

文章编号:1673-1689(2008)03-0001-05

碳水化合物网络工具和数据库

金征宇^{*1,2}, 王金鹏^{1,2}, 邓力^{1,2}, 赵建伟^{1,2}, 陈寒青^{1,2}

(1. 食品科学与技术国家重点实验室, 江南大学, 江苏 无锡 214122; 2. 江南大学 食品学院, 江苏 无锡 214122)

摘 要: 生物信息学在生命科学领域发挥着很大的作用, 但是生物信息学的工具不完全适用于糖生物学, 糖生物学的网络工具和数据库应运而生。作者依据糖生物学的生物信息资源和数据库, 搜集了用于解析碳水化合物结构常用的网络工具和数据库, 按分类进行了分析。分析结果表明: 将碳水化合物命名工具及数据库、构型分析工具及数据库、三维结构模型数据库整合起来用于解析碳水化合物, 可以简化分析工作, 并可获得可靠的数据信息。

关键词: 糖数据库; 糖组学; 糖生物学

中图分类号: TS 241

文献标识码: A

Databases and Web Tools on Carbohydrates

JIN Zheng-yu^{*1,2}, WANG Jin-peng^{1,2}, DENG Li^{1,2}, ZHAO Jian-wei^{1,2}, CHEN Han-qing^{1,2}

(1. The State Key Lab of Food Science and Technology, Jiangnan University, Wuxi 214122, China; 2. School of Food Science and Technology, Jiangnan University, Wuxi 214122, China)

Abstract: Bioinformatics has a pivotal role in life sciences. But the tools of bioinformatics aren't perfectly suitable for glycobiology, there appears new chance for the development of glycobiology, of which the information is increasing intensively. Based on resources and databases of Bioinformatics, databases and web tools were collected and analyzed on carbohydrates. Results show that if the nomenclature web tools and databases, conformation databases and 3-D databases were combined together for researching of carbohydrates, the analysis work will be predigested, and more detailed informations can be obtained.

Key words: glycodatabase; glycomics; glycobiology

20 世纪末, 人们认识到碳水化合物与生命活动息息相关, 是与蛋白质和核酸并列的重要的生命物质^[1], 对碳水化合物进行研究成了生命科学领域的研究热点^[2]。借助于强大的网络功能, 在生物信息学的基础上, 国外一些机构开发了有关碳水化合物

的网络工具和数据库。瑞士 Olivier Berteau^[3]、法国 Marchal^[4]、德国 Claus-W^[5] 和 Claus-Wilhelm von der lieth^[6] 等分别从生物信息学方面收集并介绍了这些资源。我国在糖生物学与糖工程方面起步较晚, 与国外同行还有一定差距, 对这些资源的

收稿日期: 2007-09-18.

基金项目: 国家“863”计划项目(2006AA10Z333).

作者简介: 金征宇(1960-), 男, 江苏扬州人, 教授, 工学博士, 博士生导师, 主要从事碳水化合物研究. Email: jinlab2008@yahoo.com

搜集和使用还比较有限,仅周永达等^[7]搜集了糖类的生物信息学资源。碳水化合物结构复杂性决定了解析工作的难度,而利用网络资源为科研服务,将带来很大便利。作者搜集了解析碳水化合物常

用的网络工具和数据库(表 1),分别从糖链命名、构型分析、高级结构模型等方面对这些数据库进行介绍并简要分析。

表 1 碳水化合物的命名、结构分析、三维结构及其他相关数据库的名称、数据库功能简要描述、及数据库网址
Tab. 1 Carbohydrates' nomenclature databases、structure analysis databases、3D databases and URL

