

doi:10.11835/j.issn.1000-582X.2017.03.004

# 多肽序列优化降低 c-Src 蛋白靶向药物的毒性风险

任彦荣<sup>a</sup>, 王强<sup>a</sup>, 李晓波<sup>b</sup>

(重庆第二师范学院 a.生物与化学工程系; b.人事处, 重庆 400067)

**摘要:** 避免 c-Src 蛋白的多肽类拮抗剂与多个蛋白发生混杂性结合, 对于降低抗癌药物的毒性风险具有重要作用。本研究运用生物信息学方法对氨基酸序列进行优化设计, 旨在减少混杂性结合的发生。本研究综合利用各种多肽数据库和生物信息学工具, 首先总结了多肽分子与多个蛋白 SH3 结构域之间潜在的混杂性结合, 并发现了其中的内在规律。随后, 根据所发现的规律, 对多肽的氨基酸序列进行针对性的优化设计。结果表明, 大多数多肽在经过优化后, 所结合的 c-Src 以外的蛋白数量都有所下降, 从而显著提高了多肽与 c-Src 蛋白之间结合的特异性 ( $P < 0.05$ ), 并降低了其潜在的毒性风险。本研究取得的结果将为设计具有高度特异性和低毒性的靶向性多肽药物提供参考。

**关键词:** c-Src 蛋白; SH3 结构域; 多肽; 混杂性结合; 药物毒副作用; 氨基酸序列

**中图分类号:** R978.1

**文献标志码:** A

**文章编号:** 1000-582X(2017)03-034-06

## Reducing the toxicity of c-Src targeting drugs by optimizing amino acid sequences

REN Yanrong<sup>a</sup>, WANG Qiang<sup>a</sup>, LI Xiaobo<sup>b</sup>

(a. Department of Biological and Chemical Engineering; b. Personnel Department, Chongqing University of Education, Chongqing 400067, P.R.China)

**Abstract:** It is important to avoid promiscuous binding between an anticancer peptide and multiple proteins with SH3 domain, so as to minimize the risks of unpredictable toxic and side effects. In the present study, we applied bioinformatics methods to optimize amino acid sequences, in order to reduce the probability of promiscuous binding. Relevant peptide databases and bioinformatics tools were utilized to summarize the rules of promiscuous binding between peptides and SH3 domains. Based on that, we specifically optimized the amino acid sequences of the peptides. The results suggest that most of the modified amino acid sequences exhibit lower level of binding promiscuity, which significantly improves the overall binding specificity and reduces safety risks ( $P < 0.05$ ). This study provides a reference for designing targeted peptide drugs with high binding specificity and low toxicity.

**Keywords:** c-Src protein; SH3 domain; peptide; promiscuous binding; toxic and side effect; amino acid sequence

**收稿日期:** 2016-09-14

**基金项目:** 重庆市科委前沿与应用基础研究计划项目 (cstc2014jcyjA10104); 重庆第二师范学院校级科研创新团队 (KYC-extd03-2017004); 第二批重庆市高等学校青年骨干教师资助计划。

Supported by Front and Applied Fundamental Research Plan of Chongqing Science & Technology Commission (cstc2014jcyjA10104), Program for Innovative Research Team in Chongqing University of Education (KYC-extd03-2017004) and the Second Round of Funding for the Young Key Teachers of University in Chongqing.

**作者简介:** 任彦荣 (1980-), 女, 重庆第二师范学院教授, 博士, 主要从事计算生物化学、制药工程等研究, (E-mail) ren\_sci@sina.com。

癌症是当今医学界所面临的最为严峻的挑战之一<sup>[1]</sup>。癌症的发生及转移涉及到一系列尚未明确的遗传学和分子生物学过程<sup>[2]</sup>,尤其是近年来越来越多的证据表明组织细胞的信号通路失调是导致癌症的关键因素之一,故酪氨酸蛋白激酶等关键信号蛋白作为抗癌药物的潜在靶标愈发受到重视<sup>[3-4]</sup>。

