

# Global Docking Method for Flexible Peptide Segment-Protein Interactions

Ruiying Lai, Bo Wan, Qiang Huang

School of Life Sciences, Fudan University, Shanghai  
Email: [rlai11@fudan.edu.cn](mailto:rlai11@fudan.edu.cn), [huangqiang@fudan.edu.cn](mailto:huangqiang@fudan.edu.cn)

Received: Jun. 3<sup>rd</sup>, 2014; revised: Jun. 9<sup>th</sup>, 2014; accepted: Jun. 13<sup>th</sup>, 2014

Copyright © 2014 by authors and Hans Publishers Inc.  
This work is licensed under the Creative Commons Attribution International License (CC BY).  
<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

Protein-peptide binding plays various important roles in living cells. In many cases, the peptide-binding sites of proteins are not known in prior. Then, computational prediction of the peptide-binding sites is desirable. Popular programs for protein-peptide docking usually depend strongly on the initial positions of peptides, such as Rosetta. To overcome this limitation, here we develop a global docking approach in which the peptide is initially distributed evenly on 26 surface locations of a virtual sphere around the protein, and define a selection parameter for discriminating native-like binding site from non-native sites. We used this approach to predict the native-like binding conformations of peptide-protein complexes, and in most cases the peptide-binding sites were correctly predicted, with  $C_{\alpha}$ -RMSDs below 5.5 Å with respect to the crystal structures of peptides. The results of this study suggested that our approach may be very useful for the identification of peptide-binding sites of proteins.

## Keywords

Protein-Peptide Interaction, Peptide-Binding Site, Molecular Docking

---

# 柔性多肽片段 - 蛋白质相互作用的 全局对接方法

来瑞颖, 万波, 黄强

复旦大学生命科学学院, 上海

Email: [rlai11@fudan.edu.cn](mailto:rlai11@fudan.edu.cn), [huangqiang@fudan.edu.cn](mailto:huangqiang@fudan.edu.cn)

收稿日期: 2014年6月3日; 修回日期: 2014年6月9日; 录用日期: 2014年6月13日

## 摘要

多肽 - 蛋白质的相互作用在生物细胞中发挥着各种各样重要的作用。通常情况下, 它们之间的结合信息是未知的。所以, 利用计算方法预测结合位点具有重要意义。而以Rosetta为代表的常用对接软件通常具有很强的初始位置依赖性。为克服这一局限性, 本研究提出了一种全局对接的方法, 以受体蛋白为球状系统的中心, 将多肽平均地分布在球面26个位置上; 同时定义了一个区分天然结合构象和非天然结合构象的筛选参数。用上述方法预测多肽 - 蛋白质的结合构象, 结果显示该方法能成功预测蛋白质的结合位点, 且多数多肽的预测构象的 $C_{\alpha}$ -RMSD在5.5 Å以下。因此, 研究结果表明, 所发展的方法在蛋白质多肽结合位点预测方面有很好的应用价值。

## 关键词

多肽 - 蛋白质相互作用, 多肽结合位点, 分子对接

## 1. 引言

新药设计通常基于体内重要的蛋白质 - 蛋白质相互作用位点, 其中相当一部分是由柔性多肽所介导的, 在活细胞中, 它们控制了众多至关重要的生理过程[1]。目前, 蛋白质 - 多肽相互作用已经在制药、生物技术(如蛋白质功能检测) [2]、生物标记物、生物传感器[3]以及多肽药物治疗等领域得到广泛应用[4]。因此, 预测和设计多肽 - 蛋白质相互作用对于生物技术和新药发现具有重要的意义。高分辨率的结构解析法已经在多肽 - 蛋白质的晶体结构方面得到了广泛应用, 包括: 核磁共振法和 X 射线衍射法[5]。然而, 蛋白质结晶过程(包括蛋白质纯化和条件选择)仍存在很大的困难和挑战, 而且实验周期长, 花费昂贵。

