

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 7 月 23 日 (23.07.2020)



(10) 国际公布号
WO 2020/147668 A1

(51) 国际专利分类号:
G16C 20/50 (2019.01)

(21) 国际申请号: PCT/CN2020/071663

(22) 国际申请日: 2020 年 1 月 13 日 (13.01.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910045003.X 2019 年 1 月 17 日 (17.01.2019) CN
201911412608.4 2019 年 12 月 31 日 (31.12.2019) CN

(71) 申请人: 中山大学 (SUN YAT-SEN UNIVERSITY)
[CN/CN]; 中国广东省广州市海珠区新港西路 135 号, Guangdong 510275 (CN)。

(72) 发明人: 罗海彬 (LUO, Haibin); 中国广东省广州市海珠区新港西路 135 号, Guangdong 510275

(CN)。李哲 (LI, Zhe); 中国广东省广州市海珠区新港西路 135 号, Guangdong 510275 (CN)。

(74) 代理人: 广州粤高专利商标代理有限公司 (YOGO PATENT & TRADEMARK AGENCY LIMITED COMPANY); 中国广东省广州市天河区体育西路 191 号中石化大厦 B 塔 4416 室, Guangdong 510620 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,

(54) Title: ABSOLUTE BINDING FREE ENERGY PERTURBATION METHOD FOR PREDICTING DRUG TARGET BINDING STRENGTH

(54) 发明名称: 预测药-靶结合强度的绝对结合自由能微扰方法

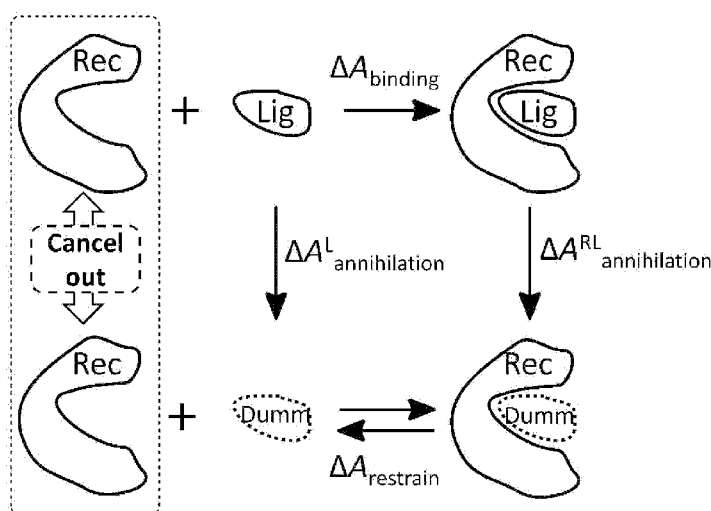


图 1

(57) Abstract: An absolute binding free energy perturbation method for predicting the drug target binding strength, comprising: on the basis of the fitting of the probability distribution of a constraint probability distribution function to a process of adding constraints, fitting the probability distribution to the remaining windows by using a linear combination of a plurality of Gaussian functions for the probability distribution obtained by counting windows of the remaining thermodynamic cycle steps, regenerating a potential difference according to the fit probability distribution, and finally counting to obtain the absolute binding free energy of a drug target. By fitting the probability distribution corresponding to the dynamics simulation trajectory in the intermediate state of the free energy perturbation, the probability distribution of sampling is more reasonable, and the convergence and accuracy of the energy calculation result are greatly improved.

ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US,
UZ, VC, VN, WS, ZA, ZM, ZW。

- (84) 指定国 (除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告 (条约第21条(3))。

(57) 摘要: 一种预测药-靶结合强度的绝对结合自由能微扰方法, 基于约束概率分布函数针对添加约束过程的概率分布的拟合, 对其余热力学循环步骤的窗口统计得到的概率分布, 利用多个高斯函数的线性组合对其余窗口的概率分布进行拟合, 根据拟合后的概率分布重新生成势能差, 最终统计得出药物-靶标绝对结合自由能。通过对自由能微扰中间状态的动力学模拟轨迹所对应的概率分布进行拟合, 使得采样的概率分布更加合理, 大幅提升了能量计算结果的收敛性及准确性。

5 预测药-靶结合强度的绝对结合自由能微扰方法

技术领域

本发明涉及药物研发技术领域，更具体的，涉及一种预测药-靶结合强度的绝对结合自由能微扰方法。

背景技术

10 药物产生药效的原因几乎都与药物和靶标大分子的相互结合密切相关。因此，药物-靶标结合强度的自由能预测技术在药物设计中扮演着极其重要的角色（Q. Rev. Biophys. 2012, 45, 1-25.; Curr. Opin. Struct. Biol. 2011, 21, 150-160.等）。若有一种方法能够准确地预测结合自由能，那么新药研发的周期与成本将大大降低。

15 尽管目前存在一些较为快速的结合自由能预测方法，例如对接打分函数，MM/PBSA(分子力学/泊松玻尔兹曼表面，药物设计中常用的一种能量计算方法)、MM/GBSA(分子力学/广义伯恩表面)等方法。但由于这些方法多引入了很多的经验参数，或者理论基础存在一定的缺陷，导致它们的准确性不高。自由能微扰(FEP)作为一种基于统计力学的自由能计算方法，理论上具有更高的准确度。如著名药
20 物设计软件公司薛定谔 Schrodinger 在 2015 年发明了一种基于自由能微扰的相对结合自由能预测方法（J. Am. Chem. Soc. 2015, 137, 2695-2703.），该方法能预测对现有药物分子只有微小结构修饰的同骨架化合物与靶标的相对结合自由能，且往往这类预测方法需要参考已知活性的小分子结构。新药研发中遇到的大多数问题情况，如虚拟筛选，骨架跃迁和全新药物设计时，需要进行绝对结合自由能的
25 预测。

发明内容

本发明为克服上述现有技术存在无法进行绝对结合自由能预测的技术问题，提供一种预测药-靶结合强度的绝对结合自由能微扰方法。

为解决上述技术问题，本发明的技术方案如下：

30 预测药-靶结合强度的绝对结合自由能微扰方法，包括以下步骤：

S1: 对受体蛋白-配体体系, 即 Rec-Lig 体系进行常规的分子动力学模拟, 并基于模拟获得的轨迹得到添加约束所需的距离、夹角、二面角数据;

S2: 构建热力学循环, 构建 $\Delta A_{annihilation}^L$ 、 $\Delta A_{annihilation}^{RL}$ 和 $\Delta A_{restrain}^L$, $\Delta A_{restrain}^{RL}$ 根据添加约束所需的距离、夹角、二面角数据得到, 并对 $\Delta A_{restrain}^L$ 的自由能进行

5 解析计算;

其中: $\Delta A_{annihilation}^L$ 表示将溶液中药物小分子全部微扰至湮灭状态的能量; $\Delta A_{annihilation}^{RL}$ 表示将药物-靶标复合物中药物小分子全部微扰至湮灭状态的能量; $\Delta A_{restrain}^{RL}$ 表示对药物-靶标复合物中药物小分子添加简谐势约束的能量变化;

S3: 利用分子动力学模拟程序, 对 $\Delta A_{annihilation}^L$ 、 $\Delta A_{annihilation}^{RL}$ 、 $\Delta A_{restrain}^L$ 步骤
10 中的每一个窗口进行分子动力学模拟采样, 得到 $\Delta A_{annihilation}^L$ 、 $\Delta A_{annihilation}^{RL}$ 、 $\Delta A_{restrain}^L$ 步骤中每一个窗口的分子动力学轨迹;

S4: 根据 $\Delta A_{annihilation}^L$ 、 $\Delta A_{annihilation}^{RL}$ 、 $\Delta A_{restrain}^L$ 步骤中的每一个窗口的分子动力学轨迹, 通过相邻窗口的力场文件进行计算, 得到每个相邻窗口之间的势能差 ΔU ;

15 S5: 对于 $\Delta A_{annihilation}^L$ 、 $\Delta A_{annihilation}^{RL}$ 这两步, 根据对应窗口的 ΔU 的概率分布, 使用高斯函数进行拟合, 并计算得到 $\Delta A_{annihilation}^L$ 、 $\Delta A_{annihilation}^{RL}$ 步骤的自由能;

S6: 对于 $\Delta A_{restrain}^{RL}$ 这一步每个相邻窗口之间的势能差 ΔU 进行概率分布统计, 利用约束概率分布函数进行拟合, 得到平滑的概率分布, 并根据拟合结果检查收敛性;

20 S7: 根据拟合后的概率分布, 重新生成 $\Delta A_{restrain}^{RL}$ 这一步中每一窗口的势能差, 通过计算得到 $\Delta A_{restrain}^{RL}$ 的能量, 并根据之前步骤得到的 $\Delta A_{annihilation}^L$ 、 $\Delta A_{annihilation}^{RL}$ 、以及 $\Delta A_{restrain}^L$ 计算得到药-靶结合自由能 $\Delta A_{binding}$ 。

其中, 所述步骤 S1 具体包括以下步骤:

S11: 准备受体蛋白 Rec-配体 Lig 复合物坐标文件;

25 S12: 准备小分子配体的电荷, 并对配体指定力场;

S13: 将受体-配体复合物中的受体指定力场, 并将该复合物在水箱中进行浸泡, 准备好复合物进行分子模拟的初始体系, 记为 Rec-Lig 体系, 并到处力场文件作为 Rec-Lig 体系初始状态参数;

S14: 将单独小分子配体浸泡在水箱中, 准备好单独的小分子的分子模拟初
30 始体系, 记为 Lig 体系, 并到处力场文件作为 Lig 体系初始状态参数;

S15: 根据 Rec-Lig 体系初始状态的力场参数, 进行短时间的分子动力学模拟, 得到运动轨迹;

S16: 基于模拟得到的轨迹, 统计与小分子配体相对位置变化较小的受体蛋白的氨基酸残基, 取该残基中与小分子配体相对距离最稳定的三个原子作为添加
5 约束所需的属于受体蛋白的三个原子, 选取小分子配体中最稳定的三个原子作为添加约束所需的属于配体小分子三个原子, 得到添加约束所需的距离、夹角、二面角数据。

其中, 所述步骤 S2 具体包括以下步骤:

S21: 根据准备好的 Rec-Lig 体系的力场参数, 根据步骤 S16 得到的 6 个原
10 子, 添加一个长度、两个夹角、三个二面角约束, 得到计算 $\Delta A_{\text{restrain}}^{RL}$ 所需窗口的分子动力学模拟参数;

S22: 进一步将 Rec-Lig 体系中的小分子的电荷以及范德华参数依次分别减小至 0, 得到计算 $\Delta A_{\text{annihilation}}^{RL}$ 所需窗口的分子动力学模拟参数;

