

现代生物技术前沿
IMMUNOINFORMATICS
PREDICTING IMMUNOGENICITY IN SILICO

〔英〕D.R. 弗劳尔 主编

吴玉章 等 译

免疫信息学

——计算机辅助预测免疫原性



科学出版社

www.sciencep.com

现代生物技术前沿

63
4022-8605-10

IMMUNOINFORMATICS PREDICTING IMMUNOGENICITY IN SILICO

(英) D.R. 弗劳尔 主编

吴玉章 等 译

免疫信息学

——计算机辅助预测免疫原性



科学出版社

北京

Q939.91
F590

图字: 01-2008-5504

内 容 简 介

随着基因组学、计算机技术以及免疫学的快速发展,免疫信息学已经成为一个新兴的且逐步完善的研究领域。免疫信息学分析就是利用免疫学的规律,对免疫学实验结果进行预测,再通过有效的免疫学实验进行验证,从而大幅度地减少免疫学研究的工作量,节约研究成本,促进现代免疫学的发展。本书系统介绍了免疫信息学的概念、产生和发展、相关数据库、研究方法及其应用等,特别着重于抗原性的预测、分析和计算机辅助疫苗设计,并介绍了一些复杂软件的使用方法,因此理论性和实用性都很强。

本书定位于免疫信息学初学者,尤其适合本科生、研究生,对于免疫学研究工作者也有很大的参考价值。

Translation from the English Language Edition;

Immunoinformatics edited by Darren R. Flower

Copyright ©2007 Humana Press Inc.

Springer is a part of Springer Science+Business Media.

All Rights Reserved.

图书在版编目(CIP) 数据

免疫信息学: 计算机辅助预测免疫原性/ (英) 弗劳尔 (Flower, D. R.) 主编;
吴玉章等译. —北京: 科学出版社, 2010

(现代生物技术前沿)

ISBN 978-7-03-026270-7

I. 免… II. ①弗…②吴… III. 免疫学-生物信息论 IV. Q939.91

中国版本图书馆 CIP 数据核字 (2009) 第 236655 号

责任编辑: 莫结胜 刘 晶/ 责任校对: 张怡君

责任印制: 钱玉芬/ 封面设计: 陈 敬

科 学 出 版 社 出 版

北京东黄城根北街16号

邮政编码: 100717

<http://www.sciencep.com>

源海印刷有限责任公司 印刷

科学出版社发行 各地新华书店经销

*

2010年1月第 一 版 开本: 787×1092 1/16

2010年1月第一次印刷 印张: 18 1/4 插页: 2

印数: 1—2 500 字数: 413 000

定价: 68.00 元

(如有印装质量问题, 我社负责调换)

编者

- MANOJ K. BHASIN • *Institute of Microbial Technology, Chandigarh, India,
and Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA*
- RAINER BLASCZYK • *Institute for Transfusion Medicine, Hannover Medical
School, Hannover, Germany*
- FRANK R. BURDEN • *SciMetrics, Harrow Enterprises Pty. Ltd., Fitzroy,
Victoria, Australia*
- PETER V. COVENEY • *Centre for Computational Science, Chemistry
Department, University College of London, London, UK*
- YANG DAI • *Bioengineering Bioinformatics, University of Illinois at Chicago,
Chicago, IL*
- MATTHEW N. DAVIES • *The Jenner Institute, University of Oxford, Berkshire,
UK*
- DAVID S. DELUCA • *Institute for Transfusion Medicine, Hannover Medical
School, Hannover, Germany*
- PIERRE DÖNNES • *Division for Simulation of Biological Systems, Eberhard
Karls University Tübingen, Tübingen, Germany*
- IRINI A. DOYTCHINOVA • *The Jenner Institute, University of Oxford,
Berkshire, UK*
- DARREN R. FLOWER • *The Jenner Institute, University of Oxford, Berkshire,
UK*
- PINGPING GUAN • *Computational Biology Group, John Innes Centre,
Norwich, UK*
- CHANNA K. HATTOTUWAGAMA • *The Jenner Institute, University of Oxford,
Berkshire, UK*
- LEI HUANG • *Bioengineering Bioinformatics, University of Illinois
at Chicago, Chicago, IL*
- PANDJASSARAME KANGUEANE • *School of Mechanical and Aerospace
Engineering, NANYANG Technological University, Singapore*
- OLEKSIY KARPENKO • *Bioengineering Bioinformatics, University of Illinois
at Chicago, Chicago, IL*
- SNEH LATA • *Institute of Microbial Technology, Chandigarh, India*

