

## 信息服务

## RNA 研究的网络资源\*

陈莉 金由辛<sup>1)</sup> 王德宝

(中国科学院上海生物化学研究所, 国家分子生物学重点实验室, 上海 200031)

**摘要** 随着 RNA 功能多样性的发现, RNA 研究领域日益受到重视, 同时 RNA 序列、二级结构和三维模建数据已有较大增长. 利用这些数据可以有效地搜索结构信息和进行功能提示. 与此同时, 计算科学的优越性以及全球网络的发展促进了 RNA 相关数据库和网络站点的建设, 为 RNA 的研究带来了极大的方便. 介绍了 RNA 序列和结构数据库、软件资源以及一些常用的在线分析工具, 文中所列网络资源均可在 <http://dna.sibc.ac.cn/bio/rna.html> 上浏览.

**关键词** 网络浏览器, 数据库, RNA, 二级结构

**学科分类号** Q522

90 年代是“RNA 的十年”, RNA 领域内的研究进展突飞猛进. 研究表明 RNA 并非简单地传递遗传信息, 还可以通过各种类型的 RNA 剪接, RNA 编辑和 RNA 再编码来控制遗传信息的表达. 它通过调控和催化蛋白质的生物合成而控制着生命中枢. 已发现的 RNA 种类及功能正在不断增加, 它们具有全部 DNA 和蛋白质的生物功能种类: 携带遗传信息 (病毒 RNA)、催化活力 (ribozyme)、调控功能 (反义 RNA、ppGpp、oxy S RNA、某些抑癌基因、调控 X 染色体活性的 Xist)、运动功能 (pRNA)、信使功能 (cAMP). 此外, snRNA 参与 mRNA 剪接、snoRNA 参与 rRNA 成熟加工、gRNA 参与 RNA 编辑、SRP-RNA 参与蛋白质的分泌、端粒 RNA 参与 DNA 端粒合成并影响细胞的生命、tmRNA 参与破损 mRNA 蛋白质合成的终止等等. 尚有很多 RNA 的功能还未鉴定<sup>[1]</sup>. 迄今为止有关 RNA 的研究已获得 7 次诺贝尔奖. RNA 功能的重要性正日益显现.

RNA 研究已积累了大量 RNA 序列、二级结构和三维模建数据. 利用这些数据可以有效地搜索结构信息和进行功能提示. 与此同时, 计算科学的优越性以及全球网络的发展促进了 RNA 相关软件、数据库和网络站点的建设, 为 RNA 的研究带来了极大的方便.

大型的 RNA 网络站点如: NCBI Entrez (<http://www.ncbi.nlm.nih.gov/Entrez/nucleotide.html>) 和 Rutgers 核酸数据库 (<http://ndbserver.rutgers.edu/>).

除此之外还有许多较有特色的站点.

RNA 的网络资源分为两类: 一类是序列和结构数据库, 另一类是在线分析工具和软件资源. 下面分别列表介绍, 并对其中较好的站点给出简要说明.

## 1 序列及结构数据库

RNA 种类繁多, 功能极具多样性. 有的 RNA 在进化过程中相当保守, 如 rRNA, 通过序列联配等方法可以获得其中的进化关系和结构信息<sup>[2]</sup>. 某些 RNA 的结构 motif 比较保守, 通过搜索基因组可以发现许多新的功能 RNA, 如 snoRNA<sup>[3]</sup>. 通过 RNA 序列和结构数据库<sup>[4]</sup> (表 1) 可以获得较丰富的 RNA 结构元件, 为深入研究 RNA 的结构功能以及与蛋白质的相互作用提供丰富的资源.

### 1.1 Fournier 酵母 snoRNA

这是个很全面的数据库, 除了指导 2'-O-甲基化的 snoRNA 外, 容括了酵母中其他所有类型 snoRNA 的信息.

