科学通报 2018 年 第 63 卷 第 33 期：3427 ~ 3441
专题：理论计算化学 评 述
《中国科学》杂志社 SCIENCE CHINA PRESS

大分子和凝聚相体系的快速量子化学计算：普适的基于能量的分块方法的发展和应用

廖康，程正，李云志，赵东波，李伟*，黎书华*

南京大学化学化工学院，介观化学教育部重点实验室，理论与计算化学研究所，南京 210023
* 联系人，E-mail: wli@nju.edu.cn; shuhua@nju.edu.cn

2018-09-01 收稿，2018-09-30 修回，2018-09-30 接受，2018-11-08 网络版发表
国家自然科学基金(21873046, 21833002, 21473087)和中央高校基本科研业务费专项资金(020514380150)资助

摘要 大分子和凝聚相体系的量子化学计算是理论化学的挑战之一，为了处理实验研究中越来越复杂的体系，线性标度的量子化学方法发展一直是理论研究的热点．由于简单有效且易于推广，基于能量的分块方法近十几年来得到较快的发展．本文主要介绍本课题组的普适的基于能量的分块量子化学方法的发展和应用，包括大体系的基态能量、结构和性质、局域激发态的计算方法及周期性体系的算法．该方法有望应用于多种类型的大体系(包括团簇、超分子、生物体系、分子晶体和溶液等)的能量、结构、振动光谱、核磁化学位移及电子吸收光谱等性质的计算．

关键词 量子化学计算，大分子，凝聚相体系，线性标度，基于能量的分块方法

近十几年来，化学及其交叉学科经过迅速发展，实验上已经能够处理越来越复杂的体系(包括大分子和凝聚相体系等)．因此也需要量子化学能够准确高效地处理这些复杂的体系(如溶液、材料、生物大分子等)的能量、结构和性质等．传统的量子化学方法，包括以Hartree-Fock(HF)方法为基础的从头算方法和密度泛函理论(density functional theory, DFT)，能够对中小尺寸的体系给出较为准确的描述，由于这些方法的标度很高，很难推广到大体系的计算．如HF和DFT方法的标度是$N^{3\sim4}$(这里N是能够表示体系尺寸大小的量，如电子数)，从头算电子相关(也称post-HF)方法的标度更高，如常用的二阶多体微扰(second-order Møller-Plesset perturbation theory, MP2)方法[1]的标度是N^5，而更高精度的耦合簇单双激发(coupled cluster singles and doubles, CCSD)方法[2]以及包含非迭代三重激发矫正的CCSD(T)方法[3]分别是N^6和N^7，因此这些方法往往只能处理包含十几到上百个原子的中小体系．目前用于复杂体系的量子化学计算方法主要包括半经验的量子化学方法和组合量子力学和分子力学方法(combined quantum mechanics and molecular mechanics, QM/MM)[4,5]．其中半经验的量子化学方法用一些经验的参数代替分子轨道积分的计算，通常用于定性或半定量的研究．QM/MM方法对复杂体系(如生物酶)的活性中心使用量子力学方法处理，而对其余环境部分使用分子力学方法处理，然而对于某些体系，如蛋白质构象演化和分子聚集体的自组装，需要对整个体系进行全量子力学处理．

线性标度的量子化学方法为大体系的全量子计算提供了可能，该方法主要包括基于第一性原理的方法和分块的量子化学方法这两大类．基于第一性原理的方法主要包括线性标度的HF/DFT方法[6~10]和局域相关方法[11~15]，前者通过积分筛选技术、减少库仑矩阵和交换矩阵的计算量、避免Fock矩阵的对角化等来降低大体系的计算标度，后者利用局域分子轨

引用格式：廖康，程正，李云志，等．大分子和凝聚相体系的快速量子化学计算：普适的基于能量的分块方法的发展和应用．科学通报，2018，63: 3427–3441
Liao K, Cheng Z, Li Y Z, et al. Fast quantum chemistry calculations for large molecules and condensed-phase systems: The developments and applications of generalized energy-based fragmentation approach (in Chinese). Chin Sci Bull, 2018, 63: 3427–3441, doi: 10.1360/N972018-00907

© 2018《中国科学》杂志社

道(localized molecular orbital, LMO)或原子轨道(atomic orbital, AO)表象下电子对相关能的局域性实现了低标度计算. 分子中的簇(cluster-in-molecule, CIM)方法[16~18]属于局域相关方法, 该方法的基本思想是, 在局域轨道下体系的总电子相关能可以表示为占据轨道相关能贡献之和, 其中每个占据轨道相关能贡献可以通过包含部分占据轨道分子簇近似获得. 分块的量子化学方法包括基于密度矩阵的分块方法[19~26]和基于能量的分块方法[27~38]. 在基于密度矩阵的分块方法中, 总体系的密度矩阵可以通过一系列子体系的密度矩阵或局域轨道组装得到, 进而得到大体系的总能量或性质, 这类方法包括分而治之方法[19,25,39]、增长法[21,40]、局域轨道组装方法[26,41]等. 在基于能量的分块方法中, 目标大体系的总能量可以通过一系列子体系的能量线性组合得到, 这类方法思想简单但非常有效, 容易推广到能量导数的计算[27~38]. 基于能量的分块方法按照子体系的构造方式又可分为基于多体展开的方法和基于容斥原理的方法这两类, 前者使用相互独立的片段单元通过多体展开"自下而上"地依次构建单体、两体、三体等较大的子体系, 后者先构建相互重叠的原始子体系, 然后使用容斥原理"自上而下"地依次产生较小的衍生子体系. 基于能量的分块方法在2005年首次用于大体系的总能量计算[27~29], 然而由于子体系中忽略了远程原子的静电和极化作用, 早期的这类方法对于带电体系(如生物分子)和极性较强的体系(如水簇)计算精度不高. 2006年, 马晶课题组[31]对于带电基团使用整数电荷作为背景点电荷, 提高了带电体系的计算精度. 2007年, 黎书华课题组[38]提出了普适的基于能量的分块(generalized energy-based fragmentation, GEBF)方法, 在同类方法中首次引入了全背景点电荷, 把大体系的总能量表示为一系列嵌入在背景电荷中的子体系的能量组合, 从而大大提高了这类方法的对于一般体系的计算精度. 经过10多年的发展, GEBF方法可以用于多种类型的大分子和凝聚相体系的能量、结构优化、振动光谱和核磁等性质计算[38,42~48].

1 大体系的基态能量

1.1 基本原理

对于一个大体系, 假设可以在空间上将其划分为多个较小且相互不重叠的"块", 则该大体系的基态能量可以通过多体展开表示如下[27]:

$$E=\sum_{i}E_{i}+\sum_{i<j}E_{i,j}+\sum_{i<j<k}E_{i,j,k}+\sum_{i<j<k<l}E_{i,j,k,l}+\cdots, \quad (1)$$

其中E_i为单体能量, $E_{i,j}$, $E_{i,j,k}$, $E_{i,j,k,l}$分别为两体、三体和四体相互作用能. 能量分解的计算表明这些的相互作用能会随着块之间距离的增大而衰减[27], 分块方法就是通过处理近程块间相互作用而忽略远程块间相互作用, 从而使用较小的计算量近似求得体系的基态能量, 如基于密度矩阵的分块方法是忽略远程块之间的密度矩阵元贡献.

基于能量的分块方法的基本思想是, 目标大体系(S)可以表示为一系列(M个)子体系$\{S_m, m=1, 2, \cdots, M\}$的线性组合[38,44]:

$$S=\sum_{m}^{M}C_mS_m, \quad (2)$$

其中C_m是第m个子体系的组合系数. 这里需要注意的是子体系中与目标体系中其余部分有共价键(单键)断裂的部分需要用氢原子饱和, 从而避免悬挂键的出现. 然后目标体系的总能量可以近似表示为所有子体系能量$\{E_m, m=1, 2, \cdots, M\}$的线性组合:

$$E=\sum_{m}^{M}C_mE_m. \quad (3)$$

能量相关的函数$f(E)$也可以相应地表示为

$$f(E)=\sum_{m}^{M}C_mf(E_m), \quad (4)$$

因此这类方法被称为基于能量的分块方法. 这里f可以是能量对坐标、外电场、外磁场等的导数, 如能量对坐标的一、二阶导数分别对应于梯度和Hessian(可用于构型优化和频率计算), 对电场的一、二阶导数分别对应于偶极矩和极化率, 依次对磁矩和外磁场的二阶导数对应于核磁共振屏蔽张量. 因此基于能量的分块方法原理上容易推广到构型优化、振动光谱等多种分子性质的计算[42,43].

基于能量的分块方法[34,37]中, 子体系的构建通常有2种方式: 一种是根据"自上而下"的容斥原理算法; 另一种是利用"自下而上"的多体展开. 在容斥原理算法中[27,28,30,32,38,49], 首先根据目标体系产生一系列(N个)相互之间重叠的子体系$\{A_i, i=1, 2, \cdots, N\}$(称为原始子体系), 任意原子至少被包含在某个子体系中, 因此目标体系可以表示为所有子体系的并集:

$$S=A_1\cup A_2\cup\cdots\cup A_N. \quad (5)$$

根据容斥原理, 总体系可以表示为

$$S=\sum_i A_i-\sum_{i<j} A_i\cap A_j+\sum_{i<j<k} A_i\cap A_j\cap A_k-\cdots$$
$$+(-1)^{N-1}\sum_{i<j\cdots<N} A_i\cap\cdots\cap A_N, \tag{6}$$

经过合并重复的子体系, 就可以通过式(2)组合得到总体系.