由原癌基因 SRC 编码的 c-Src 蛋白是非受体型酪氨酸蛋白激酶家族的成员之一,该蛋白的功能异常被认为与多种癌症相关<sup>[5-6]</sup>。c-Src 核心区由一个激酶域 CD 及两个调节域 SH2 和 SH3 构成<sup>[7]</sup>。传统 c-Src 蛋白抑制剂的研发主要围绕其激酶域,但多为毒副作用较大的化学药物。同时,c-Src 的调节域 SH3 在 c-Src 的亚细胞定位、底物识别及介导肿瘤细胞结构形态变化中扮演了重要的角色,有望成为抗癌药物的一个新靶点<sup>[8]</sup>。由于 SH3 是典型的肽识别域,专门识别配体的氨基酸片段<sup>[9]</sup>,因此可以设计和多肽分子阻断 c-Src 的正常功能并抑制肿瘤生长。这一策略的优势在于提高药物靶向特异性,相比于传统化疗最大限度地降低了毒副作用<sup>[10]</sup>。

鉴于此,众多研究致力于筛选、设计和优化针对 c-Src 蛋白的多肽类拮抗剂<sup>[11]</sup>。然而,SH3 结构域并非 c-Src 蛋白所独有。人类蛋白组中迄今已发现 300 多个 SH3 结构域<sup>[12]</sup>,其中绝大多数都倾向于识别氨基酸序列中的脯氨酸丰富区域<sup>[13]</sup>。所以一个多肽配体与 c-Src 蛋白结合的同时,也有可能与其他蛋白的 SH3 结构域存在高度的亲和力。所以,候选多肽药物必须与 c-Src 蛋白特异性结合,从而在实现 c-Src 靶向抗癌疗效的同时,尽可能避免与其他蛋白的 SH3 结构域发生混杂性结合,否则有可能干扰与疗效无关的信号通路,引发的不可预期的毒副作用<sup>[14]</sup>。

本研究围绕多肽与 c-Src 酪氨酸激酶 SH3 结构域的特异性结合进行分析,目的在于减少混杂性结合的发生。首先,通过多肽数据库收集已知与 c-Src 蛋白 SH3 结构域存在结合关系的多肽作为研究的原型。而后,基于高效的生物信息学工具发现原型多肽氨基酸序列与结合特异性高低之间存在的规律。最后,基于所发现的规律,对相关原型多肽的氨基酸序列进行了优化,使优化后的多肽在结合特异性方面具有显著提升。

## 1 材料与方法

### 1.1 搜集原型多肽

已知与 c-Src 蛋白 SH3 结构域结合的多肽信息采集自 Molecular INTeraction(简称 MINT)数据库(<http://mint.bio.uniroma2.it/>)。首先,在搜索界面输入 c-Src 蛋白编码基因的标准名称 SRC 后,可以从 MINT 数据库中获得已知与 c-Src 存在相互作用的分子(包括蛋白质、化合物和多肽等),笔者仅保留其中的多肽分子。随后,从以上多肽中进一步挑选出具有第一类脯氨酸丰富区域(class I proline-rich motifs)的多肽分子作进一步分析。第一类脯氨酸丰富区域的特征序列为  $+x\Phi P x\Phi P$ ,其中 P 代表脯氨酸, $x$  代表任意人类蛋白中常见的氨基酸, $\Phi$  代表疏水性氨基酸(包括丙氨酸,异亮氨酸,亮氨酸,蛋氨酸,苯丙氨酸,脯氨酸,色氨酸,缬氨酸和甘氨酸),+ 代表带正电荷的氨基酸(通常为精氨酸或赖氨酸)。

### 1.2 多肽与各蛋白 SH3 结构域结合的模拟计算

所有原型多肽的氨基酸序列首先被转化为 FASTA 格式,而后被输入到(modular domain peptide interaction, MoDPepInt)在线服务器所提供的 SH3PepInt 工具进行计算(<http://modpepint.informatik.uni-freiburg.de/SH3PepInt/>)。SH3PepInt 是一个基于图核方法的 SH3 结构域与多肽相互作用预测工具。通过使用已知的 SH3 与多肽之间非线性相互作用数据对支持向量机模型进行训练,SH3PepInt 工具能够快捷地预测一条多肽与 69 个人类蛋白(其中包括 c-Src 蛋白)的 SH3 结构域是否具有结合潜力。根据 SH3PepInt 的输出结果,可以评估特定多肽分子除 c-Src 蛋白以外,还与哪些其他蛋白具有潜在的相互作用关系,从而衡量该多肽与 c-Src 蛋白结合的特异性。除了 c-Src 以外,若一个多肽与越多的蛋白发生结合,则代表该多肽的结合特异性越低,即结合混杂性越高。

