

蛋白质表面功能区的嫁接

梁世德 肖 岚 毛凤楼 江 林 韩玉真 来鲁华*

(北京大学化学与分子工程学院, 物理化学研究所, 北京 100871. * 联系人, Email: lai@linux2.ipc.pku.edu.cn)

摘要 报道了一种蛋白-蛋白相互作用表面位点的嫁接方法. 首先确定被嫁接的配体表面结合区的重要残基, 在骨架蛋白中搜索适合于嫁接这些残基的位点. 基于对映残基的 C_{α} , C_{β} 原子坐标, 将骨架蛋白叠加到配体蛋白上, 评估骨架蛋白与受体蛋白的互补性并留下互补性较好的骨架蛋白. 细调骨架蛋白与受体蛋白的相对位置, 使之具有合理的界面堆积密度. 然后在骨架蛋白选出的位置上移植配体蛋白相应的残基. 另外, 对于与受体有严重碰撞的残基, 或不能形成盐桥的包埋带电侧链, 也予以突变. 突变后的骨架蛋白与受体一起用 Charmm 进行能量优化. 用刚性蛋白-蛋白复合物结构回归出一个适合于嫁接研究的打分函数. 利用打分函数对骨架蛋白的结合自由能进行评估.

关键词 嫁接 叠加 形状互补 打分函数 蛋白质工程

蛋白质工程的长远目标就是要设计出具有任意功能或结构的全新蛋白. 但由于对蛋白质结构与序列关系认识的不足, 目前还远远不能达到这个目标. 另一方面, 医学和生物学上往往需要蛋白质的某一功能却不满意该蛋白的另外一些特点, 如高分子量、抗原性、不稳定, 以及不易折叠等. 这个问题可以通过把蛋白质的功能区嫁接到另外一个合适的骨架蛋白得到解决. 目前, 蛋白质结构数据库(PDB)已积累了几千套蛋白质结构数据, 这些蛋白结构可作为嫁接研究的骨架蛋白.

国际上已发展了几种程序用于蛋白质功能嫁接的研究^[1~4]. 由于受方法本身限制, 这些程序都只能用于嫁接小分子配体结合位点, 例如金属离子结合位点. 蛋白质与蛋白质相互作用在生命过程中的意义不言而喻. 其界面包埋面积往往在 $12 \sim 20 \text{ nm}^2$ 之间, 涉及 $10 \sim 30$ 个氨基酸残基. 另一方面, 蛋白与蛋白相互作用首先要求界面形状基本互补, 但对结合起决定性作用的往往只是其中 $3 \sim 4$ 个残基. 因此, 在嫁接蛋白质相互作用表面区, 骨架蛋白必须有互补的界面, 而且要有合适的位置来移植对结合起决定作用的残基.

1 材料与方法

嫁接算法的前提是已知配体蛋白与受体蛋白的复合物晶体结构. 首先根据定点突变信息结合理论分析确定被嫁接的配体表面结合区的重要残基.

(i) 在骨架蛋白搜索适合于移植配体功能区重要残基的位点组合. 这一步骤根据文献[5]中的集合衰减算法(set reduction)略作改动. 先计算功能区中任意两个重要残基 1, 2, $C_{\alpha 1}-C_{\alpha 2}$, $C_{\alpha 1}-C_{\beta 2}$, $C_{\beta 1}-C_{\alpha 2}$, $C_{\beta 1}-C_{\beta 2}$ 4 个距离. 若在骨架蛋白中也存在两个残基 $1'$, $2'$, 使得相应 4 个距离与残基 1, 2 偏差小于 0.2 nm , 我们认为残基 $1'$, $2'$ 间的几何关系与 1, 2 间的几何关系一致. 若骨架蛋白中存在一套残基组合, 任意两个残基的几何关系都与配体功能表面区中相应的两个残基间的几何关系一致. 骨架蛋白中的这一套残基组合就能模拟功能区中重要残基的几何关系.

由于 Cys, Pro 在维持蛋白结构上的重要性, 它们不被作为突变的对象. 对于 Gly, 如果位

于 Gly 拐角中则不能被突变. 否则可以被突变, 并且产生一个假 Gly C_{β} 原子以便于计算.