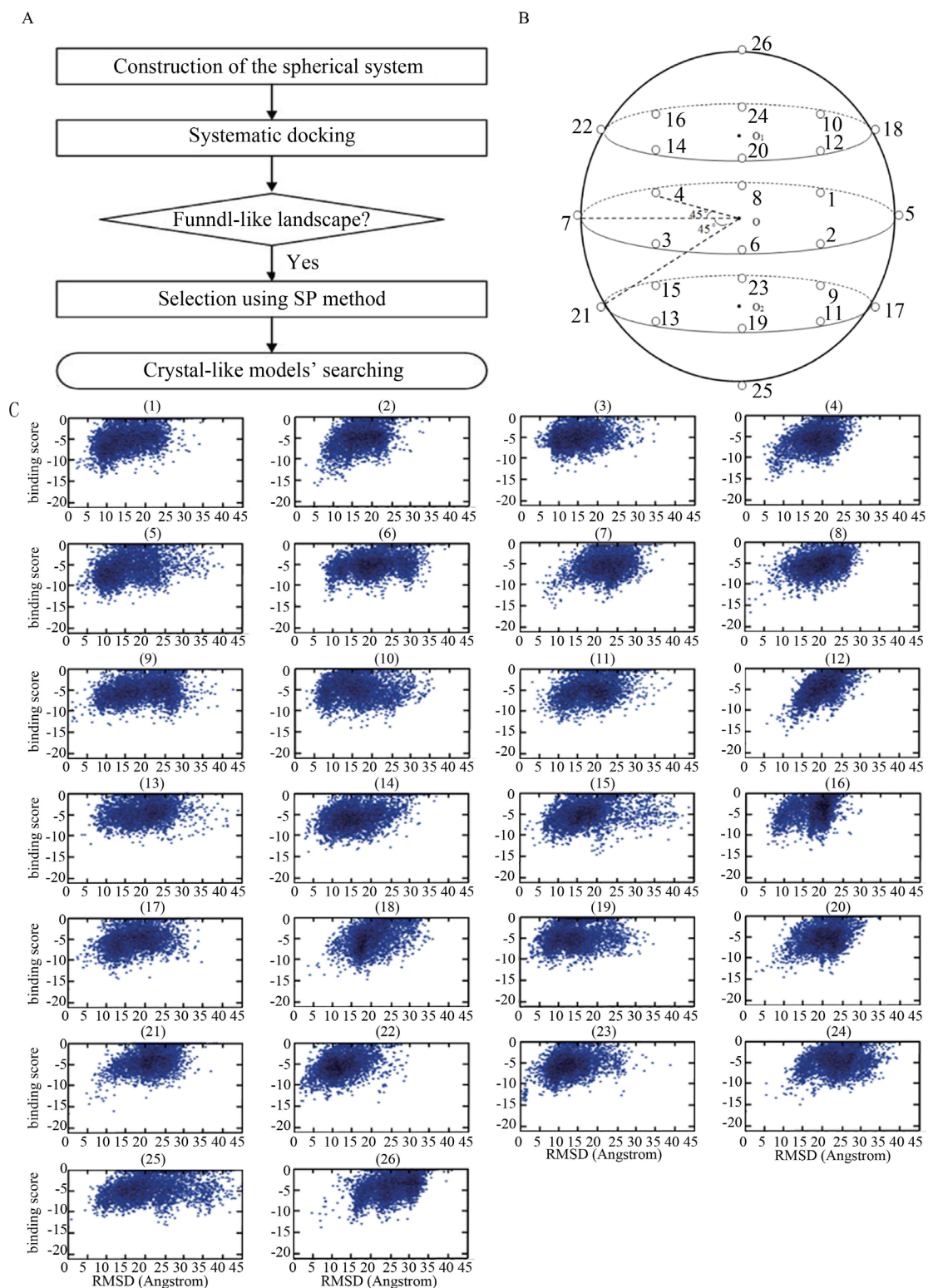
为解决上述问题, 可以通过计算模拟的方法来直接预测蛋白质 - 多肽相互作用。分子对接是预测蛋白质 - 多肽相互作用最重要的计算技术之一。目前, 相关的对接软件包括: AutoDock[6], RosettaDock[7]和 DOCK[8]等等。但是在完全未知互作信息的前提下, 进行蛋白质 - 多肽的相互作用预测尚没有较为系统的方法。AutoDock 是盲对接常用的一种软件, 其所适用的多肽全长上限仅仅为四个[9]。另外一种盲对接方法能够克服上述缺陷, 然而在入选例子中, 互作位点一般均在最大和次大的口袋中[10]。另外, 在许多例子中, 非结合态的蛋白质结构通常没有明显的口袋出现在结合位点附近[11]。

