

• 专论与综述 •

蛋白-蛋白作用界面特征及界面预测研究进展^{*}

黎明民 梁世德 张健 郭华荣 张士瑾

(中国海洋大学生命学院 青岛 266003)

摘要: 蛋白-蛋白界面与其余蛋白表面有明显的差别。本文对近年来国内外有关蛋白-蛋白界面几何学、物理学、化学、进化保守性等方面特征的研究概况及应用这些特征对单体中预测界面方法的研究进展予以综述。

关键词: 蛋白-蛋白作用; 蛋白界面特征; 蛋白界面预测

中图分类号: Q6 **文献标识码:** A

Structural Characterisation and Prediction of Protein-protein Interfaces

LI Ming-min, LIANG Shi-de, ZHANG Jian, GUO Hua-rong, ZHANG Shi-jin

Ocean university of china, Qingdao, 266003

ABSTRACT: Protein-protein interfaces have different properties from the other surface. We make a summary of the characteristics of protein interfaces and the methods to predict the interface regions from monomer structures.

Key words: protein-protein interaction; characterisation of protein interfaces; prediction of protein interfaces

蛋白质的功能, 往往是通过和其它蛋白的作用而体现出来。据估计, 自然界中每一个蛋白平均和 5 个蛋白相互作用^[1]。蛋白质在与其它蛋白发生作用时, 主要是通过界面进行的。非界面部分只是起骨架作用, 以稳定界面的构象。由于蛋白-蛋白相互作用界面与生物功能的重要关系, 研究蛋白-蛋白界面的特性, 在单体蛋白中预测有生物功能的界面就显得极为重要。实际上, 蛋白-蛋白相互作用跟蛋白分子的形状、化学组成的互补性、弹性等有很大关系。迄今为止, 已经有大量的关于蛋白-蛋白界面的几何学方面、物理学方面、化学方面、进化保守性和残基能量的理论研究, 成果主要包括以下这些方面:

1 几何学特性:

1.1 界面的面积:

蛋白-蛋白作用界面的大小一般是用溶剂可及面积的减少量, 即形成复合物的两个单体与复合物溶剂可及面积之差, 来定义的。计算溶剂可及面积的程序常用的有基于 Lee & Richards 算法的 NACCESS 程序^[2]和 SurfRace 程序^[3]等。蛋白-蛋白相互作用界面大小的平均值会因为作用类型不同或选用蛋白数据库不同而不同。Argos 等人的研究^[4]发现, 二聚物的相互作用界面平均是整个分子溶剂可及面积的 12%; 三聚物为 17.4%; 四聚物为 20.9%。这些只是平均值, 在不同的情况下可能有很大的变化。例如, 在超氧化物歧化酶二聚物中相互作用的界面为 670 Å², 约为总溶剂可及面积的 9%; 在过氧化氢酶三聚物中界面大小为 10570 Å², 约为总溶剂可及面积的 40%^[5]。Jones & Thornton 的研究^[6]表明, 对于同源二聚物 (Homodimer), 即组成复合物的两个蛋白是相同的, 溶剂可及面积大致和分子的重量线性相关。对蛋白质数据库的全面分析

表明, 同源二聚物的界面大小范围约为 370~4750 Å², 异源二聚物的界面大小范围约为 640~3230 Å²^[7]。大多数情况下, 蛋白-蛋白界面的面积大于 1200 Å²。

1.2 平面度:

界面的平面度通常是指界面原子和其最近平面的偏离程度, 偏离程度用这些原子同最近平面距离的均方根来表示。与其它暴露表面相比大部分的蛋白-蛋白作用界面要更平坦一点, 在同源二聚物中尤其明显, 在部分异源复合物也是这样, 如抗原-抗体复合物^[8,9]。当然, 例外也是存在的, 在某些单体不能稳定存在的复合物 (Obligate complex) 中, 如 γ 干扰素同源二聚物的界面, 由于亚单位存在环绕, 界面完全不平坦。

1.3 形状:

大部分的蛋白-蛋白作用界面是圆形或近似圆形的。Jones & Thornton^[9]用圆形的补缀 (patch) 作为理论界面与生物学界面比较, 在所用数据库的同源二聚体蛋白中, 理论界面与生物学界面的残基重叠程度从 54% 到 87% 不等。