S23: 对于 Rec-Lig 体系中小分子的电荷以及范德华参数依次分别减小至 0,
15 得到计算 $\Delta A_{\text{annihilation}}^L$ 所需窗口的分子动力学模拟参数;

S24: 对 $\Delta A_{\text{restrain}}^L$ 的自由能进行解析计算。

其中, 在所述步骤 S24 中, 所述 $\Delta A_{\text{restrain}}^L$ 自由能的计算公式具体为:

$$\Delta A_{\text{restrain}}^L = -k_B T \ln \left(\frac{8\pi^2 V}{r_{aA}^2 \cdot \sin \theta_a \cdot \sin \theta_A} \frac{(k_{r_{aA}} k_{\theta_a} k_{\theta_A} k_{\phi_{ba}} k_{\phi_{aA}} k_{\phi_{AB}})^{1/2}}{(2\pi k_B T)^3} \right)$$

其中, k_B 为玻尔兹曼常数, V 为系统的体积, r_{aA} 为一个两原子间的距离用于
20 距离简谐势约束, θ_a 和 θ_A 为两个三原子间的夹角用于夹角简谐势约束, 诸多 k 值为用于约束一个距离 r_{aA} 、两个夹角 θ_a, θ_A 、三个二面角 $\phi_{aA}, \phi_{AB}, \phi_{ba}$ 的简谐势力常数。

其中, 在所述步骤 S3 或 S4 中, 采用包括 AMBER、Gromacs、Namd 在内的任意分子动力学模拟软件进行计算。

其中, 所述步骤 S5 具体包括以下步骤:

S51: 针对得到每个相邻状态之间的势能差进行概率分布统计, 得到概率分布 $P(\Delta U)_{\text{data}}$;

S52: 将概率分布 $P(\Delta U)_{\text{data}}$ 利用多个高斯函数的线性组合进行拟合, 得到平滑的概率分布, 具体计算公式为:

$$P(\Delta U)_{fit} = \sum_{i=1}^5 c_i \exp\left(-\frac{(\Delta U - \mu_i)}{2\sigma_i^2}\right)$$

其中, $P(\Delta U)_{fit}$ 为经高斯函数拟合后的概率分布, c_i 为第 i 个高斯函数的参数, μ_i 为第 i 个高斯函数的平均值, σ_i 为第 i 个高斯函数的标准差;

S53: 根据得到的 $P(\Delta U)_{fit}$, 采用 Benneet Acceptance Ratio 计算方法, 即 BAR 计算方法进行 $\Delta A_{annihilation}^L$ 、 $\Delta A_{annihilation}^{RL}$ 自由能的计算, 具体计算公式为:

$$\exp(-\beta \Delta A_{i,i+1}) = \frac{\left\langle \{1 + \exp[\beta(\Delta U_{i,i+1} - C)]\}^{-1} \right\rangle_i}{\left\langle \{1 + \exp[-\beta(\Delta U_{i,i+1} - C)]\}^{-1} \right\rangle_{i+1}} \exp(-\beta C)$$

$$C = \Delta A_{i,i+1} + \frac{1}{\beta} \ln \frac{n_i}{n_{i+1}}$$

5 其中, $\Delta A_{i,i+1}$ 为两相邻窗口 i 与 $i+1$ 的结合自由能差, $\Delta U_{i,i+1}$ 为基于拟合后的概率分布新生成的势能差, $\beta=1/k_B T$, k_B 为玻尔兹曼常数, n_i 与 n_{i+1} 分别为两相邻状态 i 与 $i+1$ 采样所包含的样本量, C 为方程组之间的隐含变量;

S54: 将 $\Delta A_{annihilation}^{RL}$ 和 $\Delta A_{annihilation}^L$ 对应的计算窗口得到的 $\Delta A_{i,i+1}$ 进行加和, 得到 $\Delta A_{annihilation}^{RL}$ 和 $\Delta A_{annihilation}^L$ 的数值。

10 其中, 所述步骤 S6 具体包括以下步骤:

S61: 对计算 $\Delta A_{restrain}^{RL}$ 的窗口中每个窗口之间的势能差进行概率分布统计, 得到概率分布 $P(\Delta U)_{data}$;

S62: 基于约束概率分布函数对概率分布 $P(\Delta U)_{data}$ 进行拟合, 具体表达式为:

$$P(\Delta U)_{fit} = \sum_{i=1}^n \left(\frac{a_i}{2} \cdot b_i^{\frac{3}{2}} \cdot \Delta U^2 \cdot e^{-b_i \Delta U - \{c_i \Delta U\}^{n_1}} + \frac{1}{1 + \left(\frac{d_i}{\Delta U}\right)^{n_2}} \left(h_i \cdot \frac{1}{\sqrt{\pi} \mu_i} \cdot e^{-\frac{(\Delta U - \mu_i)^2}{2\sigma_i^2}} \right) \right)$$

15 其中, $P(\Delta U)_{fit}$ 为经过约束概率分布函数拟合后的平滑的概率分布, 其中 $a_i, b_i, c_i, d_i, h_i, \mu_i, \sigma_i$ 为第 i 个“约束概率分布函数”的参数, n 为选取拟合函数的数目, n_1 和 n_2 为两个正整数的修正因子。

其中, 所述步骤 S7 具体包括以下步骤:

S71: 根据概率分布 $P(\Delta U)_{fit}$, 重新生成每一状态的势能差 $\Delta U'$;

20 S72: 基于新的势能差 $\Delta U'$, 利用步骤 S5 的计算方法重新计算得到 $\Delta A_{restrain}^{RL}$ 的计算数值;

S73: 根据步骤 S24、步骤 S5 和步骤 S72 得到的 $\Delta A_{restrain}^L$ 、 $\Delta A_{annihilation}^{RL}$ 、

$\Delta A_{annihilation}^L$ 和 $\Delta A_{restrain}^{RL}$, 计算出药-靶结合自由能 $\Delta A_{binding}$ 。

其中, 在所述步骤 S73 中, 所述药-靶结合自由能 $\Delta A_{binding}$ 的计算过程具体为:

$$\Delta A_{binding} = \Delta A_{annihilation}^L + \Delta A_{restrain}^L - \Delta A_{annihilation}^{RL} - \Delta A_{restrain}^{RL}。$$

与现有技术相比, 本发明技术方案的有益效果是:

- 5 本发明提供一种预测药-靶结合强度的绝对结合自由能微扰方法, 基于约束概率分布函数针对添加约束过程的概率分布的拟合, 对其余热力学循环步骤的窗口统计得到的概率分布, 利用多个高斯函数的线性组合对其余窗口的概率分布进行拟合, 根据拟合后的概率分布重新生成势能差, 最终统计得出药物-靶标绝对结合自由能; 通过对自由能微扰中间状态的动力学模拟轨迹所对应的概率分布
- 10 进行拟合, 使得采样的概率分布更加合理, 大幅提升了能量计算结果的收敛性及准确性。

附图说明

图 1 为在同一靶标不同小分子之间质量差距比较小时, 不添加约束所涉及的热力学循环示意图;

- 15 图 2 为采样所得 ΔU 被转化的概率分布 $P(\Delta U)$ 示意图;

图 3 为采样所得概率分布 $P(\Delta U)$ 被一系列高斯函数拟合的结果示意图;

图 4 为经高斯函数拟合采样概率分布前后方法准确度的差异示意图;

图 5 为计算能量过程中, $\exp(-\beta\Delta U)P_0(\Delta U)$ 与 $P_0(\Delta U)$ 的关系示意图;

- 图 6 为本发明中高斯函数增强的绝对结合自由能计算方法与薛定谔公式的相对结合自由能计算方法的准确度对比示意图;
- 20

图 7 为本方法流程示意图;

图 8 为计算绝对结合自由能所涉及的热力学循环图;

图 9 为添加约束过程的状态采样所得概率被约束概率分布函数拟合的结果示意图。

具体实施方式

附图仅用于示例性说明, 不能理解为对本专利的限制;

为了更好说明本实施例, 附图某些部件会有省略、放大或缩小, 并不代表实际产品的尺寸;

- 对于本领域技术人员来说, 附图中某些公知结构及其说明可能省略是可以理解的。
- 30

下面结合附图和实施例对本发明的技术方案做进一步的说明。

实施例 1

首先利用薛定谔公司测试相对结合自由能的靶标与配体分子 (J. Am. Chem. Soc. 2015, 137, 2695-2703.) 对本专利报道的方法进行测试。

- 5 由于小分子间差异较小, 因此利用不考虑约束力的如图 1 所示的较为简单的热力学循环, 其中 Rec 代表 Receptor, 即药物-靶受体; Lig 代表 Ligand, 即药物小分子配体; Dumm 代表 Dummy, 即虚原子, 所有的虚原子与周围的环境没有任何的相互作用。通过计算得到 Lig 体系及 Rec-Lig 体系的结合自由能 $\Delta A_{binding}$, 依据如图 1 所示的热力学循环, 在分别求得 $\Delta A_{annihilation}^L$ 和 $\Delta A_{annihilation}^{RL}$
- 10 的大小后, 即可得到结合自由能 $\Delta A_{binding}$ 。

进一步的, 针对 CDK2 药物靶标的 16 个药物小分子为例, 使用不考虑约束力的如图 1 所示的较为简单的热力学循环的具体步骤如下:

S1: 构建药物小分子配体 Lig 体系以及药物-靶标的 Rec-Lig 体系, 为中间微扰状态参数准备, 具体为:

- 15 S11: 准备受体-配体复合物的结构, 利用分子对接的方法将 16 个药物小分子配体对接进大分子 CDK2 受体的结合口袋;

S12: 将小分子配体通过 Gaussian 软件在 HF/6-31G* 水平下计算 ESP 电荷, 使用 Antechamber 拟合出 RESP 电荷, 并指定 GAFF 力场参数;

- S13: 将受体-配体复合物中的受体指定 AMBER03 力场对于整个复合物添加
- 20 8 Å 的去角八面体的水箱, 添加抗衡离子使整个体系保持电中性, 将最终产生的体系的坐标与初始的力场参数均保存为力场文件, 作为 Rec-Lig 体系自由能微扰的初始状态参数;

- S14: 将单独的小分子配体添加 8 Å 的去角八面体的水箱, 将最终产生的体系的坐标与初始的力场参数均保存为力场文件, 作为 Lig 体系进行自由能微扰的初
- 25 始状态的参数;

S15: 将 Rec-Lig 体系以及 Lig 体系中的小分子电荷进行湮灭。对于这两个体系, 分别通过 5 步, 将小分子自中所有原子的电荷都减小至 0。即每一步将小分子中每个原子的电荷修改为原来的 0.8、0.6、0.4、0.2、0, 将每一步修改过电荷之后的参数文件分别保存为独立的力场文件, 作为中间微扰状态的参数;

