

从对接结构中挑选近天然构象的新方法*

于永辉

李春华

卢本卓

陈慰祖

王存新

(北京工业大学生命科学与生物工程学院, 北京 100022)

摘要 分析了包含静电能 (ΔE_{ele})、去水化自由能 (ΔG_{ACE}) 以及范德华能 (ΔE_{vdw}) 的打分函数在蛋白质-蛋白质对接中评价近天然构象的能力。对 17 种蛋白质复合物对接体系进行打分的结果表明, 包含范德华能的打分函数 ($\Delta E_{\text{ele}} + \Delta G_{\text{ACE}} + \Delta E_{\text{vdw}}$) 比通常的打分函数 (ΔE_{ele} 、 ΔG_{ACE} 、 $\Delta E_{\text{ele}} + \Delta G_{\text{ACE}}$ 、 $\Delta E_{\text{ele}} + \Delta E_{\text{vdw}}$ 、 $\Delta G_{\text{ACE}} + \Delta E_{\text{vdw}}$) 具有更好的区分近天然构象的能力。进一步的研究表明, 优化 (EM) 对接体系后再进行打分, 上面几种打分函数对对接结构的评价效果都有不同程度的改善, 其中打分函数 ($\Delta E_{\text{ele}} + \Delta G_{\text{ACE}} + \Delta E_{\text{vdw}}$) 有更明显的改善。为了进一步确定候选结构中的近天然构象, 以一种蛋白质复合物为例, 对候选结构进行分子动力学 (MD) 模拟, 根据 MD 轨迹中构象相对于初始构象的平方平均偏差 (MSD) 随时间的变化来辅助打分函数排除错误构象, 得到了较好的结果。

关键词: 打分函数, 范德华能, 能量优化 (EM), 分子动力学 (MD) 模拟, 近天然构象

中图分类号: Q617

分子对接是研究蛋白质-蛋白质相互作用与识别的重要方法之一。该方法包括两个重要因素: 一个有效的搜索程序和一个好的打分函数。对于如何从众多的对接构象中筛选出近天然构象, 一般分为两个步骤, 过滤和打分。构造打分函数来有效地区分正确与错误两种结合模式一直是人们研究的难点^[1]。目前大部分对接算法的过滤是基于几何互补性^[2-3]或简化的分子势能函数^[4-5], 或者是这二者的组合^[6]。然而在很多情况下, 简化的分子势能函数不能对近天然和错误结合模式作出很好的区分^[7], 需要在打分中进一步判断。

本文采用软对接和组合过滤算法^[8]产生复合物结合模式来进行打分研究, 考查了静电能 (ΔE_{ele})、去水化自由能 (ΔG_{ACE})^[9]和范德华能 (ΔE_{vdw}) 作为打分函数评价近天然构象的能力, 目的在于找到更加有效的打分函数。以不同类型的 17 种蛋白质复合物体系为例, 比较了采用六种打分函数 ($\Delta E_{\text{ele}} + \Delta G_{\text{ACE}} + \Delta E_{\text{vdw}}$ 、 ΔE_{ele} 、 ΔG_{ACE} 、 $\Delta E_{\text{ele}} + \Delta G_{\text{ACE}}$ 、 $\Delta E_{\text{ele}} + \Delta E_{\text{vdw}}$ 和 $\Delta G_{\text{ACE}} + \Delta E_{\text{vdw}}$) 所得到的结果, 发现打分函数 $\Delta E_{\text{ele}} + \Delta G_{\text{ACE}} + \Delta E_{\text{vdw}}$ 在挑选近天然构象方面有比较明显的优势。

然而, 只靠单纯的能量打分函数来挑选近天然构象是有困难的。因为蛋白质-蛋白质对接计算量较大, 目前还没有有效的方法充分考虑分子的柔性, 所以某些对接结构中会存在部分原子交叠, 导致构象能很大, 从而使得打分函数不能对这些结构作出合理评价。基于这一点, 我们以 5 种复合物为例, 首先优化对接结构, 消除原子交叠, 然后再进行打分, 以改善评价效果。

另外, 作为一种辅助手段, 在打分排序后, 本文还以一种蛋白质复合物为例进行了 MD 模拟, 根据 MD 轨迹构象相对于初始构象的平方平均偏差 (MSD) 随时间的变化来进一步排除不合理结构。结果表明, 这种辅助手段是有效的。