此外, 在计算结构生物学中, 预测多肽的结合构象是最具有挑战性的, 主要因为, 多肽具有很大的自由度[12], 再加上有很多自然状态下的蛋白质并没有稳定的构象[13], 其内部构象也可能会迅速地发生转变[14]。有研究表明, 考虑多肽的柔性能够增加对接预测的准确性[15] [16]。同时, 相关研究显示, Rosetta FlexPepDock 模块能够边实现多肽的从头预测, 边实现对接, 但是必须已知部分互作信息[7]。为了解决上述难题, 我们提出了一种新的盲对接方法, 即在未知互作信息的前提下, 用 Rosetta FlexPepDock 程序实现相互作用位点的系统预测, 其整个研究流程见图 1A。

## 2. 材料与方法

### 2.1. 多肽 - 蛋白质复合物结构的准备

用于对接的蛋白质 - 多肽复合物的结构均来自 PDB (Protein Data Bank)蛋白质库(见表 1)。对于每一



**Figure 1.** (A) Flowchart of the global docking approach for protein-peptide interactions; (B) The sketch map of the systematic docking with 26 initial locations; (C) The *bindingscore-RMSD* landscapes with the example of Calmodulin-MLCK docking pair

**图 1.** (A) 多肽 - 蛋白质相互作用的全局对接流程图; (B) 全局对接系统示意图(26 个初始位置); (C) 以 Calmodulin-MLCK 为例的 *bindingscore-RMSD* 散点图

**Table 1.** Structures of protein-peptide complexes for systematic docking**表 1.** 系统对接的多肽 - 蛋白质复合物结构

PDB ID	Protein	Length(AA)	Peptide	Sequence
2UZT	PKA	336	PKI	TTYADFIAS
4IMY	CycT1 <sup>1</sup>	266	AFF4	LSSRIQSML
1SSH	LAS17	60	SLA1	PPPAMPARP
1T7R	Androgen	269	FxxLF	SSRFESLFA
1CGI	Chymotrypsin A	245	TATI	TGPCRAMIS
3T04	ABL1	123	7C12	EFTVPGYYS
2O5G	Calmodulin	148	MLCK	AVRAIGRLS
1D4T	SLAM	104	CD150	SLTIYAQVQ

<sup>1</sup>表示在实现对接时，将来自 3MI9 的 Tat 多肽也考虑进来，因为有关研究显示，Tat 多肽和 AFF4 也有直接的相互作用[18]。

个结构，首先将其离子、配体以及水分子去掉。另外，如果蛋白质中存在不连续的片段，使用经典建模程序 MODELLER 将其补齐。从蛋白质 - 多肽的复合物结构中，选取连续 9 个氨基酸片段的多肽作为对接多肽片段。原因如下：在 Rosetta FlexPepDock 中，用于对接的多肽平均长度是 9 个[7]，另外，在分析蛋白质序列的疏水性时，其默认的窗口序列长度也是 9[17]。那么，9 个连续的多肽片段可以视为性质较为稳定的结构单元。

## 2.2. 球状对接系统的构建

在未知互作信息的前提下，Rosetta 程序通常不能准确预测蛋白质 - 多肽相互作用位点。为了找到对接的最佳初始位置，通过编写 Python 脚本来实现全局对接系统的构建：以蛋白质为球心，将多肽均匀地放置在球面 26 个初始位置上。Rosetta 程序中的 FlexPepDock 模块可以同时实现对接和从头预测，为简化模型，同一多肽在 26 个初始位置上的构象均设置为相同(图 1B)。

## 2.3. 构象筛选的经验参数法

基于上述构建的系统实现平行对接，并设置相关的参数[19]，每个位置对接产生 5000 个构象，并把这些对接构象按其结合自由能(即结合亲和性)大小进行排序：结合自由能越低，排序越前。在这里，结合自由能值直接使用 Rosetta 经验能量函数值表示，即 Rosetta 程序计算所得的 *binding score*。按常规的配体 - 蛋白对接分析过程[7] [20]，以最低 *binding score* 的构象为参考构象，计算得到其它构象对参考构象的 *RMSD* (均方根偏差)值(参考构象的 *RMSD* = 0)。以构象的 *RMSD* 值为横坐标，其 *binding score* 为纵坐标，获得关于 5000 个对接构象的 *binding score*-*RMSD* 散点图。

