

活性肽与 MHC 结合能力预测的免疫信息学方法研究进展

Research progress on immune-informatics methods for predicting the binding ability of active peptides to MHC

黄璐 文李许宙 陈茂龙 程云辉

HUANG Lu WEN Li XU Zhou CHEN Mao-long CHENG Yun-hui

(长沙理工大学化学与生物工程学院, 湖南 长沙 410114)

(College of Chemical and Biological Engineering, Changsha University of Science and Technology, Changsha, Hunan 410114, China)

摘要:综述活性肽与 MHC(主要组织相容性复合体, major histocompatibility complex)结合能力预测的免疫信息学方法的最新进展, 介绍常用的肽与 MHC 分子结合预测的相关工具及方法, 分析了各类方法的特点、研究重点和难点, 以期为寻找免疫活性肽提供更快捷的方法。

关键词:活性肽; MHC(主要组织相容性复合体); 分子结合预测; 免疫信息学

Abstract: In this paper, the latest advances in immune-informatics methods for predicting the binding ability of active peptides to MHC are reviewed. The related tools and methods for predicting the binding of commonly used peptides to MHC molecules are introduced. The characteristics, research focuses, and difficulties of various methods are analyzed. Finding immune-active peptides provides a faster method.

Keywords: active peptide; MHC (major histocompatibility complex); molecular binding prediction; immuno-informatics

近年来生物活性肽因其强大的功能以及来源的可靠性等优点成为保健品、药品等高科技产业的研究和开发热点。基于酶法降解食物蛋白质并结合分离纯化技术, 从酶解混合物中获得活性肽抑制被当作筛选食源性活性肽的经典方法, 但因活性功能评价方法不一致及分离纯化过程繁琐而无法实现食源性活性肽的高效筛选。免疫活性肽(Immuno-peptides)能提高机体免疫力和抵挡外界病原体的感染能力、刺激淋巴细胞增殖、增强巨噬细胞吞噬能力, 并具抗肿瘤等

功能^[1]。免疫活性肽进入抗原提呈细胞(antigen presenting cells, APC)与主要组织相容性复合体(major histocompatibility complex, MHC)结合形成结合物, 被抗原的 T 细胞受体(T cell receptor, TCR)识别后递呈给 CD4+ T 细胞, 启动 CD4+ T 细胞参与免疫应答反应, 如刺激淋巴细胞增殖、分化、成熟和增强巨噬细胞吞噬功能等^[2]。免疫信息学是一门新兴交叉学科, 它建立在现代免疫学和信息学基础上, 运用信息学的理论和方法解决免疫学及疫苗问题, 同时也研究免疫系统调控和免疫应答过程中信息传递规律^[3]。免疫信息学主要研究免疫相关分子的结构与功能, 尤其致力于抗原受体、MHC 分子以及细胞因子等方面^[4]。为了找到更快捷筛选免疫活性多肽的方法, 可结合免疫信息学工具分析免疫调节机制, 由此省略传统食源性活性肽分离纯化的诸多步骤, 减少活性肽筛选的盲目性, 有效提高筛选效率, 进而推广至其他食源性活性肽的高效筛选。

T 细胞介导的免疫反应由效应 T 细胞激活。激活过程需要 T 受体细胞识别抗原多肽-MHC 复合物。表位肽与 MHC 分子结合是细胞免疫的主要事件, 精确预测多肽与 MHC 分子间的结合是免疫信息学的一项重要任务。因此鉴定 MHC 分子结合的肽序列在理解免疫及发展基于表位的疫苗研究中具有关键作用。

由于试验测定多肽与 MHC 分子的结合亲和力耗时长且费用昂贵, 在基因组水平鉴定宿主及病原体蛋白中的潜在结合多肽是一项比较艰巨的任务。因此, 国内外已在利用免疫信息学工具识别 MHC 结合多肽的理论计算方法上做了大量的工作^[5-6]。本文综述了活性肽与 MHC 结合能力预测的免疫信息学方法的最新进展, 介绍了常用的预测多肽与 MHC 分子结合的相关工具及方法, 分析了各类方法的特点、研究重点和难点, 以期为寻找免疫活性肽提供更快捷的

基金项目:湖南省自然科学基金项目(编号:2018JJ2423)

作者简介:黄璐, 女, 长沙理工大学在读硕士研究生。

通信作者:程云辉(1964—), 女, 长沙理工大学教授, 博士。

E-mail: Chengyh6488@sina.com

收稿日期:2018-02-07

方法。

1 常用免疫信息学数据库

1.1 国际免疫遗传学信息系统(IMG)