- MARIE-PAULE LEFRANC • *Institut Universitaire de France, Laboratoire d'ImmunoGénétique Moléculaire, LIGM, Université Montpellier II, UPR CNRS 1142, Institut de Génétique Humaine, Montpellier Cedex, France*
- TONGBIN LI • *Department of Neuroscience, University of Minnesota, Minneapolis, MN*
- THY-HOU LIN • *Institute of Molecular Medicine and Department of Life Science, National Tsing Hua University, Hsinchu, Taiwan*
- WEN LIU • *Department of Neuroscience, University of Minnesota, Minneapolis, MN*
- RONNA REUBEN MALLIOS • *Grants and Research Office, Fresno, CA*
- STEVEN G. E. MARSH • *Department of Haematology, Anthony Nolan Research Institute, Royal Free Hospital, Hampstead, London, UK*
- XIANGSHAN MENG • *Department of Neuroscience, University of Minnesota, Minneapolis, MN*
- NAVEEN MURUGAN • *Bioengineering Bioinformatics, University of Illinois at Chicago, Chicago, IL*
- MARIA-DOROTHEA NASTKE • *Department of Immunology, Institute for Cell Biology, University of Tübingen, Tübingen, Germany*
- GAJENDRA P. S. RAGHAVA • *Institute of Microbial Technology, Chandigarh, India; and UAMS, BRCII, Little Rock, AR*
- SHOBA RANGANATHAN • *Department of Chemistry and Biomolecular Sciences & Biotechnology Research Institute, Macquarie University, New South Wales, Australia; and Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore*
- PEDRO A. RECHE • *Laboratory of Immunobiology and Department of Medical Oncology, Dana-Farber Cancer Institute; and Department of Medicine, Harvard Medical School, Boston, MA*
- ELLIS L. REINHERZ • *Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA*
- JAMES ROBINSON • *Anthony Nolan Research Institute, Royal Free Hospital, Hampstead, London, UK*
- SUDIPTO SAHA • *Institute of Microbial Technology, Chandigarh, India*
- MEENA KISHORE SAKHARKAR • *School of Mechanical and Aerospace Engineering, NANYANG Technological University, Singapore*
- MATHIAS M. SCHULER • *Department of Immunology, Institute for Cell Biology, University of Tübingen, Tübingen, Germany*
- RICHARD SIMON • *Biometric Research Branch, National Cancer Institute, National Institutes of Health, Rockville, MD*

- MAHENDER KUMAR SINGH • *Institute of Microbial Technology, Chandigarh, India*
- SHILPY SRIVASTAVA • *Institute of Microbial Technology, Chandigarh, India*
- STEFAN STEVANOVIĆ • *Department of Immunology, Institute for Cell Biology, University of Tübingen, Tübingen, Germany*
- MYONG-HEE SUNG • *Laboratory of Receptor Biology and Gene Expression, Staff Scientist National Cancer Institute, Bethesda, MD*
- JOO CHUAN TONG • *Department of Biochemistry, National University of Singapore, Singapore, and Institute for Infocomm Research, Singapore*
- GRISH C. VARSHNEY • *Institute of Microbial Technology, Chandigarh, India*
- JI WAN • *Department of Neuroscience, University of Minnesota, Minneapolis, MN*
- SHUNZHOU WAN • *Centre for Computational Science, Chemistry Department, University College of London, London, UK*
- DAVID A. WINKLER • *Centre for Complexity in Drug Discovery, CSIRO Molecular and Health Technologies, Clayton, Australia*
- YINGDONG ZHAO • *National Cancer Institute, National Institutes of Health, Rockville, MD*

译者序

现代生命科学的发展在很大程度上依赖于实验科学的进步,使得科学数据大量、快速积累。与实验科学相比,理论科学的发展相对滞后,因此从某种意义上看,生命科学的现状是 “We know everything, but understand nothing”。在实验研究中我们已经具备了 *in vitro*、*in vivo*、*ex vivo* 等研究体系,但缺少深入的数据分析与理论思考。后基因组时代的到来和计算机科学的发展,催生了生物信息学科,科学研究增加了 *in silico* 研究系统,可以帮助我们进行目的性更强的研究,已经开始并越来越明显地促进了我们对生命科学规律特别是分子层面的认识 and 理论思考。

