

文章编号: 1004-0374(2010)09-0946-06

构象性 B 细胞表位预测的免疫信息学方法及其网络资源

欧阳玉梅

(伊犁师范学院文理系计算机教研室, 奎屯 833200)

摘要: 抗原表位预测是免疫信息学研究的重要方向之一, 可以给实验提供重要的线索。B 细胞表位或抗原决定簇是抗原中可被 B 细胞受体或抗体特异性识别并结合的部位。实际上, 近 90% 的 B 细胞表位是构象性的。即使抗原蛋白质三级结构已知, B 细胞表位预测仍然是一大挑战。该文结合实例阐述当今主要的构象性 B 细胞表位预测方法和算法: 机器学习预测、非机器学习的计算预测、基于噬菌体展示数据的识别方法, 以及一些也可用于构象性 B 细胞表位预测的通用蛋白质-蛋白质界面预测方法; 介绍最新相关预测软件和 Web 服务资源, 说明未来的研究趋势。

关键词: 抗原; 抗体; 表位; 预测; 构象性

中图分类号: R392.12; R392-3 **文献标识码** A

An introduction to the immunoinformatics methods for conformational B-cell epitope prediction and its web-based resources

OUYANG Yu-mei

(Department of Liberal Arts & Science, Yili Normal University, Kuitun 833200, China)

Abstract: Accurate prediction of epitopes is an important goal of immunoinformatics, which can give important clues to experiments. B-cell epitopes or antigenic determinant is a part of an antigen recognized by either a particular antibody molecule or a particular B-cell receptor of the immune system. Up to 90% of B-cell epitopes are conformational in nature. Even when the tertiary structure of the antigen is available, the accurate prediction of B-cell epitopes remains challenging. First, this review illustrated current prediction methods and algorithms for the conformational B-cell epitopes by describing some examples, including machine-learning approaches, computational prediction methods independent of machine-learning approach, Phage-displayed data based approaches, along with some universal recognizing approaches of protein-protein interface prediction. Next, it presented the latest software packages and web services available online that relate to conformational B-cell epitope prediction, and projected future developments.

Key words: antigen; antibody; epitope; prediction; conformational

B 细胞表位或抗原决定簇是抗原中可被 B 细胞受体或抗体特异性识别并结合的部位。许多免疫检测和免疫治疗学应用是基于抗体识别表位的能力。表位上特定残基及其位置的确定是这些应用的关键起始步骤。确定 B 细胞表位的实验方法费时费力, 而且并不是总能成功。因此, 免疫信息学的重要目标之一就是利用计算方法预测 B 细胞表位。这个问题的解决, 即使是近似解决都将有助于设计实验和

提高研究效率。

B 细胞表位可以分为两类: 线性(连续)表位和构象性(不连续)表位。前者是肽链上一段连续的氨基酸残基序列; 后者是一级结构上不完全连续但在折叠的蛋白质三维结构表面上密切接近的一些氨基

收稿日期: 2010-03-09; 修回日期: 2010-05-10

通讯作者: E-mail: fang9352@sina.com

酸残基。Greenbaum等^[1]研究表明,在B细胞表位中只有很少一部分是线性的,而近90%是构象性的。线性表位预测可以直接用于疫苗设计和免疫诊断。构象性表位预测的目标则是设计能够模拟表位结构和免疫原性的分子,这种分子在抗体产生的过程中可以替代抗原(作为预防或治疗的疫苗),或在医学诊断或实验研究的抗体检测过程中替代抗原。线性B细胞表位预测已经历了20多年的发展,主要是使用依赖于序列的方法来构建免疫原性肽段集,目前已有多种工具,但预测精度仍有很大的提高空间;而构象性B细胞表位在没有蛋白质三维结构知识的情况下,仅从功能分析很难预测和识别。事实上,抗原和抗体相互作用是通过空间结构互补性而非共价结合形成,两者距离越小相互作用力(静电作用、氢键、疏水基和范德华力)越大,而且结合时可能互相形变。因此,即使抗原蛋白质的三级结构已知,表位预测仍然是一大挑战。鉴于过去已知晶体结构的抗原-抗体复合物数量较少、抗原-抗体结合机理研究不够清楚以及预测算法设计难度大等原因,这项研究一直以来进展十分缓慢。最近几年,随着免疫信息学时代的来临^[2]和生物学技术的发展,大量原始数据的积累为计算预测及预测系统性能评价提供了基本的素材^[3]。考虑抗原蛋白质的三维结构等信息,采用计算预测或计算预测和生物实验数据结合的免疫信息学方法进行构象性表位分析和定位已成为目前研究的热点,并呈现出了快速发展的趋势^[4]。

