生物信息学入门

2021年10月8日 13:42

为什么说 AlphaFold 2 足以改变全人类?

来自 <https://xw.qq.com/partner/vivoscreen/20210720A00CU7/20210720A00CU700?showComments=0&isNews=1>

国内外有哪些开放/免费的生物信息云计算平台？ - 知乎 (zhihu.com)

如何自学入门生物信息学 - 知乎 (zhihu.com)

生物信息学_山东大学_中国大学MOOC(慕课) (icourse163.org)
山东大学课程

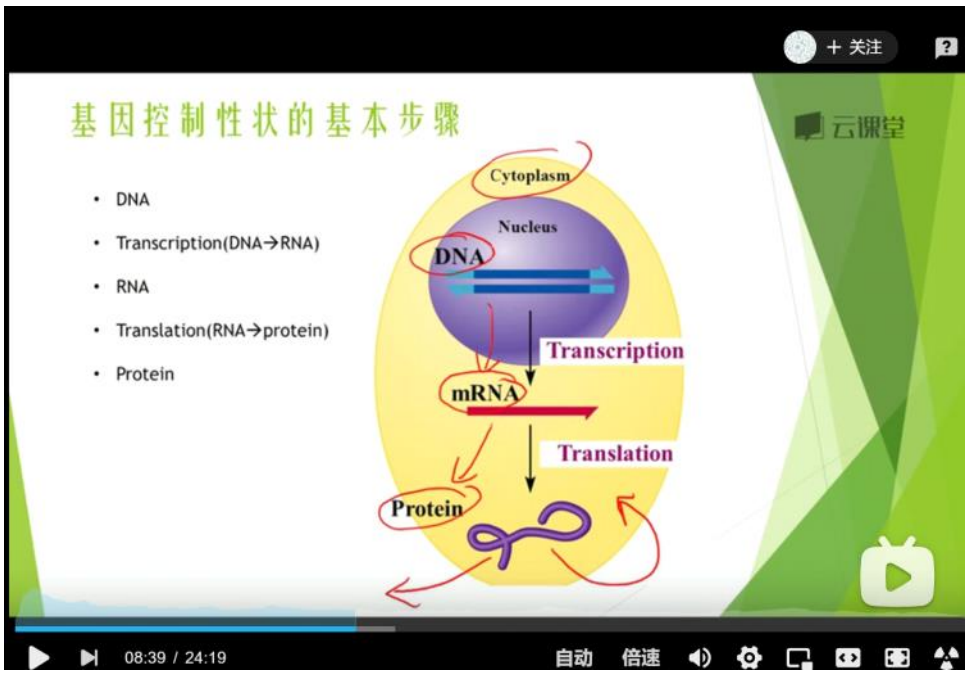
Functional genomics II | EMBL-EBI Training

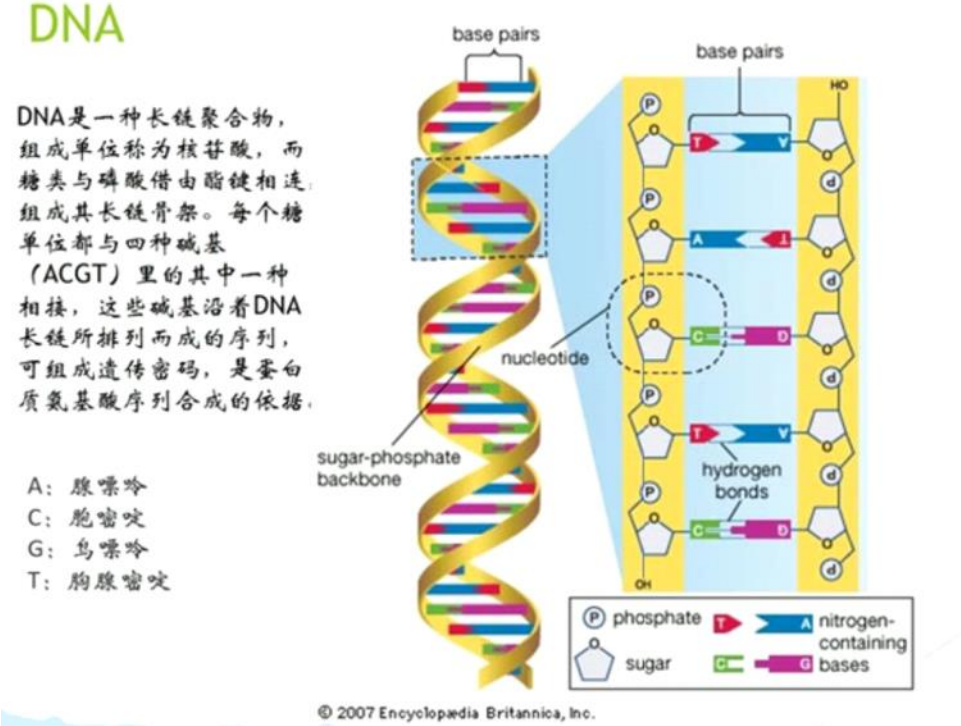
如何自学生物信息学：从菜鸟到专家 - 知乎 (zhihu.com)

