

1003-7713/2005/05-641-10

综 述

量子力学和分子力学联用方法*

魏 凯, 刘 磊, 李晓松, 郭庆祥**

(中国科学技术大学化学系, 合肥 230026)

摘 要: 溶液体系中的一些化学行为、生物大分子的结构性能, 已经成为很多理论研究的重点. 在用分子力场来处理溶液体系的时候由于没有考虑电子的运动, 所以在很多力场中, 原子间的化学键必须确定, 以确保在模拟过程中化学键不会破坏或生成. 而如果用基于 B-O 近似的量化计算来处理溶液体系, 由于计算过于耗时而不能处理太大的体系使得量化计算的方法受到了限制. 于是产生了 QM/MM 的组合方法, 并且在研究凝聚态中的化学反应和生物大分子尤其是酶催化反应的机理等方面有着广泛的应用, 也取得了很大的进展. 这一方法的优点是利用了量子力学的精确性, 又利用了分子力学的高效性. 对 QM/MM 组合理论及其发展历史进行简单的介绍, 并对其应用作一些介绍.

关键词: QM/MM; 赝势场; 广义杂化轨道理论; 连接原子法; 经验价键理论; 分子轨道价键理论; INIOM

中图分类号: O623.13

文献标识码: A

Combined Quantum Mechanism and Molecular Mechanism Method*

Wei Kai, Liu Lei, Li Xiaosong, Guo Qingxiang**

(Department of Chemistry, University of Science and Technology of China, Hefei 230026)

Abstract Till now, theoretical research has focused on some chemical behaviors, structural properties of bio-macromolecules in solutions. When the molecular force field was used to treat solution systems, not considering the movement of electrons, in many force fields the chemical bonds between atoms must be confirmed to make sure that they can't be destroyed or formed in modeling procedure. Using quantum chemical calculation based on B-O approximation to treat solution systems, because it can consume much more resources, this method can't treat large systems, which restricts the application of quantum chemical calculations. Moreover the hybrid QM/MM method has been widely used in research of chemical reaction in condense phase and biologic macromolecules especially the mechanisms of enzyme catalysis reactions, which has developed rapidly and takes advantage of the accuracy of QM method and high efficiency of MM method. This article will describe a brief introduction of the principle and development history of QM/MM method and give some application introductions.

Key words QM/MM, Pseudopotential, Generalized Hybrid Orbital principle, Linking atom method, Empirical valence bond principle, Molecular orbital valence bond principle, INIOM

1 引 言

量子力学从头算方法是化学以及化学生物学研

究中的一件重要武器. 它能够准确而又细致地为我们提供有关分子结构、能量变化以及反应活性等方面的信息, 从而对实验结果起到补充和检验的作用,

* Project supported by the National Natural Science Foundation of China (20332022, 20472079).

** Corresponding author, Email: qxguo@ustc.edu.cn Received 23 February 2005; in final form 8 June 2005.

有助于我们透过现象看清本质. 经过几十年的努力, 由于计算机硬件的快速发展和计算算法的不断优化, 目前我们已经能够有效地使用从头算方法处理含有 100 个非氢原子的体系. 然而, 由于从头算方法的时间复杂度大多在 $N^3 \sim N^7$ (N 为原子数) 的数量级, 从现在到将来相当长的一段时间内, 我们仍然无法使用从头算方法处理含有 1000 个以上非氢原子的体系.

从头算方法在体系大小方面受到的限制严重地阻碍了人们对于凝聚态体系以及生命化学体系的理论研究, 而目前有关凝聚态体系(如纳米材料)以及生命化学体系(如酶与分子机器)的研究正逐渐成为化学领域中的重点. 为此, 科研人员往往不得不采用较为粗糙的、纯属经验性质的分子力学方法来处理大型的化学体系. 这些分子力学方法简单地将体系用弹簧力加范德华力束缚起来, 的确能够提供一些有关分子构象以及静电势分布的信息. 但是, 除了计算精度上的问题以外, 分子力学方法还有一个根本的缺陷, 就是不能有效地描述化学键的形成和断裂过程^[1]. 因此, 当我们关心的是体系的“化学”问题(也即体系中的化学键形成与断裂问题)时, 传统的分子力学方法是无能为力的.

一个巧妙地解决上述一系列矛盾的方法是将量子力学从头算方法与分子力学方法有机地结合起来, 对于体系中重要的、涉及到化学键变化的部分使用量子力学从头算方法来描述, 而对于体系中次要的、不直接涉及到化学键变化的部分使用分子力学方法来描述. 这种方法就是量子力学和分子力学联用方法, 简称 QM/MM 方法. 该方法在科学计算理论中有着简单而又正确的逻辑, 也就是对于主要矛盾部分进行精算, 而对于次要矛盾部分进行粗算. 该方法在计算的可行性方面也迈出了一大步. 原则上讲, 在目前的条件下使用 QM/MM 方法处理含有上万个非氢原子的体系是完全可行的.

最早有关 QM/MM 方法的研究由 Warshel 等人上个世纪 70 年代开展^[2]的. 经过随后近 30 年的改进, QM/MM 方法已经从理论探讨逐渐走上实践应用. 直到今天, QM/MM 方法仍然尚未发展成为一种成熟可靠的、规范化和商业化的理论工具, 我们应该看到这一方法在未来化学研究中的巨大潜力和长远价值. 因此, 学习、掌握并发展 QM/MM 方法, 在目前有着重要的意义, 机不可失. 本文扼要地阐述了 QM/MM 方法的基本理论、实现方法以及使用范围,

希望能够为这方面的深入研究提供有益的信息.

2 基本原理

使用 QM/MM 方法, 第一步是将整个体系划分为两个部分. 其中一个部分我们称之为活性中心, 它是我们将要考察的关键部位, 由那些直接参与化学键形成或断裂的原子构成. 对于活性中心我们采用了量子力学从头算的方法来描述, 因此该部分又称之为 QM 区域. 活性中心之外的所有原子构成了外围环境. 对于外围环境我们采用了分子力学方法来描述, 因此该部分又称之为 MM 区域.

描述用复合势场处理的体系时, 使用有效哈密顿 H_{eff} 算符. 在这个公式中, 可以通过求解薛定谔方程, 得到给定分子结构的体系能量 E_{total} ,

$$\hat{H}_{\text{eff}} \Psi(r_i, R_\alpha, R_M) = E_{\text{total}}(R_\alpha, R_M) \Psi(r_i, R_\alpha, R_M) \quad (1)$$

其中, r_i 和 R_α 为电子和核子在 QM 区的系数; R_M 为原子在 MM 区的系数.

