基于丙氨酸扫描与相互作用熵方法的PARP1与配体的结合自由能计算

张右梓¹, 刘笑¹, 周一凡¹, 张增辉^{1,2*}

1. 精密光谱科学与技术国家重点实验室, 华东师范大学化学与分子工程学院, 上海 200062

2. NYU-ECNU计算化学联合研究中心, 上海纽约大学, 上海 200122

*通讯作者, E-mail: john.zhang@nyu.edu

收稿日期: 2017-10-10; 接受日期: 2017-12-26; 网络版发表日期: 2018-01-25

国家重点研发专项(编号: 2016YFA0501700)、国家自然科学基金(编号: 21433004, 91753103)、上海市普陀区创新项目(编号: 2014-A-02)和上海教委重大项目(编号: 201701070005E00020)资助

摘要 PARP1蛋白是真核细胞DNA损伤修复的关键因子,也是近年来肿瘤治疗的新热靶标之一. 本文运用本课题组发展的基于相互作用熵方法并结合MM/GBSA方法对蛋白-配体进行丙氨酸扫描,计算了10个抑制剂小分子与PARP1蛋白及与PARP1蛋白各残基间的相互作用结合自由能并由此获得总的结合自由能. 计算结果表明,相比传统MM/GBSA方法,新的计算方法可以显著提高蛋白-配体结合自由能计算的精度,缩小与实验值的误差.

关键词 PARP1, 抑制剂, MM/GBSA, 相互作用熵, 结合自由能, 分子动力学

1 引言

多聚腺苷二磷酸-核糖聚合酶(poly(ADP-ribose) polymerases, PARPs)广泛存在于真核细胞中,是一类重要的多功能蛋白质翻译后修饰酶. 通过对各种核蛋白进行修饰参与如DNA损伤修复、细胞凋亡等一系列细胞生理活动的调节^[1]. 依照其催化区域的序列同源性进行判定,此家族共有17位成员. 但研究表明,只有包括PARP1、PARP2、Tankyrase-1和Tankyrase-2等在内的几种具有上述介导细胞DNA损伤修复的作用^[2,3]. PARP1是目前研究比较深入的亚型,由N端的DNA结合区域、中心自修饰区域以及C端的催化区域3部分组成. 其主要参与针对单链DNA(ssDNA)断裂的

细胞信号转导和DNA碱基切除修复(BER)过程. 由于BER是真核生物最重要的DNA修复方式,对PARP1进行抑制会导致细胞内大量ssDNA损伤因为无法被及时修复而大量堆积,因此临床上最初将PARP抑制剂作为放疗以及烷化剂和铂类药物化疗的致敏剂联合使用^[4-6]. 2005年, Farmer等^[7]与Bryant等^[8]发现携带BRCA1/2任意一种缺陷突变的肿瘤细胞对PARP抑制剂存在特殊的敏感性,达到携带野生型BRCA基因的肿瘤细胞的1000倍,即PARP抑制剂可有选择性地杀伤具有BRCA1/2缺陷的肿瘤细胞而对正常体细胞不产生影响,这一发现推进了PARP抑制剂作为针对BRCA1/2突变型肿瘤单一治疗药物的研究进程^[9-11]. 目前已获美国食品药品监督管理局(Food and Drug Administra-

引用格式: Zhang Y, Liu X, Zhou Y, Zhang J Z H. Calculation of PARP1-ligand binding free energy using alanine scanning and interaction entropy method. *Sci Sin Chim*, 2018, 48: 204–209, doi: [10.1360/N032017-00161](https://doi.org/10.1360/N032017-00161)

tion, FDA)批准上市的PARP抑制剂药物包括奥拉帕尼(olaparib)、瑞卡帕布(rucaparib)与尼拉帕尼(niraparib),另有多项临床试验在进行中^[12~14], PARP1已成为治疗癌症的新热门靶点.