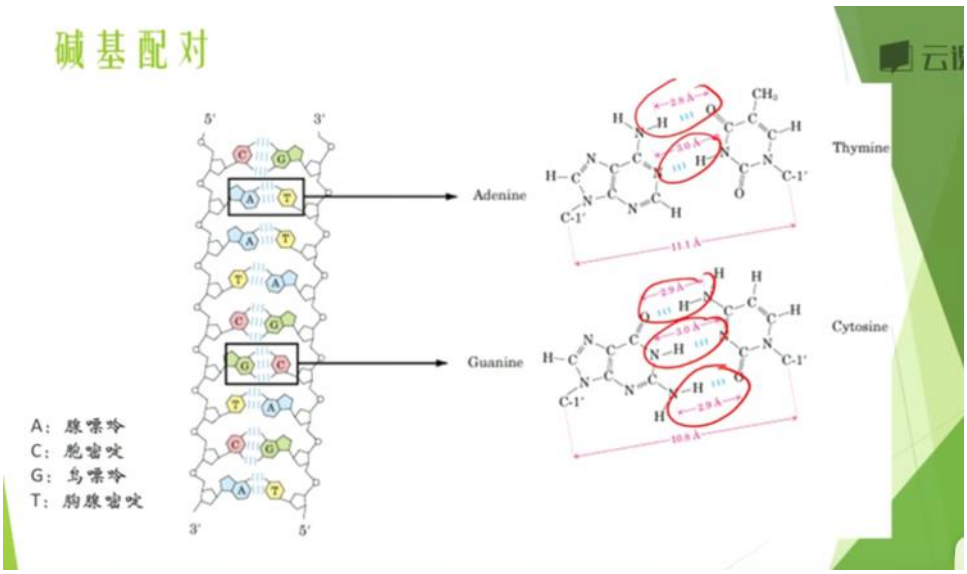
生物信息学入门必看的87个名词：每个都要记牢 - 知乎 (zhihu.com)

干货满满的生物信息学入门课程 - 知乎 (zhihu.com)


总结：到底用谁？
能否找到一致度≥30%的模板？
SWISS-MODEL
能否舒服的穿到现有的结构里？
I-TASSER
Protein Structure & Function Predictions
氨基酸序列长度是否<200？
QUARK
Ab Initio Prediction
ROBETTA
BETA
Full-chain Protein Structure Prediction Server

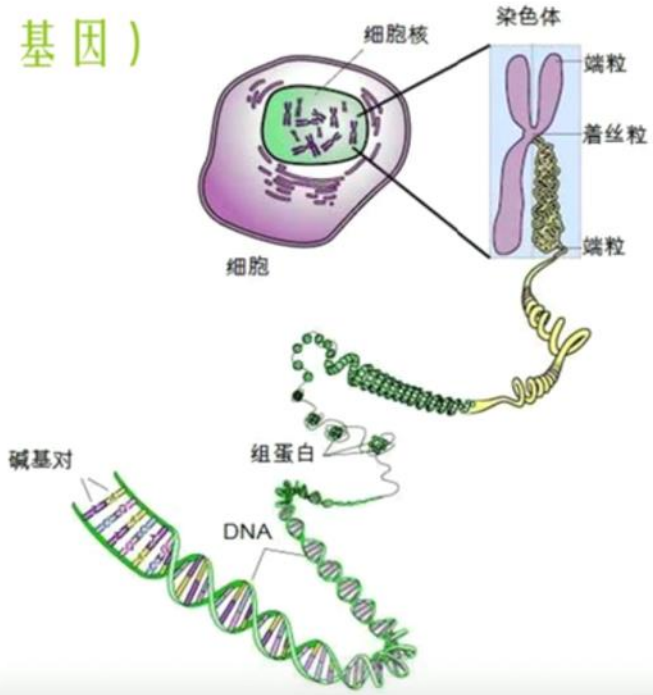

+ 关注
基因控制性状的基本步骤
云课堂
DNA
Transcription(DNA→RNA)
RNA
Translation(RNA→protein)
Protein
Cytoplasm
Nucleus
DNA
Transcription
mRNA
Translation
Protein
08:39 / 24:19
自动
倍速