整个体系的有效混合哈密顿算符可表示为:

$$\hat{H}_{\text{eff}} = \hat{H}_{\text{QM}}(\text{QM}) + \hat{H}_{\text{MM}}(\text{MM}) + \hat{H}_{\text{QM/MM}}(\text{QM/MM}) \quad (2)$$

第一项为 QM 区的哈密顿算符, 可以用标准的量子力学形式表示. 例如用 *ab initio* 处理时,

$$\hat{H}_{\text{QM}} = -\frac{1}{2} \sum_i \nabla_i^2 + \sum_{ij} \frac{1}{r_{ij}} - \sum_{\alpha} \frac{Z_\alpha}{r_{i\alpha}} + \sum_{\alpha\beta} \frac{Z_\alpha Z_\beta}{r_{\alpha\beta}} \quad (3)$$

其中, Z_α 为 α 原子核的电荷; r_{ij} 为粒子 i 和 j 之间的距离. 下标 (i, j) 和 (α, β) 分别为电子和原子核. 第二项为 MM 区哈密顿算符 H_{MM} , 而且只在 MM 原子间有影响, 体系的能量只跟原子坐标有关, 可以表示为键的伸缩 E_{bond} 、键弯曲 E_{angle} 、键扭曲 E_{dihedral} 以及非键相互作用, 如静电作用和 Lennard-Jones 相互作用^[3].

$$\begin{aligned} \hat{H}_{\text{MM}} &= E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} + E_{\text{Lennard-Jones}} + E_{\text{Coulomb}} \\ &= \sum_i K_b (b_i - b_0)^2 + \sum_i K_\theta (\theta_i - \theta_0)^2 + \\ &\quad \sum_i K_\phi \cos[n(\phi_i - \phi_0)] + \sum_{i>j} \frac{Q_i Q_j}{r_{ij}} + \\ &\quad \sum_{i>j} \epsilon_{ij} \left[\left(\frac{r_{ij}^0}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{ij}^0}{r_{ij}} \right)^6 \right] \end{aligned} \quad (4)$$

第三项为 QM 区与 MM 区之间相互作用的哈密顿算符, 它构成了 QM/MM 方法的核心, 但是它依赖于 QM 和 MM 的形式. 对此我们将作详细的介绍.

当我们在划分 QM/MM 区域时, 必然涉及边缘

边界问题. 在模拟液相体系时, 一般选取数百个溶剂分子组成的正方体或长方体箱子作为 MM 区, 溶质分子则自然成为 QM 体系, 而边界区域则通过溶剂箱的周期平移得到, 把边界区域看成连续介质模型.

这样通过式(1)~(4), 整个体系的能量(E_{total})由下式给出^[4]:

$$E_{\text{total}} = E_{\text{QM}}(\text{QM}) + E_{\text{MM}}(\text{MM}) + E_{\text{QM/MM}}(\text{QM/MM}) \quad (5)$$

式中, $E_{\text{QM}}(\text{QM})$ 为 QM 区域的能量, 取决于所需的精度, 这一部分能量可以由某一种量子力学方法(如半经验方法、Hartree-Fock 方法、Møller-Plesset 微扰法以及 DFT 方法等)给出; $E_{\text{MM}}(\text{MM})$ 为 MM 区域的能量, 通常可以使用某一力场(如 AMBER 和 CHARMM 力场)计算得到; $E_{\text{QM/MM}}(\text{QM/MM})$ 为 QM 区域与 MM 区域之间的相互作用能.

一旦总能量知道了, 原子之间的力参数很容易通过能量对原子坐标的微分得到:

$$F_{\alpha} = -\frac{\partial E_{\text{total}}}{\partial R_{\alpha}} \quad (6)$$

$$F_M = -\frac{\partial E_{\text{total}}}{\partial R_M} \quad (7)$$

得到总的体系能量以及相关的力参数后, 就可以处理分子几何优化、分子动力学研究以及 Monte Carlo 模拟等问题了.

对 $E_{\text{QM/MM}}(\text{QM/MM})$ 的计算则取决于研究的体系, 对不同的体系要使用不同的计算方法. ① QM 区域与 MM 区域之间没有共价相连. 这种情况多见于溶液体系, 其中一个溶质分子(QM 区域)被大量溶剂分子(MM 区域)包围, 而溶质分子与溶剂分子之间并没有共价相连. 在这种体系中, QM 区域与 MM 区域之间的相互作用能主要来自于它们之间的静电吸引(或排斥)和范德华作用, 也即:

$$E_{\text{QM/MM}}(\text{QM/MM}) = E_{\text{electrostatic}}(\text{QM/MM}) + E_{\text{vdw}}(\text{QM/MM}) \quad (8)$$

其中范德华作用能($E_{\text{vdw}}(\text{QM/MM})$)可以用经验的公式(如 12~6 次 Lennard-Jones 势场)求算, 其表达式为:

$$E_{\text{vdw}} = \sum_{s=1}^S \sum_{m=1}^M 4\varepsilon_{ms} \left[\left(\frac{\sigma_{ms}}{R_{ms}} \right)^{12} - \left(\frac{\sigma_{ms}}{R_{ms}} \right)^6 \right] \quad (9)$$

σ_{ms} 、 ε_{ms} 为 Lennard-Jones 参数, 且 $\sigma_{ms} = (\sigma_m \sigma_s)^{1/2}$, $\varepsilon_{ms} = (\varepsilon_m \varepsilon_s)^{1/2}$, σ_m 、 σ_s 可以从 MM 力场中获得, ε_m 、 ε_s 为 QM 原子的经验参数^[5].

而静电相互作用能($E_{\text{electrostatic}}(\text{QM/MM})$)可以

首先使用分子力场方法求出 QM 区域和 MM 区域中每个原子所带的点电荷, 然后使用库仑定律得到总的静电相互作用, 其表达式为:

$$E_{\text{electrostatic}}(\text{QM/MM}) = -\sum_{s=1}^S \sum_{i=1}^N \frac{q_s}{r_{is}} + \sum_{s=1}^S \sum_{m=1}^M \frac{q_s Z_m}{R_{ms}} \quad (10)$$

式中, q_s 为 MM 原子上的点电荷; Z_m 为 QM 原子的核电荷; r_{is} 、 R_{ms} 分别为 MM 点电荷与 QM 电子、原子核间的距离.

