

* 专题评述 *

免疫信息学

王月丹 陈慰峰*

北京大学医学部免疫学系 T 细胞研究室, 北京 100083

摘要 免疫信息学是在 21 世纪初才被建立的一门以信息学与现代免疫学为基础的交叉学科。在中外科学家的共同努力下, 免疫信息学的概念和理论体系正在逐步走向成熟, 其自身的发展也促进了现代免疫学和新型疫苗研究工作的进步。文中对免疫信息学概念的建立过程及免疫信息学的最新进展做一简要概述。

关键词 免疫信息学 抗原表位 疫苗设计

免疫信息学(Immunoinformatics)是以信息学和现代免疫学为基础的一个新的交叉学科^[1];是运用信息学的理论和方法解决免疫学问题的学科;同时也是研究免疫应答中免疫信息传递规律的科学。免疫信息学的研究内容是运用生物信息学的分析手段,利用现代免疫学实验的资料和现代免疫学理论,对免疫系统识别/传递的特有信号(特别是指抗原信号)进行研究,从而总结归纳出免疫系统信号识别/传递的规律,对免疫学现象进行科学的预测,并通过免疫学实验对预测的结果进行验证,以解决免疫诊断、免疫治疗和疫苗制备等方面的实际问题。免疫信息学分析^[2]就是利用免疫信息学的规律,对免疫学实验结果进行预测,再通过有效的免疫学实验对其预测的结果进行验证,从而能在短期内得到正确的结论。这将大幅度地减少免疫学研究的工作量,节约研究成本,促进现代免疫学的更快发展,使原来不能或不易进行的工作(如通过抗原表位分析寻找疫苗)能够顺利进行^[3-5]。因此,免疫信息学分析正在成为现代免疫学研究的重要研究工具。同时,随着生物信息学分析计算能力的提高

和现代免疫学实验技术的迅速发展,免疫信息学也得以飞速发展。在此,我们对免疫信息学与免疫信息学分析的出现与发展进行一个简要的介绍。

1 免疫信息学概念的提出及其发展

“免疫信息学”这个名词最早是由以色列科学家 Segel 在 2001 年提出的,是描述免疫系统中以细胞外化学物质(如细胞因子)为介质的各种生物反馈现象及其机制的国际合作网络^[6]。同年,亚洲大洋洲免疫学会联盟(FIMSA)理事会提出在其第三届学术大会中增设“生物信息学在免疫学研究中应用”的主题,负责学术大会筹备的我国科学家陈慰峰教授提出了用“Immunoinformatics”作为该主题讨论的题目。尽管大会由于 SARS 流行等原因被推迟了,但这是免疫学家第一次明确阐明了免疫信息学是以信息学与现代免疫学研究为基础的交叉学科,并提出现阶段的免疫信息学应主要以抗原表位(主要是肿瘤抗原表位)分析和免疫学计算国际合作网络的建设作为讨论的重点^[1]。2003 年 5 月,英国科学家 Flower 提出免疫信息学是生物信息学的一个

2004-02-16 收稿, 2004-03-26 收修改稿

* 通讯作者, E-mail: wfchen@public.bta.net.cn

新发展, 是运用信息学的手段分析免疫大分子活动的机制, 是通过计算进行疫苗或免疫学药物设计的重要工具^[3]. 同月, 我们发表了以免疫信息学分析为基础的 SARS 病毒表位研究结果, 并通过免疫信息学分析比较了 SARS 病毒与人类冠状病毒 HCV-229E 的免疫应答差异性的可能机制, 并在此基础上经过免疫信息学计算和免疫学实验, 从千余个可能的 T 细胞表位中确定了两个人 HLA-A2 分子识别的 SARS 表面蛋白特有的 T 细胞表位^[2], 以此证明了免疫信息学在研究现代病原微生物免疫学与疫苗学研究中的重要作用. 同年, 新加坡科学家 Brusic 提出免疫信息学是一门新的信息科学, 其主要发展包括了免疫数据库的建立和查询、序列分析、免疫分子结构模型分析、免疫系统数学模型分析、免疫信息学实验室研究、免疫学实验统计学和免疫基因组学(Immunogenomics)等方面的进展^[1]. 这些工作为免疫信息学理论的形成奠定了基础, 并为其今后的发展方向指明了道路. 尽管“免疫信息学”的提出仅有不足 3 年的时间, 但是在上述科学家的不懈努力下, 免疫信息学的概念已经被人们逐渐接受, 学科体系也已初步形成, 并进入了一个迅速发展的新阶段.