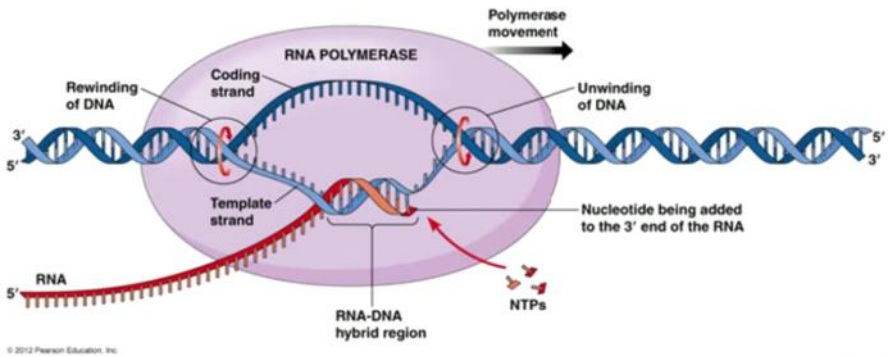
AT两建联合，结构相对没有 CG 三建耦合的结构稳定

DNA（基因）



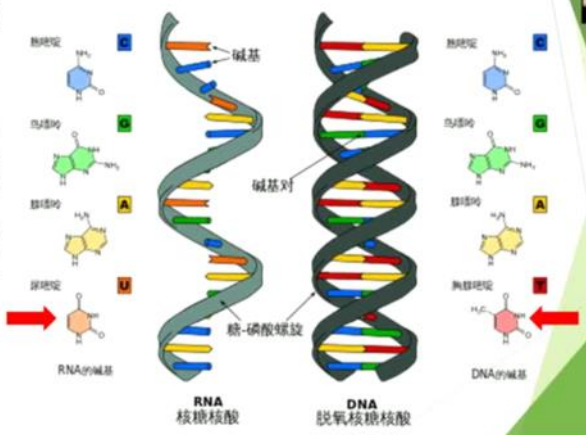
Transcription

转录中，一个基因会被读取、复制为mRNA；就是说一特定的DNA片段作为模板，以DNA依赖的核糖核酸聚合酶（RNA聚合酶或RNA合成酶）作为催化剂而合成前体mRNA的过程。

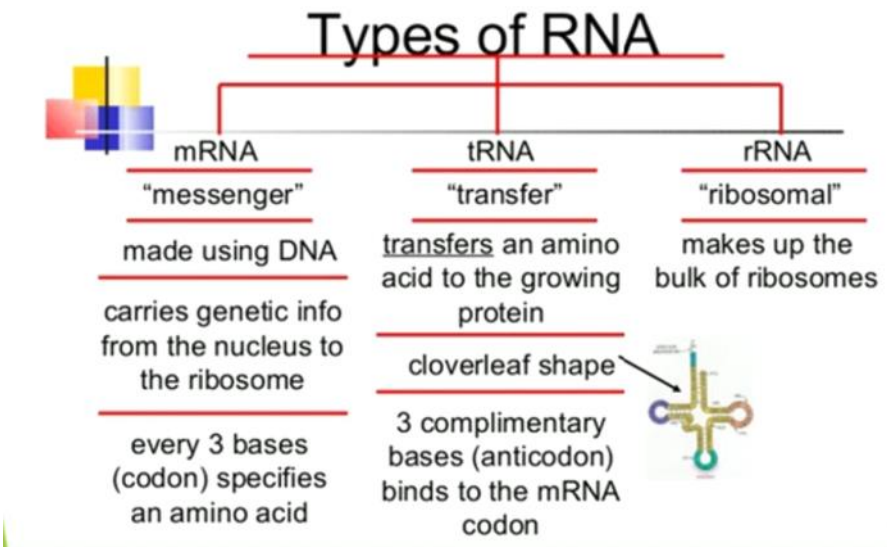


RNA

RNA有着多种多样的功能，可在遗传编码、翻译、调控、基因表达等过程中发挥作用。按RNA的功能，可将RNA分为多种类型。比如，在细胞生物中，mRNA（信使RNA）为遗传信息的传递者，它能够指导蛋白质的合成。因为mRNA有编码蛋白质的能力，它又被称为编码RNA。而其他没有编码蛋白质能力的RNA则被称为非编码RNA（ncRNA）。