名称	描述	网址
碳水化合物链式结构命名数据库		
LINUCS	糖链编码	http://www.glycosciences.de/tools/linucs/
LiGraph	图谱解析	http://www.glycosciences.de/tools/LiGraph/
IUPAC	命名	http://www.chem.qmw.ac.uk/iupac/2carb/
CabosML	碳水化合物序列的 markup 语言	http://phoenix.hydra.mki.co.jp/CabosDemo.html
GLYDE	聚糖资料交流平台	http://lsdis.cs.uga.edu/projects/glycomics/index.php?page=4
CFG	聚糖的符号命名	http://grtc.ucsd.edu/symbol.html
MonoSaccharide DB	单糖的系统性命名	http://www.dkfz.de/spec/monosaccharide-db/
ConsortiumFunctional Glycomics	聚糖命名法	http://glycomics.scripps.edu/CFGnomenclature.pdf
碳水化合物结构分析数据库		
Glycofragment	由聚糖碎片得到质谱	http://www.glycosciences.de/tools/GlycoFragments/
GlycosidIQ	糖碎片质量指纹图	http://tmat.proteomesystems.com/glycosuite/
GlycoSearchMS	质谱对照	http://www.glycosciences.de/sweetdb/start.php?action=form_ms_search/
GlycoMod	由分子峰得到聚糖结构	http://www.expasy.org/tools/glycomod/
GlycoMass	由混合物得到质谱	http://www.expasy.org/tools/glycomod/glycanmass.html
Glypeps	糖肽检测	http://dkfz-heidelberg.de/spec/glypeps/
CASPER	¹ H, ¹³ C 核磁共振预测	http://www.casper.organ.su.se/casper/
SugarBase	¹ H, ¹³ C 核磁共振检索	http://boc.chem.uu.nl/sugabase/sugabase.html
NMR—Search	¹ H, ¹³ C 核磁共振检索	http://www.glycosciences.de/sweetdb/start.php?action=form_shift_estimation/
CCRC	木聚糖核磁共振数据库	http://www.crc.uga.edu/specdb/ms/pmaa/pframe.html
Elution Coord. DB	高效液相色谱	http://www.gak.co.jp/ECD/Hpg_eng/hpg_eng.htm
GlycoBase	高效液相色谱	http://www.bioch.ox.ac.uk/glycob/glycoimmunology/glycobase.htm
碳水化合物的三维结构数据库		
SWEET-II	生成三维结构	http://www.glycosciences.de/modeling/sweet2/
GLYCAM	生成三维结构	http://glycam.crc.uga.edu/CCRC/
Monosaccharide	构象图	http://www.cermav.cnrs.fr/cgi-bin/monos/monos.cgi
Disaccharides	构象图	http://www.cermav.cnrs.fr/cgi-bin/di/di.cgi
GlycoMaps DB	构象图	http://www.glycosciences.de/modeling/glycomapsdb/
GlyVicinity	碳水化合物-蛋白质交作用的分析	http://www.glycosciences.de/tools/glyvicinity/

续表 1

名称	描述	网址
GlyTosrion	PDB 建立用于分析碳水化合物键角	http://www.glycosciences.de/tools/glytorsion/
CARP	碳水化合物的 Ramachandran 图	http://www.glycosciences.de/tools/carp/
pdb-care	在蛋白质数据库中的碳水化合物残基	http://www.glycosciences.de/tools/pdbcare/
DynamicMolecules	聚糖分子动力学	http://www.md-simulations.de/manager/
Glydict	N-连接聚糖结构	http://www.dkfz-heidelberg.de/spec/glydict
Xtaloligosaccharides	寡糖的晶体结构	http://www.cermav.cnrs.fr/cgi-bin/oligos/oligos.cgi
其他有关碳水化合物的数据库		
CCSD	复杂碳水化合物研究中心研制的复杂碳水化合物数据库	http://www.boc.chem.uu.nl/sugabase/carbbank.html
GLYCOSCIENCES.de	德国癌症研究中心研制的有关糖的综合数据库	http://www.glycosciences.de/
KEGG-Glycan	基因和基因组学京都百科全书研制的有关聚糖的数据库	http://www.genome.jp/ligand/kcam/
GlycoSuite	糖蛋白系统	http://tmat.proteomesystems.com/glycosuite/
Glycomics database	斯德哥尔摩大学汇编的糖组学信息	http://www.glycomics.com
CarbBank	雅典复杂碳水化合物研究中心研制的搜索引擎	http://www.boc.chem.uu.nl/sugabase/carbbank.html
SweetDB	解析 NMR 图谱的数据库	http://www.dkfz-heidelberg.de/spec2/sweetdb/
Kcam	通过结构搜索数据库	http://www.genome.jp/ligand/kcam/
GlycoMindsa	糖组学	http://www.glycominds.com
SPECARB	糖的拉曼图	http://www.models.kvl.dk/user/engelsen/specarb/specarb.html

