中国科学: 化学

2014年 第44卷 第6期:854~863



SCIENTIA SINICA Chimica

评 述

化学热力学专刊

www.scichina.com chem.scichina.com

自由能计算方法及其在生物大分子体系中的 适用性问题

张志森",王琦",陈尔余"?"

浙江大学化学系,浙江大学软物质研究中心,杭州 310027
 武警杭州士官学校车辆维修系油料教研室,杭州 310023
 *通讯作者, E-mail: qiwang@zju.edu.cn

收稿日期: 2014-01-03; 接受日期: 2014-04-16; 网络版发表日期: 2014-05-15 doi: 10.1360/N032014-00005

摘要 本文总结了计算科学领域中被广泛使用的自由能计算方法,详述了各种 自由能计算方法的基本原理、操作过程以及所具有的特点,并在最后对各种方法 在生物大分子体系中的适用性做了概括,以期使读者在自由能相关研究工作中, 根据自身研究体系和研究目的,选择合适的计算方法开展研究工作. 关键词 自由能计算 生物大分子 Umbrella Sampling Free Energy Perturbation Jarzynski Equality Matadynamics

1 引言

自由能的计算在计算科学领域是一个非常重要 的研究课题,所有物理、化学以及生物过程自发进行 达到的程度都是由初态(反应物)与终态(产物)之间的 自由能变化决定的. 对于一个反应来说, 反应的自由 能差对应于该反应的热力学上的可逆功.目前,自由 能的计算方法有多种类型, 经典的方法有 Umbrella Sampling^[1]、Free Energy Perturbation^[2~4]和 Thermodynamic Integration^[5]等,这类方法原理上比较严谨, 计算结果也比较精确,但是需要大量的数据采集,所 需的计算资源很多,在实际应用上受到体系大小以 及计算资源的限制. 最新发展出来的自由能计算方 法有 Jarzynski Equality^[6]、Metadynamics^[7]等, 其中 Jarzynski Equality 由大量不可逆功的系统平均值来求 解反应过程的自由能数据,但是在具体使用时,此方 法的准确度受研究体系的影响很大; 而 Metadynamics 则是不断地对研究体系施加额外的高斯型排斥 势能函数,使目标分子能够脱离自由能的势阱,访问 到高自由能区域,进而通过施加的排斥势能函数来 求解自由能数据,在求解自由能二维势能面上有很 大的优势.

在生物学领域中,生物大分子之间的相互作用^[8-10]、蛋白质折叠^[11-14]、离子通道^[15]以及生物分子 与材料的相互作用^[16-20]等都是很重要的研究课题. 在这些研究中,对生物大分子结构与功能,和对新设 计的生物分子或生物材料的评价是很重要,都可以 通过对其能量学的考察得到揭示^[21,22].然而,由于上 述各种自由能计算方法的原理不同,它们在具体操 作的步骤,计算自由能的效率,以及对研究体系的适 用性等方面都有各自的特点,因此在计算体系自由 能数据的时候,需要根据研究体系的不同,计算资源 的多少来选定特定的计算方法.下面我们对各个方 法的原理、计算效率以及在生物大分子体系中的适用 性等方面进行阐述,使读者对其能有一个较为全面 的了解.

2 自由能计算方法

2.1 Umbrella Sampling

根据统计热力学的基本原理可知,系统的自由 能与研究体系的概率分布有关.对于系统从状态0转 变为状态1的过程,其自由能差值可以由研究系统在 两个状态下的概率分布得到:

$$\Delta G = G_1 - G_0 = -k_B T \ln \frac{\rho_0}{\rho_1} \tag{1}$$

因此对于一些特殊体系的自由能数据,我们可 以统计出体系在所要研究两种状态的概率分布之比, 进而使用公式(1)来计算相关的自由能数据.在实际 应用中,人们往往对自由能数据随着给定的反应坐 标*č*的变化趋势比较关注,这种自由能随着某个反应 坐标变化的曲线称为平均力势(potential of mean force, PMF).例如,计算体相水中两个水分子分离的 自由能曲线时,可以通过分子动力学模拟,得到两个 水分子直接的径向分布函数(radial distribution function, RDF_{*i*}),然后即可通过公式(2)得到相关的 PMF 数据:

$$PMF_{\xi} = -k_{B}T\ln\frac{RDF_{\xi}}{RDF_{0}}$$
(2)

其中 ζ 为反应坐标,即水分子分离的距离, RDF₀ 为选 择的自由能参考点的数据.因此对于概率分布求解 简单的体系,可以直接通过概率分布求解相关自由 能数据.然而在实际体系中,由于受模拟时间的限制, 研究体系(目标分子)绝大多数的时间都处在自由能 低的区域,一般的动力学模拟很难采集到其在自由 能高的区域的分布,无法利用公式(1)求解相关自由 能数据.Umbrella Sampling 就是为了解决高自由能区 域的取样问题而提出来的.