为从 26 个初始位置中确定出多肽的最佳对接初始位置，我们对 *binding score*-*RMSD* 散点图进行了深入分析，不断摸索与总结，获得了一个筛选最佳初始位置的经验公式。其过程是：对从某一初始位置出发对接所得的 5000 个构象，把 *binding score* 排在前 20、且其 *RMSD* 值大于 0 且小于 7Å 的构象取出，构成一个构象总数为  $N$  ( $0 \leq N < 20$ ) 的子集，并设子集中最小 *RMSD* 值  $RMSD_{min}$ ，对应构象在子集中 *binding score* 排序为  $Rank_{min}$ 。我们发现，这些从散点图中获得的特征参数存在以下规律： $N$ 、 $RMSD_{min}$  和  $Rank_{min}$  越小，*binding score* 也越小，越有利于筛选出好的初始对接位置。因此，有如下用于筛选最佳初始位置的经验公式：

$$SP(\text{Selection Parameter}) = \frac{-BS_0}{RMSD_{min} \cdot N^2 \cdot \sqrt{Rank_{min}}} \quad (1)$$

上式中的  $BS_0$  是 5000 个构象中的最低 *binding score* 值, 因为 5000 个对接构象中绝大部分的 *binding score* 为负值, 所以这个最小值是小于零的, 为负值。还有, 我们发现上式中的  $N$  对于最佳位置的筛选作用较大, 为强化它的作用, 将其做平方处理; 类似地,  $Rankmin$  对于最终结果作用较小, 将其进行开方处理。

这样, 利用每一个对接初始位置所获得的 5000 个构象, 如果其  $N > 0$ , 可用上式确定计算该位置对应的  $SP$  值, 最大  $SP$  值的位置即为最佳对接初始位置。对  $N = 0$  的位置, 不需计算而直接认定其为非最优位置。另外, 在与晶体结构对比中, 对接能量排在前 10 低的构象当中, 其  $C_\alpha$ -RMSD 在 5.5 Å 以内的构象被视为成功预测构象。

## 2.4. 应用于小分子的 *ISR* 法

*ISR* 的全称即固有专一率, *ISR* 法是一种适用于小分子 - 蛋白质全局对接的定量虚拟筛选法[21], 它表示对接双方的特异性高低。这种 *ISR* 方法适用于小分子化合物与蛋白质的对接[22]。计算公式如下[21]:

$$ISR = \frac{\delta E}{\Delta E} \quad (2)$$

$$\delta E = |E_{\min} - \langle E_D \rangle| \quad (3)$$

$$\Delta E = \sqrt{\langle E_D^2 \rangle - \langle E_D \rangle^2} \quad (4)$$

$\delta E$  指的是能隙, 即最小能量  $E_{\min}$  与构象的平均能量  $\langle E_D \rangle$  的差值,  $\Delta E$  指的是构象的能量波动。

*ISR* 法通过 *ISR-Affinity* 散点图来进行直观展示, 将本研究提出的  $SP$  法与 *ISR* 法进行比较, 进一步证明其可靠性。

## 3. 结果与讨论

### 3.1. Bindingscore-RMSD 散点图的经验参数筛选

为了找出正确的或者近似的初始位置, 基于收敛状散点图(以 Calmodulin-MLCK 为例, 见图 1C), 按照经验公式计算筛选参数。根据经验公式(1), 将符合公式适用条件的初始位置筛选出来, 具体地包括: 位置 2, 4, 8, 10, 12, 18, 25 (PKA-PKI), 位置 7, 9, 10, 12 (CycT1\_Tat-AFF4), 位置 2, 5, 8, 11, 13, 17, 18, 23 (LAS17-SLA1), 位置 4, 5, 7, 10, 14, 15, 19 (Androgen-FxxLF), 位置 2, 6, 12, 16, 17, 18, 19 (Chymotrypsin A-TATI), 位置 1, 2, 3, 4, 5, 8, 11, 12, 15, 16, 18, 23, 24, 26 (ABL1-7C12), 位置 1, 2, 3, 4, 8, 12, 21, 24 (Calmodulin-MLCK), 位置 2, 5, 10, 12, 18, 24, 26 (SLAM-CD150)。对于上述每个对接例子的粗略筛选位置, 计算相应的  $SP$  值, 最大  $SP$  值所在的初始位置即最佳初始位置, 见表 2。结果显示, PKA-PKI 的最佳初始位置是 10 ( $SP$  value = 1.646), CycT1\_Tat-AFF4 的最佳初始位置是 7 ( $SP$  value = 0.800), LAS17-SLA1 的最佳初始位置是 2 ( $SP$  value = 1.487), Androgen-FxxLF 的最佳初始位置是 15 ( $SP$  value = 1.543), Chymotrypsin A-TATI 的最佳初始位置是 19 ( $SP$  value = 1.751), ABL1-7C12 的最佳初始位置是 23 ( $SP$  value = 1.021), Calmodulin-MLCK 的最佳初始位置是 24 ( $SP$  value = 1.142), SLAM-CD150 的最佳初始位置是 26 ( $SP$  value = 1.376)。

