一种改进的研究蛋白质-蛋白质相互作用的平均势方法

苏 煜 李文飞 张 建 王 骏 王 炜*

(南京大学固体微结构物理国家重点实验室, 南京大学物理系, 南京 210093. * 联系人, E-mali: wangwei@biophy.nju.edu.cn)

摘要 在原子水平上发展了一种距离相关的用于研究蛋白质-蛋白质相互作用的平均势(potential of mean force, 简称 PMF)方法. 与传统理论模型相比, 我们的模型考虑了蛋白质系统的复杂环境因素. 这种改进使得该模型能够给出物理上更合理和准确的势函数形式. 得到这样的势函数是正确描述蛋白质结构及相互作用的前提条件. 而且借助于改进后的方法, 还可以对蛋白质中残基相互作用的空间拓扑规律进行研究. 期望这种改进将促进平均势方法在蛋白质科学其他领域, 如蛋白质折叠识别, 结构预测及热稳定性预测中的应用和发展.

关键词 基于知识的方法 平均势(PMF) 蛋白质-蛋白质相互作用

基于知识的方法, 已经广泛地应用于蛋白质结构预测, 蛋白质折叠, 热稳定性预测以及蛋白质-蛋白质对接和绑定(docking and binding)的研究中, 并取得了一定的成功[1~9]. 这种方法的主要思想是对蛋白质结构数据库(protein data bank, 简称 PDB)[10]中的晶体结构数据进行统计分析, 进而计算原子(残基)对相互作用. 近年来, 这类方法中最吸引人和广泛使用的是距离相关(distance-dependent)的方法, 通常被称为平均势(PMF)方法. 它是流体统计力学中平均势方法[11,12]在蛋白质领域中新的发展和应用.

平均势方法分为残基水平和原子水平. 人们最先发展的是残基水平的方法, 并且已经取得了一定的成功[13,14]. 近些年来, 随着 PDB 数据库数据指数性的增长, 考虑原子水平细节的方法成为热点课题. 这种细节的考虑主要沿两个方向发展. 一种是在传统残基水平的方法中, 考虑原子方向性细节, 即在残基水平的势函数中加入角度分布因子. 研究表明其在一些应用上比传统残基水平的方法有一定的提高[15,16]. 另一种方法是直接基于原子水平的方法. 这种方法结合精细的原子类型定义, 可以得到更多、更细致的原子间相互作用信息. 近些年来, 这种方法得到了广泛的应用[17~24].

平均势方法因其构建简单计算快速等优点而非常具有吸引力. 但是这类方法的理论基础和物理解释仍然不尽清楚, 尚存在很多争论[25~30]. 这主要是因为以往工作将此方法应用于蛋白质系统时, 大多只是直接套用来源于流体统计力学中的方法. 但是蛋白质系统是一种非常复杂的软物质系统, 有很多简单流体所没有的特性和环境, 如蛋白质链的连接性. 所以将简单流体中平均势方法应用到蛋白质系统时, 应该把这种软物质系统特有的复杂性因素加以考虑.

我们的工作通过考虑蛋白质系统特有的复杂性因素以改进传统的平均势方法, 然后用势函数的合理性和保留有价值信息的能力来判断方法是否得到了改进. 这是因为一方面得到合理和准确的势函数是正确描述蛋白质结构及相互作用的前提条件, 另一方面几乎所有已出版的平均势工作都给出了所得到势函数的图像. 所以我们借由对势函数的讨论, 可以和以往平均势方法在蛋白质科学中不同领域的工作进行广泛的比较.

统计势函数是一种相似于有效物理能量函数(physical effective energy function, 简称 PEEF)[6] 又有其独特特性的函数. 一方面, 它拥有如 Lennard-Jones 势的非键粒子对相互作用的经典图像. 另一方面, 统计势函数还拥有一些特性. 这是因为它还包含了很多其他效应, 比如共价键基团间的相互影响、溶剂化作用、三体作用等. 统计势函数同物理势函数都有一些共同的性质. 我们把它称为势函数的“合理性”. 例如, 势函数中一般有一个或几个势阱, 这些势阱的极小值处, 通常为负值, 代表了粒子对在平衡位置时, 相互间存在吸引势. 此外, 势函数在长程处的渐进值应该是稳定值, 且此值通常为 0. 这代表粒子对相距很远时, 相互作用势为 0. 当把平均势方法直接应用于蛋白质这个复杂系统时, 所得到的势函数中, 有一

2007-05-28 收稿, 2007-08-24 接受

国家自然科学基金(批准号: 90403120, 10474041, 10504012)资助项目

些是符合以上性质的. 但是还有大量的势函数常表现出一些奇怪的行为, 例如很多势函数是全程排斥的. 我们认为这些异常的势函数不能反映原子对真实的相互作用. 这样的结果是由于直接套用简单流体系统的平均势方法到蛋白质系统造成的. 这种套用没有充分考虑蛋白质系统特有的复杂性.