分子动力学(molecular dynamics, MD)模拟和自由能计算已经成为目前研究蛋白与小分子配体间相互作用的重要工具,比较常用的计算结合自由能方法是基于经验方程的MM/PB(GB)SA^[15,16].而基于分子动力学模拟的丙氨酸变异(alanine-scanning mutagenesis)方法可用于定量计算单个残基对于结合自由能的贡献^[17,18].近年来本课题组将熵变计算引入丙氨酸变异,发展了一种基于相互作用熵(interaction entropy, IE)和MM/GBSA结合计算特定残基对于结合能贡献的新方法^[19],其中熵变由相互作用熵的方法得到^[20].本文采用此方法计算了PARP1蛋白与10个抑制剂小分子的结合自由能及抑制剂与PARP1蛋白各残基间的相互作用,结果表明,相比只运用MM/GBSA方法计算,引入熵变后的结合自由能预测的精度显著提高,与实验值吻合更好.

2 理论方法

2.1 单个残基的结合自由能

单个残基对蛋白-配体结合能的贡献通常会通过丙氨酸变异前后的结合能差来得到,公式可表达为

$$\Delta\Delta G_{\text{bind}}^{x \rightarrow a} = \Delta G_{\text{bind}}^a - \Delta G_{\text{bind}}^x, \quad (1)$$

其中

$$\begin{aligned} \Delta G_{\text{bind}}^x &= \Delta G(A(x) - B) \\ &= \Delta G_{\text{gas}}(A(x) - B) + \Delta G_{\text{sol}}(A(x) - B), \end{aligned} \quad (2)$$

同样地,

$$\begin{aligned} \Delta G_{\text{bind}}^a &= \Delta G(A(a) - B) \\ &= \Delta G_{\text{gas}}(A(a) - B) + \Delta G_{\text{sol}}(A(a) - B), \end{aligned} \quad (3)$$

其中A与B分别代表复合物中的蛋白与配体, x 代表位于蛋白A与其配体B界面附近的A的某残基, a 则代表原来的 x 残基替换为丙氨酸. ΔG_{bind}^x 表示野生型的蛋白配体复合物的结合自由能, ΔG_{bind}^a 则为变异后的结合自由能. 将上述公式展开可得到

$$\begin{aligned} \Delta\Delta G_{\text{bind}}^{x \rightarrow a} &= \Delta G_{\text{gas}}(A(a) - B) + \Delta G_{\text{sol}}(A(a) - B) \\ &\quad - [\Delta G_{\text{gas}}(A(x) - B) + \Delta G_{\text{sol}}(A(x) - B)] \\ &= \Delta G_{\text{gas}}(A(a) - B) - \Delta G_{\text{gas}}(A(x) - B) \\ &\quad + [\Delta G_{\text{sol}}(A(a) - B) - \Delta G_{\text{sol}}(A(x) - B)] \\ &= \Delta\Delta G_{\text{gas}}^{x \rightarrow a} + \Delta\Delta G_{\text{sol}}^{x \rightarrow a}, \end{aligned} \quad (4)$$

可作如下简化:

$$\begin{aligned} \Delta\Delta G_{\text{gas}}^{x \rightarrow a} &= \Delta G_{\text{gas}}(A(a) - B) - \Delta G_{\text{gas}}(A(x) - B) \\ &\approx \Delta G_{\text{gas}}(a - B) - \Delta G_{\text{gas}}(x - B). \end{aligned} \quad (5)$$

2.2 气相部分 $\Delta\Delta G_{\text{gas}}^{x \rightarrow a}$ 的计算

$$\Delta G_{\text{gas}}(x - B) = \langle E_{x-B}^{\text{int}} \rangle + K T \ln \langle e^{\beta \Delta E_{x-B}^{\text{int}}} \rangle. \quad (6)$$

定义相互作用熵 IE ^[19]:

$$IE = K T \ln \langle e^{\beta E_{x-B}^{\text{int}}} \rangle = -T \Delta S. \quad (7)$$

相互作用熵中的 $\langle e^{\beta \Delta E^{\text{int}}} \rangle$ 可通过数值积分得到.