IMGT 即国际免疫遗传学信息系统(<http://imgt.cines.fr>),于1989年由Laboratoire在法国蒙彼利埃创立,目的在于管理复杂的免疫信息学数据并使其标准化。它是一种高质量的整合信息资源,分别致力于以下三个方面数据的获取:① 免疫球蛋白(IG),T细胞受体(TR),人类和其他脊椎动物的主要组织相容性复合体(MHC);② 属于免疫球蛋白超家族和主要组织相容性复合体超家族(MhcSF)的蛋白质;③ 任何物种的与免疫系统相关的蛋白质(RPI)。IMGT为获取IG、TR、MHC、IgSF、MhcSF及RPI等基因组、蛋白质组、遗传学和三维结构标准化数据提供了一个通用路径^[7-8]。IMGT信息系统由数据库、工具和网上资源组成^[5],它为基因组、序列和三维结构分析提供了互动式在线工具。因此也发展了3种主要的IMGT物理学方法:基因组学、遗传学和结构研究方法。基因组学方法是以基因为中心,对基因定位及染色体定位的研究。遗传学方法是对基因及其相关的序列多态性、突变、基因表达、特异性和进化的研究。结构方法指对IG、TR、MHC和RPI的二维及三维结构以及蛋白质功能、多态性和进化有关的抗原和配体结合特征的研究。IMGT为每一种方法提供了数据库,包括1个基因组数据库IMGT/GENE-DB^[9],3个序列数据库IMGT/LIGM-DB、IMGT/MHC-DB和IMGT/PRIMER-DB^[10-11]和1个三维结构数据库IMGT/3Dstructure-DB^[12]。

IMGT/HLA数据库是针对人类白细胞抗原系统(或人类主要组织相容性复合体MHC)等位基因序列的专业数据库。超过4 Mb的复合体MHC位于人类染色体6的短臂6p21.3处,同时含有220多个基因^[13]。HLA系统的核心包括21个高度多肽的HLA基因,这些基因影响到细胞和器官移植排斥反应以及宿主针对传染性疾病的免疫反应,同时与许多慢性传染性疾病的易感性密切相关^[14-15]。IMGT/HLA数据库中的所有序列也可以在更加综合的核苷酸序列数据库如EMBL^[16]、GenBank^[17]和DDBJ^[18]中得到运用。大型综合序列数据库的优势是可利用大量的序列,包括很大范围的HLA相关数据。与其他HLA数据库相比,IMGT/HLA数据库可直接从万维网进入,同时该数据库给使用者提供许多工具,如等位基因报告、序列排列工具和来源细胞的详细数据库,使数据分析更方便^[19]。以任何形式命名的HLA等位基因序列均可从等位基因查询工具中获得,如来源个体、种族起源、参考文献、核苷酸和蛋白质序列信息。序列排列工具格式与文本排列格式相同,该格式可进入ANRI网(<http://www.anthonynolan.org.uk/hig/>)查询。在使用排列工具时,单个序列的排列无法进行,但已排列过的序列可进行,序列用完整的核苷酸序列、单个外显子的部分序列或编码蛋白的氨基酸序列表示。使用细胞查询工具来查询伴随的细胞数据库,每个等位基因的来源和特征确定的原始淋巴细胞和细胞株的描述都能在该工具中被检索。

1.2 T细胞表位查询和预测数据库 SYFPEITHI

SYFPEITHI是MHC配体及人、小鼠、大鼠、牛和鸡等物种MHC分子肽基序的数据库。该数据库包含MHC-多肽基序、MHC配体和T细胞表位。T细胞表位的预测以序列库中已出版的序列和对个体天然配体的分析为基础,尤其将锚着位点的氨基酸及其他常见氨基酸^[20]考虑在内。在二维空间的数据系列中,氨基酸的字母代表行数,位点数字代表列数。任何进入的序列均被划分为8肽、9肽或10肽;紧接着对每一个氨基酸的低聚体进行评分计算;该过程会一直持续到序列终点为止。根据T细胞表位在天然配体中的出现频率,不同的氨基酸被评价为不同的分值。10分为频繁出现在锚着位点的氨基酸,8分为出现在大量配体中的氨基酸,6分则为出现在辅助锚着位点的氨基酸。在辅助锚着位点低频率出现的氨基酸则为4分;首位氨基酸拥有1~4分,这依赖于序列库信号的长度或个体序列的频率。-1~3分为通常不出现在自然配体的各自序列位点的氨基酸。SYFPEITHI表位预测的结果来源于一系列被MHC分子高概率提呈的肽,预测结果可产生一张由可能的MHC分子组成的抗原清单。但由于肽降解及结合槽的可变性,MHC II类分子限制性T细胞表位的预测更为复杂。

