdot\text{mol}^{-1})$	Δ	$IE_{3B}/(\text{kJ}\cdot\text{mol}^{-1})$	Δ
[NME-Gly-ACE·(H ₂ O) ₃₀]-1	4495	-36.28	-37.66	-1.38	-39.08	-2.80
[NME-Gly-ACE·(H ₂ O) ₃₀]-2	4495	-80.46	-82.05	-1.59	-84.14	-3.68
[NME-Asn-ACE·(H ₂ O) ₅₀]-1	20825	-159.91	-159.24	0.67	-166.27	-6.36
[NME-Asn-ACE·(H ₂ O) ₅₀]-2	20825	-105.23	-115.19	-9.96	-123.34	-18.12
NME-Gly-Asn-ACE·(H ₂ O) ₅₀	20825	-121.92	-130.25	-8.33	-136.15	-14.23
NME-Gly-Asn-Gly-ACE·(H ₂ O) ₅₀	20825	-187.90	-199.95	-12.05	-197.86	-9.96
Maximum absolute error				12.05		18.12
Root-mean square error				7.29		10.74
Mean absolute error				5.66		9.19
Pearson correlation coefficient				0.9965		0.9950

采用 MP2/aug-cc-pVDZ(BSSE)方法计算 [NME-Gly-ACE·(H₂O)₃₀]-2 团簇的三体作用能为 -80.46 kJ/mol, 本文方法和 AMOEBABIO18 力场方法算得的该团簇的三体作用能分别为 -82.05 和 -84.14 kJ/mol, 误差分别为 -1.59 和 -3.68 kJ/mol. 采用 MP2/aug-cc-pVDZ(BSSE)方法算得的 [NME-Asn-ACE·(H₂O)₅₀]-1 团簇的三体作用能为 -159.91 kJ/mol, 本文方法和 AMOEBABIO18 力场方法算得的该团簇的三体作用能分别为 -159.24 和 -166.27 kJ/mol, 误差分别为 0.67 和 -6.36 kJ/mol. 进一步用本文方法计算同时包含甘氨酸和天门冬酰氨的多肽与周围水分子团簇的三体作用能. 计算结果表明, 本文方法仍可较准确预测团簇的总三体作用能, 本文方法及参数表现出良好的可转移性.

采用本文方法计算的上述6个团簇的总三体作用能与采用 MP2/aug-cc-pVDZ(BSSE)方法计算结果的 Pearson 相关系数为 0.9965, 采用 AMOEBABIO18 力场方法计算的上述6个团簇的总三体作用能与 MP2/aug-cc-pVDZ(BSSE)方法计算结果的 Pearson 相关系数为 0.9950. 上述结果表明, 本文方法和 AMOEBABIO18 力场方法计算多肽-水体系的总三体作用能结果均与 MP2/aug-cc-pVDZ(BSSE)方法接近, 本文方法的计算精度要稍好于 AMOEBABIO18 力场方法.

2.3 计算效率

统计了基于多体展开法, 分别使用本文方法和 AMOEBABIO18 力场方法在 Intel(R) Xeon(R) CPU E5-2650 V4@2.20GHz 机器上, 采用单核计算4个二肽-水团簇的总三体作用能所需的CPU时间, 结果如表5所示.

Table 5 CPU time needed for calculation of the total three-body interaction energies

Cluster	Number of V_{3B} units	Time/s	
		This work	AMOEBABIO18 method
[NME-Gly-ACE·(H ₂ O) ₃₀]-1	4495	2040	4477
[NME-Gly-ACE·(H ₂ O) ₃₀]-2	4495	2042	4455
[NME-Asn-ACE·(H ₂ O) ₅₀]-1	20825	12321	24031
[NME-Asn-ACE·(H ₂ O) ₅₀]-2	20825	8992	20568

以 [NME-Asn-ACE·(H₂O)₅₀]-1 体系为例, 分别统计采用本文方法和 AMOEBABIO18 可极化力场方法计算每个三体作用所需 CPU 时间. 将计算 20825 个三体作用所需的 CPU 时间加和, 得到计算该体系三体作用的总 CPU 时间. 统计结果表明, 采用本文方法和 AMOEBABIO18 力场方法计算上述4个体系三体作用能的 CPU 时间之比相差不大, 本文方法的 CPU 时间基本为 AMOEBABIO18 力场方法的 50%.

3 结 论

提出了一种基于多肽分子中极性较强的 C=O 和 N-H 化学键偶极计算分子间极化效应的方法.

在计算精度方面,本文方法计算得到的三体作用能与高精度的MP2方法的计算结果高度一致,显示出良好的预测能力.在计算效率方面,本文方法通过降低计算复杂性,显著提高了计算效率.本文方法为大规模模拟多肽在水环境中的结构与功能提供了一个高效且可靠的计算工具.未来,随着更广泛的测试和优化,本文方法有望在蛋白-水体系的模拟研究领域,尤其在进行蛋白和水分子互动的长时间分子动力学模拟研究中发挥重要作用.

支持信息见 <http://www.cjcu.jlu.edu.cn/CN/10.7503/20240518>.

