

自由能微扰理论及其应用研究进展

邹慧莹 柳高川 琚寅 郭阳涛 于涛*
江汉大学光电材料与技术学院 湖北武汉 43005

摘要: 自由能微扰 (Free Energy Perturbation, FEP) 理论是微扰理论重要分支, 通过引入微小扰动参数逐步改变系统哈密顿量, 从而计算不同系统状态之间自由能差, 从理论上提高自由能计算精确度。该方法适合从已知系统出发, 经过一系列结构状态之间转变的复杂体系, 通过FEP方法精确计算分子在不同状态之间的自由能差异, 其结论对揭示分子结构转变行为及其相互作用至关重要。本研究聚焦于自由能微扰理论基础原理及其理论演化过程, 重点分析近年来FEP理论在药物设计和分子动力学模拟中的实际应用, 以促进理论领域深入探索与计算技术进步。

关键词: 自由能微扰理论; 分子动力学模拟; 药物设计

一、自由能微扰理论发展历程

FEP最初理论建立在统计力学基础之上, 通过将分子系统从初始状态逐步转换到最终状态来计算自由能变化。该方法假定存在一个介于0和1之间的参数 λ , 用来插值从初态到终态的所有中间态。通过不同 λ 值下运行分子动力学模拟, 可以收集数据以估算自由能差^[1]。

随着量子力学/分子力学 (Quantum Mechanics/Molecular Mechanics, QM/MM) 等方法的发展, FEP理论得到了扩展, 形成了FEP+方法。允许在计算过程中同时考虑量子力学细节 (如电子密度分布) 和分子力学级别的粗粒化描述。将量子力学计算与经典分子力学计算相结合, 将为特定类型的化学反应提供更准确的自由能数据。FEP+适用于研究化学键断裂和形成的系统如酶催化反应, 这样可以提供更准确的能量描述。早期FEP方法使用隐式溶剂模型来简化计算, 但该方法忽略了真实水分子的影响。显式溶剂FEP (Explicit Solvent FEP, ES-FEP) 考虑了周围水分子的具体位置和相互作用, 从而提供了更真实的溶液环境, ES-FEP方法结合分子动力学模拟可得到更高精度的结合自由能。

图形处理单元 (GPU) 计算能力的提高和并行计算

技术的发展使FEP计算效率显著提升, 这样在合理的时间尺度内处理大规模的分子系统成为可能, 进一步拓宽了FEP的应用范围。绝对FEP (Absolute FEP, AFEP) 与相对FEP (Relative FEP, RFEP) 逐步被发展和完善, 并且在功能上相辅相成。绝对FEP只需计算单个状态自由能, 而相对FEP则计算两个状态之间自由能差, 在实际应用中更为常见。同时绝对FEP被用来校准相对FEP结果, 共同提供更全面的能量分布。

随着计算能力提升和算法发展, 为了克服传统FEP方法在采样不足问题, 研究人员开发了增强采样策略。FEP方法结合增强采样策略可以更有效地探索复杂能量分布, 提高了计算收敛速度和可靠性。常用的增强采样算法有复制交换分子动力学 (λ -REMD)^{[2][3]}、溶质回火 (REST) 复制交换^{[4][5]}以及溶质缩放 (REST2) 副本交换^[6], 上述方法可提高FEP计算结果的收敛性^[7]。Martins等人于2021年开发了PyAutoFEP开源工具, 可设置相对FEP扰动。PyAutoFEP由Python3编写, 将REST和REST2两种增强采样方法相结合, 可自动生成扰动图、双拓扑及系统构建和分子动力学模拟分析所需参数, 并且支持多种力场并允许沿扰动窗口灵活 λ 值, 用于自动设置和分析FEP计算^[8]。

近年来, 机器学习技术被引入FEP框架中, 以加速自由能计算。2023年Bhat等人描述了一种称为FEP协议生成器 (FEP-PB) 自动化工作流程, 为使用默认设置而表现不佳的系统快速生成准确FEP协议。FEP-PB使用主动学习工作迭代搜索协议参数空间, 可以在有限的人工干预下为MCLI系统快速生成准确的FEP方案, 开

