第7章 蛋白质序列分析

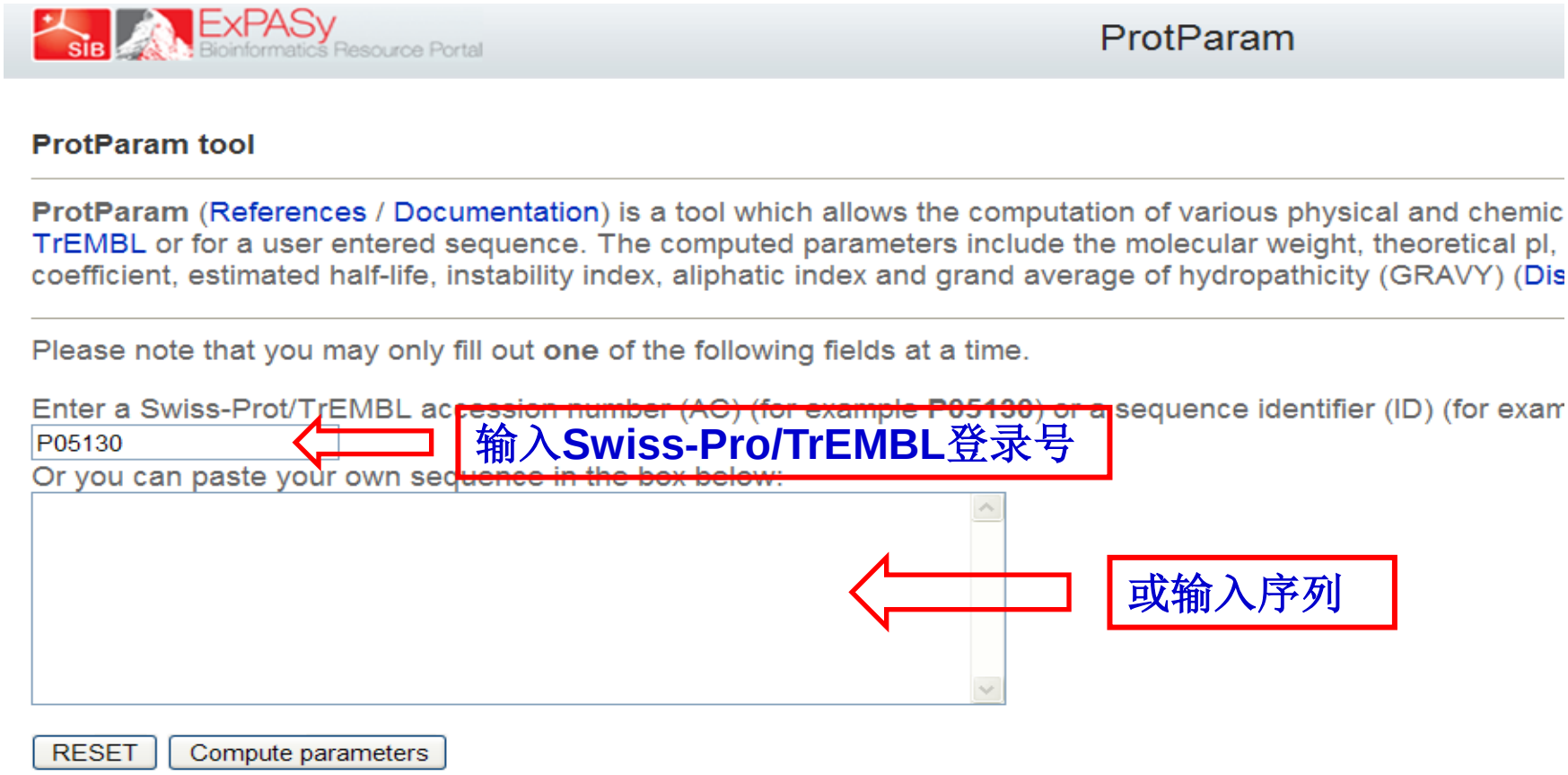
1 蛋白质序列的基本性质分析
2 蛋白质结构分析

1 蛋白质序列的基本性质分析

①理化参数分析

- 蛋白质的理化性质包括蛋白质的**分子量、分子式、氨基酸的组成、等电点、消光系数、稳定性、总平均亲水性**等。
- **ProtParam**在线分析工具：
  http://expasy.org/tools/protparam.html

输入蛋白质在Swiss-Prot/TrEMBL数据库对应的记录号或标识符：



选择全部或者感兴趣的片段：

ProtParam

Selection of endpoints on the sequence

KPC1_DROME (P05130)

Protein kinase C, brain isozyme (EC 2.7.11.13) (PKC) (dPKC53E(BR))
Drosophila melanogaster (Fruit fly).

Please select one of the following features by clicking on a pair of endpoints, and the computat
default, the complete sequence is used.
Note: Only the features corresponding to subsequences of at least 5 residues are highlighted.

FT	CHAIN	1-679	Protein kinase C, brain isozyme.
FT	DOMAIN	191-278	C2.
FT	DOMAIN	350-608	Protein kinase.
FT	DOMAIN	609-679	AGC-kinase C-terminal.
FT	ZN_FING	45-104	Phorbol-ester/DAG-type 1.
FT	ZN_FING	119-169	Phorbol-ester/DAG-type 2.
FT	NP_BIND	356-364	ATP (By similarity).

Or, if you wish to select a different sequence fragment (at least 5 amino acids long), you can e
computation will be carried out for the complete sequence).

N-terminal:
C-terminal:

The sequence KPC1_DROME consists of 679 amino acids.

RESET SUBMIT

分析结果：

Number of amino acids: 679

Molecular weight: 77694.9

Theoretical pI: 6.77

Amino acid composition: CSV format

Ala (A)	28	4.1%
Arg (R)	26	3.8%
Asn (N)	29	4.3%
Asp (D)	52	7.7%
Cys (C)	23	3.4%
Gln (Q)	28	4.1%
Glu (E)	44	6.5%
Gly (G)	48	7.1%
His (H)	16	2.4%
Ile (I)	35	5.2%
Leu (L)	51	7.5%
Lys (K)	68	10.0%
Met (M)	18	2.7%
Phe (F)	42	6.2%
Pro (P)	34	5.0%
Ser (S)	34	5.0%
Thr (T)	31	4.6%
Trp (W)	8	1.2%
Tyr (Y)	24	3.5%
Val (V)	40	5.9%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 96
Total number of positively charged residues (Arg + Lys): 94

Atomic composition:

Carbon	C	3477
Hydrogen	H	5374
Nitrogen	N	922
Oxygen	O	1018
Sulfur	S	41

分析结果

Formula: $C_{3477}H_{5374}N_{922}O_{1018}S_{41}$
Total number of atoms: 10832

Extinction coefficients:

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient	81135
Abs 0.1% (=1 g/l)	1.044, assuming all pairs of Cys residues form cystines

Ext. coefficient	79760
Abs 0.1% (=1 g/l)	1.027, assuming all Cys residues are reduced

Estimated half-life:

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 37.98
This classifies the protein as stable.

Aliphatic index: 70.60

Grand average of hydropathicity (GRAVY): -0.517

② 亲疏水性分析

- 组成蛋白质的**氨基酸其侧链组成及带电情况**有差异，导致了蛋白质的亲疏水性差异；
- 蛋白质折叠时会形成疏水内核和亲水表面，同时在潜在的跨膜区出现高疏水值区域；
- ProtScale在线分析工具：

http://expasy.org/tools/protscale.html

举例：选择氨基酸标度：

ExPASy Bioinformatics Resource Portal — ProtScale

ProtScale

ProtScale [Reference / Documentation] allows you to compute and represent the profile produced by any amino a

An **amino acid scale** is defined by a numerical value assigned to each type of amino acid. The most frequently us
scales and the secondary structure conformational parameters scales, but many other scales exist which are base
amino acids. This program provides 57 predefined scales entered from the literature.
Enter a UniProtKB/Swiss-Prot or UniProtKB/TrEMBL accession number (AC) (e.g. **P05130**) or a sequence identifie

Or you can paste your own sequence in the box below:

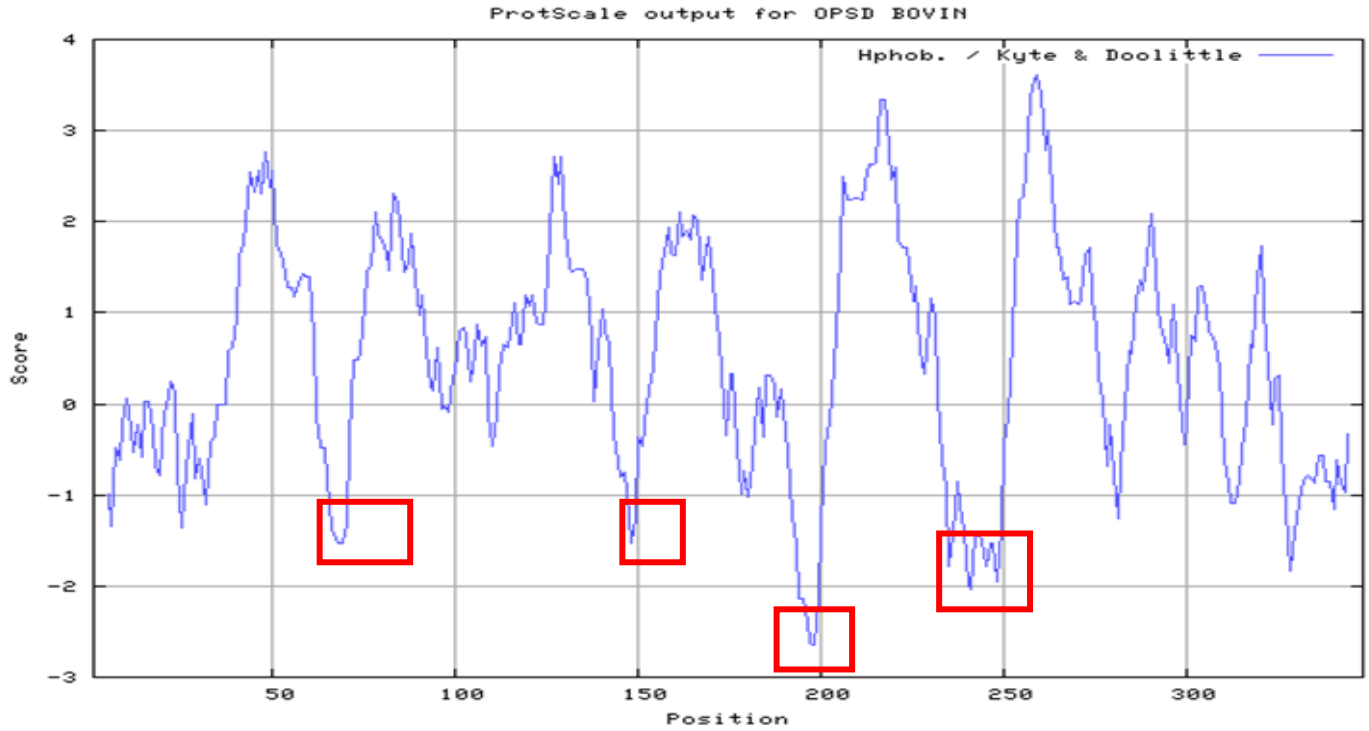
Please choose an amino acid scale from the following list. To display information about a scale (author, reference,

- ○ Molecular weight
- ○ Bulkiness
- ○ Polarity / Grantham
- ○ Recognition factors
- ○ Hphob. OMH / Sweet et al.
- ◉ Hphob. / Kyte & Doolittle
- ○ Hphob. / Abraham & Leo
- ○ Hphob. / Bull & Breese
- ○ Hphob. / Guy
- ○ Hphob. / Miyazawa et al.
- ○ Hphob. / Roseman
- ○ Number of codon(s)
- ○ Polarity / Zimmerman
- ○ Refractivity
- ○ Hphob. / Eisenberg et al.
- ○ Hphob. / Hopp & Woods
- ○ Hphob. / Manavalan et al.
- ○ Hphob. / Black
- ○ Hphob. / Fauchere et al.
- ○ Hphob. / Janin
- ○ Hphob. / Rao & Argos
- ○ Hphob. / Tanford

Window size: 9
Relative weight of the window edges compared to the window center (in %): 100
Weight variation model (if the relative weight at the edges is < 100%): ◉ linear ○ exponential
Do you want to normalize the scale from 0 to 1? ○ yes ◉ no

If you need more information about how to set these parameters, please click here.