$$\langle e^{\beta \Delta E^{\text{int}}} \rangle = \frac{1}{N} \sum_{i=1}^N e^{\beta \Delta E^{\text{int}}(t_i)}, \quad (8)$$

其中, $\Delta E^{\text{int}} = E^{\text{int}} - \langle E^{\text{int}} \rangle$, 通过式(9)得到平均相互作用能,

$$\langle E^{\text{int}} \rangle = \frac{1}{T} \int_0^T E^{\text{int}}(t) dt = \frac{1}{N} \sum_{i=1}^N E^{\text{int}}(t_i). \quad (9)$$

2.3 液相部分 $\Delta\Delta G_{\text{sol}}^{x \rightarrow a}$ 的计算

液相部分的 $\Delta\Delta G_{\text{sol}}^{x \rightarrow a}$ 通过MM/GBSA计算得到. 将去溶剂化能分成静电与非静电两部分:

$$\Delta G_{\text{sol}} = \Delta G_{\text{gb}} + \Delta G_{\text{np}}. \quad (10)$$

第一项去溶剂化能的静电项 ΔG_{gb} 通过GB模型得到; 第二项 ΔG_{np} 为去溶剂化能的非静电项, 一般认为其正比于溶剂可及表面积(solvent-accessible surface area, SASA).

$$\Delta G_{\text{np}} = \gamma \text{SASA} + \beta, \quad (11)$$

其中 γ 项与 β 项在本方法中分别取经验值0.00542 kcal/(mol Å²)与0.92 kcal/mol^[21,22].

需要指出的是, 在之前的研究中, 我们分别计算了非极性、极性与带电氨基酸内部介电常数从1到10范

围的结合能差并与实验值进行对比, 结果表明, 当使用GB模型计算溶剂化能时, 对非极性残基介电常数取1, 对极性残基介电常数取3, 对带正电或负电的残基介电常数取5时与实验值误差达到最小^[19]. 本文在计算时亦沿用这一规则对带电、极性和非极性氨基酸分别取5、3和1的介电常数.

2.4 蛋白-配体结合自由能 ΔG 的计算

以下运用从丙氨酸扫描计算得到的单个氨基酸的结合自由能来近似获得蛋白-配体的总的结合自由能,

$$\Delta G_{\text{bind}} = \sum_x \Delta \Delta G_{\text{bind}}^{x \rightarrow a}. \quad (12)$$

式(12)里的求和包括所有对蛋白-配体结合自由能有贡献的氨基酸. 由于Alanine-scanning方法计算熵变焓变均定义为变异后减变异前的值, 因此得到的差值需要变个符号来给出总的结合自由能, 所以在方程(12)中加上负号即为计算出的结合自由能值. 用以上组合方法得到的蛋白-配体结合自由能(式(12))的方法简称为AS-IE (alanine scanning with interaction entropy).

2.5 分子动力学模拟

选取人类PARP-1蛋白与不同抑制剂的复合物结构用做MD的初始构象. 10个蛋白-配体复合物的PDBID分别为: 2RCW、2RD6、3L3M、3GJW、3GN7、4L6S、4RV6、4R5W、4ZZZ、5A00. 本文分子动力学模拟和MM/GBSA计算均采用AMBER14软件包, MD利用pmemd.cuda进行, 小分子力场选用gaff力场, 蛋白质力场选用ff14SB力场. 10个小分子抑制剂的电荷处理和参数生成都使用Amber Tools中Antechamber程序, 并利用Tleap模块对复合物模型进行处理.

在进行MD之前对体系进行优化、升温和一段时间的模拟平衡. 优化首先约束溶质, 只优化小分子和水; 然后无约束地进行整体优化. 优化完成后体系在NVT系综中被缓慢地从0 K加热到300 K, 用时1 ns; 随后在NPT系综中进行共计2 ns的平衡模拟, 温度为300 K, 压强为101 kPa; 最后再在NPT系综中跑5 ns的MD, 时间步长为2.0 fs, 每100 fs保存一帧. 在分子动力学模拟过程中采用SHAKE^[23]算法控制含氢化学键的伸缩, PME^[24]方法计算长程静电相互作用, 非成键相互

作用的截断值为10 Å.