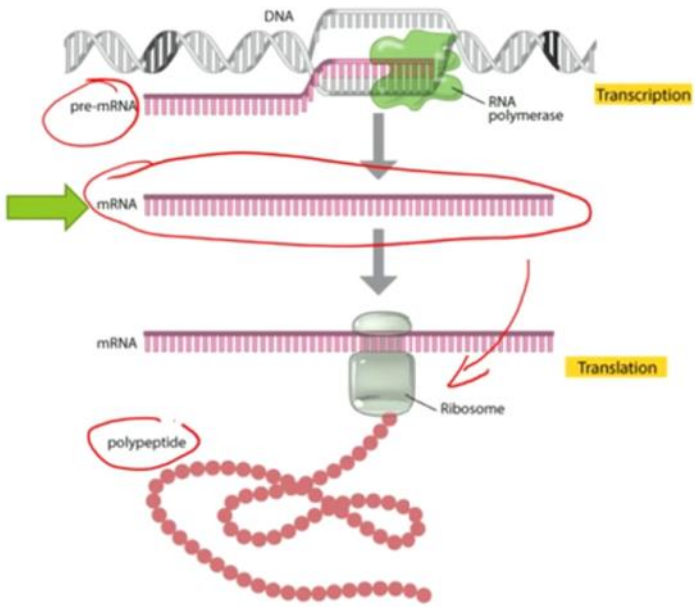


RNAs



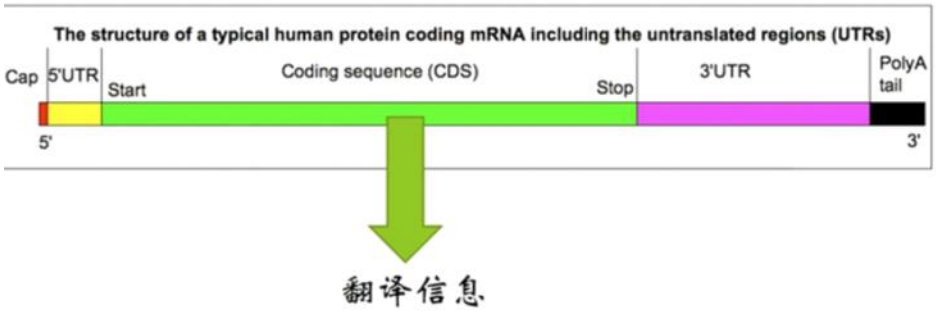
mRNA

信使RNA（messenger RNA，缩写：mRNA），是由DNA经由转录而来，带着相应的遗传讯息，为下一步翻译成蛋白质提供所需的讯息。



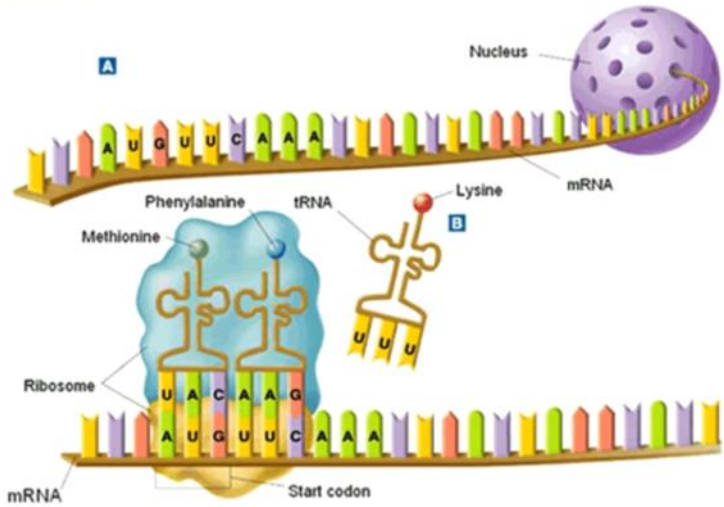
mRNA结构

成熟真核细胞的mRNA的结构。一个完整的mRNA包括有5'端帽、5'非翻译区、编码区、3'非翻译区和poly(A)尾链



翻译信息

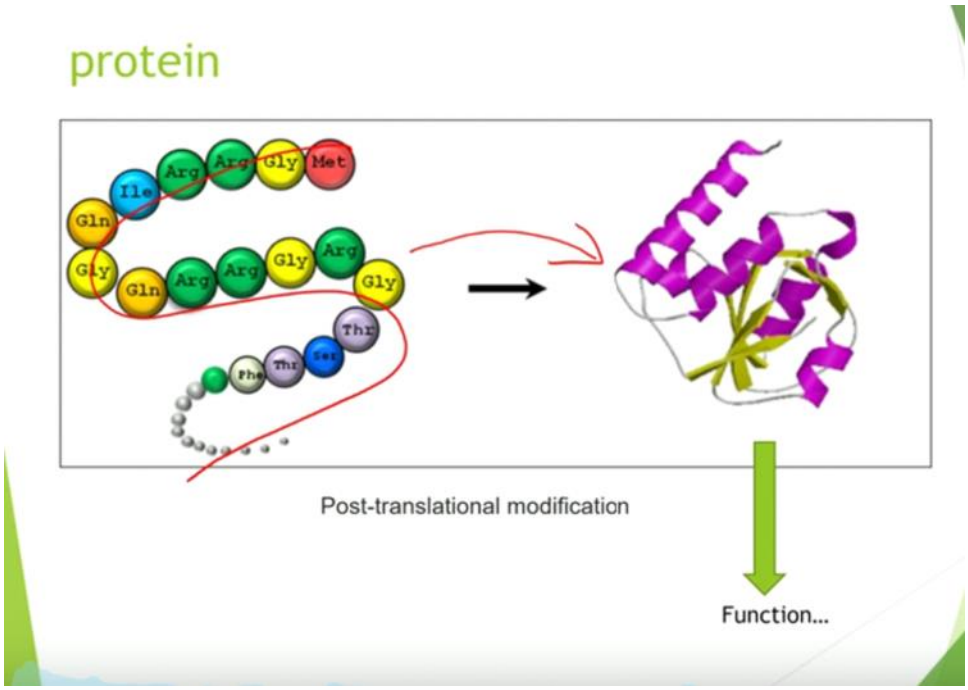
Translation



氨基酸密码子表

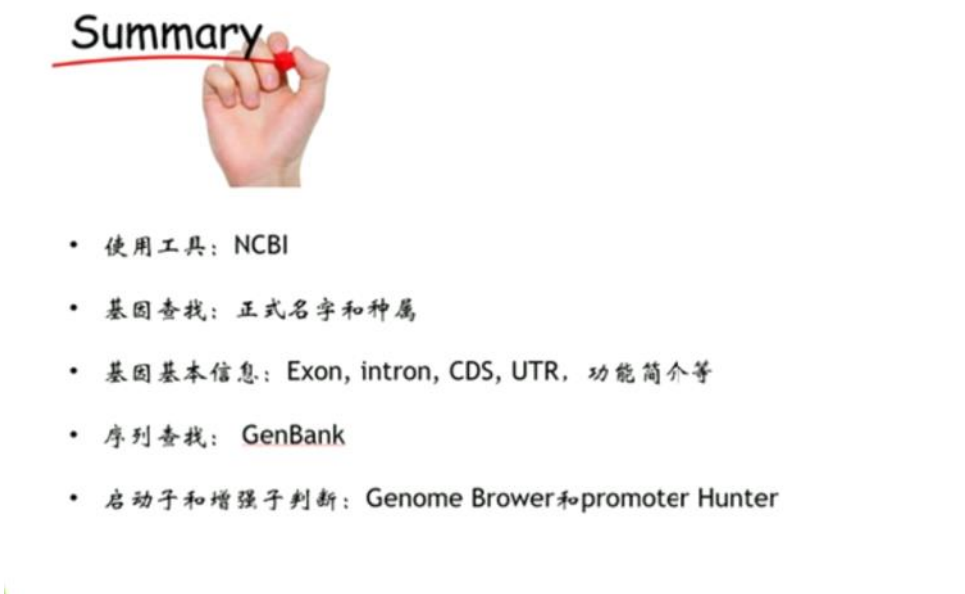
遗传密码表

第一位核苷酸	第二位核苷酸 U	第二位核苷酸 C	第二位核苷酸 A	第二位核苷酸 G	第三位核苷酸
U	UUU 苯丙氨酸 (Phe)	UCU 丝氨酸 (Ser)	UAU 酪氨酸 (Tyr)	UGU 半胱氨酸 (Cys)	U
	UUC	UCC	UAC	UGC	
