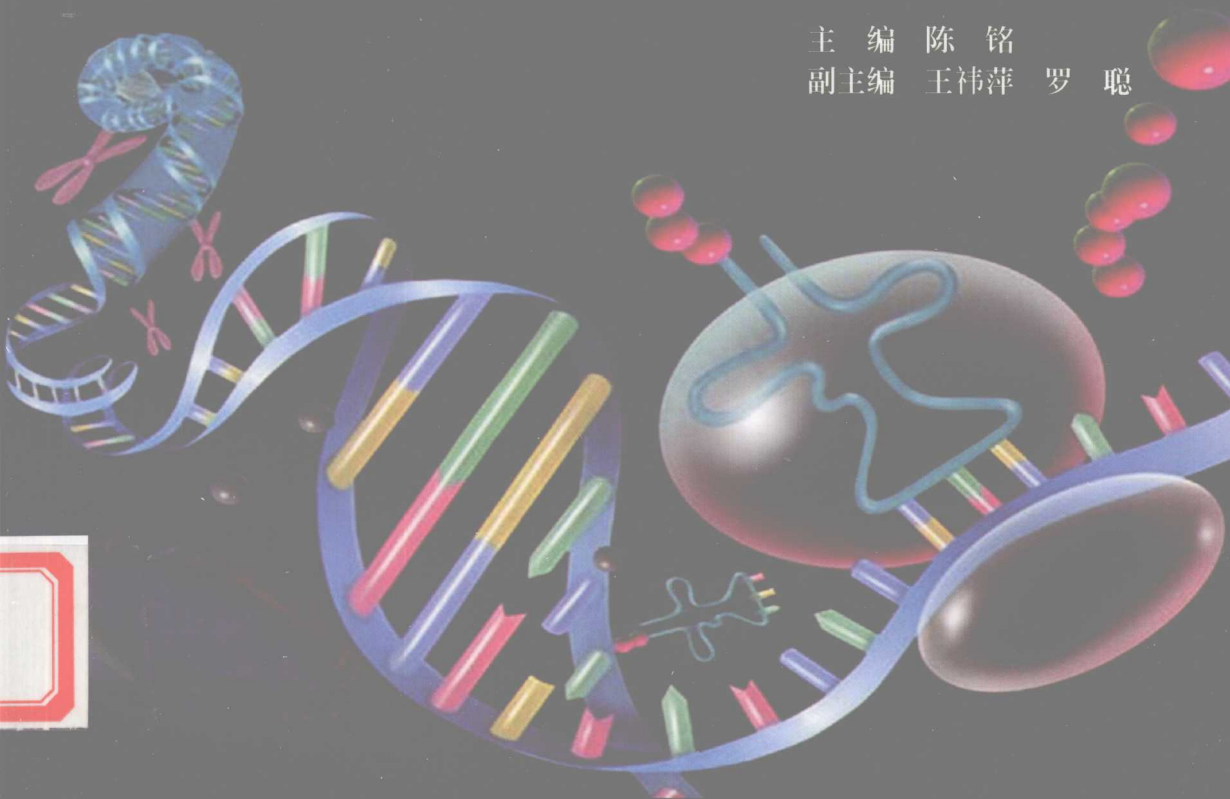
BILINGUISTIC BIOINFORMATICS GLOSSARY

[生物信息学词汇双语通解]

主 编 陈 铭
副主编 王祎苹 罗 聪



ZHEJIANG UNIVERSITY PRESS
浙江大学出版社

生物信息学词汇双语通解

主　编　陈　铭
副主编　王祎萍　罗　聪

ZHEJIANG UNIVERSITY PRESS
浙江大學出版社

图书在版编目(CIP)数据

生物信息学词汇双语通解 /陈铭主编. —杭州:浙江大学出版社,2008.9
ISBN 978-7-308-05924-4

Ⅰ.生… Ⅱ.陈… Ⅲ.生物信息论—词汇—汉、英 Ⅳ.Q811.4—61

中国版本图书馆 CIP 数据核字(2008)第 057580 号

生物信息学词汇双语通解

主　编　陈　铭

副主编　王祎萍　罗　聪

责任编辑　余健波
封面设计　张作梅
出版发行　浙江大学出版社
(杭州天目山路 148 号　邮政编码 310028)
(E-mail:zupress@mail.hz.zj.cn)
(网址:http://www.zjupress.com
http:www.press.zju.edu.cn)
电话:0571—88925592,88273066(传真)
排　　版　杭州求是图文制作有限公司
印　　刷　富阳市育才印刷有限公司
开　　本　710mm×1000mm　1/16
印　　张　14
字　　数　383 千字
版 印 次　2008 年 9 月第 1 版　2008 年 9 月第 1 次印刷
书　　号　ISBN 978-7-308-05924-4
定　　价　28.00 元