在本文中, 我们发展了一种平均势方法, 它是距离相关的, 基于原子水平的. 为了获得更多、更精细的原子对相互作用细节, 我们定义了 47 种原子类型. 通过考虑蛋白质系统的复杂环境因素, 最终获得了比传统方法有较大改善的势函数. 特别对蛋白质链的约束造成的势函数不合理问题改进明显. 而且借助改进后的方法, 还可以对蛋白质系统中残基相互作用的空间拓扑规律进行研究.

1 方法与材料

我们所用的传统方法相似于 Sippl 提出的方法[13]. 此传统方法已被广泛地应用在蛋白质及蛋白质复合物领域的研究中. 在该模型中, 相距为 r 的两个原子之间的势函数由下式给出(具体推导过程见 Sippl 的工作[13]):

$$A_{ij}(r) = -k_{\mathrm{B}}T\ln[P_{ij}(r)], \tag{1}$$

式中 k_{B} 为玻尔兹曼常数, T 为绝对温度. $P_{ij}(r)$为统计优势函数, 具体由下式给出:

$$P_{ij}(r) = \frac{q_{ij}(r)}{q_{xx}(r)}, \tag{2}$$

其中 $q_{ij}(r)$为原子对 ij 归一化后的相对频率, $q_{xx}(r)$为所有原子对归一化后的相对频率. 它们的具体定义将在下面给出.

上面方法中用所有原子的平均作用做参考状态. 但实际上, 每一种原子都有它特殊的环境. 因此用一个统一的参考状态是一种非常粗糙的简化, 显得过于平均. 在下面我们将从平均势基础理论出发, 给出改进方法的推导.

与简单流体理论中的平均势理论相似[11], 蛋白质系统中通过观测频率直接得到的平均势函数 $A_{ij}^{\mathrm{obs}}(r)$ 可以写成两部分之和: 一部分代表原子 i 和 j 直接相互作用 $A_{ij}^{\mathrm{dir}}(r)$, 另一部分代表其他环境因素带来的效果 $A_{ij}^{\mathrm{env}}(r)$, 即

$$A_{ij}^{\mathrm{obs}}(r) = A_{ij}^{\mathrm{dir}}(r) + A_{ij}^{\mathrm{env}}(r). \tag{3}$$

根据流体统计力学中平均势理论, 我们可以建立观测的原子对 ij 的相对频率和平均势之间的关系,

$$q_{ij}^{\mathrm{obs}}(r) = \exp\left[\frac{1}{-k_{\mathrm{B}}T}A_{ij}^{\mathrm{obs}}(r)\right]. \tag{4}$$

另外我们把原子 i 和其他所有原子的相对频率之和用下式表示:

$$q_{ix}^{\mathrm{obs}}(r) = \sum_j q_{ij}^{\mathrm{obs}}(r) = \sum_j \exp\left[\frac{1}{-k_{\mathrm{B}}T}A_{ij}^{\mathrm{obs}}(r)\right]. \tag{5}$$

用(4)式除以(5)式, 然后把其中的 $A_{ij}^{\mathrm{obs}}(r)$ 用(3)式替换得到

$$\begin{aligned}\frac{q_{ij}^{\mathrm{obs}}(r)}{q_{ix}^{\mathrm{obs}}(r)} &= \frac{\exp\frac{1}{-k_{\mathrm{B}}T}A_{ij}^{\mathrm{obs}}(r)}{\sum_j \exp\frac{1}{-k_{\mathrm{B}}T}A_{ij}^{\mathrm{obs}}(r)} \\ &= \frac{\exp\frac{1}{-k_{\mathrm{B}}T}[A_{ij}^{\mathrm{dir}}(r)+A_{ij}^{\mathrm{env}}(r)]}{\sum_j \exp\frac{1}{-k_{\mathrm{B}}T}[A_{ij}^{\mathrm{dir}}(r)+A_{ij}^{\mathrm{env}}(r)]},\end{aligned} \tag{6}$$