Submit Reset

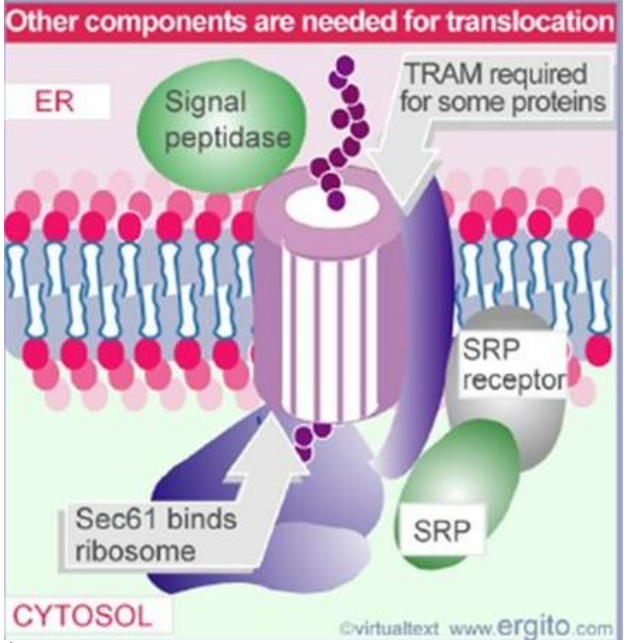


The results of your ProtScale query are available in the following formats:

- Image in GIF-format
- Image in Postscript-format
- Numerical format (verbose)
- Numerical format (minimal, to be exported into an external application)

牛视紫红质蛋白亲疏水性分布图，4个主要亲水分布区以方框标示，横坐标为序列位置，纵坐标为氨基酸的标度值，KD标度定义疏水性氨基酸有较高的打分(>0 标示疏水，<0 表示亲水性)

③ 信号肽预测

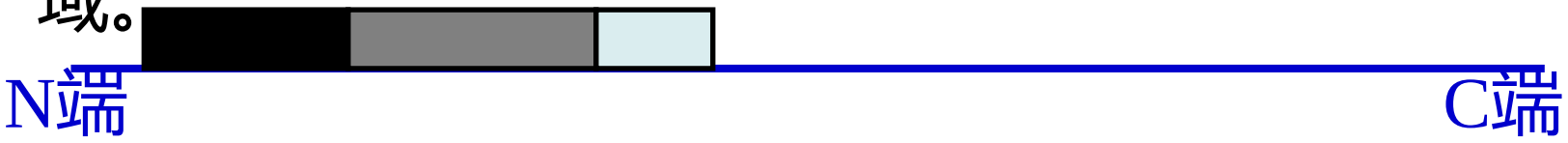


- 信号肽基本长度约为15-35个氨基酸，主要由三个区域组成：

N-region: 正电荷区域，至少含有1个精氨酸（R）或赖氨酸（K）

H-region:疏水核，一般长12-14个氨基酸

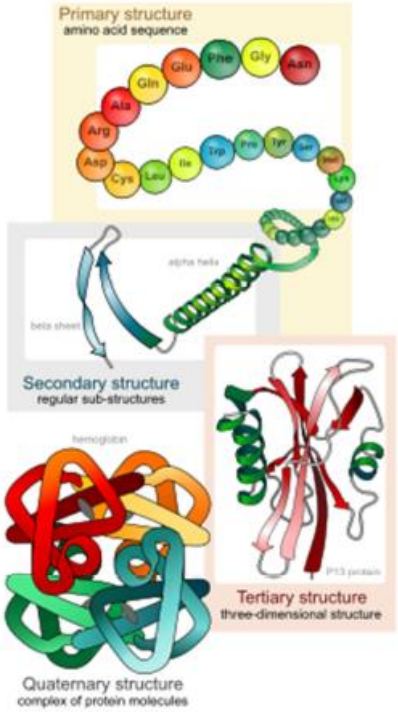
C-region:包含信号肽酶的剪切位点，在剪切位点-1和-3位上多为中性的丙氨酸，故该区域也称为富含丙氨酸区域。

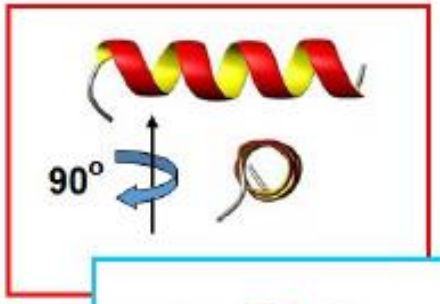


- SignalP在线分析工具：

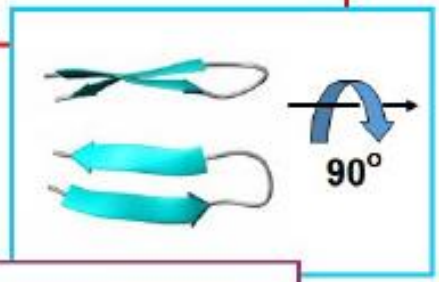
http://genome.cbs.dtu.dk/services/SignalP/

2 蛋白质结构分析

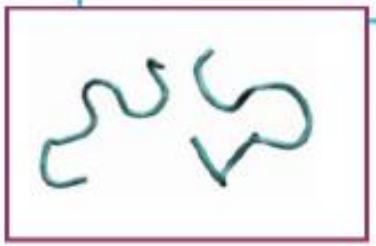




螺旋：最常见的就是α螺旋。



β折叠（β sheet）：β折叠由β折片（β strand）平行排列而成。



无规卷曲 (coil)：无规律松散结构。

β转角（turn）：如果肽链发生了急转弯（角度大于90°），这个转弯结构叫β转角。

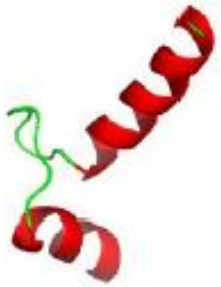
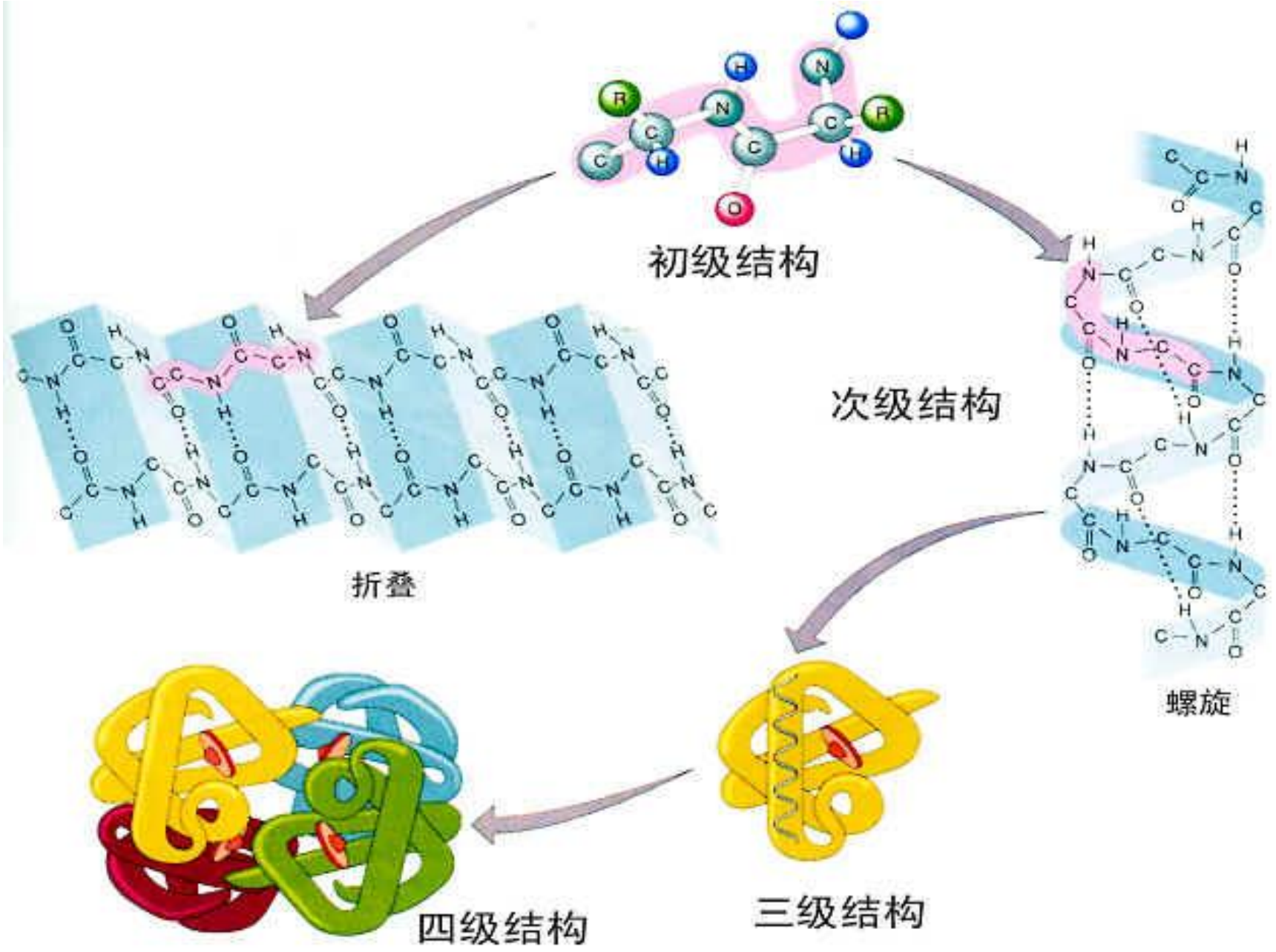