1 免疫信息学预测方法

人们通过观察发现一些蛋白质的氨基酸序列或结构特征与抗原表位有关。在线性表位预测中,通常从抗原蛋白质的氨基酸一级结构出发,综合抗原蛋白质的理化性质、结构特点、统计显著性度量等指标进行表位预测。目前已发展出氨基酸倾向性标度(propensity scale)方法、机器学习方法等^[5]。氨基酸倾向性标度方法中具有较好预测效果的有:亲水性方案、溶剂表面可及性方案、抗原性方案、可塑性方案、二级结构预测方案和电荷分布方案等。其基本思想是首先给每一个氨基酸残基赋予一个分值,然后按一定窗口宽度(通常为7~13个残基)对序列曲线进行光滑,如果某个曲线峰值所在片段的平均指标超过某个指定阈值(通常为整个抗原序列的均值),则该片段很可能是B细胞表位。这些方案

预测准确度一般低于60%。机器学习方法中的人工神经网络、隐马尔可夫模型、支持向量机、基于距离度量的K近邻、决策树、贝叶斯分类器等已被应用于预测线性B细胞表位。

刚刚兴起的构象性表位预测通常涉及表面残基定义、表位残基定义、基准数据集构建、算法设计与性能评价。目前有机器学习预测、非机器学习的计算预测、基于噬菌体展示实验数据的识别方法,以及一些最近发展起来的通用蛋白质-蛋白质界面(也称结合位点或接触面)预测方法。它们多数是基于结构的,有的联合了序列数据,有的结合了噬菌体展示数据;一些方法使用了训练集(已验证了的表位数据集),而有些方法则没有使用。这些方法一般需要输入抗原的三维结构,对于那些未知三维结构的抗原可以采用计算机预测技术得到。输出主要是被预测为表位的各残基及其位置、表位的残基数目。不同方法的输出形式不尽相同,有的反映抗原蛋白质表面残基成为蛋白质表位的概率得分;有的则预测由表面残基组成的表面块(surface patch)是否为表位,即每个残基是否是表位表面块的一部分;有的兼有输出前两者的功能;有的还提供了3D结构的可视化图形,实现了表位作图。而设计出的软件工具,多数既能够预测线性表位也能够预测构象性表位。

1.1 机器学习预测方法

机器学习方法的主要思想是将预测问题转化为分类问题。主要有两种学习方式:无监督学习和有监督学习。无监督学习,如聚类,是数据挖掘的一种基本手段,其目的是提取数据中隐含的类结构。由于类不是预先定义的,需根据实际数据的数学特征按照数据之间的相似性来定义。有监督学习则使用已知数据源训练分类器区别正、负样本,再用构造好的分类器对待识别的样本进行分类决策。已有多种有监督学习用于预测线性B细胞表位,而把它们用于预测构象性B细胞表位是最近才刚刚出现的。建立这样的预测系统通常包括基本步骤:(1)数据集的构建;(2)表位特征提取;(3)分类器的设计与性能评价。下面介绍的方法前两个采用的是无监督学习(聚类),第三个采用的是有监督学习。

CEP是一个利用抗原蛋白质的三维结构预测构象性表位和线性表位的在线服务系统^[6]。该方案没有使用训练集。首先,系统为输入的抗原结构中的每一个氨基酸残基计算相对可及表面积(残基在单体