1.3 MHCBN数据库

MHCBN数据库(<http://www.imtech.res.in/raghava/mhcbn>)包括超过23 000个肽段与MHC或TAP(transporter associated with antigen processing)分子的亲和力数据,且这些亲和力数据都已经过实验验证。该数据库对所有的肽段都提供了完整的信息,如序列、与MHC或TAP的结合、肽与MHC/TAP分子亲和力的 IC_{50} 值、T细胞活力和蛋白质来源;并且还能提供关于锚着位点的信息(肽段的每个位置对其与TAP或MHC分子的反应都非常关键)。关于肽的其他信息在注解区也可以找到,注解区主要包括已发布论文中提到的 IC_{50} 及其参考值,用来区分高、中、低和非结合。与条目有关的数据的文献出处保留在条目的发布参考区^[21],且公开发表的文献已经链接到NCBI中的PubMed数据库。

MHCBN数据库也包括所有具有相同抗原肽的蛋白质序列和结构信息,不仅数据库中所有的肽都可以在SWISS-PROT上搜索,而且含有匹配肽段的蛋白质都可以被搜索到。这些序列号以FASTA格式储存在数据库中,并被链接到GenBank和SWISS-PROT^[17-22]。含有匹配肽段的蛋白质的简略三维结构可以通过OCA browser^[23]在所链接的蛋白质数据库中找到,这些信息有助于认识抗原区和非抗原区的结构特点。

2 常用免疫信息学分析工具及方法

2.1 特征参数法预测MHC结合肽

位点特异性的分矩阵(position-specific scoring matrix, PSSM)也称为特征参数,它来自于特定MHC分子已知结合肽的比对,可以作为MHC-多肽以及T细胞表位的预测工具。其中肽序列与MHC分子的结合亲和力由其与已知

MHC 结合肽的相似性决定,可通过查询肽序列与 PSSM 相比较而得到。利用 PSSM 预测多肽与 MHC 结合需要有与特定 MHC 结合的多肽序列,这些序列可以从任意 MHC 配体数据库中获得(表 1),所用到的计算机程序见表 2。

表 1 所选 MHC 配体公共数据库网址

Table 1 Selected MHC ligand public database URL		
数据库	网址	描述
MHCPEP	http://wehih.edu.au/mhcpep	MHC 结 合 肽 数 据库 ^[24]
MHCBN	http://www.imtech.res.in/raghava/mhcbn	MHC 结合肽以及非结合肽数据库 ^[25]
ANTIGEN	http://www.jenner.ac.uk/antigen/	免疫相关的定量功能多肽数据库 ^[26]
FIMM	http://research.i2r.2a-star.edu.sg/fimm	免疫功能分子数据库 ^[27]
EPIMHC	http://bio.dfci.harvard.edu/epimhc/	MHC 配 体 数 据库 ^[28]

表 2 所用计算机程序下载地址

Table 2 Computer program download address		
程序	下载地址	描述
BLIMPS	ftp://ftp.ncbi.nih.gov/repository/blocks/unix	基序查找程序
MEME	ftp://ftp.sdsc.edu/pub/sdsc/biology/meme	基序查找程序
PROFILE-WEIGHT	ftp://ftp.ebi.ac.uk/pub/software/unix/profile.tar.Z	由 GCG/MSF 比对建立 PSSM 的程序
READSEQ	ftp://iubio.bio.indiana.edu/molbio/redseq/classic/	序列格式转化程序

从已知 MHC 分子配体得到比对和简表来预测多肽与 MHC 分子结合的方法包括 3 个基础步骤(图 1):① 多肽序列的收集并按 MHC 结合特异性与长度划分子集;② 无空位比对的生成;③ 由比对产生 PSSM。

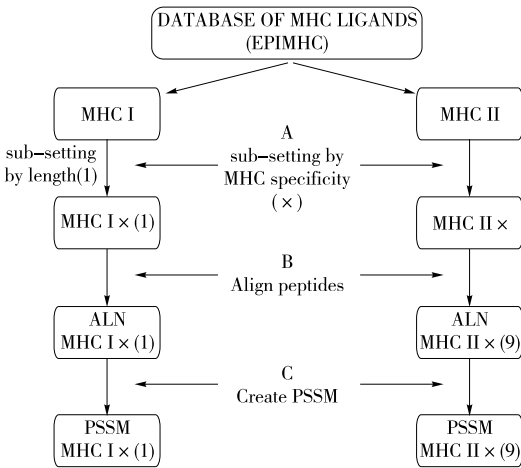


图 1 PSSM 的基本步骤

Figure 1 The basic steps of PSSM

(1) 多肽收集及子集划分:对于 MHC II 类分子配体,仅需要根据 MHC II 类分子结合肽的限制性将序列划分为不同的子集文件,长度小于 9 个残基的肽需剔除。满足上述标准的 MHC 配体可以通过 EPIMHC 数据库提供的网页界面得到,肽序列以纯文本或 FASTA 格式保存。