1 糖链命名工具及数据库

碳水化合物有关数据库没有普遍使用的一个重要的原因,就是各数据库输出的数据格式不兼容。目前国际上通用的系统命名法是国际纯粹与应用化学联合会(IUPAC, International Union of Pure and Applied Chemistry)和国际生物化学与分子生物学联盟(IUBMB, International Union of Biochemistry and Molecular Biology)生物化学命名联合委员会(JCBN, IUPAC-IUBMB Joint Commission on Biochemical Nomenclature)于1996年修订并出版的碳水化合物命名方法(1996年方案)^[8],该命名法将糖链命名为3种形式:扩展式(extended form),精简式(condensed form)和缩略式(short form)。但是这些命名形式都不能直接用于计算机程序。对于结构异常复杂的碳水化合物

来说,人工的方法对其命名是一件很繁琐的工作,并且很容易产生错误,其他命名法因此相继发展。德国癌症研究中心(DKFZ)于2001年推出了网络命名工具LINUCS(Linear Notation for Unique Description of Carbohydrate Sequences)^[9],LINUCS输出的程序可以用于碳水化合物的其它数据库,这样就可使碳水化合物相关数据库整合起来解析复杂的碳水化合物,从而能够节省大量的人力和时间。完成LINUCS命名,需要一个网络工具Ligraph,通过Ligraph可以输出机读扩展式结构,进而生成LINUCS线性命名,从而用于其它程序。

2 碳水化合物的构型分析工具及数据库

解析碳水化合物构型常用质谱(MS)和核磁共振(NMR),但是图谱的解析只能靠经验来进行,这

为碳水化合物工作者带来了很大的不便。目前,已经有很多已知结构碳水化合物的 MS 和 NMR 数据都被储存在公共数据库中,研究者只需要在相应的数据库中查询相关信息资源,对照实验结果即可得出正确结论,这样就简化了实验所得图谱的解析工作。目前国外常用 SugaBase、CASPER、SweetDB 等数据库解析 NMR 图谱。SugaBase 不断扩展,现已包含 ^{13}C -NMR 数据及其它糖类化合物信息。它们的内容也已被合并到新近开发的 SweetDB 中。SweetDB 能根据实验所得化学位移值找到相应的 NMR 图谱,并且能将这些图谱对应的扩展式给出。CASPER 可以模拟出已知结构碳水化合物的 NMR 化学位移值,也可直接进 SweetDB,从而获得模拟化学位移对应的 ^{13}C -NMR、 ^1H -NMR 图谱及结构式。图谱的模拟条件是 $70\text{ }^\circ\text{C}$ 、 D_2O 作为溶剂,如果实验条件不相符,模拟时应减去系统误差,纠正化学位移。有时由于输入结构残缺会造成模拟出来的系统误差比较大,这时模拟出的图谱只能看做是近似图。

常用解析 MS 的网络工具主要有 Glycofragment、Glyco Mod、Glycan Mass、GlycoSearch MS 等,将这些工具联合起来使用,就能够准确解析碳水化合物的 MS 图。用 GlycanMass 可计算出低聚糖碎片质量,Glyco Mod 可根据低聚糖质量碎片来预测其结构,Glyco Search MS 能够根据质量碎片找到对应碳水化合物的质谱图和扩展式结构。Glycofragment 可以根据碳水化合物的扩展式输出所有离子碎片质量,并且可以对碳水化合物质谱图的所有峰进行解释。

