

· “双清论坛”专题:理论化学家视角中的仪器创制 ·

新一代高精度生物大分子力场

高加力*

(吉林大学理论化学研究所,美国明尼苏达州大学化学系,明尼苏达州 55455)

[摘 要] 精确地描述分子内和分子间相互作用是生物分子体系动力学计算和模拟的核心问题。势能面的精确度直接影响并决定了整个计算结果的可靠性和预言的准确性。因此,发展和应用精确势能面是生命科学计算中最至关重要的一个环节。可以预言,未来生命科学定量计算的发展途径仍然会与开发新一代更精确的生物分子势能面紧密相关。本文通过经典分子力场发展过程和现状为借鉴,探讨未来分子力场发展的可能趋势。

[关键词] 分子力场;量子力场;分子力学;量子化学分块法;生物分子计算

生物分子体系动力学计算和模拟的核心是描述分子内和分子间相互作用的势能面。势能面的精确度直接影响并决定了整个计算结果的可靠性和预言性。因此,发展和应用精确势能函数是生命科学计算中最至关重要的一个环节,也是从始至今一直投入最多,关注最大,影响最广的科学研究课题^[1]。未来生命科学定量计算的发展途径仍然会与开发新一代更精确的生物分子势能面紧密相关。本文通过经典分子力场发展过程和现状为借鉴,探讨未来分子力场发展的可能趋势。这里不包括粗粒化方法。

1 生物分子力场

势能面这一概念是建立在量子力学 Born-Oppenheimer 近似之下,原理上可以直接应用量子化学的方法得到精确的结果。然而受限于生物体系的大分子特征,精确的量子化学方法目前仍无法解决生命科学中的复杂问题。大分子体系的势能面,即体系的相互作用能量随原子坐标位置变化的关系,通常是用一系列的解析函数来近似表达的。因为不再直接考虑整个体系的电子结构,这一经典力学的方法也称为分子力学(示例-1 分子力学),而表达原子间相互作用的势能函数被统称为分子力场(虽然这里所指的乃是体系的势能)。经过半个多世纪来的优化和测试,分子力场目前在其适用范围内已经

达到了非常高的精确度。结合分子动力学计算可以用来帮助解析很多的实验现象,预测新功能,设计新实验,以及辅助开发新药物和酶催化反应。而且应用范围涵盖从水溶液中的单个蛋白质分子到病毒壳层和细菌光合作用体系的整体结构^[4,5],原子数目超过了 1 亿个以上。分子力场的精确度会继续提高和发展,包括引进多体极化效应,并有效的应用到更多更广的体系。

示例-1: 分子力学

分子力场是在 20 世纪 60 年代由 Scheraga 和 Lifson 等延伸到生物分子体系^[2],其势能函数表达式(方程 1)一直延续至今:

$$V = \sum_b \frac{1}{2} K_b (R_b - R_b^0)^2 + \sum_a \frac{1}{2} K_a (\theta_a - \theta_a^0)^2 + \sum_t \sum_n \frac{V_n}{2} [1 + \cos(n\varphi_t - \varphi_t^0)] + \sum_d \frac{1}{2} K_d (\delta_d - \delta_d^0)^2 + \sum_{i < j} \left\{ \epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{R_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{R_{ij}} \right)^6 \right] + \frac{q_i q_j}{R_{ij}} \right\} \quad (1)$$

这里,体系总势能可以分成两部分。第一部分包括前四项,用来描述分子内部的共价振动运动,分别表达化学键的伸缩,键角弯曲,分子内旋转(二面角扭转)和非共面扭转。式中的各个符号对应着各种运动的力常数、变量和平衡值。第二部分描述分子间非键

收稿日期:2017-10-18;修回日期:2017-11-12

* 通信作者,Email: jiali@jialigao.org

连相互作用,包括长程静电作用和范德华作用。

虽然半个世纪以来,能量表达形式(1)没有发生基本变化,但是其中的参数发生了质的改变。该式中的参数可以从量子化学计算获得,但更精确的方法是通过对小分子模型体系的热力学性质与实验结果比较优化而得到。我们假定这些通过优化模型体系得到的参数可转移并应用到真正大分子体系之中。

值得一提的是在1960年代末Lifson提出了生物大分子力场之后,Martin Karplus利用科研休假到他的实验室访问。回到哈佛大学之后Karplus等提出了分子力场与量子化学(QM/MM)结合的一种多尺度概念^[3],为此Karplus,Levitt和Warshel获得了2013年诺贝尔化学奖。

随着经典分子力学的发展,我们面临一个重要的问题:在今后半个世纪及更长远的未来,如何发展生命体系的分子力场?很显然,这里有两个简单的答案:

(1) 沿着既定过程,继续发展优化方法,如机器学习等,提高分子力学的精确性和应用范畴。

(2) 发展新的理论框架,超越经典力学的限制,建立新一代量子力学力场,为系统化提高生命科学计算的定量性和预言性奠定基础^[6]。

两个答案都非常重要,各有千秋。前者在近期会更为活跃,因为分子力学算法和计算软件都已经发展成熟。后者则需要发展新概念,建立新算法,开发新程序,但是有望对生命科学计算产生质的跃迁。因此为提高生物分子计算精确性,同时发展这两个方向都是必要的。

1.1 分子力学的优势和困扰

首先,从计算应用上考虑,分子力学在时间和空间尺度上占有绝对优势,其特点是计算速度快,可以应用到超大型的复杂体系并做长时间的分子动力学模拟。如上面提到的整个细胞的生物化学,物理过程。目前大家公认的发展方向集中于处理分子间相互作用产生的极化效应,在(1)式中引入极化能项,并已经有了显著进展^[7]。分子间极化效应来源于电子云密度的变化,属于量子效应,因此分子力学只能近似地表现描述这一现象,没有一个公认的极化能理论计算方法。应用最多的极化模型有诱导偶极子^[8,9],化学势平衡^[10],Drude振子^[11],和ABEEM原子—键浮动电荷理论^[12],其中ABEEM理论是中国的杨忠志和同事提出的^[12],并应用到分子间相互作用和液相计算^[13]。国际上发展最完整的是AMOEBA极化力场,由美国华盛顿大学的Ponder和

德克萨斯州立大学的任彭宇合作开发^[9]。此外分别应用极化偶极子^[8]和Drude振子方法^[11],极化效应也被引进到CHARMM力场。这里需要强调的是发展生物大分子的极化力场只是近年来才有明显进展,虽然已有较长历史,但仍处于起始阶段^[11]。因此显著超越非极化力场计算结果的优势目前还没有体现出来,有待更深入研究静电和极化作用的平衡。

然而分子力学也有一系列众所周知的明显缺陷^[6],包括(a)势能函数和参数都有极大的随机性使得不同力场参数数值非常不同并无法交叉使用;(b)虽然有非谐矫正但势能函数仍然主要依靠谐振子方程表达化学键运动;(c)没有系统途径处理多体极化效应,电荷转移,化学反应,和电子激发态过程。而且最困扰的问题是没有真正可行的办法来解决这些缺陷。

示例-2: 量子力场

量子力学力场^[6],简称量子力场,是一个分块化量子化学方法,其原理建立在分子体系的“化学局域性”。首先将整个分子体系划分成 N 个分子块,通过对子体系,即分子块分别进行量子化学计算而整合整个体系的总能量及其他性质。

分块量子化学方法的第一个近似是把整体系的波函数表达成各个分子块波函数的乘积(或分块电子云密度加和)。这一近似忽略了分子块之间的电子交换排斥和远程色散作用以及电荷离域效应,因此需要引入参数并做优化处理。分子块之间的长程相互作用不再涉及电子离域效应,主要包括双电子静电相互作用。当分子块相距足够远,量子化学中的双电子积分可以简化成单电子积分,并用量子化学和分子力学(QM/MM)组合迭代方法取代。这是量子力场的第二层近似。最后的第三层近似是对单一双电子积分做近似处理,如多极矩展开,密度拟合,及半经验积分表达等。可以看出如果不对任何一个层次做近似,量子力场的方法就自然的回到了相对应的从头算法水平。重要的是对于每一个分子块都可以用任何等级的,各块不同的波函数或密度泛函理论来表达。因此量子力场是一个系统化的全量子化学QM/QM组合方法。

分块化量子化学方法是当今国际上非常活跃的领域^[14],应用范围包括中小型分子的超高精度量子化学计算到生物大分子的动力学模拟。中国的几个小组的工作都处在各自领域的领先地位。主要工作包括张增辉的分子碎片共轭帽方法^[15],黎书华与合作者开发的广义能量分块法^[16],刘文剑发展的从块

到分子整合方法^[17]，和高加力在 1997 年美国纽约州立大学时就提出的显性极化理论^[6,18]。借鉴分子力学力场的成功经验，通过对上面提到的参数优化可以矫正由于分块近似产生的误差，与实验结果相拟合而达到生物分子计算所需要的精确度。量子化学分块法已经成功的应用于各种复杂的分子体系，如分子团簇，蛋白质在水溶液中的分子动力学和蛋白-配体相互作用和光谱等体系^[19]。