2 免疫信息学的实验基础与免疫信息学理论基础的建立

免疫信息学是以免疫学实验资料为基础并通过信息学的手段进行研究, 以总结归纳出其中的一些规律, 建立相应的结构数学模型^[7]. 应用这些数学模型对以后的实验结果进行预测, 再通过免疫学实验对预测的结果进行验证, 并根据实验验证的结果对结构数学模型进行修正或对采用的信息学手段进行优化. 因此, 免疫信息学是以完整和正确的免疫学实验结果为基础的, 没有这些结果或这些结果不完善, 免疫信息学的预测结果就是不准确的, 甚至是毫无价值的^[8]. 免疫信息学本身也是在免疫学实验结果大量积累之后, 才逐渐形成和发展的. 例如, 现在被广泛使用的用于各种抗原分子表位预测的网上数据库能够较准确地预测出抗原分子上能与人白细胞抗原(HLA)分子结合的 T 细胞表位序列(如 SYFPEITHI 系统^[9, 10] 目前预测 HLA-A2 识别的表位可靠性为 80% 左右), 主要是依靠收集经过

实验验证的各种能与 HLA 分子结合的多肽序列(如 SYFPEITHI 系统目前收集了超过 33000 个能与各种组织相容性抗原复合体(MHC)分子结合的表位多肽的实验数据), 并通过一系列计算将组成这些多肽序列的氨基酸残基按其所在的位置进行评分, 据此在未知抗原分子中找到可能的抗原表位位点. 随着更多新的表位多肽的被发现和被补充入数据库, 这些网络数据库的预测准确性也在不断地提高. 然而, 这些实验数据的积累往往需要很长的时间. 例如, 1987 年人类发现了第一个 HLA-A2 识别的抗原表位(流感病毒基质蛋白 p58-66)^[11], 1994 年建立了能较准确预测抗原表位的 SYFPEITHI 数据库^[9, 10](最初的预测准确率约为 12.5%); 至 2002 年 11 月 SYFPEITHI 数据库的最佳预测可信性(以 HLA-A2 为例)达到 80% 以上. 这前后共用了 15 年的时间进行实验数据的积累和分析.

如同理论物理学是建立在对物理学理论研究的基础上一样, 免疫信息学也是建立在免疫学研究各种理论基础之上的. 细胞因子和免疫系统网络调节的理论、免疫细胞的克隆选择学说、抗原肽-MHC 复合物识别理论和免疫细胞双信号活化理论等均是免疫信息学建立的必要的理论基础. 目前发展最快的免疫信息学领域——抗原分子表位的预测, 就是建立在免疫系统抗原识别和提呈理论基础之上的, 即抗原信号是通过抗原肽-MHC-T 细胞受体(TCR)三元体^[12]传递的, 每一个个体中均具有数目庞大的 T 细胞受体库. 根据这一理论, 人们在考虑表位预测时更多地考虑了抗原肽-MHC 分子的稳定性, 而不是考虑个体 T 细胞受体的因素, 从而大大降低了预测的难度和计算量, 使表位预测成为可能. 因此, 免疫信息学的发展不仅依赖免疫学实验结果的完整与准确, 还要依靠正确的免疫学理论作为其分析和总结规律的基础. 同时, 免疫信息学的发展也为现代免疫学的持续发展提供了新的手段.

3 免疫信息学的应用与免疫信息学分析方法的探索

尽管“免疫信息学”这一名词的提出是在本世纪初, 但其应用至少可以追溯到 20 世纪 80 年代. 免疫信息学涉及的范围很广, 研究的内容也很多. 根据理论来源和研究方向的不同, 主要可以分为细

胞因子反馈网络研究、免疫分子结构数学模型及数据库的建立和使用^[13, 14]、免疫信息学实验室研究、免疫学实验统计学及与免疫基因组学有关的免疫信息学研究等多个领域。抗原分子表位的预测是目前免疫信息学发展最快的领域。抗原分子表位分析是建立在抗原提呈与识别的 MHC-抗原肽-TCR 三元体形成的 X 射线衍射实验结果^[15]和相应的理论研究基础上的。该理论最早来源于以对人类及动物 MHC 多样性研究为主要内容的免疫基因组学研究。1994 年, Rammensee 等建立了以抗原分子表位预测为主要功能的网上分析数据库^[9, 10], 这标志着抗原分子表位的免疫信息学预测开始从免疫基因组学这一有较长历史的学科^[16]中分离了出来。现在, 越来越多的科学家正在运用各种各样的网上分析数据库(见表 1)或软件分析预测各种抗原分子(如肿瘤抗原分子、与移植排斥相关的抗原分子以及各种病毒蛋白抗原分子^[17])的抗原表位。其中, 肿瘤抗原分子表位预测与 HIV^[18]和 SARS-CoV 等病毒抗原表位预测, 是研究预防和治疗这些致死性疾病抗原肽疫苗的重要工具。然而, 由于数据库中积累的实验结果太少(如 SYFPEITHI 数据库中, 数据最完整的 HLA-A2 分子识别的九肽仅不足 100