中的可及表面积与其总的可及表面积的比值); 确认抗原表面结构上具有高溶剂可及性的肽段。然后, 按它们之间的空间距离低于一个指定阈值进行聚类, 形成各个块(patch)区域。最后, 输出这些块区域作为候选表位。该系统使用蛋白质数据库(protein data bank, PDB)中63个抗原/抗体复合物组成的数据集进行测试。使用接受者操作特性曲线(ROC)下的面积(area under the ROC curve, AROC)度量系统性能, 算法的AROC评分达到0.750。该系统是第一个以抗原蛋白质的三级结构PDB文件作为输入, 以构象性表位预测为主要目的网上免费服务软件, 具有开创性的意义, 与其后出现的同类软件相比略显不足的是特征选择太少。

ElliPro是一个基于结构的蛋白质表位预测和可视化的在线服务工具^[7]。如果抗原蛋白质的三维结构已知, 则可输入其在PDB数据库中的ID或提交一个具有PDB格式的自己的PDB文件。如果其结构未知, 则要求输入其序列, 系统将在PDB数据库中使用BLAST工具搜索这个蛋白质或它的同源物; 如果没有发现匹配的目标, 则用MODELLER程序预测蛋白质的3D结构, 然后用MODELLER建立的模板作为一个输入。对于输入的蛋白质结构, 系统实现了三个算法来完成表位预测任务: (1)把蛋白质的形状近似为一个椭球; (2)计算球面残基的突出索引(protrusion index, PI); (3)基于它们的PI值聚类近邻残基。在Ponomarenko等于2007年发布的基准数据集上系统的AROC评分达到0.732。系统还利用开源分子可视化工具Jmol实现了3D表位图的输出。改进该系统的进一步研究也是应考虑更多的特征。

Rubinstein等^[8]于2009年首次将朴素贝叶斯分类器用于构象性B细胞表位预测。这项研究潜在的假设是, 抗原表面的表位部分和非表位部分可以依据它们的理化和几何结构特征加以区分。为识别基于这些大量特征的免疫原区, 系统训练了两个朴素贝叶斯分类器: 一个是基于结构的, 一个是基于序列的。对于基于结构的分类器, 系统构建了一个表位已可靠确定的抗原基准数据集, 它们来自已知晶体结构的抗原-抗体复合体集。首先, 为已验证的表位集提取它们的理化和几何结构特征, 用这些表位特征集训练分类器。然后, 对于一个待求表位的蛋白质抗原, 输入其表面的三维结构; 把该表面按典型表位尺寸划分成一些搭接的表面块, 计算每个表面块的理化和几何结构特征。最后, 使用已训练好

的分类器, 基于与先前已知表位的相似性为每个表面块计算反映其潜在免疫原性的概率得分, 再把该得分赋给表面块的中心残基, 作为推断它是否为表位的依据。在这项设计中, 系统利用各种软件工具提取了28个几何特征、16个氨基酸倾向性标度特征。其中几何特征包括每个氨基酸在表面块与其余表面中的出现率之比、二级结构元件(α 螺旋、 β 转角和环)在表面块与其余表面中的出现率之比、在探针半径为1.4Å情况下表面块到溶剂的平均相对可及性、在探针半径为9Å情况下表面块接近抗体互补决定区(complementarity-determining region, CDR)的平均可及性等。为了优化性能, 系统还对这些特征进行了筛选。采用交叉验证中的留一验证法, 系统的AROC评分从特征筛选前的0.703达到了筛选后的0.914。

机器学习方法能够揭示数据集中的复杂非线性关系, 适合数据量大、含有噪声并且缺乏统一理论的领域。该方法通常联合多个特征(有时还有实验数据)进行预测, 以克服基于较少特征的一些方法的预测识别率不高的缺陷。系统性能的进一步提高依赖于基准数据集的不断增长和更具价值特征地发现。

1.2 非机器学习的计算预测方法

非机器学习的计算预测方法, 主要指的是使用蛋白质抗原结构的几何特征或使用几何特征与理化特征的联合, 通过计算直接实现预测(不使用机器学习)。这类方法直观明确, 不存在机器学习方法中的所谓“黑箱操作”质疑, 受到一些研究人员的信赖。