前　言

生物信息学(Bioinformatics)是一门数学、统计、计算机与生物、医学交叉结合的新兴学科，它已广泛地渗透到生命科学的各个研究领域中，成为生物学、医学发展不可缺少的重要工具。近年来，生物信息学的发展日新月异，无论是从数据库构建还是新技术的形成与应用，生物信息学技术在人类疾病与功能基因的发现与识别、基因与蛋白质的表达与功能研究方面都发挥着关键的作用。生物信息学技术在基于基因与蛋白质功能缺陷的合理化药物设计方面也有着巨大的潜力。同时，生物信息学技术在生物基因资源挖掘、虚拟生命研究开发等方面都有重要的应用。正因为如此，对生物信息学的学习与研究显得尤为重要。

在这样的发展背景下，衍生出许多新的生物信息学相关专业词汇。但是，目前并没有一本合适的书籍提供这些词汇。而在欧洲生物信息学研究所 EBI 网站上有已经整理好的英文词条近 700 条，这是一个非常好的资源库。但是，由于网站处于欧洲并且为全英文的，并不能得到很好的普及应用。为了方便广大的中国学者与学习生物信息学的人能够更直接与方便地查询生物信息学专业词汇，我们经过近一年的努力，编辑了这本由浙江大学生命科学院生物信息系主任陈铭副教授主编，汇集了浙江大学生命科学院三十多名学生努力的《生物信息学词汇双语通解》。

本书共分为三部分，一部分是生物专业词汇，一部分是该词汇的英文释义，最后一部分为词汇的中文释义。这种结构方便读者查阅，即使中文解释不到位，亦可与英文解释对照理解词义。本书还可以作为 EMBOSS 应用程序的说明文档之用，所有应用程序以斜体字区别表示。

词典的翻译是一个复杂而精细的工作，对我们来说，也是一个对生物信息学学习的过程。在此过程中，我们大量地参考各种文献，综合理解词汇的意思来进行翻译，尽可能做到准确。但是，限于时间和水平，本书肯定还有不少缺点甚至错误，衷心希望得到各方面专家、学者的批评和指导。

在《生物信息学词汇双语通解》发表之际，我们对在编写过程中给予我

们帮助和支持的领导和同事表示衷心的感谢。尤其是浙江大学生物信息系的研究生们，对这本词典作了精细的检查与修改工作，为我们顺利完成编写这本书提供了很大帮助。另外，感谢浙江大学、生命科学学院的生物信息学双语课程建设项目的资助。

最后，衷心希望这本《生物信息学词汇双语通解》能为我们的生物信息学教学科研增光添彩。

编　者

2008 年 3 月 24 日

0—9

☞ 3′flanking region

A region of DNA which is not copied into the mature mRNA, but which is present adjacent to 3′ end of the gene. It was originally thought that the 3′ flanking DNA was not transcribed at all, but it was discovered to be transcribed into RNA, but quickly removed during processing of the primary transcript to form the mature mRNA. The 3′ flanking region often contains sequences which affect the formation of the 3′ end of the message. It may also contain enhancers or other sites to which proteins may bind.

☞ 3′侧翼区

指的是存在于基因的3′末端的DNA序列，虽然被转录，但是会在剪接的过程中被切除而使得其转录后的RNA序列不被包含在成熟的mRNA中。3′侧翼区常常含有能够影响mRNA 3′末端形成的信息。不仅如此，3′侧翼区可能还含有增强子或是其他能够与蛋白质结合的位点。

☞ 3′untranslated region (3′UTR)

A region of DNA which is transcribed into mRNA and becomes the 3′ end of the message, but which does not contain protein coding sequence. Everything between the stop codon and the polyA tail is considered to be 3′ untranslated. The 3′ untranslated region may affect the translation efficiency of the mRNA or the stability of the mRNA. It also has sequences which are required for the addition of the poly(A) tail to the message (including one known as the 'hexanucleotide', AAUAAA).

☞ 3′非翻译区

指的是被转录成mRNA的3′末端的DNA序列，但是它并不携带有效的氨基酸编码子序列。存在于终止密码子和多聚腺苷酸尾巴之间的所有序列都被认为是3′非翻译区。3′非翻译区可能会影响到mRNA的翻译效率或者是其稳定性。这个区域同时含有mRNA加尾过程的识别序列(包括我们所熟知的六核苷酸序列：AAUAAA)。

☞ 3Dseq Database

Providing annotation of the PDB sequences to a uniform standard, and providing cross-references to the SWISS-PROT database.