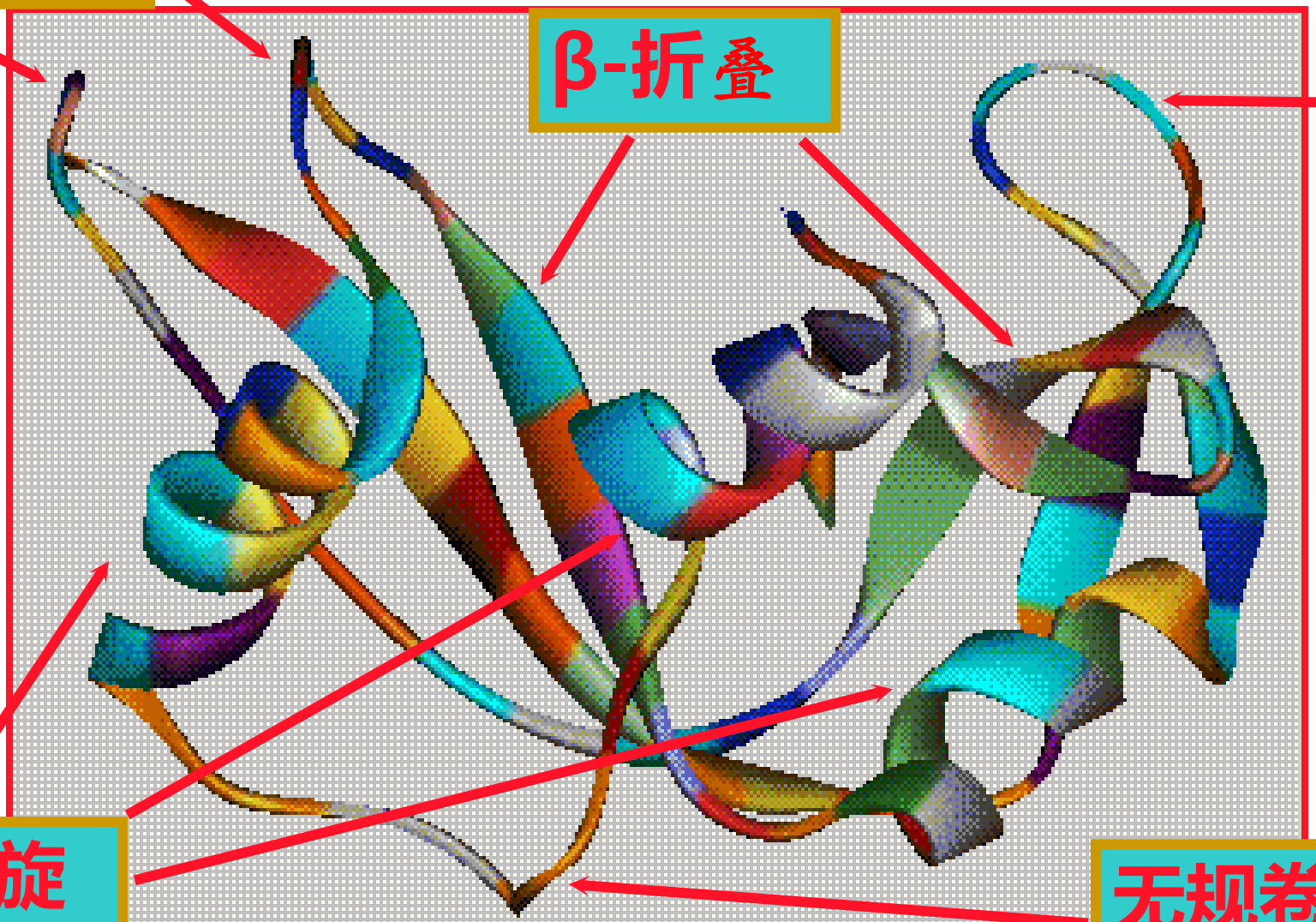
图 1. 蛋白质的二级结构


初级结构
次级结构
折叠
螺旋
三级结构
四级结构

蛋白质二级结构的主要形式

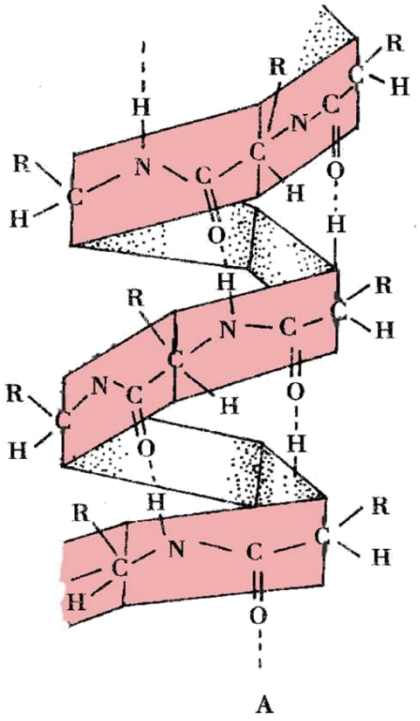
- α-螺旋 (α -helix)
- β-折叠 (β-pleated sheet)
- β-转角 （ β-turn)
- 无规则卷曲 (random coil)