1.4 对称性:

几乎所有同源多聚物的界面对称性都很高。解释这些对称性产生原因的假说有不少: “限制性组装”假说认为, 亚单位形状和相互作用可以对组合方式起挑选作用, 把不利于对称的组合方式排除掉^[10]; Blundell & Srinivasan^[11]认为高度对称的组合方式比不对称的方式在能量上更稳定; Wolynes 推测^[12], 从蛋白折叠的前景来看, 对称蛋白的动力学障碍更低。

1.5 互补性:

几何互补描述的是界面的两个表面间物理学上的合适程度。测量界面间几何互补程度的方法有很多, 现在常用到的方法有: 形状相关指数指标 (shape correlation index)^[13], 间隙指

* 基金项目: 国家自然科学基金: 蛋白-蛋白界面残基能量分析 (No. 30400096)

作者简介: 黎明民, (1979-), 男, 中国海洋大学硕士研究生, 从事结构生物学方面的研究。

山东青岛鱼山路 5 号 K314, 邮编: 266003, Tel: 13869855593, E-mail: proteinli@gmail.com

(收稿日期: 2006-04-06 接受日期: 2006-04-26)

标(gap index)^[6]和堆积密度(packing density)^[14]。Lo Conte等人综合上述三种方法做了研究^[15],发现形状互补程度在低聚物和酶-抑制剂复合物界面中要稍微高于抗原-抗体复合物界面,界面中心的堆积密度要高于界面边缘的堆积密度。

2 理化特性:

2.1 疏水组分分布:

疏水相互作用被认为是稳定蛋白-蛋白相互作用的驱动力^[16]。然而疏水相互作用又被认为对蛋白间相互作用的只有很小的影响^[17]。在那些只能临时存在的复合物(transient complexes)中,相互作用表面和非相互作用界面上疏水性质只有微小的差别^[15];然而,在单体不能稳定存在的永久性复合物(obligate complexes)中,蛋白-蛋白作用界面有明显更疏水的趋势^[18-20]。同时,疏水组分“散在分布”的趋势比集中成大块的趋势更明显^[21]。对于临时存在的复合物,界面倾向于和单体表面最大的疏水块重叠。

界面氨基酸组成的倾向性和疏水组分分布紧密相关,但又不完全相同。在蛋白-蛋白相互作用界面上的氨基酸残基在蛋白的内部和暴露表面也都有发现,与暴露表面相比,相互作用界面富含芳香族氨基酸,如His、Tyr、Phe、Trp;脂肪族氨基酸Leu、Ile、Val、Met的含量也很丰富;带电荷的氨基酸Asp、Glu、Lys的含量极少。含量相对较丰富的带电荷的氨基酸是Arg,氨基酸组成在不同种类的相互作用界面也有很大的区别。在酶和抑制物的复合物中富含Cys;抗原-抗体复合物的界面富含Tyr。但是,普遍富含Arg和缺乏Lys^[15]。

2.2 静电相互作用和氢键:

单体可以稳定存在的复合物的相互作用界面通常有较强的极性。极性组分可以使相互作用具有很强的特异性:潜在的配体必须在阴阳极和作用方向上与受体完全匹配才能结合。这种特异性的优点在于,可以使存在于细胞浆中的亚单位顺利的相互识别。然而,由于脱溶剂作用,极性组分也会阻碍复合物的形成。蛋白-蛋白相互作用需要把极性组分从水中包埋起来,以达到热力学平衡,而极性组分的脱溶剂作用对包埋是不利的^[17,22]。

氢键在单体可以稳定存在的复合物中比在单体不能稳定存在的复合物(或称永久性复合物)中更普遍。盐桥也有类似的现象。在Jones & Thornton研究的数据库中,盐桥只在约半数的同源二聚物(绝大多数是永久性复合物)中出现^[6]。不同种类的蛋白的相互作用界面的氢键出现频率差异很大:在单体不能稳定存在的同源二聚物中,平均每100 Å²氢键出现的频率为0.88;单体能稳定存在的复合物中频率为1.4;抗原-抗体复合物中频率为1.1^[15]。这些平均值实际上并没有体现个体间的显著区别:根据Lo Conte等的研究^[15],在单体可以稳定存在的复合物界面中,氢键的数目从3到50不等,个体之间的差异极大。