1 方法与研究体系

1.1 方法

不考虑分子柔性时, 蛋白质与蛋白质结合过程中自由能的变化主要为

$$\Delta G = \Delta E_{\text{ele}} + \Delta G_{\text{solv}} - T\Delta S_{\text{sc}} + \Delta E_{\text{vdw}} \quad (1)$$

其中 ΔE_{ele} 、 ΔG_{solv} 、 ΔS_{sc} 和 ΔE_{vdw} 分别表示结合过程中静电能、水化自由能、侧链构型熵及范德华能的

表 1 用于评价打分函数的 17 种蛋白质-蛋白质复合物

Table 1 The selected 17 protein-protein complexes used to test the scoring functions

Dock name	Description	PDB code
Enzyme / inhibitor		
1CHOFF	Alpha-chymotrypsin/Ovomucoid 3 rd Domain	5cha, 2ovo
2SICXX	Subtilisin BPN/Subtilisin inhibitor	2sic
1TGSXX	Trypsinogen/Pancreatic secretory trypsin inhibitor	1tgs
1BRSFF	Barnase/Barstar	1a2p, 1a19
2PTCXX	Beta-trypsin/Pancreatic trypsin inhibitor	2ptc
2SNIFF	Subtilisin Novo/Chymotrypsin inhibitor 2	1sup, 2ci2
Antibody / antigen		
1AHWFF	Antibody Fab 5G9/Tissue factor	1fgn, 1boy
1BVKFF	Antibody Hulys11 Fv/Lysozyme	1bvl, 3lzt
1BQLXF	Hyhel-5 Fab/Lysozyme	1bql, 1dkj
1WEJFF	IgG1 E8 Fab fragment/Cytochrome C	1qbl, 1hrc
1NMBXF	Fab NC10/Neuraminidase	1nmb, 7nn9
1KXQXF	Antibody Vhh Fragment/Amylase	1kxq, 1pif
1LQYXX	Antibiotic Actinonin/Peptide Deformylase	1lqy
Others		
1A0OFX	Che A/Che Y	1chn, 1a0o
1MDAFF	Methylamine dehydrogenase/Amicyanin	2bbk, 1aan
4INSXX	Pig insulin dimer	4ins
1FINFF	CDK2 cyclin-dependant kinase 2/Cyclin	1hcl, 1vin

变化, T 为绝对温度. 本文采用介电距离依赖库仑势模型计算静电能, 电荷参数来自 CHARMM 19 力场^[10]:

$$\Delta E_{\text{ele}} = \frac{q_i q_j}{4\pi\epsilon_0\epsilon(r_{ij} + c)} \quad (2)$$

其中, ϵ_0 为真空中的介电常数, 相对介电常数 $\epsilon = 4r_{ij}$, r_{ij} 为原子 i 和 j 之间的距离, q_i 、 q_j 分别为原子 i 和 j 的电荷. 用以下软心 Lennard-Jones 势^[11] 来计算范德华相互作用:

$$\Delta E_{\text{vdw}} = \frac{C_{12}}{(r_{ij} + c)^{12}} - \frac{C_6}{(r_{ij} + c)^6} \quad (3)$$

方程(2)和(3)中的常数 c 取 0.1 nm. 式(3)中 C_{12} 、 C_6 分别为范德华相互作用势排斥项和吸引项的势参数. 对去水化自由能的计算, 我们采用原子接触势模型 (ACE)^[9]. 此模型包含了侧链熵效应, 因而在很多近似计算中, 以 ΔG_{ACE} 取代 ($\Delta G_{\text{sol}} - T\Delta S_{\text{sc}}$) 项对结合自由能的贡献. 所以本文采用的打分函数是

$$\text{Score} = \Delta E_{\text{ele}} + \Delta G_{\text{ACE}} + \Delta E_{\text{vdw}} \quad (4)$$

为了改善打分效果, 以 5 种蛋白质复合物为例, 用最陡下降法对其所有对接结构进行 1000 步能量优化后再进行打分排序. 另外, 作为挑选近天然结构的辅助手段, 在打分排序后, 以 1 种蛋白质复合物为例对其候选结构进行了 50 ps 的 MD 模拟, 根据

模拟轨迹构象相对于初始构象的 MSD 随时间的变化排除不合理结构. 能量优化和 MD 模拟均采用 GROMACS 程序^[12].