- 30 S16: 在小分子电荷进行湮灭后, 小分子与外界还保留着范德华相互作用。

因此需要对 Rec-Lig 体系以及 Lig 体系中小分子的范德华相互作用进行湮灭。分子间的范德华相互作用以 Lennard-Jones (LJ) 势能计算, 计算公式为:

$$V_{LJ} = 4\varepsilon_{ij} \left[\left(\frac{r_{ij}}{r} \right)^{12} - 2 \left(\frac{r_{ij}}{r} \right)^6 \right];$$

其中, ε_{ij} 为原子 i 和原子 j 之间的 LJ 势能参数, r_{ij} 为原子 i 和原子 j 之间距离; 在本实施例中, 通过减小 ε_{ij} 来减弱原子间的范德华相互作用; 由于随着范德华相互作用的湮灭, 对这些非真实状态的采样变得越发困难, 因此当范德华作用减小到比较小时, 需让不同状态之间更加接近以提高采样的收敛性。因此每个原子的 ε 以 $(1 - 0.2i)^6$ 的形式进行减少。按照该方式进行了 5 步对于范德华相互作用的修改之后, 小分子与外界的所有作用全部消失, 达到了对小分子湮灭的效果。

10 将所有修改过 ε 的参数文件保存为独立的力场文件, 作为中间微扰状态的参数;

S2: 利用分子动力学模拟程序 AMBER, 对 Lig 体系及 Rec-Lig 体系分别用原始的力场参数文件以及步骤 S1 生活曾的中间微扰状态的力场文件进行分子动力学模拟。在模拟时不可使用 SHAKE 函数对体系进行限制, 因为 SHAKE 函数的特性会使中间状态不稳定, 因此模拟的步长选择为 1fs。此外, 需要对体系周期性边界条件, 对长程静电互相作用作用采用 PME 的方法进行计算, 对短程的范德华相互作用选取 8Å 的截断值; 对于每个体系的每个状态都在 300K 的条件下进行 4ns 的分子动力学模拟。在最后 2ns 的模拟过程中, 每隔 100fs 提取一帧的构象, 每个状态的轨迹包含了 20000 帧构象;

15

S3: 将得到的 Lig 体系及 Rec-Lig 体系每个状态的分子动力学轨迹, 通过相邻状态的力场文件进行单点能计算, 得到每个相邻状态之间的势能差 ΔU , 具体为:

20

基于第 i 状态的分子动力学轨迹为基础, 求解状态 $i + 1$ 状态与状态 i 的势能差的系统平均, 由此求解出来的为向前微扰势能差; 求解状态 $i - 1$ 状态与状态 i 的势能差的系统平均, 由此求解出来的为向后微扰势能差; 将向前微扰势能差和向后微扰势能差的结果结合起来, 使用 BAR 方法进行计算, 有效提高最终能量计算的准确性。

25

S4: 针对得到每个相邻状态之间的势能差进行概率分布统计, 利用多个高斯函数的线性组合进行拟合, 得到平滑的概率分布, 具体为:

对于所有的势能差 ΔU 的概率分布进行你和你, 经过步骤 S3 后, 由于每个状态

的轨迹都包含 20000 个构象。因此每个状态的轨迹经计算后得到 20000 个 ΔU 的数据点；如图 2 所示，具体为：

S41：先将所得势能差 ΔU 转换成对应的概率分布 $P(\Delta U)_{data}$ ；

S42：利用多个高斯函数的线性组合对概率分布 $P(\Delta U)_{data}$ 进行拟合，具体计

5 算公式为：

$$P(\Delta U)_{fit} = \sum_{i=1}^5 c_i \exp\left(-\frac{(\Delta U - \mu_i)^2}{2\sigma_i^2}\right);$$

其中， $P(\Delta U)_{fit}$ 为经高斯函数拟合后的概率分布， c_i 为第 i 个高斯函数的参数， μ_i 为第 i 个高斯函数的平均值， σ_i 为第 i 个高斯函数的标准差；共包含 15 个参数 c_i, μ_i, σ_i ，其中 $i = 1, 2, 3, 4, 5$ ，由此定义费用函数的形式为：

$$\begin{aligned} Cost(c_i, \mu_i, \sigma_i, \Delta U, P(\Delta U)_{data}) &= \sum_{\text{over all datapoints}} (P(\Delta U)_{data} - P(\Delta U)_{fit})^2 \\ &= \sum_{\text{over all data points}} \left(P(\Delta U)_{data} - \sum_{i=1}^5 c_i \exp\left(-\frac{(\Delta U - \mu_i)^2}{2\sigma_i^2}\right) \right)^2, \end{aligned}$$

10

其中， $P(\Delta U)_{data}$ 为动力学采样得到的概率分布， $P(\Delta U)_{fit}$ 为经高斯函数拟合后的概率分布， c_i 为第 i 个高斯函数的参数， μ_i 为第 i 个高斯函数的平均值， σ_i 为第 i 个高斯函数的标准差；利用随机梯度下降法对费用函数进行优化，具体为采用 SGD 法优化这 15 个参数 c_i, μ_i, σ_i ，其中 $i = 1, 2, 3, 4, 5$ ，具体为：

$$15 \quad (c_i, \mu_i, \sigma_i)_t = (c_i, \mu_i, \sigma_i)_{t-1} - \lambda \nabla_{c_i, \mu_i, \sigma_i} Cost(c_i, \mu_i, \sigma_i, \Delta U, P(\Delta U)_{data});$$

其中， $(c_i, \mu_i, \sigma_i)_t$ 为经 t 次迭代后的参数值；通过不断迭代直至第 t 次迭代所得 $(c_i, \mu_i, \sigma_i)_t$ 即为拟合后的函数的最有参数。经拟合后，结果的概率分布如图 3 所示。

S5：根据拟合后的概率分布，重新生成每一状态的势能差，通过计算得到药物与靶标的结合自由能，具体为：

20 S51：利用经过高斯函数拟合后的概率分布，重新生成 500000 至 1000000 个新的势能差 $\Delta U'$ ；

S52：并基于新生成的势能差 $\Delta U'$ 求解方程组，得到相邻状态之间的结合自由能之差 $\Delta A_{i,i+1}$ ，具体方程为：

$$\exp(-\beta \Delta A_{i,i+1}) = \frac{\langle \{1 + \exp[\beta(\Delta U_{i,i+1} - C)]\}^{-1} \rangle_i}{\langle \{1 + \exp[-\beta(\Delta U_{i,i+1} - C)]\}^{-1} \rangle_{i+1}} \exp(-\beta C);$$

$$C = \Delta A_{i,i+1} + \frac{1}{\beta} \ln \frac{n_i}{n_{i+1}};$$

其中, $\Delta A_{i,i+1}$ 为两相邻状态 i 与 $i+1$ 的结合自由能差, $\Delta U_{i,i+1}$ 为基于拟合后的概率分布新生成的势能差, $\beta=1/k_B T$, k_B 为玻尔兹曼常数, n_i 与 n_{i+1} 分别为两相邻状态 i 与 $i+1$ 采样所包含的样本量, C 为方程组之间的隐含变量;

- 5 S53: 对 Rec-Lig 体系所有状态势能差进行加和运算, 得到能量 $\Delta A_{annihilation}^{RL}$; 对 Lig 体系所有状态能量差进行加和运算, 得到能量 $\Delta A_{annihilation}^L$, 计算得到药物与靶标的结合自由能 $\Delta A_{binding}$, 具体计算公式为:

$$\Delta A_{binding} = \Delta A_{annihilation}^L - \Delta A_{annihilation}^{RL};$$

其中, $\Delta A_{annihilation}^L$ 表示将溶液中药物小分子全部微扰至湮灭状态的能量;

- 10 $\Delta A_{annihilation}^{RL}$ 表示将药物-靶标复合物中药物小分子全部扰乱至湮灭状态的能量; 其中, 湮灭状态为药物小分子的电荷以及范德华力 L-J 势能为 0, 周围院子无法感知到药物小分子的存在。

- 在具体实施过程中, 根据步骤 S1-S5, 对于 CDK2 体系的 16 个小分子进行了计算, 这 16 个小分子的结构已知, IC_{50} 可以由实验测定, 结合自由能的实验值的计算公式为:

$$\Delta G = -RT \ln IC_{50};$$

得到的结果如图 4 所示, 对比可得到, 本发明所述的方法计算得到的绝对结合自由度与实验值之间具有非常高的相关性, 因而该方法所得的结果精确度高。

- 在具体实施过程中, 所述的预测配体受体间绝对结合自由能的高斯函数增强的自由能微扰方法具有较高的准确性, 现有的自由能微扰方法仅能用于计算相对结合自由能, 本发明能够计算绝对结合自由能, 因而本发明所述的方法能够更广泛地应用于理性药物设计的工作中。

在具体实施过程中, 拟合采样概率分布 $P(\Delta U)_{data}$ 能使自由能微扰准确性及收敛性提升的方法为:

- 25 由自由能微扰的基本公式:

$$\Delta A = -\beta \ln \int \exp(-\beta \Delta U) P_0(\Delta U) d\Delta U;$$

- 其中, β 为常数, ΔA 表示两状态自由能差; 采样得到的 $P(\Delta U)$ 需要与 $\exp(-\beta \Delta U)$ 相乘, 结果如图 5 能差 ΔA ; 在 $P(\Delta U)$ 与 $\exp(-\beta \Delta U)$ 进行相乘时, $P(\Delta U)$ 处于较小 ΔU 的部分被与 $\exp(-\beta \Delta U)$ 的乘积放大, 导致该部分的采样准确性对于能量计算的计算有非常大的影响; 而该部分由于数据点出现的概率小, 导致采样

往往不充分，因此在面临较大的微扰时会导致收敛性及准确性不足。对于采样得到的 $P(\Delta U)$ 进行拟合时，考虑到 $P(\Delta U)$ 类高斯分布的性质，用一系列高斯函数的加和进行拟合，该拟合函数充分考虑到 $P(\Delta U)$ 的整体性质， $P(\Delta U)$ 处于较小 ΔU 的部分也通过外推法得到了优化，优化后的能量计算的准确性与收敛性随之极大的提升。所示，对该乘积进行积分，得到的曲线面积正比于自由

在具体实施过程中，通过本发明所述方法利用自由能微扰进行绝对自由能计算的准确度大幅提升。以针对 CDK2 药物靶标的 16 个药物小分子为例，以现有的自由能微扰方法不对概率分布 $P(\Delta U)$ 进行拟合计算，得到的实验值与计算值的线性回归结果如图 4 所示，实验与计算值之间的回归系数为 0.40；而使用本发明所述方法对概率分布进行拟合计算后，得到的实验与计算值之间的回归系数为 0.88，对比可知该发明专利中的自由能微扰方法比起现有的自由能微扰方法在进行药物与靶标的绝对结合自由能计算时准确度大幅提升。