参 考 文 献

- [1] Ren P., Chun J., Thomas D. G., Schnieders M. J., Marucho M., Zhang J., Baker N. A., *Q. Rev. Biophys.*, **2012**, 45(4), 427—491
- [2] Liu C., Qi R., Wang Q., Piquemal J. P., Ren P., *J. Chem. Theory Comput.*, **2017**, 13(6), 2751—2761
- [3] Oshima H., Kinoshita M., *J. Chem. Phys.*, **2015**, 142(14), 145103
- [4] Guo H., Gresh N., Roques B. P., Salahub D. R., *J. Phys. Chem. B*, **2000**, 104(41), 9746—9754
- [5] Wu X. H., Zhang H., Wu Y. D., *Proteins: Struct., Funct., Bioinf.*, **2010**, 78(5), 1186—1194
- [6] Ejtehadi M. R., Ejtehadi S. P., Plotkin S. S., *Proc. Natl. Acad. Sci. USA*, **2004**, 101(42), 15088—15093
- [7] Ghosh A., Hochstrasser R. M., *Chem. Phys.*, **2011**, 390(1), 1—13
- [8] Li S., Tan P., Li J., Tang M., Hong L., *Phys. Rev. Res.*, **2022**, 4(2), L022003
- [9] Yamasaki K., Tsuzuki S., Tateno H., *J. Phys. Chem. B*, **2024**, 128(30), 7401—7406
- [10] Cisneros G. A., Wikfeldt K. T., Ojamäe L., Lu J., Xu Y., Torabifard H., Bartók A. P., Csányi G., Molinero V., Paesani F., *Chem. Rev.*, **2016**, 116(13), 7501—7528
- [11] Kaminski G. A., Stern H. A., Berne B. J., Friesner R. A., Cao Y. X., Murphy R. B., Zhou R., Halgren T. A., *J. Comput. Chem.*, **2002**, 23(16), 1515—1531
- [12] Cisneros G. A., Karttunen M., Ren P., Sagui C., *Chem. Rev.*, **2014**, 114(1), 779—814
- [13] Ponder J. M., Wu C., Ren P., Pande V. S., Chodera J. D., Schnieders M. J., Haque I., Mobley D. L., Lambrecht D. S., DiStasio R. A. Jr., Head-Gordon T., *J. Phys. Chem. B*, **2010**, 114(8), 2549—2564
- [14] Stukan M. R., Asmadi A., Abdallah W., *J. Mol. Liq.*, **2013**, 180, 65—69
- [15] Lemkul J. A., Huang J., Roux B., MacKerell A. D. Jr., *Chem. Rev.*, **2016**, 116(9), 4983—5013
- [16] Patel S., Brooks C. L., *J. Comput. Chem.*, **2004**, 25(1), 1—16
- [17] Patel S., MacKerell Jr A. D., Brooks C. L., *J. Comput. Chem.*, **2004**, 25(12), 1504—1514
- [18] Li S. S., Huang C. Y., Hao J. J., Wang C. S., *J. Comput. Chem.*, **2014**, 35(6), 415—426
- [19] Hao J. J., Wang C. S., *RSC Adv.*, **2015**, 5(9), 6452—6461
- [20] Gao X. C., Hao Q., Wang C. S., *J. Chem. Theory Comput.*, **2017**, 13(6), 2730—2741
- [21] Li X. L., Sun Y. J., Tang Y., Wang C. S., *Chem. J. Chinese Universities*, **2021**, 42(12), 3664—3671(李晓蕾, 孙云娇, 唐颖, 王长生. 高等学校化学学报, **2021**, 42(12), 3664—3671)
- [22] Li X. L., Li C. M., Zhu J. Y., Zhou Z., Hao Q., Wang C. S., *J. Comput. Chem.*, **2023**, 44(5), 677—686
- [23] Jankiewicz W., Podeszwa R., Witek H. A., *J. Chem. Theory Comput.*, **2019**, 14(10), 5079—5089
- [24] Góra U., Podeszwa R., Cencek W., Szalewicz K., *J. Chem. Phys.*, **2011**, 135(22), 224102
- [25] James G. S., *Lange's Handbook of Chemistry, 16th Edition*, McGraw-Hill, New York, **2005**, Chapter 1.8, 171
- [26] Hickey A. L., Rowley C. N., *J. Phys. Chem. A*, **2014**, 118(20), 3678—3687
- [27] Haynes W. M., *CRC Handbook of Chemistry and Physics, 97th Edition*, CRC Press, Boca Raton, **2016**, Chapter 9, 60—66
- [28] Jiao D., Golubkov P. A., Darden T. A., Ren P., *Proc. Natl. Acad. Sci. USA*, **2008**, 105(17), 6290—6295
- [29] Rackers J. A., Zhi W., Liu C., Laury M. L., Lagardere L., Schnieders M. J., Piquemal J. P., Ren P., Ponder J. M., *J. Chem. Theory Comput.*, **2018**, 14(10), 5273—5289
- [30] Shi Y., Xia Z., Zhang J., Best R., Wu C., Ponder J. M., Ren P., *J. Chem. Theory Comput.*, **2013**, 9(9), 4046—4063
- [31] Zhang C. S., Chao L., Jing, Z. F., Wu C. J., Piquemal J. P., Ponder J. M., Ren P., *J. Chem. Theory Comput.*, **2018**, 14(4), 2084—2108

(Ed.: Y, K, M)