3 进化保守性:

长链蛋白质分子经过了长时间的进化。由于长时间的突变、重组、遗传漂变、选择压等,造成蛋白质氨基酸具有多样性的特征。根据中性漂变学说^[23],蛋白质分子产生的功能性突变(位于界面),绝大多数下都是有害的,在选择中被淘汰掉。保留下来的是在选择上呈中性的突变(位于其它表面)^[23]。因此,蛋白质界面残基,特别是那些对结合能有重大贡献的残基,应该比其它表面残基更为保守^[24]。但是界面残基比起非

界面残基,只是稍微有点保守^[25]。很多情况下,界面残基比起其它残基更为保守,仅仅是因为界面刚好也是极其保守的酶活性中心^[26]。

界面残基较低的保守性,可能起源于蛋白-蛋白作用本身并不保守。相对于蛋白质折叠类型(fold),蛋白-蛋白作用种类更为繁多。即使属于同一种折叠类型,不同家族(family)的蛋白质,可能拥有完全不同的作用位点和配体蛋白(partner protein)。如果用已知的所有氨基酸序列来计算保守性,这些位点就不是保守的了。为了克服这个缺点,Lichtarge等^[27]发展了一种能够更好的描述残基保守性的方法,称之为“Evolutionary Tracing (ET)”。只有属于同一种家族的蛋白序列才用来计算残基的保守性。这个方法成功的前提是,能得到同一家族的蛋白序列。但是,目前为同源序列建立进化树的方法,还不够准确。因此,ET方法的用途也是受限制的。另外,我们在应用ET计算的过程中也发现,具有相同蛋白-蛋白作用类型的同一家族蛋白质,在序列上往往极其相似。也就是说,其它表面残基还没有经历中性漂变的积累。在这种情况下,即使ET方法也是无能为力的。

4 残基能量

组成蛋白质的氨基酸残基可以分成两类:一是纯粹结构上需要,二是和功能有关。纯粹结构上需要的残基,采取能量较低的状态,这样天然构象才能稳定。与功能有关的残基,受到功能上的限制,有时必须采取能量较高的状态。Hertzberg and Moult发现,采用罕见骨架构象的残基,往往位于功能区^[28]。Elcock指出静电能较高的残基往往是功能残基^[29]。另外,蛋白-蛋白界面还富含反常暴露的残基和采取反常二级结构的残基^[30]。实际上,蛋白-蛋白作用主要发生在界面区。如果界面残基在单体状态下能量较高,形成复合物时,体系能量可以大幅度下降,从而稳定复合物结构。我们用经验能量函数^[31],分析蛋白质界面残基的能量,并与其它表面残基的能量比较,发现蛋白质界面残基的能量,确实比其他表面残基高^[32]。蛋白表面能量最高的残基,很可能就是对结合起关键作用的残基。最近,David Baker研究小组,用他们原来用于蛋白设计的能量函数,分析蛋白质功能区的残基。他们发现功能残基,往往不是结构上最稳定的残基^[33]。也就是说,功能残基突变成其他残基有可能使蛋白质的结构更为稳定。

5 蛋白-蛋白相互作用界面的预测

在单个蛋白上,特别是非结合单体(unbound monomer)。预测蛋白-蛋白的结合位点具有重要的实际意义。目前,蛋白质的结构主要是通过晶体衍射得到的。由于技术上的原因,单体能稳定存在的临时性复合物的晶体很难得到。迄今为止,蛋白质结构数据库(PDB)中仅有一百个左右代表性的结构。这极大限制了蛋白-蛋白作用的研究。如果能在非结合单体结构表面预测蛋白质的结合位点,会对实验工作起到很大的促进作用。前文论述的蛋白-蛋白界面特征是界面预测的基础。但是无论哪一个特征都不足以预测所有类型的界面。常用的是用线性方法把这些特征组合起来^[34,35],或者通过训练人工智能(比如神经网络)综合利用这些特征^[36-38]。由于临时性复合物晶体结构数量有限,只能用永久性的复合物来训练人工智能,再用临时性复合物检验。一般地说,永久性复合物的单体界面特征更为明显,预测精度也更高。由于各种方法对成功预测的定义不同,再加上有些方法^[35,37]强调