☞ 3Dseq 数据库

为PDB(蛋白质数据库)中的序列提供一个一致、标准的注释，并且提供SWISS-PROT数据库的交叉访问。

☞ 5′ flanking region

A region of DNA which is not transcribed into RNA, but rather is adjacent to 5′ end of the gene. The 5′ flanking region contains the promoter, and may also contain enhancers or other protein binding sites.

☞ 5′侧翼区

指的是不会被转录的，但是相当接近于基因5′末端的DNA序列。这个区域含有启动子，而且可能含有增强子或者其他蛋白质的结合位点。

☞ 5′untranslated region (5′UTR)

A region of a gene which is transcribed into mRNA, becoming the 5′ end of the message, but which does not contain protein coding sequence. The 5′ untranslated region is the portion of the DNA starting from the cap site and extending to the base just before the ATG translation initiation codon. While not itself translated, this region may have sequences which alter the translation efficiency of the mRNA, or which affect the stability of the mRNA.

☞ 5′非翻译区

指的是能够被转录为 mRNA 的 5′末端，但是不携带氨基酸编码序列的区域。这个区域是一段从帽子结构开始到起始密码子之间的 DNA 序列。虽然这段序列本身无法被翻译，但是它能够改变 mRNA 的翻译效率或者可能会影响 mRNA 的稳定性。

A a

☞ Abstract Syntax Notation (ASN.1)

A language that is used to describe structured data types formally. Within bioinformatics, it has been used by the National Center for Biotechnology Information to encode sequences, maps, taxonomic information, molecular structures, and biographical information in such a way that it can be easily accessed and exchanged by computer software.

☞ 抽象语法标记

一种用来描述数据结构类型的格式。在生物信息学领域里，NCBI（美国国立生物技术信息中心）用它编码、定位序列，统计分类学信息、分子结构以及生物图谱信息，计算机软件可方便地访问它并与之交互。

☞ Accession number

An identifier supplied by the curators of the major biological databases upon submission of a novel entry that uniquely identifies that sequence (or other) entry.

☞ 登录号

由主要的生物数据库的管理员根据条目所提供的标识，它是对这个序列（或其他）条目的独一无二的标识。

☞ Accession number line (EMBL)

The AC (Accession Number) line lists the accession numbers associated with this entry.

☞ 登录号行(EMBL)

登录号行列出所有与条目相关的登录号。

☞ Accession number line (SWISS-PROT)

The AC (ACcession number) line lists the accession number(s) associated with an entry.

☞ 登录号行(SWISS-PROT)

登录号行列出与一个（或多个）条目相关的登录号。

☞ **Acquired mutations**

Gene changes that arise within individual cells and accumulate throughout a person's lifetime; also called somatic mutations.

☞ **获得性突变**

个体细胞引起,并在一代中积累的基因变化,也叫做体细胞突变。

☞ **Acrylamide gels**

A polymer gel used for electrophoresis of DNA or protein to measure their sizes (in daltons for proteins, or in base pairs for DNA). Acrylamide gels are especially useful for high resolution separations of DNA in the range of tens to hundreds of nucleotides in length.

☞ **丙烯酰胺凝胶**

一种用于 DNA 或者蛋白质电泳的聚合物凝胶,用来测量它们的相对分子质量(蛋白质:道尔顿;DNA:碱基对)。丙烯酰胺凝胶特别用于高分辨率分离 DNA,在几十到几百碱基对长度之间。

☞ **Adenine**

A purine base found in DNA and RNA.

☞ **腺嘌呤**

在 DNA 和 RNA 中都存在的嘌呤碱基。

☞ **Agarose gels**

A polysaccharide gel used to measure the size of nucleic acids (in bases or base pairs). This is the gel of choice for DNA or RNA in the range of thousands of bases in length, or even up to 1 megabase if you are using pulsed field gel electrophoresis.

☞ **琼脂糖凝胶**

一种用于测量核苷酸大小(碱基或碱基对)的多聚糖凝胶。用来分离几千碱基大小的 DNA 和 RNA,如果用脉冲凝胶电泳可以分离到数量级为百万碱基的核酸。

☞ **Alanine**

A non-essential amino acid in mammals. It contains an aliphatic methyl group as its sidechain.