1.2 研究体系

本文以 17 个蛋白质-蛋白质复合物对接体系为研究对象, 它们是用本研究小组在 Janin 等人的刚性对接程序 DOCK^[13] 基础上发展的软对接程序^[8], 经过构象搜索和过滤后产生的. 每一个复合物对接体系有 800~1000 个结合模式. 这些复合物包含 3 种类型 (见表 1): 酶/抑制剂类 6 个、抗原/抗体类 7 个 [其中 1KXQXF 和 1LQYXX 分别是 CAPRI Round 2^[14] (<http://capri.ebi.ac.uk>) 中的测试靶结构 06、07], 另外还有 4 个其它类型的复合物. 若对接的配体或受体选自非复合态的晶体结构 (自由态), 则在 PDB 代码后加字母 F 表示, 若配体或受体的结构数据选自复合物结构, 则在 PDB 代码后加字母 X 表示.

2 结果与讨论

2.1 六种打分函数的排序结果

表 2 总结了用 6 种不同的打分函数 ΔE_{ele} 、 ΔG_{ACE} 、 $\Delta E_{\text{ele}} + \Delta G_{\text{ACE}}$ 、 $\Delta E_{\text{ele}} + \Delta E_{\text{vdw}}$ 、 $\Delta G_{\text{ACE}} + \Delta E_{\text{vdw}}$ 和 $\Delta E_{\text{ele}} + \Delta G_{\text{ACE}} + \Delta E_{\text{vdw}}$ 对 17 个蛋白质-蛋白质复

表 2 17 种蛋白质-蛋白质复合物的打分结果

Table 2 Scoring results in docking for the 17 protein-protein complexes^{*}

Dock name	Scoring											
	Rank ^a	RMSD ^a	Rank ^b	RMSD ^b	Rank ^c	RMSD ^c	Rank ^d	RMSD ^d	Rank ^e	RMSD ^e	Rank ^f	RMSD ^f
Enzyme/inhibitor												
1CHOFF	32	0.30	1	0.01	11	0.21	1	0.01	1	0.07	1	0.01
2SICXX	26	0.33	1	0.18	4	0.18	7	0.41	4	0.37	1	0.18
1TGSXX	25	0.04	1	0.04	1	0.04	39	0.46	13	0.46	1	0.04
1BRSFF	86	0.20	-	-	-	-	35	0.44	56	0.50	69	0.23
2PTCXX	76	0.35	17	0.07	6	0.07	2	0.07	1	0.07	1	0.07
2SNIFF	95	0.40	26	0.17	12	0.37	56	0.44	35	0.40	37	0.25
Antibody/antigen												
1AHWFF	1	0.37	-	-	-	-	9	0.50	51	0.50	13	0.22
1BVKFF	-	-	91	0.32	-	-	-	-	-	-	-	-
1BQLXF	87	0.46	21	0.07	-	-	99	0.46	52	0.07	75	0.07
1WEJFF	94	0.39	-	-	-	-	123	0.35	-	-	-	-
1NMBXF	-	-	-	-	-	-	-	-	-	-	-	-
1KXQXF	4	0.35	-	-	-	-	18	0.47	42	0.47	1	0.09
1LQYXX	84	0.47	-	-	-	-	-	-	-	-	-	-
Others												
1A00FX	1	0.37	-	-	1	0.37	14	0.31	49	0.31	12	0.31
1MDAFF	-	-	5	0.25	68	0.25	76	0.47	5	0.40	3	0.20
4INSXX	18	0.28	12	0.04	4	0.28	11	0.29	9	0.04	1	0.04
1FINFF	31	0.48	40	0.49	-	-	23	0.46	31	0.48	78	0.48

^{*}The highest ranking position of a near-native docked structure (before 200) and the corresponding RMSD (unit in nm). Dashes indicate that no near-native docked structures were ranked within the top 200; a) using ΔE_{ele} as the score function; b) using ΔG_{ACE} as the score function; c) using $\Delta E_{\text{ele}} + \Delta G_{\text{ACE}}$ as the score function; d) using $\Delta E_{\text{ele}} + \Delta E_{\text{vdw}}$ as the score function; e) using $\Delta G_{\text{ACE}} + \Delta E_{\text{vdw}}$ as the score function; f) using $\Delta E_{\text{ele}} + \Delta G_{\text{ACE}} + \Delta E_{\text{vdw}}$ as the score function

合物对接体系进行排序的结果。表中列出了第一个近天然结构的排位(截止到 200 位)及该结构主链原子相对于晶体结构的均方根偏差(RMSD)。当对接结构的 $\text{RMSD} \leq 0.5 \text{ nm}$ 时,就认为该结构是近天然构象结构。