### 1.2 指导 2'-O-甲基化的 snoRNA 数据库

snoRNA 指导 2'-O-核糖的甲基化. Eddy 等<sup>[3]</sup>实验室开发了隐马可夫模型 (Hidden Markov Model, HMM) 算法, 搜索基因组中的 snoRNA 基因序列. 该数据库提供了他们在酵母基因组中预测

\* 中国科学院重大项目资助 (KJ951-B1-610).

<sup>1)</sup> 通讯联系人.

Tel: (021) 64374430-222, E-mail: yxjin@sunm.shnc.ac.cn

收稿日期: 1999-11-04, 修回日期: 2000-02-21

的结果以及实验方法, 包括引物延伸法检测 rRNA 中 2'-O-甲基化位点, 酵母 rRNA 图谱定位引物和

snoRNA 搜索程序 (含源代码). 数据库中还包括以往鉴定的 snoRNA 信息.

表 1 RNA 序列及结构数据库

| 序列及结构数据库                              | WWW 地址                                                                                                                                                                                                                                     |
|---------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| * Fournier 酵母 snoRNA 数据库              | <a href="http://www.bio.umass.edu/biochem/rna-sequence/Yeast_snoRNA_Database/snoRNA_DataBase.html">http://www.bio.umass.edu/biochem/rna-sequence/Yeast_snoRNA_Database/snoRNA_DataBase.html</a>                                            |
| * 指导 2'-O-甲基化的 snoRNA 数据库             | <a href="http://rna.wustl.edu/snoRNAdb/">http://rna.wustl.edu/snoRNAdb/</a>                                                                                                                                                                |
| * 核糖核酸酶 P 数据库                         | <a href="http://www.mbio.ncsu.edu/RNaseP/home.html">http://www.mbio.ncsu.edu/RNaseP/home.html</a>                                                                                                                                          |
| * 核糖核酸酶 P 3D 模型                       | <a href="http://jwbrown.mbio.ncsu.edu/RNaseP/RNA/threeD/threeD.html">http://jwbrown.mbio.ncsu.edu/RNaseP/RNA/threeD/threeD.html</a>                                                                                                        |
| * 核糖体 RNA 数据库                         | <a href="http://rna.uia.ac.be/lsu/">http://rna.uia.ac.be/lsu/</a><br><a href="http://rna.uia.ac.be/ssu/">http://rna.uia.ac.be/ssu/</a>                                                                                                     |
| * 核糖体数据库计划                            | <a href="http://www.cme.msu.edu/RDP/html/index.html">http://www.cme.msu.edu/RDP/html/index.html</a>                                                                                                                                        |
| * 类 mRNA 非编码 RNA 数据库                  | <a href="http://www.man.poznan.pl/SSData/ncRNA/index.html">http://www.man.poznan.pl/SSData/ncRNA/index.html</a>                                                                                                                            |
| * tmRNA 数据库                           | <a href="http://sunflower.bio.indiana.edu/~kwilam/tmRNA/home.html">http://sunflower.bio.indiana.edu/~kwilam/tmRNA/home.html</a><br><a href="http://psyche.uthct.edu/dbs/tmRDB/tmRDB.html">http://psyche.uthct.edu/dbs/tmRDB/tmRDB.html</a> |
| tRNA 序列及 tRNA 基因序列                    | <a href="http://www.unibayreuth.de/departments/biochemie/trna">http://www.unibayreuth.de/departments/biochemie/trna</a>                                                                                                                    |
| 高等植物线粒体 tRNA 及 tRNA 基因                | <a href="http://www.ebi.ac.uk/service">http://www.ebi.ac.uk/service</a>                                                                                                                                                                    |
| RNA 编辑站点                              | <a href="http://www.lifesci.ucla.edu/RNA/index.html">http://www.lifesci.ucla.edu/RNA/index.html</a>                                                                                                                                        |
| RNA 编辑站点(U 插入/删除)                     | <a href="http://www.lifesci.ucla.edu/RNA/trypanosome/index.html">http://www.lifesci.ucla.edu/RNA/trypanosome/index.html</a>                                                                                                                |
| RNA 修饰数据库                             | <a href="http://medlib.med.utah.edu/RNAmods">http://medlib.med.utah.edu/RNAmods</a>                                                                                                                                                        |
| 5S 核糖体 RNA 数据库                        | <a href="http://rose.man.poznan.pl/SSData/5SRNA.html">http://rose.man.poznan.pl/SSData/5SRNA.html</a>                                                                                                                                      |
| 16S rRNA, 23S rRNA 和 Group I 内含子的二级结构 | <a href="http://pundit.icmb.utexas.edu">http://pundit.icmb.utexas.edu</a>                                                                                                                                                                  |
| 30S 核糖体亚基和 16S rRNA 3D 模型             | <a href="http://www-smi.stanford.edu/projects/helix/ribo3dmodels/index.html">http://www-smi.stanford.edu/projects/helix/ribo3dmodels/index.html</a>                                                                                        |
| 核糖体 RNA 突变数据库                         | <a href="http://www.fandm.edu/Departments/Biology/Databases/RNA.html">http://www.fandm.edu/Departments/Biology/Databases/RNA.html</a>                                                                                                      |
| gRNA 数据库(3.0)                         | <a href="http://www.biochem.mpg.de/~goeringe/">http://www.biochem.mpg.de/~goeringe/</a>                                                                                                                                                    |
| 小 RNA 数据库                             | <a href="http://mber.bcm.tmc.edu/smallRNA/smallrna.html">http://mber.bcm.tmc.edu/smallRNA/smallrna.html</a>                                                                                                                                |
| 信号识别颗粒数据库                             | <a href="http://psyche.uthct.edu/dbs/SRPDB/SRPDB.html">http://psyche.uthct.edu/dbs/SRPDB/SRPDB.html</a>                                                                                                                                    |
| uRNA 数据库                              | <a href="http://psyche.uthct.edu/dbs/uRNADB/uRNADB.html">http://psyche.uthct.edu/dbs/uRNADB/uRNADB.html</a>                                                                                                                                |
| 类病毒和类病毒样 RNA 序列库                      | <a href="http://www.callisto.si.usherb.ca/~jpperra">http://www.callisto.si.usherb.ca/~jpperra</a>                                                                                                                                          |
| 氨酰 tRNA 合成酶数据库                        | <a href="http://rose.man.poznan.pl/aars/index.html">http://rose.man.poznan.pl/aars/index.html</a>                                                                                                                                          |