DiscoTope是一个基于序列-结构混合方法的构象性B细胞表位预测在线服务系统^[9]。它是一个基于训练集且利用了更多表位特征的方案, 所用特征涉及氨基酸倾向性、残基溶剂可及性、空间分布和分子相互作用接触。当抗原氨基酸残基具有与抗体原子距离在4Å之内的原子时, 则被定义为表位残基。训练样本集(DiscoTope数据集)来源于对76个抗原/抗体复合物的X射线晶体衍射分析得到的非连续性表位。基于该表位数据集, 系统计算出了参与抗原相互作用的每一个氨基酸残基的统计特征——对数让步比矩阵(log-odds ratios matrix); 计算出了表面氨基酸定位和结构突出部分的联合度量(一种溶剂表面可及性度量)——残基接触数目。残基接触数目是指距该残基 C_{α} 原子10Å之内的抗原中的 C_{α} 原子

的数目。系统使用了该矩阵和联合度量的得分来预测构象性表位残基的定位。与单独的基于Parker的亲水性参数的线性表位预测方法和基于抗原三维结构推断出的溶剂表面可及性预测方法相比均显示出其优越性。算法的AROC评分达到0.711。

PEPITO是一个联合氨基酸倾向性得分和多个距离上的半球暴露值的构象性表位预测系统^[10]。其中,氨基酸倾向性标度采用Andersen等^[9]的描述;半球暴露值与接触数目的相关性采用了Hamelryck的描述。系统为目标蛋白质肽链中的每一个残基 r 计算出一个表位得分 $E(r)$ 。 $E(r)$ 越大,表明残基 r 越可能是一个表位残基。 $E(r)$ 值计算方法:首先,对于每一个 k ($k=8, 10, \dots, 16\text{\AA}$),分别计算出下列三项:残基 r 的 k 范围内的所有残基的倾向性标度得分(滑动窗口宽度为9个残基)求和、残基 r 的 k 范围内上半球中的 C_{α} 原子的数目(上半球暴露值,反映其侧链是指向蛋白质表面还是中心)、残基 r 的 k 值范围内下半球中的 C_{α} 原子的数目(下半球暴露值)。然后,对这三项进行线性叠加。最后,对所有 k 值求得的三项线性叠加再求和,即得 $E(r)$ 值。PEPITO在DiscoTope数据集上AROC评分达到0.754。另外,在更新的Epitome数据集上检验PEPITO与DiscoTope,AROC评分分别达到了0.683和0.660。

1.3 基于噬菌体展示数据的识别方法

噬菌体展示技术(phage display technology)经过多年的发展和完善,已能够为表位研究提供一个操作简便、高通量、低成本和可调控的实用性平台。因为用单克隆抗体从噬菌体肽库中竞争免疫筛选得到的线性肽序列即模拟肽(mimotope),在一定条件下可以引发识别原始抗原的抗体反应,所以这样得到的模拟肽集,特别是通过比对模拟肽集得到的模体(motif)集含有一定的表位信息;但是噬菌体形态发生过程中伴随着氨基酸序列的偏向性,而且筛选得到的模拟肽在蛋白质抗原上通常不是以线性序列形式存在的^[5]。因此,确定这些肽段模拟的抗原区域还需要特殊的工具。一般地,这种基于模拟肽分析的计算方法可采用下列步骤^[11]:(1)描述抗原的表面残基;(2)在抗原三维结构分子表面搜索与模拟肽或模体最佳匹配的氨基酸序列;(3)基于聚类算法和可视化技术输出候选表位。因为在代表抗原暴露残基的表面图上求解所有给定长度的简单路径(即一个由相邻残基构成的序列)是NP-hard问题,所以精确求解时面临过高的时间复杂度。为了克服传统方法的

局限,已提出多种启发式算法。

EMT是利用表位模体(epitopic motif)数据库预测蛋白表位的工具^[12]。通过比对从噬菌体肽库中竞争免疫筛选捕捉到的模拟肽,收集氨基酸表位模体,构建表位模体表。对于抗原中的氨基酸残基,结合对其上是否存在这些表位模体的考察和相对表面可及性的评价,对其抗原性进行预测。