1977 年, Umbrella Sampling 最早由 Torrie 和 Valleau 提出^[1]. 在这个方法中,为了解决普通分子 动力学在高自由能区域的取样问题,整个反应路径 被分为多个区域,这些区域被称为 Umbrella Sampling 的窗口. 在每一个窗口 *i* 中,一个额外的势能项 被添加到研究体系的哈密顿量中:

$$H(x,p) = H(x,p) + h_i(r)$$
 (3)

其中x和p分别表示研究体系的位置和动量,而额外加的项为一个静态的谐振势,它是反应坐标r = r(x)的函数,对于给定的中心位置 r_0 ,则有:

$$h_i(r) = \frac{1}{2}k(r - r_0)$$
(4)

谐振势 *h_i*(*r*)的作用是把研究体系的位置限制在中心 位置 *r*₀ 附近, 从而加强研究体系在这段区域的取样.

整个反应路径被分割成许多段很小的区域,这样的区域在 Umbrella Sampling 中称为窗口.对每一 个窗口,系统都被谐振势限制在对应的位置 r_0 .在对 应的谐振势的限制下,通过分子动力学模拟,我们就 能够得到研究体系在反应路径上的一系列的有偏差 的概率分布 $\rho_i(r)$.需要指出的是,这里相邻的两个窗 口的概率分布之间要有足够重叠的部分,才能得到 准确的自由能曲线.然后这一系列的概率分布 $\rho_i(r)$ 再 通过权重直方分析法(weighted histogram analysis method, WHAM)^[23, 24]组合起来,去除掉额外添加 的谐振势的影响,从而得到研究体系沿着反应坐 标 ξ 的准确的概率分布,进而求解得相关的自由能 曲线.

由于 Umbrella Sampling 计算自由能具有严格的 理论依据,因此可以得到非常准确的自由能数据.例 如:本课题组^[25]在研究纳米碳管封装多肽的工作中, 采用了 Umbrella Sampling 的方法计算了多肽分子在 不同手性的纳米碳管中封装过程的自由能数据,揭 示了纳米碳管的手性对其封装多肽分子的影响. 然 而,取样需要研究体系大量的平衡轨迹和计算资源 进行数据采集,因此对计算资源以及研究体系的大 小有很多要求和限制. 在生物体系中, 对于 Umbrella Sampling 来说, 溶于水溶液的生物大分子体系是非 常大的,所需的大量的计算资源是 Umbrella Sampling 在该研究领域推广的最大阻力.此外, Umbrella Sampling 还要求相邻窗口之间拥有足够的 重叠部分.对于很多体系而言,某些自由能较高的区 域需要很多次的尝试,才能得到较好的取样,这不仅 增加了所需的计算资源,拖延了计算周期,更降低了 操作的便捷性.因而,在生物大分子领域,相对于在 实际研究体系中的应用, Umbrella Sampling 更多地应 用于对新提出的自由能计算方法准确性的检验 上[26~30]

2.2 Free Energy Perturbation

Free Energy Perturbation 由 Zwanzig 于 1954 年提出^[3]. Zwanzig 将研究体系的两个状态(初态和终态)的自由能差与其能量差值的系统平均联系到了一起:

$$\Delta G = G_1 - G_0 = -kT \ln \left\langle \exp \frac{-(E_1 - E_0)}{kT} \right\rangle$$
(5)

之后, Zwanzig 又进一步得到了计算自由能差值的幂 级数展开式. 将初态和终态的能量分别定义为 *E*₀ 和 *E*₁, 这两个状态由一个很小的干扰势能 *V* 连接起来, 就得到公式(5)的二阶展开项:

$$E_1 = E_0 + V \tag{6}$$

$$\Delta G = G_1 - G_0 = \langle V \rangle - \frac{1}{2kT} (\langle V^2 \rangle - \langle V \rangle^2)$$
 (7)

这样研究体系初态和终态之间的自由能差值就能由 干扰势能的系统平均值计算得到.这种方法的有效 性已经被很多体系所证实,尤其是波动服从高斯概 率分布体系^[31-34].

上述公式只能用来计算初态和终态之间只有较 小差异的研究体系,即干扰势能 V 很小的体系.而在 实际应用过程中,所计算自由能差的初态和终态往 往有比较大的差异,这就使得上述公式不能直接应 用到研究体系中.解决方法是在初态和终态之间加 入一系列的中间态:引入耦合变量 J,通过微调 J 人 为制造出一系列只有微小差异的中间态,每一个中 间态的能量为初态和终态的耦合:

$$E(\lambda_i) = E_0 \lambda_i + E_2 (1 - \lambda_i)$$
(8)

然后通过公式(7)求解相邻两个中间态的自由能差值 Δ*G_i*,最后将所有中间态之间的自由能差值加和,就 得到初态和终态之间的自由能差:

$$\Delta G = \sum_{i}^{n} \Delta G(\lambda_{i \to i+1}) \tag{9}$$

Free Energy Perturbation 计算自由能的基本思想就是 从一个已知的状态出发,通过一系列的微小变化,使 之达到终态,在每一个中间态上进行分子动力学模 拟,得到体系的势能变化,最终求解得到初态和终态 的自由能差.

根据λ的间隔大小,可以将 Free Energy Perturbation 分为三种方法:固定步长,慢增长以及动态步 长^[21, 22].在固定步长方法中,相邻两个中间态的间隔 是固定的.在慢增长方法中,中间态的数据趋近于无 穷多个,相邻两个中间态之间的差异接近于零.在动 态步长方法中,每一步微扰的步长是可变的,系统根 据上一次微扰的自由能数据变化大小确定下一次微 扰的步长.