\* 在正文对该数据库进行了简要说明.

### 1.3 核糖核酸酶 P 数据库

核糖核酸酶 P 是核糖核蛋白, 在细菌及一些古核生物中 RNA 本身具有体外催化活性. 该数据库收集了各种 RNase P 的序列 (包括细菌, 古核生物, 真核生物核细胞器等来源)、序列联配、二级结构、三维建模和一些附加信息等等. RNase P RNA 及蛋白质的二级结构图以 GIF 图象格式显示, 也可以下载 Postscript (.ps) 或 Acrobat (.pdf) 格式文件. 三维结构模型为动画 GIF 或 JPEG 格式. 数据库新添加了序列联配编辑和分析软件 BIOEDIT, 所有文件都可以免费下载.

### 1.4 核糖核酸酶 P 三维模型

该数据库收集了 Chen<sup>[5]</sup> 及 Westhof<sup>[6]</sup> 等建模的 RNase P 3D 全原子模型, 并新增了枯草杆菌 RNase P 蛋白亚基的晶体结构<sup>[7]</sup>.

### 1.5 rRNA 数据库

位于比利时 Antwerp 的核糖体大亚基 (LSU rRNA) 数据库将 607 条大亚基序列进行联配, 提

供了较全面的 LSU rRNA 结构信息, 每条序列均标明二级结构元件. 序列联配定期更新. 服务器还提供软件进行 RNA 序列联配和结构预测. Antwerp 核糖体小亚基 (SSU rRNA) 数据库收集了来自 3 166 种真核生物、7 336 种细菌、324 种古核生物、120 种质体和 601 种线粒体的 SSU rRNA 序列. 所有的 SSU rRNA 序列均以序列联配的形式贮存, 并含有预测的二级结构模式.