在具体实施过程中，使用的高斯函数增强的自由能微扰方法进行绝对自由能计算，其结果与现有的结合自由能计算方法进行对比，对于 CDK2、TYK2、JNK1、Thrombin 这几个药物靶标以及相应的 63 个药物小分子，其对比结果如图 6 所示，由图可知使用高斯函数增强的自由能微扰方法进行绝对自由能计算的结果准确度与相对结合自由能计算方法相当，但绝对自由能计算不需要已知的参考药物分子，其应用范围远大于现有的相对结合自由能计算方法。

实施例 2

进一步的，利用如图 7、图 8 所示的完整版热力学循环，对本专利报道的绝对结合自由能计算方法进行测试，步骤如下：

S1：对受体蛋白-配体体系，即 Rec-Lig 体系进行常规的分子动力学模拟，并基于模拟获得的轨迹得到添加约束所需的距离、夹角、二面角数据；

S2：构建热力学循环，构建 $\Delta A_{\text{annihilation}}^L$ 、 $\Delta A_{\text{annihilation}}^{RL}$ 和 $\Delta A_{\text{restrain}}^L$ ， $\Delta A_{\text{restrain}}^{RL}$ 根据添加约束所需的距离、夹角、二面角数据得到，并对 $\Delta A_{\text{restrain}}^L$ 的自由能进行解析计算；

其中： $\Delta A_{\text{annihilation}}^L$ 表示将溶液中药小分子全部微扰至湮灭状态的能量； $\Delta A_{\text{annihilation}}^{RL}$ 表示将药物-靶标复合物中药小分子全部微扰至湮灭状态的能量； $\Delta A_{\text{restrain}}^{RL}$ 表示对药物-靶标复合物中药小分子添加简谐势约束的能量变化；

S3：利用分子动力学模拟程序，对 $\Delta A_{\text{annihilation}}^L$ 、 $\Delta A_{\text{annihilation}}^{RL}$ 、 $\Delta A_{\text{restrain}}^L$ 步骤

中的每一个窗口进行分子动力学模拟采样，得到 $\Delta A_{\text{annihilation}}^L$ 、 $\Delta A_{\text{annihilation}}^{RL}$ 、 $\Delta A_{\text{restrain}}^L$ 步骤中每一个窗口的分子动力学轨迹；

S4: 根据 $\Delta A_{\text{annihilation}}^L$ 、 $\Delta A_{\text{annihilation}}^{RL}$ 、 $\Delta A_{\text{restrain}}^L$ 步骤中的每一个窗口的分子动力学轨迹，通过相邻窗口的力场文件进行计算，得到每个相邻窗口之间的势能差 ΔU ；

S5: 对于 $\Delta A_{\text{annihilation}}^L$ 、 $\Delta A_{\text{annihilation}}^{RL}$ 这两步，根据对应窗口的 ΔU ，分别计算得到 $\Delta A_{\text{annihilation}}^L$ 、 $\Delta A_{\text{annihilation}}^{RL}$ 步骤的自由能；

S6: 对于 $\Delta A_{\text{restrain}}^{RL}$ 这一步每个相邻窗口之间的势能差 ΔU 进行概率分布统计，利用约束概率分布函数进行拟合，得到平滑的概率分布，并根据拟合结果检查收敛性；

S7: 根据拟合后的概率分布，重新生成 $\Delta A_{\text{restrain}}^{RL}$ 这一步中每一窗口的势能差，通过计算得到 $\Delta A_{\text{restrain}}^{RL}$ 的能量，并根据之前步骤得到的 $\Delta A_{\text{annihilation}}^L$ 、 $\Delta A_{\text{annihilation}}^{RL}$ 以及 $\Delta A_{\text{restrain}}^L$ 计算得到药-靶结合自由能 $\Delta A_{\text{binding}}$ 。

在具体实施过程中，本发明提供一种基于约束概率分布函数优化的自由能微扰方法，实现全自动得到添加约束所需的复合物拓扑文件，约束概率分布函数用于拟合实验中的概率分布，使自由能微扰（FEP）得到的约束窗口内的自由能变化自由值更精准，在保证计算精度的前提下，极大减少了约束步骤的窗口数，减少了计算量，提高了计算效率。

在具体实施过程中，根据以上步骤，基于本发明提供的自动添加约束方法和约束概率分布函数拟合约束步骤的概率分布，以人类磷酸二酯酶 4D 及其小分子抑制剂罗氟司特的复合物体系（PDB 代码为 1XOQ）和人类 BRD4 蛋白及其小分子配体阿普唑仑（PDB 代码为 3U5J）为例，将约束窗口由现有文献中使用的 10 个（Chem. Sci., 2016, 7, 207-218; J. Am. Chem. Soc. 2017, 139, 946-957 等）减少到 1 个，通过对比考察 10 步方案计算所得的 $\Delta A_{\text{restrain}}^{RL}$ 与 1 步方案差距，获得的测试结果如表 1 所示。

表1. 选择不同数目的窗口计算得出的 $\Delta A_{restrain}^{RL}$

PDB	数目的窗口	$\Delta A_{restrain}^{RL}$ (Kcal/mol)
1XOQ	1	0.761
	10	0.716
3U5J	1	1.006
	10	1.178

从上表可以看出,在都以本发明提供的自动添加约束程序进行简谐势约束的前提下,设置不同的约束步骤扰动窗口数,经过约束概率分布函数拟合其概率分布后,计算得到 $\Delta A_{restrain}^{RL}$ 相差较小,差值仅为 0.2 kcal/mol 以内。该对比说明在
5 使用本发明提供的自动添加约束程序及“约束概率分布函数”拟合基础上,只设置 1 个约束扰动窗口的快速约束方案是可行的。

更具体的,所述步骤 S1 具体包括以下步骤:

S11: 准备受体蛋白 Rec-配体 Lig 复合物坐标文件;

S12: 准备小分子配体的电荷,并对配体指定力场;

10 S13: 将受体-配体复合物中的受体指定力场,并将该复合物在水箱中进行浸泡,准备好复合物进行分子模拟的初始体系,记为 Rec-Lig 体系,并到处力场文件作为 Rec-Lig 体系初始状态参数;

S14: 将单独小分子配体浸泡在水箱中,准备好单独的小分子的分子模拟初始体系,记为 Lig 体系,并到处力场文件作为 Lig 体系初始状态参数;

15 S15: 根据 Rec-Lig 体系初始状态的力场参数,进行短时间的分子动力学模拟,得到运动轨迹;

S16: 基于模拟得到的轨迹,统计与小分子配体相对位置变化较小的受体蛋白的氨基酸残基,取该残基中与小分子配体相对距离最稳定的三个原子作为添加约束所需的属于受体蛋白的三个原子,选取小分子配体中最稳定的三个原子作为
20 添加约束所需的属于配体小分子的三个原子,得到添加约束所需的距离、夹角、二面角数据。

更具体的,所述步骤 S2 具体包括以下步骤:

S21: 根据准备好的 Rec-Lig 体系的力场参数,根据步骤 S16 得到的 6 个原子,添加一个长度、两个夹角、三个二面角约束,得到计算 $\Delta A_{restrain}^{RL}$ 所需窗口的
25 分子动力学模拟参数;

S22: 进一步将 Rec-Lig 体系中的小分子的电荷以及范德华参数依次分别减小至 0, 得到计算 $\Delta A_{annihilation}^{RL}$ 所需窗口的分子动力学模拟参数;

S23: 对于 Rec-Lig 体系中小分子的电荷以及范德华参数依次分别减小至 0, 得到计算 $\Delta A_{annihilation}^L$ 所需窗口的分子动力学模拟参数;

5 S24: 对 $\Delta A_{restrain}^L$ 的自由能进行解析计算。

其中, 在所述步骤 S24 中, 所述 $\Delta A_{restrain}^L$ 自由能的计算公式具体为:

$$\Delta A_{restrain}^L = -k_B T \ln \left(\frac{8\pi^2 V}{r_{aA}^2 \sin \theta_a \sin \theta_A} \frac{(k_{r_{aA}} k_{\theta_a} k_{\theta_A} k_{\phi_{ba}} k_{\phi_{aA}} k_{\phi_{AB}})^{1/2}}{(2\pi k_B T)^3} \right)$$

其中, k_B 为玻尔兹曼常数, V 为系统的体积, r_{aA} 为一个两原子间的距离用于距离简谐势约束, θ_a 和 θ_A 为两个三原子间的夹角用于夹角简谐势约束, 诸多 k 值为用于约束一个距离 r_{aA} 、两个夹角 θ_a, θ_A 、三个二面角 $\phi_{aA}, \phi_{AB}, \phi_{ba}$ 的简谐势力常数。

更具体的, 所述步骤 S5 具体为:

根据对应窗口的 ΔU , 采用 Benneet Acceptance Ratio 计算方法, 即 BAR 计算方法进行 $\Delta A_{annihilation}^L$ 、 $\Delta A_{annihilation}^{RL}$ 自由能的计算, 具体计算公式为:

$$\exp(-\beta \Delta A_{i,i+1}) = \frac{\left\langle \{1 + \exp[\beta(\Delta U_{i,i+1} - C)]\}^{-1} \right\rangle_i}{\left\langle \{1 + \exp[-\beta(\Delta U_{i,i+1} - C)]\}^{-1} \right\rangle_{i+1}} \exp(-\beta C)$$

$$C = \Delta A_{i,i+1} + \frac{1}{\beta} \ln \frac{n_i}{n_{i+1}}$$

15 其中, $\Delta A_{i,i+1}$ 为两相邻窗口 i 与 $i+1$ 的结合自由能差, $\Delta U_{i,i+1}$ 为基于拟合后的概率分布新生成的势能差, $\beta=1/k_B T$, k_B 为玻尔兹曼常数, n_i 与 n_{i+1} 分别为两相邻状态 i 与 $i+1$ 采样所包含的样本量, C 为方程组之间的隐含变量;

将 $\Delta A_{annihilation}^{RL}$ 和 $\Delta A_{annihilation}^L$ 对应的计算窗口得到的 $\Delta A_{i,i+1}$ 进行加和, 得到 $\Delta A_{annihilation}^{RL}$ 和 $\Delta A_{annihilation}^L$ 的数值。

20 更具体的, 所述步骤 S6 具体包括以下步骤:

S61: 对计算 $\Delta A_{restrain}^{RL}$ 的窗口中每个窗口之间的势能差进行概率分布统计, 得到概率分布 $P(\Delta U)_{data}$;