2.6 MM/GBSA计算焓变

结合自由能焓和熵贡献的计算分别由AMBER14中的perl脚本mm_pbsa.pl和nmode模块完成, GB模型的“igb”设置为2, 从5 ns的轨迹中均匀的取50帧用于MM/GBSA的计算. 首先计算野生型的蛋白配体复合物的结合能, 然后计算变异后的蛋白配体复合物的结合能. 变异后的结合能减去变异前的结合能即为该残基由于丙氨酸变异引起的焓变.

2.7 相互作用熵计算熵变

从5 ns的轨迹中均匀地取1000帧, 首先计算残基(变异前以及变异后的丙氨酸)与配体的气相下的结合能, 然后用式(7)计算熵变. 变异后的熵变减去变异前的熵变即为该残基由于丙氨酸变异引起的熵变. 完整过程如图1所示, 概括即为: 从PDBbank下载并检查初始构象; 用Amber软件进行分子动力学模拟; 对轨迹进行丙氨酸变异; 用MM/GBSA方法计算焓变; 用IE方法计算熵变; 计算变异前后的结合自由能差^[19].

3 结果与讨论

3.1 体系的稳定性与自由能计算

为了了解体系动力学过程的稳定性以及结合自由能取样是否合理, 我们以初始构象作为参考对象, 计算了PARP1蛋白主链的均方根偏差(root-mean-square displacement, RMSD), 并绘制了RMSD随时间变化的图

图1 AS-IE方法流程图

像(图2). 如图2所示, 经过5 ns的MD模拟, 10个体系的RMSD基本都稳定在2 Å左右, 说明所有体系都处于相对平衡的状态, 因此我们认为对这段轨迹采样计算蛋白配体结合自由能是合理的.

表1列出了10个抑制剂与PARP1蛋白的结合自由能, 包括MM/GBSA与Alanine-scanning方法的预测值与实验值. 结合自由能的实验值由下列公式得到: $\Delta G_{\text{bind}} = RT \ln K_i$, 其中 R 为理想气体常数, T 为反应温度, 本文取298.15 K即室温25°C; K_i 为抑制常数, K_d 为解离常数, 在无其他底物参与竞争靶标时, 抑制剂的 $K_i = K_d$. 二者都可代入公式中计算结合自由能^[25].

由表1可看出, 10个PARP1抑制剂的结合自由能范围为-8.46~-12.07 kcal/mol, 即4RV6结构中的配体Rucaparib结合蛋白相对最牢固. 而除4R5W、4ZZZ两个体系结合自由能大于-10 kcal/mol外, 其他8个体系的自由能浮动范围并不大, 且多在-11 kcal/mol左右.

3.2 两种方法计算结合自由能的比较

为了便于验证MM/GBSA与AS-IE两种方法计算得到的自由能与实验值是否存在显著差异, 本文使用了以下3种误差评估方法:

平均绝对误差(MAE)定义为

$$\text{MAE} = \langle |\Delta\Delta G_{\text{bind}}(\text{cal}) - \Delta\Delta G_{\text{bind}}(\text{exp})| \rangle, \quad (13)$$

均方根误差(RMSE)定义为

$$\text{RMSE} = \sqrt{\langle ((\Delta\Delta G_{\text{bind}}(\text{cal}) - \Delta\Delta G_{\text{bind}}(\text{exp}))^2) \rangle}, \quad (14)$$

平均符号偏差(MSD)定义为

图2 MD模拟中10个体系蛋白主链原子的RMSD (网络版彩图)

$$\text{MSD} = \langle (\Delta\Delta G_{\text{bind}}(\text{cal}) - \Delta\Delta G_{\text{bind}}(\text{exp})) \rangle \quad (15)$$