### 3.2. 基于初始位置的构象选择

基于筛选得到的初始位置, 计算结合自由能排在前 10 低的构象的  $C_\alpha$ -RMSD 值(与晶体结构相比, 见表 3)。与晶体结构相比, 其中 6 个例子的最终  $C_\alpha$ -RMSD 值在 5.5 Å 以内, 被视为正确的预测构象, 包括: Chymotrypsin A-TATI, Androgen-FxxLF, LAS17-SLA1, Calmodulin-MLCK, ABL1-7C12 和 CycT1\_Tat-AFF4。只有 PKA-PKI 和 SLAM-CD150 ( $C_\alpha$ -RMSD 分别为 7.71 Å 和 8.05 Å) 不在正确预测的范畴内。总之,

**Table 2.** The *SP* value results docking from 26 initial locations of each protein-peptide pair  
**表 2.** 每对蛋白质 - 多肽基于 26 个初始位置的 *SP* 值

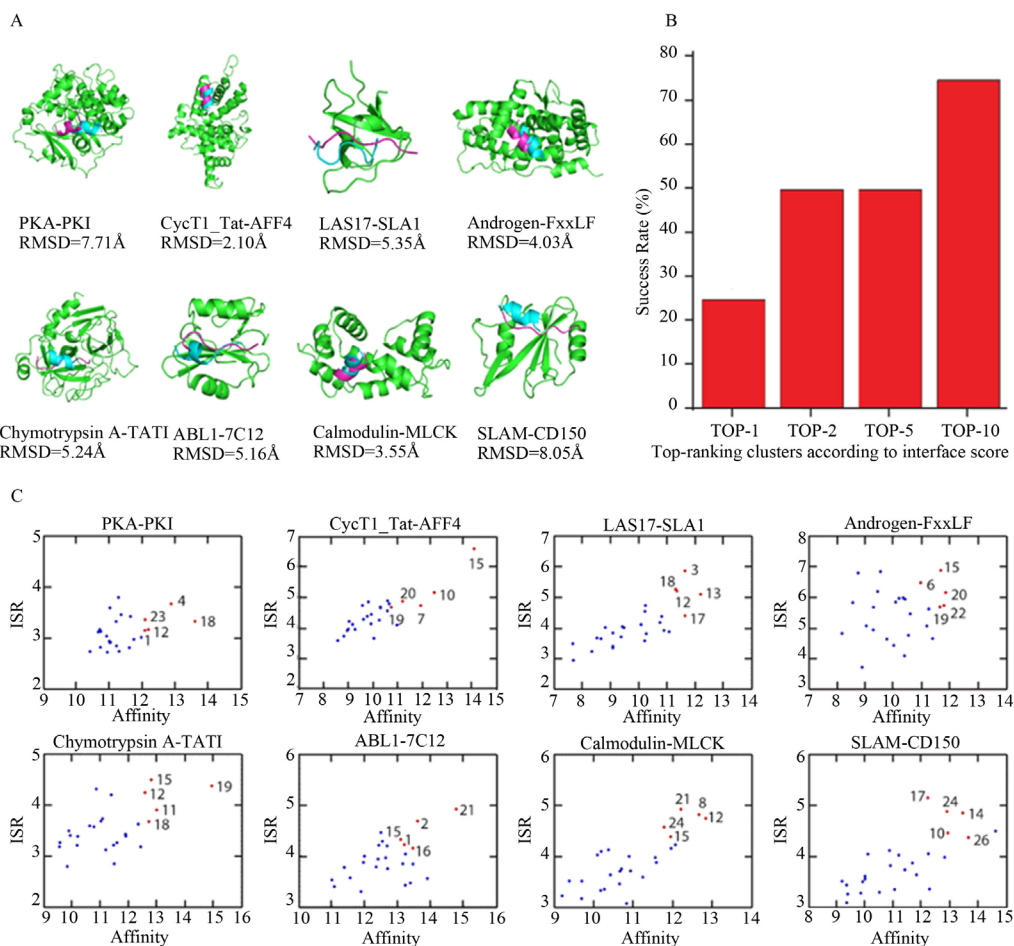
Location	PKA-PKI	CycT1_Tat-AFF4	LAS17-SLA1	Androgen-FxxLF	ChyA-TATI	ABL1-7C12	Calmodulin-MLCK	SLAM-CD150
1	-	-	-	-	-	0.245	0.145	-
2	0.195	-	1.487	-	0.101	0.424	0.200	0.106
3	-	-	-	-	-	0.160	0.030	-
4	0.036	-	-	0.048	-	0.180	0.401	-
5	-	-	0.082	0.068	-	0.027	-	0.156
6	-	-	-	-	0.063	-	-	-
7	-	0.800	-	0.729	-	-	-	-
8	0.027	-	0.183	-	-	0.275	0.149	-
9	-	0.109	-	-	-	-	-	-
10	1.646	0.727	-	0.052	-	-	-	0.440
11	-	-	0.054	-	-	0.111	-	-
12	0.119	0.506	-	-	0.035	0.016	0.027	0.096
13	-	-	0.326	-	-	-	-	-
14	-	-	-	0.077	-	-	-	-
15	-	-	-	1.543	-	0.622	-	-
16	-	-	-	-	0.037	0.025	-	-
17	-	-	0.132	-	0.109	-	-	-
18	0.335	-	0.029	-	1.246	0.142	-	0.059
19	-	-	-	1.066	1.751	-	-	-
20	-	-	-	-	-	-	-	-
21	-	-	-	-	-	-	0.201	-
22	-	-	-	-	-	-	-	-
23	-	-	0.020	-	-	1.021	-	-
24	-	-	-	-	-	0.252	1.142	0.625
25	0.074	-	-	-	-	-	-	-
26	-	-	-	-	-	0.021	-	1.376