在这里, 我们认为所有拥有相同原子 i 的原子对 ij 都拥有同样的环境因素. 即当 j 取不同原子时, 所有 $A_{ij}^{\mathrm{env}}(r)$ 都相等, 这是个合理的假设. 根据这个近似, 我们可以消去(6)式中的此项, 从而得到

$$\frac{q_{ij}^{\mathrm{obs}}(r)}{q_{ix}^{\mathrm{obs}}(r)} = \frac{\exp\frac{1}{-k_{\mathrm{B}}T}A_{ij}^{\mathrm{dir}}(r)}{\sum_j \exp\frac{1}{-k_{\mathrm{B}}T}A_{ij}^{\mathrm{dir}}(r)}. \tag{7}$$

把上面的等式两边取自然对数得到

$$A_{ij}^{\mathrm{dir}}(r) = -k_{\mathrm{B}}T\ln\frac{q_{ij}^{\mathrm{obs}}(r)}{q_{ix}^{\mathrm{obs}}(r)} + C, \tag{8}$$

式中的常数 C 由下式定义,

$$C = -k_{\mathrm{B}}T\ln\sum_j \exp\frac{1}{-k_{\mathrm{B}}T}A_{ij}^{\mathrm{dir}}(r), \tag{9}$$

在这里我们和以往的工作一样把(8)式中常数项 C 忽略. 事实上, 我们可以通过定义势函数 $A_{ij}^{\mathrm{dir}}(r)$ 在远距离处为0(即粒子对在相距较远时相互作用为0)来得到这个常数. 它不改变函数形状, 只是使得最后得到的势函数在纵向有一个平移. 并且从后面的结果看, 这个常数非常接近于 0.

经过近似, 我们得到一个简单的表达式,

$$A_{ij}^{\mathrm{dir}}(r) = -k_{\mathrm{B}}T\ln\frac{q_{ij}^{\mathrm{obs}}(r)}{q_{ix}^{\mathrm{obs}}(r)}, \tag{10}$$

式中 $q_{ij}^{\mathrm{obs}}(r)$ 和 $q_{ix}^{\mathrm{obs}}(r)$ 的具体的定义将在下面给出.

把这个方法和传统方法相比较, 我们可以看出它的物理意义. 在传统方法中, 用所有原子对的观测频率作为参考状态, 这样做是认为所有原子都拥有相同的环境因素. 而在改进的方法中, 用所有以原子 i 为基点的观测频率作为参考状态. 这样可以考虑到不同的原子 i 拥有的不同环境因素, 从而避免了对所有原子采用一个统一的参考状态这样过于平均的作法. 相比传统方法我们能够得到更多的相互作用细节.

下面给出在具体统计计算中, 如何得到 $q_{ij}(r)$, $q_{ix}(r)$和 $q_{xx}(r)$. 首先, 我们从训练集的所有结构中, 通过统计获得一对原子 i 和 j 在相距为 r 时所发生的数目. 统计范围从 0~12 Å, 间隔为 0.1 Å.

$$N_{ij}(r)=\sum_{p}\delta(r_{ij}-r)\,, \tag{11}$$

式中$\delta(x)$是δ函数.

然后我们对观测数目归一化, 得到相应的相对频率为

$$q_{ij}(r)=\frac{N_{ij}(r)}{\sum\limits_{r}N_{ij}(r)}\,. \tag{12}$$

另外 $q_{ix}(r)$和 $q_{xx}(r)$由 $q_{ij}(r)$得到

$$q_{ix}(r)=\sum_{j}q_{ij}(r)\,,\quad q_{xx}(r)=\sum_{i}\sum_{j}q_{ij}(r)\,. \tag{13}$$

在我们的工作中, 对20种氨基酸残基的所有的重原子定义了 47 种原子类型, 见表 1. 原子类型的定义是依据蛋白质中原子的物理化学性质、连接性和环境. 在定义原子类型的时候, 为了获得更多的相互作用细节, 一方面我们要尽量定义更多的原子类型, 但是另一方面, 为了满足有足够数据统计的需要, 我们定义的数目又不能过多. 最终, 定义的原子类型数目是这两个因素的折中. 此外为了获得统计上稳定的势函数, 我们只考虑总观测数目大于 1000 次的原子对. 这个标准曾在前人的工作中使用[21].