另一种更为精确的求算 $E_{\text{electrostatic}}(\text{QM/MM})$ 的方法是将 MM 区域中每个原子所带的经验点电荷直接写进 QM 区域的 Schrödinger 方程里. 由此我们得到有效 Hamilton 方程:

$$H_{\text{eff}} = -\frac{1}{2} \sum_i^{N_{\text{eff}}} \nabla_i^2 + \sum_{i \neq j}^{N_{\text{eff}}} \frac{1}{r_{ij}} - \sum_i^{N_{\text{eff}}} \sum_{\alpha \in \text{QM}} \frac{Z_{\alpha}}{r_{\alpha i}} - \sum_i^{N_{\text{eff}}} \sum_{\beta \in \text{MM}} \frac{q_{\beta}}{r_{i\beta}} + \sum_{\alpha_1 \neq \alpha_2 \in \text{QM}} \frac{Z_{\alpha_1} Z_{\alpha_2}}{r_{\alpha_1 \alpha_2}} + \sum_{\alpha \in \text{QM}, \beta \in \text{MM}} \frac{Z_{\alpha} q_{\beta}}{r_{\alpha\beta}} \quad (11)$$

式中, N_{eff} 为 QM 区域总的电子数目; r_{ij} 为电子 i 到电子 j 的距离; Z_{α} 为 QM 区域中原子 α 的核电荷; $r_{\alpha i}$ 为电子 i 到原子核 α 的距离; q_{β} 为 MM 区域中原子 β 所带的经验点电荷; $r_{i\beta}$ 为电子 i 到原子 β 的距离; $r_{\alpha_1 \alpha_2}$ 为原子核 α_1 到原子核 α_2 的距离; $r_{\alpha\beta}$ 为原子核 α 到原子 β 的距离. 显而易见, 使用上式求解 QM 区域的 Schrödinger 方程时, $E_{\text{electrostatic}}(\text{QM/MM})$ 已经被包含在体系总的能量中, 即:

$$E_{\text{QM}}(\text{QM}) + E_{\text{electrostatic}}(\text{QM/MM}) = \langle \psi | H_{\text{eff}} | \psi \rangle \quad (12)$$

② QM 区域与 MM 区域之间有共价相连. 这种情况较前一种情况更为常见, 如在酶化学中反应中心与整个蛋白质分子并非是两个彼此独立的单元, 而是通过共价键相互连接在一起. 在这种情况下我们不能将 QM 区域与 MM 区域之间的化学键简单切断, 否则造成末端为自由基的分子片断, 从而导致荒谬的计算结果^[6].

为了饱和 QM 区域中与 MM 区域相键连的原子(以下称之为 QM 末端原子)的化学价, 人们提出了以下几种方法. 第一, 使用氢原子来饱和 QM 末端原子, 也就是由 Singh、Field 等人发展的连接原子法^[7,8], 以及在此基础上由 Morokuma 及其合作者发展的 IMOMM 方法和 INIOM 方法^[9-11]. 这些方法简明易行, 但也有先天的缺陷, 而且它对连接原子的选

择有严重的依赖性^[12]. 产生缺陷的主要原因来自于氢原子独特的成键能力(即使用 1s 轨道成键)及其较小的电负性. 这使得 QM 末端原子不仅在成键后电子云分布与真实情况有着较大的差异,而且其电荷转移情况(即由 QM 区域向 MM 区域转移电子,或者 QM 区域从 MM 区域得到电子)往往失真. 第二,使用有效原子对 QM 末端原子进行“盖帽”. 该原子带有一定数目的电子、一定量的核电荷以及一定的有效电势场. 使用一定数目的电子以满足两条基本要求,一是该有效原子只能与 QM 末端原子(通常是 sp^3 杂化的碳原子)形成一根 σ 单键,二是该有效原子的外层电子云基本上保持球对称. 使用一定量的核电荷则是为了使得该有效原子的总电荷平衡. 最后,使用一定的有效电势场可以调节 QM 区域与 MM 区域之间的电荷转移.

值得指出的是,在目前主要有两种方法来实现有效电势场. 一种方法是固定轨道法,即在有效原子上人为地添加一些系数一定的已经填充好电子的原子轨道,比较有代表性的是 Rivail 及其合作者的 LSCF 方法^[13-16],以及在此基础上由 Gao 发展的广义杂化轨道理论(GHO)^[17]. 由于在组成分子轨道时这些原子轨道的系数并不被优化,这相当于人为地固定了一些具有一定空间分布的负电荷势场在有效原子周围. 另一种方法是有效核势场法. 在这种方法中,人为地强加一些描述电势场的函数在有效原子的核上. 通过调节函数的性质和参数可以很好地模拟再现 QM 区域与 MM 区域之间的电荷转移,这种方法比较有代表性就是 Yang 及其合作者的 pseudobond^[18]法.

显而易见,使用有效原子的方法的根本思路就是最大程度地再现 QM 区域与 MM 区域之间的共价键特征. 尽管我们可以使用多种方法来添加有效原子,但实质上各种方法都是通过某种参数调节来实现最优的拟合. 因此,好的有效原子方法不仅要拟合程度高,而且要参数化过程简单.

3 实现方法

根据以上的基本原理和简便参数化的思想,我们在这里具体地给出了一种 QM/MM 的实现方法. 该方法的核心是解决 QM 区域与 MM 区域之间的共价相连问题. 而针对 QM 区域与 MM 区域之间没有共价相连这种简单情况,我们假设下述的方法也可以直接应用.

首先,考虑到计算精度与计算速度的关系,我们决定采用 DFT^[19]方法(B3LYP)来描述 QM 区域. 这种方法与 Hartree-Fock^[20]方法有着近似的时间复杂度,却能够较好描述体系的电子相关能. 其次,由于我们的 QM/MM 方法主要针对于酶化学,我们决定采用 CHARMM^[21]力场. 这种力场在处理多肽与蛋白质体系时有着较好的能力. 最后,为了简化参数化的过程,我们将“盖帽”原子定义为拟卤素原子. 该原子有 7 个核电荷, 7 个价电子,一套核势场函数. 显然,该原子能且仅能与 sp^3 杂化的碳原子形成一根 σ 单键. 形成单键后,其 $8(7+1)$ 个价电子保证了球对称性,并最大程度地模拟了饱和成键后的碳原子.