**Table 3.** The lowest  $C_{\alpha}$ -RMSD with respect to the crystal structure, binding score and rank of the binding score  
**表 3.** 与晶体结构相比的最小  $C_{\alpha}$ -RMSD 值, binding score 及其排序

PDB ID	The lowest $C_{\alpha}$ -RMSD values ( $\text{\AA}$ ) from the top 10 lowest binding score conformations	Binding score	Rank
2UZT	7.71	-15.1	2
4IMY	2.10	-16.4	1
1SSH	5.35	-12.6	9
1T7R	4.03	-16.0	2
1CGI	5.24	-19.4	2
3T04	5.16	-13.7	10
2O5G	3.55	-16.2	1
1D4T	8.05	-16.1	5

大多数例子的预测结果是成功的, 见图 2A。

统计数据显示(见图 2B), 在结合自由能 TOP1 的构象中, 预测正确率为 25%, 包括 CycT1·Tat-AFF4 和 Calmodulin-MLCK, 对应的  $C_{\alpha}$ -RMSD 值分别是 2.10 Å (Rank1)和 3.55 Å (Rank1)。在结合自由能 TOP2 的构象中, 预测正确率为 50%, 包括 CycT1·Tat-AFF4, Androgen-FxxLF, Chymotrypsin A-TATI, Calmodulin-MLCK, 对应的  $C_{\alpha}$ -RMSD 值分别是 2.10 Å (Rank1), 4.03 Å (Rank2), 5.24 Å (Rank2)和 3.55 Å (Rank1)。在结合自由能 TOP5 的构象中, 预测正确率为 50%, 包括 CycT1·Tat-AFF4, Androgen-FxxLF, Chymotrypsin A-TATI, Calmodulin-MLCK, 对应的  $C_{\alpha}$ -RMSD 值分别是 2.10 Å (Rank1), 4.03 Å (Rank2), 5.24 Å (Rank2)和 3.55 Å (Rank1)。在结合自由能 TOP10 的构象中, 预测正确率在 75%, 包括 CycT1·Tat-AFF4, LAS17-SLA1, Androgen-FxxLF, Chymotrypsin A-TATI, ABL1-7C12, Calmodulin-MLCK, 对应的  $C_{\alpha}$ -RMSD 值分别是 2.10 Å (Rank1), 5.35 Å (Rank9), 4.03 Å (Rank2), 5.24 Å (Rank2), 5.16 Å (Rank10)和 3.55 Å (Rank1)。因此, *SP* 法能够成功识别天然结合构象。当然, 为了能够更加全面地检验该方法的可靠性, 需要更多的例子加以验证。



**Figure 2.** (A) The crystal and predicted conformations. Shown are the native binding pose (magenta) and the final predicted binding pose (blue); (B) The percentages of near-native models of  $C_{\alpha}$ -RMSD values below 5.5 Å with respect to crystal structure in binding score poses of top 1, top 2, top 5 and top 10; (C) *ISR-Affinity* plots of the eight docking pairs