在多体展开算法中[29,50~54], 目标体系先划分为N个不相连的块$\{F_i, i=1, 2, \cdots, N\}$, 总体系可以展开为单体、两体、三体等的组合:

$$S=\sum_i F_i+\sum_{i<j} F_{i,j}+\sum_{i<j<k} F_{i,j,k}+\sum_{i<j<k<l} F_{i,j,k,l}+\cdots. \tag{7}$$

这里两体和三体贡献对应的子体系可以表示为

$$F_{i,j}=F_i\cup F_j-F_i-F_j, \tag{8}$$

$$F_{i,j,k}=F_i\cup F_j\cup F_k-F_{i,j}-F_{i,k}-F_{j,k}-F_i-F_j-F_k. \tag{9}$$

四体或更高的多体可以类似得到, 经过合并子体系(单键断裂的部分加氢饱和), 也可以得到式(2).

普适的基于能量的分块(即GEBF)方法[38,44]在同类方法中首次将子体系嵌入背景电荷中, 从而可以考虑长程静电作用和极化作用, 其中对于任意子体系S_m, 在目标体系中的其余原子都表示为背景点电荷, 我们标记该包含背景电荷的子体系为$\tilde{S}_m$. 为了讨论方便, 我们在M个背景电荷嵌入(也称为静电嵌入)的子体系基础上, 构建一个目标体系中所有原子均表示为点电荷的子体系$\tilde{S}_{M+1}$, 其对应的系数为C_{M+1}, 进一步我们将目标体系表示为M+1个静电嵌入(包括纯电荷体系)的子体系的线性组合,

$$S=\sum_m^{M+1} C_m\tilde{S}_m. \tag{10}$$

如果将点电荷当作虚拟的“原子”, 这M+1个子体系均可以看作为虚拟“总体系”, 为了求和后只得到一个有效的“总体系”, 需要满足$\sum_m^{M+1} C_m=1$, 或$C_{M+1}=1-\sum_m^{M} C_m$. 由于公式(10)中前M项的组合保证了实际的原子只出现1次, 因此包含第M+1项后就抵消了由于背景电荷的引入导致的多余相互作用(除了额外的氢原子引入的误差). 与式(3)类似, 相应的目标体系的总能量可以近似表示为包含背景电荷的子体系能量(含点电荷的自能)$\{\tilde{E}_m, m=1,2,\cdots,M+1\}$ (其中$\tilde{E}_{M+1}$为纯电荷的自能)的线性组合[38,44]:

$$E=\sum_m^{M} C_m\tilde{E}_m+\left(1-\sum_m^{M} C_m\right)\sum_A\sum_{B>A}\frac{Q_AQ_B}{R_{AB}}, \tag{11}$$

其中Q_A表示在原子A位置的电荷. 公式(11)中的第二项被称为GEBF能量矫正方案, 该方案比直接的库伦矫正方案精度更高[55], 因此被多种基于能量的分块方法使用[55,56]. 这里背景电荷可以使用分子力学或量子力学电荷. 量子力学电荷有2种方式可以得到, 一是从未嵌入背景电荷的原始子体系的中心部分取电荷迭代一次得到, 二是通过子体系的线性组合得到:

$$Q_A=\sum_m^{M} C_mQ_A(S_m). \tag{12}$$

这两种方式结果差别不大, 但在原始子体系较小的时候通常使用后者. 在GEBF中, 我们使用了自然布居数分析(natural population analysis, NPA)[57,58]得到的自然电荷作为背景电荷, 尽管NPA电荷会高估分子的偶极矩从而影响长程相互作用, 但前期研究表明, GEBF方法使用NPA电荷能给出更为准确的结果[38], 可能的原因尚待未来进一步研究. 实际计算中, HF和DFT使用相同的计算水平下的电荷, 基于HF的电子相关计算使用HF水平的电荷.

1.2 分块和构建子体系算法

以图1(a)中一个包含6个氨基酸的小的肽链分子(序列DAADAA, 其中D为天冬氨酸、A为丙氨酸)为例, 描述GEBF构建子体系[38,43,44]的过程如下.

(ⅰ) 把目标大体系分成一系列互不重叠的块. 分块方式根据精度和计算量的需求可以不同, 可以自动或手动分块. 自动分块算法通常以某一个较小的官能团或小分子作为块, 如单个水分子, 乙烯基、苯基等, 可以切割单键(尽可能碳–碳单键), 但不能切割双键和三键, 对蛋白质我们通常切断肽键中的羰基碳和α碳之间的键, 对于含较大的侧链的氨基酸残基, 该侧链也可切断. 对于图1(a)中的六肽, 自动分块方法会在每个氨基酸的羰基碳和α碳之间切1次, 总共切6次, 如图1(a)中的实线和虚线均为自动分块断键部分, 从而产生7个块, 后面的步骤中会生产较多子体系. 为了在例子中尽可能以较少的子体系方便讨论, 我们合并了折叠处的2块, 图1(a)中的虚线部分为不断键, 从而将该分子分为6块, 如图1(b)所示, 其中块与块之间相连的部分实线表示共价单键, 虚线表示氢键.

(ⅱ) 给每一个块加上周围的环境块构建原始子体系. 通常用2个参数ζ和λ控制, 对每一个块(称中

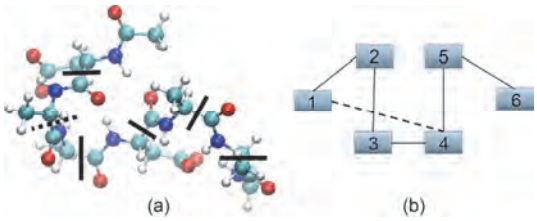

图 1 (网络版彩色)六肽(DAADAA)及其手动分块示意图. (a) 分子结构及断键, 粗实线表示切断键的部分, 点线表示在自动分块中会额外切的部分; (b) 方框示意图, 实线和虚线分别表示共价单键和氢键

Figure 1 (Color online) Hexapeptide (DAADAA) and its manual fragmentation scheme. (a) Molecular structure and the bond breaking. The bold solid line represents the places for the cut bond and the dotted line represents the additional cut bond in an automatic fragmentation. (b) The fragmentation scheme with box models. The solid and dotted lines represent covalent single bonds and hydrogen bonds, respectively

心块), 选取距离中心块小于ζÅ的块(两块中原子之间的最短距离)作为环境块. 如果中心块加环境块的总数超过λ, 只取最近的λ–1个块作为环境块. 在实际计算中, ζ通常取值3.0~4.0, λ一般取4~8, 通常这两个参数取值越大, 计算结果精度越高. 图2是以DAADAA的6个块为中心构建的6个原始子体系, 如对于第一个块, 由于块**2**与之直接键连, 块**3**块**4**与之在空间中的距离小于参数ζ, 因此**2**, **3**以及**4**都作为环境块和中心块**1**共同形成子体系(1234).

(ⅲ) 删除被完全包含在其他原始子体系中的冗余子体系. 在构建的原始子体系中, 如果一个相对较小的子体系完全被包含在另一个较大的子体系中, 由于前者中所有的多体相互作用项已经完全出现在后者中, 我们将删除前者, 相应的中心块也归属到后者中. 在六肽的原始子体系中, 我们注意到第2个子体系(123)被完全包含在第1个子体系(1234)中, 因此(123)将被删除, 同样子体系(234)和(56)也将被删除, 最后只保留3个原始子体系(1234), (345)和(456), 但(1234)中分配到3个中心块**1**, **2**和**3**(456)中得到2个中心块**5**和**6**, 如图3(a)中的前3个子体系.

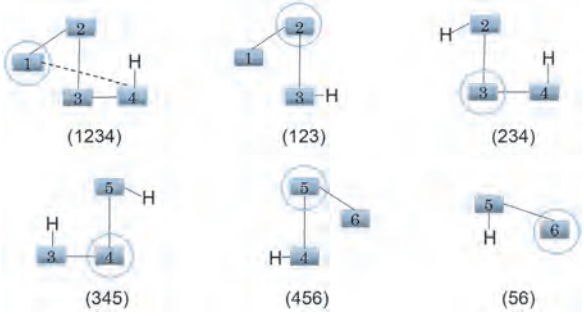

图 2 (网络版彩色)以每个块为中心构建的原始子体系(未合并). 圆圈里面表示的是中心块, 每个子体系包含的块的标号列在括号中

Figure 2 (Color online) Unmerged primitive subsystems, each of which is built with each fragment as center. The fragment inside the circle represents central fragment. The fragments contained in each subsystem are listed in parentheses

(ⅳ) 增加额外的三体和两体原始子体系(可选). 如果存在一些三块项, 两两之间的距离都较接近(小于1.5ζ), 如果该三体项没有被包含在前期产生的原始子体系中, 我们将以这3个块构建额外的原始子体系. 类似地, 如果存在两块项, 块间距离小于2ζ, 如果该两体项未被包含在之前构建的任何原始子体系中, 我们则以这2个块构建额外的原始子体系. 最终得到全部原始子体系$\{A_i, i=1, 2, \cdots, N\}$. 在六肽的体系中, 我们没有发现需要构建的三体原始子体系, 但发现块**1**和块**5**之间的距离小于2ζ, 且**1**, **5**未共同出现在前期产生的原始子体系中, 因此我们构建相应的两体原始子体系(15), 最终该体系共有4个保留的子体系, 相应的系数均为+1, 如图3(a)所示.

(ⅴ) 使用容斥原理产生全部子体系. 根据最终保留的原始子体系, 使用式(6)中的容斥原理算法, 就可以得到全部的子体系即相应的系数(+1或–1), 合并子体系后, 可以得到全部M个子体系$\{S_m, m=1, 2, \cdots, M\}$及系数. 如在六肽例子中, 使用容斥原理产生并合并后, 我们可以得到4个衍生子体系: (34), (45), (1)和(5), 对应的系数均为–1, 如图3(b)所示.

(ⅵ) 计算目标体系的电荷. 如前所述, 目标体系的电荷可以通过两种方式得到: 一是从包含中心块的原始子体系(未嵌入背景电荷)的计算中提取中心块的电荷得到; 二是计算所有的子体系(未嵌入背景电荷)的电荷后使用公式(12)得到. 两种方式都可以使用第1次得到的全部电荷构建静电嵌入的子体系迭代1次得到更准确的电荷. 在六肽的例子中, 可以从图3(a)中的前3个子体系的计算结果中提取中心块的电荷组合得到全部电荷, 也可以计算图3中所有子体系的电荷通过组合得到全部电荷.