S62: 基于约束概率分布函数对概率分布 $P(\Delta U)_{data}$ 进行拟合, 具体表达式为:

$$P(\Delta U)_{fit} = \sum_{i=1}^n \left(\frac{a_i}{2} \cdot b_i^{\frac{3}{2}} \cdot \Delta U^2 \cdot e^{-b_i \Delta U - (c_i \Delta U)^{n_1}} + \frac{1}{1 + (\frac{d_i}{\Delta U})^{n_2}} \left(h_i \cdot \frac{1}{\sqrt{\pi} \mu_i} \cdot e^{-\frac{(\Delta U - \mu_i)^2}{2\sigma_i^2}} \right) \right)$$

其中, $P(\Delta U)_{fit}$ 为经过约束概率分布函数拟合后的平滑的概率分布, 其中 $a_i, b_i, c_i, d_i, h_i, \mu_i, \sigma_i$ 为第 i 个“约束概率分布函数”的参数, n 为选取拟合函数的数目, n_1 和 n_2 为两个正整数的修正因子, 在本申请给出的计算案例中, 取 n, n_1 和 n_2 分别为 1, 10 和 4。为说明拟合效果, 图 9 为对添加约束过程的状态采样所得概率分布 $P(\Delta U)$ 被一系列约束概率分布函数拟合的结果示意图。图中灰点为采样所得概率分布点, 黑线为最终拟合所得的分布曲线, 灰实线和灰虚线分别为公式中的前后两项。

更具体的, 所述步骤 S7 具体包括以下步骤:

S71: 根据概率分布 $P(\Delta U)_{fit}$, 重新生成每一状态的势能差 $\Delta U'$;

S72: 基于新的势能差 $\Delta U'$, 利用步骤 S5 的计算方法重新计算得到 $\Delta A_{restrain}^{RL}$ 的计算数值;

S73: 根据得到的 $\Delta A_{restrain}^L$ 、 $\Delta A_{annihilation}^{RL}$ 、 $\Delta A_{annihilation}^L$ 和 $\Delta A_{restrain}^{RL}$, 计算出药-靶结合自由能 $\Delta A_{binding}$ 。

其中, 在所述步骤 S73 中, 所述药-靶结合自由能 $\Delta A_{binding}$ 的计算过程具体为:

$$\Delta A_{binding} = \Delta A_{annihilation}^L + \Delta A_{restrain}^L - \Delta A_{annihilation}^{RL} - \Delta A_{restrain}^{RL}。$$

实施例 3

实施例 1 的结果说明了高斯函数拟合可以增加 $\Delta A_{annihilation}^{RL}$ 和 $\Delta A_{annihilation}^L$ 计算的收敛性, 实施例 2 的结果说明自动添加约束的方法结合约束概率分布函数进行拟合, 可以使得 $\Delta A_{restrain}^{RL}$ 项能量结算结果达到非常好的收敛性。基于实施例 1 和实施例 2 的基础上, 以两个针对 BRD4 蛋白靶标的药物小分子为例。采用自动添加约束并在约束过程中只设置一个扰动窗口的约束方案, 在处理约束过程中的概率分布数据时, 采用本发明中的约束概率分布函数对其概率分布散点图进行拟合。并基于图 8 提供的完整版热力学循环, 计算绝对结合自由能 $\Delta A_{binding}$ 。具体

步骤如下:

S1: 下载 BRD4 蛋白晶体结构 3MXF 和 4HBV, 准备 BRD4 的力场参数, 选定 Rec-Lig 体系以及 Lig 体系, 对 Rec-Lig 体系进行一小段常规的分子动力学模拟, 并基于模拟获得的轨迹得到添加约束所需距离、夹角、二面角数据;

S2: 根据图 8 所示构建热力学循环, 构建 $\Delta A_{annihilation}^L$ 、 $\Delta A_{annihilation}^{RL}$ 、 $\Delta A_{restrain}^{RL}$

步骤所需的分子动力学模拟参数，其中对 $\Delta A_{\text{restrain}}^{RL}$ 这一步添加约束使用 S1 步骤得到的距离、夹角、二面角数据，并对 $\Delta A_{\text{restrain}}^L$ 这一步的自由能进行解析解计算；

S3：利用分子动力学模拟程序，对 $\Delta A_{\text{annihilation}}^L$ 、 $\Delta A_{\text{annihilation}}^{RL}$ 、 $\Delta A_{\text{restrain}}^{RL}$ 步骤中的每一个窗口进行分子动力学模拟采样，得到 $\Delta A_{\text{annihilation}}^L$ 、 $\Delta A_{\text{annihilation}}^{RL}$ 、

5 $\Delta A_{\text{restrain}}^{RL}$ 步骤中的每一个窗口的分子动力学轨迹；

S4：根据 $\Delta A_{\text{annihilation}}^L$ 、 $\Delta A_{\text{annihilation}}^{RL}$ 、 $\Delta A_{\text{restrain}}^{RL}$ 步骤中的每一个窗口的分子动力学轨迹，通过相邻窗口的力场文件进行计算，得到每个相邻窗口之间的势能差 ΔU ；

S5：对于 $\Delta A_{\text{annihilation}}^L$ 、 $\Delta A_{\text{annihilation}}^{RL}$ 这两步，根据对应窗口的 ΔU ，分别计算
10 得到 $\Delta A_{\text{annihilation}}^L$ 、 $\Delta A_{\text{annihilation}}^{RL}$ 步骤的自由能。

S6：对于 $\Delta A_{\text{restrain}}^{RL}$ 这一步每个相邻窗口之间的势能差 ΔU 进行概率分布统计，利用本专利给出的“约束概率分布函数”进行拟合，得到平滑的概率分布，并根据拟合结果检查收敛性；

S7：根据拟合后的概率分布，重新生成 $\Delta A_{\text{restrain}}^{RL}$ 这一步中每一窗口的势能差，
15 通过计算得到 $\Delta A_{\text{restrain}}^{RL}$ 的能量，并根据之前步骤得到的 $\Delta A_{\text{annihilation}}^L$ 、 $\Delta A_{\text{annihilation}}^{RL}$ 、以及 $\Delta A_{\text{restrain}}^L$ 计算得到药-靶结合自由能 $\Delta A_{\text{binding}}$ 。

在具体实施过程中，根据以上步骤，考察针对 BRD4 蛋白靶标与两个两个药物小分子的计算结合自由能 $\Delta A_{\text{binding}}$ 与实验值的差别，获得的结果如表 2 所示：

表2. 药-靶结合自由能计算结果与实验结果对比^a

PDB	ΔA_{cal} (Kcal/mol) ^b	ΔA_{exp} (Kcal/mol) ^c	$\Delta A_{\text{cal}} - \Delta A_{\text{exp}}$ (Kcal/mol)
3MXF	10.5	9.6	0.9
4HBV	6.5	6.3	0.2

^a 计算过程中使用自动添加约束方法

^b ΔA_{cal} 为计算所得药靶结合自由能

^c ΔA_{exp} 为实验所得药靶结合自由能

20 从计算与实验结果的对比可知，计算结果与实验结果非常接近，差距在 2 kcal/mol 之内。

在最新文献中，计算 $\Delta A_{\text{restrain}}^{RL}$ 至少需 10 个扰动窗口，且需要手动添加约束（Chem. Sci., 2016, 7, 207-218; J. Am. Chem. Soc. 2017, 139, 946-957）。基于本发明的技术方案，可实现全自动得到添加简谐势约束所需的复合物拓扑文件，约

束概率分布函数可用于拟合实验所得的 ΔU 与其概率的散点图，使得 FEP 计算得到的约束窗口内的自由能变化值更精准。FEP 方法是目前准确度较高的方法，但该方法需要非常大的计算量，而使用本发明的技术方案，可以在保证计算精度的前提下，极大减少约束步骤的窗口数，使得计算窗口数目从 10 降至 1，这
5 同时也极大减少了计算量，计算量仅为之前的 1/10。以上实施例也说明了结果的准确性。

显然，本发明的上述实施例仅仅是为清楚地说明本发明所作的举例，而并非是对本发明的实施方式的限定。对于所属领域的普通技术人员来说，在上述说明的基础上还可以做出其它不同形式的变化或变动。这里无需也无法对所有的实施方式予以穷举。凡在本发明的精神和原则之内所作的任何修改、等同替换和改进
10 等，均应包含在本发明权利要求的保护范围之内。

权 利 要 求 书

1. 预测药-靶结合强度的绝对结合自由能微扰方法，其特征在于，包括以下步骤：

5 S1：对受体蛋白-配体体系，即 Rec-Lig 体系进行常规的分子动力学模拟，并基于模拟获得的轨迹得到添加约束所需的距离、夹角、二面角数据；

S2：构建热力学循环，构建 $\Delta A_{annihilation}^L$ 、 $\Delta A_{annihilation}^{RL}$ 和 $\Delta A_{restrain}^L$ ， $\Delta A_{restrain}^L$ 根据添加约束所需的距离、夹角、二面角数据得到，并对 $\Delta A_{restrain}^L$ 的自由能进行解析计算；

10 其中： $\Delta A_{annihilation}^L$ 表示将溶液中药物小分子全部微扰至湮灭状态的能量；

$\Delta A_{annihilation}^{RL}$ 表示将药物-靶标复合物中药物小分子全部微扰至湮灭状态的能量；

$\Delta A_{restrain}^{RL}$ 表示对药物-靶标复合物中药物小分子添加简谐势约束的能量变化；

S3：利用分子动力学模拟程序，对 $\Delta A_{annihilation}^L$ 、 $\Delta A_{annihilation}^{RL}$ 、 $\Delta A_{restrain}^L$ 步骤中的每一个窗口进行分子动力学模拟采样，得到 $\Delta A_{annihilation}^L$ 、 $\Delta A_{annihilation}^{RL}$ 、

15 $\Delta A_{restrain}^L$ 步骤中每一个窗口的分子动力学轨迹；

S4：根据 $\Delta A_{annihilation}^L$ 、 $\Delta A_{annihilation}^{RL}$ 、 $\Delta A_{restrain}^L$ 步骤中的每一个窗口的分子动力学轨迹，通过相邻窗口的力场文件进行计算，得到每个相邻窗口之间的势能差 ΔU ；

S5：对于 $\Delta A_{annihilation}^L$ 、 $\Delta A_{annihilation}^{RL}$ 这两步，根据对应窗口的 ΔU 的概率分布，，
20 使用高斯函数进行拟合，并计算得到 $\Delta A_{annihilation}^L$ 、 $\Delta A_{annihilation}^{RL}$ 步骤的自由能；

S6：对于 $\Delta A_{restrain}^{RL}$ 这一步每个相邻窗口之间的势能差 ΔU 进行概率分布统计，利用约束概率分布函数进行拟合，得到平滑的概率分布，并根据拟合结果检查收敛性；