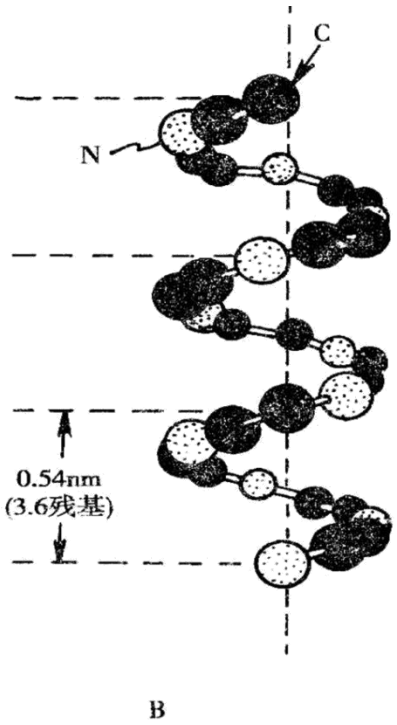
RNA酶的分子结构

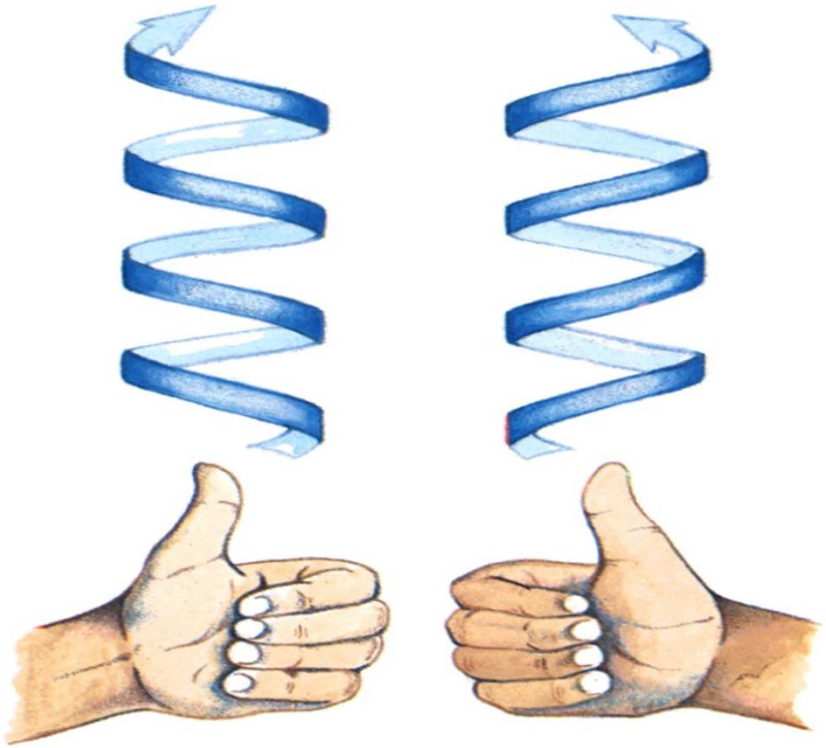


α-螺旋

- 指肽链主链骨架围绕中心轴盘绕折叠所形成的有规则的结构。
- 大多数蛋白质中的α-螺旋为右手螺旋。





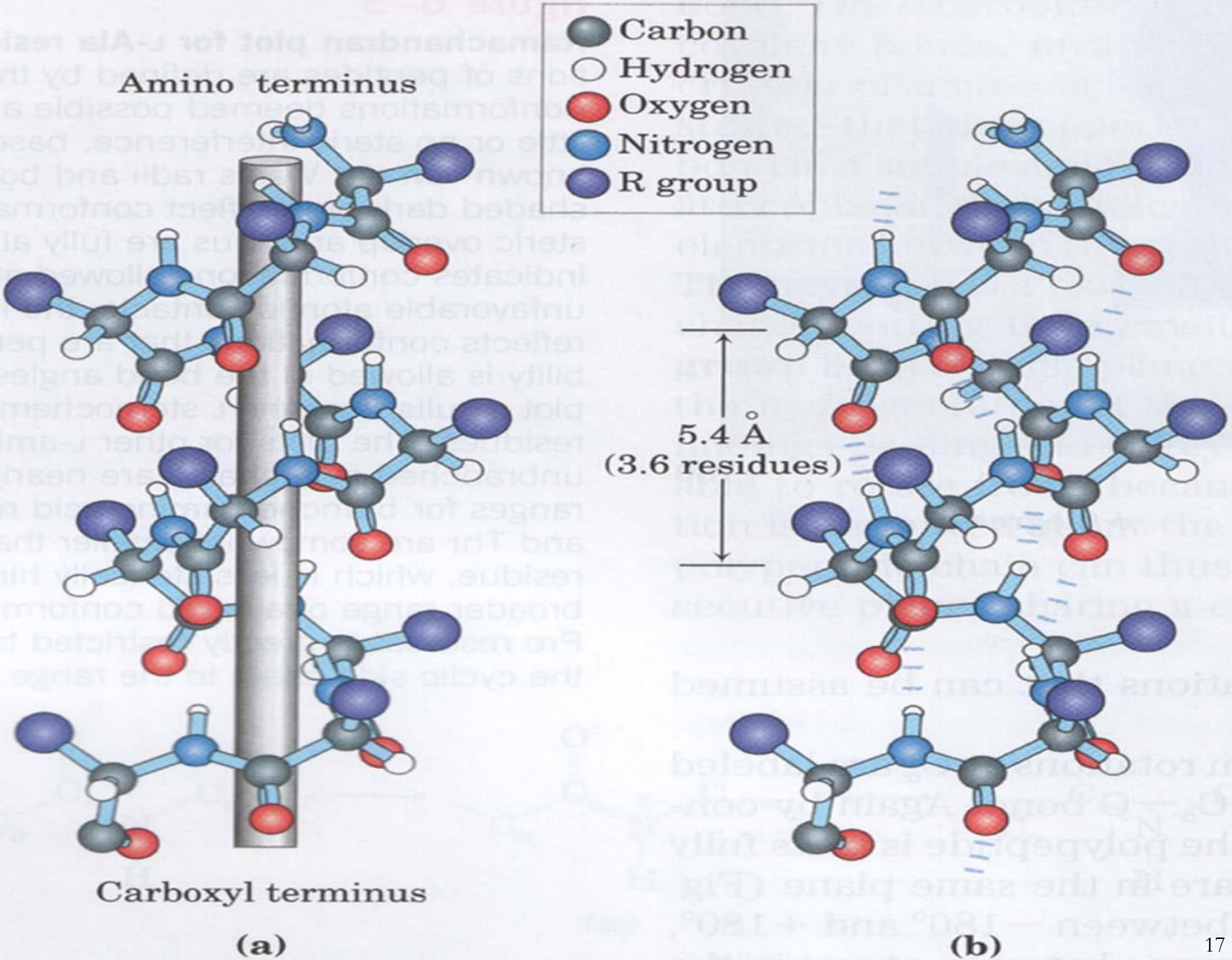


α-螺旋

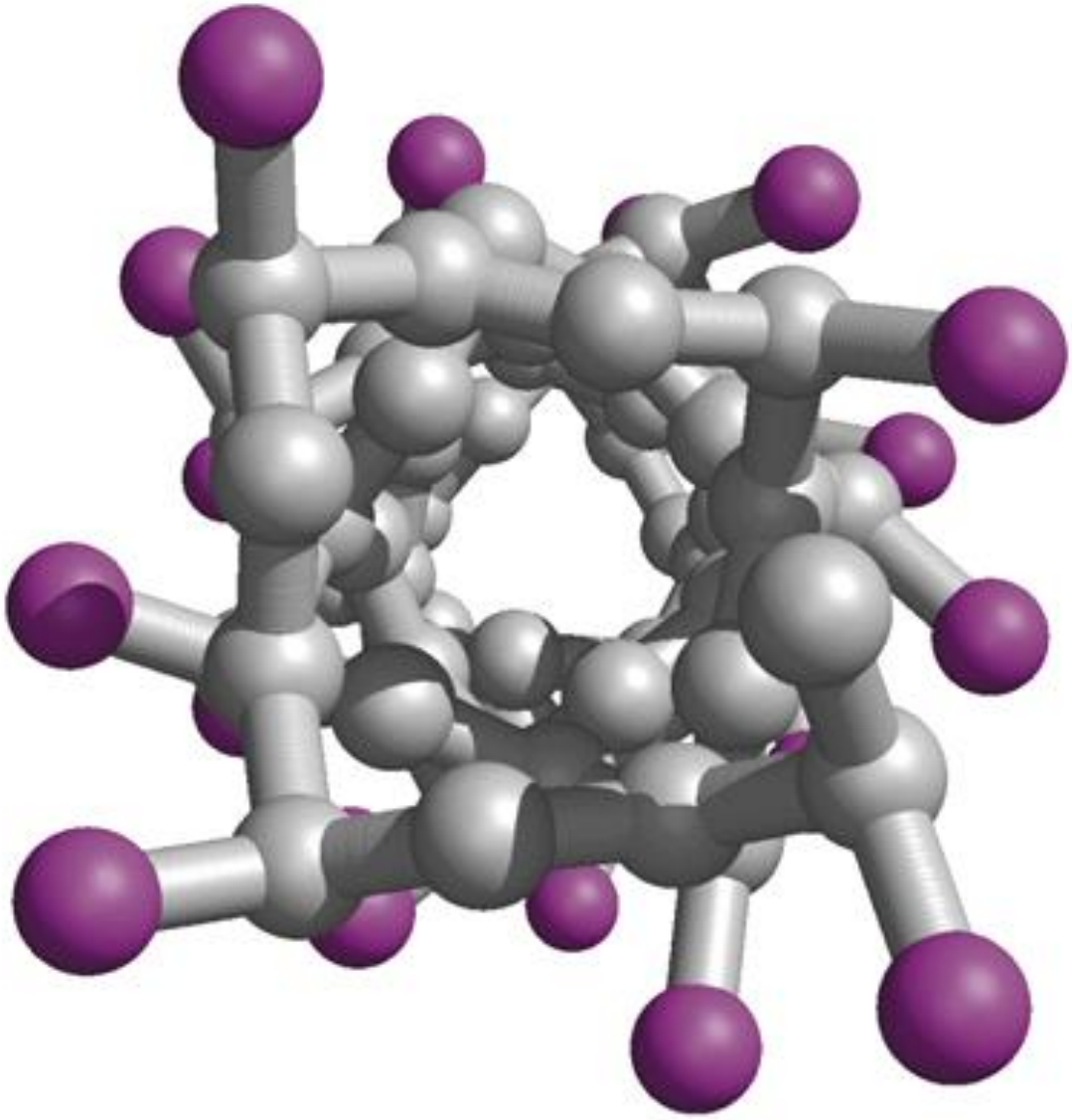
结构特征：

- 类似棒状结构，紧密卷曲的多肽链构成棒的中心部 分，侧链R伸出到螺旋的外面，完成一个螺旋需3.6个氨基酸残基，螺旋每上升一圈，螺距为0.54nm，相邻两个氨基酸残基之间的轴心距为0.15nm；

- α-螺旋结构的稳定主要靠链内氢键，氢键形成于第一个氨基酸残基的羧基与线性顺序中的第四个氨基酸的氨基，氢键环内包含13个原子


Carbon
Hydrogen
Oxygen
Nitrogen
R group
Amino terminus
Carboxyl terminus
5.4 Å
(3.6 residues)
(a)
(b)

α-螺旋顶面观



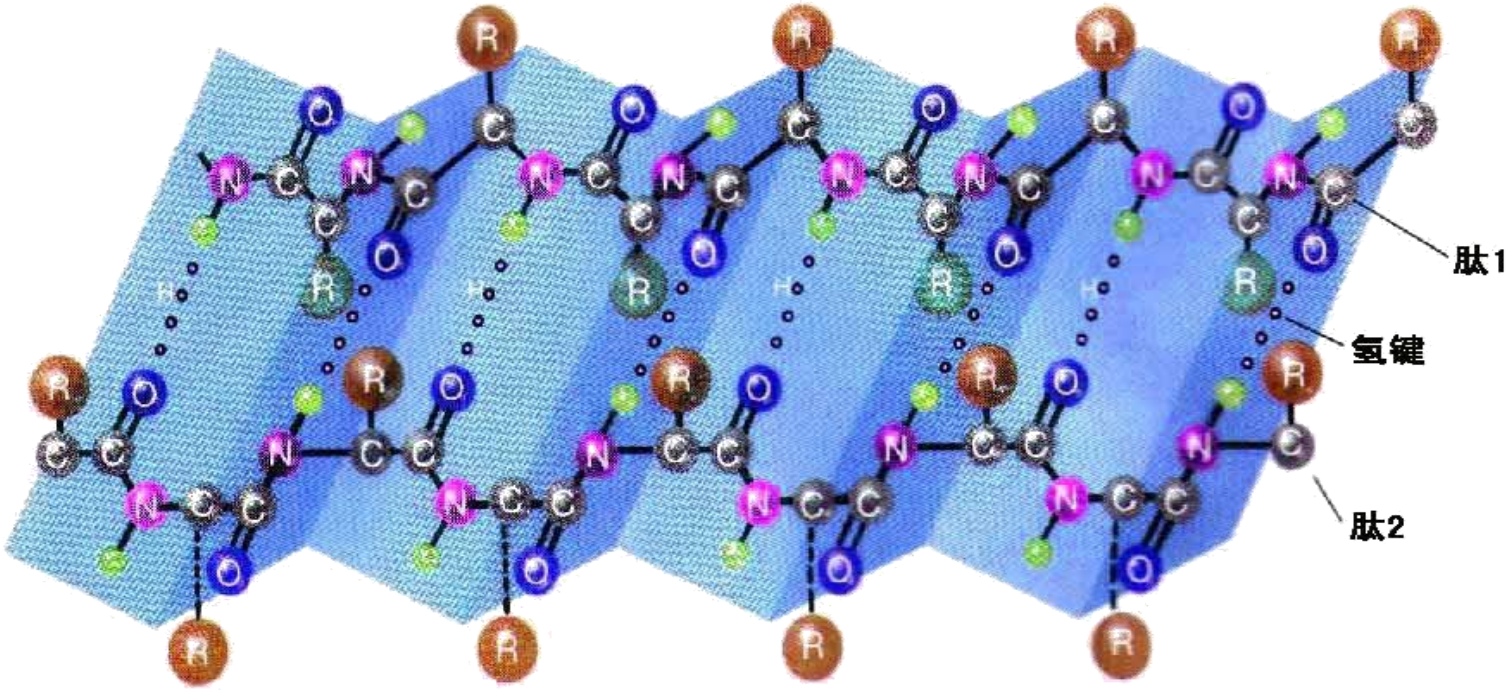
(c)

- 有些结构不易形成α-螺旋

  - ✓Gly和Pro不易形成α-螺旋；
  - ✓多聚大侧链氨基酸；
  - ✓多聚相同电荷氨基酸。

β-折叠

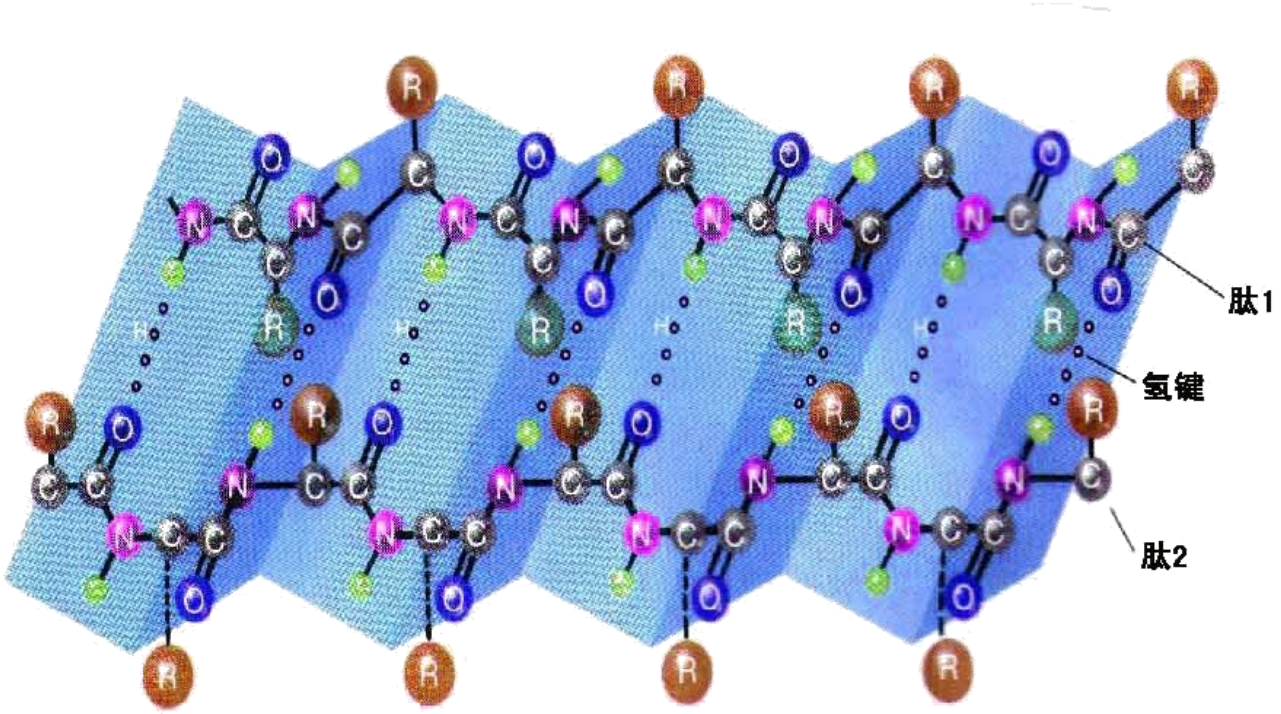
- 由两条或两条以上几乎完全伸展的肽链平行排列，通过链间氢键交联而成
- 肽链的主链成锯齿状折叠构象