(ⅶ) 构建静电嵌入的子体系. 得到目标体系的全部电荷后, 可以对每个子体系, 周围原本有其他原子的部分加上背景点电荷, 得到静电嵌入的子体系. 如在六肽的例子中, 对于子体系(1234), 我们在目标体系剩余原子位置(5, 6)加上背景点电荷, 得到相应的静电嵌入子体系(图4). 图中最后1个子体系(Q)表示包含全部电荷的纯电荷子体系, 根据公式(10)的讨

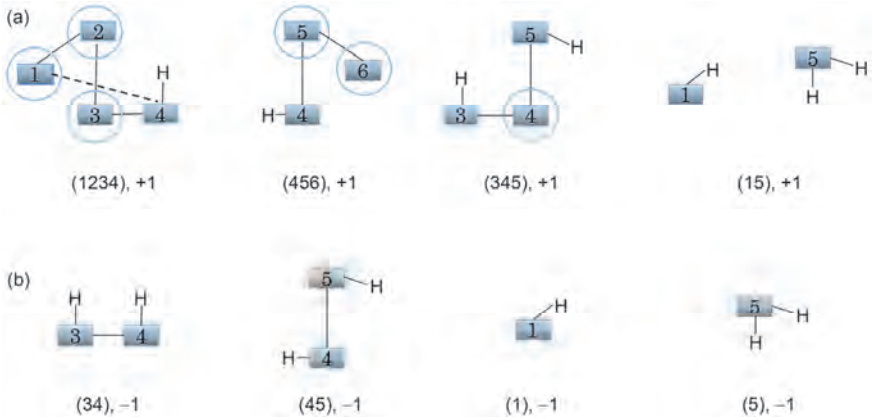

图 3 (网络版彩色)GEBF方法产生的子体系. (a) 删除冗余子体系和增加额外两体子体系后得到最终保留的原始子体系; (b) 使用容斥原理得到的衍生子体系. 圆圈中的表示中心块, 括号中数字为子体系中所包含的块, 后面的数字为该子体系的系数

Figure 3 (Color online) The GEBF subsystems. (a) The retained primitive subsystem after removing the redundant subsystems and adding the extra two-body subsystems; (b) the derivative subsystems using the inclusion-exclusion principle. The fragment in circle represents the central fragment. The fragments contained in each subsystem are listed in parentheses. The number outside the parentheses is the coefficient of the subsystem

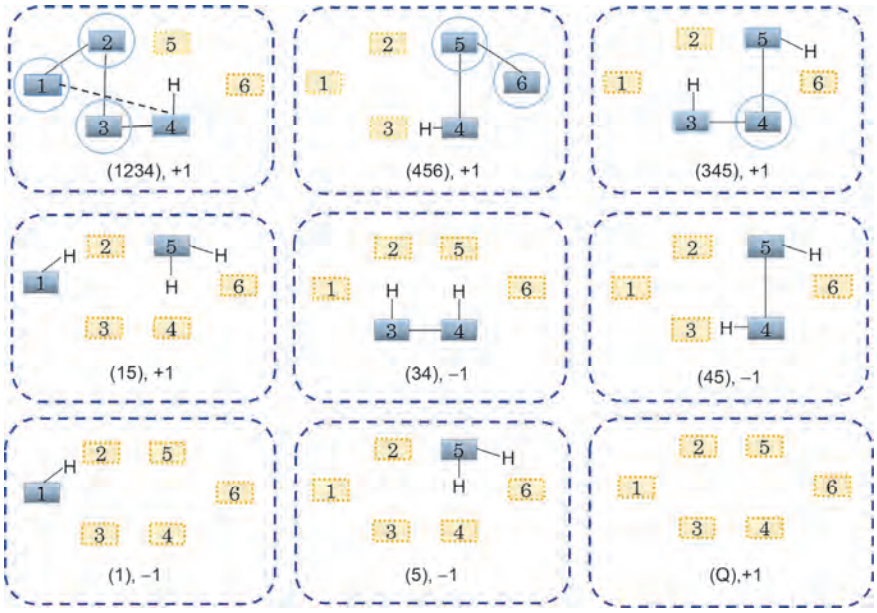

图 4 静电嵌入的子体系. 其中实线框表示块, 虚线框表示背景电荷. 圆圈中的表示中心块, 括号中数字为子体系中所包含的块, 后面的数字为该子体系的系数

Figure 4 Electrostatically embedded subsystems. The solid box represents the fragment, and the dotted box represents the background charges. The fragment in circle represents the central fragment. The fragments contained in each subsystem are listed in parentheses. The number outside the parentheses is the coefficient of the subsystem

论, 该电荷体系的系数为+1.

(ⅷ) 计算子体系并得到总体系能量. 构建好静电嵌入的子体系后, 就可以根据GEBF能量式(11)组合得到总体系的能量. 表1中列出了图4中所有子体系在B3LYP方法和6-311G(d,p)基组下的能量, 对9个子体系的能量通过其系数进行组合, 就可以得到总体系的GEBF能量(−2109.00552 a.u.), 相应的传统方法能量为−2109.00665 a.u., GEBF的偏差仅仅为0.71 kcal/mol, 在化学精度范围内.

1.3 应用算例

GEBF方法被广泛用于多种类型的大体系的基态能量(包括相对能量、结合能)计算, 包括团簇、超分子、生物体系等[38,43,59~67], 目前已经被包含在免费的

表 1 六肽(DAADAA)的子体系以及总体系的GEBF能量[a)]

Table 1 The subsystems of the Hexapeptide (DAADAA) and the GEBF energy

m	S_m	C_m	$\tilde{E}_m$	m	S_m	C_m	$\tilde{E}_m$
1	(1234)	+1	−1654.52152	2	(456)	+1	−1220.58293
3	(345)	+1	−1180.97200	4	(15)	+1	−727.54120
5	(34)	−1	−934.16354	6	(45)	−1	−1012.90692
7	(1)	−1	−479.62172	8	(5)	−1	−252.93539
9	(Q)	+1	−5.01535	合计	–	–	−2109.00552

a) 方法基组分别为B3LYP和6-311G(d,p), 能量单位为a.u., 传统B3LYP能量为−2109.00665 a.u.

LSQC程序中[68,69]. 通过与Gaussian[70,71]或Molpro程序[72]的接口, GEBF子体系的计算可以使用传统的电子结构方法, 包括HF, DFT, MP2, CCSD和CCSD(T)等, 或显式相关的电子结构方法[73,74], 包括MP2-F12和CCSD(T)-F12x(x=a,b)[75]等, 显式相关的F12方法由于在波函数中引入电子间距离相关的算符, 基组收敛比传统的电子相关方法更快, 结合中等大小的基组如aug-cc-pVDZ等就可以得到接近基组极限的结果.

图5是我们使用GEBF方法计算的部分极性或带较多电荷的体系, 这些体系使用不考虑背景电荷的分块方法难以准确描述. 图5(a)~(e)是包含水分子数目分别为20, 20, 32, 64和144的水团簇, 我们使用了高精度耦合簇方法计算了它们的相对能量或结合能[64~66]. 如2个水20簇异构体(图5(b))相对于(图5(a))的相对能量使用32种泛函在6-311++G(d,p)基组下得到的结果从−0.99变化到9.05 kcal/mol, 使用GEBF-CCSD(T)-F12a在aug-cc-pVDZ下得到的相对能量2.31 kcal/mol可以作为基准[64]. 目前在耦合簇水平下GEBF计算的最大的水簇为图5(e)中的水144簇, 使用GEBF-CCSD(T)-F12b方法在cc-pVDZ-F12基组(7000左右基函数)下计算了该体系与其他构象的相对能量和结合能, 并与不同泛函的GEBF-DFT比较, 研究发现不同泛函对于该大体系给出差异较大的结果, GEBF耦合簇方法可以作为泛函的标定[65,66]. 图5(f)是使用ONIOM模型计算的甲醇在水中的结合能, 其中内层(彩色)、中层(灰色)和外层(线型)分别使用了GEBF-MP2-F12, GEBF-MP2和半经验DFTB方法, 研究表明甲醇在水溶液中的长程作用达到10 Å左右, 因此需要使用大的团簇描述[63].

图5(g)是包含18个氨基酸残基(260个原子)的蛋白质(PDB ID: 1FGD)[76], 该蛋白质中含有8个电荷中心(1个正电荷, 7个负电荷), GEBF-HF在6-311G(d)基组下的能量为−7610.99529 a.u., 与传统的HF结果偏差仅为1.4 kcal/mol, 此外GEBF和传统方法得到的偶极矩也非常接近, 分别为144.71和144.78 Debye[38]. 图6(h)是包含18个碱基的DNA分子(PDB ID: 6ASF), 该体系包含570个原子, 16个负电荷中心, 以每个碱基(包含脱氧核糖)作为一块, 使用GEBF(4.0, 5)-M06-2X方法(考虑了额外的两体原始子体系)在6-31G(d)基组下得到的能量为−23776.29899 a.u., 与传统M06-2X结果偏差仅为2.1 kcal/mol. 因此, 由于使用背景电荷考虑了极化和长程作用, GEBF方法可以用来计算多种类型的大体系(包括极性体系和带较多电荷中心的体系)的基态能量.

