文章编号：1004-0374(2009)02-0320-04

B 细胞表位预测方法研究进展

梁 瑾，王靖飞*
（中国农业科学院哈尔滨兽医研究所兽医生物技术国家重点实验室，哈尔滨 150001）

摘 要：B 细胞抗原表位预测方法的研究对基础免疫学的研究及实际应用有着重要的意义。本文归纳了理论预测 B 细胞表位的常用方法，并对目前预测 B 细胞表位方法存在的问题进行了分析。

关键词：B 细胞；表位；预测
中图分类号：R392.1 **文献标识码**：A

Progress in the studies of protein B-cell epitope prediction

LIANG Jin, WANG Jing-fei*
(State Key Laboratory of Biotechnology, Harbin Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Harbin 150001, China)

Abstract: The studies of protein B-cell epitope prediction play a significant role in the theory of immune recognition and the practical application. In the paper, the several efficient methods for B-cell epitope prediction were summarized and the problems with the method of B-cell epitope prediction were discussed.

Key words: B-cell; epitope; prediction

抗体通常特异性地识别抗原蛋白质表面的特定区域，这些区域被称为 B 细胞表位[1]。从空间结构上看，B 细胞表位可分为连续性的线性表位和不连续性的空间构象性表位。线性表位，是由肽链上顺序连续的氨基酸组成；构象性表位，是由那些在空间结构上接近，但顺序上不连续的氨基酸组成，具有高度的空间依赖性。表位之所以区别于非表位是因为其具有自身的特点，Rubinstein等[2]根据PDB中筛选的53个抗原抗体复合物总结了抗原表位区别于非抗原表位的一些特点：(1)表位分布的大小及范围。75% 的表位是由跨越 600 — 1 000Å 面积的 15 — 25 个氨基酸组成。(2)与表位结合的抗体的区域。抗体的互补决定区(complementary determining region, CDR)分析表明，平均 90% 的表位区域的残基是与 CDR 的残基相互作用的。换言之，仅有 10% 的表位区域的残基是与 CDR 外的残基相互作用的。(3)表位富含的氨基酸。表位和非表位区的氨基酸组成有很大的不同，表位富含带电荷的极性氨基酸，而脂肪族疏水性氨基酸在表位出现的频率比较低，酪氨酸和色氨酸在表位出现频率很高，但缬氨酸出现在表位的频率几乎为零。Bogan 和 Thorn[3]称，酪氨酸、色氨酸和带电荷的氨基酸残基之所以多出现在蛋白质与蛋白质相互作用界面是由于它们发生相互作用的能力比较强。(4)表位的二级结构。表位富含环状结构，而缺失螺旋和折叠结构，因为环状结构要比其他形式的二级结构灵活，利于抗体的结合。

表位是蛋白质抗原性的基础，正确而详细地绘制抗原表位图谱不仅有助于基础免疫学地研究，而且对生物活性药物及表位疫苗设计方面也具有重要的指导意义。

确定 B 细胞表位主要有三种方法：X- 射线衍射方法、实验方法和表位预测方法，其中 X- 射线衍射技术是用于测定蛋白质结构最精确的方法[4]，目前 PDB 中已知蛋白质结构的 80% 是通过这种技术得到的，但这种方法耗时长，其所需的单克隆抗体以

收稿日期：2008-11-25；**修回日期**：2008-12-04
***通讯作者**：Tel: 0451-85935090; E-mail: jfwang@hvri.ac.cn

及纯抗原蛋白单克隆抗体的结晶复合物制备都有很大的困难[5]，而且实验的方法比较繁琐，工作量大。B 细胞表位预测方法是人们在总结抗原表位的序列及结构特征的基础上，得出的蛋白质B细胞抗原位点位置的经验规则[6]。应用B细胞表位预测方法可以减少烦琐的实验工作量，提高命中率[7]。

目前的绝大部分B细胞表位预测方法都是从抗原蛋白的一级结构出发，以线性表位预测为主。近几年来，随着PDB 中抗原抗体复合物的增加，对抗原表位的空间特征有了一定的了解，因此构象表位预测方法的研究也取得了一定的进展。