β-折叠

结构特征：

①在β-折叠中，氨基酸的α-碳原子总是处于折叠线上，侧链都垂直于折叠片平面，交替地分布在片状平面的上面或下面，以避免R基团的空间障碍；

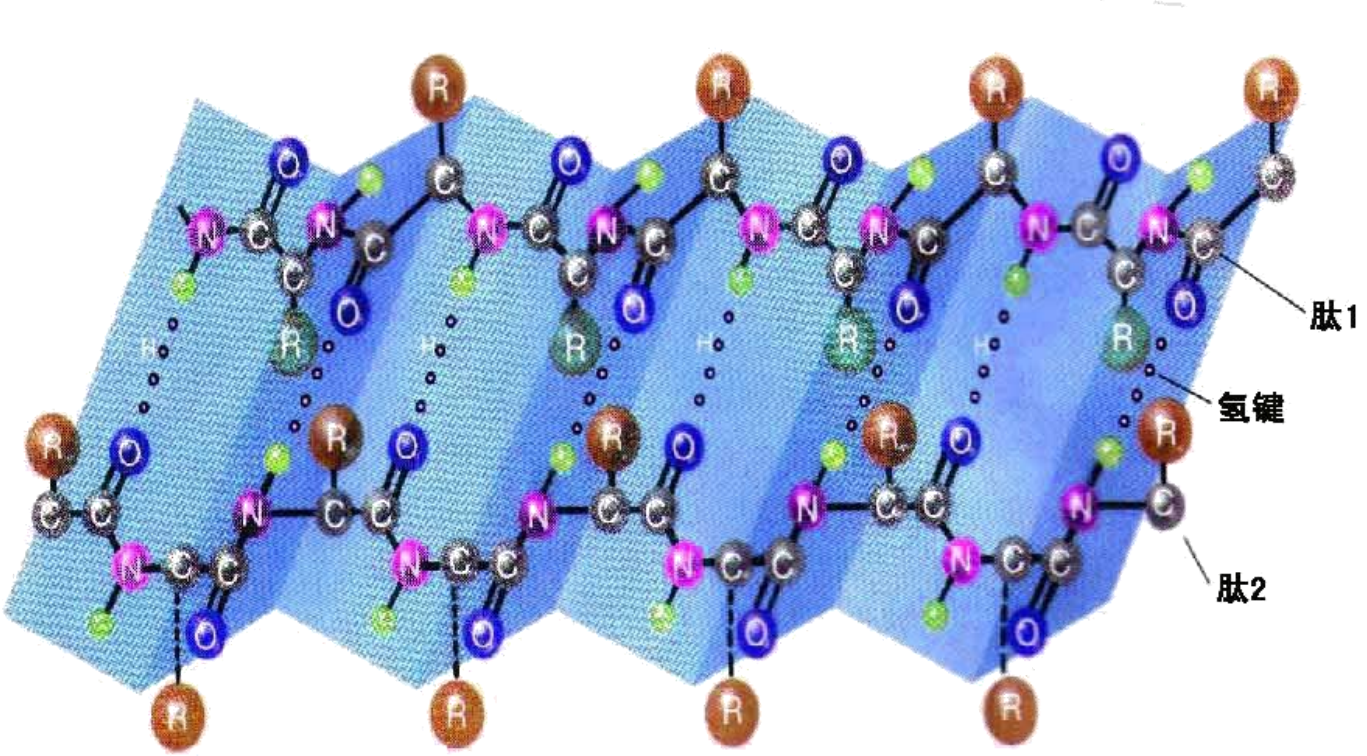


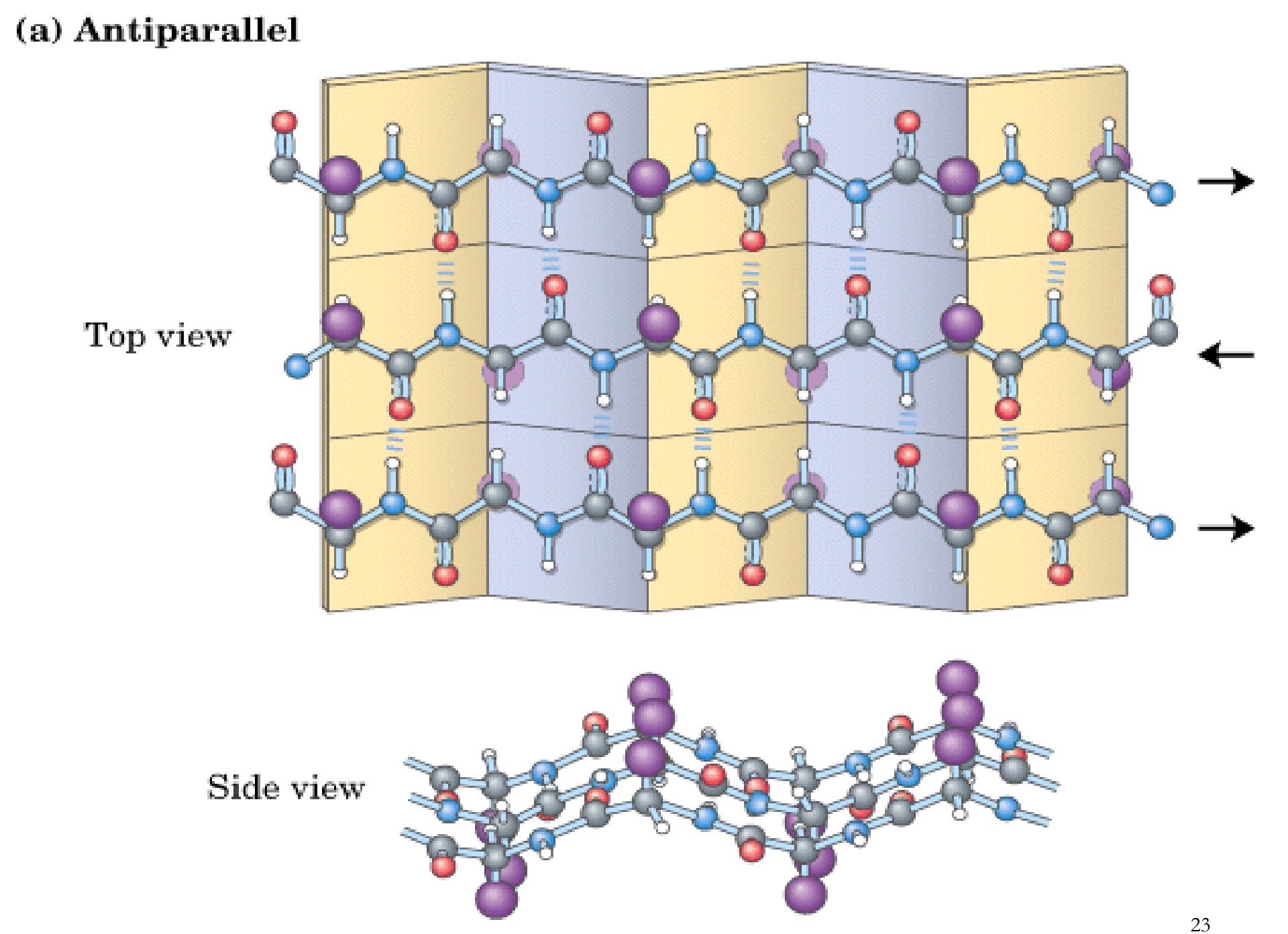
β-折叠

结构特征：

② 几乎所有肽键都参与链间氢键的交联，氢键与链的长轴接近垂直。

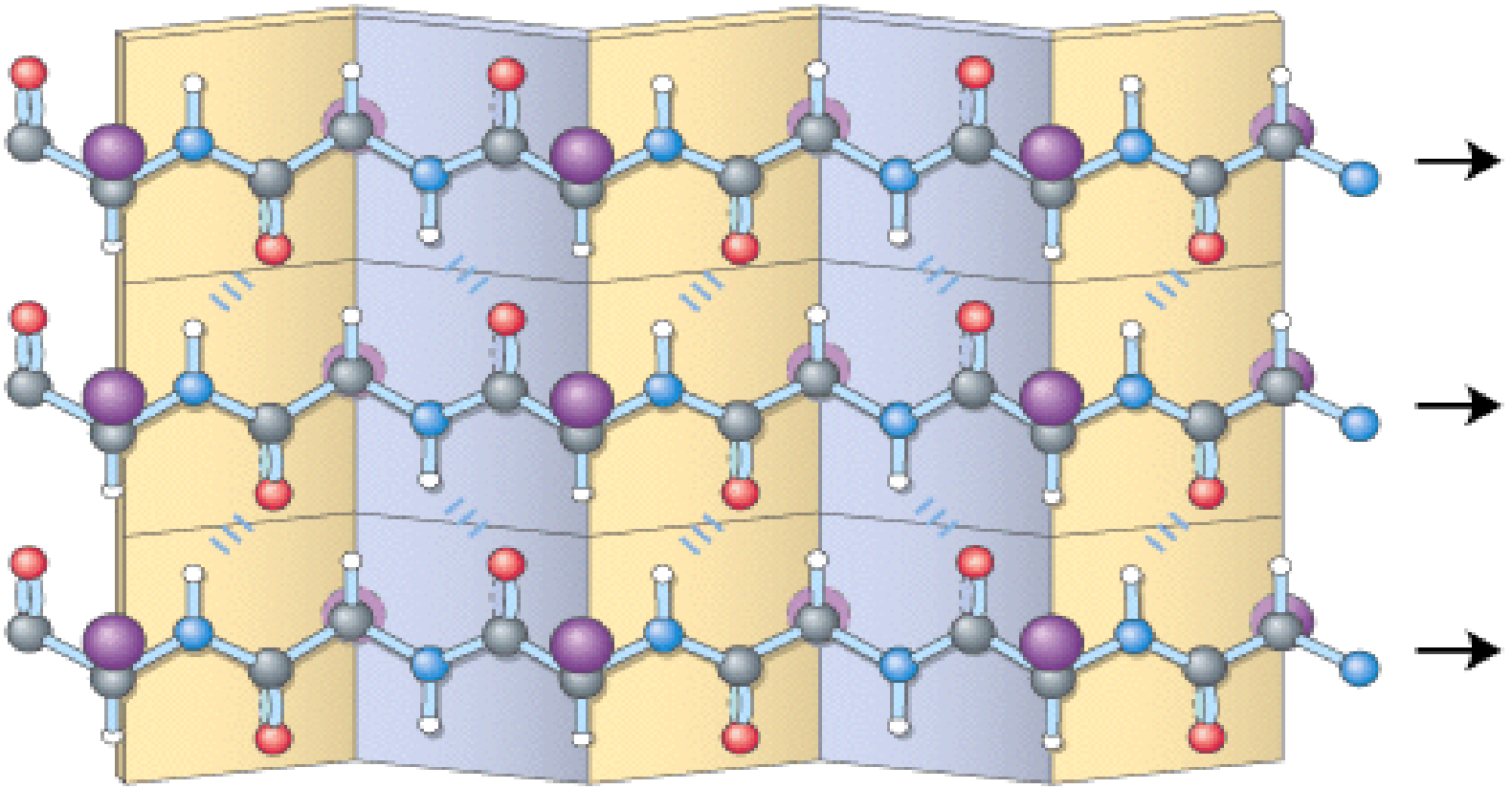
③ β-折叠有两种类型。一种为平行式，即所有肽链的N-端都在同一边。另一种为反平行式，即相邻两条肽链的方向相反。