(2) 无空位基序比对的生成:利用 MEME 程序的命令进行比对。MEME 的输出文件(mhcii_lig.meme)包括 1 个对数值以及 1 个 MHC II 类分子配体结合核心的 PSSM 概率矩阵,该概率矩阵可用来预测多肽与 MHC II 类分子的结合。

(3) 利用 MHC 配体比对生成 PSSM:BLIMPS PSSM 通过依次使用以下 3 个程序得到,包括 mablock 将 FASTA 格式的比对转换为 BLOCK 格式、blweight 将比对中的序列进行加权、Blk2pssm 产生真正的 PSSM 矩阵。

PSSM 是强大的工具,不仅可以用来确定与起始比对序列(MHC 分子结合肽)功能相关的新的多样性序列,也可用于鉴定那些与 MHC 分子结合的多肽。

2.2 3D-QSAR 模型预测 MHC-多肽亲和力

模拟肽抑制剂和呈递多肽的 MHC II 类分子和 HLA-DR4 受体间的相互作用,可以用 Cerius2 软件和比较分子力场分析(CoMFA)等三维定量构效关系(3D-QSAR)方法进行建模。多肽配体的结构应基于 PDB 数据库中 MHC-多肽复合物 X 射线结构进行构建。MHC 活性位点中的配体构建可由 SYBYL 完成,用其他残基替换模板侧链来构建配体结构^[29](Build/Edit>>Sketch Molecule>>Draw)。对每个结构加氢,在将配体对接到 MHC 分子之前,这些结构必须置于 X 射线配体结构的坐标框架中;然后在 SYBYL 中使用 Tripos60 力场对每个受体-配体复合物结构进行简单的能量最小化(Compute>>Minimize)。在 SYBYL 中将配体结构从复合物中提取出来(Build/Edit>>Extract>>Substructures),其生物活性由 pIC50 表示。

(1) 使用 SYBYL 软件计算 CoMFA 参数:清除显示区域中的所有分子 Build/Edit>>Zap(Delete)Molecule;为构建的多肽配体结构建立一个数据库 File>>Database>>New>>Put Molecule;使用公共结构模板对数据库进行结构叠合 File>>Align Database>>>Database to Align:>>>Template Molecule:>>>Location of Substructures:>>>Put Molecules Into:>>>Align;为叠合数据库中的所有配体结构建立一个分子表格 File>>Molecular Spreadsheet>>New>>>Database>>>Open,数据库中的所有配体结构作为行读入此表格;对所有叠合分子添加 CoMFA 力场来建立环绕区域并计算超过 33 000 的能量值,在 MSS 面板上选择 empty column2 并点击 Autofile,选择 COMFA 作为新的列类型;通过点击 MSS 执行对生成的 CoMFA 列进行 PLS 分析 QSAR>>>Partial Least Standard CoMFA Field,出现 PLS 分析对话框;输入 COMFA2 列作为自变量,然后输入 ACTIVITY 列作为因变量,其值可以用 PLS 分析结构进行预测;当数据集的个数与所选数据行数相同时,利用留一法交互检验进行 PLS 分析,关掉 SAMPLS 选项;使用图形方法显示最好的交互检验结果,并在预测中使用最优预测模型。

文本窗口将显示拟合的 r^2 、静电和空间场的百分比贡献;在 MSS 控制板中点击 QSAR > > > View CoMFA 查看 CoMFA 的结果,在 Display 选项菜单中选择最终建立模型。

(2) 使用 Cerius2 软件进行 QSAR 预测:在 Unix 提示中输入 Cerius2 打开一个新的 Cerius2 对话框;进入 Build/3D-Sketcher 面板,选择所需结构模板按钮绘制结构;进入 OFF SETUP 选项卡对绘制结构进行优化,然后选择 Load Force Field 并选择 cvff950_1_0_1 作为所用力场;点击 Energy Minimization 选项卡上的单选按钮开始计算;进入 QSAR/Show Study 表格,将分子导入此表格,在 Molecules 下拉菜单中点击 Add all,输入活性数据并以 Activity 标记列;在 Study 表格中点击列头选择标记为 Activity 的列,选择菜单栏的 Variables/Set Y 选项将此列设为因变量 Y;选择 Study 表格菜单栏中的 Variables/Set X 选项,将所有描述子列为