从表 2 列出的数据可以看出,对于不同的体系,不同的打分函数各有优势。如对 17 种复合物对接体系,用打分函数 ΔG_{ACE} 可以找到 10 种复合物的近天然构象,其中有 4 种排在前 10 位,其 RMSD 一般较小;用打分函数 $(\Delta E_{\text{ele}} + \Delta G_{\text{ACE}})$ 可以找到 8 种复合物的近天然构象,有 5 种排在前 10 位,但 RMSD 一般较大;用打分函数 ΔE_{ele} 可以找到 14 种复合物的近天然构象,其中只有 3 种排在前 10 位,而且 RMSD 也较大。用打分函数 $\Delta E_{\text{ele}} + \Delta E_{\text{vdw}}$ 和 $\Delta G_{\text{ACE}} + \Delta E_{\text{vdw}}$ 分别找到了 14、13 种复合物的近天然构象,其中排在前 10 位的个数分别为 4、5。而用打分函数 $(\Delta E_{\text{ele}} + \Delta G_{\text{ACE}} + \Delta E_{\text{vdw}})$ 进行打分,可以找到 13 种复合物的近天然构象,有 7 种排在前 10 位,其中 6 种排在第 1 位,且 RMSD 一般都很小。很明

显,对酶/抑制剂类及其它类复合物体系,打分函数 $(\Delta E_{\text{ele}} + \Delta G_{\text{ACE}} + \Delta E_{\text{vdw}})$ 的效果较好,而打分函数 ΔG_{ACE} 对酶/抑制剂类复合物的近天然构象也显示出了较强的评价能力。Hu 等人^[15]指出,酶/抑制剂复合物经历了长期的进化过程,不断地优化其界面适应一定的生物功能,因而其界面具有较强的保守性,这可以解释为什么从酶/抑制剂导出的 ΔG_{ACE} 对这一类复合物打分是很成功的。另外,从表 2 可以看到,单独的 ΔG_{ACE} 作为打分函数对于抗原/抗体类复合物并不奏效,加上 ΔE_{ele} 后,结果更不好(见表 2, Rank^c)。在 7 种抗原/抗体类复合物中,对于 1NMBXF,六种打分函数均未筛选到近天然构象;对余下的 6 种,相对来说,用单独的静电相互作用 ΔE_{ele} 作为打分函数对近天然结构有较强的区分能力,找到了其中 5 种复合物的近天然构象,而用打分函数 ΔG_{ACE} 、 $\Delta E_{\text{ele}} + \Delta E_{\text{vdw}}$ 、 $\Delta G_{\text{ACE}} + \Delta E_{\text{vdw}}$ 和 $\Delta E_{\text{ele}} + \Delta G_{\text{ACE}} + \Delta E_{\text{vdw}}$ 分别找到了 2 种、4 种、3 种和 3 种近天然构象。然而,打分函数 ΔE_{ele} 加上范德华作用后评价效果反而被削弱了。这可能是由于抗

表3 5种蛋白质-蛋白质复合物对接体系经过能量优化(EM)后的打分结果

Dock name	Scoring											
	Rank ^a	RMSD ^a	Rank ^b	RMSD ^b	Rank ^c	RMSD ^c	Rank ^d	RMSD ^d	Rank ^e	RMSD ^e	Rank ^f	RMSD ^f
1BRSFF	46	0.32	-	-	96	0.32	4	0.20	9	0.20	3	0.20
2SNIFF	36	0.36	19	0.38	11	0.25	13	0.40	8	0.47	11	0.25
1BVKFF	94	0.47	74	0.42	166	0.42	47	0.42	38	0.42	47	0.42
1NMBXF	168	0.45	-	-	116	0.45	104	0.45	150	0.45	124	0.45
1BQLXF	-	-	16	0.07	73	0.07	40	0.07	32	0.07	36	0.07

* See foot note of table 2.

原/抗体的带电性较强, 静电作用在其识别中有很大的贡献, 因而用 ΔE_{ele} 来挑选近天然构象具有一定的优势.

从上面的分析可知, 在筛选近天然构象的能力上, $\Delta E_{\text{ele}} + \Delta G_{\text{ACE}} + \Delta E_{\text{vdw}}$ 要优于其它的各种打分函数.