根据热力学原理构建热力学循环以后, Free

Energy Perturbation 就可以用来计算相关反应过程的 自由能变化,对于一般的研究体系,需要使用 Free Energy Perturbation 进行两次自由能计算, 再根据构 建好的热力学循环,得到最终的自由能变化.然而, 由于需要对一系列的中间态进行取样, Free Energy Perturbation 计算自由能的过程就需要消耗大量的计 算资源,在计算效率上与其他方法有较大的差距.由 于计算过程只涉及到初态和终态,中间的状态是人 为设置的, 所以最终结果只有一个自由能变化的数 据,不像其他方法那样具有反应坐标以及对应的自 由能数据变化趋势.因此对于小分子体系来说,使用 Free Energy Perturbation 方法可以得到较为准确的自 由能数据^[35~38],但是涉及到生物大分子体系时,此方 法会给计算带来非常大的压力,可以截取关键片段 进行相关计算,或者使用其他计算效率相对较高的 方法.

2.3 Thermodynamic Integration

Thermodynamic Integration 的基本计算公式是由 Kirwood 提出的^[5]. 在这个方法中,与 Free Energy Perturbation 比较类似的是,研究系统的两个状态的 自由能变化也由一个耦合变量 λ 连接,初态时 $\lambda = 0$, 终态时 $\lambda = 1$. 通过连续改变 λ ,得到一系列介于初态 和终态的中间态,通过对这些中间态的取样,最终计 算得到自由能变化数据. 以 *H* 表示研究体系在某个 状态 λ 下的能量,则其具体的自由能计算公式为:

$$\Delta G = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_{\lambda} \mathrm{d}\lambda \tag{10}$$

在实际操作中,自由能能量的动力学部分并没有被 包含在这里面,这是因为在计算中,可以根据相关基 本原理明确地得到动力学部分的解析解,同时在利 用热力学循环求解研究体系自由能变化时,动力学 的贡献会相互抵消.

在 Thermodynamic Integration 方法中, 计算自由 能的基本思想是研究系统在无限长时间内的能量均 值等价于其系综平均值, 因而可以通过有限时间的 取样来估计系综平均值, 进而通过上述公式求解得 到相关自由能数据. 由于在实际应用中, 无法得到λ 连续变化的相关数据, 只能为λ设定一些差异较小的 离散值. 通过对研究体系在不同λ对应的状态下的多 次取样, 估计出研究体系的能量以后, 由下面的自由 能计算公式, 就可得到相关的自由能变化数据:

$$\Delta G \approx \sum_{k=1}^{N} \left(\frac{1}{n_{\lambda}} \sum_{i=1}^{n_{\lambda}} \frac{\mathrm{d}V_{\lambda}\left(r_{i,k}\right)}{\mathrm{d}\lambda} \right) \Delta \lambda \tag{11}$$

其中 N 为计算过程中涉及到的研究体系状态的个数, 也即λ的个数; n_λ为对研究体系的某个状态λ的取样次 数; V_λ则是研究体系在状态λ下的势能; 坐标 r_{i,k}则与 研究体系在状态λ下的玻尔兹曼概率相关, 其中下标 *i* 和 *k* 分别指对研究体系每个状态的取样数以及涉及 到的研究体系状态的总个数.

尽管 Thermodynamic Integration 计算自由能的相 关理论出现得很早,其准确、通用的计算方法却是在 最近才提出的^[39-41]. Thermodynamic Integration 方法 与 Free Energy Perturbation 原理以及操作流程上比较 类似, Thermodynamic Integration 方法在原理上比较 严谨,计算所得结果较为精确,计算相关研究体系自 由能变化的时候需要构建热力学循环,通过两次自 由能数据的计算,可以求得自由能变化的数据. 同样, Thermodynamic Integration 也需要对研究体系的各个 状态进行大量的样本采集,需要消耗大量的机时. 因 而对于较为简单的研究体系来说,采用 Thermodynamic Integration 方法计算自由能数据的可以得到很好 的结果^[42-47],而对生物大分子体系来说,对计算资源 的需求较大.

2.4 Jarzynski Equality

Jarzynski Equality 是由 Jarzynski 于 1997 年提出 的^[6]. 与 Umbrella Sampling 类似的是,在使用 Jarzynski Equality 计算自由能的时候,也要用到额外 的谐振势引导研究体系访问高自由能区域,以采集



图 1 Free Energy Perturbation 和 Thermodynamic Integration 计算自由能数据所需的热力学循环示意图^[48]

一般分子动力学模拟中研究体系无法访问的高自由 能区域的样本.而与 Umbrella Sampling 方法不同的 是,外加的谐振势不是静止的,而是随时间做恒速运 动的.该谐振势的具体表述如下:

$$h(x,t) = \frac{1}{2}k \left[r(x) - (\mu_0 - \nu t)\right]^2$$
(12)

其中,μ₀为谐振势的初始位置,v为谐振势的运动速度. 这种谐振势做恒速运动的模拟称为操控式分子动力 学模拟(Steered Molecular Dynamics Simulation). 操 控式分子动力学模拟是受原子力显微镜中使用的单 分子操纵的启发而开发的,是单分子操纵对应的理 论模拟. 在操控式分子动力学模拟中,谐振势引导研 究体系从自由能低的区域访问自由能高的区域时, 会对研究体系做功. 由于自由能差是反应过程的热 力学上的可逆功,因此操控式分子动力学中谐振势 做的不可逆功 W 就是该反应过程自由能差 Δ*G* 的上 限,即:

$$\Delta G \le W \tag{13}$$

Jarzynski 通过大量的理论模拟,证实了在有足够多的不可逆功 W_i 的取样时,反应过程的自由能差 ΔG 与不可逆功 W_i 有如下关系,即Jarzynski Equality:

$$e^{-\beta\Delta G} = e^{-\beta W_i} \tag{14}$$

其中β为温度因子.