用于统计势函数的非冗余训练集来源于蛋白质结构数据库(PDB)[10]. 我们只包含X射线的蛋白质-蛋白质复合物结构. 分辨率要小于 2.5 Å. 基于这样的标准, 共有 438 个条目被选出. 为了排除近似结构, 我们根据 PDB 条目里 REMARK 中的分子信息及所引用的参考文献对结构进行逐一筛选. 在筛选过程中也依靠三维分子图形软件 RasMol. 对相同的复合物我们只取其中解析度最高的一个结构. 最终, 训练

表 1 标准氨基酸残基的重原子类型定义

原子类型	类型定义
1	Cα (所有残基, 除了 Gly)
2	Gly-Cα
3	N(所有残基, 除了 Pro)
4	C (所有残基)
5	O (所有残基)
6	Val-Cγ1, Val-Cγ2, Leu-Cδ1, Leu-Cδ2, Ile-Cγ2, Ile-Cδ, Thr-Cγ
7	Leu-Cγ, Ile-Cγ1, Gln-Cγ, Lys-Cγ, Lys-Cδ, Glu-Cγ, Arg-Cγ
8	Cβ (所有残基, 除了 Pro, Ser, Thr, Cys)
9	Met-Sδ
10	Pro-N
11	Phe-Cγ, Tyr-Cγ
12	Phe-Cδ1, Phe-Cδ2, Phe-Cε1, Phe-Cε2, Phe-Cζ, Tyr-Cδ1, Tyr-Cδ2, Tyr-Cε1, Tyr-Cε2
13	Trp-Cγ
14	Trp-Cε2
15	Ser-Cβ
16	Ser-Oγ, Thr-Oγ
17	Thr-Cβ
18	Asn-Nδ2; Gln-Nε2
19	Cys-Sγ
20	Lys-Nζ
21	Arg-Cξ
22	Arg-Nη1, Arg-Nη2
23	His-Cγ
24	His-Cδ2
25	His-Nε2
26	His-Cε1
27	Asp-Cγ; Glu-Cδ
28	Asp-Oδ1, Asp-Oδ2; Glu-Oε1, Glu-Oε2
29	Cys-Cβ
30	Met-Cε
31	Tyr-Cζ
32	Pro-Cδ
33	Asn-Cγ; Gln-Cδ
34	Asn-Oδ1; Gln-Oε1
35	Lys-Cε
36	Arg-Nε
37	Arg-Cδ
38	His-Nδ1
39	Trp-Nε1
40	Tyr-Oη
41	OXT (链终端的氧原子)
42	Pro-Cβ
43	Pro-Cγ
44	Met-Cγ
45	Trp-Cε3, Trp-Cζ2, Trp-Cζ3, Trp-Cη2
46	Trp-Cδ1
47	Trp-Cδ2

数据库包含了来源于 128 个 PDB 条目的 174 个蛋白质-蛋白质界面, 见表 2.

另外直接计算出的势函数要用一个宽为 0.5 Å 的移动窗进行平滑处理. 平滑窗中有 5 个间隔, 其权重分别为 1:2:4:2:1. 这种平滑方法是取自 Mitchell 和 Thornton 的工作[22].

2 结果与讨论

下面将比较传统方法和改进方法所得到的势函数. 改进的方法考虑了蛋白质系统的复杂性因素, 从而使得到的势函数较传统方法变得更加合理.

原子对势可分为 3 类: 主链-主链(B-B), 主链-侧链(B-S), 侧链-侧链(S-S). 我们分别对它们进行观察和讨论. 每一类势函数都由两种方法分别计算出的 3 个势函数作为例子. 结果分别在图 1, 2 和 3 中给出.

图 1 为主链-主链(B-B)原子对势函数 (它们分别是原子对 4-2, 3-1, 1-1. 这里的数字是表 1 中定义的原子类型). 图 1(a)势函数是用传统方法计算的势函数, 其值在大部分距离处是正的, 代表原子对(4-2)之间有较强的排斥势. 而用改进方法计算得到的势函数(图 1(b))则出现了明显的吸引阱, 其势函数图像很符合经典的 Lennard-Jones 势图像, 变得更加合理.