☞ **丙氨酸**

哺乳动物中非必需氨基酸。含有一个脂肪甲基基团侧链。

☞ **Algorithm**

A series of steps defining a procedure or formula for solving a problem, that can be coded into a programming language and executed. Bioinformatics algorithms typically are used to process, store, analyse, visualise and make predictions from biological data.

☞ **算法**

用来定义一个解决问题的过程或公式的一系列步骤,能够通过计算机编程语言来编码及执行。生物信息学算法一般用于处理、储存、分析,实现可视化以及通过生物数据来做预测。

☞ ***align***

An EMBOSS application. This program uses the Needleman-Wunsch global alignment algorithm to find

☞ ***align* 程序**

EMBOSS 应用程序之一。该程序采用 Needleman-Wunsch 全局比对

the optimum alignment (including gaps) of two sequences when considering their entire length. The Needleman-Wunsch algorithm is a member of the class of algorithms that can calculate the best score and alignment in the order of mn steps, (where 'n' and 'm' are the lengths of the two sequences). These dynamic programming algorithms were first developed for protein sequence comparison by Needleman and Wunsch, though similar methods were independently devised during the late 1960's and early 1970's for use in the fields of speech processing and computer science. What is the optimal alignment? Dynamic programming methods ensure the optimal global alignment by exploring all possible alignments and choosing the best. It does this by reading in a scoring matrix that contains values for every possible residue or nucleotide match. Needle finds an alignment with the maximum possible score where the score of an alignment is equal to the sum of the matches taken from the scoring matrix. An important problem is the treatment of gaps, i. e., spaces inserted to optimise the alignment score. A penalty is subtracted from the score for each gap opened (the 'gap open' penalty) and a penalty is subtracted from the score for the total number of gap spaces multiplied by a cost (the 'gap extension' penalty). Typically, the cost of extending a gap is set to be 5 - 10 times lower than the cost for opening a gap.

算法来寻找考虑两个序列全长时两者的最佳比对方案(包括空位)。Needleman-Wunsch 算法是以 mn 步(设 n 和 m 为两个序列的长度)的顺序计算最佳比对的算法之一。这类动态规划算法是由 Needleman 和 Wunsch 首次用于蛋白质序列比对的,尽管类似方法是在 20 世纪 60 年代末、70 年代初为语音处理和计算机科学领域的应用分别独立地发明的。什么是最佳比对?动态规划方法可通过考察所有可能的比对方式并选择最佳方案找到最佳全局配对。该方法通过读入包含所有可能残基或核苷配对的打分矩阵来实现,Needle 发现比对最大可能打分值,其配对值等于得分矩阵中匹配值的总和。一个重要的问题是如何处理空位,也即为最优化配对而插入的空位。每一空位将从总分中扣除罚分(空位罚分),总位数乘以罚分也将从总分中扣除罚分(空位延伸罚分)。典型的空位扩展罚分设为低于空位罚分的 1/10 - 1/5。

☞ Alignment

The result of a comparison of two or more gene or protein sequences in order to determine their degree of base or amino acid similarity. Sequence alignments are used to determine the similarity, homology, function or other degree of relatedness between two or more genes or gene products.

☞ 比对

两个甚至更多的基因或者蛋白质序列进行比较的结果,用以计算它们碱基或者氨基酸的相似度。序列比对用来决定两个甚至更多基因或基因产物的相似度、同源性、功能以及其他的相关度。

☞ **Alignment score**

The alignment score, represents the likelihood that the described alignment is not random, providing an indication of its validity. They are calculated by totaling the scores for each matched pair of residues at each position in the alignment, plus unmatched residues are given the gap open penalty, (the gap penalty for non-affine searches), or the gap extension penalty, if appropriate in the alignment, and if the affine search is running.

☞ **比对得分**

比对得分,代表着相似性,说明描述的比对并不是随机的,并为比对提供正确的依据。计算时,将比对中所有配对的分值相加,并给予非配对残基对空位打开罚分(非仿射搜索的空间罚分),或者空位间隔长度罚分(在适当的情况下采用,并使用仿射搜索)。

☞ **Allele**

A given form of a gene that occupies a specific position or locus on a chromosome. Variant forms of genes occurring at the same locus are said to be alleles of one another.

☞ **等位基因**

基因的一种形式,它位于染色体上一个特定的位置或基因座。在同一基因座上发生的不同形式的基因被称做一个基因的等位基因。

☞ **Alleles**

Variant forms of the same gene. Different alleles produce variations in inherited characteristics such as eye colour or blood type.