1　线性表位的预测方法

B 细胞表位的预测方法主要集中于线性表位，在20世纪70—80年代发展起来的大量的预测B细胞表位的算法都是基于蛋白质序列。这些算法包括：(1)蛋白质的亲水性方案(hydrophilicity)，以Hopp 和 Woods[8]方案为代表，此方案认为蛋白质抗原各氨基酸残基可分为亲水残基和疏水残基两类。在机体内，疏水性残基一般埋在蛋白内部，而亲水性残基位于表面，因此蛋白的亲水部位与蛋白抗原表位有密切的联系。(2)可及性方案(accessibility)，如Janin可及性参数，指蛋白质抗原中氨基酸残基被溶剂分子接触的可能性[9]。它反映了蛋白质抗原内、外各层残基的分布情况。(3)蛋白质可塑性预测方案(flexibility)认为蛋白抗原构象不是刚性不变的，其多肽链骨架有一定程度的活动性，活动性强的氨基酸残基即可塑性大的位点，易形成抗原表位[10]。(4)蛋白质二级结构预测方案(secondary structure)，认为蛋白质二级结构分析与蛋白质表位的分布关系密切，α 螺旋、β 折叠结构规则，形态固定，常处于蛋白质的内部，难以与抗体嵌合，而β转角和无规则卷曲多处于蛋白质的表面，结构突出，有利于与抗体嵌合，成为抗原表位的可能性大[11]。(5)蛋白质抗原性方案(antigenicity)，即Welling等[12]通过对20个已研究得很透的蛋白质的69个连续位点的606个氨基酸统计分析，用各氨基酸残基在已知B细胞表位中出现的百分率与其通常在蛋白质中出现的百分率比值的对数建立了抗原性刻度，并以此计算蛋白中各亚序列的抗原性。这些方法的代表软件有PEOPLE[13]、PREDITOP[14]、BEPITOPE[15] 、BcePred[16]等；但是最近Blythe和Flower[17]对氨基酸的性质与线性表位的关系做了一个评估，结果表明基于氨基酸序列信息来预测线性表位，即使很好地结合了氨基酸的各种性质，其预测结果仅略强于随机预测。近年来，一些应用隐形马尔可夫模型(HMM)、人工神经网络(ANN)、支持向量机算法(SVM)及其他技术的机器研究方法[18]已经被引入来预测B细胞表位，取得了较好的结果。代表软件有ABCpred[19]、BepiPred[20]、APP[21]等。ABCpred采用人工神经网络来预测线性表位，从Bcipep和SwissProt数据库中提取非冗余的表位肽和非表位肽作为训练集，采用5-折交叉验证，预测敏感性约为67%，特异性约为64%。BepiPred结合氨基酸的性质(亲水性、柔韧性、可及性、极性、暴露表面、转角)和隐形马尔可夫模型来预测线性表位，预测结果表明，同那些仅依赖于氨基酸性质的预测方法相比，BepiPred预测结果的准确性有一定程度的提高。Chen等[21]发现氨基酸通常成对出现在抗原表位的频率要比其出现在非表位肽段的频率高，基于此，并联合支持向量机算法建立了APP方法。应用此方法在872个表位肽和872个非表位肽数据集中，采用5-折交叉验证，预测准确度为71%。EL-Manzalawy等[22]采用同一数据集对这三种方法进行比较，结果表明ABCpred预测表位的准确性略高于BepiPred及APP。

2　构象表位的预测方法

目前，绝大多数B细胞表位预测方法都是基于蛋白质的一级或二级结构，但这些方法只能用来预测由连续的氨基酸残基构成的线性表位，而基于蛋白质的三级结构来预测构象表位的方法比较少，这是因为各种抗原的构象表位可获得的数据要远远少于线性表位，并且到目前为止，几乎没有哪个抗原的所有表位都能够彻底地研究清楚[23]。

2.1　基于蛋白质三级结构来预测构象表位的方法

CEP[24](conformational epitope prediction)是第一个以抗原蛋白的三级结构PDB 文件作为输入条件，以构象性表位预测为主要目的的网上免费服务软件。它提供了一个预测构象表位的web界面，这种方法除了能够预测构象表位，同时也能预测线性表位，该软件主要根据氨基酸残基的溶剂可及性及空间距离截值来预测表位，其公布的预测精度达75%。