这个公式的成立很大程度上依赖于一些不可逆 功接近自由能差时的取样,而这类至关重要的取样 是随着操控式分子动力学的速度增长呈指数形式递 减的.因此从理论上讲,在使用 Jarzynski Equality 计 算自由能数据的时候,不可逆功数据的采样越多,得 到的结果就越准确.在实际应用中,尽管理论上自由 能差与操控式分子动力学的速度无关,Jarzynski Equality 的可靠性仍仅限于不可逆功的标准偏差在 *k*_BT 的数量级上的慢速过程.而且在计算自由能数据 的时候,需要一系列的不可逆功的数据,不可逆功的 数目要足够多才能得到可靠的结果.

在此基础上, Schulten 课题组^[26, 27]使用一个模型 多肽对 Jarzynski Equality 进行了详细的研究,发现, 在外加的谐振势的弹性系数足够大时,不可逆功呈 高斯分布,此时 Jarzynski Equality 的泰勒展开表达式 可以只保留前两项,即:

$$\Delta G = -\frac{1}{\beta} \ln \langle e^{-\beta W_i} \rangle = \langle W_i \rangle - \frac{\beta}{2} \left(\langle W_i^2 \rangle - \langle W_i \rangle^2 \right)$$
(15)

857

其中β为温度因子. 此泰勒二阶展开表达式又称为硬 弹簧模型. 此方法主要采用不可逆功的系统平均, 而 非指数形式, 因此可以在一定程度上缓和 Jarzynski Equality 准确性依赖于少部分较低的不可逆功的 问题.

在谐振势的弹性系数较小且不可调的时候(更多 是在实验中),谐振势的引入会对研究体系产生干扰, 为了解决这个问题, Hummer 和 Szabo^[49]在 Jarzynski Equality 中添加了一个狄拉克 δ 函数,最终自由能的 表达式为:

$$G(r) = -\frac{1}{\beta} \ln \frac{\sum_{t} \frac{\delta(r-r_{t})\exp(-\beta W_{t})}{\exp(-\beta W_{t})}}{\sum_{t} \frac{\exp[-\beta h(r,t)]}{\exp(-\beta W_{t})}}$$
(16)

然后他们进行单分子操纵实验,通过与实验数据的 对比,验证了此方程的准确性.

另外, Crooks^[50]在 Jarzynski Equality 的基础上, 发展了 Jarzynski Equality, 使用双向的拉伸, 对反应 过程的正向和反向过程都进行取样, 提出了 Crooks Equation:

$$e^{-\beta\Delta G} = \frac{\left\langle f(+W) \right\rangle_F}{\left\langle f(-W)e^{-\beta W} \right\rangle_F}$$
(17)

其中下标 F 和 R 分别代表反应方向的正向和反向.

自 Jarzynski Equality 提出十几年以来,大量的研 究工作说明 Jarzynski Equality 在求解简单体系自由 能数据的时候,能够得到与 Umbrella Sampling 准确 度相当的结果^[46, 51-54].并且, Jarzynski Equality 在操 作的便捷性与计算效率上面都较 Umbrella Sampling 有更好的表现. 然而在体系稍微复杂一些的时候,功 的耗散以及外加谐振势对生物大分子构型的干扰都 会对所得自由能数据的准确度产生较大的影响^[55-57].

鉴于 Jarzynski Equality 的计算效率和操作的便 捷性,对 Jarzynski Equality 的改进的研究工作也开始 出现.Ytreberg^[58]通过对研究体系添加额外的限制(图 2(a)),利用 Jarzynski Equality 计算了蛋白-配体的结 合自由能,最后得到了准确的自由能数据.本课题 组^[59]在使用 Jarzynski Equality 计算生物分子在羟基 磷灰石表面吸附自由能的时候,发现固液界面的结 构水层会使操纵式分子动力学模拟产生大量的耗散 功,导致 Jarzynski Equality 严重高估生物分子的吸附 自由能.通过详细分析,我们对 Jarzynski Equality 进 行了改进,提出了一个混合算法,消除了固液界面结构水层对自由能数据的干扰(图 2(b)).