☞ **等位基因**

同一基因的不同形式。不同的等位基因产生遗传特征上的变异,例如眼色或者血型。

☞ **Alpha helix**

The alpha helix is the most abundant type of secondary structures in proteins. In brief, alpha helices are formed from stretches of consecutive amino acid residues with phi, psi angle pairs which correspond to the bottom left quadrant of the Ramachandran plot. The mean phi, psi angles for alpha helices found in proteins are −62 degrees and −41 degrees respectively. The alpha helix has 3.6 residues per turn, with a hydrogen bond between the CO of residue n and the NH of residue n + 4. The closed loop formed by one of these hydrogen bonds and the intervening stretch of backbone contains 13 atoms (including the hydrogen). Hence the nomenclature for an alpha helix is 3.6(13)-helix,

☞ **α-螺旋**

α-螺旋是蛋白质二级结构中最常见的一种类型。简而言之,α-螺旋是由与 Remachandran 结构左下象限构象对应的 phi/psi 角的一段连续氨基酸残基对组成的。蛋白质的 α-螺旋中 phi/psi 角的平均值分别为 −62°和 −41°。α-螺旋每圈包含 3.6 个氨基酸残基,在第 n 个残基 CO 和第 n + 4 个残基 NH 间由氢键连接。这个封闭的环由一个氢键和其间的 13 个原子(包括氢原子)的骨架延长组成。因此,α-螺旋的命名是 3.6(13)-螺旋,3.6代表了一圈氨基酸残基的数

where the 3.6 is the number of residues per turn and 13 is the number of atoms in the hydrogen bonded loop. Some alpha helices are curved or show distinct kinks, for example those caused by the presence of proline residues.

量，13 代表了氢键环中的原子数。一些 α-螺旋会发生弯曲或明显的扭曲，比如含有脯氨酸残基的 α-螺旋。

☞ Alternative splicing

One of the alternate combinations of a folded protein that are possible due to recombination of multiple gene segments during mRNA splicing that occurs in higher organisms.

☞ 选择性剪接

折叠蛋白质多种剪接可能性的一种，取决于在高等生物 mRNA 剪接时多基因片段的重组。

☞ Alu family

A common set of dispersed DNA sequences found throughout the human genome; each is about 300 bases long and they are repeated at least 500,000 times. Alu sequences are speculated to have originated from viral RNA sequences that integrated into human DNA thousands of years ago.

☞ Alu 家族

在人类基因组中找到的一种常见的分散 DNA 片段；每个片段大概有 300 碱基长，并且至少被重复了 500000 次。Alu 序列被推测可能起源于几千年前整合到人类 DNA 中的病毒 RNA。

☞ Alzheimer's disease

A disease that causes memory loss, personality changes, dementia and, ultimately, death. Not all cases are inherited, but genes have been found for familial forms of Alzheimer's disease.

☞ 阿尔兹海默症

一种能够引起记忆丢失、个性改变、痴呆以及最终死亡的疾病。并不是所有的病例都是遗传性的，但是该基因是在家族遗传性疾病中找到的。

☞ Amino acid

One of the 20 chemical building blocks that are joined by amide (peptide) linkages to form a polypeptide chain of a protein.

☞ 氨基酸

20 种化学单体的统称，能够通过肽键连接结合形成多肽链，进而形成蛋白质。

☞ Amino acids with acidic side chains

These have a carboxylic acid group in their side chain and are very hydrophilic.

☞ 酸性侧链氨基酸

在支链上有羧酸基团的氨基酸，具有很高的亲水性。

☞ Amino acids with aliphatic hydrophobic side chains

The hydrophobic side chains of these amino acids will

☞ 脂肪族侧链氨基酸

这些氨基酸的疏水支链并不会形

not form hydrogen bonds or ionic bonds with other groups. These hydrophobic amino acids tend to be buried in the centre of proteins away from the surrounding aqueous environment.

成氢键或者与其他基团形成离子键。这些疏水氨基酸通常处在蛋白质的中央，远离外界的水环境。

☞ Amino acids with basic side chains

The positive charge on these side chains makes them hydrophilic and they are likely to be found at the protein surface.

☞ 碱性侧链氨基酸

在支链上的正电荷使得它们亲水，因此它们一般处在蛋白质的表面。

☞ Amino acids with neutral side chains

The single hydrogen atom side chain has no strong hydrophobic or hydrophilic properties.

☞ 中性氨基酸

它只有一个氢原子支链，因此没有很强的亲水或疏水特性。

☞ Amino acids with uncharged but polar side chains

The side chains of these amino acids are uncharged at physiological pH.