1.2 为什么发展新一代量子力场

分子力学中的理论困难完全来源于基于经典力学的近似。因此，如果我们能够应用量子化学方法求解生物大分子体系的薛定谔方程，用来表达分子势能面，即使是近似的甚至是半经验方法，分子力学所面临的这些几乎无法克服的困难都可以迎刃而解了。我们把直接应用量子化学表达分子体系势能面的方法定义为量子力学力场(简称量子力场)，而不是优化经典分子力学的参数重复量子化学计算结果(示例-2 量子力场)。量子化学具有通用性，适用于任何生物和非生物分子体系。非谐振势能面，电子波函数极化，电荷转移效应都自然而然的包括在量子化学方法之中。这样的一个量子分子力场又是一个可以完全处理基态和激发态化学过程的反应力场。

量子力场在两方面不同于量子化学。第一，量子力场是一种量子化学的近似方法，把整个分子体系做分块化近似处理，从而使得各种量子化学方法可以有效的应用到生物大分子体系。第二，量子力场包含经验参数，通过引入具有明确定义的参数来弥补由分块产生的近似误差。因为每个分子块的量子化学计算可以单独分别进行，量子力场可以很容易进行大规模的并行计算，有效利用当前国家超算中心的设备。借鉴于分子力学力场成功的经验，量子力场也可以通过优化少数，类似于密度泛函理论用来处理色散力的原子参数取得液相和生物大分子计算的精确度。量子力场面临的挑战是发展和建立系统化的近似方法，广泛应用各种量子化学波函数分子轨道理论(WFT)和密度泛函理论(DFT)计算方法，与实验结果相比较验证计算的精确性。

量子力场的分块化方法与量子化学计算的线性标度方法不尽相同。即使是后者真正达到了线性标度，其处理分子间相互作用的结果是不会超过所用的量化计算的精确性。因此，这些方法仍然满足不了液相和生物大分子动力学计算的需求。而量子力场包含的少数参数提供了优化凝聚态分子动力学计算结果的可能性。分子力学在生物大分子体系的成

功应用恰恰是因为有效利用了并优化了(1)式中的经验参数。换句话说，量子力场应用量子化学方法保证了正确描述生物分子体系势能面的定性性质，如电子极化，电荷转移，化学键生成和断裂，而这些性质恰恰是分子力学无法有效表达的。通过分块法，一方面降低计算标度，另一方面通过引进分块参数而优化量子力场的定量结果。所以量子力场同时涵盖了量子化学和分子力学二者的优势，使得生物大分子体系的精确量子化学计算成为可能。

2 建立和开发新一代分子力场

如何成功发展新一代生物大分子量子力场呢？我们仍然借助分子力场成功的先例来探讨这一实际并且重要的问题。如果以使用的广泛性作为标度的话，目前应用于生物大分子体系上最成功的分子力场有 AMBER,CHARMM,GROMACS,OPLS 非极化力场，以及 AMOEBA 极化力场。尽管每个力场的起点和参数优化路径都不尽相同，但是这些分子力场的成功并得到广泛应用有两个惊人的共同点：

(1) 结合力场的开发，同时发展自己的计算应用软件；

(2) 有长期且相对稳定的来源于政府部门的科研经费支持。

第一点其实很明显，前三个分子力学力场的名字和对应的软件完全相同，而后两个力场也有其应用软件，分别为 BOSS 和 Tinker。或许软件的开发促进了力场的发展，或许由于力场的成功使得计算软件得到广泛应用，可以肯定二者同时发展和开发起到了相互促进的作用，并为发展其它分子动力学计算理论和方法提供了条件。另外一个有趣的现象是，与量子化学软件不同，直到近期应用于生物大分子计算的各自软件只能应用自己的分子力场。这里再次体现了分子力场的局限性；即使力场所用的公式完全一样，力场参数还是不能直接应用到其他软件。但是这也说明软件开发和发展分子动力学算法之间的紧密关系。

第二点也极为重要，因为软件和力场的开发和优化是一个冗长和艰苦的工作，且需要大量人力物力又不一定是当时最热门的科研课题。在 1980 年代末发展当代 CHARMM 力场的过程中有 30 多人同时参与不同参数的优化。我们知道 CHARMM 力场到目前还在一直不断改进，包括修改参数和增加应用范围。因此稳定的课题支持也是方法和软件成功发展的前提。