基金项目: 湖北省自然科学基金 (2022CFB474); 江汉大学科研项目 (2021yb026, 2022XKZX12, 2023KJZX33)。

作者简介: 邹慧莹 (2000.8-), 女, 汉族, 江西樟树人, 硕士研究生, 研究方向: 新一代电子信息技术。

通讯作者: 于涛 (1982.5-), 男, 汉族, 山东高青人, 副教授, 博士, 研究方向: 计算物理化学。

发准确FEP协议并增加适合该技术靶标数量^[9]。利用深度学习的方法如(Deep Potential, DP)和机器学习力场(Machine Learning Force Fields, MLFFs),可以在降低计算量的同时依然保持对分子间相互作用力的高精度预测。2024年, Crivelli-Decker等人利用机器学习引导绝对自由能扰动工作流程,提出了一个集成工作流程,可以有效地筛选大型和多样化的化学数据库,将主动学习与基于快速绝对自由能扰动方法和基于物理配分函数相结合,实现了迄今为止最大的免费能源模拟数据集^[10]。

二、自由能微扰理论基本原理

(一) 基本原理

微扰理论是一种结合量子力学与统计力学的基础理论方法,用于处理复杂系统内部相互作用。其基本思想为在已知系统基础上,将系统哈密顿量 H 分解为已知部分和一个微小扰动项。通过对扰动部分进行统计分析,可以得到扰动后系统哈密顿量变化差异。这种方法要求系统中扰动部分很小以获得更高的结果准确性。系统哈密顿量 H 可以表示为: $H(\lambda) = H_0 + V$ 。其中 H_0 是未扰动系统的哈密顿量, V 是扰动项, λ 为介于0和1之间的微扰参数,用于控制扰动项强度。当 $\lambda=0$,系统处于未扰动状态 H_0 ,当 $\lambda=1$ 时,系统完全转变为扰动状态 $H_1 = H_0 + V$ 。

自由能(Free Energy)是衡量系统在一定条件下对外做功的能力,它包含了系统内能、熵和温度等因素,是一个综合状态函数。在温度为 T 时,自由能定义为: $F(\lambda) = -k_B T \ln Z(\lambda)$ 。其中 $Z(\lambda)$ 为系统的配分函数, k_B 为玻尔兹曼常数, T 为系统温度。自由能微扰理论是一种用于计算不同状态之间自由能差异的方法,一般通过计算从 $\lambda=0$ 到 $\lambda=1$ 之间的自由能变化 $\Delta F = F(1) - F(0)$ 来得到两种状态之间的自由能差。根据微扰理论,自由能的变化可以表示为

$$\Delta F = -k_B T \ln \exp\left(-\frac{H_1 - H_0}{k_B T}\right)_0$$

(二) 自由能计算方法

计算结合自由能的方法分为三大类:(1)终点方法,只需要模拟初始和最终状态;(2)炼金术方法,蛋白质配体复合物在连接两个端点的热力学状态下模拟;(3)严格物理学运动路径采样方法。

终点方法通过提供构象采样准确性和计算成本之间实现了良好的平衡,常用线性相互作用能(LIE)方法、分子力学泊松-玻尔兹曼表面积(MM-PBSA)和分子力

学广义Born表面积(MM-GBSA)^[11]。MM-PBSA和MM-GBSA只需要模拟一个终点,但是无法区分彼此靠近的结合能。由后续发展出炼金术方法,涵盖了熵值变化和(去)水合产生的影响,采样算法在炼金术方法中可实现多种软件解决方案。其中哈密顿副本交换方案应用广泛,得到了如GROMACS^[12]、GROMOS^[13]、FEP+^[14]和NAMD^[15]等软件支持,每次特定模拟可以通过Metropolis-Hastings接受准则与相邻 λ 值进行交换。这类算法不需要预先了解必须施加偏差势能自由度^{[16][17]}。此外还考虑变化过程中化学效应,例如质子化态和互变异构体分布。沿着炼金术路径,引入并模拟离散中间状态,将过渡期分为更小步骤。状态之间重叠增加并允许自由能计算在相对较短时间内收敛。

三、自由能微扰理论主要应用领域

自由能微扰理论作为一种强大的计算工具,在药物设计和分子动力学模拟等领域应用广泛,极大推动了科学研究和问题解决。在药物设计领域,FEP可以帮助研究人员计算不同配体或药物分子与靶标的结合自由能差异,通过该方法可以评估药物分子有效性和选择性,优化药物结构,提高药物设计成功率。在分子动力学模拟领域中,FEP方法被广泛应用于研究复杂分子系统的自由能变化。通过结合分子动力学模拟动态信息,统计系统在不同状态下的自由能差异,进而可以研究分子在溶液中的稳定性、蛋白质折叠过程、大分子结构间的相互作用等。