☞ 不带电的极性氨基酸

这些氨基酸的支链在生理 pH 值下不带电。

☞ Analogy

Reasoning by which the function of a novel gene or protein sequence may be deduced from comparisons with other gene or protein sequences of known function. Identifying analogous or homologous genes via similarity searching and alignment is one of the chief uses of bioinformatics.

☞ 类比

一个新基因或蛋白质序列的功能可以通过与其他功能已知的基因或蛋白质比对的方式推测出来。通过相似性搜索和序列比对来识别同源基因是生物信息学中的主要方法。

☞ Angiosperms

Any member of the more than 250,000 species of flowering plants. Angiosperms are often differentiated from gymnosperms by their production of seeds within a closed chamber (ovary). The Angiosperms division is composed of two classes, monocotyledons and dicotyledons.

☞ 被子植物

属于开花植物，超过 250000 种。被子植物通常是从裸子植物分化而来的，其种子有子房包被。被子植物包括两类：单子叶植物和双子叶植物。

☞ Annotation

A combination of comments, notations, references, and citations, either in free format or utilising a controlled vocabulary, that together describe all the experimental

☞ 注释

一种由评论、符号、参考文献和引用组成的结合体，使用自定的格式或者限制性词表，它描述了关于基

and inferred information about a gene or protein. Annotations can also be applied to the description of other biological systems. Batch, automated annotation of bulk biological sequence is one of the key uses of Bioinformatics tools.

☞ Antibiotic resistance

Plasmids generally contain genes which confer on the host bacterium the ability to survive a given antibiotic. If the plasmid pBR322 is present in a host, that host will not be killed by (moderate levels of) ampicillin or tetracycline. By using plasmids containing antibiotic resistance genes, the researcher can kill off all the bacteria which have not taken up this plasmid, thus ensuring that the plasmid will be propagated as the surviving cells divide.

☞ *antigenic*

An EMBOSS application. Antigenic predicts potentially antigenic regions of a protein sequence, using the method of Kolaskar and Tongaonkar. Analysis of data from experimentally determined antigenic sites on proteins has revealed that the hydrophobic residues Cys, Leu and Val, if they occur on the surface of a protein, are more likely to be a part of antigenic sites. A semi-empirical method which makes use of physicochemical properties of amino acid residues and their frequencies of occurrence in experimentally known segmental epitopes was developed by Kolaskar and Tongaonkar to predict antigenic determinants on proteins. Application of this method to a large number of proteins has shown that their method can predict antigenic determinants with about 75% accuracy which is better than most of the known methods. This method is based on a single parameter and thus very simple to use.

因或者蛋白质的所有实验和推断信息。注释同样可以应用于其他生物系统的描述。批处理生物序列的注释是生物信息学工具的一个重要应用。

☞ 抗生素抗性

质粒通常包含基因，这种基因使宿主细菌能够在抗生素存在的情况下存活。如果 pBR322 质粒存在于宿主中，宿主就不会被氨苄霉素或者四环素（中等水平）杀死。通过利用包含抗生素抗性基因的质粒，研究者能够杀死所有没有转入质粒的细菌，从而确保质粒转入存活的细胞分布中。

☞ *antigenic* 程序

EMBOSS 应用程序之一。Antigenic 使用 Kolaskar 和 Tongaonkar 的方法预测一段蛋白质序列中潜在的抗原域。蛋白质抗原表位的实验数据显示，如果含有疏水残基的半胱氨酸、亮氨酸、缬氨酸出现在蛋白质表面，则它们很可能是抗原表位的一部分。Kolaskar 和 Tongaonkar 开发了半经验法，这种方法利用氨基酸残基的物理化学性质和它们在实验已知的抗原决定簇中出现的频率来预测蛋白质抗原决定簇。通过大量蛋白质应用这种方法的实验结果表明，该法预测抗原决定簇的准确率达到 75%，优于已知的其他方法。它基于单一的参数，因而使用简便。

☞ Antisense

DNA or RNA composed of the complementary sequence to the target DNA/RNA. Also used to describe a therapeutic strategy that uses antisense DNA or RNA sequences to target specific gene DNA sequences or mRNA implicated in disease, in order to bind and physically inhibit their expression by physically blocking them.

☞ 反义

目标 DNA 或 RNA 序列的互补片段。也用来描述使用反义序列进行治疗的方法，即利用 DNA 或 RNA 反义序列与疾病中特定基因序列或 mRNA 结合，从而阻断或抑制它们的表达。

☞ Arginine

An essential amino acid in rats and young humans. It contains a hydrophilic positively charged sidechain.

☞ 精氨酸

鼠和年轻人中的一种必需氨基酸，它含有亲水的正电性侧链。

☞ ArrayExpress

ArrayExpress is a public repository for microarray based gene expression data.