自变量 X;将 Methods 弹出窗口设为 GFA,利用遗传函数近似(GFA)方法产生 QSAR 方程,点击 RUN 进行 GFA 计算;利用交互检验法选择最好的方程进行验证,结果将显示在文本窗口中;Plot Equation 按钮可以查看预测活性对实测活性的 2D 散点图。

2.3 基于 MHC 分子模型预测表位肽

因为 MHC 识别病原性和致癌性多肽的作用使得基因在高强度的环境压力下具有高度多态性^[11]。根据每个等位基因可用的准确结合肽试验数据的数目可预测与特定 MHC 分子结合的多肽。模块是特定 MHC 等位基因中构成结合位点的氨基酸序列。对于 9 肽,一个给定的等位基因含有 9 个模块(P1,P2,...,P9)。由于 MHC 等位基因间的相似性,当不同的 MHC 在定义的位点拥有相同的氨基酸时,表明它们共享一个模块(见表 3)。

表 3 A * 0101 和 A * 7401 的 P1 模块
Table 3 P1 Modules for A * 0101 and A * 7401

表位	位点	氨基酸	该模块的其他表位
A * 0101	5 7 33 59 62 63 66 99 159 163	M Y F Y Q E	A * 0102,A * 0103,A * 0106,A * 0107,A * 0110
	167 171	N Y Y R G Y	
A * 7401	5 7 33 59 62 63 66 99 159 163	M Y F Y Q E	A * 0256,A * 0301-14,A * 1104,A * 3001-6,8,9,11,12,A * 3101,3,4,6,9,A * 3201-4,6-8,A * 3601-3,A * 7402,3,5-10
	167 171	N Y Y R G Y	

表 3 中所列出的位点是 HLA 蛋白中可能影响多肽 P1 位置氨基端结合的位点,氨基酸分别是 A * 0101 和 A * 7401 中给定位点的氨基酸。这些非连续氨基酸的类别就是 P1 位点的模块,标记有“Other alleles with this module”等位基因是在这些位点拥有相同氨基酸的等位基因,因此拥有相同的 P1 模块,建立模块的概念是为了扩展可预测 MHC 结合肽等位基因的数量。

基于模块结合肽预测方法需要收集已由试验证明可与 MHC 结合的多肽,数据库如 SYFPRITHI^[20]、MHCBN^[25] 和 Antigen^[26] 均可查询已由试验证明可与 MHC 结合的多肽。HLA 蛋白序列可以从 IMGA/HLA 数据库中获得^[11],序列可从 <ftp://ebi.ac.uk/pub/database/imgt/mhc/hla/> 下载。核酸序列和蛋白质序列有多种格式如 FASTA 和 PEPTIDE 格式。结合肽模块预测最简单的方法是使用打分矩阵,当预测 9 肽的结合时,9×20 矩阵包含多肽中每个氨基酸在各个位点的数值,该方法可在 PeptideCheck 网点(<http://www.peptidecheck.org>)使用。输入一条多肽序列并选择一个等位基因,结果代表此肽与给定等位基因结合可能性的得分;也可直接输入一个蛋白质的序列,所有生成的多肽均被打分;也可以同时选择多个基因,结果为此肽与多个不同基因的打分。为了判断该得分是表示结合还是非结合,需要将此得分与阈值相比较。阈值的选择要依靠试验情况,如果要找最有可能结合的多肽,则选择较高阈值。PeptideCheck 建议的阈值是灵敏度曲线和特异性曲线的交叉点。

3 免疫信息学的应用

当前免疫信息学致力于指导修改 PCR 条件^[30]、抗体制

备^[31]、疫苗设计^[32]、预测癌症抗原及基因^[33-34]、细胞洗脱肽的比对^[35]以及各数据库的方法更新^[36-37]等研究。在食源性活性肽的研究上,刘盟梦等^[38]运用 BP 神经网络数学模型拟合蛋白水解过程与产物活性之间的关系,构建以加酶量、料水比、酶解体系游离谷氨酸浓度为神经网络输入和酶解产物·OH 清除率为输出的模型,该模型可用于杜蛎抗氧化活性肽的在线监控与可控制备。陈艳艳等^[39]采用生物传感器-人工神经网络建立了基于游离氨基酸含量对胶原蛋白酶酶解进程的预测模型,可实现对酶解过程的在线监控,以获得最大量的目标活性肽。孙国威等^[40]以“活性肽搜寻与蛋白模拟水解系统”为工具,选择碱性蛋白酶和中性蛋白酶对大豆蛋白进行模拟水解得到不同水平的肽段,并结合水解试验,以还原力和 DPPH 自由基清除率为指标评价两者的相关性。陈征松等^[41]利用 Microsoft Office XP 中的 Access XP 数据库软件建立了 3 个数据库系统,并与编制的“生物活性肽搜寻与酶解模拟系统”程序配合,实现单条、多条活性肽序列在食物蛋白质中的批量搜寻,并能找出活性肽含量的链长百分比、活性肽在蛋白质中的位置和前后氨基酸的种类,实现利用活性肽的不完全归纳预测其活性、利用单酶或复酶的模拟水解并标出水解产物中的活性肽及其功能。庞广昌等^[42]通过软件 The Swiss-PdbViewer(v.3.7)根据免疫活性肽序列的组成、氨基酸之间的键合力、非键合力、静电引力、扭矩等参数对其空间结构进行了模拟。当前免疫信息学主要在解决免疫学问题及疫苗设计方面发挥其强大的作用,目前食品领域也已开展结合信息学手段的各项相关研究,迈出了创新的第一步。将信息学手段与工程制备相结合,探寻出