EPIMAP方法是由Mumey等^[13]于2003年和2006年提出的。在这个方法中,从噬菌体肽库中筛选得到的每个模拟肽都通过动态规划算法与其父抗原进行比对。抗原蛋白质每一个残基的比对频率结果图揭示出与三维空间中自然折叠接近的片段。

MEPS是一个基于肽段有效模拟的蛋白质构象性表位识别的在线服务系统^[14],其具体步骤为:(1)产生肽表面体。对于给定的一个目标蛋白质的结构,首先,选择溶剂暴露氨基酸。可设最小溶剂可及性 SA_{min} 。其次,计算每对暴露氨基酸的 C_{β} (对于甘氨酸是 C_{α})间的距离并存入一个矩阵。最后,用该矩阵作图:每一个节点代表一个表面氨基酸;当两节点的距离低于一个最大阈值 d 时,它们之间则存在一条边。用图表示邻接表的集合。(2)为一个目标蛋白质或蛋白质集找到所有的潜在表位。把肽表面体图处理为能够提取模拟给定蛋白质表面的所有的可能的肽序列(序列最大长度为 L)。对于每一个邻接表,迭代搜索 5.5\AA 距离内的近邻,直到用户选择的 L 值。肽序列被交互地检查并存入一个FASTA格式的文件中,以便能够直接用于肽表面体的BLAST或FASTA相似性搜索。(3)搜索构象性表位。当用户向系统输入一个蛋白质结构、模拟表位的肽序列和一个最小错配氨基酸数目后,如果它的确存在,通过搜索将得到相应表位的定位。也可以搜索固定大小肽序列,得到相应的表位定位。系统还利用RasMol可视化工具实现了可视化表位图的输出。

Pep-3D-Search是基于模拟肽分析的B细胞构象性表位预测工具^[11]。其算法流程如下。首先,输入一个抗原的三维结构和一个模拟肽集(或一个模体),识别抗原的表面残基并创建一个表面残基图。其次,对于输入为模拟表位集的情况,通过使用蚁群优化(ant colony optimization,ACO)算法在抗原表面图上搜索各模拟肽;计算每一个匹配的路径的 P 值;从匹配的路径集中选择结果路径;为结果路径创建一个权重图。最后,使用深度优先(depth-first search,DFS)算法聚类结果路径得到候

选表位。而对于输入为模体的情况,通过使用 ACO 算法在抗原表面图上定位模体,得到高分的那些路径被直接输出为候选表位。

把噬菌体展示技术构建的随机肽库和计算机建模结合起来进行构象性表位预测,有较坚实的实验基础,预测更可靠,受到了研究者的广泛青睐。这些方法的主要局限在于只能预测出确定的单克隆抗体的亲和表位,算法没有考虑抗原-抗体结合的构象契合关系和分子相互作用机制,不能进行更广泛的表位预测^[5]。

1.4 一些通用蛋白质-蛋白质界面预测方法

蛋白质-蛋白质界面预测是指对参与蛋白质相互作用的特定残基及其位置的预测。它们依赖于在蛋白质复合物界面中发现的残基所具有的特性。近年来, PDB 经历了快速发展。为了最大化利用存储在 PDB 中的高分辨率的蛋白质和蛋白质复合物的三维结构数据,已开发出众多蛋白质-蛋白质界面数据库。而其中的数据使得比较蛋白质表面的界面和非界面成为可能。在界面残基特征分析中发现序列保守性、20 种氨基酸的比例、二级结构、溶剂可及性、侧链构象熵等是比较突出的特征。但这些涉及典型蛋白质结构的特征属性对于伴随复合物形成的局部构象改变不是特别敏感。应该注意,随着抗体的成熟,一个蛋白质抗原表面的许多部分可以成为与抗体结合的表位。因此,早期的仅依赖于这些普通特征的界面预测系统并不能很好地预测蛋白质抗原表位。