在 SH3PepInt 预测结果的基础上,本研究进一步使用 CABS-dock 在线工具(<http://biocomp.chem.uw.edu.pl/CABSdock>)对特定原型多肽与 SH3 结构域进行一对一的模拟对接。通过输入原型多肽的氨基酸序列和 SH3 结构域在 Protein Data Bank 数据库(即 PDB)的编号,CABS-dock 能够模拟出一系列可能的结合模型。结合模型的准确度以均方根差(root-mean-square deviation)为衡量指标,均方根差越低则代表一系列结合模型的整体准确度越高。



多肽序号	脯氨酸富集模体	混杂性结合的蛋白数量
3	R A L P L P P	3
6	R P L P L L P	9
7	R P L P L P P	10
8	R P L P A L P	13
1	R P L P P V P	14
10	R P L P P C P	17
2	R R L P A L P	18
4	R A L P P L P	18
5	R P L P P L P	25
9	R P L P R L P	25

图 2 原型多肽中脯氨酸丰富区域的第 5 个氨基酸与结合特异性之间存在一定的关联

Fig.2 The fifth amino acid of proline-rich motif is associated with the binding specificity of prototype peptides

### 2.3 多肽序列优化显著提高 SH3 结构域结合的特异性

为了验证以上假设,笔者对除了第 3、第 6 和第 7 号以外的其他原型多肽进行了优化,将其脯氨酸丰富区域的第 5 个氨基酸替换为亮氨酸。优化后的原型多肽被重新输入到 SH3PepInt 工具中与 SH3 结构域进行模拟结合。结果表明,全部原型多肽在经过优化后不仅保留了对 c-Src 蛋白的相互作用,而且大多数所结合的其他蛋白数量也有所下降(见图 3)。笔者将优化后的多肽序列与未优化的原型多肽进行了一对一的比较,并进行了单侧对偶 T 检验(见图 4)。结果表明,氨基酸优化后出现的混杂性结合的下降具有统计学意义上的显著性( $P=0.013$ )。以上数据充分说明,多肽的氨基酸序列与结合特异性之间存在关联,并且这种关联关系能够指导氨基酸序列的优化设计,从而有利于降低候选多肽的毒性风险。

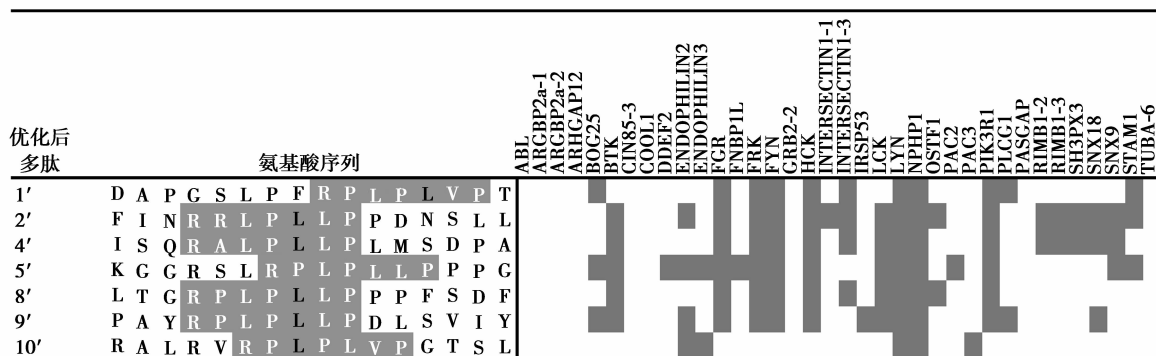


图 3 原型多肽经过氨基酸序列优化,将脯氨酸丰富区域的第 5 个氨基酸替换为亮氨酸后的预测结合靶点

Fig.3 After optimization, the fifth amino acid of proline-rich motif is replaced with Leucine