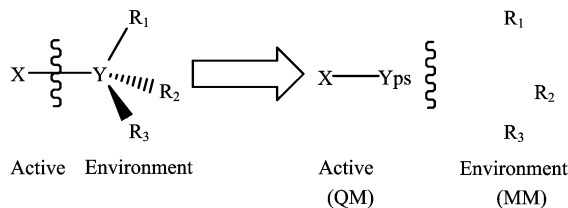


图 1 QM 与 MM 体系的划分

Fig. 1 The division between QM and MM parts

我们先将分子体系做如图 1 的划分^[18]. 图 1 中, Active 区表示反应的活性中心,用 QM 来处理, Environment 区表示反应的活性中心以外的部分,用 MM 来处理. 在 pseudobond 的方法中, Y 原子(MM)将被单自由价的边界原子 Y_{ps} 所取代,而 Y_{ps} 原子就是前面提到的拟卤素原子,形成的 $X-Y_{ps}$ 将会拟合原来的 $X-Y$ 键. 这个时候将包括 Y_{ps} 原子在内的 Active 区域表示为 QM 体系,而将剩下的部分表示为 MM 体系. 则系统能量式中 $E_{QM/MM}$ (QM/MM) 可以由下式给出:

$$E_{QM/MM}(QM/MM) = E_{\text{electrostatics}}(QM/MM) + E_{\text{vdW}}(QM/MM) + E_{\text{mm-bonded}}(QM/MM) \quad (13)$$

其中第三项表示 QM 部分与 MM 部分成键的作用能,用的是分子力场的方法. 所以,在计算的过程中 $E_{\text{vdW}}(QM/MM)$ 、 $E_{\text{mm-bonded}}(QM/MM)$ 和 $E_{\text{mm}}(MM)$ 是使用分子力场的方法来计算的,而用量子化学方法计算的只有 $E_{\text{QM}}(QM)$ 和 $E_{\text{electrostatics}}(QM/MM)$ 可以用式(12)来得到,其中有效 Hamiltonian 算符描述的是一组有效电子 N_{eff} , 它包括 Active 部分的所有电子以及边界原子 Y_{ps} 的价电子, H_{eff} 可以由下式给出:

$$H_{\text{eff}} = -\frac{1}{2} \sum_i^{N_{\text{eff}}} \nabla_i^2 + \sum_{i \neq j}^{N_{\text{eff}}} \frac{1}{r_{ij}} - \sum_i^{N_{\text{eff}}} \sum_{\alpha \in \text{QM}} \frac{Z_\alpha}{r_{\alpha i}} - \sum_i^{N_{\text{eff}}} \sum_{\beta \in \text{MM}} \frac{q_\beta}{r_{\beta i}} + \sum_i^{N_{\text{eff}}} \sum_{\gamma \in Y_{\text{ps}}} V_{Y_{\text{ps}}}^{\text{eff}}(r_{i\gamma}) + \sum_{\alpha_1 \neq \alpha_2 \in \text{QM}} \frac{Z_{\alpha_1} Z_{\alpha_2}}{r_{\alpha_1 \alpha_2}} + \sum_{\alpha \in \text{QM}, \beta \in \text{MM}} \frac{Z_\alpha q_\beta}{r_{\alpha \beta}} \quad (14)$$

其中, Z_α 为 QM 部分的原子核电荷; q_β 为 MM 部分原子的点电荷; $V_{Y_{\text{ps}}}^{\text{eff}}$ 为边界原子的有效核电势 (effective pseudopotential). 所以, 要想得到 H_{eff} , 其中有两项是未知的, 就是 q_β 和 $V_{Y_{\text{ps}}}^{\text{eff}}$, 求解这两项的方法如下: ① 要得到 q_β , 需要注意的两个问题是: MM 体系的部分电荷的总和加上 QM 体系的电荷应该等于整个体系的电荷; QM 计算中怎么样降低由于存在非常靠近 QM 体系的 MM 原子而产生大的静电作用. 解决的方法就是将直接和 QM 体系成键的 MM 原子以及和 QM 体系通过两个键相连的原子的点电荷视为 0, 这种方法的原理是利用了分子力场中的电荷形式. 通过这种定义就解决了如上提到的两个问题. ② 要得到有效核电势 $V_{Y_{\text{ps}}}^{\text{eff}}$. Y_{ps} 原子有效核电势的作用就是 $Y_{\text{ps}}-X$ 键可以模拟原来的 sp^3 杂化的 $X-Y \sigma$ 单键, 也就是说 $Y_{\text{ps}}-X$ 键应该和原来的 $X-Y$ 单键有相似的键长、键的强度以及对 QM 部分的影响, 其表达式如下:

$$\begin{cases} V^{\text{eff}}(r) = V_L^{\text{eff}}(r) + \sum_{l=0}^{L-1} [V_\lambda^{\text{eff}}(r) - V_L^{\text{eff}}(r)] \times \sum_m |\lambda m \rangle \langle \lambda m| \\ V_L^{\text{eff}}(r) = \frac{1}{r} a_L e^{-b_L r^2} \\ V_\lambda^{\text{eff}}(r) - V_L^{\text{eff}}(r) = a_\lambda e^{-b_\lambda r^2} \end{cases} \quad (\lambda = 0, 1, \dots, L-1) \quad (15)$$

其中 L 为基组中的最大 λ 值; a_i 和 b_i 分别为适当的参数. 有效核电势的参数化过程就是让 $Y_{\text{ps}}-X$ 键的参数和原来的 $X-Y$ 键的参数拟合得最好, 这样在 QM 的计算过程中使得 Y_{ps} 和 MM 部分的作用是相当的. 当得到这些参数后再对其他的化学体系进行检验. 而且值得注意的是有效核电势的参数化过程不是在分子力场的计算中得到的, 而是在纯 QM 计算过程中得到的. 这种算法的和其他方法相比, 明显的优势就是参数的通用性很强, 而且对不同体系计算得到的结果与用纯 QM 计算得到的结果是相近的.

4 操作过程

Pseudobond QM/MM 计算的流程如图 2. 具体的步骤如下: ① 从蛋白质、核酸等数据库下载相应的分子晶体结构文件, 然后通过一系列的能量最小化、平衡步骤, 从中选择几个比较合理的分子结构, 对每一个分子进行如下的步骤. ② 先用 Tinker 程序计算 MM 部分的能量、电荷以及能量梯度等数据, 它们将在下面的计算过程中起到非常重要的作用. ③ 用 Gaussian98 程序对 QM 部分进行优化, 在这个优化过程中 MM 部分是固定不动的. 并且这个过程中 Y_{ps} 原子用 F 原子来表示, 它的赝势场目前只有 3-21G 和 6-31G* 两个基组下的参数, 这些参数在文献 [18] 中有详细的说明. ④ 用 tinker 中的 Newton 程序对 MM 部分进行构型的优化, 同样在这个过程中 QM 部分是固定不动的. 循环进行上述三个过程直到系统总的能量以及 $RMSD$ (均方根偏差) 均收敛为止, 其中 $RMSD$ 的计算表达式如下:

$$RMSD = \sqrt{\frac{1}{N_{\text{atoms}}} \sum_{i=1} (r_i - r_{i0})^2} \quad (16)$$