	UUA 亮氨酸 (Leu)	UCA	UAA 终止密码	UGA 终止密码	
	UUG	UCG	UAG	UGG 色氨酸 (Trp)	
C	CUU 亮氨酸 (Leu)	CCU 脯氨酸 (Pro)	CAU 组氨酸 (His)	CGU 精氨酸 (Arg)	C
	CUC	CCC	CAC	CGC	
	CUA	CCA	CAA 谷氨酰胺 (Gln)	CGA	
	CUG	CCG	CAG	CGG	
A	AUU 异亮氨酸 (Ile)	ACU 苏氨酸 (Thr)	AAU 天冬酰胺 (Asn)	AGU 丝氨酸 (Ser)	A
	AUC	ACC	AAC	AGC	
	AUA	ACA	AAA 赖氨酸 (Lys)	AGA 精氨酸 (Arg)	
	AUG 蛋氨酸 (Met) 或起始密码	ACG	AAG	AGG	
G	GUU 缬氨酸 (Val)	GCU 丙氨酸 (Ala)	GAU 天冬氨酸 (Asp)	GGU 甘氨酸 (Gly)	G
	GUC	GCC	GAC	GGC	
	GUA	GCA	GAA 谷氨酸 (Glu)	GGA	
	GUG	GCG	GAG	GGG	



[Homo sapiens (ID 51) - Genome - NCBI (nih.gov)]