2 大体系的结构和性质

2.1 能量导数的算法

在GEBF方法中, 目标体系的能量导数也可以通过一系列子体系的能量导数组合得到, 如能量对坐标的一阶导数(即梯度)可以近似表示为[44]

$$\frac{\partial E}{\partial \boldsymbol{q}_A}=\sum_m^M C_m\left(\frac{\partial \tilde{E}_m}{\partial \boldsymbol{q}_A}-\boldsymbol{F}_{m,a}Q_a-\sum_b \boldsymbol{f}_{ab}\right)+\left(\sum_m^M C_m-1\right)\sum_{b\in \mathrm{all}}\boldsymbol{f}_{ab}, \tag{13}$$

这里$\boldsymbol{q}_A$表示原子A的坐标, a, b表示在原子A, B位置的点电荷中心, $\boldsymbol{F}_{m,a}$为第m个子体系作用在中心a位置上的电场, $\boldsymbol{f}_{a,b}$表示电荷a与电荷b之间的库伦作用力, 最后一项中 $b\in all$ 表示求和变量b取目标体系对应的全部背景点电荷. 前期研究表明, GEBF能量梯度可以进一步近似为[43]

$$\frac{\partial E}{\partial \boldsymbol{q}_A}=\sum_m^M C_m\frac{\partial \tilde{E}_m}{\partial \boldsymbol{q}_A}. \tag{14}$$

有了目标体系的能量梯度, 可以实现大体系的几何构型优化和从头算分子动力学计算. 类似地, 能量二

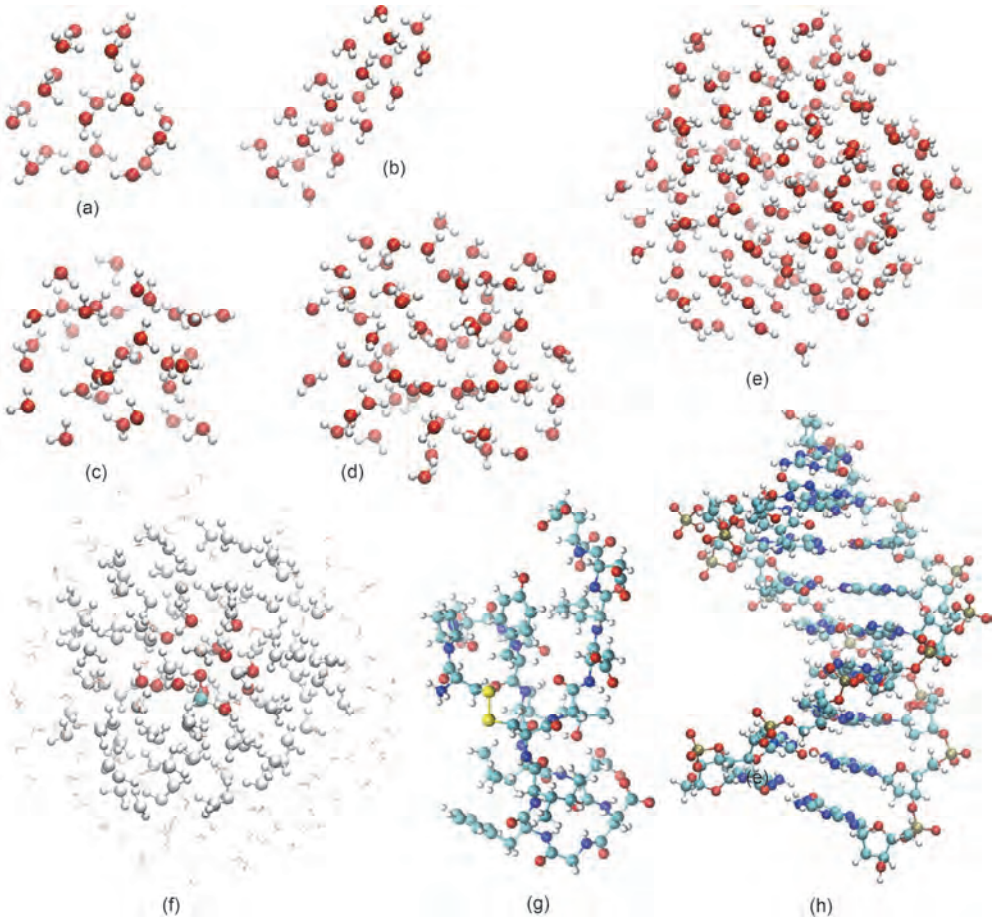

图 5 (网络版彩色)使用GEBF方法进行基态能量计算的部分团簇和大分子. (a), (b) 水簇$(H_2O)_{20}$的 2 种结构; (c)~(e) 水簇$(H_2O)_n$(n=32, 64, 144); (f) 甲醇在水溶液中的团簇(使用了多层计算模型); (g) 蛋白质(PDB ID: 1FGD); (h) 脱氧核糖核酸(PDB ID: 6ASF)

Figure 5 (Color online) Some clusters and macromolecules calculated with the GEBF method for the ground state energies. (a), (b) Two water 20-mers; (c)–(e) water 32-mer, 64-mer, and 144-mer; (f) the clusters of methanol in aqueous solution with a multi-layer model; (g) a protein (PDB ID: 1FGD); (h) A DNA molecule (PDB ID: 6ASF)

阶导数(Hessian)也可以近似表示为[43]

$$\frac{\partial^2 E}{\partial \boldsymbol{q}_A \boldsymbol{q}_B} = \sum_m^M C_m \frac{\partial^2 \tilde{E}_m}{\partial \boldsymbol{q}_A \boldsymbol{q}_B}. \tag{15}$$

该Hessian可以用于振动频率的计算, 从而实现大体系红外光谱、拉曼光谱、零点振动能、热力学性质(如焓和自由能)的计算.

通常情况下, 目标体系的性质如偶极矩、静态极化率、化学屏蔽张量、静态超极化率等(用Ω表示), 可以使用以下公式近似获得[38]:

$$\Omega = \sum_m^M C_m \tilde{\Omega}_m \quad (\Omega = \mu_i, \alpha_{ij}, \sigma_{ij}^A \cdots), \tag{16}$$

其中 $\tilde{\Omega}_m$ 为包含背景点电荷的第m个子体系的性质. 如在核磁计算中[48], σ_{ij}^A 是在原子核A上的化学屏蔽张量, 可以表示为能量对外磁场($\boldsymbol{B}$)和磁偶极矩($\boldsymbol{m}$)的二阶导数:

$$\sigma_{ij}^A = \frac{\partial^2 E}{\partial B_i \partial m_{A_j}}. \tag{17}$$

原子核A上的各向同性屏蔽常数可以表示为屏蔽张量矩阵的迹的1/3:

$$\sigma_{\text{iso}}^A = \frac{1}{3} \boldsymbol{tr}\left(\boldsymbol{\sigma}^A\right). \tag{18}$$

最后, 原子核A的化学位移可以表示为该各向同性屏蔽常数与参考化合物的屏蔽常数结果的差值:

$$\delta^A = \sigma_{\text{iso}}^A - \sigma_{\text{ref}}. \tag{19}$$

2.2 应用算例

GEBF方法可以在HF, DFT和MP2水平上进行大体系的几何构型优化、从头算动力学、红外、拉曼光

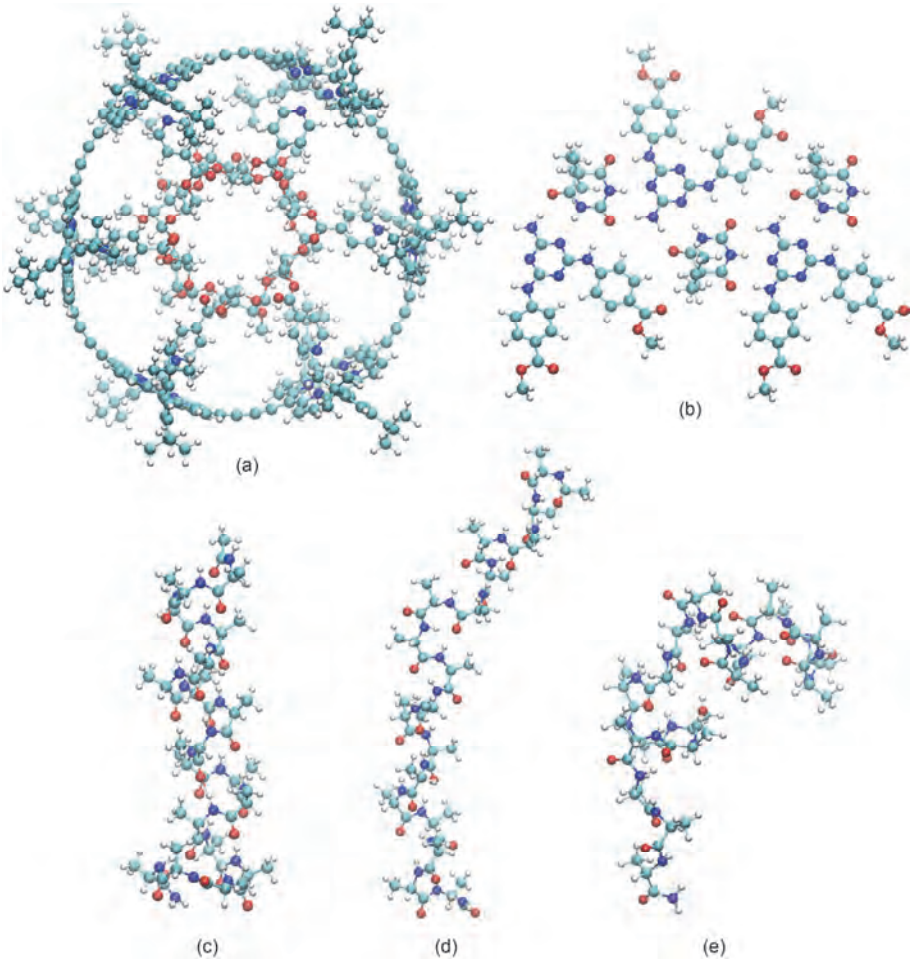

图 6 (网络版彩色)使用GEBF进行结构和性质计算的部分体系. (a) α环糊精模板的卟啉纳米环; (b) CA3·M3 超分子聚集体, 其中CA代表氰尿酸, M代表三聚氰胺; (c)~(e) 3_{10}螺旋 18 丙氨酸多肽分别使用GEBF-M06-2X方法、AMBER99 力场和半经验DFTB-D方法运行 5 ps动力学后的结构