了活性肽批量搜寻与可控制备的新方法,为筛选功能性肽并应用到食品领域提供了一个新思路,开辟了一条新渠道。但从目前看来,食品领域所利用的信息学工具和范围仍比较局限,因此将信息学应用到食品领域的空间相当广阔。将免疫信息学分析工具应用到免疫活性肽的预测和鉴定上,可拓宽功能性蛋白质和多肽的研究渠道,减少活性肽筛选的盲目性,有效提高筛选效率,进而推广至其他食源性活性肽的高效筛选,为开发新型功能性食品提供可靠依据。

4 展望

免疫信息学正以前所未有的速度发展,越来越多的研究小组参与到免疫信息学相关数据库和算法的改进中。3 种免疫信息学常用数据库中,IMGT 数据库因其资源的高度整合化,被认为是免疫信息学的国际参照,因此其查询和使用率极高,被广泛应用于基础医学研究、兽医研究、基因进化研究、诊断学和治疗学方法等多个领域,在免疫遗传学网络服务器的发展中发挥着重要作用。SYFPEITHI 数据库涵盖了许多 MHC 结合肽的信息,包括及时更新的 T 细胞表位信息,但由于 MHC II 类分子的结合槽结构相对 MHC I 类分子更为复杂,且 SYFPEITHI 数据库的预测结果以已出版的序列为基础,使得 SYFPEITHI 数据库更适用于 MHC I 类分子 T 细胞表位的预测,可靠性约 50%,在 MHC II 类分子的 T 细胞表位预测方面存在局限性。MHCBN 数据库可以链接一系列的网络工具进行交互式复杂查询从而获得更精确的信息,但数据库中的部分功能需要注册才能被使用。同其他数据库相比,MHCBN 数据库可直接进行 BLAST 比对以检验所查询序列的正确性,同时 MHCBN 数据库提供的锚着位点信息也更为详细,对 T 细胞表位预测更具针对性。

3 种活性肽与 MHC 分子结合能力的预测方法中,特征参数法针对性强,准确性高,且所有用户可免费使用。利用此方法首先需要从其他相关数据库中获得与特定 MHC 结合的多肽序列再产生相应的算法矩阵,该方法对数据的敏感程度相对机器算法较低,因此可节约大量时间成本,且结果与机器算法类似甚至更优。3D-QSAR 建模方法中,所有的结构均由理论生成并利用不同力场分析进行建模,模型通过最优拟合结果的叠合结构构建,并可剔除由计算产生的差异值;但前期参数的计算主成分分析过程较复杂,需较坚实的计算机基础。基于 MHC 分子模型的预测方法中,只需构建多肽结合能力的打分矩阵。但该方法的弊端是,打分矩阵的分值是假定多肽特定位点的特定氨基酸的出现频率与这些氨基酸在结合中的作用相关,如果肽库中的序列是合成的,合成肽对上述假定无效,需寻求另外的办法;且在选择阈值时需根据试验情况决定,建立特异性曲线与灵敏度曲线进行交叉分析。目前针对 MHC 分子的结合预测主要集中在 MHC I 类分子,MHC II 类分子由于其结构的复杂性及多肽结合槽的可变性使得预测工作变得更为困难。

尽管越来越多的算法被开发,但这些算法还需进行确实有效的实验验证,将体外试验和计算机分子方法结合起来,以提高预测方法的适用范围和能力。并且目前的文献报道

主要集中在疫苗设计、抗体制备等医学诊断和治疗学方面,在食源性功能活性肽筛选和预测中的应用报道还不多见。利用免疫信息学工具,不仅可以在分子或基因水平获得更直观的结果,归纳出结果的规律形式,更清晰地认识免疫活性的调控机制,也可在借鉴前人研究经验的基础上将这种信息预测和模拟方法与试验结合并应用到各个工程领域,增强免疫信息学工具的可信度,简化工程工艺步骤,并开发出针对食品领域的新方法或新工具,有望将食品领域推向新的发展巅峰。