### 1.6 核糖体数据库计划

该数据库由美国自然科学基金和能源部支持, 为研究小组提供有关核糖体的数据分析服务和相应计算机程序, 包括 rRNA 序列在线数据分析、衍生系统发生树、rRNA 序列联配、序列注释 (annotation)、rRNA 二级结构图以及各种联配分析和显示软件.

### 1.7 类 mRNA 的非编码 RNA 数据库

不同的细胞均合成多种 RNA 转录物, 它们缺乏编码蛋白质的能力, 很多功能和作用机制还所知

甚少。该数据库首次将所有已知的 RNA 转录物收集制作成数据库以利于进一步的结构功能研究。

## 8 tmRNA 数据库

该数据库现有 46 种 tmRNA (分别来自 10 个细菌种系及叶绿体), 是 tmRNA 结构与功能研究的良好资源。除文献综述和搜索新 tmRNA 序列的指南外, 还提供了序列联配以及预测的标记肽 C 端序列。站点 <http://psyche.uthct.edu/dbs/tmRDB/tmRDB.html> 的特点是提供了小部分的 RNA 二级结构图以及 PDB 格式的三维模建。

## 2 RNA 分析工具和软件资源

从趋势来看, 下个十年将在 RNA 的结构方面取得重要进展。尽管目前许多先进的技术, 如 NMR, X 射线晶体衍射等在 RNA 研究领域已普遍使用, 但所获得的 RNA 三维数据仍然很少, 绝大多数是低分辨率数据<sup>[8]</sup>。原因主要是 RNA 分子存

在很明显的构象多态性, 并且常常与一些小分子或大分子配体成复合物, 因此, 即使一些化学纯度很高的 RNA 也极难获得结晶<sup>[1]</sup>。而借助于计算生物学则可以很好地利用已知的 RNA 序列进行二级结构预测乃至三维模建。目前较为成熟并实现自动化的软件主要在二级结构预测水平。三维模建还需综合一些低分辨率的晶体数据和其他生物化学、生物物理方法探测的信息。众所周知, RNA 的功能由它的二级结构和三级结构来决定。要理解 RNA 的功能仅仅有二级结构的信息是不够的, 但是二级结构预测无疑是迈向远为复杂的三维模建的第一步。运用预测的二级结构和一些附加的三级结构约束条件可以进行 RNA 的全原子结构模建<sup>[9]</sup>, 从而为 RNA 的结构功能研究的方案设计提供快捷途径。表 2 列出了一些 RNA 在线分析工具和提供免费下载的软件资源<sup>[10]</sup>, 可以进行在线分析或免费下载相应软件。

表 2 RNA 分析工具和软件资源

| RNA 分析工具和软件资源               | WWW 网址                                                                                                                      |
|-----------------------------|-----------------------------------------------------------------------------------------------------------------------------|
| * Vienna RNA 软件包 (Ver. 1.3) | <a href="http://www.tbi.univie.ac.at/~ivo/RNA/">http://www.tbi.univie.ac.at/~ivo/RNA/</a>                                   |
| * MFOLD (Ver. 3.0)          | <a href="http://www.ibc.wustl.edu/~zucker/rna/">http://www.ibc.wustl.edu/~zucker/rna/</a>                                   |
| RNAstructure                | <a href="http://128.151.176.70/RNAstructure.html">http://128.151.176.70/RNAstructure.html</a>                               |
| * RNAdraw                   | <a href="http://rnadraw.base8.se">http://rnadraw.base8.se</a>                                                               |
| RNA 二级结构预测服务器               | <a href="http://www.genebee.msu.su/services/rna2_reduced.html">http://www.genebee.msu.su/services/rna2_reduced.html</a>     |
| RNA 二级结构分析                  | <a href="http://www-bia.inra.fr/T/essa/Doc/essa_home.html">http://www-bia.inra.fr/T/essa/Doc/essa_home.html</a>             |
| RnaViz                      | <a href="http://rna.uia.ac.be/rnaviz/">http://rna.uia.ac.be/rnaviz/</a>                                                     |
| * RNA 世界                    | <a href="http://www.imb-jena.de/RNA.html">http://www.imb-jena.de/RNA.html</a>                                               |
| tRNAscan-SE 1.1             | <a href="http://www.genetics.wustl.edu/eddy/tRNAscan-SE/">http://www.genetics.wustl.edu/eddy/tRNAscan-SE/</a>               |
| tRNAeditor                  | <a href="http://www.uni-bayreuth.de/departments/biochemie/trna/">http://www.uni-bayreuth.de/departments/biochemie/trna/</a> |