它表明不同构象的两个分子间的构象差异. 至此, 对

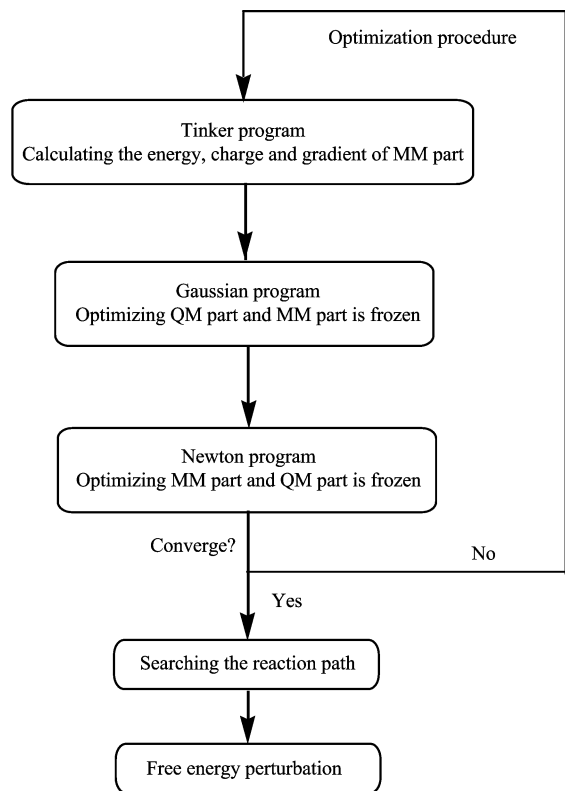


图 2 QM/MM 流程图
Fig. 2 QM/MM procedure

此构象分子的优化过程结束. ⑤ 要研究一个反应的机理, 也就是反应路径的搜索步骤, 必须选取合适的反应坐标, 反应坐标可以是两个原子间的距离、两组原子的距离差、三个原子间的夹角以及四个原子间的二面角等参数. 一旦反应坐标确定后, 就可以对反应路径进行一系列的研究. 在这个过程中对反应坐标上的每一个构型都要进行上述②~④步骤的优化. 在优化的过程中为了让反应坐标稳定在我们预先设定的某个值上, 必须在体系的能量函数中引入约束能这一能量项, 它的表达式如下:

$$E_{\text{Restrain}} = k(R - s)^2 \quad (17)$$

其中, k 为力常数; R 为反应坐标; s 为反应坐标变化的步长. 这样做的目的就是迫使分子体系向预定反应路径的方向进行优化. 通过如上的步骤就可以得到整个反应路径的能量-反应坐标曲线. ⑥ 得到了反应路径, 就相当于掌握了反应的整个过程, 还可以对这个反应路径进行一系列的计算, 比如可以计算整个反应过程中的自由能垒, 反应中的自由能变化, 这个可以用自由能微扰来得到, 自由能微扰就是将整个反应路径分成很多的点, 然后计算相邻两点间的自由能, 这样反应路径中任何两点间的自由能变化等于它们之间所有相邻两点间的自由能变化的叠加, 对反应路径划分的越细, 得到的自由能也越准确. 在步骤⑤中, 我们刚好得到了反应路径-反应坐标的曲线, 通过对这个曲线进行自由能微扰的计算, 就可以得到整个反应过程的自由能变化和自由能垒等重要参数, 而这些参数是有实验数据可以考证的, 所以这样反过来又可以检验得到的这个反应路径是否合理. 在进行自由能微扰的计算过程中值得注意的一点就是对反应路径上的每一个分子假定 QM 部分对整个分子体系自由能的贡献是固定的, 它们有前面得到的固定的分子几何构型和电荷分布.

5 应用范围

在此, 我们重点讨论一下利用复合 QM/MM 方法, 进行的分子模拟研究. 目前, 最主要的应用是溶液体系中的化学现象模拟和生物分子的研究; QM/MM 方法的最大优势就是能够处理庞大而又复杂的体系. 这主要包括两个方面, 即溶液体系和生物大分子的处理^[5, 22-31].

5.1 溶液体系中的化学现象

利用 QM/MM 方法, 可以很好地分析整个反应过程中参与反应原子的电荷分布, 进一步可以说明

溶剂分子对反应的影响. 而且在模拟溶液体系中的化学行为时, QM/MM 边界划分显的比较明显, 参与反应的分子用 QM 的方法处理, 而反应体系周围的溶剂分子将用 MM 的方法处理; 有些情况下, 溶剂分子也参与反应过程, 特别是水作为溶剂的反应过程, QM/MM 的边界划分将显得比较复杂.

1976 年, Warshel 和 Levitt 介绍了 QM/MM, 并且用它研究了溶菌酶的作用机理^[2]. Singh 和 Kollman 发展了用 *ab initio* 分子轨道理论和 AMBER MM 力场相组合的 QM/MM 方法, 即 *ab initio*/AMBER 的 QM/MM 方法, 并且研究了氯甲烷/氯负离子的交换反应, 溶液中以及气相中的聚醚的质子化作用 (protonation of polyethers)^[7]. 气态条件下, $\text{CH}_3\text{Cl} + \text{Cl}^-$ 形成复合物时, 能量有 36 kJ/mol 的降低; 但是动力学显示, 稳定的对称复合物之间有 50 kJ/mol 的能垒. 当考虑溶剂化效应时, 情况变得很简单, 但是没有找到稳定极小, 只有一个简单能垒为 113 kJ/mol. Singh 和 Kollman 在 QM 部分采用了 *ab initio* 的方法来优化反应途径中各种结构, 发现了一个更高的反应能垒, 以及两个没有气态条件下稳定的化合物. 但是这种方法的缺点是 *ab initio* 的 QM 方法太耗时, 并且准确的同时也排除了分子动力学和 Monte Carlo 模拟中势能的使用, 并且仅限于静态势能曲线的研究.

为了解决这个问题, Karplus 及合作者发展了 AM1/MNDO 的半经验 QM 方法和 CHARMM 的 MM 力场的组合方法^[8, 21, 32, 33]. 并且也进行了氯甲烷/氯负离子交换反应的研究^[34].

后来由 Warshel 等人提出的 EVB (Empirical Valence Bond) 方法^[24, 35-37], 在处理溶液体系中得到了广泛的应用^[36-39], 并且研究了许多关于溶液和酶的反应^[35]. Voth 及其合作者对 EVB 方法做了进一步的扩展, 并且应用于水溶液体系中的质子迁移反应的模拟^[40, 41]. 由 Gao 等人在 EVB 的基础上发展的 MOVB (Molecular Orbital and Valence Bond) 方法^[42]. 对反应体系 PES (Potential Energy Surfaces) 的处理有很好效果.