为了进一步验证以上结果,笔者使用 CABS-dock 工具<sup>[17]</sup>模拟了原型多肽与 SH3 结构域的具体结合构型。例如,第 10 号原型多肽经过序列优化后,SH3PepInt 工具的预测其不再与 HCK 蛋白的 SH3 结构域发生混杂性结合。CABS-dock 模拟结果表明,第 10 号多肽与 HCK 蛋白的 SH3 结构域的结合构型在优化前与优化后存在直观的差异(见图 5(a)、(b))。并且,通过 CABS-dock 计算的均方根差,可以衡量结合模型的准确性。对比优化前和优化后的 10 个可能性最高的结合构型,可以发现优化后的结合模型均方根差数值显著高于优化前(见图 5(c), $P=4.64 \times 10^{-6}$ )。这一差别说明,第 10 号多肽优化后难以与 HCK 蛋白形成稳定的结合构型,表现为均方根差的增大,从而进一步验证了氨基酸序列优化提升结合特异性的结论。

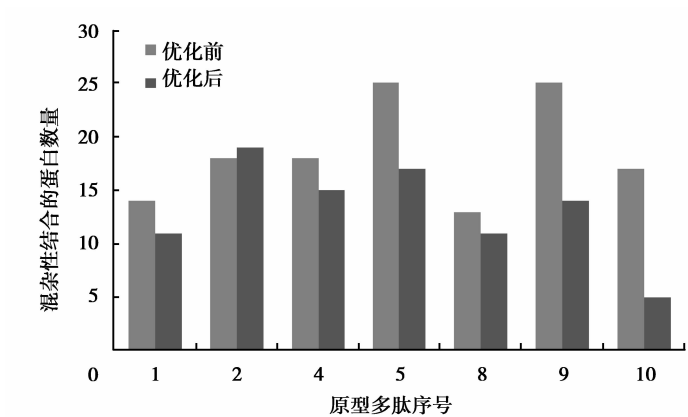
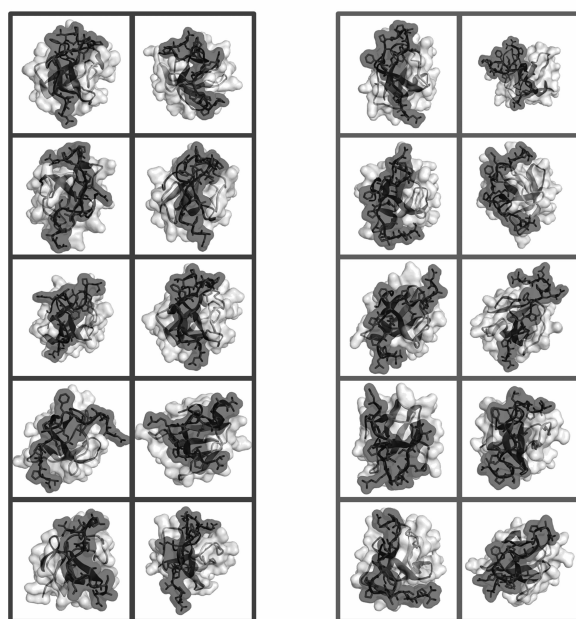
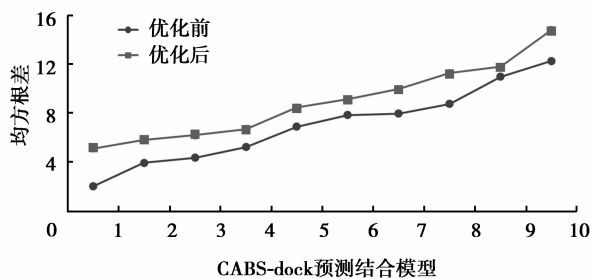


图 4 多数原型多肽(除第 2 号多肽)在优化脯氨酸丰富区域的第 5 个氨基酸后结合特异性得到提升  
 Fig.4 Most of the prototype peptides, except for No. 2, showed higher binding specificity after optimization



(a) 优化前的 10 个可能性最高的结合构型  
 (b) 优化后的 10 个可能性最高的结合构型



(c) 第 10 号原型多肽优化后的结合模型均方根差数值显著高于优化前

图 5 第 10 号原型多肽与 HCK 蛋白的 SH3 结构域的模拟对接结果

Fig.5 The simulated binding between prototype peptide No. 10 and the SH3 domain of HCK protein