Figure 6 (Color online) Some systems calculated by GEBF method for structures and properties. (a) α-cyclodextrin-templated porphyrin nanoring; (b) CA3·M3 supramolecular aggregate, in which CA represents cyanuric acid and M represents melamine; (c)–(e) the structures of 3_{10}-helix polyalanines, $AcA_{18}NH_2$, after 5-picosecond dynamics using GEBF-M06-2X, AMBER99, and semi-empirical DFTB-D methods, respectively

谱和核磁化学位移计算, 其中GEBF从头算动力学是通过LSQC与Tinker程序[77]结合实现的, 图6是我们使用GEBF进行结构和性质研究的3个体系, 对于图6(a)中的α环糊精模板的卟啉纳米环(包含582个原子), 使用GEBF-M06-2X方法在6-31G(d)基组(6126个基函数)上进行了构型优化, 使用16个CPU线程优化一步仅需要40 min, 而传统的M06-2X方法需要260 min. 此外对于该体系使用相同的方法在6-31G基组水平上进行了优化和红外光谱的计算, 同用16个线程的计算振动频率仅需15 h[68]. 图6(b)中的体系为氰尿酸和三聚氰胺的聚集体CA3·M3(CA代表氰尿酸, M代表三聚氰胺), 使用GEBF-B97-2在pcSseg-2基组下的NMR计算表明, 不同位置的H原子的化学位移与实验值基本一致, 平均偏差仅为0.6 ppm[48]. 图6(c)~(e)是以3_{10}螺旋18丙氨酸多肽为初始结构分别使用GEBF-M06-2X方法、AMBER99力场和半经验DFTB-D方法运行5 ps气相动力学后得到的结构, 这里为了加速动力学过程, 使用了600 K下的正则NVT系综. 图6(c)~(e)显示GEBF-M06-2X动力学仍能保持较为稳定的螺旋结构, 而AMBER99力场给出了解螺旋的结构, DFTB-D方法给出了蜷曲的结构[78]. 能量计算表明[78], GEBF预测的3_{10}螺旋和α螺旋的相对能量与大基组下的MP2相对能量非常接近, 远高于AMBER99力场的相对能量, 所以GEBF下的结构变化能垒可能也较高, 从而5 ps内能保持螺旋的动力学稳定. 因此, GEBF方法可以用来快速预测大体系的结构和性质.

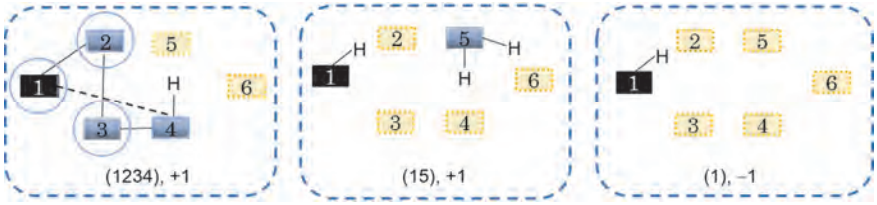

图 7 (网络版彩色)用于激发态计算的静电嵌入的活性子体系, 其中黑底白字块表示活性块, 实线框表示非活性块, 虚线框表示电荷. 圆圈中的表示中心块, 括号中数字为子体系中所包含的块, 后面的数字为该子体系的系数

Figure 7 (Color online) Electrostatically embedded active subsystems for excited-state calculations. The black block with white number represents the active fragment, the solid box represents the inactive fragment, and the dashed box represents the point charges. The fragment in circle represents the central fragment. The fragments contained in each subsystem are listed in parentheses. The number outside the parentheses is the coefficient of the subsystem

3 大体系的局域激发态

3.1 局域激发态方法

GEBF方法也被推广到了大体系的局域激发态计算[46], 在激发态算法中, 目标体系的局域激发中心被定义为活性区域, 相应的块被称为"活性块". 在构建子体系的时候, 包含活性块的子体系被称为活性子体系, 其余子体系被称为非活性子体系. 然后目标体系的激发态总能量E_{ES}可以表示为M_1个活性子体系的激发态能量和M_2个非活性子体系的基态能量线性组合[46], 这里所有子体系都嵌入在背景点电荷中, 因此还需要加上GEBF能量矫正如下:

$$E_{\mathrm{ES}}=\sum_{m}^{M_1}C_m\tilde{E}_m^*+\sum_{m}^{M_2}C_m\tilde{E}_m+\left(1-\sum_{m}^{M_1+M_2}C_m\right)\sum_{A}\sum_{B>A}\frac{Q_AQ_B}{R_{AB}}, \tag{20}$$

这里, $\tilde{E}_m^*$表示第m个静电嵌入的活性子体系的激发态能量, 其余符号同式(11). 这里需要注意的是在目标体系对应的背景电荷中, 活性块部分取的是激发态电荷, 其余块取的是基态电荷. 如对于图1中的六肽, 假设在第1块中有局域激发中心, 则我们定义块**1**为活性块, 图4中包含块**1**的子体系共有3个(图7)均为活性子体系, 图7中活性块用红色表示出来, 这些子体系将使用激发态方法计算. 有了目标体系的激发态能量, 就可以计算相应的电子吸收光谱, 这里需要指出的是振子强度是直接取自以活性块为中心块的静电嵌入原始子体系, 如图7中的(1234)体系.

3.2 应用算例

目前GEBF可以在含时密度泛函(time-dependent DFT, TDDFT)和近似的耦合簇CC2水平上进行大体系局域激发态计算, 其中CC2的计算使用Psi4程序[79]实现. 图8是我们使用GEBF-TDDFT研究的一些大体系,

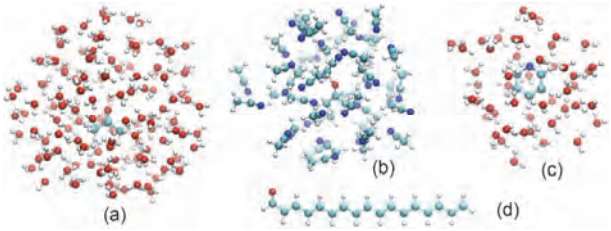

图 8 (网络版彩色)使用GEBF计算的部分局域激发体系. (a) 丙酮-水团簇; (b) 丙酮-乙腈团簇; (c) 吡啶-水团簇; (d) 正十七醛

Figure 8 (Color online) Some systems with localized excitation calculated by the GEBF method. (a) Acetone-water clusters; (b) acetone-acetonitrile clusters; (c) pyridine-water cluster; (d) normal 17-aldehyde

其中图8(a)~(c)分别为从丙酮在水溶液中、丙酮在乙腈溶液中和吡啶在水溶液中的MD模拟得到的团簇之一, 使用了GEBF-TD-ωB97XD方法在6-311++G(d,p)基组下对上百个团簇的激发态(丙酮的$n\rightarrow\pi^*$激发, 吡啶的$\pi\rightarrow\pi^*$激发)进行计算并进行统计平均, 得到这些溶质分子在从气相到溶液中发生的吸收光谱的位移(即溶致位移)分别为0.16, 0.06和−0.16 eV, 与相应的实验结果0.17, 0.06, −0.17 eV非常吻合[46]. 静电嵌入的QM/MM计算也能够较好地预测丙酮在溶液中的$n\rightarrow\pi^*$吸收光谱位移, 但低估了吡啶在水溶液中的$\pi\rightarrow\pi^*$吸收光谱位移. 图8(d)中的正十七醛的局域激发中心在醛基, 使用GEBF-TD-ωB97XD方法在不同的基组下得到的吸收光谱都与传统的TD-ωB97XD计算结果非常接近[46]. 因此, GEBF方法可以用于大体系的局域激发态计算, 从而得到这些体系的电子吸收光谱.

4 周期性凝聚相体系的计算

4.1 周期性GEBF算法

为了处理凝聚相体系如分子晶体和溶液, 结合周期性边界条件(periodic boundary condition, PBC),

GEBF方法被推广到适用于周期性体系的PBC-GEBF算法[45,47,80]. 该方法中, 为了精确描述凝聚相晶胞受其相邻镜像的影响, 目标晶胞被放置于由该晶胞形成的较大的超胞的中心. 超胞的尺寸通常取决于目标晶胞的尺寸, 对于较小尺寸的晶体, 需要选取7×7×7超胞, 而对于较大尺寸的晶体或溶液体系, 可选择较小的5×5×5或3×3×3超胞. 在PBC-GEBF框架下, 周期性体系的晶胞能量可以表示为[45,47,80]

$$E_{\text{unit-cell}}=\sum_{m}^{M}C_m\tilde{E}_m-\left(\sum_{m}^{M}C_m\right)\sum_{A\in K}\sum_{(B>A)\in K}\frac{Q_AQ_B}{R_{AB}}+E_{\text{Ewald}}, \tag{21}$$

其中, $\tilde{E}_m$ 表示在超胞K中嵌入于有限背景点电荷中的第m个子体系的总能量(包含电荷的自能), C_m, Q_A, R_{AB}的意义与非周期性GEBF式(11)中的一致, E_{Ewald}是通过Ewald求和方法得到的晶胞中电荷的经典静电相互作用:

$$E_{\text{Ewald}}=\frac{1}{2}\sum_{n}\sum_{A,B}Q_AQ_B\frac{\operatorname{erfc}\left(\alpha r_{AB,n}\right)}{r_{AB,n}}+\frac{2\pi}{V}\sum_{\boldsymbol{k}\neq 0}\frac{\mathrm{e}^{-\boldsymbol{k}^2/4\alpha^2}}{\boldsymbol{k}^2}\left(\sum_{A}Q_A\mathrm{e}^{i\mathbf{k}\cdot\mathbf{r}_A}\right)^2-\frac{\alpha}{\sqrt{\pi}}\sum_{A}Q_A^2, \tag{22}$$

其中, $r_{AB,n}$表示原始晶胞中的原子A和晶胞$n=(n_1, n_2, n_3)$中的原子B之间的距离, $\boldsymbol{k}$表示倒空间中的矢量, α是为了平衡实空间项和倒空间项的计算量的预设正参数, V是晶胞的体积.