(一) FEP在药物发现中的应用

FEP计算在药物发现中潜在应用很早就已开展。1935年Kirkwood理论为自由能差异计算估计奠定了基础^[18]。随后FEP数学计算方式取得了进步,将系统初始状态和最终状态之间自由能差与通过对初始状态进行采样来评估能量差的函数平均值联系起来^[19]。1985年,Jorgensen首次应用FEP计算来估计有机分子溶剂化自由能^[20]。近年来,在GPU计算技术发展背景下,MD软件计算所用时间大幅减少,具体取决于系统大小和模拟设置。Jiang等人采用了FEP与哈密顿复制交换分子动力学(FEP/H-REMD)结合方法,为计算配体与受体绝对结合自由能提供了一个有效技术路径^[7]。这种方法通过在不同化学状态之间转换,允许更精确地估算配体在复杂生物分子系统中结合强度。通过改进在NAMD软件中自由能扰动理论计算方法,以提高配体和受体结合亲和力预测精度。该方法能够更精确地模拟和计算不同药物分子

与生物大分子之间的相互作用，这对药物筛选和优化过程至关重要。

FEP还可以通过模拟药物分子与受体相互作用，高效评估药物分子与受体蛋白结合模式和活性。Singh等人在2020年介绍了使用FEP计算高度灵活蛋白质MDM2及其抑制剂之间的绝对结合自由能^[22]。该研究采用综合计算协议，结合了绝对结合自由能计算和自由能分布方法，显著提高了复杂系统结合自由能预测精度。该方法能够更准确地模拟和理解药物分子与蛋白质靶点之间的相互作用，从而优化药物设计。Wu等人在2021年研究中，对EAI001和EAI045与EGFR复合物进行了分子动力学模拟和自由能计算，揭示了EAI001和EAI045作为表皮生长因子受体（EGFR）变构抑制剂的详细抑制机制，为EGFR变构抑制剂的合理药物设计提供依据^[23]。Vinícius等人在2021年得到FEP在预测可逆共价抑制剂结合亲和力方面应用，通过FEP计算得到了与实验数据高度一致的结果^[23]。这表明FEP作为一种高效计算方法能够帮助优化配体分子构造，在药物设计领域展现了显著应用价值。

FEP方法在药物发现研究过程中得到进一步拓展。Li等人在2023年利用FEP进行了大规模虚拟筛选，寻找针对特定靶标的潜在药物分子。研究通过计算数千种化合物与靶标蛋白的结合自由能，成功筛选出了一批具有高活性候选药物^[24]。研究人员可快速排除那些理论上不适合作为药物候选的化合物。通过计算潜在配体与靶标蛋白结合自由能，优先选择具有较高结合亲和力的分子进行进一步研究。

（二）分子动力学模拟中应用

在分子模拟领域，MacKerell团队在2018年研究中通过CHARMM-GUI Ligand Binder构建了对二甲苯和正丁基苯的广义CHARMM力场参数^[25]。利用FEP计算对二甲苯和正丁基苯与T4溶菌酶L99A突变体非极性空腔的绝对结合自由能。研究表明，该方法有效增强结合袋周围主链和侧链扭转运动采样，并加速了自由能计算收敛^[26]。这一研究验证了FEP在计算化学中有效性，为分子模拟提供了可靠基础。

2019年，新冠病毒（SARS-CoV-2）引发全球冠状病毒流行疾病（COVID-19），其表面刺突蛋白（S蛋白）是病毒入侵宿主细胞的关键。为了更好地理解S蛋白的功能以及其与宿主细胞的相互作用，来自韩国首尔国立大学、英国剑桥大学和美国里海大学研究人员共

同合作，建立了首个全长S蛋白开源全原子模型^[27]。在这项工作中，研究团队采用了CHARMM力场对蛋白质、脂类和碳水化合物进行建模，并使用Solution Builder和Membrane Builder模块构建模型。该研究主要集中在S蛋白建模和模拟。通过结合FEP模拟计算可以进一步探究S蛋白与其他分子之间结合模式，预测潜在药物结合位点，并优化现有抑制剂活性，从而为预防和治疗COVID-19提供理论支持。Zhang等人在2020年研究中，通过自由能微扰理论结合分子动力学模拟，预测了瑞德西韦（Remdesivir）与SARS-CoV-2 RNA聚合酶（RdRp）之间结合模式^[28]。研究预测了瑞德西韦可能的结合位点，还指出了关键氨基酸，为开发更高活性抑制剂提供了理论指导。在2022年，陈前及其团队运用微扰理论与蒙特卡罗（Monte Carlo N Particle Transport Code, MCNP）模拟，提出高效伽马-伽马密度计算方法，并以双源距补偿密度仪器标准参数为参考，构建了不同基准密度地层条件扰动灵敏度函数数据库^[29]。