(ii) 叠加及几何互补性的评估. 在骨架蛋白中搜索到能模拟配体蛋白表面功能区重要残基的残基组合后, 基于这些对映残基的 C_{α} , C_{β} 原子, 将骨架蛋白叠加到配体蛋白上. 几何互补性的评估参考文献[6]. 文献[6]用格点空间表示蛋白质形状并进行对接研究. 在这里, 蛋白质被放在一个步长为 0.1 nm, 一维格点数为 100 的三维格点空间. 对于受体, 蛋白内的格点赋予数值 1, 蛋白外的格点赋予 0. 对骨架蛋白, 由于认定主链在嫁接前后保持不变, 主链原子占有的格点赋予-10, 侧链原子占有的格点赋予-2, 蛋白表面外的两层格点赋予 1, 其余格点赋予 0. 蛋白互补性表示为骨架蛋白与受体蛋白重叠格点所赋数值乘积的加和. 如果最后得分较高, 留下骨架蛋白作进一步分析.

(iii) 利用界面堆积密度细调骨架蛋白与受体蛋白的相对位置. 由于蛋白质在结合受体前后会有微小的构象变化, 因此在评估互补性时对碰撞的惩罚不能太高. 否则, 即使是配体蛋白的单体结构叠加到它的复合物晶体结构上, 它与受体蛋白的互补分数也有可能是负值. 因此对碰撞的低惩罚是必需的. 但由此带来的后果是骨架蛋白与受体蛋白的相对位置有可能很不合理. 我们用界面堆积密度^[7]进一步细调骨架蛋白与受体蛋白的相对位置. 原子堆积密度定义为原子 van der Waals 体积与实际占有空间之比. 界面堆积密度为界面上原子堆积密度的平均值. 个别原子由于碰撞较厉害堆积密度可能大于 1, 我们把它定义为 1 以减小对平均堆积密度计算的影响. 另外确定充分暴露于溶剂中的原子所占有空间随意性较大. 我们只计算界面上溶剂可及面积低于 0.15 nm² 的原子的平均堆积密度, 溶剂可及面积的计算采用 NACCESS 程序, 溶剂探针半径取为 0.14 nm. 我们先对 PDB 中蛋白-蛋白复合物结构的界面堆积密度进行分析, 发现蛋白界面平均堆积密度分布较为狭窄, 大部分在 0.70 ~ 0.76 之间. 由于这里评估的界面还没有经过优化, 还存在碰撞, 我们把骨架蛋白与受体蛋白界面的允许堆积密度定为 0.70 ~ 0.78. 当堆积密度大于 0.78 时, 骨架蛋白沿着受体界面原子中心指向骨架蛋白界面原子中心的矢量依次位移 0.01 nm, 直至满足要求. 每次位移后计算骨架蛋白与受体蛋白的界面包埋面积, 若包埋面积小于 12 nm², 则放弃该骨架蛋白. 当堆积密度小于 0.70 时, 沿相反方向位移. 位移完毕后, 计算骨架蛋白与配体蛋白相应重要残基的 C_{α} , C_{β} 原子均方根偏差 (RMS), 若 RMS 大于 0.1 nm, 则放弃骨架蛋白. 另外, 若骨架蛋白任一 $C_{\alpha} \rightarrow C_{\beta}$ 矢量与相应配体蛋白 $C_{\alpha} \rightarrow C_{\beta}$ 矢量偏角大于 $\pi/3$, 我们也放弃该骨架蛋白.

(iv) 突变与能量优化. 运用 Quanta 图形软件, 将筛选出来的骨架蛋白上的位点突变成配体蛋白上相应位点的残基. 调整突变后残基的侧链二面角, 使之与受体形成的相互作用尽可能模拟配体蛋白. 另外, 对与受体有严重碰撞的残基, 或不能形成盐桥的包埋带电侧链, 也予以突变. 突变的原则是尽可能维持几何互补, 形成有利疏水相互作用和氢键, 避免不必要的带电原子脱溶剂能量. 突变后的骨架蛋白与受体一起用 Charmm 进行能量优化. 我们加了一项距离限制能量项来模拟配体蛋白与受体的强原子原子相互作用, 限制势函数为

$$E(r) = 10 \times (r - r_0)^2.$$

第 1 步先用 100 步最陡下降法(SD)进行优化. 所有二面角给予 21 kJ · mol⁻¹ · rad⁻² 的限制, 以避免结构破坏. 第 2 步, 去除二面角限制, 用 50 步共轭梯度法(ABNR)优化. 优化采用随距离变化的介电系数.