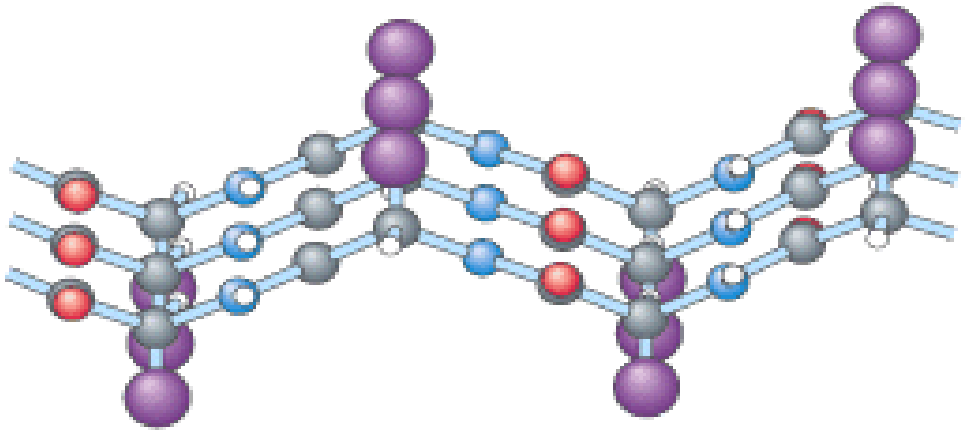

(a) Antiparallel
Top view
Side view

(b) Parallel

Top view

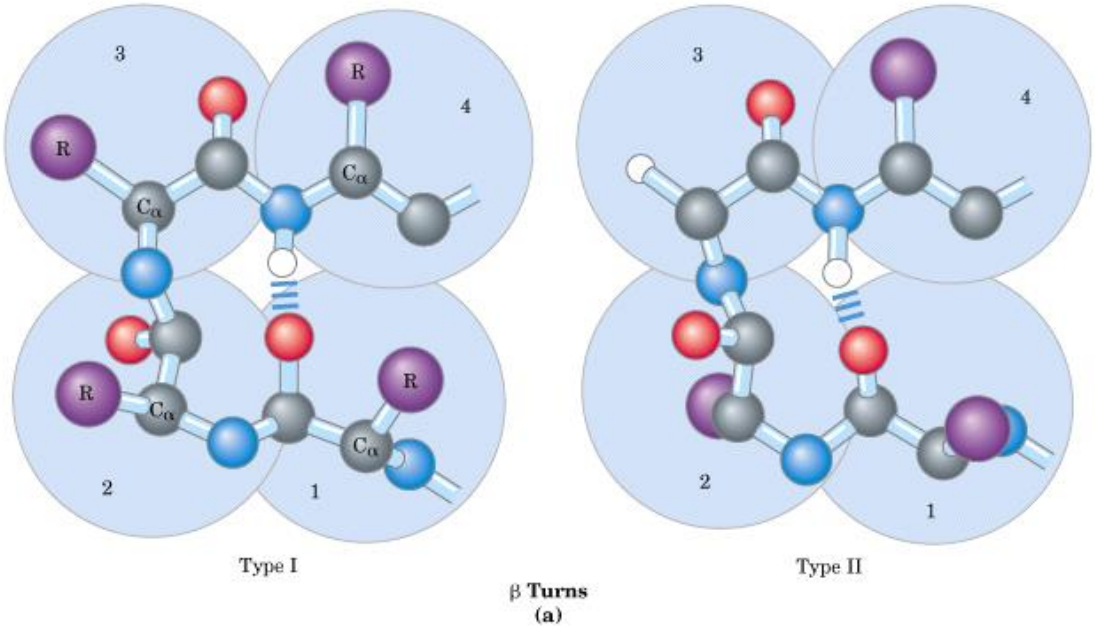


Side view

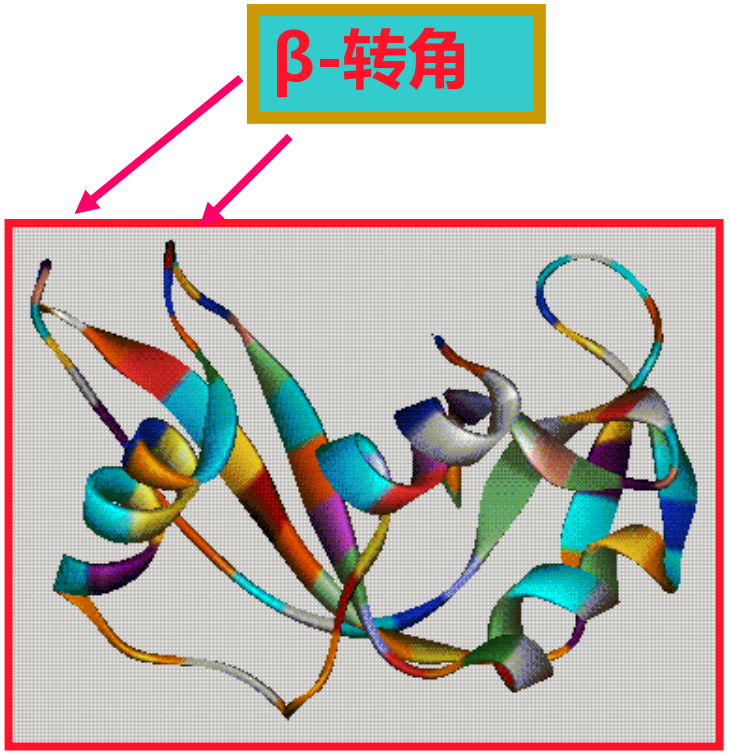


β-转角

- 也称β-回折或发夹结构，存在于球状蛋白中。
- 在β-转角部分，由四个氨基酸残基组成;
- 弯曲处的第一个氨基酸残基的 C=O 和第四个残基的N-H 之间形成氢键，形成一个不很稳定的环状结构。



β Turns
(a)

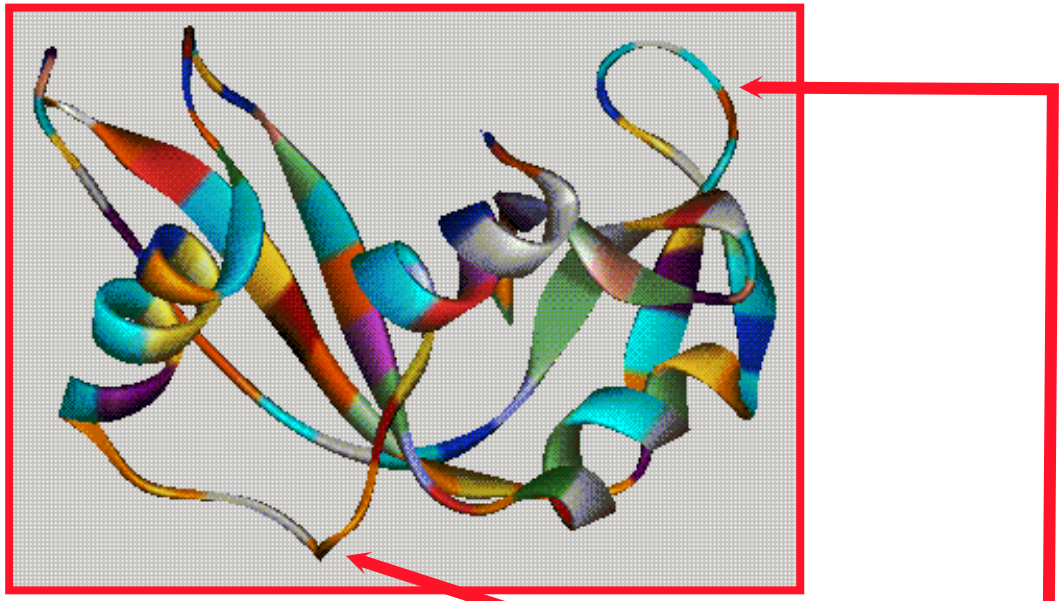


无规则卷曲

- 指没有明确规律性的肽链构象，但仍然是紧密有序的稳定结构，可通过主链间氢键，甚至在主链与侧链间形成氢键而维持其构象。
- 无规则卷曲的类型大体上可以分为两大类

（1）紧密环

（2）连接条带



① 蛋白质结构数据库

- 蛋白质晶体结构数据库（**PDB**）
- 蛋白质结构分类数据库（**SCOP**）
- 蛋白质二级结构数据库（**DSSP**）
- 蛋白质序列模式数据库（**PROSITE**）
- 结构域数据库（**CATH**、**Dali**、 **FSSP**）

CATH：收录蛋白质结构域的数据库，分四个等级，

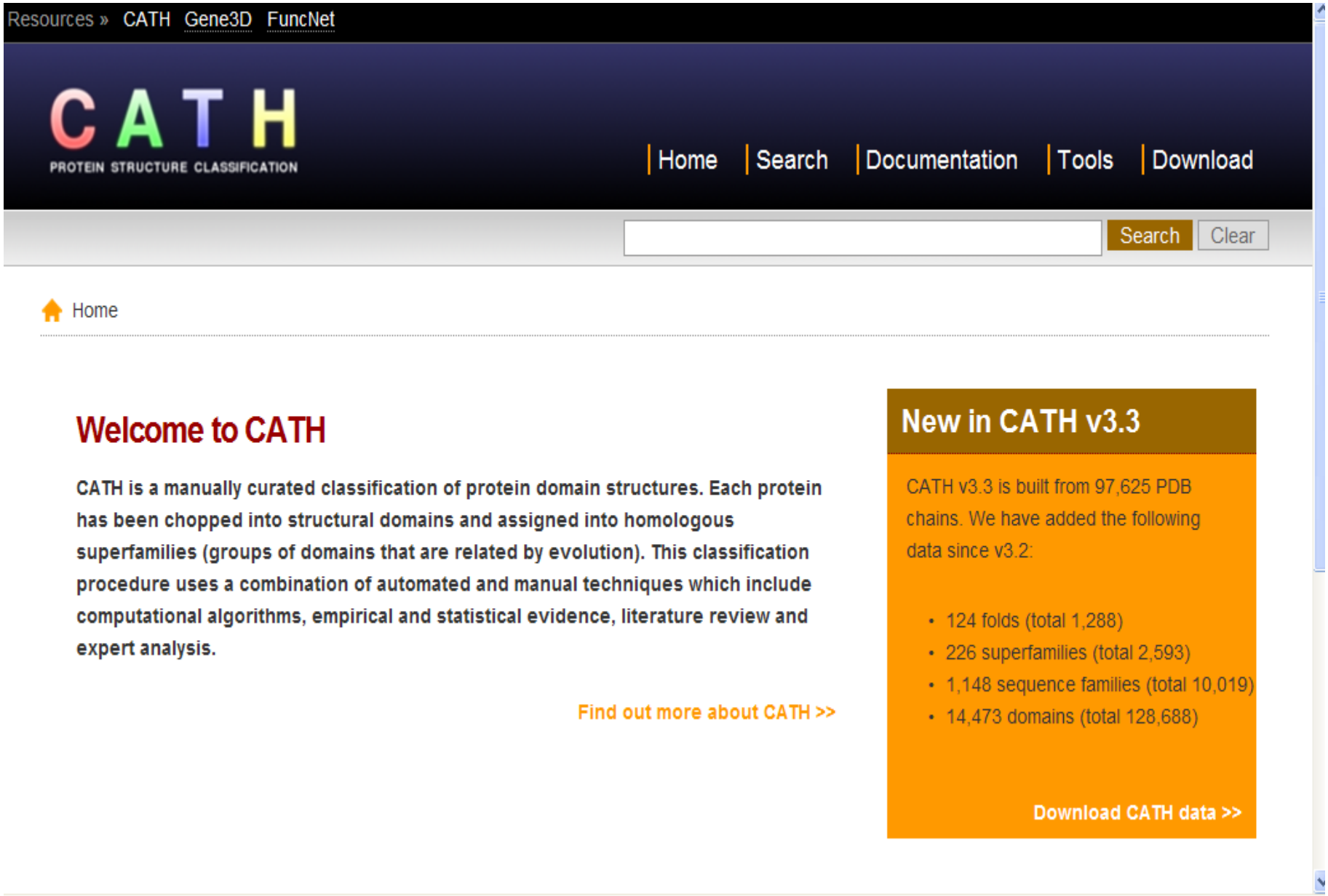
C（Class，类）：按结构域中二级结构的组成与包装方式进行划分；

A（Architecture，构架）：按二级结构的排列方向进行划分，但忽略其连接方式；

T（Topology，拓扑）：按二级结构的总体外观及连接方式进行划分；

H （Homologous Superfamily，同源超家族）：在进化上亲缘关系较近的类群。

- http://biochem.ucl.ac.uk/cath



② 蛋白质结构显示软件

- Swiss-PdbViewer：分子建模和可视化工具
- RasMol: 蛋白质结构数据库PDB的文件格式（.pdb）为默认的可以显示的格式。
- Chime：用于显示网页中的生物大分子三维结构。
- Cn3D：只能显示在NCBI中的MMDB数据库格式的生物大分子空间结构。

③ 蛋白质结构的实验测定方法

- **X射线晶体衍射**
  - 培养并挑选合适的蛋白质结晶；
  - 测定x射线衍射点的距离来确定晶胞轴长；
  - 测定衍射强度计算结构振幅；
  - 衍射数据的测量和处理；
  - 相位的计算；
  - 电子密度图的计算和解释；
  - 确定原子在晶胞中的坐标；
  - 蛋白质结构的修正
- **核磁共振技术**
- 电子显微镜
- 其他方法：中子衍射、紫外光谱法、红外光谱法、拉曼光谱法等。