表1中列出了每种方法与实验值误差的具体数值. 可以看出, 基本对于每个体系(仅除4ZZZ外) AS-IE方法相对于MM/GBSA方法计算值与实验值的误差都有所降低, 如5A00体系, MM/GBSA计算值与实验值相差10.77 kcal/mol, 而AS-IE计算值仅与实验值差2.85 kcal/mol. 为了给出更直观的评价, 我们绘制了两种方法与实验值上述3种误差的条形图(图3), 其中MM/GBSA方

表1 MM/GBSA与AS-IE方法计算所得到的结合自由能以及与实验结果的比较

PDBID ^{a)}	ΔG_{bind} (kcal/mol)		
	Expt ^{b)}	MM/GBSA ^{c)}	AS-IE ^{d)}
5A00	-10.26	-21.03	-13.11
4R5W	-8.46	-5.17	-11.44
4ZZZ	-9.63	-8.94	-12.84
4RV6	-12.07	-18.27	-11.67
4L6S	-11.62	-19.78	-12.67
3L3M	-11.21	-20.52	-13.83
3GJW	-11.32	-17.59	-12.23
3GN7	-11.32	-6.39	-10.45
2RCW	-11.04	-12.8	-12.62
2RD6	-11.32	-14.88	-8.01
	MAE	5.5	2.0
与实验的误差	RMSE	6.3	2.3
	MSE	-3.7	-1.1

a) 除了4R5W与4ZZZ外, 其余体系的配体都带电; b) K_i/K_d 值经公式得到的结合自由能实验值; c) 标准MM/GBSA计算得到的结合自由能; d) AS-IE计算得到的结合自由能

图3 两种方法计算值与实验值的3种误差的平均值(网络版彩图)

法计算值计算全部10个体系与实验值的平均绝对误差为5.49 kcal/mol, 均方根误差为6.31 kcal/mol, 平均符号偏差为-3.71 kcal/mol; 而AS-IE方法计算全部10个体系与实验值的平均绝对误差为1.98 kcal/mol, 均方根误差为2.25 kcal/mol, 平均符号偏差为-1.1 kcal/mol. 这些结果表明, 通过计算相互作用熵的方式将熵引入可以降低计算值与实验值的误差. 这些误差可以更直接地显示在图3中.

4 结论

本文对10个抑制剂小分子与PARP1蛋白的复合物

分别进行了5 ns的分子动力学模拟, 用MM/GBSA方法和本课题组新发展的基于相互作用熵的AS-IE方法计算了PARP1蛋白与抑制剂的结合自由能, 新方法利用MM/GBSA方法计算焓变, 利用相互作用熵的方法计算熵变, 并通过丙氨酸扫描来计算单个氨基酸对结合自由能的贡献. 计算结果显示, MM/GBSA方法计算值与实验值的平均绝对误差为5.49 kcal/mol, 而AS-IE方法计算值与实验值的平均绝对误差为2.0 kcal/mol. 结论为用相互作用熵的方法将熵引入丙氨酸变异可有效降低预测值与实验值间的误差. 本文同时也为基于PARP1靶点的药物设计和小分子数据库虚拟筛选奠定了基础.

致谢 感谢华东师范大学超级计算机中心给予的帮助.