在原子对(3-1)的势函数中, 用传统方法计算的势函数(3-1)显示了较强的排斥势(图 1(c)), 并且代表范德瓦尔斯作用的势阱很不明显. 而用改进后的方

表 2 训练集所含的 128 个 PDB 结构

12gs	1axd	1a09	1abo	1bhf	1efn	1flt	1ir3	1lck	1a3b
1a2c	1a4w	1a46	1a5g	1a61	1an1	1brb	1brc	1avw	1mct
1sfi	1slu	1smf	1a5s	1ab9	1gbb	1gl1	1tec	1sib	1scn
3sic	2tgp	3cyh	3nse	3sgb	5csm	1a14	1a1n	1a2y	1a3r
1a9e	1aqd	1bd2	1bj1	1e4x	1fdl	1gc1	1h0d	1ikf	1itb
1jhl	1kip	1ld9	1mel	1mlc	1nmc	1oau	1oak	1oga	1ogt
1osp	1osz	1qew	1qo3	1sm3	1wej	2h1p	2hrp	2jel	2seb
2vaa	1vad	1a2k	1a2x	1abr	1abw	1ak4	1aqc	1aqv	1axi
1bnd	1bt6	1dkz	1dzb	1e96	1eay	1eer	1efu	1exf	1gg2
1got	1gua	1gux	1gzs	1h2s	1he1	1hwg	1jhg	1lfd	1oby
1obz	1oey	1ohz	1okk	1okv	1okv	1ol5	1qja	1qls	1rst
1rsu	1shd	1slg	1spp	1taf	1tbg	1tx4	1tze	1upt	1uzx
1www	1x11	1ycs	1zfp	2cbl	2fib	2prg	2trc		