S7：根据拟合后的概率分布，重新生成 $\Delta A_{restrain}^{RL}$ 这一步中每一窗口的势能差，
25 通过计算得到 $\Delta A_{restrain}^{RL}$ 的能量，并根据之前步骤得到的 $\Delta A_{annihilation}^L$ 、 $\Delta A_{annihilation}^{RL}$ 、以及 $\Delta A_{restrain}^L$ 计算得到药-靶结合自由能 $\Delta A_{binding}$ 。

2. 根据权利要求1所述的预测药-靶结合强度的绝对结合自由能微扰方法，其特征在于，所述步骤S1具体包括以下步骤：

S11：准备受体蛋白 Rec-配体 Lig 复合物坐标文件；

30 S12：准备小分子配体的电荷，并对配体指定力场；

S13: 将受体-配体复合物中的受体指定力场, 并将该复合物在水箱中进行浸泡, 准备好复合物进行分子模拟的初始体系, 记为 Rec-Lig 体系, 并到处力场文件作为 Rec-Lig 体系初始状态参数;

5 S14: 将单独小分子配体浸泡在水箱中, 准备好单独的小分子的分子模拟初始体系, 记为 Lig 体系, 并到处力场文件作为 Lig 体系初始状态参数;

S15: 根据 Rec-Lig 体系初始状态的力场参数, 进行短时间的分子动力学模拟, 得到运动轨迹;

10 S16: 基于模拟得到的轨迹, 统计与小分子配体相对位置变化较小的受体蛋白的氨基酸残基, 取该残基中与小分子配体相对距离最稳定的三个原子作为添加约束所需的属于受体蛋白的三个原子, 选取小分子配体中最稳定的三个原子作为添加约束所需的属于配体小分子的三个原子, 得到添加约束所需的距离、夹角、二面角数据。

3. 根据权利要求 2 所述的预测药-靶结合强度的绝对结合自由能微扰方法, 其特征在于, 所述步骤 S2 具体包括以下步骤:

15 S21: 根据准备好的 Rec-Lig 体系的力场参数, 根据步骤 S16 得到的 6 个原子, 添加一个长度、两个夹角、三个二面角约束, 得到计算 $\Delta A_{\text{restrain}}^{RL}$ 所需窗口的分子动力学模拟参数;

S22: 进一步将 Rec-Lig 体系中的小分子的电荷以及范德华参数依次分别减小至 0, 得到计算 $\Delta A_{\text{annihilation}}^{RL}$ 所需窗口的分子动力学模拟参数;

20 S23: 对于 Rec-Lig 体系中小分子的电荷以及范德华参数依次分别减小至 0, 得到计算 $\Delta A_{\text{annihilation}}^L$ 所需窗口的分子动力学模拟参数;

S24: 对 $\Delta A_{\text{restrain}}^L$ 的自由能进行解析计算。

4. 根据权利要求 3 所述的预测药-靶结合强度的绝对结合自由能微扰方法, 其特征在于, 在所述步骤 S24 中, 所述 $\Delta A_{\text{restrain}}^L$ 自由能的计算公式具体为:

$$25 \quad \Delta A_{\text{restrain}}^L = -k_B T \ln \left(\frac{8\pi^2 V}{r_{aA}^2 \cdot \sin \theta_a \cdot \sin \theta_A} \frac{(k_{r_{aA}} k_{\theta_a} k_{\theta_A} k_{\phi_{ba}} k_{\phi_{aA}} k_{\phi_{AB}})^{1/2}}{(2\pi k_B T)^3} \right)$$

其中, k_B 为玻尔兹曼常数, V 为系统的体积, r_{aA} 为一个两原子间的距离用于距离简谐势约束, θ_a 和 θ_A 为两个三原子间的夹角用于夹角简谐势约束, 诸多 k 值为用于约束一个距离 r_{aA} 、两个夹角 θ_a, θ_A 、三个二面角 $\phi_{aA}, \phi_{AB}, \phi_{ba}$ 的简谐势力常数。

5. 根据权利要求4所述的预测药-靶结合强度的绝对结合自由能微扰方法，其特征在于，在所述步骤S3或S4中，采用包括AMBER、Gromacs、Namd在内的任意分子动力学模拟软件进行计算。

6. 根据权利要求4所述的预测药-靶结合强度的绝对结合自由能微扰方法，
5 其特征在于，所述步骤S5具体包括以下步骤：

S51：针对得到每个相邻状态之间的势能差进行概率分布统计，得到概率分布 $P(\Delta U)_{data}$ ；

S52：将概率分布 $P(\Delta U)_{data}$ 利用多个高斯函数的线性组合进行拟合，得到平滑的概率分布，具体计算公式为：

$$P(\Delta U)_{fit} = \sum_{i=1}^5 c_i \exp\left(-\frac{(\Delta U - \mu_i)}{2\sigma_i^2}\right)$$

10 其中， $P(\Delta U)_{fit}$ 为经高斯函数拟合后的概率分布， c_i 为第 i 个高斯函数的参数， μ_i 为第 i 个高斯函数的平均值， σ_i 为第 i 个高斯函数的标准差；

S53：根据得到的 $P(\Delta U)_{fit}$ ，采用 Benneet Acceptance Ratio 计算方法，即 BAR 计算方法进行 $\Delta A_{annihilation}^L$ 、 $\Delta A_{annihilation}^{RL}$ 自由能的计算，具体计算公式为：

$$\exp(-\beta \Delta A_{i,i+1}) = \frac{\left\langle \{1 + \exp[\beta(\Delta U_{i,i+1} - C)]\}^{-1} \right\rangle_i}{\left\langle \{1 + \exp[-\beta(\Delta U_{i,i+1} - C)]\}^{-1} \right\rangle_{i+1}} \exp(-\beta C)$$

$$C = \Delta A_{i,i+1} + \frac{1}{\beta} \ln \frac{n_i}{n_{i+1}}$$

15 其中， $\Delta A_{i,i+1}$ 为两相邻窗口 i 与 $i+1$ 的结合自由能差， $\Delta U_{i,i+1}$ 为基于拟合后的概率分布新生成的势能差， $\beta=1/k_B T$ ， k_B 为玻尔兹曼常数， n_i 与 n_{i+1} 分别为两相邻状态 i 与 $i+1$ 采样所包含的样本量， C 为方程组之间的隐含变量；

S54：将 $\Delta A_{annihilation}^{RL}$ 和 $\Delta A_{annihilation}^L$ 对应的计算窗口得到的 $\Delta A_{i,i+1}$ 进行加和，得到 $\Delta A_{annihilation}^{RL}$ 和 $\Delta A_{annihilation}^L$ 的数值。

7. 根据权利要求6所述的预测药-靶结合强度的绝对结合自由能微扰方法，
20 其特征在于，所述步骤S6具体包括以下步骤：

S61：对计算 $\Delta A_{restrain}^{RL}$ 的窗口中每个窗口之间的势能差进行概率分布统计，得到概率分布 $P(\Delta U)_{data}$ ；

S62：基于约束概率分布函数对概率分布 $P(\Delta U)_{data}$ 进行拟合，具体表达式为：

$$P(\Delta U)_{fit} = \sum_{i=1}^n \left(\frac{a_i}{2} \cdot b_i^{\frac{3}{2}} \cdot \Delta U^2 \cdot e^{-b_i \Delta U - (c_i \Delta U)^{n_1}} + \frac{1}{1 + \left(\frac{d_i}{\Delta U}\right)^{n_2}} \left(h_i \cdot \frac{1}{\sqrt{\pi} \mu_i} \cdot e^{-\frac{(\Delta U - \mu_i)^2}{2\sigma_i^2}} \right) \right)$$

其中, $P(\Delta U)_{fit}$ 为经过约束概率分布函数拟合后的平滑的概率分布, 其中 $a_i, b_i, c_i, d_i, h_i, \mu_i, \sigma_i$ 为第 i 个“约束概率分布函数”的参数, n 为选取拟合函数的数目, n_1 和 n_2 为两个正整数的修正因子。

- 5 8. 根据权利要求 7 所述的预测药-靶结合强度的绝对结合自由能微扰方法, 其特征在于, 所述步骤 S7 具体包括以下步骤:

S71: 根据概率分布 $P(\Delta U)_{fit}$, 重新生成每一状态的势能差 $\Delta U'$;

S72: 基于新的势能差 $\Delta U'$, 利用步骤 S5 的计算方法重新计算得到 $\Delta A_{restrain}^{RL}$ 的计算数值;

- 10 S73: 根据步骤 S24、步骤 S5 和步骤 S72 得到的 $\Delta A_{restrain}^L$ 、 $\Delta A_{annihilation}^{RL}$ 、 $\Delta A_{annihilation}^L$ 和 $\Delta A_{restrain}^{RL}$, 计算出药-靶结合自由能 $\Delta A_{binding}$ 。

9. 根据权利要求 8 所述的预测药-靶结合强度的绝对结合自由能微扰方法, 其特征在于, 在所述步骤 S73 中, 所述药-靶结合自由能 $\Delta A_{binding}$ 的计算过程具体为: $\Delta A_{binding} = \Delta A_{annihilation}^L + \Delta A_{restrain}^L - \Delta A_{annihilation}^{RL} - \Delta A_{restrain}^{RL}$ 。