④ 蛋白质二级结构预测算法

➢ 点模式方法

步骤：

a) 给出氨基酸残基的亲疏水性。

b) 采用八个残基确定二级结构的间隔方式，用二进制数字表示，疏水残基用1表示，亲水残基用0表示，两性残基用1或0表示。

c) 把这个有8个数位的二进制数按“左边是低位，右边是高位”的原则化成十进制数字。

d) 根据十进制结果确定八残基片段的二级结构。

e) 如果含有两性残基，取1和0各算一次，得出的结构不同的话，则以α-螺旋优先，β-折叠其次，最后才是无规则卷曲。

- 残基的亲疏水性：

9种疏水残基：C I L F M V W H Y

9种亲水残基：K E R S Q D N P T

两性残基：A G

- 计算结果确定：

二级结构	十进制数字
α-螺旋	9、12、13、17、18、19、25、27、29、31、34、36、38、44、45、46、47、50、51、54、55、59、61、62、77、201、205、217、219、237
β-折叠	连续的1或交替的01（或10）构成
无规则卷曲	不属于以上两种情况

举例

利用点模式法预测氨基酸序列LKMVRSCH**的二级结构。**

LKMVRSCH

↓

10110011

↓

$$1\times2^0+0\times2^1+1\times2^2+1\times2^3+0\times2^4+0\times2^5+1\times2^6+1\times2^7$$

$$=1+0+4+8+0+0+64+128=205$$

α-螺旋

➢Chou-Fasman方法

Chou和Fasman（1978）利用统计方法得出每种氨基酸形成α-螺旋、β-折叠、β-转角的**二级结构倾向性因子**，再通过分析待测蛋白质序列中各氨基酸的倾向性因子，最后获得该序列的二级结构，这种方法称为Chou-Fasman方法。

二级结构倾向性因子的计算方法为：

$$P_i=A_i/T_i$$

*其中，i为二级结构，包括α-螺旋（α）、β-折叠（β）、无规则卷曲（c）、和β-转角（t），A_i表示第A种残基处于结构i的比例，T_i是所有被统计残基处于结构i的比例

*另外，每个氨基酸同时也有4个转角参数***f（i）**，代表这种氨基酸出现在转角第一至四位的频率

20种氨基酸的相关信息

氨基酸	残基总数(N)	α-螺旋			β-折叠			β-转角		
		Nα	fα (Nα/N)	Pα(fα /<fα>)	Nβ	fβ (Nβ/N)	Pβ(fβ /<f>)	Nt	ft (Nt/N)	Pt(ft/ <ft>)
Ala	434	234	0.54	1.42	71	0.16	0.83	85	0.20	0.66
Arg	142	53	0.37	0.98	26	0.18	0.93	40	0.28	0.95
Asn	230	58	0.25	0.66	40	0.17	0.89	106	0.46	1.56
Asp	273	105	0.39	1.01	29	0.11	0.54	118	0.43	1.46
Cys	94	25	0.27	0.70	22	0.23	1.19	33	0.35	1.19
Gln	162	68	0.42	1.11	35	0.22	1.10	47	0.29	0.98
Glu	234	134	0.57	1.51	17	0.07	0.37	51	0.22	0.74
Gly	422	91	0.22	0.57	62	0.15	0.75	194	0.46	1.56
His	129	49	0.38	1.00	22	0.17	0.87	36	0.28	0.95
Ile	233	95	0.41	1.08	73	0.31	1.60	32	0.14	0.47
Leu	358	164	0.46	1.21	91	0.25	1.30	62	0.17	0.59
Lys	347	153	0.44	1.16	50	0.14	0.74	103	0.30	1.01
Met	73	40	0.55	1.44	15	0.21	1.05	13	0.18	0.60
Phe	170	73	0.43	1.13	46	0.27	1.38	30	0.18	0.60
Pro	176	38	0.22	0.57	19	0.11	0.55	79	0.45	1.52
Ser	367	107	0.29	0.77	54	0.15	0.75	155	0.42	1.43
Thr	278	87	0.31	0.83	65	0.23	1.19	79	0.28	0.96
Trp	78	32	0.41	1.08	21	0.27	1.37	22	0.28	0.96
Tyr	184	48	0.26	0.69	53	0.29	1.47	62	0.34	1.14
Val	357	144	0.40	1.06	119	0.33	1.70	53	0.15	0.50
Total	4741		1798			930			1400	
<f>	1.00		<fα> = 0.38			<fβ> = 0.20		<ft> =0.30		

氨基酸的二级结构倾向性因子

氨基酸名称	氨基酸代码	P_α	P_β	P_t	f_i	f_{i+1}	f_{i+2}	f_{i+3}	P_t
丙氨酸	A	142	83	66	0.060	0.076	0.035	0.058	70
半胱氨酸	C	70	119	119	0.149	0.050	0.117	0.128	118
天冬氨酸	D	101	54	146	0.147	0.110	0.179	0.081	120
谷氨酸	E	151	37	74	0.056	0.060	0.077	0.064	84
苯丙氨酸	F	113	138	60	0.059	0.041	0.065	0.065	71
甘氨酸	G	57	75	156	0.102	0.085	0.190	0.152	150
组氨酸	H	100	87	95	0.140	0.047	0.093	0.054	106
异亮氨酸	I	108	160	47	0.043	0.034	0.013	0.056	66
赖氨酸	K	114	74	101	0.055	0.115	0.072	0.095	98
亮氨酸	L	121	130	59	0.061	0.025	0.036	0.070	68
甲硫氨酸	M	145	105	60	0.068	0.082	0.014	0.055	58
天冬酰胺	N	67	89	156	0.161	0.083	0.191	0.091	135
脯氨酸	P	57	55	152	0.102	0.301	0.034	0.068	159
谷氨酰胺	Q	111	110	98	0.074	0.098	0.037	0.098	86
精氨酸	R	98	93	95	0.070	0.106	0.099	0.085	104
丝氨酸	S	77	75	143	0.120	0.139	0.125	0.106	132
苏氨酸	T	83	119	96	0.086	0.108	0.065	0.079	107
缬氨酸	V	106	170	50	0.062	0.048	0.028	0.053	62
色氨酸	W	108	137	96	0.077	0.013	0.064	0.167	75
酪氨酸	Y	69	147	114	0.082	0.065	0.114	0.125	106

资料来源 Chou 和 Fasman(1978)的文献

*Chou-Fasman方法使用较为简便，其预测准确率为50%-60%。

预测规则：

α-螺旋规则：先找α-螺旋核，寻找相邻6个残基中至少4个的Pα≥100的区域；再进行螺旋延伸，直至末端连续4残基P_α平均值小于100，并计算得到序列片段中所有氨基酸的P_α总和及P_β总和，如果所得序列长度大于6个残基且$\sum P_\alpha > \sum P_\beta$，则该片段的二级结构为α-螺旋

β-折叠规则：先找β-折叠核，寻找相邻6个残基中至少4个的$P_\beta \geq 100$的区域；再进行延伸，直至末端连续4残基P_β平均值小于100，并计算得到序列片段中所有氨基酸的P_α总和及P_β总和，如果所得序列的所有氨基酸的P_β的平均值大于105且$\sum P_\alpha < \sum P_\beta$，则该片段的二级结构为β-折叠

转角规则：寻找连续4残基，若这四个残基能满足以下三个条件，则可以预测这样连续4残基形成转角：$f_i \times f_{i+1} \times f_{i+2} \times f_{i+3}$大于0.000075；4残基的$P_t$平均值大于100；$\sum P_t > \sum P_\alpha$和$\sum P_t > \sum P_\beta$

重叠规则：比较重叠区域的P_α平均值和P_β平均值，若重叠区域的P_α平均值 > P_β平均值，则该重叠区域为α-螺旋，反之为β-折叠

举例

利用Chou-Fasman算法，预测下面序列中α-螺旋区域、β-折叠区域和发夹转角区域：

CAENKLDHVADCCILFMTWYNDGPCIFIYDNGP

重叠规则： P_α平均值 > P_β平均值，则该重叠区域为α-螺旋，反之为β-折叠

	C	A	E	N	K	L	D	H	V	A	D	C	C	I	L	F	M	T	W	Y	N	D	G	P	C
Pα	70	142	151	67	114	121	101	100	106	142	101	70	70	108	121	113	145	83	108	69	67	101	57	57	70
Pβ	119	83	37	89	74	130	54	87	170	83	54	119	119	160	130	138	105	119	137	147	89	54	75	55	119
Pt	119	66	74	156	101	59	146	95	50	66	146	119	119	47	59	60	60	96	96	114	156	146	156	152	119
fi	0.149	0.060	0.056	0.161	0.055	0.061	0.147	0.140	0.062	0.060	0.147	0.149	0.149	0.043	0.061	0.059	0.068	0.086	0.077	0.082	0.161	0.147	0.102	0.102	0.149
fi+1	0.050	0.076	0.060	0.083	0.115	0.025	0.110	0.047	0.048	0.076	0.110	0.050	0.050	0.034	0.025	0.041	0.082	0.108	0.013	0.065	0.083	0.110	0.085	0.301	0.050
fi+2	0.117	0.035	0.077	0.191	0.072	0.036	0.179	0.093	0.028	0.035	0.179	0.117	0.117	0.013	0.036	0.065	0.014	0.065	0.064	0.114	0.191	0.179	0.190	0.034	0.117
fi+3	0.128	0.058	0.064	0.091	0.095	0.070	0.081	0.054	0.053	0.058	0.081	0.128	0.128	0.056	0.070	0.065	0.055	0.079	0.167	0.125	0.091	0.081	0.152	0.068	0.128
Pc	118	70	84	135	98	68	120	106	62	70	120	118	118	66	68	71	58	107	75	106	135	120	150	159	118
α	70	142	151	67	114	121	101	100	106	142	101	70	70	108	121	113	145	83	108	69	67	101	57	57	70
β	119	83	37	89	74	130	54	87	170	83	54	119	119	160	130	138	105	119	137	147	89	54	75	55	119
t				8E-05	7E-05	2E-05	5E-05	1E-05	3E-05	1E-05	2E-05	1E-04	1E-04	5E-05	7E-06	1E-05	4E-06	3E-06	5E-05	6E-05	1E-05	8E-05	2E-04	2E-04	5E-05
													1									1	1	1	
													383									345	294	282	
													375									427	365	273	
													450									512	572	610	
	C	A	E	N	K	L	D	H	V	A	D	C	C	I	L	F	M	T	W	Y	N	D	G	P	C
	H	H	H	H	H	H	H	H	H	H	E	E	E	E	E	E	E	E	E	E	T	T	T	T	E

寻找相邻6个残基中至少4个的Pα≥100的区域；再进行螺旋延伸，直至末端连续4残基P_α平均值小于100，并计算得到序列长度大于6个残基且$\sum P_\alpha > \sum P_\beta$

寻找相邻6个残基中至少4个的$P_\beta \geq 100$的区域；再进行延伸，直至末端连续4残基P_β平均值小于100，并计算得到序列的所有氨基酸的P_β的平均值大于105且$\sum P_\alpha < \sum P_\beta$

连续4残基f_i乘积大于0.000075；且平均值大于100；$\sum P_t > \sum P_\alpha$和$\sum P_t > \sum P_\beta$

➢GOR方法

- 由Garnier, Osguthorpe和Robson（1978）研制的基于信息论原理的蛋白质二级结构预测方法，名称是三位作者名字的首字母缩写。目前，一般使用该法的第四版，即GORIV。
- 预测某一氨基酸残基处于哪种二级结构时，同时要考虑与其相邻的左右各8个氨基酸残基，计算17个残基的信息差值之和，取信息差值之和最大的二级结构作为该氨基酸的二级结构。
- GOR方法预测蛋白质二级结构的准确率约为65%