PBC-GEBF中, 晶胞能量对原子坐标和对晶格平移矢量的一阶导数分别可以表示为[45,47]

$$\frac{\partial E_{\text{unit-cell}}}{\partial \boldsymbol{q}_{A,0}}=\sum_{i\in\text{Box}}\left[\sum_{m}C_m\boldsymbol{G}_{A,i,m}+\left(\sum_{m}C_m\right)\sum_{j\in\text{Box},b\in\text{all}}\boldsymbol{f}_{ai,bj}\right]+\frac{\partial E_{\text{Ewald}}}{\partial \boldsymbol{q}_{A,0}}, \tag{23}$$

$$\frac{\partial E_{\text{unit-cell}}}{\partial \boldsymbol{L}}=\sum_{i\in\text{Box}}i_L\left[\sum_{m}C_m\sum_{A}\mathbf{G}_{A,i,m}+\left(\sum_{m}C_m\right)\sum_{a}\sum_{j\in\text{Box},b\in\text{all}}\boldsymbol{f}_{ai,bj}\right]+\frac{\partial E_{\text{Ewald}}}{\partial \boldsymbol{L}}, \tag{24}$$

其中

$$\boldsymbol{G}_{A,i,m}=\frac{\partial\tilde{E}_m}{\partial \boldsymbol{q}_{A,i}}-\boldsymbol{F}_{m,a,i}Q_a-\sum_{j\in\text{Box},b}\boldsymbol{f}_{ai,bj}. \tag{25}$$

在式(23)~(25)中, 除了部分符号同公式(13)一致外, i代表晶胞指标, $\boldsymbol{L}$表示晶胞矢量, $\boldsymbol{q}_{A,i}$表示晶胞i中原子A的坐标, $\boldsymbol{F}_{m,a}$为第m个子体系作用在晶胞i的中心a位置上的电场, $\boldsymbol{f}_{ai,bj}$表示晶胞i中电荷a与晶胞j中电荷b之间的库伦作用力, $b\in all$ 表示b取超胞中的电荷中心位置. 与GEBF方法类似, 式(23)也可以近似为

$$\frac{\partial E_{\text{unit-cell}}}{\partial \boldsymbol{q}_{A,0}}=\sum_{i\in\text{Box}}\sum_{m}C_m\frac{\partial\tilde{E}_m}{\partial \boldsymbol{q}_{A,i}}. \tag{26}$$

通过晶胞能量对于原子坐标和晶格平移矢量的一阶导数, 就可以优化分子晶体的结构.

进一步, 为了实现PBC-GEBF下的周期性体系的振动光谱的计算[47,80], 目标体系的Γ点动力学矩阵可以表示为子体系的质量权重的Hessian矩阵的组合并对晶格指标求和:

$$D\left(\boldsymbol{q}_A,\boldsymbol{q}_B\right)=\frac{1}{\sqrt{m_Am_B}}\sum_{i}\sum_{m}C_m\frac{\partial^2\tilde{E}_m}{\partial \boldsymbol{q}_{A,0}\boldsymbol{q}_{B,i}}. \tag{27}$$

使用该动力学矩阵求解以下方程:

$$|D(\boldsymbol{q}_A,\boldsymbol{q}_B)-\omega^2\boldsymbol{I}|=0 \tag{28}$$

就可以得到周期性体系振动频率和相应的振动模.

周期性体系的其他一些性质(如偶极矩和极化率)对原子坐标的导数也可以类似表示为

$$\frac{\partial\Omega}{\partial \boldsymbol{q}_A}=\sum_{m}C_m\frac{\partial\Omega_m}{\partial \boldsymbol{q}_A}\quad(\Omega=\mu,\alpha,\cdots). \tag{29}$$

使用这些导数, 就可以计算红外和拉曼光谱的强度等. 类似地, 周期性体系中原子的核磁化学位移也可以求得.

4.2 周期性体系的子体系

在PBC-GEBF方法框架下[45,47,80], 周期性体系晶胞的能量或能量导数也是通过一系列静电嵌入的子体系的能量或导数的线性组合得到, 其中子体系的构建比非周期性的体系稍复杂. 对于周期性体系, 对单个晶胞中的体系进行分块, 与非周期性体系的子体系构建方法的区别主要包括: 第(2)步构建原始子体系时, 对于目标晶胞中每个块, 环境块可以取超胞中的块; 第(5)步产生衍生子体系时, 由于不同子体系的块分布在不同的盒子里, 因此需要在程序上判断位于不同位置但等价的衍生子体系; 第(7)步构建静电嵌入的子体系时, 需要把子体系周围分布在超胞中的电荷全部作为背景点电荷, 并且需要在超胞的边界使用补偿点电荷以保持单胞的偶极矩为零, 从而合理模拟晶体环境中的长程静电作用. 以图9(a)中的尿素分子晶体为例, 图9(b)是使用PBC-GEBF以

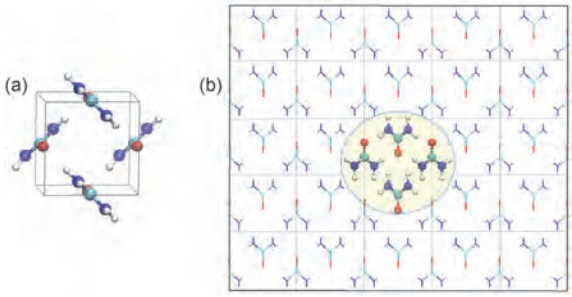

图 9 (网络版彩色)PBC-GEBF原始子体系示例. (a) 尿素分子晶体; (b) 静电嵌入的原始子体系之一, 其中用线状模型表示的尿素分子为该子体系在 5×5×5 的超胞内的点电荷

Figure 9 (Color online) Illustration of PBC-GEBF primitive subsystem. (a) Molecular crystal of urea; (b) an electrostatically embedded primitive subsystems. The wireframe represents the point charges on the urea molecules in a 5×5×5 supercell

5×5×5的超胞构建的静电嵌入的原始子体系, 其中线状模型表示点电荷. 计算所有PBC-GEBF产生的静电嵌入的子体系, 就可以计算通过式(21), (23), (24)等计算晶胞的能量和能量导数等.

4.3 应用算例

PBC-GEBF方法可以用于凝聚相体系, 如分子晶体、离子液体晶体、溶液等的能量计算、晶体结构预测和振动光谱等性质的计算[45,47,65,80,81]. 图10是使用PBC-GEBF方法研究的部分体系, 其中图10(a)是盒子包含64个水分子的冰的I_h相晶体, 图10(b)是在(a)溶解后的周期性水, 我们使用了PBC-GEBF-CCSD(T)-F12b方法计算了包含(a)和(b)的冰融化过程中的10个结构的相对能量和结合能. 研究表明, 由于相变中的体系微观结构变化较大, 不同的泛函给出差异较大的结果, PBC-GEBF结合高精度电子相关方法可以给出合理的描述[65]. 图10(c)是使用PBC-GEBF-MP2方法在6-311++G(d,p)基组下优化得到的甲胺硝酸盐([$MeNH_3$][NO_3])离子液体晶体结构[47], 在表2中, 我们将该晶体优化后的部分N-O距离(d_1, d_2和d_3)、键角(θ)、晶胞参数与相应的实验值进行了比较. 通过表2可以发现, 对于3个N-O距离的偏差均小于0.05 Å, O–N–N键角的距离小于1°, 晶胞参数中的盒子边长和角度偏差分别不超过0.05 Å和0.2°, 体积(V)的偏差小于0.05%. 由于传统的PBC-MP2很难用于晶体的优化, PBC-GEBF方法有望结合晶体结构搜索算法, 广泛应用于分子材料和溶液等凝聚相体系的结构和性质计算中.

5 结论和展望

在普适的基于能量的分块方法中, 大分子或凝聚相体系的总能量(及能量导数)可以表示为一系列静电嵌入的子体系的能量和电荷能量矫正项(及相应的能量导数)的线性组合, 从而克服了传统量子化学方法的标度困难, 可以在HF, DFT, MP2, MP2-F12, CCSD(T)及CCSD(T)-F12等水平上实现线性标度的量子化学计算. 该方法已经可以用于团簇、超分子、生物体系、溶液和分子晶体等体系的基态能量(包括相对能量、结合能、晶格能)、偶极矩、静态极化率、静态超极化率、几何构型优化、从头算分子动力学、振动光谱(红外和拉曼光谱)、热力学性质、核磁化学位移、电子吸收光谱等. 未来, 需要进一步改进GEBF和PBC-GEBF算法, 提高计算精度和速度, 如引入长程块之间的偶极相互作用和色散相互作用, 从而更准确地描述长程相互作用能; 扩展到更多分子性质的计算, 如实现大体系的发射光谱、分子晶体和溶液的动力学模拟等; 进一步完善相应的程序, 将该方法推广应用到更多类型的大分子和凝聚相体系的计算和模拟.