然而,最近出现的一些通用蛋白质-蛋白质界面预测方法如蛋白质-蛋白质对接方法、用各种原理和不同数据集训练的基于结构的方法和把残基序列保守性映射到表面的方法等却取得了较好地预测效果。其中,蛋白质-蛋白质对接是指由诸蛋白质单体结构配位构成复合体结构的过程,分为刚性对接和柔性对接两大类。在这个过程中,不论是局部的还是大规模的构象改变都是极其复杂的。潜在姿势的巨大搜索空间和排列对接姿势的困难是两个主要问题。例如, ClusPro (DOT 程序)^[15, 16]在线服务实现了基于快速傅立叶变换相关方法的刚体蛋白质-蛋白质对接, PatchDock^[17]在线服务实现了基于局部形状特征匹配的刚体蛋白质-蛋白质对接。

2 预测软件和 Web 服务资源

最新预测软件和 Web 服务资源见表 1。

3 结语

抗原表位预测是免疫信息学研究的重要方向之一,可以给实验提供重要的线索。构象性 B 细胞表位预测具有广阔的应用前景,但同时也是一个难度很大的课题。首先,应尽可能多地收集抗原表位数据,并按标准格式存放在数据库中;其次,开发用于表位预测的新方法、新软件,为临床和基础生物医学研究人员提供服务。最近,出现了各种构象性 B 细胞表位预测软件工具和相关 Internet 资源。由于它们选择的数据集和提取的特征不同和采用的策略差异较大,呈现出了各自的优势与局限,但总体上预测的准确性都还不够高。这些系统性能的进一步提高依赖于基准数据集的增长、更具价值特征的发现、计算策略的改进以及各种方法间的联合,对这些方法进行评估与比较亦有助于其发展^[1]。采用生物实验和计算机预测相结合的方法,把分子相互作用理论、抗原-抗体的构象契合关系以及抗原-抗体结合机制融合到预测系统中来,提高预测的准确性和预测方法的普遍性是未来研究的方向^[5]。

[参 考 文 献]

- [1] Greenbaum JA, Andersen PH, Blythe M, et al. Toward a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools. *J Mol Recognit*, 2007, 20: 75-82
- [2] Korber B, LaBute M, Yusim K. Immunoinformatics comes of age. *PLoS Comput Biol*, 2006, 2(6): 484-92
- [3] Ponomarenko JV, Bourne PE. Antibody-protein interactions: benchmark datasets and prediction tools evaluation. *BMC Struct Biol*, 2007, 7: 64-82
- [4] Yang X, Yu X. An introduction to epitope prediction methods and software. *Rev Med Virol*, 2009, 19: 77-96
- [5] 黄艳新. 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): 857-61
- [6] Kulkarni-Kale U, Bhosle S, Kolaskar AS. CEP: a conformational epitope prediction server. *Nucleic Acids Res*, 2005, 33: W168-71
- [7] Ponomarenko J, Bui H, Li W, et al. ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics*, 2008, 9: 514-21
- [8] Rubinstein ND, Mayrose I, Pupko T. A machine-learning approach for predicting B-cell epitopes. *Mol Immunol*, 2009, 46: 840-7
- [9] Andersen PH, Nielsen M, Lund O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci*, 2006, 15: 2558-67
- [10] Sweredoski MJ, Baldi P. PEPITO: improved discontinuous B-cell epitope prediction using multiple distance thresholds and half sphere exposure. *Bioinformatics*, 2008, 4(12): 1459-60