2.5 Metadynamics

Metadynamics 是 Parrinello 等^[7]于 2002 年提出的 一个计算多维自由能势能面的方法.这个方法与 Umbrella Sampling 以及 Jarzynski Equality 一样, 都是 利用外加额外的势能,能够更快地对研究体系的高 自由能区域取样. 与前两种方法不同的是, Metadynamics 在计算自由能的时候, 对研究体系所 加的额外的势能为高斯型的排斥势能. Metadynamics 方法类似于"添坑"的过程,通过对反应路径上的低 自由能区域持续地添加排斥势能,阻止研究体系再 次访问反应坐标上已经采集过数据的区域,不断地 促使研究体系向高自由能区域运动,最后再根据所 添加的一系列排斥势能来计算相关的自由能数据. 这个方法最初应用在粗粒度模拟中研究相关体系在 变量空间中的动力学行为. 之后, Metadynamics 的排 斥势函数持续地应用在分子动力学模拟当中[60, 61]. 这里要讨论的是 Metadynamics 最新发展的计算自由



图 2 (a) Ytreberg 改进 Jarzynski Equality 所用到的限制项 U_a(*φ*, *θ*)^[58]; (b) 本课题组提出的混合算法得到的自由能数 据^[59]与 Umbrella Sampling 结果比较吻合

能的方法.

假设 Metadynamics 要对 N 维的反应路径求解自由能势能面,使反应坐标 S 为 N 维的向量:

$$S(R) = (S_1(R), ..., S_N(R))$$
 (18)

那么体系的哈密顿量就为

$$\hat{H}(x, p, t) = H(x, p) + V_G(S, t)$$
 (19)

其中 V_G(S,t)为额外添加的排斥势能,具体表达式为:

$$V_G(S,t) = \int_0^t dt' \omega \exp\left(-\sum_{i=1}^N \frac{\left(S_i(R) - S_i(R(t'))\right)^2}{2\sigma_i^2}\right)$$
(20)

其中ω分别为σ_i为第*i*个变量排斥势能的高斯分布的 高度和宽度.最终根据额外添加的一系列排斥势能, 就可以计算相关自由能数据

$$F(S) = -\frac{1}{\beta} \ln \left(\int dR \delta \left(S - S(R) \right) e^{-\beta U(R)} \right)$$
(21)

其中 U(R)为势能函数.

Metadynamics 在求解自由能的时候拥有很多优势,首先,它通过一系列的外加排斥势能,一点点地把研究体系从低自由能区域推向高自由能区域,加速了对研究体系高自由能事件的取样;其次,此方法可以设定高纬度的反应路径,进而求解高纬度的自由能数据,例如二维的自由能势能面(图3);再次,由于研究体系是从低自由能区域一点点地推高到高自由能区域,此方法在寻找最优反应路径方面拥有特有的优势;最后,与上一点相关的是,此方法不需要



图 3 由 Metadynamics 方法得到的配体-蛋白相互作用体系的二维自由能数据^[62]

事先对研究体系的反应路径有预判,而 Umbrella Sampling 和 Jarzynski Equality 则在计算自由能数据之 前,就要设置好反应路径.鉴于上述优点,Metadynamics 迅速成为一种流行的自由能计算方法,有很多 工作使用此方法计算相关的一维以及二维的自由能 数据^[62-64].一开始,Metadynamics 计算自由能数据是 由独立于常用的分子动力学模拟软件之外的插件 PLUMED^[28]执行的.现在此方法已经集成到了多个 计算模拟软件中,例如:NAMD^[65],ORAC^[66],CP2K, 以及 CPMD(IBM, Armonk, NY)等.

3 自由能计算方法在生物大分子研究中的 适用性

在简单体系以及小分子体系中,由于不受计算 效率的限制,自由能数据的计算方法的选择就比较 多,可以根据研究目的选择准确度高的方法.然而, 对于生物大分子体系来说,由于一般都需要水溶液 环境,研究体系粒子数往往都在几十万以上,这就对 计算资源有着很高的要求,而这在一般情况下很难 得到满足.因此生物大分子体系的自由能计算往往 面临着计算效率和结果准确度的权衡问题.

总体来说, 经典的自由能计算方法 Umbrella Sampling, Free Energy Perturbation 以及 Thermo-dynamic Integration 对计算资源有着很高的需求, 很少直 接用于求解生物大分子体系的自由能数据. 由于 Umbrella Sampling 在求解 PMF 时与 Jarzynski Equality 比较相近, 因此多用于对 Jarzynski Equality 及其 改进方法所得结果可靠度以及计算效率的检验[26~30]. Free Energy Perturbation 和 Thermodynamic Integration 在原理及操作上都比较相近,由于对计算资源有 很高的需求,多用于求解小分子体系溶剂化自由能、 构型自由能以及化学反应自由能等^[36, 67, 68]. Jarzynski Equality 操作简便, 可移植性强, 同时计算效率也相 对较高,在生物大分子体系中的应用比较广泛[26, 27, 69]. Metadynamics 则在求解高纬自由能数据, 以及探索 全局最优反应路径方面有着很大的优势, 现已有较 为全面的综述文章可以参考[64].