预测精度(specificity),而另一些方法^[36]覆盖率(recovery)较高。客观评价这些方法比较困难,但它们都明显优于随机预测。比如说 Bordner 等人的方法^[36],对 43 个临时的异源复合物的预测精度、覆盖率、和随机预测精度分别是 26%, 57%, 和 12%。

我们研究已经证实,蛋白质界面残基较其余表面残基能量高^[32]。另一方面,界面残基较其余表面残基更为保守。根据进化规律,能量高的残基应该更容易被突变,除非是受到功能上的限制。这样,具有较高能量而且又保守的残基,很可能就是界面残基。基于这个思想,我们把界面残基能量、保守性分数、和残基的界面倾向性分数线性组合起来,用来预测 Neuvieth 等人收集的 57 个临时性复合物的非结合单体^[35]。与他们的结果相比,我们的方法具有类似的精度,和更高的覆盖率。和 Neuvieth 一样,我们假设每一个蛋白只有一个界面。但在实际应用过程中,我们发现许多蛋白都有超过一个的结合位点,这导致预测精度大大降低。今后,这个领域的一个研究方向就是要确定蛋白质结合位点的数目并预测出多个结合位点。

参考文献

- [1] Bork P, et al. Protein interaction networks from yeast to human[J]. *Curr Opin Struct Biol*, 2004, 14(3): p. 292–9
- [2] Hubbard S J., J M Thornton, NACCESS [computer program]. 1993
- [3] Tsodikov O V, M T Record Jr, Y V Sergeev, Novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature[J]. *J Comput Chem*, 2002, 23(6): p. 600–9
- [4] Argos P, An investigation of protein subunit and domain interfaces[J]. *Protein Eng*, 1988, 2(2): p. 101–13
- [5] Jones S, A Marin, J M Thornton, Protein domain interfaces: characterization and comparison with oligomeric protein interfaces[J]. *Protein Eng*, 2000, 13(2): p. 77–82
- [6] Jones S, J M Thornton, Principles of protein–protein interactions[J]. *Proc Natl Acad Sci U S A*, 1996, 93(1): p. 13–20
- [7] Henrick K, J M Thornton, PQS: a protein quaternary structure file server[J]. *Trends Biochem Sci*, 1998, 23(9): p. 358–61
- [8] Porstingl H, et al., Morphological aspects of oligomeric protein structures[J]. *Prog Biophys Mol Biol*, 2005, 89(1): p. 9–35
- [9] Jones S, J M Thornton, Analysis of protein–protein interaction sites using surface patches[J]. *J Mol Biol*, 1997, 272(1): p. 121–32
- [10] Goodsell D S, A J Olson, Structural symmetry and protein function[J]. *Annu Rev Biophys Biomol Struct*, 2000, 29: p. 105–53
- [11] Blundell T L, N Srinivasan, Symmetry, stability, and dynamics of multidomain and multicomponent protein systems[J]. *Proc Natl Acad Sci U S A*, 1996, 93(25): p. 14243–8
- [12] Wolynes P G, Symmetry and the energy landscapes of biomolecules[J]. *Proc Natl Acad Sci U S A*, 1996, 93(25): p. 14249–55
- [13] Lawrence M C, P M Colman, Shape complementarity at protein/protein interfaces[J]. *J Mol Biol*, 1993, 234(4): p. 946–50
- [14] Gerstein M, J Tsai, M. Levitt, The volume of atoms on the protein surface: calculated from simulation, using Voronoi polyhedra[J]. *J Mol Biol*, 1995, 249(5): p. 955–66
- [15] Lo Conte L, C Chothia, J Janin, The atomic structure of protein–protein recognition sites[J]. *J Mol Biol*, 1999, 285(5): p. 2177–98
- [16] Dill K A, Dominant forces in protein folding[J]. *Biochemistry*, 1990, 29(31): p. 7133–55
- [17] Sheinerman F B, R Norel, B Honig, Electrostatic aspects of protein–protein interactions[J]. *Curr Opin Struct Biol*, 2000, 10(2): p. 153–9
- [18] Chothia C, J Janin, Principles of protein–protein recognition[J]. *Nature*, 1975, 256(5520): p. 705–8
- [19] Young L, R L Jemigan, D G Covell, A role for surface hydrophobicity in protein–protein recognition[J]. *Protein Sci*, 1994, 3(5): p. 717–29