### 3 结 语

设计多肽拮抗剂使之特异结合 c-Src 蛋白,且尽可能不与其他蛋白的 SH3 结构域发生相互作用,一直以来都是一项高难度的研究课题。笔者通过分析多条已知与 c-Src 的 SH3 结构域存在结合的原型多肽,发现了多肽序列与蛋白结合特异性之间的规律,从而对原型多肽的氨基酸序列进行一定的优化。计算机模拟的结果表明,优化显著提高了多肽的结合特异性,从而有利于降低毒副作用的风险。

本研究虽然取得了一定的初步结果,但提高多肽与 SH3 结构域结合的特异性仍需要一系列深入的后续工作。现有的结果主要基于计算机模拟数据,还需要经过生物化学实验的验证,最终确定序列优化在何种程度上提高了多肽与 c-Src 结合的特异性,从而为设计更安全、毒性更低的多肽类新药提供指导和依据。

#### 参考文献:

- [1] 邱惠,张艳,雷海科,等.恶性肿瘤危险因素的 Logistic 回归分析[J].重庆医科大学学报,2012,37(3):251-255.  
QIU Hui, ZHANG Yan, LEI Haike, et al. Risk factors related to cancer by Logistic regressive analysis[J]. Journal of Chongqing Medical University, 2012, 37(3):251-255. (in Chinese)
- [2] Koutsogiannouli E, Papavassiliou A, Papanikolaou N. Complexity in cancer biology: is systems biology the answer? [J]. Cancer Medicine, 2013, 2(2): 164-177.
- [3] Takeuchi K, Ito F. Receptor tyrosine kinases and targeted cancer therapeutics[J]. Biological & Pharmaceutical Bulletin, 2011, 34(12): 1774-1780.
- [4] 李霞.抗癌药物研究的新进展[J].中国药房,2001,12(7):429-431.  
LI Xia. The research development of cancer therapy drug[J]. Journal of Practical Oncology, 2001, 12(7):429-431. (in Chinese)
- [5] Ishizawa R, Parsons S. c-Src and cooperating partners in human cancer[J]. Cancer Cell, 2004, 6(3):209-214.
- [6] Guarino M. Src signaling in cancer invasion[J]. Journal of Cellular Physiology, 2010, 223(1):14-26.
- [7] Roskoski R. Src protein-tyrosine kinase structure and regulation [J]. Biochemical and Biophysical Research Communications, 2004, 324(4): 1155-1164.
- [8] Kim L, Song L, Haura E. Src kinases as therapeutic targets for cancer[J]. Nature Reviews Clinical Oncology, 2009, 6(10): 587-595.
- [9] 何菁,夏添松,王水.皮动蛋白与肿瘤侵袭性研究进展[J].中华肿瘤防治杂志,2015,22(1):72-75.  
HE Jing, XIA Tiansong, WANG Shui. Cortactin and tumor invasiveness[J]. Chinese Journal of Cancer Prevention and Treatment, 2015, 22(1):72-75. (in Chinese)
- [10] Craik, Fairlie D, Liras S, et al. The future of peptide-based drugs[J]. Chemical Biology & Drug Design, 2013, 81(1): 136-147.
- [11] Sparks A, Quilliam L, Thorn J, et al. Identification and characterization of Src SH3 ligands from phage-displayed random peptide libraries[J]. Journal of Biological Chemistry, 1994, 269(39): 23853-23856.
- [12] Kärkkäinen S, Hiipakka M, Wang J, et al. Identification of preferred protein interactions by phage-display of the human Src homology-3 proteome[J]. Embo Reports, 2006, 7(2):186-191.
- [13] Yu H, Chen J, Feng S, et al. Structural basis for the binding of proline-rich peptides to SH3 domains[J]. Cell, 1994, 76(5): 933-945.
- [14] Rudmann D. On-target and off-target-based toxicologic effects[J]. Toxicologic Pathology, 2013, 41(2):310-314.
- [15] Licata L, Briganti L, Peluso D, et al. MINT, the molecular interaction database: 2012 update [J]. Nucleic Acids Research, 2012, 40(Database Issue): D857-D861.
- [16] Kundu K, Mann M, Costa F, et al. MoDPepInt: an interactive web server for prediction of modular domain-peptide interactions[J]. Bioinformatics, 2014, 30(18): 2668-2669.
- [17] Kurcinski M, Jamroz M, Blaszczyk M, et al. CABS-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site[J]. Nucleic Acids Research, 2015, 43(W1): W419-W424.