免疫系统是一个复杂的系统。众多的细胞和分子参与了特异性免疫应答,而更多的细胞和分子构筑了天然免疫防线。据估计,自然界至少有 10^{17} 种抗原,其中已知有 10^{11} 种以上由 9 肽组成的线形表位和更多的构象型表位;机体至少有 10^{13} 个 MHC 分子、 $10^7 \sim 10^{15}$ 个不同的 T 细胞受体、 10^{12} 个不同的 B 细胞克隆。如此复杂的系统组分通过相互作用在不同时空表现不同功能状态,从而与生命现象和疾病相联系。近 30 年来,免疫学发展迅速,其数据和知识积累约每 4 年增加 1 倍,直到今日,增长的速度也没有减慢。在此背景下,我们迎来了“组学时代”,免疫组学 (immunomics)、免疫信息学 (immunoinformatics) 应运而生。它们能帮助我们认识免疫应答的全貌,解读免疫应答的规律,进行目的性更强的研究。

全军免疫学研究所的主要研究方向之一就是“免疫识别、分子设计与抗原工程”。在 20 年的研究过程中,我们采用了信息学理论和技术,指导我们进行目的性更强的实验研究,特别是在表位预测、表位描图、表位改造、抗原提呈等方面的研究获益匪浅,但苦于没有一本免疫信息学方面的专著供初学者和研究生们系统学习相关知识。正在我们撰写中文《免疫信息学导论》一书的时候,英国免疫信息学家 Darren R. Flower 主编的 *Immunoinformatics: Predicting Immunogenicity In Silico* 出版问世了。我们受科学出版社委托翻译这本国际上第一部免疫信息学专著。这本专著系统介绍了免疫信息学的概念、产生和发展、相关数据库、研究方法及其应用等,特别着重于抗原性的预测、分析和计算机辅助疫苗设计,介绍了一些复杂软件的使用方法,使得该书理论性、实用性都很强。该书内容涉及学科、概念众多,但表述简练、图文并茂,便于理解,看了该书,读者就可以动手应用免疫信息学的相关技术了。该书定位于初学者,尤其适合本科生、研究生,对于免疫学研究工作者也有很大的参考价值,作为免疫学研究的工具书,必将对我国免疫学教学、科研起到很好的促进作用。

该书中文版由长期从事免疫信息学研究工作的尚小云博士、林治华教授、王莉副教授、王书峰博士和王远强老师承担主要翻译工作。在中文版出版之际,谨向他们表示深

切谢意。在翻译过程中还得到了科学出版社特别是莫结胜编辑的大力支持,特表由衷谢意。

由于译者的学识所限，中文版中肯定存在不妥甚至错误之处，欢迎广大读者提出宝贵意见和建议，以便今后改进。

吴玉章

全军免疫学研究所

2009 年 6 月 27 日

前言

本书阐述的是免疫信息学相关内容及其在免疫原性预测中日益重要的作用。所谓免疫原性,即当病原体暴露于免疫系统监视下时,诱导特异性免疫应答的能力。免疫信息学是将基于计算机科学的信息技术应用于免疫系统中分子之间相互作用的一门学科。近年来,免疫信息学在科学前沿领域和实际应用方面的发展为世人瞩目,除用于免疫原性的预测之外,还可参与解决免疫学的几乎所有问题。尽管许多人对免疫信息学在疫苗发现和研究中的实际应用尚存疑问,但能理解并充分利用其潜能的人已将该方法视为一种重要工具。《免疫信息学——计算机辅助预测免疫原性》一书是渴望紧跟此新兴技术者的入门读物。本书并不打算无所不包,而是重点关注免疫信息学中的重要领域,试图让读者对免疫信息学的地位及其发展趋势有所了解。我们希望本书不仅能吸引读者,而且可为免疫信息学在免疫学和疫苗学中的应用提供深厚的背景知识。由于系统生物学正高速发展并影响着其他领域,毫无疑问,生物科学包括免疫学的未来必然属于那些有能力将生物科学的实验和理论紧密结合的人,因为他们能一直轻松地将实验技术和计算科学相结合。作为一门交叉学科,免疫信息学要取得支配地位尚需时日,但这天终将到来,正如那些身穿白大褂、只知道手持吸液管反对自动化的勒德分子(Luddite)和只会傻瞪着电脑屏幕的笨拙的人最后都将被时代抛弃。科学研究作为一个整体,应当尽可能全身心地接受免疫信息学带来的变化。相信假以时日,人们对免疫信息学的怀疑和排斥就将消除。因此,翻开《免疫信息学——计算机辅助预测免疫原性》这本书吧,获取其中的智慧,品味其中的精华,收集深刻的见解,同时忘记其中诸多的瑕疵、矛盾、缺点及遗漏,铭记最重要的是从中学习并尽量享受学习的乐趣。

感谢本书所有作者,没有他们参加撰写,本书不可能问世,他们的工作质量也极为突出。感谢《分子生物学方法》丛书的总编 John Walker 教授,他的帮助和鼓励从未间断,使得本书从开始到出版以来的压力大为缓解。同时,要感谢 Humana 全体职员们在本书的编审和出版中给予的宝贵帮助,他们对本书的内容都给予了补充和完善。对于大家的帮助和建议我深表感激,本书中出现的一切错误以及遗漏由本人承担责任。