- [20] Jones S, J M Thornton, Protein–protein interactions: a review of protein dimer structures[J]. *Prog Biophys Mol Biol*, 1995, 63(1): p. 31–65
- [21] Larsen T A., A J Olson, D S Goodsell, Morphology of protein–protein interfaces[J]. *Curr. Biol.*, 1997, 6: p. 421–7
- [22] McCoy A J, V Chandana Epa, P M Colman, Electrostatic complementarity at protein/protein interfaces[J]. *J Mol Biol*, 1997, 268(2): p. 570–84
- [23] 阎隆飞, 孙之荣. 蛋白质分子结构[M]. 北京: 清华大学出版社, 1999: 279–281
- [24] Hu, Z., et al. Conservation of polar residues as hot spots at protein interfaces[J]. *Proteins*, 2000, 39(4): p. 331–42
- [25] Grishin N V, M A Phillips, The subunit interfaces of oligomeric enzymes are conserved to a similar extent to the overall protein sequences[J]. *Protein Sci*, 1994, 3(12): p. 2455–8
- [26] Caffrey D R, et al. Are protein–protein interfaces more conserved in sequence than the rest of the protein surface? [J]. *Protein Sci*, 2004, 13(1): p. 190–202
- [27] Lichtarge O, H R Boume, F E Cohen, An evolutionary trace method defines binding surfaces common to protein families[J]. *J Mol Biol*, 1996, 257(2): p. 342–58
- [28] Herzberg O, J Moult, Analysis of the steric strain in the polypeptide backbone of protein molecules[J]. *Proteins*, 1991, 11(3): p. 223–9
- [29] Elcock A H., Prediction of functionally important residues based solely on the computed energetics of protein structure[J]. *J Mol Biol*, 2001, 312(4): p. 885–96
- [30] Hoskins J, S Lovell, T L Blundell, An algorithm for predicting protein–protein interaction sites: Abnormally exposed amino acid residues and secondary structure elements[J]. *Protein Sci*, 2006, 15(5): p. 1017–29
- [31] Liang S, N V Grishin, Effective scoring function for protein sequence design[J]. *Proteins*, 2004, 54(2): p. 271–81
- [32] Liang S, et al. Prediction of the interaction site on the surface of an isolated protein structure by analysis of side chain energy scores[J]. *Proteins*, 2004, 57(3): p. 548–57
- [33] Cheng G, et al. Improvement in protein functional site prediction by distinguishing structural and functional constraints on protein family evolution using computational design[J]. *Nucleic Acids Res*, 2005, 33(18): p. 5861–7
- [34] Jones S, J M Thornton, Prediction of protein–protein interaction sites using patch analysis[J]. *J Mol Biol*, 1997, 272(1): p. 133–43
- [35] Neuvieth H, R Raz, G Schreiber, ProMate: a structure based prediction program to identify the location of protein–protein binding sites[J]. *J Mol Biol*, 2004, 338(1): p. 181–99
- [36] Bordner A J, R Abagyan, Statistical analysis and prediction of protein–protein interfaces[J]. *Proteins*, 2005, 60(3): p. 353–66
- [37] Bradford J R, D R Westhead, Improved prediction of protein–protein binding sites using a support vector machines approach[J]. *Bioinformatics*, 2005, 21(8): p. 1487–94
- [38] Chen H, H X Zhou, Prediction of interface residues in protein–protein complexes by a consensus neural network method: test against NMR data[J]. *Proteins*, 2005, 61(1): p. 21–35