☞ ArrayExpress 数据库

ArrayExpress 是一个公开的基于微阵列的基因表达数据库。

☞ Asparagine

A non-essential amino acid in mammals. It is an uncharged amide derivative of the amino acid aspartic acid.

☞ 天冬酰胺

哺乳动物中的一种非必需氨基酸，是天冬氨酸的去电荷氨基衍生物。

☞ Aspartic Acid

A non-essential amino acid in mammals. It contains a hydrophilic negatively charged carboxylase group sidechain.

☞ 天冬氨酸

哺乳动物中的一种非必需氨基酸，带有一个亲水的负电性羧基侧链。

☞ Assay

A method for measuring a biological activity. This may be enzyme activity, binding affinity, or protein turnover. Most assays utilize a measurable parameter such as color, fluorescence or radioactivity to correlate with the biological activity.

☞ 检验

对生物活性物质的分析测定，包括酶活力、亲和力及蛋白质折叠。很多化验分析利用与生物活性物质相关的参数来度量，如颜色、荧光性、放射性等。

☞ Autoradiography

A method used to locate radioisotope-labeled materials which have been separated in gels or are present in blots. The location of the radiolabeled material is determined by overlaying the test material with a photographic film that is sensitive to the radioisotope.

☞ 放射自显影

一种定位凝胶电泳或杂交后经放射标记样品的技术。对放射性敏感的照片上放射性标记物的位置即反映了待测样品的位点。

☞ **Autosome**

Any of the non-sex-determining chromosomes. Human cells have 22 pairs of autosomes.

☞ **常染色体**

非性别决定相关的染色体。人类细胞含有 22 对常染色体。

B b

☞ ***backtranseq***

An EMBOSS application. It takes a protein sequence and makes a best estimate of the likely nucleic acid sequence it could have come from. It does this by using a codon frequency table. For each amino acid, the corresponding most frequently occuring codon is used in the construction of the nucleic acid sequence.

☞ ***backtranseq* 程序**

EMBOSS 应用程序之一。它对给定的一段蛋白质序列，运用密码子频率来分析编码该蛋白质序列的最可能的核酸序列。对于每个氨基酸，相应的出现频率最高的密码子会被该程序采用。

☞ **Bacteriophage**

A virus that infects bacteria. The bacteriophage DNA has served as a basis for cloning vectors, and is also utilized to create phage libraries containing human or other genes.

☞ **噬菌体**

一种可以感染细菌的病毒。噬菌体 DNA 可作为一种基本的克隆载体，还可以用来创建人类或其他物种的基因噬菌体文库。

☞ ***banana***

An EMBOSS application. It predicts bending of a normal (B) DNA double helix, using the method of Goodsell & Dickerson, This program calculates the magnitude of local bending and macroscopic curvature at each point along an arbitrary B-DNA sequence, using any desired bending model that specifies values of twist, roll and tilt as a function of sequence. The data, based on the nucleosome positioning data of Satchwell, et al. 1986, correctly predicts experimental A-tract curvature as measured by gel retardation and cyclization kinetics and successfully predicts curvature in regions containing phased GGGCCC sequences. (This is the model 'a' described in the Goodsell & Dickerson paper). This model-showing local bending

☞ ***banana* 程序**

EMBOSS 运用程序之一。它运用 Goodsell & Dickerson 方法（NAR 1994）预测一个普通的（B 型）DNA 双螺旋的弯曲度。这个程序可对任意 B-DNA 序列的每一个点计算局部的弯曲程度大小和肉眼可见的曲率大小，并运用任何具有特征螺旋扭曲值的弯曲度模型进行预测。1986 年 Satchwell 等人的关于核小体分布的数据成功地预测了试验中根据凝胶延滞和环化动力学方法得到的 A-tract 弯曲度，同时成功预测了含有 GGGCCC 序列的弯曲度（这是 Goodsell &

at mixed sequence DNA, strong bends at the sequence GGC, and straight, rigid A-tracts is the only model, out of six models investigated in Goodsell & Dickerson paper, that is consistent with both solution data from gel retardation and cyclization kinetics and structural data from X-ray crystallography. The consensus sequence for DNA bending is 5 As and 5 non-As alternating. 'N' is an ambiguity code for any base, and 'B' is the ambiguity code for 'not A' so 'BANANA' is itself a bent sequence-hence the name of this program. The program outputs both a graphical display and a text file of the results.