3 总 结

本文以经典分子力场发展过程和现状为借鉴，

可以预言开发新一代更精确的分子间相互作用势能函数是保证提高生物大分子定量计算必要条件。未来生物分子力场的发展有两条途径。其一是在现有分子力学框架下继续优化力场参数,包括机器学习,增大应用范畴,进一步引进并优化分子间相互作用的极化效应。第二是引进新范式,超越经典力学限制,建立新一代的量子力学力场。后者是一个全新的概念,需要发展计算方法和软件,目前在上也刚刚处于兴起之势,具有发展潜力。此外,我们发现,成功发展并被广泛应用的分子动力学计算方法和分子力场具有两个共同特征,为今后方向提供了极好的借鉴。第一是方法和计算软件的发展紧密相连,相辅相成,相互促进从而保证二者的广泛应用。第二是具有长期的课题支持。生物大分子力场的开发是一个耗时,耗力,虽然极为重要但不一定是当时的热门科研课题,因此对这一战略科研方向的稳定支持亦是至关重要的。

参 考 文 献

- [1] MacKerell AD. Empirical force fields for proteins: Current status and future directions. *Ann Rep Comput Chem*, 2005, 1: 91 - 102.
- [2] Bixon M, Lifson S. Potential functions and conformations in cycloalkanes. *Tetrahedron*, 1967, 23(2): 769 - 784.
- [3] 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(16): 5612 - 5625.
- [4] Zhao G, Perilla JR, Yufenyuy EL, et al. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature*, 2013, 497(7451): 643 - 646.
- [5] Sener M, Strumpfer J, Singharoy A, et al. Overall energy conversion efficiency of a photosynthetic vesicle. *ELife*, 2016, 5: e09541.
- [6] Xie W, Gao J. Design of a next generation force field; the X-POL potential. *J Chem Theory Comput*, 2007, 3(6): 1890 - 1900.
- [7] Jorgensen WL. Special issue on polarization. *J. Chem. Theory Comput*, 2007, 3: 1877.
- [8] Gao J, Pavelites JJ, Habibollazadeh D. Simulation of liquid amides using a polarizable intermolecular potential function. *J Phys Chem*, 1996, 100(7): 2689 - 2697.
- [9] Shi Y, Xia Z, Zhang J, et al. Polarizable atomic multipole-based AMOEBA force field for proteins. *J Chem Theory Comput*, 2013, 9(9): 4046 - 4063.
- [10] Kaminski GA, Stern HA, Berne BJ, et al. Development of a polarizable force field for proteins via ab initio quantum chemistry: first generation model and gas phase tests. 2002, 23(16): 1515 - 1531.
- [11] Lemkul JA, Huang J, Roux B, et al. An empirical polarizable force field based on the classical drude oscillator model: development history and recent applications. *Chem. Rev.* 2016, 116(9): 4983 - 5013.
- [12] Yang ZZ, Wang CS. Atom-bond electronegativity equalization method and its applications based on density functional theory. *J Chem Theory Comput*, 2003, 2(02): 273 - 299.
- [13] Li H, Sun T T, Zhang C, et al. QM/MM (ABEEM) Study on the Ligand Substitution Processes of Ruthenium (III) Complex NAMI-A. *Chin J Chem*, 2017, 35(3): 354 - 362.
- [14] Gao J, Zhang ZH, Houk KN. Beyond QM/MM: Fragment quantum mechanical methods. *Acc Chem Res*, 2014, 47, 2711.
- [15] Zhang DW, Zhang JZH. Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein-molecule interaction energy. *J Chem Phys*, 2003, 119(7), 3599 - 3605.
- [16] Li S, Li W, Ma J. Generalized energy-based fragmentation approach and its applications to macromolecules and molecular aggregates. *Acc Chem Res*, 2014, 47(9): 2712 - 2720.
- [17] Li Z, Li H, Suo B, et al. Localization of molecular orbitals; from fragments to molecule. *Acc Chem Res*, 2014, 47(9): 2758 - 2767.
- [18] Gao J. Toward a molecular orbital derived empirical potential for liquid simulations. *J Phys Chem B*, 1997, 101(4): 657 - 663.
- [19] Xie W, Orozco M, Truhlar DG, et al. X-Pol potential: An electronic structure-based force field for molecular dynamics simulation of a solvated protein in water. *J. Chem. Theory Comput*, 2009, 5(3): 459 - 467.

Next generation force fields for biological macromolecules

Gao Jiali

(Theoretical Chemistry Institute, Jilin University Department of Chemistry, University of Minnesota Minneapolis, MN 55455 USA)

Abstract In the heart of molecular dynamics simulations and modeling of biological molecules is the potential energy surface that describes the internal and intermolecular interactions. The accuracy of the potential energy surface directly influences the reliability and predictability of these computations. Consequently, it is a critically important step to develop and apply accurate potential energy functions for macromolecular computations. Undoubtedly, future progresses towards quantitative computational biology will also coincide with the development of more accurate, next generation force fields. Inspired by the successes of classical molecular mechanical force fields, this article analyzes possible future progresses.

Key words molecular mechanics; quantum mechanical force field; next generation force field; fragmental quantum chemistry; computation of biological molecules