除了对溶剂化能进行了详细的研究外, Gao 通过 QM/MM 方法对溶液中溶剂与溶质间的相互作用的研究为溶剂极性不同对溶质光谱性质的影响提供了有力的理论依据^[43]. Thompson 及其合作者也对溶剂对分子光谱特性的影响进行了研究^[23], 主要包括羰基化合物中 $v \rightarrow n^*$ 光谱位移. 这项工作尽管采

用了 AM1 半经验的方法,但是它包含了 MM 原子的极化作用,并且发展了一套完整的方法来计算 MM 原子的诱导偶极与 QM 原子上电荷的相互作用.利用这种方法,他们研究了 *Phodopseudomonas viridis* 光合成反应中心的细菌叶绿素(bacteriochlorophyll) 激发态的光谱特性.他们还研究了金属 K^+ 和二甲醚(dimethyl ether)复合物的光谱特性,但是没有包括 MM 原子的极化作用.

Bakowies 和 Thiel 也发展了考虑 MM 原子极化作用的 QM/MM 集合的方法,研究了一些气态条件下的反应.

由 Morokuma 及其合作者提出的 ONION 模型在处理大分子 Diels-Alder 反应时,给出了和试验数据较接近的结果^[11].

早期的 QM/MM 方法一般使用半经验的方法,但是,随着计算机的发展以及算法的优化,使得对 QM 部分一般使用 DFT 方法或者 *ab initio* 方法.

5.2 酶催化反应

QM/MM 复合的方法为我们模拟酶催化反应提供了可能.而且,可以这么说, QM/MM 方法将来最大的应用是来解决酶催化反应的模拟.所以近年来 QM/MM 复合方法在酶催化反应中得到了广泛的应用.在这个体系中,酶催化的活性中心一般用 QM 处理,而酶分子的其他部分以及溶剂分子一般用 MM 的方法处理.

与溶液体系中的反应相比,酶催化反应的模拟有两点很不一样.其一, QM/MM 边缘的区分将会变得很不明显,我们要尽量多的用 QM 的方法来活性中心;其二,对于蛋白质分子,由于 QM/MM 间有化学键的断裂,而在溶液体系中这种情况是比较少的.所以在酶催化反应体系中, QM/MM 的处理方法和溶液体系有很大的不同.

Warshel 和 Levitt 在这方面作出了开创性的贡献,早期他们研究了溶菌酶 lysozyme 催化的糖苷键的 glycosidic 裂解过程中,碳正阳离子中间体的稳定性,结果表明静电作用的稳定,对酶活性的保持起着重要的作用.在他们对酶作用机理的研究中,得出了影响酶反应的各种因素,包括反应中化学键的重排、电荷的重新分布、酶与酶作用物的空间及静电相互作用、酶以及周围水分子由于极化作用而导致的介电效应.并且对酶催化反应中 QM/MM 区域的划分提出了不少建议^[2].随后 Warshel 提出了经验价键理论(EVB)方法,利用此方法,他们研究了一系列

的酶催化反应,包括丝氨酸蛋白酶和胰蛋白酶在溶液中的催化反应^[44,45]. Antonczak 及其合作者通过 Rivail 提出的 LSCF 方法对嗜热菌蛋白酶的作用机理进行了研究,并且总结了酶所在的环境对酶反应过程的影响因素^[46].

6 应用实例

6.1 互变异构酶催化机理

4-草酰丁烯酸酯互变异构酶(4-OT)可以催化非轭的 α -酮酸转变为共轭的同分异构体,例如可以将 2-羧基-4-烯己二酸酯通过 2-羟基-2,4-二烯己二酸酯中间产物转变为 2-羧基-3-烯己二酸酯.反应过程如图 3.

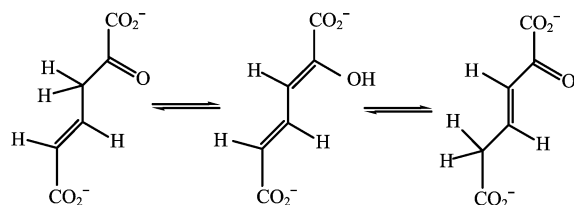


图3 4-OT催化的异构化反应

Fig. 3 Isomerization catalyzed by 4-OT

4-OT 由 6 个亚单元组成,其中每一个亚单元由 62 个氨基酸残留基团构成.4-OT 含有 6 个活性中心,他们分布在每一个亚单元中 Pro-1 周围的疏水区域,同时 Arg-11'、Arg-39'、Arg-61' 也参与同分异构反应的催化过程.由于 4-OT 尺寸很小,而且它可以在没有任何金属离子的作用下进行质子迁移反应的催化,所以对它的催化机理的研究也就显得非常的重要.实验结果显示反应机理是一个广义酸碱反应,而且其中的 Pro-1 氨基酸残留基团起到广义碱的作用,但是广义酸的角色一直受到争议.反应中心如图 4.

理论计算得到 Arg39' 的 pKa 在 9.8 ~ 10.3,可以起到广义酸的作用,但是实验证明上面 3 个 Arg 氨基酸残留基团没有一个 pKa 在 9 附近,所以寻找广义酸基团便成了这个工作的重点.之前有人提出在这个反应过程中,如图 4 所示, Arg-39' 和一个水分子通过和过渡态中的氧负离子形成两个氢键来稳定过渡态.

在这个工作中,体系一共加入 2407 个水分子,总共由 13197 个原子,在分子动力学平衡的过程中,发现有两个定序的水分子与底物和 Arg-61' 形成氢键.为了寻找反应过程中的广义酸,在这个工作中对

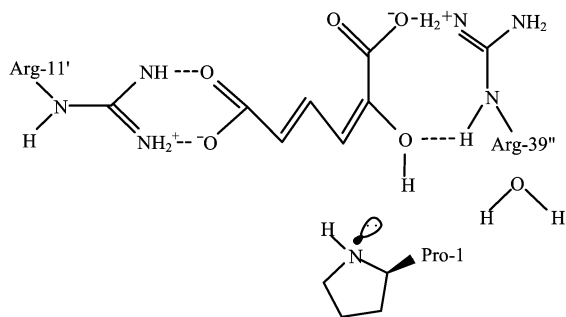


图 4 4-OT 的催化反应中心

Fig. 4 Active site of the 4-OT catalyzed reaction

反应体系进行四组不同 QM/MM 界线的划分, 然后分别对这四组体系进行 QM/MM 的计算, 在这个过程中先在 HF/3-21G 的基础上对体系进行构象优化, 然后在 B3LYP/6-31G* 的基础上对体系进行单点能的计算, 然后进行自由能微扰的计算, 得出整个反应的自由能垒为 67.76 kJ/mol, 比实验值约 54.35 kJ/mol 低 14.21 kJ/mol 左右. 而这个实验值也是一个近似的结果. 并且通过对反应能垒的分解, 可以得到单个氨基酸残留基团对过渡态稳定的贡献. 通过如上的对比研究可以得知在这个反应中没有氨基酸残留基团或者水分子起到广义酸的角色, 而是通过两个定序的水分子和 Arg-39'' 通过静电相互作用来稳定底物分子的负电荷, 而且在这个反应中 Arg-11' 在稳定过渡态的过程中起到了非常重要的作用, 同时这些结果也使以往在寻找广义酸过程中解释不了的现象得到了比较满意的解释^[47].