参考文献

- [1] MEISEL Hans. Overview on milk protein-derived peptides[J]. International Dairy Journal, 1998, 8(5/6): 363-373.
- [2] 程媛, 曹慧, 徐斐, 等. 食源性蛋白中免疫活性肽的研究进展[J]. 食品科学, 2015, 36(17): 296-299.
- [3] 陈兆国, 冯新港, 米荣升, 等. 免疫信息学在抗原疫苗研发中的应用进展[J]. 中国人兽共患病学报, 2010, 26(5): 495-498.
- [4] ORTUTAY Csaba, SIERMALA Markku, VIHINEN Mauno. Molecular characterization of the immune system: emergence of proteins, processes, and domains[J]. Immunogenetics, 2007, 59(5): 333-348.
- [5] FLOWER Darren R. Vaccines in silico: the growth and power of immunoinformatics[J]. Biochemist, 2004, 26(4): 17-20.
- [6] GROOT Anne S De, BERZOFKY Jay A. From genome to vaccine—new immunoinformatics tools for vaccine design[J]. Methods, 2004, 34(4): 425-428.
- [7] LEFRANC Marie-Paule. IMGT®, the International Immunogenetics Information System® [J]. Nucleic Acids Research, 2009, 37(Database issue): D1 006.
- [8] LEFRANC Marie-Paule, CLEMENT O, KAAS Quentin, et al. IMGT-Choreography for Immunogenetics and Immunoinformatics[J]. Immunogenetics, 2004, 27: 55-77.
- [9] OKUBO K, SUGAWARA H, GOJOBORI T, et al. DDBJ in preparation for overview of research activities behind data submissions [J]. Nucleic Acids Research, 2006, 34 (Database issue): D6.
- [10] FOLCH G, BERTRAND J. IMGT/PRIMER-DB[J/OL]. Information System, 2004. [2018-02-17]. <http://www.imgt.org>.
- [11] ROBINSON James, WALLER Matthew, PARHAM Peter, et al. IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex[J]. Nucleic Acid Research Medical Journal, 2003, 31(1): 311-314.
- [12] KAAS Quentin, RUIZ Manuel, LEFRANC Marie-Paule. IMGT/3Dstructure-DB and IMGT/Structural Query, a database and a tool for immunoglobulin, T cell receptor and MHC structural data [J]. Nucleic Acids Research, 2004, 32(1): 208-210.
- [13] HORTON Roger, WILMING Laurens, RAND Vikki, et al. Gene map of the extended human MHC[J]. Nature Reviews Genetics, 2004, 5(12): 889.
- [14] MCGINNIS M B C, CHADWICK R, CONRAD M, et al. Ge-