**图 2.** (A) 晶体构象(紫色)和预测构象(蓝色); (B) 结合自由能的排序分别在 top 1, top 2, top 5 和 top 10, 近似天然结合构象(与晶体结构相比,  $C_{\alpha}$ -RMSD 在 5.5 Å 以内的构象)的预测成功率; (C) 八个例子的 *ISR-Affinity* 散点图

**Table 4.** Comparison of *SP* selection results those of *ISR* selection  
**表 4.** *SP* 法与 *ISR* 法比较

PDB ID	<i>SP</i> selected points	<i>ISR</i> selected points
2UZT	10	<i>1</i> , 4, 12, 18, 23
4IMY	7	<u>7</u> , 10, 15, 19, 20
1SSH	2	3, <i>12</i> , 13, 17, 18
1T7R	15	6, <u>15</u> , 19, 20, 22
1CGI	19	<i>11</i> , 12, 15, 18, <u>19</u>
3T04	23	1, 2, <i>15</i> , 16, 21
2O5G	24	8, 12, 15, 21, <u>24</u>
1D4T	26	<i>10</i> , 14, 17, <u>24</u> , <u>26</u>

注：粗体划线的位置表示 *SP* 法与 *ISR* 法的精准匹配，斜粗体位置表示 *SP* 法与 *ISR* 法的模糊匹配。

### 3.3. 与 *ISR* 法的比较

最终结果显示，*SP* 筛选法能够在大多数例子中找到最优构象。其他相似的研究揭示，*ISR* 值能够反映小分子 - 蛋白质对接的特异性。通过对比 *SP* 筛选法与 *ISR* 筛选法，我们找到了两者间存在的合理关系。在 *ISR-Affinity* 的散点图中，处于右上方的点，其 *ISR* 值以及亲和性都是最大的。在 8 个例子当中，*SP* 法选出的点分别在 *ISR-Affinity* 散点图的右上方能够找到准确或者模糊的对应(图 2C, 表 4)。与 *ISR* 法相比，准确预测结果包括：CycT1\_Tat-AFF4(位置 7)，Androgen-FxxLF(位置 15)，Chymotrypsin A-TATI(位置 19)，Calmodulin-MLCK(位置 24)，SLAM-CD150(位置 26)。模糊预测结果包括：PKA-PKI(位置 10 与 *ISR* 法预测的位置 1 相邻)，LAS17-SLA1 (位置 2 与 *ISR* 法预测的位置 12 相邻)，ABL1-7C12 (位置 23 与 *ISR* 法预测的位置 15 相邻)。结果显示，*SP* 法与 *ISR* 法具有 62.5% 的精准匹配率和 100% 的模糊匹配率。此外，与 *SP* 筛选法相比，*ISR* 筛选法得到的结果并不精准，它能够在前 5 个结果中找到较为匹配的位置，无法直接找到最优的位置。因此，*SP* 筛选法具有更大的精准度和优势。

## 4. 结论

本研究提出了一种多肽 - 蛋白质系统盲对接法，将多肽均匀地分布到以蛋白质为中心的球面上实现平行对接，并提出了经验参数筛选法(即 *SP* 法)来筛选预测构象，最终结果显示，多数多肽的预测构象的  $C_{\alpha}$ -RMSD 在 5.5 Å 以下，得以成功识别和预测多肽 - 蛋白质的相互作用，得到成功的预测结合构象。另外已被证明，我们所提出的这种新的研究方法，与适用于小分子 - 蛋白质对接的 *ISR* 法相比，具有高度的一致性，而且可以将预测结果的范围更进一步地缩小。综上，在未知任何互作信息的前提下，利用该方法能够成功预测蛋白质 - 多肽的结合位点，在蛋白质与多肽结合位点的预测方面，所发展的方法将有很好的应用价值。

## 项目基金

上海市重点学科建设项目(B111)。

## 参考文献 (References)

- [1] Petsalaki, E. and Russell, R. (2008) Peptide-mediated interactions in biological systems: New discoveries and applications. *Current Opinion in Biotechnology*, **19**, 344-350.
- [2] Karanicolas, J. and Kuhlman, B. (2009) Computational design of affinity and specificity at protein-protein interfaces.



*Current Opinion in Structural Biology*, **19**, 458-463.