参考文献

- Miwa M, Masutani M. *Cancer Sci*, 2007, 98: 1528–1535
- Rouleau M, Patel A, Hendzel MJ, Kaufmann SH, Poirier GG. *Nat Rev Cancer*, 2010, 10: 293–301
- Schreiber V, Dantzer F, Ame JC, de Murcia G. *Nat Rev Mol Cell Biol*, 2006, 7: 517–528
- Underhill C, Toulmonde M, Bonnefoi H. *Ann Oncol*, 2011, 22: 268–279
- Ferraris D, Ficco RP, Pahutski T, Lautar S, Huang S, Zhang J, Kalish V. *Bioorg Med Chem Lett*, 2003, 13: 2513–2518
- Herceg Z, Wang ZQ. *Mutat Res*, 2001, 477: 97–110
- Farmer H, McCabe N, Lord CJ, Tutt ANJ, Johnson DA, Richardson TB, Santarosa M, Dillon KJ, Hickson I, Knights C, Martin NMB, Jackson SP, Smith GCM, Ashworth A. *Nature*, 2005, 434: 917–921
- Bryant HE, Schultz N, Thomas HD, Parker KM, Flower D, Lopez E, Kyle S, Meuth M, Curtin NJ, Helleday T. *Nature*, 2005, 434: 913–917
- Brody LC. *N Engl J Med*, 2005, 353: 949–950
- Iglehart JD, Silver DP. *N Engl J Med*, 2009, 361: 189–191
- Fisher AEO, Hochegger H, Takeda S, Caldecott KW. *Mol Cellul Biol*, 2007, 27: 5597–5605
- Curtin NJ. *Expert Rev Mol Med*, 2005, 7: 1–20
- Fong PC, Boss DS, Yap TA, Tutt A, Wu P, Mergui-Roelvink M, Mortimer P, Swaisland H, Lau A, O'Connor MJ, Ashworth A, Carmichael J, Kaye SB, Schellens JHM, de Bono JS. *N Engl J Med*, 2009, 361: 123–134
- Ang YLE, Tan DSP. *Curr Probl Cancer*, 2017, 41: 273–286
- Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE. *Acc Chem Res*, 2000, 33: 889–897
- Brown SP, Muchmore SW. *J Chem Inf Model*, 2006, 46: 999–1005
- Moreira IS, Fernandes PA, Ramos MJ. *J Comput Chem*, 2007, 28: 644–654
- Martins SA, Perez MAS, Moreira IS, Sousa SF, Ramos MJ, Fernandes PA. *J Chem Theor Comput*, 2013, 9: 1311–1319
- Yan Y, Yang M, Ji CG, Zhang JZH. *J Chem Inf Model*, 2017, 57: 1112–1122
- Duan L, Liu X, Zhang JZH. *J Am Chem Soc*, 2016, 138: 5722–5728
- Gohlke H, Case DA. *J Comput Chem*, 2004, 25: 238–250
- Case D A, Babin V, Berryman J T, Betz R M, Cai Q, Cerutti DS, Cheatham TE, Darden TA, Duke R E, Gohlke H, Goetz A W, Gusarov S, Homeyer N, Janowski P, Kaus J, Kolossváry I, Lee T S, LeGrand S, Luchko T, Luo R, Madej B, Merz KM, Paesani F, Roe DR, Roitberg A, Sagui C, Salomon-Ferrer R, Seabra G, Simmerling C L, Smith W, Swails J, Walker, Wang J, Wolf RM, Wu X, Kollman PA. Amber 14 Reference

Manual, 2014. 613

23 Ryckaert JP, Ciccotti G, Berendsen HJC. *J Comput Phys*, 1977, 23: 327–341

24 Darden T, York D, Pedersen L. *J Chem Phys*, 1993, 98: 10089–10092

25 Schneider G, Baringhaus KH. *Molecular Design: Concepts and Applications*. Weinheim: Wiley-VCH, 2008

Calculation of PARP1-ligand binding free energy using alanine scanning and interaction entropy method

Youzi Zhang¹, Xiao Liu¹, Yifan Zhou¹, John Z.H. Zhang^{1,2*}

¹ State Key Laboratory of Precision Spectroscopy, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai 200062, China

² NYU-ECNU Center For Computational Chemistry, New York University Shanghai, Shanghai 200122, China

*Corresponding author (email: john.zhang@nyu.edu)

Abstract: Poly(ADP-ribose) polymerase 1 (PARP1) is a novel target of cancer medication for its important role in DNA repair. In this article, we use interaction entropy (IE) approach combined with the MM/GBSA method to compute residue-specific protein-ligand binding free energies for ten ligands and these residue-specific binding free energies are combined to produce total protein-ligand binding free energy. Our calculated binding free energies are in much better agreement with the corresponding experimental data than those using standard MM/GBSA approach.

Keywords: PARP1, inhibitor, MM/GBSA, interaction entropy, binding free energies, MD simulation

doi: [10.1360/N032017-00161](https://doi.org/10.1360/N032017-00161)