3 碳水化合物的三维结构模型数据库

由于碳水化合物的结构复杂性,其相应的分离纯化技术存在局限,对碳水化合物的三级结构研究得比较少,但是三级结构是碳水化合物的活性状态,因此对三级结构进行描述很有必要。对于单糖来说,由于单糖的结构比较简单,Jon Maher 建立了一个专门用于检索单糖结构的数据库 Monosaccharide。在这个数据库里,可以通过单糖的名称检索到该糖的 Fischer 投影式、三维图和 Haworth 结构。双糖也有专用的数据库:Disaccharides。从这个数据库,可以得到常用双糖最低能量状态下结构的 Ramachandran 图,并且能够显示该状态下的扭角。此外,SWEETII 是快速生成糖的三维结构的一个数据库,它能将聚合度 10 以内的低聚糖的三维模型瞬间生成,并且能使该三维模型在空间旋转,给

科研工作者很直观的结构视图(见图 1)。SWEET-II 输出的数据可以被 ChemOffice 软件识别,为进一步对多糖进行分子水平上的分析奠定了基础。

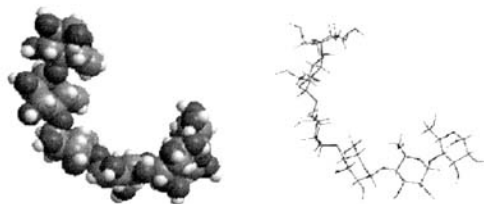


图 1 六个 $\alpha\text{-D}$ 葡萄糖以 (1-4) 糖苷键相连组成的淀粉,在 SWEETII 数据库中的三维模型及分子结构视图

Fig. 1 A 3D figure of starch composed of ($\alpha\text{-D-glcp}$, 1-4), from SWEETII database

4 有关碳水化合物的其他数据库

1985 年,格鲁吉亚大学(UGA)建立了复合碳水化合物研究中心(CCRC),它是一个基于世界范围网络资源的复合碳水化合物结构识别机构,科研工作者只要提供一张碳水化合物的图谱,就可以从 CCRC 获得它的结构信息。CCRC 提供链接,可以链接到 CCSD(复合碳水化合物结构数据库)和 CarbBank。目前,CCSD 已含有大约 50 000 条记录,其中包括结构、来源、所用分析方法及参考文献,现在该数据库作为 SweetDB 的一部分,也可以从荷兰 Utrecht 大学的 Bijvoet 研究中心的主页上进行检索。CarbBank^[10]是允许进入 CCSD 数据库的搜索引擎。GlycoSuite 数据库(GlycoSuite DB)是有关糖结构的数据库^[11],包括 1990—2003 年以来文献上发表的大多数 O-连接的糖和 N-连接的糖,结构信息包括糖的类型、糖链、构型、质量及组成等。糖相关数据库随着分析方法和计算方法的发展而迅速发展,需要一种能被广泛接受的计算机语言搜集并分析这些巨大的数据信息,一个基于 XML(Extensible Markup Language)的网络工具 GLYDE(聚糖数据交换)应运而生^[12]。GLYDE 为每个数据库设计了标识符,只有将正确的 ID 输入后该数据库才能够被识别,包含标识符的主要有:Carb-Bank、KEGG、LINUCS、Glyco。为了促进 GLYDE 的发展,他们又制作了 XML 数据式定义(DTD)和 XML 图解网络,DTD 上有很多注解和说明,帮助使用者理解 GLYDE 的结构。最近 GLYDE 正和 EuroCarbDB 合作建立一个碳水化合物残基的命名标准,并致力于对聚糖结构进行描述。

5 网络资源的整合

随着糖组学的发展,碳水化合物数据库和网络工具不断发展完善,将这些数据库整合起来,用于解析碳水化合物,将会为科研工作带来很大的便利,这些数据库及网络工具之间的关系如图2所示。