查询基因数据信息网站



如何研究基因

想要

- 检测基因表达水平
- 克隆基因/基因片段在体外做功能分析
- 改变基因的序列
- 测序
- 检测基因突变
- 等等

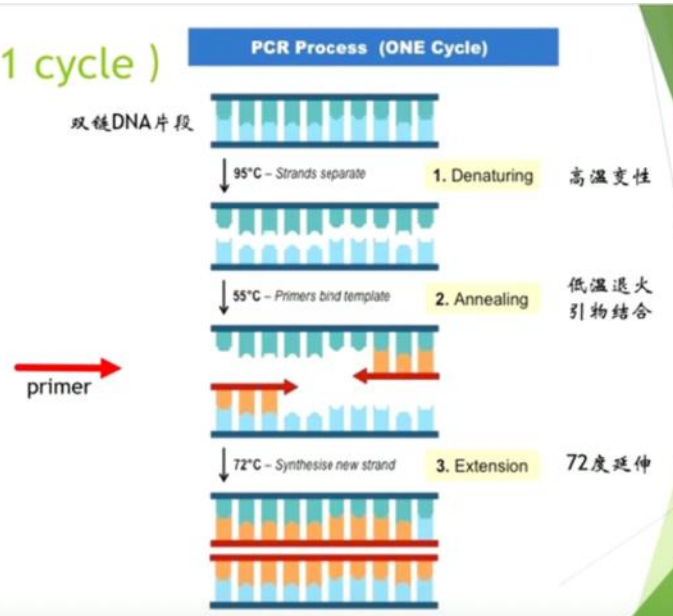
PCR

什么是PCR?

聚合酶链式反应（Polymerase Chain Reaction，PCR），是一种分子生物学技术，用于扩增特定的DNA片段，这种方法可在生物体外进行。

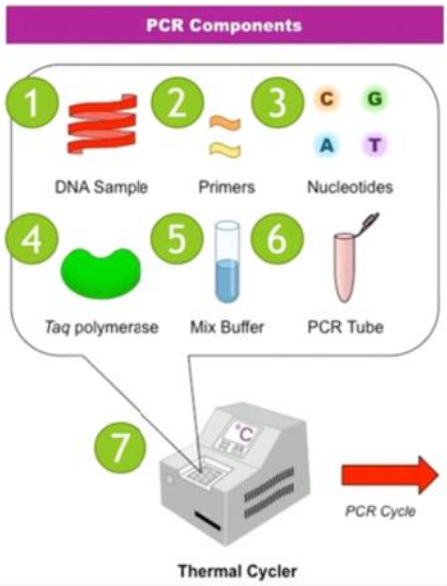
不是这种链式反应

PCR原理（1 cycle）



PCR必需品

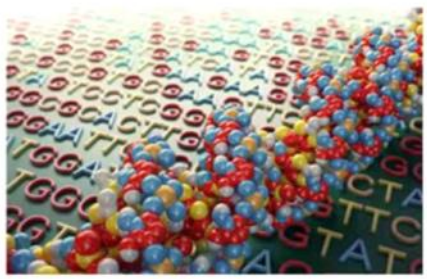
1. 模板DNA
2. 引物
3. dNTP
4. 聚合酶
5. 缓冲液
6. PCR管
7. PCR仪



为什么要DNA测序?

DNA序列是分子生物学研究的基础

测序是建立基础的最主要手段



应用范围:

基因组序列分析（如人类基因组计划）

基因突变分析

PCR片段/质粒序列分析

DNA测序发展历程

云课堂

第一代

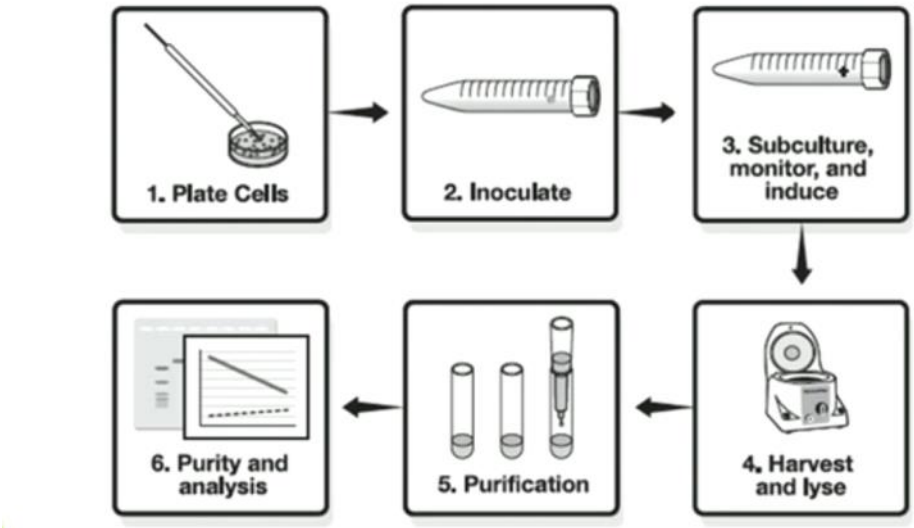
第二代

第三代

第四代

测序技术	原理	公司或仪器	测序通量	测序时间	准确率	读长	优缺点	应用状况
第一代	Sanger 双脱氧法	ABI的3730XL	0.2Mb	1.6m	>99%	400-900	高读长、高精确，一次性达标率高；成本相对高、通量相对低	成本高、速度慢，应用少
第二代	边合成边测序，可逆终止法	Illumina的Solexa，Roche的454，ABI的SOLID	400Mb-1.8T	2h-3d	>99%	50-300	高通量、低成本，但存在模板扩展和序列读长的缺陷	目前应用最广泛的技术
第三代	单分子合成测序	BioScience的Heloscope，PacBio的SMRT	0.2-30Gb	2h	<90%	>1000	高通量、高读长，低成本，但准确性不高	研发阶段，未真正商业应用
第四代	纳米孔外切酶测序	Oxford的MinION、GrindION	5-50Gb	1.2-2h	>90%	>1000	高通量、高读长，低成本、小型化	研发阶段，未真正商业应用

蛋白质表达方法



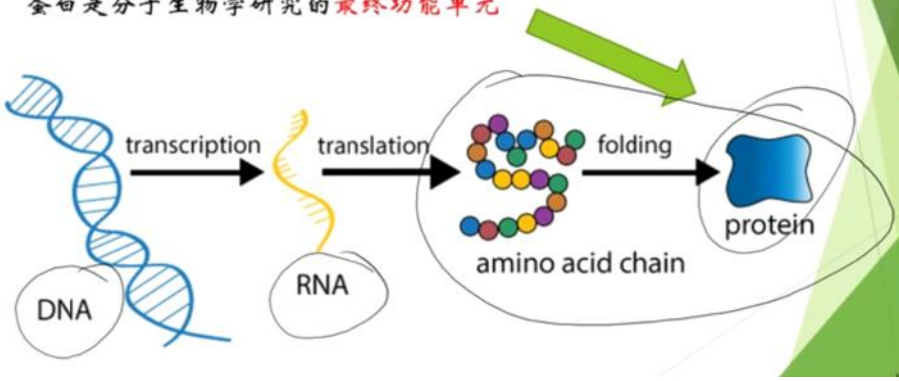
蛋白质表达步骤

1. 基因设计
2. 编码优化
3. 细胞选择
4. SDS验证
5. 蛋白纯化

为什么要研究蛋白质?

基因是分子生物学研究的起始和基本单元

蛋白是分子生物学研究的最终功能单元



蛋白质结构

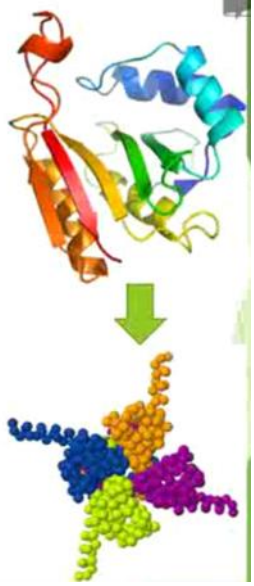
蛋白质一级结构：组成蛋白质多肽链的线性氨基酸序列。一个蛋白质是一个聚酰胺。

蛋白质二级结构：依靠不同氨基酸之间的C=O和N-H基团间的氢键形成的稳定结构，主要为α螺旋和β折叠。因为二级结构是局部的，不同的二级结构的许多区域可存在于相同的蛋白质分子。

蛋白质结构

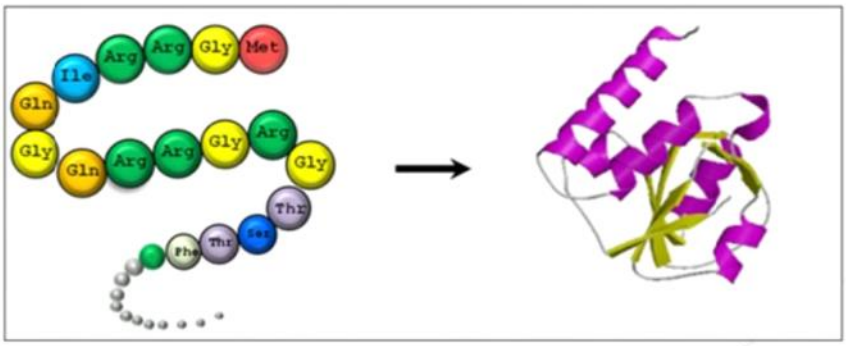
蛋白质三级结构：通过多个二级结构元素在三维空间的排列所形成的一个蛋白质分子的三维结构，是单个蛋白质分子的整体形状。蛋白质的三级结构大都有一个疏水核心来稳定结构，同时具有稳定作用的还有盐桥(蛋白质)、氢键和二硫键，甚至翻译后修饰。“三级结构”常常可以用“折叠”一词来表示。三级结构控制蛋白质的基本功能。