\* 在正文对该数据库进行了简要说明。

### 2.1 Vienna RNA 软件包

该软件包综合了两种算法来预测 RNA 二级结构: 一种是 Zuker<sup>[9]</sup>关于最小自由能的动态规划算法, 另一种是 McCaskill 的分割函数 (partition function) 算法。除了 RNA 折叠 (RNAfold) 外, 还可以计算给定二级结构的 RNA 能量 (RNAeval), RNA 比热 (RNAheat), 以及采用字符串联配或树-编辑计算二级结构间的距离 (非相似性) (RNAdistance, RNAdpdist), 该软件包还为反折叠提供了一种算法, 搜索给定二级结构的 RNA 序列 (RNAinverse)。Vienna RNA 软件包可以免费下载 C 源代码, 在任何 Unix 环境下均可编译运行。该主页还列出其他 RNA 站点以及生物数据库的超连接。

### 2.2 MFOLD

Zuker 的主页含有众多 RNA 结构站点的超连

接, 可以作为 RNA 相关网站的导航站点。内容主要集中在 RNA 二级结构预测理论和软件上。其他部分涵盖了 RNA 折叠的热力学、RNA 相关数据库、三维模建等等, 而且详细列出许多书籍、会议、学会和杂志的超连接。该站点可以下载最新的 mfold version 3.0, 也可以将序列提交给 Zuker 的 mfold 服务器来完成。其中 RNAstructure 是 Zuker 预测 RNA 二级结构的 Windows9x/Windows NT 版本, 也可以免费下载。

### 2.3 RNAdraw

RNAdraw 的主页非常详尽地介绍了程序的安装、原理、使用和前景。大多数的 RNA 二级结构预测均在大型计算机上实现, 如运行在 SUN、SGI 和 IBM 的 Unix 系统, 一般实验室不具备这些条件。Ole Matzura 成功地解决了这个问题。他采用 Vienna RNA 软件包内的 RNA 折叠程序, 制作了

友好的图形化界面, 可以在 Intel x86 兼容 PC 机上进行 RNA 优化结构、碱基配对-概率矩阵和热曲线计算。

## 2.4 RNA 世界

“RNA 世界”可能是最全面的 RNA 站点, 其中列出的超连接包括各种数据库站点、网络工具、序列、二级结构以及相关软件。这个站点可以很方便地根据 PDB (protein data bank) 代码或者 NDB (nucleic acids database) 代码来查找所需要 RNA 的结构信息, 同时提供包括研究方法、参考文献、可视化图象软件及相关数据库等信息。其中结构图象给出了便于操作的详细显示模式。

信息在当今的 RNA 科学中起着非常重要的作用。由于极其丰富的 RNA 序列所包含的结构类型是有限的, 因此不必将所有 RNA 结构都测定出来。而生物信息学可以为二级结构预测和三维建模提供一个框架, 揭示出更多的结构信息, 从而进一步指导设计实验。我们相信, 随着 RNA 结构数据的积累, 有朝一日, 我们也会象研究蛋白质一样, 能够根据 RNA 的一级序列在数据库中搜索 RNA 结构 motif, 以及进行精确的三维结构建模, 从而实现序列-结构-功能的自动化信息提示。充分利用这些网络资源不仅为 RNA 研究和 RNA 应用提供有益的帮助, 显著地提高研究效率, 而且会使研究 RNA 的模式发生深刻的改变。