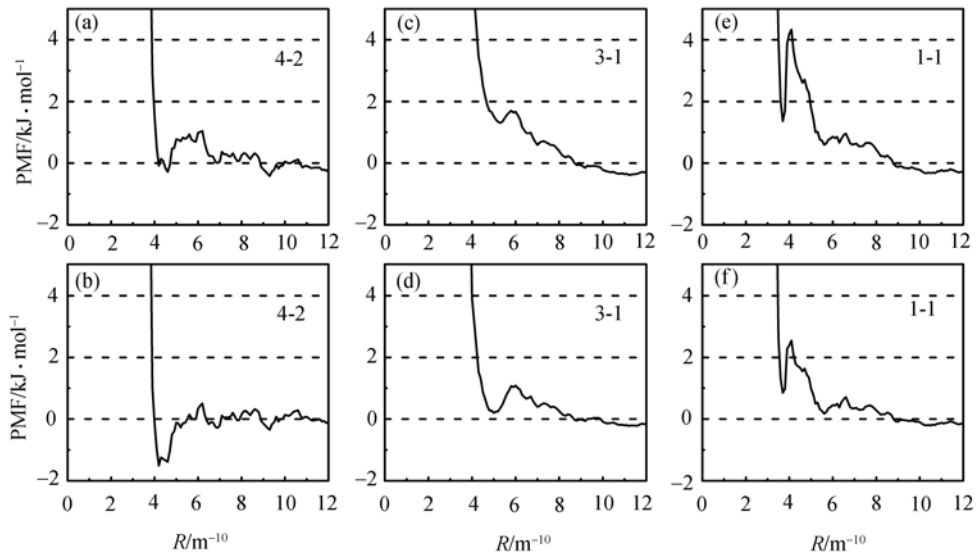

图 1 分别用传统方法和改进方法计算得到的 3 个主链-主链原子对势函数

3 个对势分别为 4-2, 3-1, 1-1, 其中(a), (c), (e)为传统方法计算所得, (b), (d), (f)为改进方法计算所得

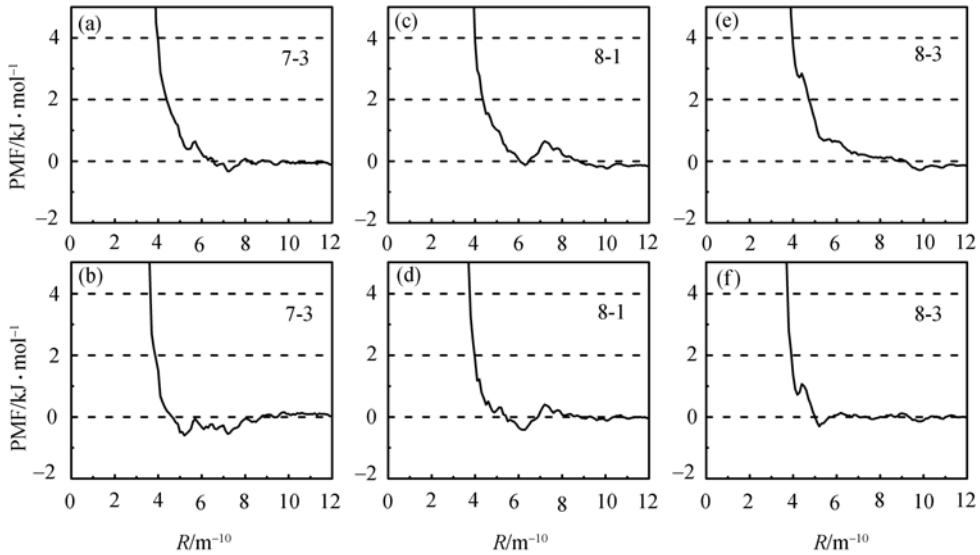

图 2 分别用传统方法和改进方法计算得到的 3 个主链-侧链原子对势函数

3 个对势分别为 7-3, 8-1, 8-3, 其中(a), (c), (e)为传统方法计算所得, (b), (d), (f)为改进方法计算所得

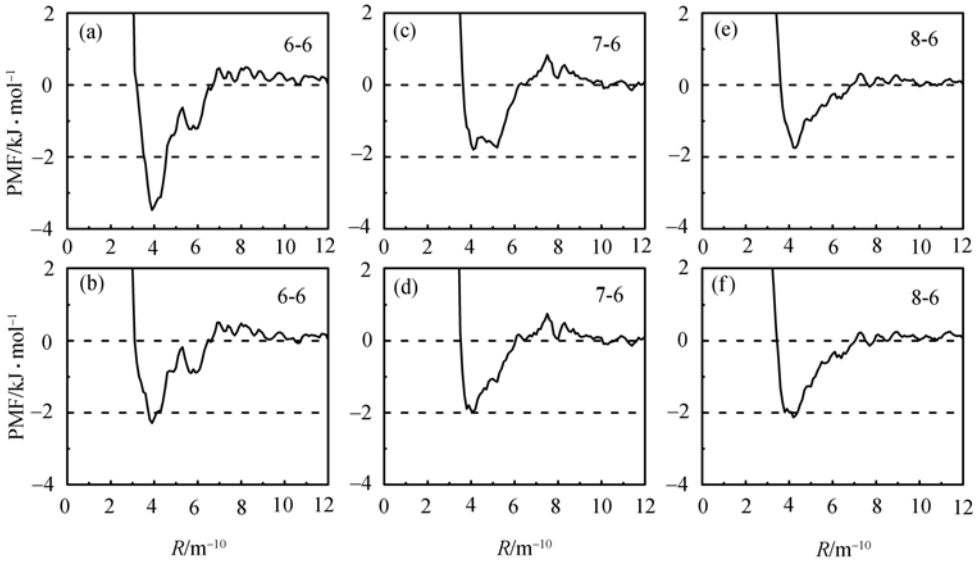

图 3 分别用传统方法和改进方法计算得到的 3 个侧链-侧链原子对势函数

3 个对势分别为 6-6, 7-6, 8-6, 其中(a), (c), (e)为传统方法计算所得, (b), (d), (f)为改进方法计算所得

法计算则得到了较明显的阱(图 1(d)).

在原子对(1-1)中, 我们可以观测到用传统方法计算的对势有很强的排斥(图 1(e)). 第一极小值大小是 1.4 kJ/mol, 出现在 3.7 Å 处. 它相应于两个碳原子的范德瓦尔斯半径之和. 从图 1 中可以看出这个极小值及其周围表现出一个明显的势阱图像, 代表了两个碳原子在这个距离处的接触. 在阱的后面有一个高度为 4.3 kJ/mol 垒, 在距离 4.1 Å 处. 这个垒的含义同简单流体理论中相同, 是因为两个碳原子以范德瓦尔斯接触造成的频率优势使之后留下了较少的观测频率. 而改进方法得到的势(图 1(f))较传统方法得到的势排斥变弱: 阱和其后垒的位置不变, 但是值变小了, 分别为 0.8 kJ/mol, 2.5 kJ/mol. 这显示了排斥效应得到了一定程度的改善.

总的来说, 在传统方法中, 大部分主链-主链势函数显示了很强的排斥. 而改进后的方法使这种排斥作用被削弱了.

图 2 为主链-侧链(B-S)势函数(它们分别为 7-3, 8-1,

8-3). 在图 2(a)中, 传统方法得到的势函数第一极小值为正值, 而改进后(图 2(b))其值变为负, 势函数图像也变得较合理. 势函数(8-1), (8-3)在改进前后也体现了同样的趋势. 另外, 从整体上看, 用传统方法得到的 B-B 势函数(图 1(a), (c), (e)), 排斥都比同方法得到的 B-S 势函数(图 2(a), (c), (e))要强. 原因在后面我们将会讨论.