个^[19]), 依据的免疫学理论体系还存在不足(如抗原加工相关转运体(TAP)识别的多肽序列特异性的规律及其机制还不清楚), 采用的计算方法还不完善(如为减少计算量, 大多数系统只考虑了每个氨基酸残基自己的因素而较少考虑不同位置的氨基酸残基之间相互作用对三元体形成的影响), 因此其预测准确性受到很大影响。在期待数据库自身完善的同时, 运用免疫信息学进行分析时, 往往要采用多种方法, 遵循一定的规律, 对预测的结果进行综合评价, 才能获得比较好的效果, 最终达到提高效率 and 节约成本的效果。我们在利用免疫信息学分析 SARS 病毒(基因组 RNA 约为 29 kb^[20])抗原 HLA-A2 相关表位时, 首先查询 SYFPEITHI 数据库和 EPITOPEPREDICT 数据库^[21, 22], 从近 1 万个可能的抗原表位中筛选出 1000 余个可能的九肽序列; 然后, 通过欧洲生物信息研究所(EBI)的 CLUSTAL W Multiple Sequence Alignments 序列比较分析系统^[23], 将这 1000 余个九肽序列与已知的人类冠状病毒 HCV-229E 比较, 去除其中相同的序列; 在此基础上, 利用多肽结构分析软件 HYPERCHEM^[24]对多肽结构

表 1 常用的网上免疫信息学数据库名称及网址

数据库名称	网 址	用 途
PREDEP	http://bioinfo.md.huji.ac.il/marg/Teppred/mhc-bind/	
MMBPred	http://www.imtech.res.in/raghava/mmbpred/	
SVMHC	http://www.sbc.su.se/svmhc/information.html	MHC I 分子结合肽预测数据库
ProPred-I	http://www.imtech.res.in/raghava/propred1/	
BIMAS	http://bimas.dert.nih.gov/molbio/hla_bind/	
ProPred	http://www.imtech.res.in/raghava/propred/	
EPIPREDICT	http://www.epipredict.de/index.html	MHC II 分子结合抗原肽预测数据库
HLA Ligand Motif Database	http://hlligand.ouhsc.edu/prediction.htm	HLA 配体、 提呈抗原肽及 CTL 表位预测数据库
RANKPEP	http://mif.dfci.harvard.edu/Tools/rankpep.html	
nHLAPred	http://www.imtech.res.in/raghava/nhlapred/	
MHC BN	http://www.imtech.res.in/raghava/mhcbn/	MHC 及 TAP 结合肽数据库
JenPep	http://www.jenner.ac.uk/jenpep/	
MHC PEP	http://wehih.wehi.edu.au/mhcepep/	MHC 结合肽数据库
SYFPEITHI	http://www.uni-tuebingen.de/uni/kxi/	能被提呈的 MHC 结合肽数据库, 具有 抗原肽预测功能
HIV Molecular Immunology Database	http://www.hiv.lanl.gov/content/immunology/	HIV 上的 CTL 和 Th 细胞识别表位数据库
BCIPEP	http://www.imtech.res.in/raghava/bcipep/	B 细胞识别表位数据库
FIMM	http://sdmc.lit.org.sg:8080/fimm/	抗原、MHC 及 TCR 等信息的免疫基因 组学数据库
IMGT	http://www.ebi.ac.uk/imgt/	

进行分析, 排除不同氨基酸残基之间相互影响形成二级结构的多肽, 并从中选取分子弹性较好的多肽; 然后, 利用分析小分子与大分子结合稳定性的 AUTODOCK 软件^[25] 从中筛选出可以与 HLA-A2 分子抗原结合槽结合并形成稳定结构的肽, 这一步不仅是免疫信息学计算分析的最后一步, 也是其中计算量最大的一部分. 利用上述计算的结果, 我们合成了部分抗原表位多肽, 进行免疫信息学的实验工作, 先后通过 T2 细胞结合实验、 γ -干扰素 ELISPOT 分析实验^[26, 27] 和 HLA-A2: Ig DIMERX 标记实验^[28], 确定了 SARS 病毒的 HLA-A2 相关抗原表位^[29]. 与传统的合成抗原肽穷举法^[30] 和酸洗脱抗原肽鉴定法^[31] 相比, 免疫信息学分析法不仅大大减少了抗原肽的合成, 降低了技术难度, 而且节约了研究时间, 提高了研究的准确性. 在进行免疫信息学分析时, 应该根据自己研究的内容制定切实可行的研究方案. 目前, 将免疫信息学计算与免疫信息学实验相结合的研究结果, 更具有可信性和说服力.