要特别感谢我的研究组成员。没有他们的帮助,本书就不可能出版,他们是: Dr. Irini Doytchinova、Valerie Walshe、Martin Blythe、Christianna Zygouri、Debra Clayton (née Taylor)、Shelley Hemsley、Christopher Toseland、Kelly Paine、Dr. Ping-ping Guan、Dr. Paul Taylor、Dr. Helen McSparron、Dr. Matthew N. Davies、Dr. Channa Hattotuwigama 和 Dr. Shunzhou Wan。也要感谢其他一些成员所提供的帮助和极富见地的讨论,他们是: Prof. Peter Beverley、Dr. Persephone Borrow、Dr. Shirley Ellis、Dr. Simon Wong、Dr. Helen Bodmer、Dr. Sam Hou、Dr. Lisa Hy-

land, Dr. David Tough、Dr. Elma Tchillian 和 Dr. Josef Walker。还要感谢我在 EJIIVR 以及 Compton 动物卫生研究所 (IAH) 的同事和合作者所给予的密切合作。最后, 感谢 Prof. Terri Attwood、Prof. Peter Coveney、Prof. Vladimir Brusic 和 Dr. Anne DeGroot。

Darren R. Flower

(吴玉章译 尚小云校)

目 录

译者序

前言

1 免疫信息学与计算机辅助预测免疫原性: 导论	1
-------------------------------	---

第 1 部分 数 据 库

2 国际免疫遗传学信息系统 (IMGT)	10
3 IMGT/HLA 数据库	27
4 免疫多态性数据库: IPD	39
5 T 细胞表位查询和预测数据库: SYFPEITHI	49
6 T 细胞表位、MHC 结合肽和 TAP 结合肽的搜寻及描图	61
7 Bcipep 数据库中 B 细胞表位的搜寻及描图	75
8 半抗原、载体蛋白和抗半抗原抗体的检索	84

第 2 部分 HLA 超型鉴定

9 基于 GRID/CPCA 和层次聚类法的 HLA 超型分类	95
10 HLA-A2 超型的结构基础	104
11 基于 MHC 结合肽库定义 MHC 超型	110
12 基于肽结合凹槽静电分布图的 HLA-I 类等位基因分型	117

第 3 部分 肽与 MHC 结合能力的预测

13 特征参数法预测 MHC 结合肽	123
14 机器学习技术预测 MHC 结合肽	133
15 人工智能方法预测 T 细胞表位	144
16 小鼠 MHC-多肽亲和力的预测	150
17 3D-QSAR 模型预测 MHC-多肽亲和力	164
18 基于 MHC 分子模型预测表位肽	175
19 基于支持向量机预测 MHC 结合肽	182
20 应用 SVRMHC 方法预测多肽与 MHC 分子的结合亲和力	189
21 基于结构及分子模拟预测 HLA 结合肽	195
22 基于结构预测 MHC 结合肽操作指南	200
23 MHC-I 及 MHC-II 与肽结合的静态能分析	206

• v •

24	分子动力学模拟.....	214
25	一种预测 MHC-II 类分子结合肽的迭代方法.....	229
26	MHC-II 类分子结合肽综合预测方法	237
27	基于贝叶斯神经网络的 MHC-II 类分子-肽复合物非线性预测模型	244

第 4 部分 免疫系统其他特性的预测

28	TAPPred 法预测抗原中的 TAP 结合肽.....	253
29	B 细胞表位的预测方法.....	257
30	一种 MHC 分子结构功能相似性分析平台: HistoCheck	263
31	免疫相关性毒力因子的预测.....	271

索引.....	278
---------	-----

彩图

1		1
2		2
3		3
4		4
5		5
6		6
7		7
8		8

附录 A 免疫学基础

9		9
10		10
11		11
12		12

附录 B MHC 分子结构

13		13
14		14
15		15
16		16
17		17
18		18
19		19
20		20
21		21
22		22
23		23

1 免疫信息学与计算机辅助预测免疫原性：导论

Darren R. Flower

摘 要 免疫信息学是将信息技术应用于免疫分子的研究的一门学科。其主要目标之一是准确预测表位、亚单元疫苗或减毒病原体等不同层次的免疫原性。免疫原性是指病原体及其组分在初次暴露于免疫系统监视下时诱导特异性免疫应答的能力，而抗原性是指在再次免疫应答中被获得性免疫反应识别的能力。本书介绍了上述内容，并讨论了免疫信息学和免疫原性计算机预测方法的现状。

