華中科技大學

二〇〇二年招收硕士研究生入学考试试题

考试科目：生物信息学

适用专业：生物物理学

（除画图题外，所有答案都必须写在答题纸上，写在试题上及草稿纸上无效，考完后试题随答题纸交回）

密封线内不要答题

准考证号码：

报考学科、专业：

姓名：

一、填空题（25×1’ =25’）

1. 基因组作图需要得到三个图谱，它们是_____图谱、_____图谱和_____图谱。
2. NIH 维护的基因序列数据库叫_____。EMBL 核酸序列数据库由欧洲生物信息学研究所(EBI)维护的核酸序列数据构成，由_____数据库系统管理维护。日本 NGI 维护的_____和上面提到的两个数据库是最常用的核苷酸及蛋白质序列数据库，它们储存了大量的公共分子生物学信息。
3. _____是 GenBank 数据库的基本信息单位，也是最广泛地用以表示生物序列地格式之一，DDBJ flatfele 格式与其类似。EMBL 格式则每行都带有前缀，以表明本行地信息类型。所有这些格式实际上都是由更结构化地_____生成的。
4. MMDB 的数据库记录运用标准的_____，其中记录了氨基酸、核酸残基这样以聚合体形式存在，具有末端多样性的分子中所有_____、_____信息。
5. NCBI 数据模型有四个核心元素：_____，DNA 序列，蛋白质序列和_____。
6. 分子相似性是比较_____的定量尺度，分子相似性以相似性指数表示，其值在 0—1 之间，如果两个分子完全相同，其值为 _____。
7. 当前最有生命力的结构预测方法是_____和 threading 方法。
8. 在基因区域预测时，衡量一个算法的优劣的标准是_____和_____。
9. EST 是从_____（cDNA）上生成的大量很短的序列（300-500bp）。它们代表了在特定组织或发育阶段表达的基因。它们代表在给定的 cDNA 文库中的_____（有些是编码的，有些不是）。这些记录通常很少有注释，只有文库和生物来源信息。很多数据库中都有这样的记录，包括 DDBJ/EMBL/GenBank，dbEST，Unigene。
10. 序列比对的理论基础是_____。早期序列比对是全局的序列比对，但是由于蛋白质具有_____性质，所以现在用得更多得是局部得序列比对。常用

试卷编号：535

_____来描述序列两两比对。

11. 在进行序列同源比较时，常用的算法有 Smith—Waterman 算法、FASTA 和 BLAST。其中，_____可以处理碱基的替换和间隔，速度最慢，而_____的敏感性最差，但是速度最快。

二、简答题（5×8' =45'）

1. 下面的斜体部分是从 GenBank 中查出的一条记录，你从中得到什么信息？GenBank 和哪两个数据库每天都交换数据？

   LOCUS BI842584 612 bp mRNA EST 04-OCT-2001

2. 从下面 6 个名词中任意选取 4 个，对它们进行简要说明（多于 4 个的则以前 5 个为准）

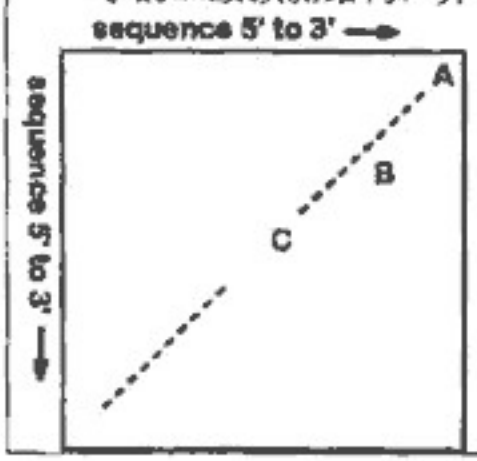
   ①local sequence alignment
   ②dynamic programming
   ③molecular clock
   ④extreme value
   ⑤gap penalty
   ⑥algorthm

3. 现有的蛋白质结构预测软件能够预测蛋白质的一些特殊结构或者结构特征（如 α 螺旋和 β 折叠），试指出其他可以预测的特殊结构和结构特征。
4. 在进行核酸功能预测的时候，需要探测的 DNA 功能位点有哪些？
5. 作为蛋白质组研究的主要工具，请简要描述双向凝胶电泳的优点。

三、问答题（3×10' =30'）

1. 在建立系统发育模型的时候，要进行一些假定，请尽可能多的指出这些假定。
2. 回答有关 RNA 结构预测的问题。

   ①在判断 RNA 分子能否形成二级结构时，第一步要做什么？
   ②用什么类型的能量值来预测 RNA 结构？如何使用？
   ③RNA 二级结构和三级结构之间的区别是什么？
   ④下面是一个 RNA 分子序列与自身的点阵图，显示的是匹配碱基。画出这个分子的二级预测结构，并标出与点阵图对应的 A、B、C 部分。



3. 用下面给出的 Blosum62 打分矩阵和公式， 回答下面的问题：

①计算下面长为 330 氨基酸的两个蛋白质序列的 log odds 值。

FWLEVEGNSMTAPTG

FWLDVQGDSMTAPAG

②长度为 330 个氨基酸的序列的期望分值是多少？

③这个分值具有显著性吗？你是如何判断的？

④出现比①中的分值更大的分值的概率是多少？

可以使用的公式：

$S = \log_2(mn)$ ，m 和 n 指的是核酸序列的长度

$P(S>x) = 1 - \exp[-Kmne^{-\lambda x}]$，其中 $K = 0.1$， $\lambda = 0.30$.

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	4																			
R	-1	5																		
N	-2	0	6																	
D	-2	-2	1	6																
C	0	-3	-3	-3	9															
Q	-1	1	0	0	-3	5														
E	-1	0	0	2	-4	2	5													
G	0	-2	0	-1	-3	-2	-2	6												
H	-2	0	1	-1	-3	0	0	-2	8											
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4										
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4									
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5								
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5							
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6						
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7					
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4				
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5			
W	-3	-3	-4	-4	2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11		
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

注：分值是以半位/分的形式给出的

计算时，请写出计算步骤，否则无分！

$\log_2 330 = 8.366$

$e^{-26.1} = 1.865 \times 10^{-9}$

$e^{(-2.031 \times 10^{-5})} = 1.00002031$