DiscoTope[23]是通过蛋白质三级结构数据来预测构象表位的一种新方法，这种方法通过对X射线晶体衍射确定的76个抗原抗体复合物所组成的构象表位数据集进行大量统计度量、空间特征分析和表面可及性计算，对B细胞构象性表位进行预测，最终

对组成蛋白序列的每个氨基酸打分，通过分值来反映某一氨基酸成为表位的可能性，并提供了阈值来确定组成表位的氨基酸残基。

2.2 预测蛋白质与蛋白质相互作用位点的方法 除以上两种方法之外，还有最近发展起来的一些预测蛋白质与蛋白质相互作用位点的方法。由于抗原抗体之间的相互作用属于蛋白质与蛋白质之间相互作用中的一种，因此，可以参考这些方法来预测B细胞表位。

分子对接主要用来研究分子间的相互作用与识别，进而预测复合物结构。一般情况下，对接包括四个阶段：搜索受体与配体分子的结合模式；利用生物学信息或简单的分子表面几何互补性等评价标准过滤排除不合理的结构；对获得的结构进行能量优化，允许氨基酸残基侧链和骨架的运动；用精细的打分函数评价、排序对接模式并挑选近天然构象[25]。常用的分子对接软件有ZDOCK[26]、DOT[27]、DOCK[28]、ClusPro[29]等。其中ClusPro是一个提供网上服务的分子对接软件，其能够根据形状互补快速地筛选ZDOCK和DOT程序产生的对接结果，并对对接结果聚类，根据聚类情况对对接结果打分，最终返回10个得分最高的对接结果，再根据这些对接结果来确定蛋白质相互作用的位点。

PPI-Pred[30](protein-protein interface prediction)将支持向量机的方法同曲面分析结合在一起预测蛋白质相互作用位点。

ProMate[31](predicting the location of potential protein-protein binding sites for unbound proteins)是将一些蛋白质相互作用界面的重要性质综合起来预测蛋白质相互作用位点。这些性质包括结合位点通常偏向位于β片层及非结构的链、芳香族氨基酸的侧链常会参与蛋白质与蛋白质的相互作用、疏水氨基酸和极性氨基酸常聚集在蛋白质与蛋白质相互作用的界面，以及在晶体结构中结合位点的周围有更多的水分子与之结合。

Ponomarenko和Bourne[18]采用以上几种方法预测构象表位并使用同一评估体系对其进行了比较，结果表明，这些方法的准确性均未超过40%，如果用ROC[30](relative operating characteristic)曲线下面积的值来评估这些方法，则DiscoTope和PPI-PRED的值大约是0.6，ClusPro的值高于0.65，但未超过0.7，而其他的方法接近于随机预测。

尽管这些年来B细胞表位预测的方法得到了一定的发展和应用，但这些研究方法还存在一定的问题。首先，所有预测表位的方法都缺乏标准的ROC[32]评估，这使得各种预测方法的结果难以比较与评估；其次，大多数预测线性表位的方法都具有一定的局限性，它们仅仅是根据少数的几个表位的特征(氨基酸的性质、残基的表面可及性、空间分布、分子间接触)来预测表位[18]，而最近对各种线性表位预测方法进行评估的结果表明，仅根据氨基酸的性质来预测线性表位的方法并不可靠。要想提高预测的准确性，需将更多表位区别于非表位的特征结合起来预测；最后，目前预测表位的方法大多数是针对于线性表位的，而据Barlow等[33]研究表明，90%以上的表位为构象表位，因此，在进一步完善线性B细胞表位预测研究的基础上，从蛋白质的三级结构入手，深入对构象性B细胞表位预测算法与程序的研究[7]。同时，我们也相信随着PDB数据库中抗原抗体复合物地增加，能够对各种抗原的构象表位进行更广泛的分析，人们对蛋白质抗原表位的研究将更加透彻。

[参 考 文 献]

[1] Schlessinger A, Ofran Y, Yachdav G, et al. Epitome: database of structure-inferred antigenic epitopes. Nucleic Acids Res, 2006, 34: 777-80

[2] Rubinstein ND, Mayrose I, Halperin D, et al. Computational characterization of B-cell epitopes. Mol Immunol, 2008, 45(12): 3477-89

[3] Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. J Mol Biol, 1998, 280(1): 1-9