4 免疫信息学的现状和前景展望

免疫信息学是本世纪才被提出的一个全新的学科, 至今未满 3 年, 必定还存在着很多不完善的方面. 由于免疫信息学分析方法的流程还未得到全面的优化, 在进行免疫信息学预测时人们往往只采用单一的分析方法, 所得到的结论还难免有所偏差. 如通过免疫信息学进行抗原表位的预测时, 其准确率一直是人们质疑的焦点, 因此纯粹的免疫信息学理论分析与计算还一时难以被主流的免疫学家所接受. 但随着实验数据的积累和对免疫应答信息内在规律的揭示, 免疫信息学理论分析终将成为免疫学研究的重要组成部分. 现在, 免疫信息学理论分析和免疫信息学实验相结合的研究方法正在逐步取代合成多肽穷举法和 MHC 分子洗脱肽鉴定法, 成为寻找抗原表位的主要手段, 并为人们所广泛接受. 由于免疫信息学将现代信息学技术与现代免疫学技术紧密地结合在一起, 因此它的迅速发展极大地促进了现代免疫学和疫苗学研究的发展. 从 2003 年 5 月起, 至今已经有十余篇采用免疫信息学或免疫信

息学分析为关键词的论文发表. 而实际上采用免疫信息学理论或免疫信息学分析技术为主要内容或重要分析手段的论文则有近百篇, 这充分显示了其旺盛的生命力. 我国从事免疫学研究工作的科学家和负责免疫学科研工作管理的领导一直十分重视信息学在免疫学研究中的重要地位, 并力争在该领域取得国际领先地位. 从 2001 年起, 根据陈慰峰、沈倍奋和张学光教授等免疫学专家的提议, 国家自然科学基金委员会生命科学部对如何促进信息学在免疫学研究中的应用组织了多次的研讨, 这些研讨不仅在我国科学工作者中普及了免疫信息学的知识, 而且最终使“免疫识别相关的结构信息的研究”这一课题列入了“十五”第二批国家自然科学基金重大项目之一¹⁾, 奠定了发展我国免疫信息学研究的基础. 同时, 国家自然科学基金委员会还大力支持我国科学工作者利用免疫信息学的手段进行应用的研究, 批准资助进行了以免疫信息学为主要研究内容的“SARS 相关冠状病毒抗原肽疫苗的基础研究”项目. 随着免疫信息学的飞速发展, 人们免疫学和信息学的研究领域越来越关注. 2004 年 9 月将在意大利的 Catania 市举行第一届世界免疫信息学会议, 由于我国学者首先报道了运用免疫信息学技术分析 SARS 病毒抗原表位变化规律的结果, 也受到了会议的邀请.

免疫信息学是伴随着 21 世纪的到来而产生的新兴现代科学, 越来越多的人对其的态度由怀疑转向肯定和期待. 在包括我国免疫信息学工作者在内的免疫信息学专家的努力下, 免疫信息学及免疫信息学分析将不断走向成熟, 并最终推动现代免疫学发展进入一个新的时代.

参 考 文 献

- 1 Busic V, et al. Immunoinformatics—the new kid in town. *Novartis Found Symp*, 2003, 254: 3
- 2 Wang Y, et al. Immunoinformatic analysis for the epitopes on SARS virus surface protein. *Beijing Da Xue Xue Bao*, 2003, 35 (Suppl): 70 (in Chinese)
- 3 Flower D R. Databases and data mining for computational vaccinology. *Curr Opin Drug Discov Devel*, 2003, 6: 396
- 4 Gulukota K. Immunoinformatics in personalized medicine. *Novartis*

1) 见国家自然科学基金委员会相关网页 <http://www.nsf.gov.cn/nsf/cen/02/htmlcreated/2003zd/002.htm>