图 3 为侧链-侧链(S-S)势函数 (它们分别为 6-6, 7-6, 8-6). 从总体可以看出用传统方法得到的势函数(图 3(a), (c), (e)), 已经出现了明显的吸引阱. 而改进后的方法对势函数的影响不大. 但值得注意的是, 在侧链-侧链势函数(7-6)和(8-6)中, 改进方法比传统方法多了一些细节. 例如在传统方法得到的势函数中(图 3(c)), 势函数第一极小值在 4.1 Å 处, 第二极小值在 5.2 Å 处. 而改进方法得到的势函数(图 3(d))除了这两个极小值外, 还可以观察到在更近的距离 3.8 Å 处出现了一个极小值. 它恰好是两个碳原子的范德瓦尔斯半径之和, 代表了分叉的侧链头部原子类型 6 和根部原子类型 7 是可以以范德瓦尔斯表面直接接触的. 这种接触模式以前也被 Singh 和 Thornton 所观测到[31]. 可是如果仅从传统方法观察, 我们可能得到错误的结论: 在蛋白质中, 分叉碳链头部和根部的两个碳原子由于空间的障碍, 不能直接以范德瓦尔斯表面接触. 这样错误的结论将强烈地影响到人们对蛋白质中残基相互作用的空间拓扑规律的认识.

综合图 1, 2 和 3 的结果, 我们看到在传统方法中, 主链-主链和主链-侧链势函数经常会表现出很强的排斥势. 这种特性也被以前用相似方法的工作所观测到[14,20,22,32]. 而在改进方法得到的势函数中, 这种排斥在很大程度上被削弱了, 并且其中一些势函数出现了更多的细节. 从上面对侧链-侧链势函数的讨论中, 我们可以知道这些新的细节是合理的. 它们特别有助于正确认识蛋白质中残基相互作用的空间拓扑规律.

在传统方法中, 主链-主链和主链-侧链势函数的普遍排斥是由一些因素导致的, 我们认为这其中蛋白质链造成的约束是一个重要的原因. 在简单流体中, 可自由接触的原子对之间的观测频率可以直接反映它们之间的相互作用强度. 但是在蛋白质系统中, 非共价键原子对之间相互作用被它们所在的链约束了. 所以最终观测到的粒子对频率不但受到非共价键原子对间相互作用的影响, 还被同一条链中共价相连的原子所约束影响. 这种约束使得本该观测到有优势频率的地方只出现很少的观测频率. 从而使最终得到的势函数表现出很强的排斥.

另外我们还观测到在传统方法中, 势函数排斥的强度顺序为主链-主链>主链-侧链>侧链-侧链. 这个顺序也被以前的工作所观察到(在文献[2]的图 1 中及在文献[14]的图 3 中). 我们认为这种强度顺序也是因为约束强度的不同造成的. 在蛋白质中, 主链起到蛋白质的骨架的作用, 其上原子受到的约束较大, 自由度较低. 而侧链约束能力相对较小, 其上原子自由度较高. 正是由于这种约束强度的顺序, 最终造成了上述势函数排斥强度的顺序.

3 结论

基于知识的方法已经被广泛地应用于蛋白质领域的研究, 我们发展了距离相关的方法用于研究蛋白质-蛋白质相互作用. 通过将蛋白质系统的复杂性因素考虑到传统的方法中, 得到了比以往的研究拥有更合理物理图像的和更多细节的势函数. 而且借助改进后的方法我们还可以对蛋白质中残基相互作用的空间拓扑规律作进一步研究.

另外我们发现在传统的平均势方法中, 来源于单链蛋白质内部的势函数[14,18,20,32]和本文计算的来源于蛋白质-蛋白质界面的势函数非常相似. 所以虽然改进的方法在本文中是用于研究蛋白质-蛋白质相互作用的, 但是我们也期望它会促进平均势方法在蛋白质折叠识别, 结构预测及热稳定性预测等领域的应用和发展.