关键词 抗原提呈，生物信息学，计算化学，计算疫苗学，免疫信息学，MHC 结合，疫苗设计

免疫学很重要，因为传染病主要领域也属于免疫学的主要领域。免疫学就是关于机体如何抵御传染病的科学。从人类疾病角度出发，正确评价获得性免疫和固有免疫是人类对抗传染性疾病的重要基础，这是降低人类死亡率和发病率最重要因素。免疫学的社会重要性毫无疑问，因为它与免疫系统在健康或患病时的生理机能紧密联系。各种分子和细胞结构构成整个机体免疫力的外在表现，相关免疫学知识推动了新的临床及非临床技术的发展。同样，生物科学向自动化和高通量的发展也不可逆转，这对免疫学有着重要影响：在 100 多年实验研究之后，免疫学正在将自身重塑为建立在基因组基础上的定量科学。因此，不管免疫学家们是否承认，经过长期发展后免疫学正面临转折，如同大多数生物科学，免疫学也将不可抗拒地接受高通量技术带来的新的信息。

由于免疫学关注的许多问题都是以人类为中心的，而在脊椎动物中，获得性免疫在系统中处于中心地位，因此免疫学被视为一门非常重要，甚至是最为重要的科学。免疫学在科学界中具有很高的地位——免疫学杂志具有高影响因子，免疫学是一个很大的学科并有充足的经费支持。正因如此，免疫学研究范围广泛、包罗万象。免疫紊乱中的免疫系统问题包括：自身免疫疾病、异体排斥和免疫缺陷以及免疫器官在体内、体外及原位的理化和功能特性。

利用计算机抵御传染性疾病和其他疾病状态对一些人而言过于牵强，但计算机方法早已应用到小分子药物的设计中，从而对人类健康发挥作用。现在我们正开始探讨应用计算机技术对免疫治疗和预防性疫苗设计产生影响的方式。

计算机方法和体外实验的结合产生了免疫信息学这门学科。免疫信息学是计算科学中的一个新分支，意义深远。通过计算科学，人们能极大地推进新型免疫制剂的研究速度和有效性。近年来，免疫信息学开始以生物信息学领域内子学科的面貌出现。因此，免疫信息学也就是将生物信息学方法应用于解决免疫学和疫苗学中特有的问题。作为初期免疫组学技术的主要构成，免疫信息学也正在开始对免疫学本身产生影响。最终，免疫信息学将对后基因组革命产生深刻的影响。到目前为止，实验研究已经对免疫信息学的发展产生束缚，免疫信息学已经开始将自身从实验的桎梏中释放出来，显然，这种高

通量方法将引起研究模式从假说驱动向数据驱动的转变。尽管在初期获得的海量数据将显得十分复杂、扑朔迷离,但通过对这些数据的分析,人们将最终产生新的理解和认识。

为应对大量数据带来的上述压力,人们对有效运用信息工具来分析免疫学数据的兴趣与日俱增,由此产生了两种免疫信息学的方法。第一种是直接基于生物信息学的方法,该方法在技术上与生物学其他技术没有区别,如包括对人类和微生物基因组的注释。另外一种免疫信息学自身所特别关注的,该方法建立在计算机科学基础上,还包括了多学科技术,如生物化学、生物物理学、计算机化学、药物化学、分析化学、结构生物学及蛋白质同源建模,且涉及的相关学科还在逐渐增多。到目前为止,这一令人兴奋不已、快速发展的专门学科的主要任务是准确预测免疫原性,具体表现为表位预测或整个蛋白质免疫原的预测,这正是本书的重点。

免疫系统十分复杂,包括许多层次,各层次都表现出惊人的自主行为。毫无疑问,免疫系统所呈现的复杂性令人困惑不已,尽管许多人不承认,但我们对基础免疫学的疑问仍大量存在。不过,对于免疫学的中心问题——分子识别事件,我们已有了较深的认识:两个或两个以上分子结合,构成一定时间内稳定的复合物;表位与 MHC 蛋白的结合或 T 细胞受体和 MHC-多肽复合物的结合是以理化性质为基础的,这与生命科学其他领域中分子间相互作用没有差异。只是在更高层次,即当成千上万或以百万计数的不同分子相互作用时,免疫系统在时空上才表现出自发特性。免疫原性就是这样一种固有的特性。