（三）自由能微扰理论应用前景

集成深度学习和人工智能领域：随着机器学习和人工智能技术的不断成熟，将上述技术应用于FEP计算中可以显著提高模拟准确性和效率。深度学习模型中卷积神经网络（CNN）可用于处理分子结构化数据；循环神经网络（RNN）和长短期记忆网络（LSTM）可用于处理序列数据，将这些技术加入到FEP理论中可以更为预测分子间相互作用结果，进而优化模拟参数和提高计算速度。

多尺度模拟发展领域：传统FEP方法在处理大规模、复杂生物系统时面临着计算成本高和精度有限问题。为了克服这些挑战，FEP将越来越多地与多尺度建模方法相结合，从而在原子水平上精确描述分子间相互作用，同时保持计算可行性。通过将FEP与多尺度建模方法相结合，可以显著提高复杂生物分子系统模拟精度和计算效率。量子力学/分子力学混合方法、粗粒化模型、多尺度采样技术和机器学习辅助的多尺度建模都是实现这一目标有效途径。未来研究将进一步探索这些方法的结合，以应对更广泛的生物物理问题，并推动计算生物学和药物设计领域的发展。

自动化计算软件工具领域：传统FEP计算通常需要复杂参数设置和手动操作，为了减少耗时提高效率，运用高度自动化和易于使用FEP软件工具将成为趋势。CHARMM-GUI服务器是一个基于Web图形用户界面，旨

在简化复杂分子模拟任务, 提供了从系统构建到模拟设置一站式解决方案。极大地降低了用户使用门槛, 使非专家也能轻松进行复杂自由能计算, 从而扩大FEP方法应用范围。CHARMM-GUI致力于提升自动化水平、改善用户体验、扩展功能与适用范围, 以及高效管理计算资源和完善数据管理与共享机制, 为用户提供强大易用平台, 推动FEP方法在药物设计和生物物理研究等领域广泛应用与进步。

结论

本文系统介绍了FEP的基本原理和发展历程, 并与现代多尺度建模方法结合, 不断优化和完善以解决复杂生物分子系统结构稳定性。着重探讨FEP近年来在药物设计和分子动力学模拟中的应用。在药物发现、筛选及优化方面展现显著优势, 结合分子动力学模拟加速收敛有效采样。随着高效算法和GPU技术发展, 基于用户web平台CHARMM-GUI服务器实现自由能扰动模拟自动化运行, 并且能够处理大规模计算任务以及更复杂生物分子系统, 提高模拟准确性和可行性。未来FEP将结合多尺度建模方法和机器学习进一步提升计算准确性和效率。

参考文献

[1]Kollman P. Free energy calculations: applications to chemical and biochemical phenomena[J]. Chemical Reviews, 1993, 93(12):2395-2417.

[2]Jiang W, Phillips J C, Huang L, et al. Generalized scalable multiple copy algorithms for molecular dynamics simulations in NAMD[J]. Computer Physics Communications, 2014, 185(3):908-916.

[3]Woods C J, Essex J W, King M A. The development of replica-exchange-based free-energy methods[J]. The Journal of Physical Chemistry B, 2003, 107(49):13703-13710.

[4]Liu P, Kim B, Friesner R, et al. Replica exchange with solute tempering: A method for sampling biological systems in explicit water[J]. Proceedings of the National Academy of Sciences of the United States of America, 2005, 102(39):13749-13754.

[5]Fratev F, Sirimulla S. An improved free energy perturbation FEP+ sampling protocol for flexible ligand-binding domains[J]. Scientific Reports, 2019, 9(1):16829.

[6]Wang L, Friesner R A, Berne B J. Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2) [J]. The Journal of Physical Chemistry B, 2011, 115(30):9431-9438.

[7]Jiang W, Chipot C, Roux B. Computing relative binding affinity of ligands to receptor: An effective hybrid single-dual topology free energy perturbation approach in NAMD[J]. Journal of Chemical Information and Modeling, 2019, 59(9):3794-3802.

[8]Martins L C, Cino E A, Ferreira R S. PyAutoFEP: An automated free energy perturbation workflow for GROMACS integrating enhanced sampling methods[J]. Journal of Chemical Theory and Computation, 2021, 17(7):42662-4273.