鉴于上文对各个自由能计算方法原理的阐述, 以及各自特点的分析,现总结如表1所示,以供读者 在做自由能研究的时候参考.

| 方法 | 原理 | 优点 | 缺点 |
|------------------------------|---|---|--|
| Umbrella Sampling | 使用谐振势强制体系在高自由能区域 取样,去除谐振势的影响以后,根据 概率密度求解自由能数据 | 数据准确度高,具有反应坐标,可以得到自由能数据随反应坐标 的变化趋势 | 非常消耗计时,计算效率不高,操作较 为繁琐,在计算资源有限的情况下不适 于用来计算生物大分子体系的自由能 |
| Free Energy Perturbation | 在初态和终态之间人为插入很多中间 态,通过计算中间态之间的自由能变 化,进而得到反应过程的自由能变化 | 数据准确度高,不需要为研究体 系规定反应路径 | 非常消耗计时,不能观察自由能随变量 的变化趋势,在计算资源有限时不适于 计算生物大分子体系的自由能 |
| Thermodynamic Integration | 在初态和终态之间设定很多中间态, 利用研究体系在各个状态下能量的系 综平均值计算自由能数据 | 数据准确度高,不需要预先为研 究体系设定反应路径 | 非常消耗计时,不能观察自由能随变量 的变化趋势,在计算资源有限时不适于 计算生物大分子体系的自由能 |
| Jarzynski Equality | 反应过程不可逆功的系统平均是反应 的自由能变化的上限 | 计算效率高,操作简便,可以用 来计算较大的生物分子体系的自 由能数据 | 受研究体系的性质影响较大,对不同的 研究体系需要采取额外的措施才能得到 可靠的结果 |
| Metadynamics | 在体系低自由能区域添加排斥的势能 函数,使体系不能再访问已经采集过 数据的区域,进而加快体系访问高自 由能区域的速度 | 不需要预先设定反应路径,可以 用来探索最优反应路径,可以方 便地计算高纬度的自由能数据, 适于计算自由能的二维势能面 | 添加的排斥势能并不收敛于真实的自由 能数据,而是在其附近震荡;与一些主 流计算软件的对接需要额外下载插件编 译 |

表1 各个自由能计算方法的原理以及自身优缺点

4 总结与展望

本文对几种影响较大的自由能计算方法进行了 总结,详述了每个方法的基本原理,各自的特点以及 研究的现状,并对各个方法在生物大分子体系中的 适用性进行了概括.然而,尽管现有的自由能方法可 以解决一部分研究工作的需要,生物大分子体系的 自由能数据的计算依然是一个非常消耗计算资源的 工作,在计算高纬度的自由能数据时,这个问题尤为 突出.因此,开发具有高计算效率的自由能计算方法 以处理生物大分子体系遇到的问题,在自由能计算 领域依然是一个重要的研究契机和挑战.

致谢 本工作得到国家自然科学基金(21273200)资助,特此致谢.

参考文献_

- 1 Torrie GM, Valleau JP. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. J Comput Phys, 1977, 23: 187–199
- 2 Essex JW, Severance DL, Tirado-Rives J, Jorgensen WL. Monte Carlo simulations for proteins: binding affinities for trypsin-benzamidine complexes via free-energy perturbations. J Phys Chem B, 1997, 101: 9663–9669
- 3 Zwanzig RW. High-temperature equation of state by a perturbation method. i. nonpolar gases. J Chem Phys, 1954, 22: 1420
- 4 Jorgensen WL, Blake JF, Buckner JK. Free energy of TIP4P water and the free energies of hydration of CH4 and Cl- from statistical perturbation theory. *Chem Phys*, 1989, 129: 193–200
- 5 Kirkwood JG. Statistical mechanics of fluid mixtures. J Chem Phys, 1935, 3: 300
- 6 Jarzynski C. Nonequilibrium Equality for Free Energy Differences. Phys Rev Lett, 1997, 78: 2690
- 7 Laio A, Parrinello M. Escaping free-energy minima. Proc Natl Acad Sci USA, 2002, 99: 12562–12566
- 8 Zoete V, Michielin O, Karplus M. Protein-ligand binding free energy estimation using molecular mechanics and continuum electrostatics. Application to HIV-1 protease inhibitors. J Comp-Aided Mol Des, 2003, 17: 861–880
- 9 Jiao D, Golubkov PA, Darden TA, Ren P. Calculation of protein-ligand binding free energy by using a polarizable potential. *Proc Natl Acad Sci USA*, 2008, 105: 6290–6295
- 10 Suever JD, Chen Y, McDonald JM, Song Y. Conformation and free energy analyses of the complex of calcium-bound calmodulin and the