为实现免疫原性的预测,免疫信息学提出了免疫原性最终建立在简单、可理解的分子事件基础上的观点。所谓免疫原性就是分子或超分子物质诱导免疫应答能力的特性,此处的分子物质可以是蛋白质、脂肪、碳水化合物或其他组合物,超分子物质可为病毒、细菌或寄生虫。免疫原——具有免疫原性的物质——是可诱导特异免疫应答的物质,而抗原——具有抗原性的物质——是在再次反应中,获得性免疫应答的效应细胞或分子如 T 细胞或抗体可识别的物质。因此抗原性就是抗原被抗体或 T 细胞受体库的一个或若干个部分识别的特性,同时,免疫原性是免疫原暴露于免疫系统初期诱导特异性免疫应答的能力。两种性质相互联系,但要恰当理解二者如何关联却绝非易事。

要预测复杂蛋白质的抗原性和(或)免疫原性仍存疑问。这取决于哪个蛋白质被提呈以及识别复杂蛋白质的免疫库的性质,其中任一因素或同时两种因素都可能十分重要。许多免疫原或抗原的免疫应答都集中在少数优势性免疫结构上,而其他大多数分子可能无法引起免疫应答,从而无法来评估其免疫原性。要评价免疫原性,我们必须同时考虑宿主和致病生物的特性,而不仅是抗原自身的固有特性。免疫库的构成将影响对特定表位的应答并改变其对特定靶的识别。这些组分可能在体内启动应答时通过免疫库与宿主抗原的交叉反应被删除。此外抗体库的局限性也可限制免疫应答,因为编码抗体的抗原结合位点的基因数量是一定的。总之,抗原性和免疫原性都有许多相互联系的因素。免疫反应的诱导需要固有免疫系统和获得性免疫系统间的相互作用,前者以反应快速和非特异为特征,后者可进行特异性表位识别。

为保护宿主抵御传染性疾病,免疫系统必须识别诸多微生物病原体(细菌、真菌、寄生虫和病毒),这主要是通过识别完整或者降解的蛋白质这些生物大分子实现的。但

是,表位也不必是蛋白质,碳水化合物、脂肪、核酸可单独(或与多肽结合)成为表位。作为一种翻译后修饰,无论是否经过体液或细胞免疫调节,它都是构成蛋白质免疫原性最常见的方式。糖基化表位可以和抗体结合、被 I 类或 II 类 MHC 分子提呈、和糖多肽特异性 T 细胞克隆结合,因此,总体上免疫原性可通过获得性免疫应答得到体现:体液的(通过抗体和全抗原结合来调节)和细胞免疫(通过 T 细胞识别蛋白酶解多肽来调节)。

体液免疫由可溶或膜表面抗体介导,这可通过多种方法检测。酶联免疫吸附实验(ELISA)或竞争性抑制结合方法可检测抗体的滴度,即血液中抗体与抗原结合能力达到半数最大值时的浓度;也可直接通过饱和和透析法检测抗体或抗原的亲合力。细胞免疫测定的方法很多,最直接的方法是检测 T 细胞杀伤效应。细胞毒性 T 细胞(CTL)可溶解靶细胞,可采用铬-51 放射性同位素进行测定,铬-51 放射性同位素被靶细胞吸收并在溶破时释放。 $CD4^+$ T 细胞的增殖反应会反过来激活巨噬细胞或 B 细胞,可通过将含氟胸苷整合到 T 细胞 DNA 中的量进行测定;还可检测 T 细胞产生的细胞因子。近来,人们已经将注意力转移到四聚体,将其作为检测 T 细胞应答的工具^[1]。

大多免疫原性由表位决定,表位是免疫系统识别的主要化学物质。因此,B 细胞和 T 细胞表位的准确预测是对免疫信息学最主要的挑战。将表位预测视为免疫信息学研究最前沿的领域以及最伟大的科学挑战恰如其分:免疫信息学尽管存在困难,却更让人激动不已,作为推动疫苗改进和治疗学发展的主要工具,它还具有实用价值。

尽管人们对非多肽表位,如碳水化合物和脂肪作用的关注日益增加,但 B 细胞和 T 细胞表位(通过体液或细胞免疫分别调节)仍然是检测和调节免疫应答复杂性的主要工具,因为 T 细胞、B 细胞和可溶性抗体对表位的识别才是获得性免疫应答的核心。上述初始应答将依次激活细胞和体液免疫系统,并最终杀死病原生物。

“表位”一词在生物学中被广泛应用。从词源上来说,“epitope”的词根是希腊语,和许多其他的词一样,它的意思具有多样性且不断变化,该词最常用于描述生物大分子被另一个生物大分子识别或结合的区域。对于免疫学家来说,它的意思多限于指被免疫系统以特定方式识别的结构,如 B 细胞表位指蛋白质分子中被抗体识别的区域,T 细胞表位指被 MHC 分子结合后被 T 细胞识别的短肽。