6.2 乙酰胆碱酯酶催化机理^[48]

乙酰胆碱酯酶(简称 AChE)对于催化神经递质乙酰胆碱(ACh)的水解有着很高的效率, 反应生成胆碱 ROH 和乙酸. 反应路线分为两步, 酰化反应和脱酰基反应. AChE 的反应核心是一个催化三联体(Ser203, His447, Glu334)和氧负离子孔(oxyanion hole). 其中 Ser203 氨基酸残基起着亲核攻击的作用, 而 His447 起着广义酸碱催化的作用, 但是 Glu334 氨基酸残基的作用机理还不是太清楚, 虽然氧负离子孔能通过氢键来稳定高能中间体和过渡态分子, 但是在整个反应过程中这些氢键是否始终存在也是这个工作所要解决的问题.

首先为了研究 Glu334 的作用机理, 在划分体系的过程中有两种方案, 如图 5 所示. 一种是将 Glu334 氨基酸残基规划到 QM 部分, 一种则是将 Glu334 氨基酸残基规划到 MM 部分, 其他部分都是

一样的. 在 QM/MM 的计算过程中, 以 Ser203 氧原子为中心 20 Å 为半径的球外的原子在分子优化的过程中是固定不动的, 这样大大减少了计算量. 以 $d_{O-H} - d_{N-H} - d_{C-O}$ 为反应路径的坐标, 在 HF/3-21G 的基础上面对反应路径上的分子进行优化, 然后分别在 B3LYP/6-31G* 和 MP2/6-31G* 的基础上进行单点能的计算, 通过对能量 Hessian 矩阵中求得的分子振动频率的分析, 确认了反应路径上过渡态的正确性. 通过对反应路径的研究, 可以得到两种划分下体系的势能分别为 52.67 kJ/mol(QM1 划分)和 43.89 kJ/mol(QM2 划分), 差别为 8.78 kJ/mol, 而实验数据为约 50.16 kJ/mol. 并且可以得到单个氨基酸残基对反应过渡态稳定性的贡献, 从中可以发现 Glu334 稳定过渡态的作用能(范德华作用和静电相互作用的和)为 -47.23 kJ/mol, 是所有氨基酸残基中贡献最大的基团, 而其中又以静电相互作用为主要稳定过渡态的作用, 作用能为 -48.07 kJ/mol, 所以 Glu334 氨基酸残基的主要作用是通过静电相互作用来稳定反应过渡态. 而通过对氧负离子孔的反应路径的研究, 发现在整个反应过程中, 氧负离子孔和底物形成了两个分子间氢键, 分别是 Gly121、Gly122 和底物形成氢键, 而 Ala204 也只是在过渡态的时候和底物形成了微弱的氢键, 这样氧负离子孔的作用原理也就清楚了.

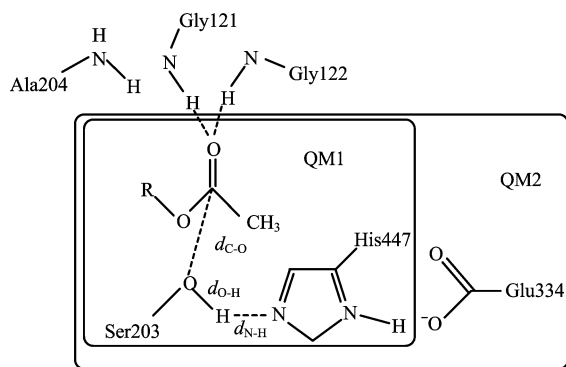


图 5 AChE QM 部分划分图

Fig. 5 Illustration of the division of AChE QM part

6.3 磷酸丙糖异构酶催化机理

磷酸丙糖异构酶(简称 TIM)是糖酵解循环中的关键酶, 可以将二羟基丙酮磷酸盐(DHAP)同分异构化成甘油醛-3-磷酸盐(GAP), 反应活性中心如图 6. 整个分子体系有 6051 个原子, 其中 752 个水分子. 以 $R_{C1-H} - R_{O-H}$ 为反应坐标, 对反应路径进行研究, 并且对得到的反应路径进行自由能微扰的计

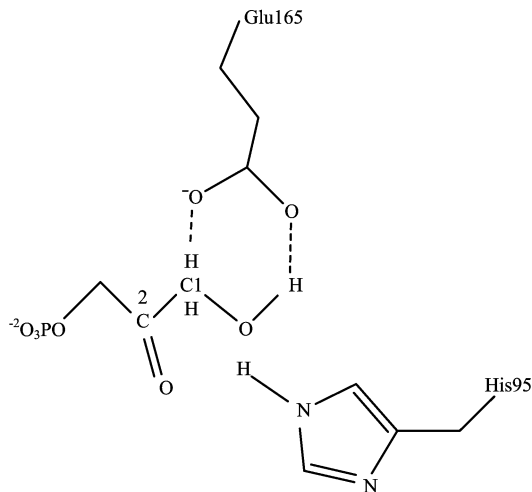


图 6 磷酸丙糖异构酶催化 DHAP 反应活性中心

Fig. 6 Activate site of (triosephosphate isomerase) TIM catalyzed reaction

算. 所得到的自由能曲线收敛的非常的好, 酶的存在大大降低了反应的能垒, 计算得到的反应自由能能垒 91.54 kJ/mol, 而实验得到的值为 58.52 kJ/mol, 这个可以归结为 HF/3-21G 方法参数的设置, 如果用 B3LYP/6-31G*, 结果会更好. 通过对反应路径的研究, 解释了磷酸丙糖异构酶催化 DHAP 异构化第一步反应是一个质子迁移的反应, 在这个反应中 Glu165 起着夺质子的作用, 而其它的残留基团起着稳定过渡态的作用, 这样就大大的降低了反应的能垒. 而且这也在实验中得到证实^[49].