Fas death domain. Biophys J, 2008, 95: 5913-5921

- 11 Andrec M, Felts AK, Gallicchio E, Levy RM. Protein folding pathways from replica exchange simulations and a kinetic network model. Proc Natl Acad Sci U S A, 2005, 102: 6801–6806
- 12 Finkelstein A, Galzitskaya O. Physics of protein folding. Phys Life Rev, 2004, 1: 23-56
- 13 Ghosh K, Ozkan SB, Dill KA. The ultimate speed limit to protein folding is conformational searching. J Am Chem Soc, 2007, 129: 11920–11927
- 14 Dill KA, Ozkan SB, Weikl TR, Chodera JD, Voelz VA. The protein folding problem: when will it be solved? *Curr Opin Struct Biol*, 2007, 17: 342–346
- 15 Piccinini E, Affinito F, Brunetti R, Jacoboni C, Ceccarelli M. Exploring free-energy profiles through ion channels: comparison on a test case. J Comput Electr, 2007, 6: 373–376
- 16 Wei Y, Latour RA. Determination of the adsorption free energy for peptide–surface interactions by SPR spectroscopy. *Langmuir*, 2008, 24: 6721–6729
- 17 de Graaf J, Dijkstra M, van Roij R. Adsorption trajectories and free-energy separatrices for colloidal particles in contact with a liquid-liquid interface. *J Chem Phys*, 2010, 132: 164902
- 18 Thyparambil AA, Wei Y, Latour RA. Determination of peptide–surface adsorption free energy for material surfaces not conducive to SPR or QCM using AFM. *Langmuir*, 2012, 28: 5687–5694
- 19 Wright LB, Walsh TR. Facet selectivity of binding on quartz surfaces: free energy calculations of amino-acid analogue adsorption. J Phys Chem, 2012, 116: 2933–2945
- 20 Mijajlovic M, Penna MJ, Biggs MJ. Free energy of adsorption for a peptide at a liquid/solid interface via nonequilibrium molecular dynamics. *Langmuir*, 2013, 29: 2919–2926
- 21 Kollman P. Free energy calculations: applications to chemical and biochemical phenomena. Chem Rev, 1993, 93: 2395–2417
- 22 Jorgensen WL. Free energy calculations: a breakthrough for modeling organic chemistry in solution. Acc Chem Res, 1989, 22: 184–189
- 23 Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. The weighted histogram analysis method for free-energy calculations on biomolecules. I The method. *J Comput Chem*, 1992, 13: 1011–1021
- 24 Hub JS, Groot BLD, Spoel DVD. g wham—A free weighted histogram analysis implementation including robust error and autocorrelation estimates. J Chem Theory Comput, 2010, 6: 3713–3720
- 25 Zhang Z-S, Kang Y, Liang L-J, Liu Y-C, Wu T, Wang Q. Peptide encapsulation regulated by the geometry of carbon nanotubes. *Biomaterials*, 2014, 35: 1771–1778
- 26 Park S, Khalili-Araghi F, Tajkhorshid E, Schulten K. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. J Chem Phys, 2003, 119: 3559
- 27 Park S, Schulten K. Calculating potentials of mean force from steered molecular dynamics simulations. J Chem Phys, 2004, 120: 5946
- 28 Bonomi M, Branduardi D, Bussi G, Camilloni C, Provasi D, Raiteri P, Donadio D, Marinelli F, Pietrucci F, Broglia RA, Parrinello M. PLUMED: a portable plugin for free-energy calculations with molecular dynamics. *Comput Phys Comm*, 2009, 180: 1961–1972
- 29 Piccinini E, Ceccarelli M, Affinito F, Brunetti R, Jacoboni C. Biased molecular simulations for free-energy mapping: a comparison on the KcsA channel as a test case. J Chem Theory Comput, 2008, 4: 173–183
- 30 Kästner J. Umbrella sampling. Wiley Interdiscip. Rev-Comput Mol Sci, 2011, 1: 932–942
- 31 Levy RM, Belhadj M, Kitchen DB. Gaussian fluctuation formula for electrostatic free-energy changes in solution. J Chem Phys, 1991, 95: 3627
- 32 Del Buono GS, Figueirido FE, Levy RM. Intrinsic pKas of ionizable residues in proteins: an explicit solvent calculation for lysozyme. *Proteins: truct Funct Bioinform*, 1994, 20: 85–97
- 33 Simonson T, Carlsson J, Case DA. Proton binding to proteins: pK_a calculations with explicit and implicit solvent models. J Am Chem Soc, 2004, 126: 4167–4180
- 34 Simonson T, Wong CF, Brünger AT. Classical and quantum simulations of tryptophan in solution. J Phys Chem A, 1997, 101: 1935–1945
- 35 Radmer RJ, Kollman PA. Free energy calculation methods: a theoretical and empirical comparison of numerical errors and a new method qualitative estimates of free energy changes. *J Comput Chem*, 1997, 18: 902–919
- 36 Jorgensen WL, Thomas LL. Perspective on free-energy perturbation calculations for chemical equilibria. J Chem Theory Comput, 2008, 4: 869–876
- 37 Åqvist J. Ion-water interaction potentials derived from free energy perturbation simulations. J Phys Chem, 1990, 94: 8021–8024
- 38 Shivakumar D, Williams J, Wu Y, Damm W, Shelley J, Sherman W. Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. J Chem Theory Comput, 2010, 6: 1509–1519