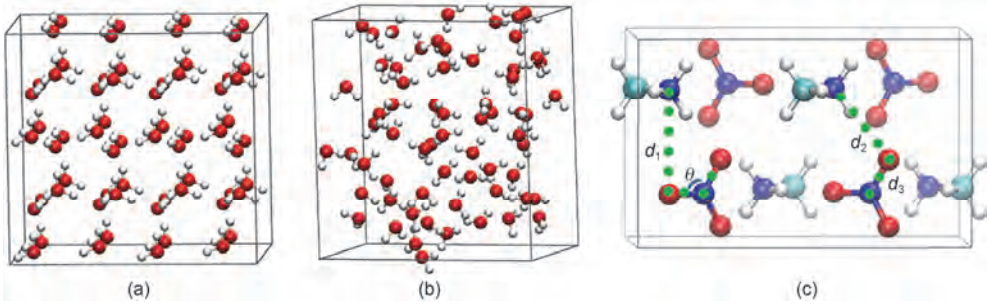

图 10 (网络版彩色)使用PBC-GEBF计算的部分周期性体系. (a) 冰I_h晶体; (b) 周期性水; (c) 甲胺硝酸盐离子液体晶体

Figure 10 (Color online) Some periodic systems calculated by PBC-GEBF method. (a) Ice-I_h crystal; (b) periodic water liquid; (c) methyl ammonium nitrate ionic liquid crystal

表 2 使用PBC-GEBF-MP2 优化得到甲胺硝酸盐晶体的部分N–O距离(d_1, d_2和d_3)、键角(θ)、晶胞参数并与相应的实验值的偏差,计算中的基组为 6-311++G**

Table 2 Comparison of some N–O distances (d_1, d_2, d_3), bond angle (θ), and lattice parameters of optimized methyl ammonium nitrate ionic liquid crystal at the PBC-GEBF-MP2/6-311++G** level with the corresponding experimental values

类型	实验值	PBC-GEBF	偏差
d_1(Å)	3.286	3.330	0.044
d_2(Å)	2.885	2.925	0.040
d_3(Å)	1.233	1.256	0.023
θ(°)	120.5	121.3	0.8
a(Å)	10.408	10.452	0.044
b(Å)	6.570	6.558	–0.012
c(Å)	6.404	6.391	–0.013
α(°)	90.0	90.1	0.1
β(°)	90.0	90.2	0.2
γ(°)	90.0	90.2	0.2
V(Å^3)	437.9	438.1	0.2

参考文献

1 Møller C, Plesset M S. Note on an approximation treatment for many-electron systems. Phys Rev, 1934, 46: 618–622

2 Purvis G D, Bartlett R J. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. J Chem Phys, 1982, 76: 1910–1918

3 Raghavachari K, Trucks G W, Pople J A, et al. Reprint of: A fifth-order perturbation comparison of electron correlation theories. Chem Phys Lett, 2013, 589: 37–40

4 Warshel A, Karplus M. Calculation of ground and excited state potential surfaces of conjugated molecules. I. Formulation and parametrization. J Am Chem Soc, 1972, 94: 5612–5625

5 Warshel A, Levitt M. Theoretic studies of enzymatic reactions: Dielectric electrostatic and steric stabilization in the reaction of lysozyme. J Mol Biol, 1976, 103: 227–249

6 Li X P, Nunes R W, Vanderbilt D. Density-matrix electronic-structure method with linear system-size scaling. Phys Rev B, 1993, 47: 10891–10894

7 Strain M C, Scuseria G E, Frisch M J. Achieving linear scaling for the electronic quantum Coulomb problem. Science, 1996, 271: 51–53

8 Schwegler E, Challacombe M, Head-Gordon M. Linear scaling computation of the Fock matrix. II. Rigorous bounds on exchange integrals and incremental Fock build. J Chem Phys, 1997, 106: 9708–9717

9 Scuseria G E. Linear scaling density functional calculations with Gaussian orbitals. J Phys Chem A, 1999, 103: 4782–4790

10 Millam J M, Scuseria G E. Linear scaling conjugate gradient density matrix search as an alternative to diagonalization for first principles electronic structure calculations. J Chem Phys, 1997, 106: 5569

11 Pulay P. Localizability of dynamic electron correlation. Chem Phys Lett, 1983, 100: 151–154

12 Förner W, Ladik J, Otto P. Coupled-cluster studies. II. The role of localization in correlation calculations on extended systems. Chem Phys, 1985, 97: 251–262

13 Saebo S, Pulay P. Fourth-order Møller-Plessett perturbation theory in the local correlation treatment. I. Method. J Chem Phys, 1987, 86: 914–922

14 Hampel C, Werner H. Local treatment of electron correlation in coupled cluster theory. J Chem Phys, 1996, 104: 6286–6297

15 Ayala P Y, Scuseria G E. Linear scaling second-order Møller-Plesset theory in the atomic orbital basis for large molecular systems. J Chem Phys, 1999, 110: 3660–3671

16 Li S, Shen J, Li W. An efficient implementation of the “cluster-in-molecule” approach for local electron correlation calculations. J Chem Phys, 2006, 125: 74109

17 Li W, Piecuch P, Gour J R, et al. Local correlation calculations using standard and renormalized coupled-cluster approaches. J Chem Phys, 2009, 131: 114109

18 Li S, Li W, Ni Z. Cluster-in-molecule local correlation method for post-Hartree-Fock calculations of large systems. Mol Phys, 2016, 114: 1447–1460

19 Yang W, Lee T S. A density-matrix divide-and-conquer approach for electronic structure calculations of large molecules. J Chem Phys, 1995, 103: 5674–5678

20 Kitaura K, Ikeo E, Asada T, et al. Fragment molecular orbital method: An approximate computational method for large molecules. Chem Phys Lett, 1999, 313: 701–706

21 Gu F L, Aoki Y, Korchowiec J, et al. A new localization scheme for the elongation method. J Chem Phys, 2004, 121: 10385–10391

22 Exner T E, Mezey P G. The field-adapted ADMA approach: Introducing point charges. J Phys Chem A, 2004, 108: 4301–4309

23 He X, Zhang J Z H. A new method for direct calculation of total energy of protein. J Chem Phys, 2005, 122: 31103

24 Akama T, Kobayashi M, Nakai H. Implementation of divide-and-conquer method including hartree-fock exchange interaction. J Comput Chem, 2007, 28: 2003–2012

25 Yang W. Direct calculation of electron density in density-functional theory. Phys Rev Lett, 1991, 66: 1438–1441

26 Li W, Li S. A localized molecular-orbital assembler approach for Hartree-Fock calculations of large molecules. J Chem Phys, 2005, 122: 194109

27 Li S, Li W, Fang T. An efficient fragment-based approach for predicting the ground-state energies and structures of large molecules. J Am Chem Soc, 2005, 127: 7215–7226

28 Deev V, Collins M A. Approximate *ab initio* energies by systematic molecular fragmentation. J Chem Phys, 2005, 122: 154102

29 Huang L, Massa L, Karle E. Kernel energy method illustrated with peptides. Int J Quantum Chem, 2005, 103: 808–817

30 He X, Zhang J Z H. The generalized molecular fractionation with conjugate caps/molecular mechanics method for direct calculation of protein energy. J Chem Phys, 2006, 124: 184703

31 Jiang N, Ma J, Jiang Y. Electrostatic field-adapted molecular fractionation with conjugated caps for energy calculations of charged biomolecules. J Chem Phys, 2006, 124: 114112

32 Ganesh V, Dongare R K, Balanarayan P, et al. Molecular tailoring approach for geometry optimization of large molecules: Energy evaluation and parallelization strategies. J Chem Phys, 2006, 125: 104109

33 Bettens R P A, Lee A M. A new algorithm for molecular fragmentation in quantum chemical calculations. J Phys Chem A, 2006, 110: 8777–8785

34 Raghavachari K, Saha A. Accurate composite and fragment-based quantum chemical models for large molecules. Chem Rev, 2015, 115: 5643–5677

35 Gao J, Cembran A, Mo Y. Generalized X-pol theory and charge delocalization states. J Chem Theory Comput, 2010, 6: 2402–2410

36 Gordon M S, Fedorov D G, Pruitt S R, et al. Fragmentation methods: A route to accurate calculations on large systems. Chem Rev, 2012, 112: 632–672

37 Collins M A, Bettens R P A. Energy-based molecular fragmentation methods. Chem Rev, 2015, 115: 5607–5642

38 Li W, Li S, Jiang Y. Generalized energy-based fragmentation approach for computing the ground-state energies and properties of large molecules. J Phys Chem A, 2007, 111: 2193–2199

39 Kobayashi M, Imamura Y, Nakai H. Alternative linear-scaling methodology for the second-order Møller-Plesset perturbation calculation based on the divide-and-conquer method. J Chem Phys, 2007, 127: 74103

40 Liu K, Korchowiec J, Aoki Y. Intermediate electrostatic field for the generalized elongation method. ChemPhysChem, 2015, 16: 1551–1556

41 Guo Y, Li W, Li S. An improved localized molecular-orbital assembler approach for Hartree-Fock calculations of general large molecules. Chem Phys Lett, 2012, 539-540: 186–190

42 Hua W, Fang T, Li W, et al. Geometry optimizations and vibrational spectra of large molecules from a generalized energy-based fragmentation approach. J Phys Chem A, 2008, 112: 10864–10872

43 Hua S, Hua W, Li S. An efficient implementation of the generalized energy-based fragmentation approach for general large molecules. J Phys Chem A, 2010, 114: 8126–8134

44 Li S, Li W, Ma J. Generalized energy-based fragmentation approach and its applications to macromolecules and molecular aggregates. Acc Chem Res, 2014, 47: 2712–2720

45 Fang T, Li W, Li S. Accurate prediction of lattice energies and structures of molecular crystals with molecular quantum chemistry methods. J Chem Theory Comput, 2015, 11: 91–98

46 Li W, Li Y, Lin R, et al. Generalized energy-based fragmentation approach for localized excited states of large systems. J Phys Chem A, 2016, 120: 9667–9677

47 Fang T, Li Y, Li S. Generalized energy-based fragmentation approach for modeling condensed phase systems. Wiley Interdiscip Rev Comput Mol Sci, 2017, 7: e1297

48 Zhao D, Song R, Li W, et al. Accurate prediction of NMR chemical shifts in macromolecular and condensed-phase systems with the generalized energy-based fragmentation method. J Chem Theory Comput, 2017, 13: 5231–5239