大分子负责识别表位的部位称之为抗原互补位。就结合的化学角度来说,我们只需要考虑结合反应中的对应的部分,但把蛋白质和机体功能的联系考虑在内,互补位本身就是一个目的性更强、甚至是拟人化的结构,而其确切内涵尚不清楚。因此从其内涵出发来认识表位和互补位,就在认识论上比现实更深入了。

B 细胞表位是蛋白质或其他大分子中被可溶性抗体或者膜结合抗体识别的区域。B 细胞表位分为线性和非连续性表位。线性表位由蛋白质序列中的一段连续氨基酸组成,非连续性表位则由在序列上分开但在蛋白质结构上相近的氨基酸构成。大多数 B 细胞表位是非连续性的,实验鉴定的表位多为线性表位。线性表位产生的抗体可以与其蛋白质发生交叉反应,第 29 章介绍了 B 细胞表位的预测方法。

T 细胞表位是一段可以与 MHC 和 TCR 结合成三分子复合物的多肽,该复合物是 T 细胞活化的初始,但并不是唯一的分子识别事件。除了 CD4 和 CD8 分子,其他很多共受体分子和辅助分子都参与了 T 细胞的识别。这个识别过程仍有许多细节还没有弄

清楚,但已知该识别过程涉及由多种受体高度组织构成的免疫突触。这些辅助分子虽是参与突触形成的必要成分,但对其参与机制仍未完全阐明。最终弄清表位提呈的过程还需要建立这些复合物形成过程的精确模型。

现在已经有大量的、准确成功的方法预测 T 细胞表位,但是,对 B 细胞表位准确可靠的预测仍然不成熟。这些方法由早期的基序方法发展到定性和以人工神经网络为代表的半定量方法,以及很多的定量方法。大多数 T 细胞表位预测方法是基于肽与 MHC 亲和力的预测。

我们知道,MHC 结合外源或自身蛋白质通过降解或蛋白酶酶解得到多肽。外来表位主要来源于有益的或致病的微生物(如病毒和细菌),而自身表位主要来源于宿主蛋白。蛋白质降解为可以被 MHC 结合的多肽的蛋白质水解途径相当复杂,至今仍有许多问题有待澄清。多肽经过翻译后修饰与 MHC 形成复合物被 TCR 识别的提呈途径同样也很复杂,另外,MHC 不限于和多肽结合,还可以和化学修饰的肽、模拟肽以及药物样分子结合^[2]。

抗原提呈有许多替代途径,但主要与 I 类和 II 类 MHC 分子有关。MHC-I 类分子表达在几乎所有核细胞表面,被 CD8⁺ T 细胞识别;MHC-II 类分子表达于抗原提呈细胞,被 CD4⁺ T 细胞识别。MHC 具有多态性,大部分人群主要包括 6 种经典的 MHC 分子——3 种 MHC-I 类分子(HLA-A、HLA-B 和 HLA-C)和 3 种 MHC-II 分子(HLA-DR、HLA-DP 和 HLA-DQ)。不同的等位基因具有和不同肽结合的特性,研究这一现象最简单的方式就是弄清 MHC 结合肽的序列模式。人群中有大量的差异基因编码的 MHC 分子,每个都表现出不同的肽结合特性。与 MHC-多肽以较低亲和力结合的 TCR 也表现出很大的差异性。MHC 与 TCR 选择的组合是免疫系统的识别基础,也是免疫识别的核心所在,因此成为免疫治疗关注的焦点,这些都将在随后的章节中讨论。

被 I 类和 II 类 MHC 分子提呈的肽的主要差别在于它们的长度。MHC-I 类分子提呈的肽主要来源于内源性蛋白(如病毒),这些蛋白质被蛋白酶或其他辅助酶降解成为不同片段的短肽,其中以 8~11 个氨基酸长度的短肽为主。最近,也发现有长达 15 个氨基酸的短肽可以被 MHC-I 分子提呈给 TCR 识别。MHC-II 类分子主要提呈受体介导的内吞的外源蛋白,这些蛋白质被组织蛋白酶降解为 15~20 个氨基酸长度的多肽,但比这个长度稍短或者稍长的肽片段均可被 MHC-II 类分子提呈^[3]。

除此之外,免疫原性的许多其他特性已经被实验研究,但这些证据还不足以修正预测方法。一般而言,蛋白质越大、化学结构越复杂,与自身蛋白序列差异越大,抗原性就越强。然而,其他一些因素如抗原与内吞途径中的组件的亲和力及免疫原性强弱的问题还未弄清。理解病原体与弄清宿主一样重要,我们已经清楚细菌病原体有很多种方式攻击人体,同时很多病原体也使用同样的毒力因子。关于这方面的内容将在第 31 章讨论。微生物蛋白质的其他性质,如亚细胞定位和表达水平等同样重要。免疫原性既不是由宿主也不是也由病原体单独决定,而是由两者的性质共同决定的。因此,我们需要整合从分子到有机体的各种信息来解决这个问题。