- 39 Den Otter W, Briels W. The calculation of free-energy differences by constrained molecular-dynamics simulations. J Chem Phys, 1998, 109: 4139
- 40 Darve E, Wilson MA, Pohorille A. Calculating free energies using a scaled-force molecular dynamics algorithm. *Mol Simul*, 2002, 28: 113–144
- 41 Darve E, Pohorille A. Calculating free energies using average force. J Chem Phys, 2001, 115: 9169
- 42 Wescott J, Fisher L, Hanna S. Use of thermodynamic integration to calculate the hydration free energies of n-alkanes. *J Chem Phys*, 2002, 116: 2361
- 43 Straatsma T, Berendsen H. Free energy of ionic hydration: analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations. *J Chem Phys*, 1988, 89: 5876
- 44 Mezei M. The finite difference thermodynamic integration, tested on calculating the hydration free energy difference between acetone and dimethylamine in water. *J Chem Phys*, 1987, 86: 7084
- 45 Riniker S, Christ CD, Hansen N, Mark AE, Nair PC, van Gunsteren WF. Comparison of enveloping distribution sampling and thermodynamic integration to calculate binding free energies of phenylethanolamine N-methyltransferase inhibitors. J Chem Phys, 2011, 135: 024105
- 46 Rodriguez-Gomez D, Darve E, Pohorille A. Assessing the efficiency of free energy calculation methods. *J Chem Phys*, 2004, 120: 3563–3578
- 47 Lartillot N, Philippe H. Computing Bayes factors using thermodynamic integration. Syst Biol, 2006, 55: 195-207
- 48 Mobley DL, Chodera JD, Dill KA. Confine-and-release method: obtaining correct binding free energies in the presence of protein conformational change. *J Chem Theory Comput*, 2007, 3: 1231–1235
- 49 Hummer G, Szabo A. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. Proc Natl Acad Sci U S A, 2001, 98: 3658–3661
- 50 Crooks GE. Path-ensemble averages in systems driven far from equilibrium. Phys Rev E, 2000, 61: 2361
- 51 Oostenbrink C, van Gunsteren WF. Calculating zeros: non-equilibrium free energy calculations. Chem Phys, 2006, 323: 102–108
- 52 Ytreberg FM, Zuckerman DM. Efficient use of nonequilibrium measurement to estimate free energy differences for molecular systems. J Comput Chem, 2004, 25: 1749–1759
- 53 Lua RC, Grosberg AY. Practical applicability of the jarzynski relation in statistical mechanics: a pedagogical example[†]. *J Phys Chem B*, 2005, 109: 6805–6811
- 54 Oberhofer H, Dellago C, Geissler PL. Biased sampling of nonequilibrium trajectories: can fast switching simulations outperform conventional free energy calculation methods? *J Phys Chem B*, 2005, 109: 6902–6915
- 55 Kato M, Pisliakov AV, Warshel A. The barrier for proton transport in aquaporins as a challenge for electrostatic models: the role of protein relaxation in mutational calculations. *Proteins*, 2006, 64: 829–844
- 56 Baştuğ T, Chen P-C, Patra SM, Kuyucak S. Potential of mean force calculations of ligand binding to ion channels from Jarzynski's equality and umbrella sampling. J Chem Phys, 2008, 128: 155104
- 57 Warshel A, Sharma PK, Kato M, Parson WW. Modeling electrostatic effects in proteins. Biochim Biophys Acta, 2006, 1764: 1647–1676
- 58 Ytreberg FM. Absolute FKBP binding affinities obtained via nonequilibrium unbinding simulations. J Chem Phys, 2009, 130: 164906
- 59 Zhang Z, Wu T, Wang Q, Pan H, Tang R. Impact of interfacial high-density water layer on accurate estimation of adsorption free energy by jarzynski's equality. J Chem Phys, 2014, 140: 034706
- 60 Laio A, Rodriguez-Fortea A, Gervasio FL, Ceccarelli M, Parrinello M. Assessing the accuracy of metadynamics. J Phys Chem B, 2005, 109: 6714–6721
- 61 Iannuzzi M, Laio A, Parrinello M. Efficient exploration of reactive potential energy surfaces using Car-Parrinello molecular dynamics. *Phys Rev Lett*, 2003, 90: 238302
- 62 Provasi D, Bortolato A, Filizola M. Exploring molecular mechanisms of ligand recognition by opioid receptors with metadynamics. Biochemistry, 2009, 48: 10020–10029
- 63 Marini F, Camilloni C, Provasi D, Broglia RA, Tiana G. Metadynamic sampling of the free-energy landscapes of proteins coupled with a Monte Carlo algorithm. *Gene*, 2008, 422: 37–40
- 64 Barducci A, Bonomi M, Parrinello M. Metadynamics. Wiley Interdiscip Rev-Comput Mol Sci, 2011, 1: 826-843
- 65 Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. Scalable molecular dynamics with NAMD. *J Comput Chem*, 2005, 26: 1781–1802
- 66 Marsili S, Signorini GF, Chelli R, Marchi M, Procacci P. ORAC: a molecular dynamics simulation program to explore free energy surfaces in biomolecular systems at the atomistic level. *J Comput Chem*, 2010, 31: 1106–1116

- 67 Radmer RJ, Kollman PA. Free energy calculation methods: a theoretical and empirical comparison of numerical errors and a new method qualitative estimates of free energy changes. *J Comput Chem*, 1997, 18: 902–919
- 68 Reddy MR, Singh U, Erion MD. Ab initio quantum mechanics based free energy perturbation method for calculating relative solvation free energies. *J Comput Chem*, 2007, 28: 491–494
- 69 Jensen MØ, Park S, Tajkhorshid E, Schulten K. Energetics of glycerol conduction through aquaglyceroporin GlpF. Proc Natl Acad Sci U S A, 2002, 99: 6731–6736

Free energy calculation and its application in bio-complex system

ZHANG ZhiSen¹, WANG Qi^{1*}, CHEN ErYu^{1, 2}

1 Department of Chemistry and Soft Matter Research Center, Zhejiang University, Hangzhou 310027, China 2 Hangzhou NCO School, Hangzhou 310023, China *Corresponding author (email: qiwang@zju.edu.cn)

Abstract: In this work, we reviewed the free energy calculation methods used in computational science. The theoretical base and the feature of the free energy calculation methods were summarized. It's expected that this work could help the new comers select proper method for their personal systems when they calculate free energy data.

Keywords: free energy calculation, bio-complex, Umbrella Sampling, Free Energy Perturbation, Jarzynski Equality, Matadynamics