2.2 能量优化对打分效果的影响

表3列出了5种蛋白质复合物对接体系(1BRSFF、2SNIFF、1BVKFF、1NMBXF和1BQLXF)经过能量优化(EM)后的打分排序结果. 对比表2和表3, 可以看出经过EM后, 各种打分函数挑选近天然构象的效果均有不同程度的改善, 尤其是含有范德华作用的打分函数的评价效果的改善更为明显. 如对复合物1BVKFF, 除了 ΔG_{ACE} 外, 其他各种打分函数在能量优化后均找到了其近天然结构, 而在优化前都未找到其近天然结构. 对2SNIFF和1NMBXF两种复合物, 用 $\Delta E_{\text{ele}} + \Delta E_{\text{vdw}}$ 和 $\Delta G_{\text{ACE}} + \Delta E_{\text{vdw}}$ 在能量优化后进行打分筛选到的近天然结构的排序明显提前, 且对于1BQLXF, 在能量优化后, 这两种打分函数均优先找到了RMSD较小的构象; 而用 $\Delta E_{\text{ele}} + \Delta G_{\text{ACE}} + \Delta E_{\text{vdw}}$ 对这两种复合物进行打分排序, 在能量优化后, 同一近天然结构分别由优化前的37位提升到11位、由75位提升到36位. 更明显的例子是复合物1BRSFF, 用 $\Delta E_{\text{ele}} + \Delta G_{\text{ACE}} + \Delta E_{\text{vdw}}$ 打分在能量优化前筛选到的近天然构象排在69位, 其RMSD=0.23 nm, 而在优化后则排在第3位, RMSD=0.20 nm, 更接近于晶体结构, 这主要是由于能量优化消除了部分原子交叠, 而使范德华能降低所致. 比如对复合物1BRSFF的RMSD为0.20 nm的近天然结构, 其范德华能由优化前的98873.71 kJ·mol⁻¹减少到了优化后的-432.39 kJ·mol⁻¹. 可见, EM确实有助于打分函数尤其是包含范德华作用的打分函数找到RMSD较小的近天然结构.

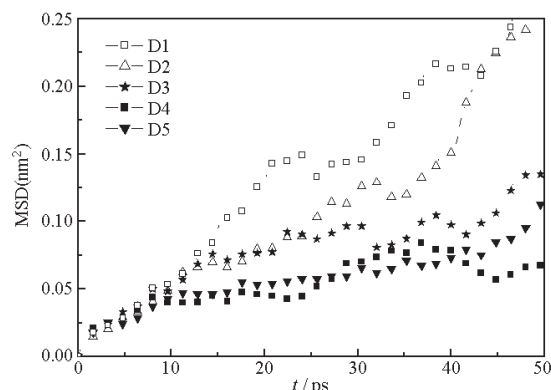


图1 4INSXX的5个候选结构的MD轨迹中的MSD随时间的变化

Fig. 1 The variations upon time in MD for the MSD of five selected structures of 4INSXX

2.3 用MD模拟辅助挑选近天然结构

为了检验MD模拟是否可以辅助挑选近天然结构, 我们选取了4INSXX体系排在前面的3个对接结构(其中有2个非近天然构象、1个近天然构象)和后面依次出现的2个近天然结构进行MD模拟, 这五个对接结构的RMSD分别为1.09、1.13、0.28、0.20和0.24 nm. 用GROMACS程序对以上5个体系分别进行了50 ps的MD模拟. 这5种不同体系的MD模拟结果分别用D1、D2、D3、D4、D5来表示(见图1). 由图1可以看出, 非近天然构象D1、D2的MSD变化很大, 随时间迅速升高, 说明其构象不稳定; 而近天然构象D3、D4和D5的MSD则相对较小, 而且随时间的变化不大, 这说明近天然构象是比较稳定的. 所以, MD模拟可以在一定程度上辅助打分函数挑选蛋白质-蛋白质对接中的近天然结构.

3 结论

在蛋白质-蛋白质对接中, 经验的结合自由能打分函数 $\Delta E_{\text{ele}} + \Delta G_{\text{ACE}} + \Delta E_{\text{vdw}}$ 具有较好的评价近天然构象的能力, 相对于打分函数 ΔE_{ele} 、 ΔG_{ACE} 、

$\Delta E_{\text{ele}} + \Delta G_{\text{ACE}}$ 、 $\Delta E_{\text{ele}} + \Delta E_{\text{vdw}}$ 和 $\Delta G_{\text{ACE}} + \Delta E_{\text{vdw}}$, 更能筛选到 RMSD 较小的近天然构象. 对构象进行能量优化可以使不同的能量打分函数有更好的区分近天然构象的效果, 尤其是包含范德华相互作用的打分函数 $\Delta E_{\text{ele}} + \Delta G_{\text{ACE}} + \Delta E_{\text{vdw}}$ 更是如此, 甚至可以筛选到在优化之前无法找到的近天然构象. 另外, 在打分得到的候选结构中, 根据其 MD 轨迹相对于初始构象的 MSD 随时间的变化可以排除不合理构象, 辅助打分函数找到近天然结构.