- [3] Hao, J., Serohijos, A., Newton, G., Tassone, G., Wang, Z., Sgroi, D., Dokholyan, N. and Basilion, J. (2008) Identification and rational redesign of peptide ligands to CRIP1, a novel biomarker for cancers. *PLoS Computational Biology*, **4**, e1000138.
- [4] Vlieghe, P., Lisowski, V., Martinez, J. and Khrestchatisky, M. (2010) Synthetic therapeutic peptides: Science and market. *Journal of Neuroscience Methods*, **15**, 40-56.
- [5] Doyle, D., Lee, A., Lewis, J., Kim, E., Sheng, M. and MacKinnon, R. (1996) Crystal structures of a complexed and peptide-free membrane protein-binding domain: Molecular basis of peptide recognition by PDZ. *Cell*, **85**, 1067-1076.
- [6] Morris, G., Goodsell, D., Huey, R. and Olson, A. (1996) Distributed automated docking of flexible ligands to proteins: Parallel applications of AutoDock 2.4. *Journal of Computer-Aided Molecular Design*, **10**, 293-304.
- [7] Raveh, B., London, N., Zimmerman, L. and Schueler-Furman, O. (2011) Rosetta FlexPepDock *ab-initio*: Simultaneous folding, docking and refinement of peptides onto their receptors. *Plos One*, **6**, e18934.
- [8] Shoichet, B. and Kuntz, I. (1993) Matching chemistry and shape in molecular docking. *Protein Engineering Design & Selection*, **6**, 723-732.
- [9] Hetényi, C. and van der Spoel, D. (2009) Efficient docking of peptides to proteins without prior knowledge of the binding site. *Protein Science*, **11**, 1729-1737.
- [10] Aita, T., Nishigaki, K. and Husimi, Y. (2010) Toward the fast blind docking of a peptide to a target protein by using a four-body statistical pseudo-potential. *Computational Biology and Chemistry*, **34**, 53-62.
- [11] Coleman, R. and Sharp, K. (2010) Protein pockets: Inventory, shape, and comparison. *Journal of Chemical Information and Modeling*, **50**, 589-603.
- [12] Dagliyan, O., Proctor, E., D'Auria, K., Ding, F. and Dokholyan, N. (2011) Structural and dynamic determinants of protein-peptide recognition. *Structure*, **19**, 1837-1845.
- [13] Vallee-Belisle, A., Ricci, F. and Plaxco, K. (2009) Thermodynamic basis for the optimization of binding-induced biomolecular switches and structure-switching biosensors. *Proceedings of the National Academy of Sciences India Section B*, **106**, 13802-13807.
- [14] Uversky, V. and Dunker, A. (2010) Understanding protein non-folding. *BBA-Proteins Proteom*, **1804**, 1231-1264.
- [15] Humphris, E. and Kortemme, T. (2008) Prediction of protein-protein interface sequence diversity using flexible backbone computational protein design. *Structure*, **16**, 1777-1788.
- [16] Ding, F., Yin, S. and Dokholyan, N. (2010) Rapid flexible docking using a stochastic rotamer library of ligands. *Journal of Chemical Information and Modeling*, **50**, 1623-1632.
- [17] Chou, S., Upton, H., Bao, K., Schulze-Gahmen, U., Samelson, A., He, N., Nowak, A., Lu, H., Krogan, N., Zhou, Q. and Alber, T. (2012) HIV-1 tat recruits transcription elongation factors dispersed along a flexible AFF4 scaffold. *Proceedings of the National Academy of Sciences India Section B*, **110**, E123-E131.
- [18] Schulze-Gahmen, U., Upton, H., Birnberg, A., Bao, K., Chou, S., Krogan, N. and Zhou, Q. (2013) Building a super elongation complex for HIV. *Elife*, **2**, e00577.
- [19] Bradley, P. (2005) Toward high-resolution de novo structure prediction for small proteins. *Science*, **309**, 1868-1871.
- [20] Zhu, J., Yang, Q., Dai, D. and Huang, Q. (2013) X-ray crystal structure of phosphodiesterase 2 in complex with a highly selective, nanomolar inhibitor reveals a binding-induced pocket important for selectivity. *Journal of the American Chemical Society*, **135**, 11708-11711.
- [21] Yan, Z., Zheng, X., Wang, E. and Wang, J. (2013) Thermodynamic and kinetic specificities of ligand binding. *Chemical Science*, **4**, 2387.
- [22] Yan, Z. and Wang, J. (2012) Specificity quantification of biomolecular recognition and its implication for drug discovery. *Science Report-UK*, **2**, srep00309.