- Found Symp, 2003, 254: 43
- 5 Groot A S, et al. Genome-derived vaccines. *Expert Rev Vaccines*, 2004, 3: 59
 - 6 Segel L A. Controlling the immune system; Diffuse feedback via a diffuse informational network. *Novartis Found Symp*, 2001, 239: 31
 - 7 Rammensee H G. Immunoinformatics: Bioinformatic strategies for better understanding of immune function. *Introduction Novartis Found Symp*, 2003, 254: 1
 - 8 Blythe M J, et al. JenPep: A database of quantitative functional peptide data for immunology. *Bioinformatics*, 2002, 18: 434
 - 9 Rammensee H, et al. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics*, 1999, 50: 213
 - 10 Rammensee H, et al. MHC ligands and peptide motifs: 1st listing. *Immunogenetics*, 1995, 41: 178
 - 11 Gotch F, et al. Cytotoxic T lymphocytes recognize a fragment of influenza virus matrix protein in association with HLA-A2. *Nature*, 1987, 326: 881
 - 12 陈慰峰主编. 医学免疫学. (第三版)北京: 人民卫生出版社, 2000. 147
 - 13 Lefranc M P, et al. IMGT, the international ImMunoGeneTics information system, <http://imgt.cines.fr>; the reference in immunoinformatics. *Stud Health Technol Inform*, 2003, 95: 74
 - 14 Lefranc M P. IMGT-ONTOLOGY and IMGT databases: tools and web resources for immunogenetics and immunoinformatics. *Mol Immunol*, 2004, 40: 647
 - 15 Fremont D H, et al. Crystal structures of two viral peptides in complex with murine MHC class I H-2Kb. *Science*, 1992, 257: 919
 - 16 Levine P. Blood groups and immunogenetics; Rare red cell genotypes; Some illustrative cases. *Acta Genet Stat Med*, 1956~1957, 6: 515
 - 17 Petrovsky N, et al. The future for computational modelling and prediction systems in clinical immunology. *Novartis Found Symp*, 2003, 254: 23
 - 18 Claverie J M, et al. T-immunogenic peptides are constituted of rare sequence patterns. Use in the identification of T epitopes in the human immunodeficiency virus gag protein. *Eur J Immunol*, 1988, 18: 1547
 - 19 Petrovsky N, et al. Bioinformatic strategies for better understanding of immune function. *Silico Biology*, 2003, 3: 34
 - 20 Rota P A, et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science*, 2003, 300: 1394
 - 21 Hansson L, et al. T-cell epitopes within the complementarity-determining and framework regions of the tumor-derived immunoglobulin heavy chain in multiple myeloma. *Blood*, 2003, 101: 4930
 - 22 Wang Q J, et al. Identification of an HLA A *0201-restricted CD8 (+) T-cell epitope for the glycoprotein B homolog of human herpesvirus 8. *Blood*, 2002, 99: 3360
 - 23 Ramirez E, et al. Genetic characterization and phylogeny of human T-cell lymphotropic virus type I from Chile. *Virus Res*, 2003, 84: 135
 - 24 Bartulewicz D, et al. Molecular modelling, synthesis and antitumour activity of carbocyclic analogues of netropsin and distamycin—new carriers of alkylating elements. *Acta Biochim Pol*, 2000, 47: 23
 - 25 Jois S D, et al. A peptide derived from LFA-1 protein that modulates T-cell adhesion binds to soluble ICAM-1 protein. *J Biomol Struct Dyn*, 2003, 20: 635
 - 26 Zhang H G, et al. Functional supertype of HLA-A2 in the presentation of Flu matrix p58-66 to induce CD8+ T-cell response in a Northern Chinese population. *Tissue Antigens*, 2003, 62: 285
 - 27 Gnjatic S, et al. Cross-presentation of HLA class I epitopes from exogenous NY-ESO-1 polypeptides by nonprofessional APCs. *J Immunol*, 2003, 170: 1191
 - 28 Schneck J P, et al. Monitoring antigen specific T cells using MHC-Ig dimers. In: *Current Protocols in Immunology*. New York: John Wiley and Sons, 2000. 17. 2~17. 2, 17
 - 29 Wang Y D, et al. T cell epitopes in severe acute respiratory syndrome (SARS)-CoV virus spike protein elicit specific T cell immune response in recovered SARS patients. *J Virol*, 2004, 78: 5612
 - 30 Kast W M, et al. Human leukocyte antigen-A2. 1 restricted candidate cytotoxic T lymphocyte epitopes of human papillomavirus type 16 E6 and E7 proteins identified by using the processing-defective human cell line T2. *J Immunother*, 1993, 14: 115
 - 31 Storkus W J, et al. Identification of T-cell epitopes: Rapid isolation of class I-presented peptides from viable cells by mild acid elution. *J Immunother*, 1993, 14: 94