(V) 结合自由能评估. 我们用经验方法预测骨架蛋白与受体的亲和力. 由于嫁接研究

的前提是主链构象在结合前后保持不变,我们用配体和受体都是刚性蛋白的复合物线性回归结合自由能.回归变量为界面包埋的疏水溶剂可及面积、亲水溶剂可及面积和亲水对.我们定义两个分子间亲水原子(N或O)距离小于0.34 nm时形成一个亲水对.

2 结果

2.1 被嫁接的功能区及骨架蛋白的选择

barnase 是一种 110 残基的胞外核酸酶.它同 89 肽抑制剂, barstar, 具有异乎寻常的高亲和力, $K_d = 1.3 \times 10^{-14}$, 在 25°C 时结合自由能为 -79.1 kJ/mol. PDB 中已存有 barnase-barstar 的复合物晶体结构.运用定点突变技术及双突变技术,界面残基对结合自由能的贡献,以及界面残基-残基相互作用能量也已经清楚^[8].因此 barnase-barstar 是研究蛋白-蛋白相互作用的理想模型.这里把 barstar 结合区嫁接到与之大小相仿的骨架蛋白.文献[8]表明, barstar 结合界面上的 Asp39, Asp35, Tyr29 突变成 Ala, 结合自由能变化分别为 32.2, 18.8, 14.2 kJ/mol. 我们把 Asp39 和 Asp35 同 barnase 的强相互作用归结为它们同 barnase 形成的氢键和盐桥. Tyr29 与 barnase 的强相互作用归结于 3 对相邻疏水原子. barnase-barstar 复合物晶体结构(PDB 代码 1brs) 用共轭梯度法优化 200 步.所有非氢原子都给予 $2\ 100\text{ kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-2}$ 的弹性限制.用优化的结构作为模板,记录强相互作用原子间的距离,用于以后骨架蛋白与受体蛋白能量优化的目标距离.

我们用 PDB 中大小与 barstar 相仿,残基数为 81 ~ 100 的晶体结构或优化的平均核磁结构作为骨架蛋白,共有 128 个.

2.2 搜索的结果、界面残基突变及能量优化

在 SGI O₂ 工作站上经 15.5 h 搜索,共找到 27 个互补分数超过 200 的骨架蛋白.它们中有些是同源蛋白,我们取出互补分数较高的 4 个代表性蛋白作进一步分析.它们是: 巯基蛋白酶抑制剂(1dvc)、质体蓝素(1plb)、转录起始起因子的 DNA 结合结构域(2hts),以及 barstar 的核磁结构(1bta).括号中为 PDB 代码,以下用 PDB 代码代表相应的蛋白.1bta 同它在复合物的结构相比,主链原子 rms 偏差 0.09 nm,这里我们把它作为对照.表 1 列出了搜索的结果.运用 Quanta 图形软件对结果进行分析.除了移植所嫁接的残基,还进行 1dvc 的 D115A, K116M, D119G, 1plb 的 D42G, E83I, K91N, 2hts 的 R251N, R229V 等额外的突变.这些突变主要是避免不利的静电作用,减少脱溶剂能量,并形成有利的疏水作用及氢键.而且,带电残基往往充分暴露于溶剂中,当两个分子靠近时易发生严重碰撞,因此更容易是突变的对象.对骨架蛋白的突变,都是在蛋白表面,而且突变残基的侧链朝外,并不会破坏骨架蛋白的结构.用 Charmm 程序对骨架蛋白和 barnase 的复合物进行能量优化,优化前后骨架蛋白主链原子均方根偏差一般在 0.02 nm 左右.计算优化后骨架蛋白与 barnase 强相互作用的原子间距离与目标距离(见 2.1)的平均偏差.

表 1 骨架蛋白与 barnase 界面概况

PDB	平均界面堆积密度	界面包埋面积/nm ²	互补分数	RMS ^{a)} /nm
1bta	0.75	15.26	586	0.021
1plb	0.71	19.48	545	0.096
1dvc	0.78	20.07	440	0.084
2hts	0.77	17.96	259	0.075

a) 骨架蛋白嫁接位点与相应 barstar 界面重要残基 C_α, C_β坐标的均方根偏差

2.3 结合能评估

我们用 19 个配体和受体都是刚性蛋白的复合物结构回归结合自由能. 这 19 个复合物 PDB 代码是 2ptc, 1tec, 4sgb, 1hbs, 1tpa, 2sni, 4tpi, 3sgb, 1cho, 1brs, 1vfb, 1mlc, 1fdl, 3hfl, 1axi, 1mel, 1dvf, 1nca, 1nmb. 回归得到的方程如下:

$$\Delta G_{cal} = -2.16 S_{pho} + 6.36 S_{phi} - 4.36 N_{pair} - 16.6.$$

S_{pho} 为界面包埋的疏水溶剂可及面积, S_{phi} 为包埋的亲水溶剂可及面积, N_{pair} 为包埋的分子间亲水对数目. 回归的相关系数 r 为 0.91,交叉验证系数 q 等于 0.87. 利用得到的打分函数对 4 个骨架蛋白的结合自由能计算结果如表 2. 从表 2 可以看出 1plb 具有较低的结合自由能, 平均偏差也小, 是理想的骨架蛋白. 图 1 列出 1bta 和 1plb 同 barnase 的优化后结构.

表 2 结合自由能和平均偏差

PDB	结合自由能/ kJ mol^{-1}	平均偏差 ^{a)} / nm
1bta	-93.7	0.003
1dvc	-96.7	0.039
1plb	-81.6	0.024
2hts	-59.4	0.024

a) 骨架蛋白与 barnase 强相互作用的原子间距离同 barstar-barnase 相应原子间距离偏差的平均值

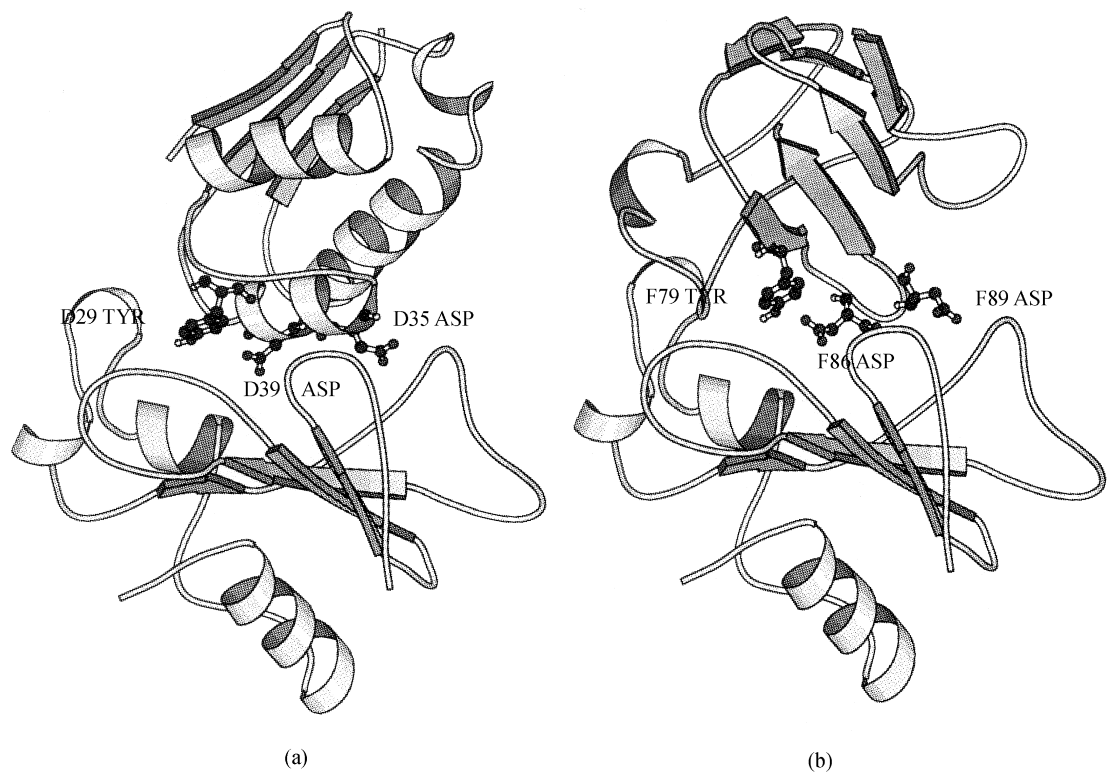


图 1 骨架蛋白和 barnase 的复合物结构
(a) 1bta 和 barnase; (b) 1plb 和 barnase

3 讨论

越来越多的研究者认为,蛋白质的功能由表面区决定,其内核只起骨架作用.我们建立了一套方法用来嫁接蛋白-蛋白相互作用表面区,取得了较好结果.我们的嫁接方法只移植对结合起决定作用的残基.这些残基的一般是大的疏水残基或带电残基.蛋白界面包埋的带电残基往往形成盐桥,若其中一个残基发生突变,另一个残基的巨大脱溶剂能量不能被盐桥的静电作用所补偿,从而大大降低亲和力.不带电的亲水原子则没有这个问题.在搜索骨架蛋白时,我们并不考虑氢键作用.实际上,蛋白界面相互作用并非完美.统计表明,蛋白-蛋白界面上 17.4% 的完全包埋的亲水原子不参与形成氢键^[9].有些氢键的键长,键角并不有利.被包埋的不形成氢键的亲水原子在结果上只是减弱了与之相邻的疏水原子对结合能的贡献.蛋白-蛋白界面巨大的包埋面积使得有利的作用可以累加,从而形成较强的亲和力.因此我们只接受界面包埋的溶剂可及面积大于 12 nm² 的骨架蛋白.

除了嫁接 barstar 的结合部位,我们还做了其他一些算例,得到较好结果.我们把细胞程序死亡调控蛋白 Bcl-x_L 的结合区嫁接到别的骨架蛋白上. Bak 同 Bcl-x_L 的相互作用调控着细胞程序死亡,它们的复合物结构还不知道.从 Bcl-x_L 切下一肽段,仍能结合上 Bak^[10].该肽段与 Bak 的复合物晶体结构表明,肽段呈螺旋状,主要由螺旋一侧的 5 个疏水残基同 Bak 相互作用.但在自由状态,肽段呈无序状.这 5 个残基被作为重要残基,在 PDB 中搜索骨架蛋白.找到的骨架蛋白的嫁接位点也都位于一个长螺旋的一侧,并且骨架蛋白其余部分同 Bak 没有严重的碰撞.由于结合时减少了熵损失,我们期望突变后的骨架蛋白比原来随机肽段对 Bak 具有更高的亲和力.

由于目前还没有可靠的打分函数可以评估蛋白-蛋白亲和力,我们得到的骨架蛋白与受体的复合物也并非处于能量最低态,我们只用简单的打分函数给出初步的定量结果.最后得到的复合物还要借助图形软件进行分析,以决定取舍.完全自动的分子设计程序是不切实际的,经验和直觉的作用也应该得到重视.

致谢 本工作为国家自然科学基金(批准号:29525306)、国家杰出青年基金和国家“攀登”计划预选资助项目.

参 考 文 献

- 1 Hornischer K, Blöcker H. Grafting of discontinuous sites: a protein modeling strategy. *Protein Eng*, 1996, 9: 931 ~ 939