- netic diversity of HLA; functional and medical implication[J]. *Human Immunology*, 1997, 59(9): 580-587.
- [15] MARSH Steven, PETER Parham, LIN Barber. HLA facts book[M]. [S.l.]: Academic Press, 2000: 208-236.
- [16] WU Xin, MICHAEL G Walker, LUO Jing-chu, et al. GBA server: EST-based digital gene expression profiling[J]. *Nucleic Acids Research*, 2005, 33(Web Server issue): 673-676.
- [17] DENNIS A Benson, ILENE Karsch-Mizrachi, DAVID J Lipman, et al. GenBank: update[J]. *Nucleic Acids Research*, 2007, 36(Database issue): 25-30.
- [18] TATENO Yoshio, SAITOU Naruya, OKUBO Kousaku, et al. DDBJ in collaboration with mass-sequencing teams on annotation[J]. *Nucleic Acids Research*, 2005, 33(Database issue): 25-28.
- [19] 孙继丽, 张工梁, 陈仁彪. HLA 数据库和 HLA 命名系统[J]. *中国输血杂志*, 2005, 18(2): 174-176.
- [20] RAMMENSEE Hans-Georg, BACHMANN J, EMMERICH N P N, et al. SYFPEITHI: database for MHC ligands and peptide motifs[J]. *Immunogenetics*, 1999, 50(4): 213-219.
- [21] WHEELER David L, CHURCH Deanna M, LASH Alex E, et al. Database resources of the National Center for Biotechnology Information[J]. *Nucleic Acids Research*, 2007, 35(Database issue): 5-12.
- [22] ROBINSON James, WALLER Matthew, PARHAM Peter, et al. IMGT/HLA Database: a sequence database for the human major histocompatibility complex[J]. *Nucleic Acids Research*, 2000, 55(3): 210-213.
- [23] BERMAN Helen, WESTBROOK John, FENG Z, et al. The protein data bank[J]. *Genetica*, 2000, 28(1): 235-242.
- [24] BRUSIC Vladimir, RUDY George, KYNE Anthony P, et al. MHCPEP, a database of MHC-binding peptides: update 1996[J]. *Nucleic Acids Research*, 1998, 26(1): 3 663-3 665.
- [25] BHASIN Manoj, SINGH Harpreet, RAGHAVA Gajendra Pal Singh. MHCBN: a comprehensive database of MHC binding and non-binding peptides[J]. *Bioinformatics*, 2003, 19(5): 665-666.
- [26] BLYTH Martin J, DOYTCHINOVA Irini A, FLOWER Darren. JenPep: a database of quantitative functional peptide data for immunology[J]. *Bioinformatics*, 2002, 18(3): 434-439.
- [27] SCHÖNBACH Christian, KOH Judice L Y, FLOWER Darren, et al. FIMM, a database of functional molecular immunology[J]. *Nucleic Acids Research*, 2002, 30(1): 226-229.
- [28] RECHE Pedro A, ZHANG Hong, GLUTTING John-Paul, et al. EPIMHC: a curated database of MHC-binding peptides for customized computational vaccinology [J]. *Bioinformatics*, 2005, 21(9): 2 140-2 141.
- [29] Tripos Associates Inc. SYBYL Software, Version 6.9[CP]. Morris, Peters, St. Louis: Tripos Associates Inc, 1985.
- [30] HOLCOMB Cherie L, RASTROU Melinda, WILLIAMS T C, et al. Next-generation sequencing can reveal in vitro-generated PCR crossover products: some artifactual sequences correspond to HLA alleles in the IMGT/HLA database[J]. *Tissue Antigens*, 2014, 83(1): 32-40.
- [31] LEFRANC Marie-Paule, EHRENMANN Francois, GINES-TOUX Chantal, et al. Use of IMGT(®) databases and tools for antibody engineering and humanization[J]. *Methods in Molecular Biology*, 2012, 907: 3-37.
- [32] LIAO Shu-jie, ZHANG Wei-na, HU Xiao-ji, et al. Preparation of HPV18 E7 peptide plus CpG vaccine and its immunologic effects in vitro[J]. *Chinese Medical Journal*, 2012, 92(23): 1 641-1 645.
- [33] ZHAO Wei-peng, LONG Hai-xia, ZHU Bo, et al. Prediction of HLA-A 2.1-restricted CTL epitopes from IGFBP7 antigen of lung carcinoma[J]. *中国人民解放军军医大学学报: 英文版*, 2009, 24(2): 63-68.
- [34] LI Zhang-qiu, ZHANG Mei-xia, HU Hai-yan, et al. [On predicting the T cell and B cell epitopes of platelet membrane glycoprotein II b/ III a antibody from human and mice][J]. *Journal of Biomedical Engineering*, 2010, 27(5): 1 146-1 151.
- [35] EVA Stodulková, NOVÁK Peter, DEININGER Sören-Oliver, et al. LC MALDI-TOF MS/MS and LC ESI FTMS analyses of HLA-B27 associated peptides isolated from peripheral blood cells[J]. *Immunology Letters*, 2008, 116(1): 79-85.
- [36] BHASIN Manoj, LATA Sneha, RAGHAVA Gajendra Pal Singh. Searching and mapping of T-cell epitopes, MHC binders, and TAP binders[J]. *Methods in Molecular Biology*, 2007, 409(409): 95-112.
- [37] SCHULER Mathias M, NASTKE Maria-Dorothea, STEVANOVIC Stefan. SYFPEITHI: database for searching and T-cell epitope prediction[J]. *Methods in Molecular Biology*, 2007, 409: 75-93.
- [38] 刘盟梦, 李银平, 延海莹, 等. 基于 BP 神经网络的牡蛎抗氧化活性肽制备工艺优化[J]. *食品工业科技*, 2016, 37(20): 206-210.
- [39] 陈艳艳, 侯虎, 陈铁军, 等. 基于酶传感器-人工神经网络的胶原蛋白酶解监控模型[J]. *中国食品学报*, 2016, 16(5): 167-173.
- [40] 孙国威, 乐国伟, 施用晖. 模拟酶解大豆 7S、11S 蛋白及其抗氧化活性的研究[J]. *食品工业科技*, 2010(7): 101-104.
- [41] 陈征松, 施用晖, 乐国伟, 等. 活性肽搜寻与蛋白模拟水解数据库的建立[J]. *计算机与应用化学*, 2007, 24(3): 331-334.
- [42] 庞广昌. 一种新型具抗菌作用的免疫活性肽及其抗菌机理[D]. 天津: 天津大学, 2007: 68-76.