7 结 语

QM/MM 组合的方法由于对大体系各部分根据其反应活性中心的影响进行分区处理, 既保证了结果的可靠性, 又大大的简化了计算, 因此在凝聚态反应和生物大分子等方面得到了非常广泛的应用, 并且取得了许多瞩目的结果. 在处理 QM/MM 体系的过程中, 产生了许多方法, 其中有连接原子法、经验价键理论法、分子轨道与价键理论法、广义杂化轨道法、赝势场法、INIOM 等方法. 它们各有优缺点, 但是在处理 QM/MM 有共价键连接的情况时, 有效电势场是一种比较合理的方法, 这在实际的检验中也得到了相应的证明.

本文扼要地阐述了 QM/MM 方法的基本理论、实现方法、以及使用范围. 希望本介绍能够为这方面的深入研究提供有益的信息.

参 考 文 献

- [1] Burket U, Allinger N L. *Molecular Mechanics*, A. C. S. Monographs 177, Washington D. C., 1982.
- [2] Warshel A, Levitt M. *J. Mol. Biol.*, 1976, **103**: 227
- [3] Levitt M. *J. Mol. Biol.*, 1974, **82**: 393
- [4] Gao J. *Rev. Comput. Chem.*, 1995, **7**: 119
- [5] Gao J. *Acc. Chem. Res.*, 1996, **29**: 298
- [6] Das D, Eurenus K P, Billings E M. *J. Chem. Phys.*, 2002, **117**: 23
- [7] Singh U C, Kollman P A. *J. Comput. Chem.*, 1986, **7**: 718
- [8] Field M J, Bash P A, Karplus M. *J. Comput. Chem.*, 1990, **11**: 700
- [9] Maseras F, Morokuma K. *J. Comput. Chem.*, 1995, **16**: 1170
- [10] Humbel S, Sieber S, Morokuma K. *J. Chem. Phys.*, 1996, **16**: 1959
- [11] Svensson M, Humbel S, Froese R D J, Matsubara T, Sieber S, Morokuma K. *J. Phys. Chem.*, 1996, **100**: 19357
- [12] Eurenus K P, Chatfield D C, Brooks B R, Hodoscek M. *Int. J. Quantum Chem.*, 1996, **60**: 1189
- [13] Thery V, Rinald D, Rivail J L. *J. Comput. Chem.*, 1994, **15**: 269
- [14] Monard G, Loos M, Thery V, Baka K, Rivail J L. *J. Quantum Chem.*, 1996, **58**: 153
- [15] Gorb L G, Rivail J L, Thery V, Rinaldi D. *Int. J. Quantum Chem.*, 1996, **30**: 1525
- [16] Assfeld X, Rivail J L. *Chem. Phys. Lett.*, 1996, **263**: 100
- [17] Gao J, Amara P, Alhambra C, Field M J. *J. Phys. Chem. A*, 1998, **102**: 4714
- [18] Zhang Y, Lee T S, Yang W. *J. Chem. Phys.*, 1999, **110**: 1
- [19] Parr R, Yang W. *Density-Functional Theory of Atoms and Molecules*, Oxford University, New York, 1989.
- [20] Hehre W J, Radom L, Schleyer P V R, Pople J A. *Ab Initio Molecular Orbital Theory*, Wiley, New York, 1986.
- [21] Brooks B R, Bruccoleri R E, Olafson B D, States D J, Swaminathan S, Karplus M. *J. Comput. Chem.*, 1983, **4**: 187
- [22] Gao J, Xia X. *Science*, 1992, **258**: 631
- [23] Gao J, Thompson M A. *Combined Quantum Mechanical and Molecular Mechanical Methods*; American Chemical

Society: Washington, DC, Vol. 712, 1998.

- [24] Gao J, Furlani T R. *IEEE Comput. Sci. Eng.* , 1995, **2**: 24
- [25] Alhambra C, Wu L, Zhang Z Y, Gao J. *J. Am. Chem. Soc.* , 1998, **120**: 3858
- [26] Hartsough D S, Merz K M Jr. *J. Phys. Chem.* , 1995, **99**: 384
- [27] Merz K M Jr, Banci L. *J. Phys. Chem.* , 1996, **100**: 17414
- [28] Assfeld X, Ferre N, Rivail J L. *ACS Symp. Ser.* , 1998, **712**: 234
- [29] Liu H, Mueller-Plathe F, van Gunsteren W F. *J. Mol. Biol.* , 1996, **261**: 454
- [30] Chatfield D C, Brooks B R. *J. Am. Chem. Soc.* , 1995, **117**: 5561
- [31] Alhambra C, Gao J, Corchado J C, Villa J, Truhlar D G. *J. Am. Chem. Soc.* , 1999, **121**: 2253
- [32] Dewar M J S, Zorbisch E G, Healy E F, Stewart J J P. *J. Am. Chem. Soc.* , 1985, **107**: 3902
- [33] Dewar M J S, Thiel W. *J. Am. Chem. Soc.* , 1977, **99**: 4899
- [34] Bash P A, Field M J, Karplus M. *J. Am. Chem. Soc.* , 1987, **109**: 8092
- [35] Warshel A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*. John Wiley & Sons, New York, 1991.
- [36] Aqvist J, Warshel A. *Chem. Rev.* , 1993, **93**: 2523
- [37] Warshel A, Weiss R M. *J. Am. Chem. Soc.* , 1980, **102**: 6218
- [38] Chang Y T, Miller W H. *J. Phys. Chem.* , 1990, **94**: 5884
- [39] Kim H J, Hynes J T. *J. Chem. Phys.* , 1990, **93**: 5211
- [40] Schmitt U W, Voth G A. *J. Phys. Chem. B*, 1998, **102**: 5547
- [41] Vuilleumier R, Borgis D. *Chem. Phys. Lett.* , 1998, **284**: 71
- [42] Mo Y, Gao J. *J. Phys. Chem. A*, 2000, **104**: 3012
- [43] Gao J. *J. Am. Chem. Soc.* , 1994, **116**: 9324
- [44] Weiner S J, Singh U C, Kollman P A. *J. Am. Chem. Soc.* , 1985, **107**: 2219
- [45] Weiner S J, Seibel G L, Kollman P A. *Proc. Natl. Acad. Sci. USA.* , 1986, **83**: 649
- [46] Antonczak, Monard S, Ruiz Lopez G M F, Rivail J L. *J. Am. Chem. Soc.* , 1998, **120**: 8825
- [47] Cisneros G A, Liu H, Zhang Y, Yang W. *J. Am. Chem. Soc.* , 2003, **125**: 10384
- [48] Zhang Y, Kua J, McCammon J A. *J. Am. Chem. Soc.* , 2002, **124**: 10572
- [49] Zhang Y, Liu H, Yang W. *J. Chem. Phys.* , 2000, **112**: 3483