49 Mayhall N J, Raghavachari K. Molecules-in-molecules: An extrapolated fragment-based approach for accurate calculations on large molecules and materials. J Chem Theory Comput, 2011, 7: 1336–1343

50 Dahlke E E, Truhlar D G. Electrostatically embedded many-body expansion for simulations. J Chem Theory Comput, 2008, 4: 1–6

51 Richard R M, Herbert J M. A generalized many-body expansion and a unified view of fragment-based methods in electronic structure theory. J Chem Phys, 2012, 137: 64113

52 Rezac J, Salahub D R. Multilevel fragment-based approach (MFBA): A novel hybrid computational method for the study of large molecules. J Chem Theory Comput, 2010, 6: 91–99

53 Wen S, Nanda K, Huang Y, et al. Practical quantum mechanics-based fragment methods for predicting molecular crystal properties. Phys Chem Chem Phys, 2012, 14: 7578

54 Wang B, Yang K R, Xu X, et al. Quantum mechanical fragment methods based on partitioning atoms or partitioning coordinates. Acc Chem Res, 2014, 47: 2731–2738

55 Isegawa M, Wang B, Truhlar D G. Electrostatically embedded molecular tailoring approach and validation for peptides. J Chem Theory Comput, 2013, 9: 1381–1393

56 Pruitt S R, Addicoat M A, Collins M A. The fragment molecular orbital and systematic molecular fragmentation methods applied to water clusters. Phys Chem Chem Phys, 2012, 14: 7752

57 Foster J P, Weinhold F. Natural hybrid orbitals. J Am Chem Soc, 1980, 102: 7211–7218

58 Reed A E, Weinstock R B, Weinhold F. Natural population analysis. J Chem Phys, 1985, 83: 735–746

59 Dong H, Hua S, Li S. Understanding the role of intra- and inter-molecular interactions in the formation of single- and double-helical structures of aromatic oligoamides: A computational study. J Phys Chem A, 2009, 113: 1335–1342

60 Yang Z, Hua W, Li S. Low-lying structures and stabilities of large water clusters: Investigation based on the combination of the AMOEBA potential and generalized energy-based fragmentation approach. J Phys Chem A, 2010, 114: 9253–9261

61 Hua S, Li W, Li S. Cooperativity in long α- and 3_{10}-helical polyalanines: Both electrostatic and van der Waals interactions are essential. J Phys Chem B, 2011, 115: 11462–11469

62 Hua S, Li W, Li S. The generalized energy-based fragmentation approach with an improved fragmentation scheme: Benchmark results and illustrative applications. ChemPhysChem, 2013, 14: 108–115

63 Li W. Linear scaling explicitly correlated MP2-F12 and ONIOM methods for the long-range interactions of the nanoscale clusters in methanol aqueous solutions. J Chem Phys, 2013, 138: 14106

64 Wang K, Li W, Li S. Generalized energy-based fragmentation CCSD(T)-F12a method and application to the relative energies of water clusters $(H_2O)_{20}$. J Chem Theory Comput, 2014, 10: 1546–1553

65 Zhang L, Li W, Li S. Accurate relative energies and binding energies of large ice-liquid water clusters and periodic structures. J Phys Chem A, 2017, 121: 4030–4038

66 Yuan D, Li Y, Ni Z. Benchmark relative energies for large water clusters with the generalized energy-based fragmentation method. J Chem Theory Comput, 2017, 13: 2696–2704

67 Li Y, Yuan D, Wang Q. Accurate prediction of the structure and vibrational spectra of ionic liquid clusters with the generalized energy-based fragmentation approach: Critical role of ion-pair-based fragmentation. Phys Chem Chem Phys, 2018, 20: 13547–13557

68 Li W, Chen C, Zhao D, et al. LSQC: Low scaling quantum chemistry program. Int J Quantum Chem, 2015, 115: 641–646

69 Li S, Li W, Jiang Y, et al. LSQC Program, Version 2.3. Nanjing: Nanjing University, 2018

70 Frisch M J, Trucks G W, Schlegel H B, et al. Gaussian09 Revision B.01, Gaussian Inc. Wallingford CT, 2009

71 Frisch M J, Trucks G W, Schlegel H B, et al. Gaussian16 Revision A.03, Gaussian, Inc., Wallingford CT, 2016

72 Werner H J, Knowles P J, Knizia G, et al. Molpro: A general-purpose quantum chemistry program package. Comput Mol Sci, 2012, 2: 242–253

73 Ten-no S. Initiation of explicitly correlated Slater-type geminal theory. Chem Phys Lett, 2004, 398: 56–61

74 Kutzelnigg W. r12-Dependent terms in the wave function as closed sums of partial wave amplitudes for large l. Theor Chim Acta, 1985, 68: 445–469

75 Adler T B, Knizia G, Werner H J. A simple and efficient CCSD(T)-F12 approximation. J Chem Phys, 2007, 127: 221106

76 Berman H M. The protein data bank. Nucleic Acids Res, 2000, 28: 235–242

77 Ponder J W, Richards F M. An efficient Newton-like method for molecular mechanics energy minimization of large molecules. J Comput Chem, 1987, 8: 1016–1024

78 Zhang L, Li W, Fang T, et al. *Ab initio* molecular dynamics with intramolecular noncovalent interactions for unsolvated polypeptides. Theor Chem Acc, 2016, 135: 34

79 Turney J M, Simmonett A C, Parrish R M, et al. Psi4: An open-source *ab initio* electronic structure program. Comput Mol Sci, 2012, 2: 556–565

80 Fang T, Jia J, Li S. Vibrational spectra of molecular crystals with the generalized energy-based fragmentation approach. J Phys Chem A, 2016, 120: 2700–2711

81 Li Y, Wang G, Li W, et al. Understanding the polymorphism-dependent emission properties of molecular crystals using a refined QM/MM approach. Phys Chem Chem Phys, 2017, 19: 17516–17520

Summary for “大分子和凝聚相体系的快速量子化学计算: 普适的基于能量的分块方法的发展和应用”

Fast quantum chemistry calculations for large molecules and condensed-phase systems: The developments and applications of generalized energy-based fragmentation approach

Kang Liao, Zheng Cheng, Yunzhi Li, Dongbo Zhao, Wei Li* & Shuhua Li*

Institute of Theoretical and Computational Chemistry, Key Laboratory of Mesoscopic Chemistry of Ministry of Education, School of Chemistry and Chemical Engineering, Nanjing University, Nanjing 210023, China

* Corresponding authors, E-mail: wli@nju.edu.cn; shuhua@nju.edu.cn

It is a challenge in theoretical chemistry to perform the quantum chemistry calculations for large molecules and condensed-phase systems due to the high scalings of the conventional quantum chemistry methods. In order to treat more and more complicated systems in experiments, the development of linear scaling quantum chemistry methods is an active area in theoretical studies. There are two categories of these methods. One is the category of first-principle base methods, including linear scaling Hartree-Fock (HF) and density functional theory (DFT) methods and local correlation methods. Another is the category of fragment-based methods, including density matrix based methods and energy-based fragmentation methods. The energy-based fragmentation approach has been developed rapidly in the last decade due to its simplicity, effectiveness, and extendibility.

In this article, we review the developments and applications of the generalized energy-based fragmentation (GEBF) approach proposed in our research group. The basic idea of the GEBF approach is that the total ground-state energy of a large system can be linearly combined by the corresponding ground-state energies of various subsystems, each of which is embedded in the background charges on the places of those atoms outside the subsystem. In the GEBF approach, the target system is divided into various medium-sized fragments, each of which (called central fragment) is capped by its neighboring fragments to form a primitive subsystem. Then, the inclusion-exclusion principle is applied to all primitive subsystems to generate the derivative subsystems. The point charges, from natural population analysis, on all atoms are iteratively obtained by extracting the charges on the central fragments in primitive subsystems or combining the charges on all subsystems. Then the energies or energy derivatives of the target system could be combined by the corresponding values of all electrostatically embedded subsystems.

The GEBF approach could be applied to the ground-state energies (including the relative energies or binding energies) of a broad range of large systems including molecular clusters, supramolecular systems, proteins, nucleic acids, and etc. The computational levels include HF, DFT, second-order Møller-Plesset perturbation theory (MP2), coupled cluster singles and doubles (CCSD), CCSD with triples correction [CCSD(T)], and explicitly correlated MP2-F12 and CCSD(T)-F12x (x=a,b) methods. With the GEBF energy derivatives (including energy gradients, Hessians, and so on), the approaches could be employed for optimizing the molecular geometries, performing *ab initio* molecular dynamics, computing the vibrational spectroscopies (including the IR and Raman spectroscopies) and nuclear magnetic resonance (NMR) chemical shifts of large systems at the HF, DFT, and MP2 levels. GEBF approach is also extended to the localized excited states by defining those subsystems including local excitation as the active subsystems, which are treated by the excited-state calculations at the time-dependent DFT or approximate equation-of-motion CCSD (CC2) levels. Then the GEBF-TDDFT or GEBF-CC2 could be used to compute electronic absorption spectra of solutions or large molecules with local excitations. The GEBF approach under the periodic boundary conditions (PBC) is also been implemented by constructing the subsystems in a super cell and employing the Ewald summation and compensation field methods to take the effect of the long-range electrostatic interaction of the crystal environment into account. The PBC-GEBF approach has been used to compute the lattice energies, crystal structures, IR and Raman spectra, and NMR chemical shifts of various condensed-phase systems, including molecular crystals, liquids, ionic liquid crystals, and solutions. Thus, the GEBF and PBC-GEBF approaches are expected to be widely applied to the energies, structures, and molecular properties of a broad range of large systems and condensed-phase systems.

quantum chemistry calculation, large molecule, condensed-phase system, linear scaling, energy-based fragmentation approach

doi: 10.1360/N972018-00907