- 2 Hearst D P, Cohen E. GRAFTER: a computational aid for the design of novel proteins. *Protein Eng*, 1994, 7: 1411 ~ 1421
- 3 Hellinga H W, Richards F M. Construction of new ligand binding sites in proteins of known structure. I. Computer aided modeling of sites with predefined geometry. *J Mol Biol*, 1991, 222: 763 ~ 785
- 4 Hellinga H W, Caradonna J P, Richards F M. Construction of new ligand binding sites in proteins of known structure. II. Grafting of a buried transition metal binding site into Escherichia coli thioredoxin. *J Mol Biol*, 1991, 222: 787 ~ 803
- 5 Brint A T, Willett P. Pharmacophoric pattern matching in files of 3D chemical structures: comparison of geometric searching algorithms. *J Mol Graph*, 1987, 5: 49 ~ 56
- 6 Katchalski-Katzir E, Shariv I, Eisenstein M, et al. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc Natl Acad Sci USA*, 1992, 89: 2195 ~ 2199
- 7 Richards F M. The interpretation of protein structures: total volume, group volume distributions and packing density. *J Mol Biol*, 1974, 82: 1 ~ 14
- 8 Schreiber G, Fersht A R. Energetics of protein-protein interactions: analysis of the barnase-barstar interface by single

- mutations and double mutant cycles. J Mol Biol, 1995, 248: 478 ~ 486
- 9 Xu D, Tsai C J, Nussinov R. Hydrogen bonds and salt bridges across protein-protein interfaces. Protein Engi, 1997, 10: 999 ~ 1012
- 10 Sattler M, Liang H, Nettlesheim D, et al. Structure of Bcl-xL-Bak peptide complex: recognition between regulators of apoptosis. Science, 1997, 275: 983 ~ 986

(1999-10-21 收稿, 2000-02-20 收修改稿)