蛋白质四级结构：由几个蛋白质分子（多肽链），通常称为蛋白质亚基所形成的结构，在功能上作为一个蛋白质复合体。



蛋白质序列决定蛋白质功能的第一步

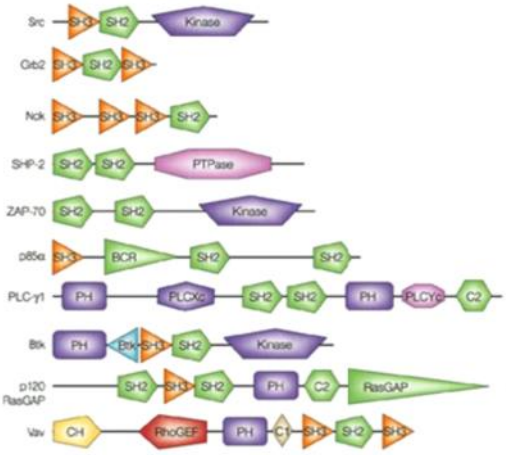
- 氨基酸序列
- 剪接
- 折叠
- 转运
- 激活



蛋白质序列包含不同的结构域（domain）

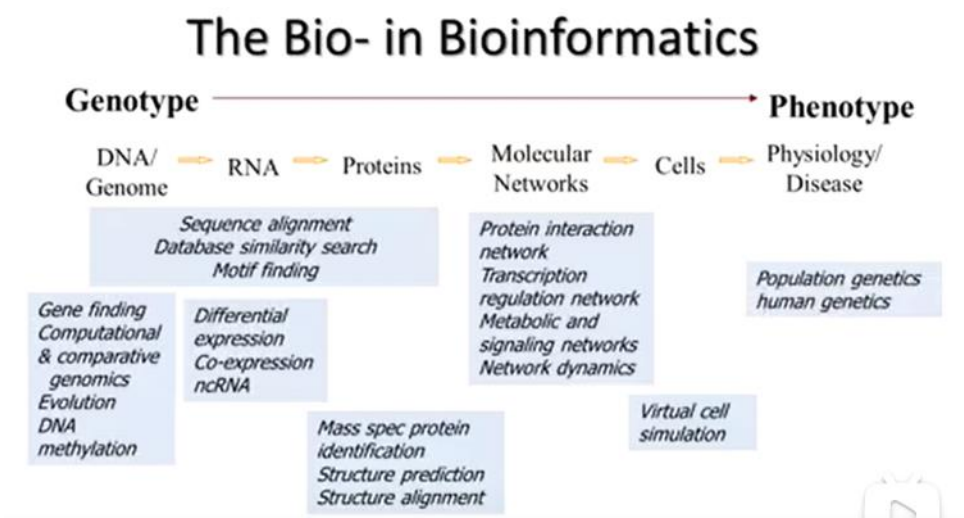
蛋白质结构域（英语：protein domain）是蛋白质中的一类结构单元，是构成蛋白质（三级）结构的基本单元。

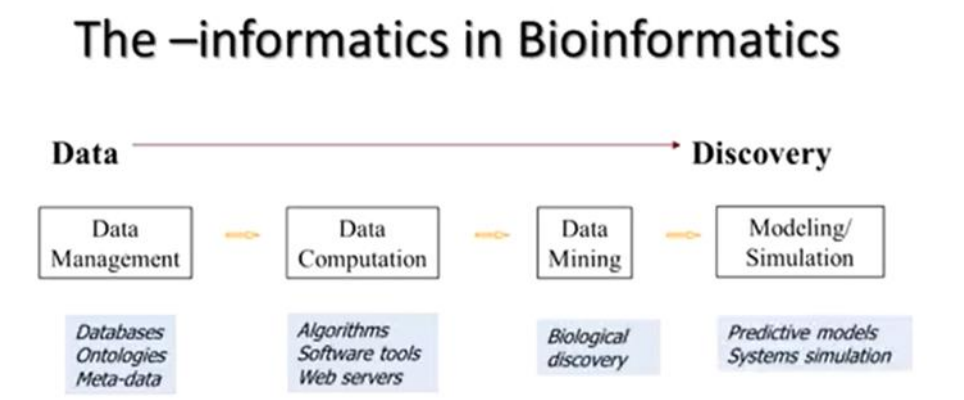
- DNA结合区
- 激酶区域
- 蛋白降解结构域
- 蛋白激活区域
- 等等



Nature Reviews | Molecular Cell Biology

[生物信息学全套视频课程 - 知乎 (zhihu.com)]





[生物信息快速入门_哔哩哔哩_bilibili]