- [11] Huang YX, Bao YL, Guo SY, et al. Pep-3D-Search: a method for B-cell epitope prediction based on mimotope analysis. *BMC Bioinformatics*, 2008, 9: 538-54
- [12] Batori V, Friis EP, Nielsen H, et al. An *insilico* method using an epitope motif database for predicting the location of antigenic determinants on proteins in a structural context. *J Mol Recognit*, 2006, 9: 21-9
- [13] Mumei BM, Bailey BW, Kirkpatrick B, et al. A new method for mapping discontinuous antibody epitopes to reveal structural features of proteins. *J Comput Biol*, 2003, 10: 555-67
- [14] Castrignanò T, De Meo PD, Carrabino D, et al. The MEPS server for identifying protein conformational epitopes. *BMC Bioinformatics*, 2007, 8(Suppl 1): S6
- [15] Mandell JG, Roberts VA, Pique ME, et al. Protein docking using continuum electrostatics and geometric fit. *Protein Eng*, 2001, 14(2): 105-13
- [16] Comeau SR, Gatchell DW, Vajda S, et al. ClusPro: an automated and discrimination method for the prediction of protein complexes. *Bioinformatics*, 2004, 20(1): 45-50
- [17] Schneidman-Duhovny D, Inbar Y, Polak V, et al. Taking geometry to its edge: fast unbound rigid (and hinge bent) docking. *Proteins*, 2003, 52(1): 107-12
- [18] Mayrose I, Shlomi T, Rubinstein ND, et al. Epitope mapping using combinatorial phage-display libraries: a graph-based algorithm. *Nucleic Acids Res*, 2007, 35(1): 69-78
- [19] Zhang Q, Wang P, Kim Y, et al. Immune epitope database analysis resource (IEDB-AR). *Nucleic Acids Res*, 2008, 36: W513-8
- [20] Bradford JR, Westhead DR. Improved prediction of protein-protein binding sites using a support vector machines approach. *Bioinformatics*, 2005, 21: 1487-94
- [21] Kufareva I, Budagyan L, Raush E, et al. PIER: protein interface recognition for structural proteomics. *Proteins*, 2007, 67(2): 400-17
- [22] 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
- [23] Landau M, Mayrose I, Rosenberg Y, et al. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res*, 2005, 33: W299-302

表1 构象性B细胞表位预测软件和Web服务资源

名称(年份)	描述	网址
CEP (2005) ^[6]	基于残基溶剂可及性和空间分布的构象性表位预测工具	http://bioinfo.ernet.in/cep.htm
ElliPro (2008) ^[7]	基于几何特性的蛋白质表位预测和可视化工具	http://tools.immuneepitope.org/tools/ElliPro
EpitopePrediction (2009) ^[8]	基于机器学习(朴素贝叶斯分类器)的B-cell表位预测	http://www.tau.ac.il/~talp/EpitopePrediction
DiscoTope (2006) ^[9]	基于氨基酸统计、残基溶剂可及性和空间分布的构象性表位预测	http://www.cbs.dtu.dk/services/DiscoTope
PEPITO (2008) ^[10]	联合氨基酸倾向性标度和多个距离上的半球暴露值的构象性表位预测	http://pepito.proteomics.ics.uci.edu/
EMT (2006) ^[11]	基于噬菌体展示数据的连续/非连续表位预测工具	elro@novozymes.com
EPIMAP (2006) ^[12]	基于噬菌体展示数据的非连续表位识别工具	mumei@cs.montana.edu
MEPS (2007) ^[13]	基于肽段有效模拟的蛋白质构象性表位识别的在线服务系统	http://www.caspur.it/meps
Pep-3D-Search (2008) ^[14]	基于模拟肽分析的B细胞构象性表位预测系统	http://kyc.nenu.edu.cn/Pep3DSearch
PepSurf (2007) ^[18]	一个基于图的使用组合噬菌体展示肽库的表位作图算法html	http://pepitope.tau.ac.il/sources.html
IEDB-AR (2008) ^[19]	免疫抗原表位数据库分析资源(含7个B细胞表位预测工具)	http://tools.immuneepitope.org
ClusPro (DOT) (2004) ^[16]	基于快速傅立叶变换相关方法的刚体蛋白质-蛋白质对接	http://structure.bu.edu
PatchDock (2003) ^[17]	基于局部形状特征匹配的刚体蛋白质-蛋白质对接	http://bioinfo3d.cs.tau.ac.il/sources.html
PPI-PRED (2005) ^[20]	基于有效结构和序列界面特性的蛋白质-蛋白质结合界面预测	http://bioinformatics.leeds.ac.uk/ppi-pred
PIER (2007) ^[21]	基于蛋白质表面(原子团水平)局部统计特性的蛋白质结合界面预测	http://abagyan.ucsd.edu/PIER/
ProMate (2004) ^[22]	基于有效结构和序列界面特性的蛋白质-蛋白质结合界面预测	http://biointo.weizmann.ac.il/promate/
ConSurf (2005) ^[23]	映射系统发生信息到已知三维结构的蛋白质表面的界面预测	http://consurf.tau.ac.il