6 结 语

目前糖生物信息学的发展仍处于初期阶段,基于Internet的资源也在不断的更新完善,将这些资源整合起来用于解析碳水化合物,能够简化分析工作。

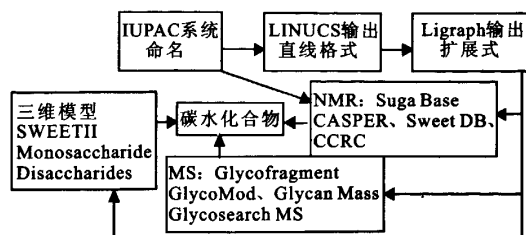


图2 解析碳水化合物从系统命名出发,各数据库相互结合,最终准确得到碳水化合物的结构

Fig. 2 Begin with system nomenclature of carbohydrates, carbohydrates' structure was obtained exactly through all databases combined

参考文献(References):

- [1] 王克夷. 糖类研究的进展和前沿[J]. 生命的化学, 1994, 14(5): 4-5.
WANG Ke-yi. Progress and frontier for carbohydrate[J]. **Chemistry of Life**, 1994, 14(5): 4-5. (in Chinese)
- [2] 金城, 张树政. 糖生物学与糖工程的兴起与前景[J]. 生物工程进展, 1995, 15(3): 14-17.
JIN Cheng, ZHANG Shu-zheng. Advancement and future in glycobiology and glycotecnology[J]. **China Biotechnology**, 1995, 15(3): 14-17. (in Chinese)
- [3] Olivier Berteau, Roland Stenutz. Web resources for the carbohydrate chemis[J]. **Carbohydrate Research**, 2004, 339: 929-936.
- [4] Marchal I, Golfier G, Dugas O, et al. Bioinformatics in glycobiology[J]. **Agricultural and Food Chemistry**, 2000, 48: 5455-5458.
- [5] Claus-W von der Lieth, Thomas L teke 1, Martin Frank, et al. The role of informatics in glycobiology research with special emphasis on automatic interpretation of MS spectra [J]. **Biophysica et Biophysia Acta**, 2006, 1760: 568-577.
- [6] Claus-Wilhelm von der lieth, Andreas Bohné-Lang, Karl Lohmann, et al. Bioinformatics for glycomics: Status, methods, requirements and perspectives[J]. **Bioinformatics for Glycomics**, 2004(2): 166-178.
- [7] 周永达, 时春娟, 张剑波, 等. 糖类的生物信息学资源[J]. 中国生物化学与分子生物学报, 2006, 22(3): 184-189.
ZHAOU Yong-da, SHI Chun-juan, ZHANG Jian-bo, et al. Bioinformatics on carbohydrates[J]. **Chinese Journal of Biochemistry and Molecular Biology**, 2006, 22(3): 184-189. (in Chinese)
- [8] McNaught A D. Nomenclature of carbohydrates[J]. **Carbohydrates Research**, 1997, 297: 1-90.
- [9] Andreas Bohné-Lang, Elke Lang, Thomas Forster, et al. LINUCS: Linear notation for unique description of carbohydrate sequences[J]. **Carbohydrate Research**, 2001, 336: 1-11.
- [10] Faramarz Valafar, Homayoun Valafar. CCRC-Net: an internet-based spectral database for complex carbohydrates using artificial neural networks search engines[J]. **Trends in Analytical Chemistry**, 1999, 18(8): 508-512.
- [11] Catherine A, Mathew J, Marc R, et al. GlycoSuite DB: a new curated relational database of glycoprotein glycan structures and their biological sources[J]. **Nucleic Acids Research**, 2001, 29(1): 332-335.
- [12] Satya S, Christopher T, Amit S, et al. GLYDE-an expressive XML standard for the representation of glycan structure[J]. **Carbohydrate Research**, 2005, 340: 2802-2807.

(责任编辑: 朱明)