Dickerson 的文章中所描述的"a"模型)。在 Goodsell & Dickerson 所研究的6个模型中,这个模型是唯一一个显示了在混合 DNA 序列中的局部弯曲度的模型, GGC 序列有紧密结合、直的、刚性的 A-tract 的模型,这个与由凝胶后滞和环化动力学以及从 X-晶体射线得到的结论是一致的。这个 DNA 结合的调和序列是由 5′-A 端和非 5′-A 端交互形成。"N"是指任意一个氨基酸碱基,"B"是任意不是"A"的碱基,因此,"BANANA"本身是一个具有倾向性的序列,这也是该程序的名字由来。这个程序具有视图和文本两种输出结果。

☞ Base pair

A pair of nitrogenous bases (a purine and a pyrimidine), held together by hydrogen bonds, that form the core of DNA and RNA i. e., the A-T, G-C and A-U interactions.

☞ 碱基对

一对含氮的碱基(包括一个嘌呤和一个嘧啶),碱基之间由氢键连接,进而构成 DNA 和 RNA 的核心,如腺嘌呤和胸腺嘧啶结合,腺嘌呤和尿嘧啶结合。

☞ Beta-Alanine

The only naturally occurring beta amino acid.

☞ β-丙氨酸

唯一的天然的β-氨基酸。

☞ Beta barrel

In some instances large anti-parallel (or parallel) sheets can roll up completely to join edges and form a cylinder or closed 'barrel', in which the first strand is hydrogen bonded to the last. The strands form the 'staves' of the barrel.

☞ β-桶状物

在一些实例中,反向平行的(或平行的)折叠会完全卷曲而与边缘结合,第一列通过氢键与最后一列连接,从而形成一个圆筒或者一个类似的"桶"。这些行列组成了桶的"木板"。

☞ Beta propeller

This fold comprises 4,6,7 or 8 antiparallel beta sheets which have a radial arrangement that gives them the

☞ β-折叠片

这种折叠由4、6、7或8个反向平行的β-折叠组成,这些放射状排

appearance of a propeller. The beta sheets pack face to face, and form a closed structure, in which each beta sheet packs against two neighbours. The sheets have a mean twist and orientation that depends on the number of beta sheets that form the propellor. The beta sheets have the same topology, i. e., the propeller blades are structurally similar units (4 antiparallel strands with +1, +1, +1 connectivity).

列的β-折叠具有螺旋状外观。β-折叠是用面对面的方式折叠的,形成一个封闭结构,每个β-折叠都有两个毗邻的折叠。折叠的平均弯曲度和方向取决于这个螺旋桨形所具有的β-折叠数。β-折叠有相同的拓扑学结构,例如螺旋桨形都有相似的结构单位(4个反向平行的列具有+1,+1,+1的连通性)。

☞ Beta sandwich

This protein architecture comprises two beta sheets that pack together, face-to-face, in a layered arrangement. The connections between the strands in the sheet differ between beta sandwiches with different topologies.

☞ β-链夹层

这种蛋白质结构由两个相对的β-折叠组成,分层排列。从拓扑学的角度看,β-折叠中的各列的连接方式与β-链夹层中的连接方式不同。

☞ Beta sheet (beta pleated sheet)

A secondary protein structure in which two or more extended polypeptide chains are hydrogen-bonded to one another in a planar array; the 'pleats' are formed by angles of bonds in the peptide chains. In addition there is a left-handed twist between adjacent strands when looking at right angles to the strand direction. The beta strands in a sheet can be arranged to form parallel, antiparallel or mixed sheets.

☞ β-折叠

一种蛋白质的二级结构,当两个或更多多肽链通过氢键相互结合时形成的平面排列;"褶皱"是由多肽链的键角形成的。另外,当沿着列的方向向右边角度看时在邻近的列上有一个左手扭曲。折叠中的β-列会排列形成平行、反向平行或两种兼有的折叠。

☞ Binding site

A place on cellular DNA to which a protein (such as a transcription factor) can bind. Typically, binding sites might be found in the vicinity of genes, and would be involved in activating transcription of that gene (promoter elements), in enhancing the transcription of that gene (enhancer elements), or in reducing the transcription of that gene (silencers). NOTE that whether the protein in fact performs these functions may depend on some condition, such as the

☞ 结合位点

可供一个蛋白质分子(如一个转录因子)结合的DNA位点。一般地,结合位点可以在基因的附近,也可以是在基因转录激活(启动元件)、基因转录增强(增强元件)、基因转录减弱(沉默子)区域。注意,蛋白质是否能够真正发挥作用还依赖于一些其他条件,如激素的存在、基因所在的组织。结