15

20

25

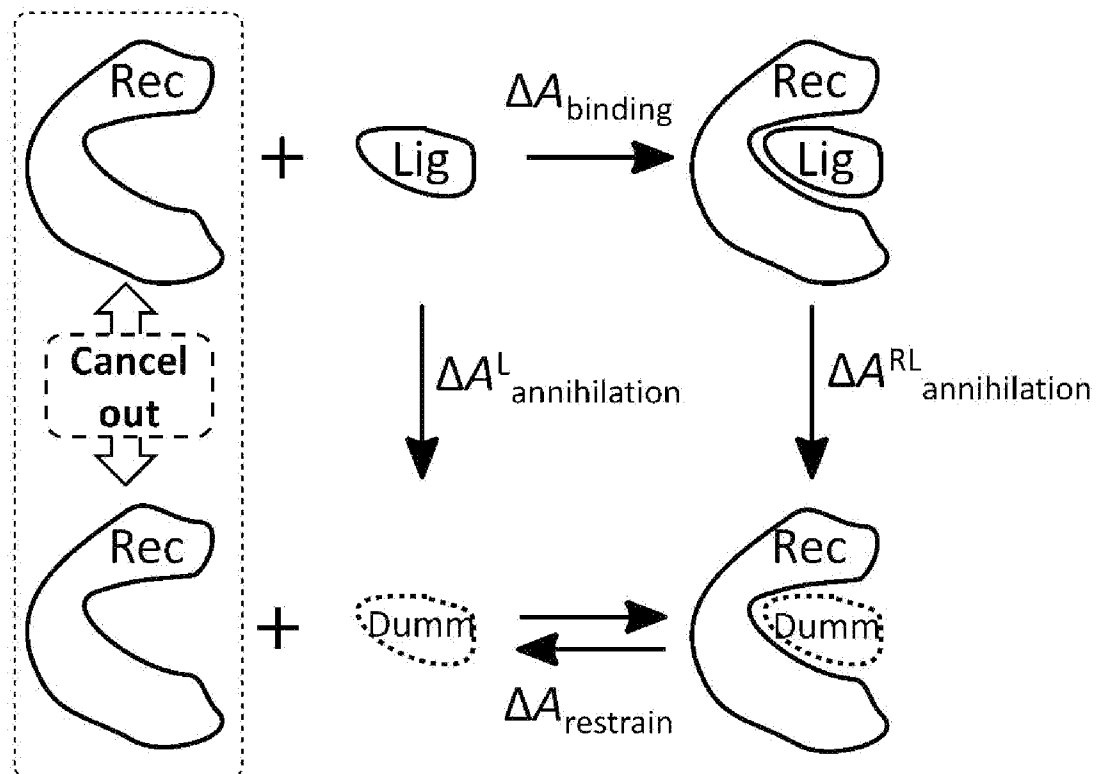


图 1

5

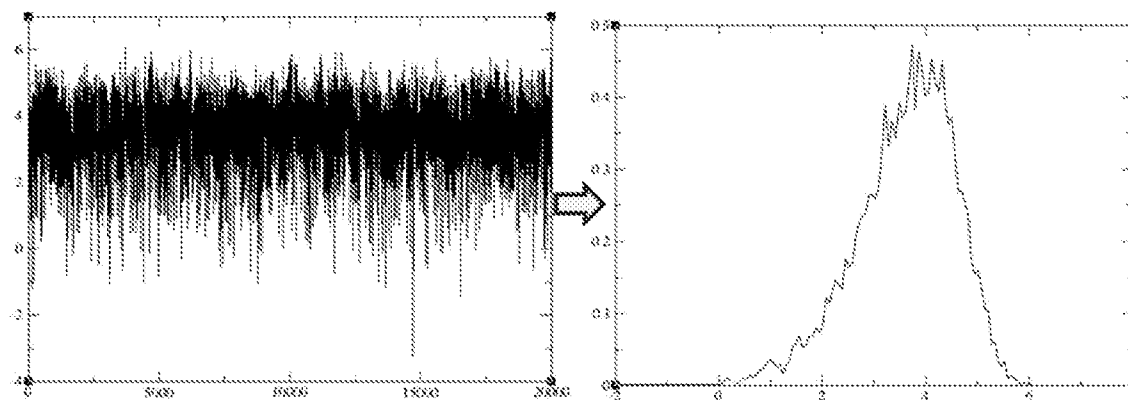


图 2

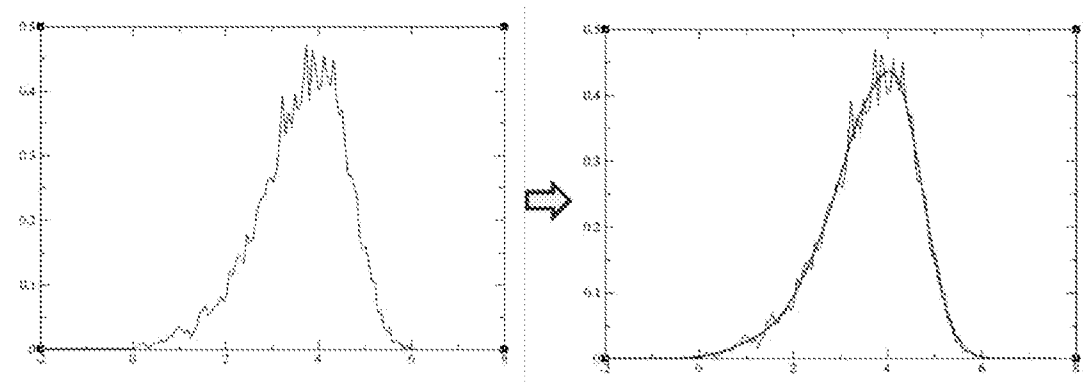


图 3

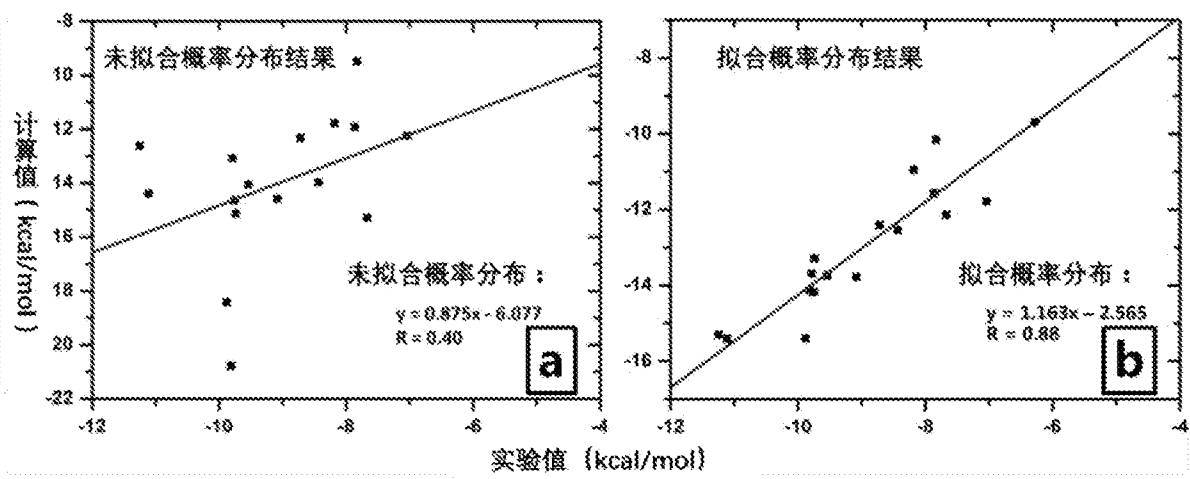
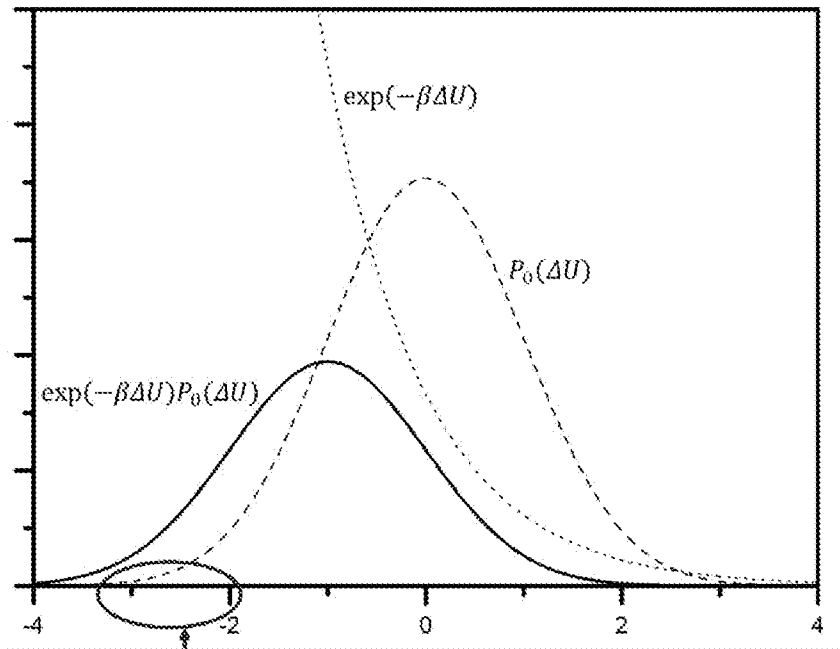