多肽与 MHC 的结合只有达到某一阈值时才表现为 T 细胞表位的这一观点现已普遍接受,在某种程度上,多肽与 MHC 结合的亲和力与 T 细胞应答强度相关,这一问题目前还比较复杂,涉及许多免疫信息学的重要问题,还需要进行系统的研究。人工合

成的增强亲和力的环化肽提高了 T 细胞反应性, 为这个问题提供了令人瞩目的证据。然而, 也许有人会认为结合亲和力不仅在免疫识别中重要, 也是整个免疫反应所必需的, 其重要性被一些人尤其是对免疫信息学有异议的人质疑。然而, 它的实用性是很清楚的, 实验免疫学家和疫苗学家们常常通过免疫信息学方法选择、筛选和修饰候选肽以确定功能性表位。

关于免疫原性的许多问题依然存在: 导致免疫原性的内在分子机制是什么? 对于细胞免疫, 表位免疫原性是由多肽和 MHC 结合的半衰期(解离率)决定, 还是由 MHC-多肽结合亲和力决定? 与抗原提呈细胞表面的 MHC-多肽数量有关, 还是由形成的免疫突触的持续作用决定? 需要强调的是, 免疫原性并不是单由肽与 MHC 分子的结合发挥功能, 而是由机体识别的一系列信号构成, 多肽只是其中之一。肽的结合是必要的, 但并不是唯一的。

一段时间以来, 数据库成了生物信息学的通用语言。虽然生物相关信息的数据库的数据类型可能改变, 但是数据库的使用、构建和操作都反映了生物信息学的同一特征, 免疫信息学的数据库也具有同样的特征, 这些在其对后基因组生物科学的支持可以得到证实。21 世纪, 免疫学进入信息经济时代, 数据库中的数据得以广泛传播。以上所述, 虽然对免疫学数据库没有特别新的东西, 但是免疫信息学的深入发展显示了生物信息学在免疫学中的应用开始扩展和成熟。

例如, IMGT 数据库(将在第 2~4 章描述)已关注免疫重要相关大分子的序列分析很多年了。定位于表位功能或者以记录表位 MHC 分子亲和力为主的数据库相对较新, 但也已初具规模。描述单一病毒性质的 HIV 分子免疫数据库被认为是最易用的数据库^[4], 这个数据库包括了该病毒几乎所有 T 细胞表位和 B 细胞表位的信息。另一个被广泛应用的数据库是 SYFPEITHI(将在第 5 章描述), 这个数据库包括了及时更新的 T 细胞表位信息, 还包括了许多 MHC 结合肽的信息。

MHCPEP 数据库包括了 T 细胞表位和 MHC 结合资料的信息^[5], 根据多种方法将肽分为“结合者”或“非结合者”。结合者根据亲和力大小又分为高结合、中等结合和低结合。最近, Brusic 及其同事开发了一个更复杂和精巧的数据库——FIMM^[6], 这个系统整合了 T 细胞表位及许多 MHC 与多肽相互作用的数据, 还包括了许多 MHC 分子及其等位基因与疾病相关关系的数据。

数据库现在变得更多样化并包括了更丰富的免疫学数据。AntiJen(以前叫 Jen-Pep)是最近开发的包含了许多免疫生物学中动力学、热力学、功能和细胞学数据的数据库^[7~9], 它关注于 T 细胞表位和 B 细胞表位。AntiJen 是第一个包含许多免疫分子相互作用的连续定量数据的功能数据库, 包括肽与转运相关蛋白(TAP)、MHC 以及 MHC-多肽复合物与 TCR 分子相互作用的热力学和动力学数据, 另外还包括诸如共受体和超抗原的其他免疫相关蛋白质间的作用。虽然 AntiJen 的数据结构设置不同于其他免疫数据库, 但仍然存在与其他系统的数据重叠。在第 6~8 章中, 由 Raghava 小组创建的数据库主要包含以下三个方面的内容: 细胞免疫(MHCBN, 第 6 章)、体液免疫(BCIPEP, 第 7 章)、小分子半抗原(HAPTENDB, 第 8 章)。很明显, 这些为了满足免疫信息学发展建立的各有侧重的数据库联合起来时将显示出巨大的能力。

多态性使 MHC 介导的肽特异性的研究变得复杂, 事实上, 由于多种原因, HLA