[9]Bhat M B S, Emailprotected E, et al. FEP protocol builder: Optimization of free energy perturbation protocols using active learning[J]. Journal of Chemical Information and Modeling, 2023, 63(17):5592-5603.

[10]Crivelli-Decker J E, Beckwith Z, Tom G, et al. Machine learning guided AQFEP: A fast and efficient absolute free energy perturbation solution for virtual screening[J]. Journal of Chemical Theory and Computation, 2024, 20(16):7188-7198.

[11]Wang E, Sun H, Wang J, et al. End-point binding free energy calculation with MM/PBSA and MM/GBSA: Strategies and applications in drug design[J]. Chemical Reviews, 2019, 119(16):9478-9508.

[12]Berendsen H J C, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation[J]. Computer Physics Communications, 1995, 91(1-3):43-56.

[13]Kunz A P E, Allison J R, Geerke D P, et al. New functionalities in the GROMOS biomolecular simulation software[J]. Journal of Computational Chemistry, 2011, 33(3):340-353.

[14]Wang L, Wu Y, Deng Y, et al. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field[J]. Journal of the American Chemical Society, 2015, 137(7): 2695-2703.

[15]Phillips J C, Braun R, Wang W, et al. Scalable

molecular dynamics with NAMD[J]. *Journal of Computational Chemistry*, 2005, 26(16): 1781–1802.

[16]Jiang W, Roux B. Free Energy Perturbation Hamiltonian Replica-Exchange Molecular Dynamics (FEP/H-REMD) for Absolute Ligand Binding Free Energy Calculations[J]. *Journal of Chemical Theory & Computation*, 2010, 6(9):2559–2565.

[17]Meng Y, Sabri Dashti D, Roitberg A E. Computing alchemical free energy differences with hamiltonian replica exchange molecular dynamics (H-REMD) simulations[J]. *Journal of Chemical Theory & Computation*, 2011, 7(9):2721–2727.

[18]Kirkwood J G. Statistical mechanics of fluid mixtures[J]. *Journal of Chemical Physics*, 1935, 3(5): 300–313.

[19]Zwanzig R W. High-temperature equation of state by a perturbation method. I. Nonpolar gases[J]. *Journal of Chemical Physics*, 1954, 22(8):1420–1426.

[20]Jorgensen W L, Thomas L L. Perspective on free-energy perturbation calculations for chemical equilibria[J]. *Journal of Chemical Theory and Computation*, 2008, 4(6):869–876.

[21]Singh N, Li W. Absolute binding free energy calculations for highly flexible protein MDM2 and its inhibitors[J]. *International Journal of Molecular Sciences*, 2020, 21(13):4765.

[22]Wu X, Guo Q, Li Q, et al. Molecular mechanism study of EGFR allosteric inhibitors using molecular dynamics simulations and free energy calculations[J]. *Journal of Biomolecular Structure & Dynamics*, 2021(6):1–10.

[23]Vin ícius B, Shamim A, Rocho F D R, et al. Predicting the relative binding affinity for reversible covalent inhibitors by free energy perturbation calculations[J]. *Journal of Chemical Information and Modeling*, 2021, 61(9):4733–4744.

[24]Li Z, Wu C, Li Y, et al. Free energy perturbation-based large-scale virtual screening for effective drug discovery against COVID-19[J]. *International journal of high performance computing applications*, 2023, 37(1):45–57.

[25]Vanommeslaeghe K, Hatcher E, Acharya C, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields[J]. *Journal of Computational Chemistry*, 2009, 31(4):671–690.

[26]Jiang W, Thirman J, Jo S, et al. Reduced free energy perturbation/hamiltonian replica exchange molecular dynamics method with unbiased alchemical thermodynamic axis[J]. *The Journal of Physical Chemistry B*, 2018, 122(41):9435–9442.

[27]Woo H, Park S J, Choi Y K, et al. Developing a fully-glycosylated full-length SARS-CoV-2 spike protein model in a viral membrane[J]. *The Journal of Physical Chemistry B*, 2020, 124(33):7128–7137.

[28]Zhang L, Zhou R. Structural basis of the potential binding mechanism of remdesivir to SARS-CoV-2 RNA-dependent RNA polymerase[J]. *The Journal of Physical Chemistry B*, 2020, 124(32):6955–6962.

[29]陈前, 张锋, 梁启轩, 等. 基于微扰理论的伽马-伽马密度测井快速正演方法[J]. *地球物理学报*, 2022, 65(4): 8.