## 参 考 文 献

- 1 金由辛. 面向 21 世纪的 RNA 研究. 见: 张 焘主编. 科学前沿与未来. 北京: 科学出版社, 1999. 86~ 101  
Jin Y X. RNA Research Challenging to the 21th Century. In: Zhang T ed. Frontier and Future of Science. Beijing: Science Press, 1999. 86~ 101
- 2 Olsen G J, Woese C R. Ribosomal RNA: a key to phylogeny. FASEB J, 1993, 7 (1): 113~ 123
- 3 Lowe T M, Eddy S R. A computational screen for methylation guide snoRNAs in Yeast. Science, 1999, 283 (5405): 1168~ 1171
- 4 Biological Database issue. Nucleic Acids Res, 1999, 27 (1): 156~ 183; 186~ 187; 192~ 197; 314; 332~ 335
- 5 Chen J L, Nolan J M, Harris M E, *et al.* Comparative photocross linking analysis of the tertiary structures of Escherichia coli and Bacillus subtilis RNase P RNAs. EMBO J, 1998, 17 (5): 1515~ 1525
- 6 Massire C, Jaeger L, Westhof E. Derivation of the three-dimensional architecture of bacterial ribonuclease P RNAs from comparative sequence analysis. J Mol Biol, 1998, 279 (4): 773~ 793
- 7 Stams T, Niranjanakumari S, Fierke C A, *et al.* Ribonuclease P protein structure: evolutionary origins in the translational apparatus. Science, 1998, 280 (5364): 752~ 755
- 8 Holbrook S R. Crystallographic analysis of RNA structure. In: Simons R W, Grunberg M M, eds. RNA Structure and Function. New York: Cold Spring Harbor Laboratory Press, 1998. 147~ 175
- 9 Zuker M. Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In: Jan B, Brian F C C eds. RNA Biochemistry and Biotechnology. Dordrecht: Kluwer Academic Publishers, 1999. 11~ 43
- 10 Szymanski M, Clark B F C, Barciszewski J. RNA on the web. In: Jan B, Brian F C C eds. RNA Biochemistry and Biotechnology. Dordrecht: Kluwer Academic Publishers, 1999. 347~ 352

## 欢迎订阅 2001 年《植物学通报》

《植物学通报》是中国植物学会创办的植物学综合性科技刊物。国内统一刊号 CN11-1945/Q, 国际标准连续出版物号 ISSN 1003-2266。

主要刊登内容: 1. 植物学各分支学科及其有关的农、林、牧、药、环保和轻工等方面有一定理论水平和应用价值的专论与综述; 2. 新技术、新方法; 3. 研究论文和简报; 4. 高等院校植物学教学研讨; 5. 信息动态; 6. 科学家园地。全国性植物学专题学术讨论会论文或摘要以专辑发表。

《植物学通报》的主要读者对象是从事植物学以及有关的农、林、牧、医药、轻工、环保等方面的科技、教学人员, 深受广大读者的欢迎。据中国科学院上海文献情报中心对生命科学 450 种中文期刊近几年流通频次的统计, 在流通率最高的 59 种生命科学中文期刊中, 《植物学通报》排在第 14 位, 居植物学类刊物之首。该刊为“中国自然科学核心期刊”, 据中国科学引文数据库统计数据, 被引频次排名前 500 名期刊中, 该刊在近 3 年内名次提高了 186 名次, 现位于第 211 名。

该刊为双月刊, 每双数月月末出版, 单价每册 14 元, 全年 84 元。欢迎全国各地图书情报单位及广大读者在当地邮局订阅。若错过邮局订阅, 请直接与编辑部联系订阅。除现刊外, 尚有自创刊以来的全部过期刊物, 半价优惠, 共计 180 元 (含邮费)。欲订阅者, 与编辑部联系。该刊编辑部地址: 北京市香山南辛村 20 号, 《植物学通报》编辑部, 邮政编码: 100093, 电话: (010) 62591431-6135。