图 4



该处概率分布采样不足，但对能量计算结果有很大的影响

图 5

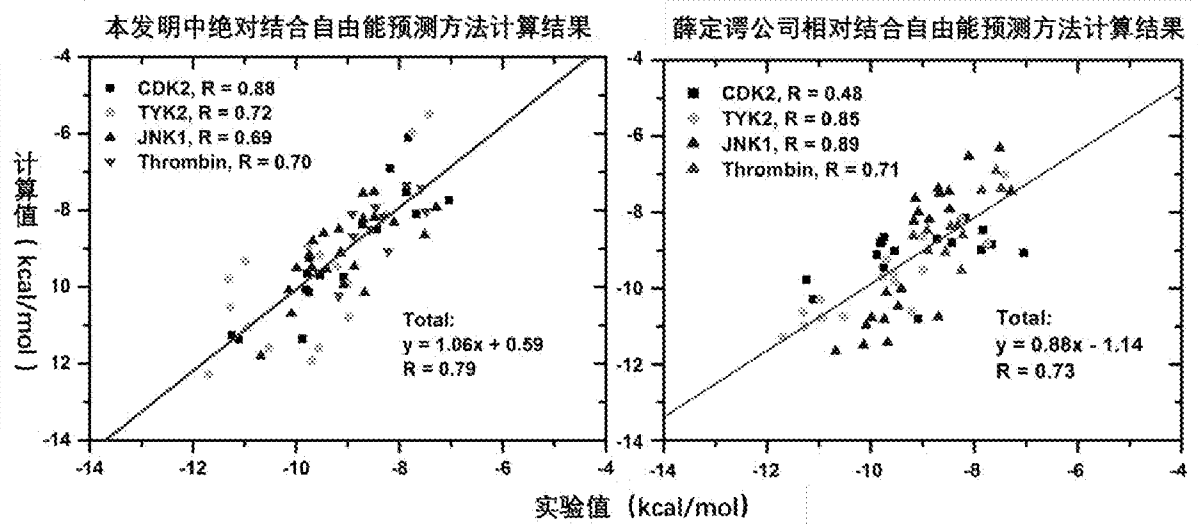


图 6

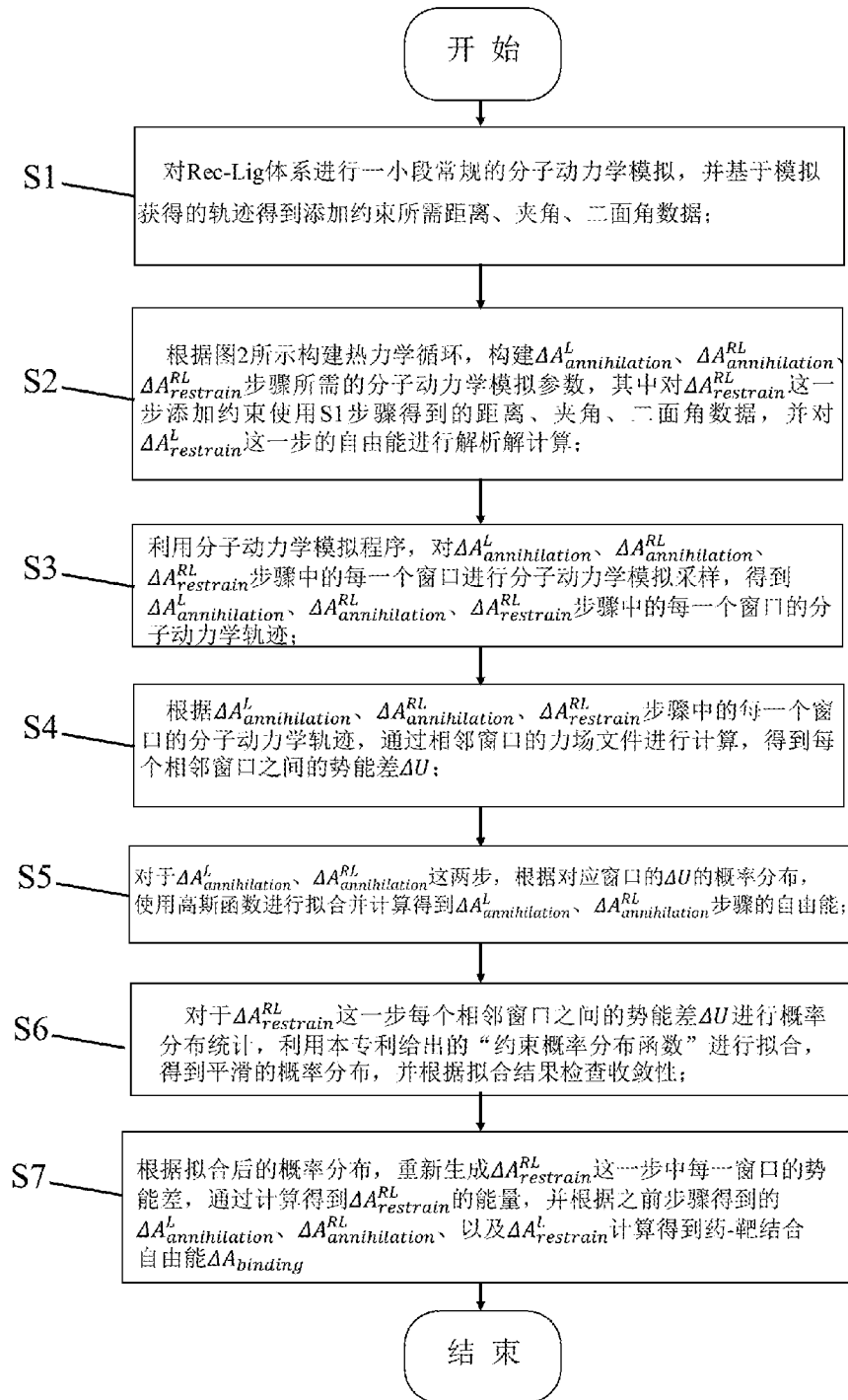


图 7

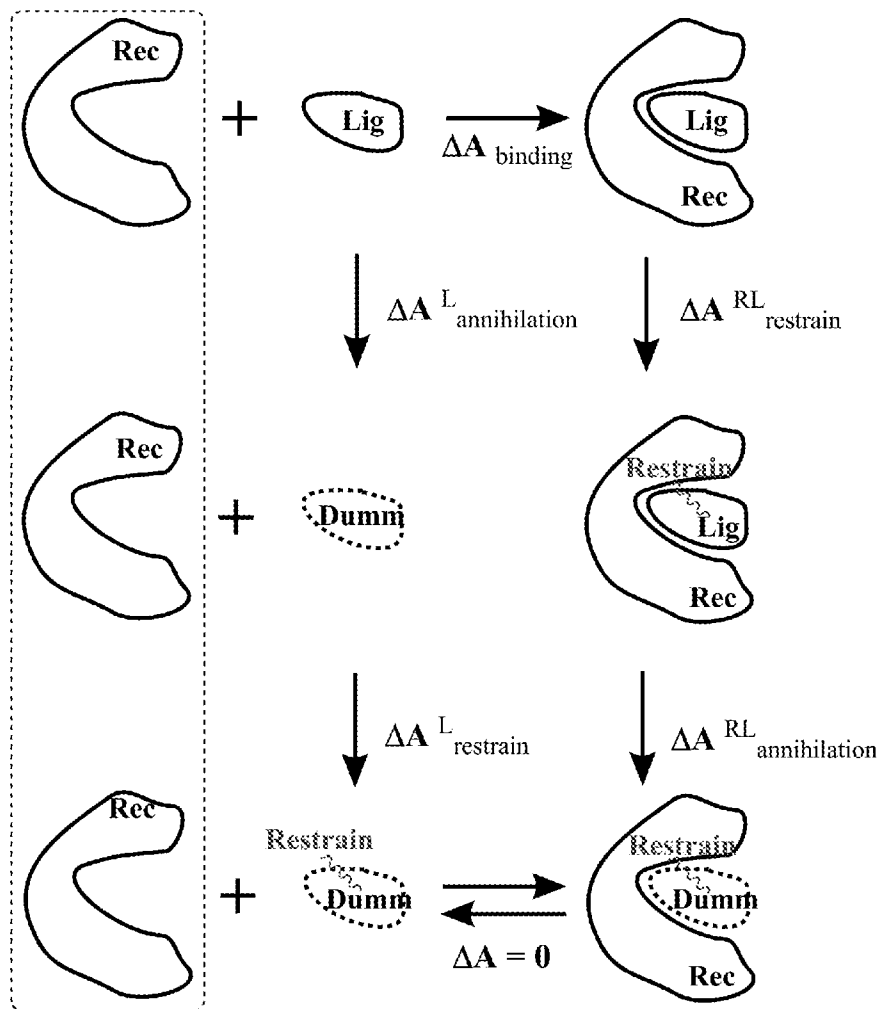


图 8

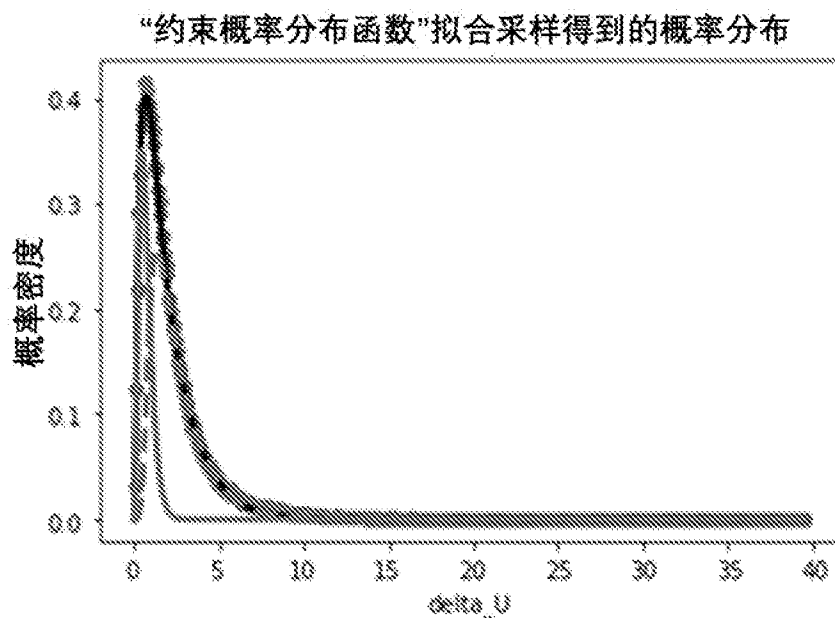


图 9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/071663

A. CLASSIFICATION OF SUBJECT MATTER

G16C 20/50(2019.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G16C

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, EPODOC, WPI, GOOGLE, CNKI: 预测, 药, 靶, 结合强度, 绝对结合, 自由能, 微扰, 受体蛋白, 配体, 湮灭, 状态, 能量, FEP, alchemical, thermodynamic, cycle, target, ligand, atom, annihilated, ABFE, hypothesized, binding

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 109859806 A (SUN YAT-SEN UNIVERSITY) 07 June 2019 (2019-06-07) claims 1-6, description, paragraphs 51-90	1-9
PA	ZHE, Li et al. "Absolute Binding Free Energy Calculation and Design of a Subnanomolar Inhibitor of Phosphodiesterase-10" <i>ACS Publications</i> , 28 January 2019 (2019-01-28), pp. 2107-2108	1-9
A	CN 102930152 A (SHANGHAI INSTITUTE OF MATERIA MEDICA, CHINESE ACADEMY OF SCIENCES et al.) 13 February 2013 (2013-02-13) entire document	1-9
A	WO 2004102159 A3 (THE PENN STATE RESEARCH FOUNDATIO) 27 January 2005 (2005-01-27) entire document	1-9

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

01 April 2020

Date of mailing of the international search report

13 April 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2020/071663

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	109859806	A	07 June 2019	None			
CN	102930152	A	13 February 2013	None			
WO	2004102159	A3	27 January 2005	US	2005027458	A1	03 February 2005

国际检索报告

国际申请号

PCT/CN2020/071663

A. 主题的分类

G16C 20/50 (2019.01) i

按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献 (标明分类系统和分类号)

G16C

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用))

CNPAT, EPDOC, WPI, GOOGLE, CNKI: 预测, 药, 靶, 结合强度, 绝对结合, 自由能, 微扰, 受体蛋白, 配体, 湮灭, 状态, 能量, FEP, alchemical, thermodynamic, cycle, target, ligand, atom, annihilated, ABFE, hypothesized, binding

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 109859806 A (中山大学) 2019年 6月 7日 (2019 - 06 - 07) 权利要求1-6, 说明书第51-90段	1-9
PA	ZHE, Li 等. "Absolute Binding Free Energy Calculation and Design of a Subnanomolar Inhibitor of Phosphodiesterase-10" ACS Publications, 2019年 1月 28日 (2019 - 01 - 28), 第2107-2108页	1-9
A	CN 102930152 A (中国科学院上海药物研究所 等) 2013年 2月 13日 (2013 - 02 - 13) 全文	1-9
A	WO 2004102159 A3 (THE PENN STATE RESEARCH FOUNDATION) 2005年 1月 27日 (2005 - 01 - 27) 全文	1-9

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2020年 4月 1日

国际检索报告邮寄日期

2020年 4月 13日

ISA/CN的名称和邮寄地址

中国国家知识产权局 (ISA/CN)

中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

王晶

电话号码 86-(10)-53961318

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/071663

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	109859806	A	2019年 6月 7日	无	
CN	102930152	A	2013年 2月 13日	无	
WO	2004102159	A3	2005年 1月 27日	US 2005027458 A1	2005年 2月 3日