[4] Scott JK, Smith GP. Searching for peptide ligands with an epitope library. Science, 1990, 249(4967): 386-90

[5] 黄艳新, 鲍永利, 李玉新. 抗原表位预测的免疫信息学方法研究进展. 中国免疫学杂志, 2008, 24(9): 851-67

[6] 孙建宏, 曹殿军. 细胞的抗原表位研究方法. 动物医学进展, 2004, 25(5): 18-21

[7] 黄 健, 郭建巍, 蔡美英. B细胞表位预测. 微生物学免疫学进展, 2004, 32(1): 40-2

[8] Hoop TP, Woods KR. Prediction of protein antigenic determinants from amino acid sequences. Proc Natl Acad Sci USA, 1981, 78(6): 3824-8

[9] Rudolph R, Tschesche H. Modern methods in protein and nucleic acid research[M]. New York: Walter de Gruger Berlin, 1990: 231

[10] Karplus PA, Schulz GE. Prediction of chain flexibility in proteins. Immunology, 1985, 72(2): 212

[11] 来鲁华. 蛋白质的结构预测与分子设计[M]. 北京:北京大学出版社, 1993: 49

[12] Welling GW, Weijer WJ, Van der Zee R, et al. Prediction of sequential antigenic regions in proteins. FEBS Lett, 1985, 188(2): 215-8

[13] Alix AJ. Predictive estimation of protein linear epitopes by using the program PEOPLE. Vaccine, 1999, 18(3-4): 311-4
[14] Pellequer JL, Westhof E. PREDITOP: a program for antigenicity prediction. J Mol Graph, 1993, 11(3): 204-210, 191-2
[15] Odorico M, Pellequer JL. BEPITOPE: predicting the location of continuous epitopes and patterns in proteins. J Mol Recognit, 2003, 16(1): 20-2
[16] Saha S, Raghava GP. BcePred: prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. Lect Notes Comput Sci, 2004, 3239: 197-204
[17] Blythe MJ, Flower DR. Benchmarking B cell epitope prediction: underperformance of existing methods. Protein Sci, 2005, 14(1): 246-8
[18] Ponomarenko JV, Bourne PE. Antibody-protein interactions: benchmark datasets and prediction tools evaluation. BMC Struct Biol, 2007, 7: 64
[19] Saha S, Raghava GP. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins, 2006, 65(1): 40-8
[20] Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. Immunome Res, 2006, 2: 2
[21] Chen J, Liu H, Yang J, et al. Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. Amino Acids, 2007, 33(3): 423-8
[22] EL-Manzalawy Y, Dobbs D, Honavar V. Predicting linear B-cell epitopes using string kernels. J Mol Recognit, 2008, 21(4): 243-55
[23] Andersen PH, Nielsen M, Lund O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. Protein Sci, 2006, 15(11): 2558-67
[24] Kulkarni-Kale U, Bhosle S, Kolaskar AS. CEP: a conformational epitope prediction server. Nucleic Acids Res, 2005, 33(Web Server issue): W168-71
[25] 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): 616-21
[26] Chen R, Li L, Weng Z. ZDOCK: an initial-stage protein-docking algorithm. Proteins, 2003, 52(1): 82-7
[27] Shoichet BK, Kuntz ID. Protein docking and complementarity. J Mol Biol, 1991, 221(1): 327-46
[28] Mandell JG, Roberts VA, Pique ME, et al. Protein docking using continuum electrostatics and geometric fit. Protein Eng, 2001, 14(2): 105-13
[29] Comeau SR, Gatchell DW, Vajda S. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. Bioinformatics, 2004, 20(1): 45-50
[30] Bradford JR, Westhead DR. Improved prediction of protein-protein binding sites using a support vector machines approach. Bioinformatics, 2005, 21(8): 1487-94
[31] Neuvirth H, Raz R, Schreiber G. ProMate: a structure based prediction program to identify the location of protein-protein binding sites. J Mol Biol, 2004, 338(1): 181-99
[32] Swets JA. Measuring the accuracy of diagnostic systems. Science, 1988, 240(4857): 1285-93
[33] Barlow DJ, Edwards MS, Thomton JM. Continuous and discontinous protein antigenic determinants. Nature, 1986, 322(6081): 747-8