参 考 文 献

1 Sippl M J. Knowledge-based potentials for proteins. Curr Opin Struct Biol, 1995, 5: 229—235[DOI]

2 Jernigan R L, Bahar I. Structure-derived potentials and protein simulations. Curr Opin Struct Biol, 1996, 6: 195—209[DOI]

3 Jones D T, Thornton J M. Potential energy functions for threading. Curr Opin Struct Biol, 1996, 6: 210—216[DOI]

4 Vajda S, Sippl M J, Novotny J. Empirical potentials and functions for protein folding and binding. Curr Opin Struct Biol, 1997, 7: 222—228[DOI]

5 Moult J. Comparison of database potentials and molecular mechanics force fields. Curr Opin Struct Biol, 1997, 7: 194—199[DOI]

6 Lazaridis T, Karplus M. Effective energy functions for protein structure prediction. Curr Opin Struct Biol, 2000, 10: 139—145[DOI]

7 Gohlke H, Klebe G. Statistical potentials and scoring functions ap-

plied to protein-ligand binding. Curr Opin Struct Biol, 2001, 11: 231—235[DOI]
8 Buchete N, Straub J E, Thirumalai D. Development of novel statistical potentials for protein fold recognition. Curr Opin Struct Biol, 2004, 14: 225—232[DOI]
9 李春华, 马晓慧, 陈慰祖, 等. 蛋白质-蛋白质分子对接方法研究进展. 生物化学与生物物理进展, 2006, 33(7): 616—621
10 Berman H M, Westbrook J, Feng Z, et al. The protein data bank. Nucleic Acids Res, 2000, 28: 235—242[DOI]
11 Chandler D. Introduction to Modern Statistical Mechanics. New York: Oxford University Press, 1987
12 McQuarrie D. Statistical Mechanics. New York: Harper Collins, 1976
13 Sippl M J. Calculation of conformational ensembles from potentials of mean force. an approach to the knowledge-based prediction of local structures in globular proteins. J Mol Biol, 1990, 213: 859—883[[DOI]
14 Bahar I, Jernigan R L. Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation. J Mol Biol, 1997, 266: 195—214[DOI]
15 Buchete N, Straub J E, Thirumalai D. Orientational potentials extracted from protein structures improve native fold recognition. Protein Sci, 2004, 13: 862—874[DOI]
16 Hoppe C, Schomburg D. Prediction of protein thermostability with a direction- and distance-dependent knowledge-based potential. Protein Sci, 2005, 14: 2682—2692[DOI]
17 Melo F, Feytmans E. Novel knowledge-based mean force potential at atomic level. J Mol Biol, 1997, 267: 207—222[DOI]
18 Samudrala R, Moult J. An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction. J Mol Biol, 1998, 275: 895—916[DOI]
19 Robert C H, Janin J. A soft, mean-field potential derived from crystal contacts for predicting protein-protein interactions. J Mol Biol, 1998, 283: 1037—1047[DOI]
20 Gohlke H, Hendlich M, Klebe G. Knowledge-based scoring function to predict protein-ligand interactions. J Mol Biol, 2000, 295: 337—356[DOI]
21 Muegge I, Martin Y C. A general and fast scoring function for protein-ligand interactions: A simplified potential approach. J Med Chem, 1999, 42: 791—804[DOI]
22 Mitchell J B O, Laskowski R A, Alex A, et al. Bleep-potential of mean force describing protein-ligand interactions: I. Generating potential. J Comput Chem, 1999, 20: 1165—1176[DOI]
23 Jiang L, Gao Y, Mao F, et al. Potential of mean force for protein-protein interaction studies. Proteins, 2002, 46: 190—196[DOI]
24 Zhou H, Zhou Y. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. Protein Sci, 2002, 11: 2714—2726[DOI]
25 Godzik A, Kolinski A, Skolnick J. Are proteins ideal mixtures of amino acids? Analysis of energy parameter sets. Protein Sci, 1995, 4: 2107—2117
26 Mirny L A, Shakhnovich E I. How to derive protein folding potential? A new approach to an old problem. J Mol Biol, 1996, 264: 1164—1179[DOI]
27 Thomas P D, Dill K A. Statistical potentials extracted from protein structures: how accurate are they? J Mol Biol, 1996, 257: 457—469[DOI]
28 Ben-Naim A. Statistical potentials extracted from protein structures: are these meaningful potentials? J Chem Phys, 1997, 107: 3698—3706[DOI]
29 Skolnick J, Jaroszewski L, Kolinski A, et al. Derivatlon and testing of pair potentials for protein folding. When is the quasichemical approximation correct? Protein Sci, 1997, 6: 676—688
30 Keskin O, Bahar I, Badretdinov A Y, et al. Empirical solvent-mediated potentials hold for both intra-molecular and inter-molecular inter-residue interactions. Protein Sci, 1998, 7: 2578—2586
31 Singh J, Thornton J M. Atlas of Protein Side-Chain Interactions. Vols. I and II. Oxford: IRL Press, 1992
32 Sippl M J. Helmholtz free energy of peptide hydrogen bonds in proteins. J Mol Biol, 1996, 260: 644—648[DOI]