References

- Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. *Proteins*, **2002**, *47*: 409
- Jiang, F.; Kim, S. H. *J. Mol. Biol.*, **1991**, *219*: 79
- Lin, S. L.; Nussinov, R.; Fischer, D.; Wolfson, H. J. *Proteins*, **1994**, *18*: 94
- Waals, P. H.; Sternberg, M. J. *J. Mol. Biol.*, **1992**, *228*: 277
- Gabb, H. A.; Jackson, R. M.; Sternberg, M. J. *J. Mol. Biol.*, **1997**, *272*: 106
- Shoichet, B. K.; Kuntz, I. D. *J. Mol. Biol.*, **1991**, *221*: 327
- Camacho, C. J.; Gatchell, D. W.; Kimura, S. R.; Vajda, S. *Proteins*, **2000**, *40*: 525
- Li, C. H.; Ma, X. H.; Chen, W. Z.; Wang, C. X. *Acta Biophysica Sinica*, **2003**, *19*: 47 [李春华, 马晓慧, 陈慰祖, 王存新. 生物物理学报 (*Shengwu Wuli Xuebao*), **2003**, *19*: 47]
- Zhang, C.; Vasmatzis, G.; Cornette, J. L. DeLisi, C. J. *Mol. Biol.*, **1997**, *267*: 707
- Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.*, **1983**, *4*: 187
- Palma, P. N.; Krippahl, L.; Wampler, J. E.; Moura, J. J. G. *Proteins*, **2000**, *39*: 178
- van der Spoel, D.; van Buuren, A. R.; Apol, E.; Meulenhoff, P. J.; Tieleman, D. P.; Sijbers, A. L. T. M.; Hess, B.; Feenstra, K. A.; Lindahl, E.; van Drunen, R.; Berendsen, H. J. C. GROMACS user manual, version 3.0. Nijenborgh 4, 9747 AG Groningen. The Netherlands, 1991
- Cherfils, J.; Duquerroy, S.; Janin, J. *Proteins*, **1991**, *11*: 271
- Janin, J. *Proteins*, **2002**, *47*: 257
- Hu, Z.; Ma, B.; Wolfson, W.; Nussinov, R. *Proteins*, **2000**, *39*: 331

A New Method to Select the Near-native Conformations from the Docked Structures*

Yu Yong-Hui Li Chun-Hua Lu Ben-Zhuo Chen Wei-Zu Wang Cun-Xin
(College of Life Science and Bioengineering, Beijing University of Technology, Beijing 100022)

Abstract A scoring function that contains the electrostatic, desolvation and van der Waals energy was used to discriminate the near-native conformations in protein-protein docking. The analysis results of the docked structures for seventeen protein complexes showed that the scoring function including van der Waals energy ($\Delta E_{\text{ele}} + \Delta G_{\text{ACE}} + \Delta E_{\text{vdw}}$) had much more potentiality in distinguishing the near-native conformations from the incorrect ones compared with other scoring functions, such as ΔE_{ele} , ΔG_{ACE} , $\Delta E_{\text{ele}} + \Delta G_{\text{ACE}}$, $\Delta E_{\text{ele}} + \Delta E_{\text{vdw}}$ and $\Delta G_{\text{ACE}} + \Delta E_{\text{vdw}}$. It was found that energy minimization could improve the discrimination results for all the scoring functions mentioned above to some extent, especially for $\Delta E_{\text{ele}} + \Delta G_{\text{ACE}} + \Delta E_{\text{vdw}}$. Finally, to further determinate the near-native structure, we performed the molecular dynamics(MD) simulations for the candidate structures of 4INSXX. According to the variations of the mean square deviation(MSD) of the structures in MD trajectories relative to the initial structures, the false structures could be excluded.

Keywords: Scoring function, van der Waals energy, Energy minimization, Molecular dynamics(MD) simulation, Near-native conformation