GOR方法涉及到的公式：

信息的计算式子 $I(S;A) = \log\frac{P(S \mid A)}{P(S)}$

信息差 $I(\Delta S;A) = I(S;A) - I(S';A)$

17个氨基酸的信息差之和 $\sum I(\Delta S;A) = I(\Delta S;A_1) + I(\Delta S;A_2) + \ldots + I(\Delta S;A_{17})$

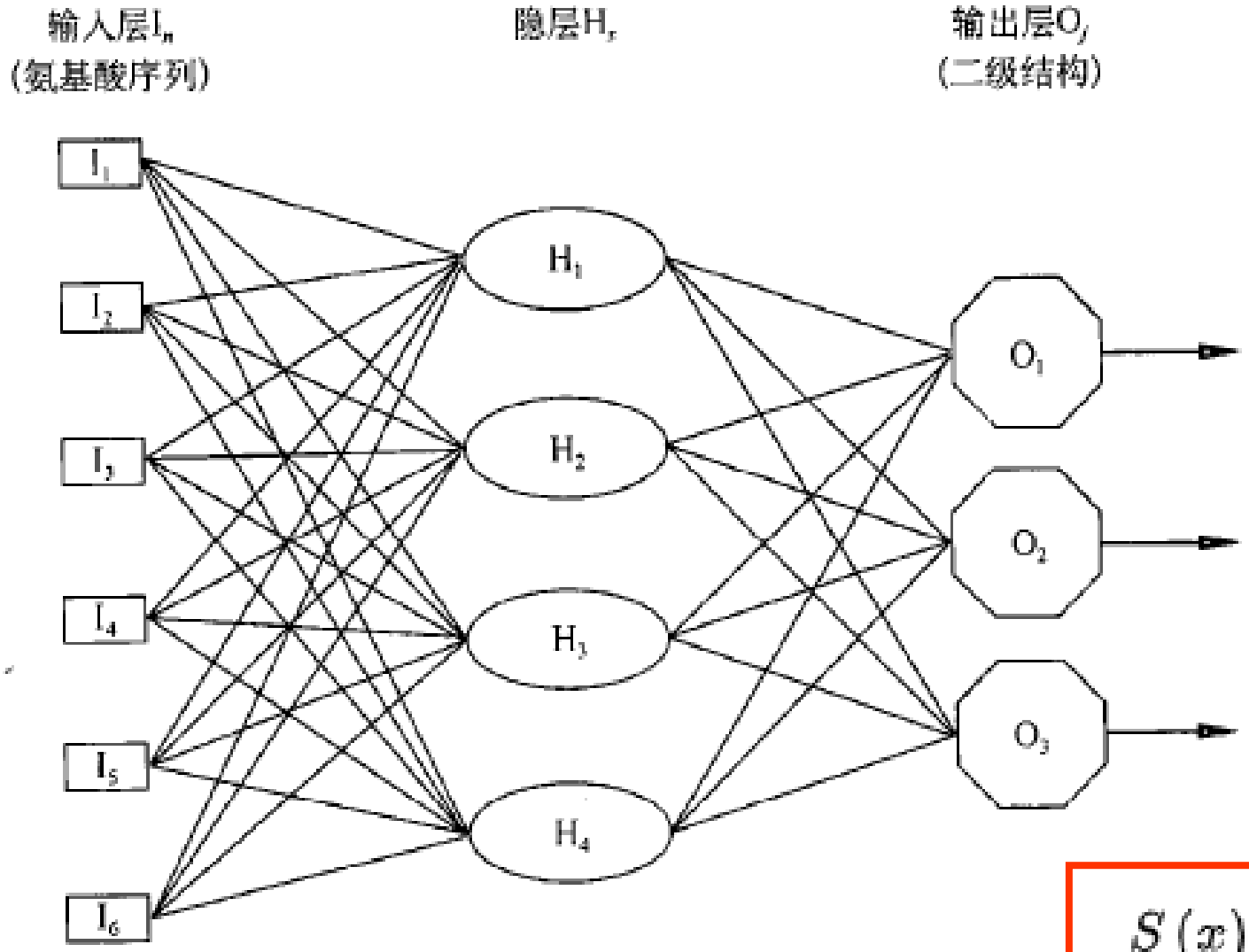
*其中，$P(S \mid A) = P(S,A) / P(A)$

$P(S,A) = f_{S,A} / N$；$P(S) = f_S / N$；$P(A) = f_A / N$

参数	含义	参数	含义
A	氨基酸	P(S)	所有氨基酸中S的发生概率
S	二级结构种类	P(S,A)	同时观察到S与A的联合概率
S'	除S以外的二级结构	P(S | A)	当氨基酸是A时，二级结构S的发生概率
N	氨基酸残基总数	I(S;A)	氨基酸A处于S时的信息值
f_A	氨基酸残基A的总数	I(S';A)	氨基酸A处在S'时的信息值
f_S	二级结构S的总数	I(ΔS;A)	氨基酸A处在S时的信息差
$f_{S,A}$	氨基酸A处在二级结构S的总数	$\sum I(\Delta S;A)$	信息差之和
P(A)	氨基酸A出现的概率	$\sum I(\Delta S;A_i)$	氨基酸A_1, A_1, …, A_{17}处于S时的信息差

➢神经网络模型

- 原理：先构建由函数式连接组成的神经网络，再输入已知二级结构的氨基酸序列，不断调节有关参数，使输出的二级结构与已知的相符，由此得出应用模型，再用此模型预测待测蛋白质的结构。
- **PHD（Profile network from HeiDelberg）**是德国的Rost和Sander研制的用神经网络预测蛋白质二级结构的方法。该法利用多序列比对得到的信息，再结合神经网络方法预测蛋白质二级结构，准确率超过70%。



$$S(x) = \frac{1}{1+e^{-x}}$$

一个蛋白质二级结构的BP神经网络模型

*** 蛋白质二级结构预测不仅是联系其一级结构和三级空间结构的桥梁和纽带，而且也是从一级结构预测其三级结构的关键步骤，其发展过程大致可分为三个阶段：**

第一阶段：以单残基、单一序列分析为重点，以Chou-Fasman方法和GOR等方法为代表。

第二阶段：考虑了局部残基的相互影响，人工神经网络方法得到应用，以PHD为代表，预测正确率提高至70%左右。

第三阶段：提出了多重序列比对的思想，尤其在近年使用蛋白质结构数据库（PDB）搜索比对，使预测正确率有了明显提高, 如PSIPRED算法。

***蛋白质二级结构预测算法的输出一般和下面的结果相似，H和h都表示预测为螺旋的区域，E和e都表示预测为伸展结构的区域：**

APAFSVSPASGASDGQSVSVSVAAAGETYYIAQCAPVGGQDACNPAT

- - - - - - - - - HHHHHH- HHHhhh - - - - - EEEEEeee- - - - EEEEee

⑤蛋白质二级结构预测软件

代表性软件	网址
PredictProtein	http://www.predictprotein.org/
PSIPRED	http://bioinf.cs.ucl.ac.uk/psipred/
Jpred3	http://www.compbio.dundee.ac.uk/www-jpred/index.html
SSpro	http://scratch.proteomics.ics.uci.edu/
PORTER	http://distill.ucd.ie/porter/

- SSpro 是第一个使用神经网络与同源分析混合进行蛋白质二级结构预测的软件，预测结果比较准确，可以作为蛋白质二级结构预测的首选之一。
- PORTER 也是综合应用了神经网络和多序列比对的新型蛋白质二级结构的预测方法，使蛋白质二级结构预测的准确率有了较大的提高。

➢ PredictProtein：

- PROFsec：分析序列二级结构
- PROFacc：分析残基溶剂可及性
- PHDhtm：分析序列潜在的跨膜拓扑结构
- COILS：预测卷曲结构
- PROSITE：搜索模体，预测序列潜在的功能
- ProDom：预测结构功能域

➢PSIPRED

- PSIPRED v3.0：预测二级结构
- MEMSAT3＆MEMSAT-SVM：预测跨膜区拓扑结构
- MEMPACK：预测跨膜拓扑和螺旋结构
- GenTHREADER和pGenTHREADER：基于折叠识别法预测蛋白质结构
- pDomTHREADER：基于结构域比来提高预测精度，适用于超家族结构域识别

UCL Department Of Computer Science

Bioinformatics Group

Bloomsbury Centre for Bioinformatics

UCL

The PSIPRED Protein Sequence Analysis Workbench

Input | Sequence Filter

Choose Prediction Methods

- [x] PSIPRED v3.3 (Predict Secondary Structure)
- [ ] pGenTHREADER (Profile Based Fold Recognition)
- [ ] BioSerf v2.0 (Automated Homology Modelling)
- [ ] FFPred 3 (Eukaryotic Function Prediction)
- [ ] MEMPACK (SVM Prediction of TM Topology and Helix Packing)
- [ ] DomSerf v2.0 (Automated Domain Modelling by Homology)
- [ ] DISOPRED3 (Disorder Prediction)
- [ ] MEMSAT3 & MEMSAT-SVM (Membrane Helix Prediction)
- [ ] DomPred (Protein Domain Prediction)
- [ ] GenTHREADER (Rapid Fold Recognition)
- [ ] pDomTHREADER (Fold Domain Recognition)

选中预测软件

Help...

Input Sequence (Single sequence or Multiple Sequence alignments; as raw sequence or fasta format)

```
>3CIG:A|PDBID|CHAIN|SEQUENCE
QCTVRYNVADCSHLKLTHIPDDLPSNITVLNLTHNQLRRLPPTNFTRYSQLAILDAGFNS
ISKLEPELCQILPLLKVLNLQHNELSQISDQTFVFCTNLTELDLMSNSIHKIKSNPFKNQ
KNLIKLDLSHNGLSSTKLGTGVQLENLQELLLAKNKILALRSEELEFLGNSSLRKLDLSS
NPLKEFSPGCFQTIGKLFALLLNNAQLNPHLTEKLCWELSNTSIQNLSLANNQLLATSES
```

psipred.fasta

Help...

If you wish to test these services follow this link to retrieve a test fasta sequence.

Submission Details

Email Address for job completion alert (optional) Help...

gongjing@sdu.edu.cn

结果会发送到邮箱里，也可以在线等待，大约需要30分钟。注意：不支持免费的商业邮箱，比如：hotmail，gmail，139，QQ邮箱等。

Password (only required for licenced commercial e-mail addresses) Help...

Short identifier for submission Help...

3CIG

给预测任务起名

Predict Clear form

PSIPRED 预测结果和 DSSP 真实二级结构的比较

氨基酸序列	ADFNGTWEMLSNDNFEDVMKALDIDFATRKIAVHLKQTKVIVQNGDKFETKTLSTF
实验测定结果	ccccceeeeeeeechhhhhhhhccccchhhhhhhhhccccceeeeeeeccceeeeeeeccc
Porter	ccccceeeeeeeechhhhhhhhccccchhhhhhhhhccccceeeeeeeccccceeeeeeccc
Jpred	ccccceeeeeeccchhhhhhhhccccchhhhhhhhhccccceeeeeeccceeeeeeeccc
DSC	ccccceeeeeeccchhhhhhhhccccchhhhhhhhhccccceeeeeeccccceeeeeeccc
PHD	ccccceeeeeeccchhhhhhhhccchhhhhhhhhhccccceeeeeeccccceeeeeeccc
SSpro	ccccceeeeeeccchhhhhhhhcchchhhhhhhhhccccceeeeeeccccceeeeeeccc
MLRC	ccccceeeeeeechhhhhhhhhhccchhhhhhhhhccccceeeeeeccccceeeeeeccc
SIMPA96	ccccceeeeeeechhhhhhhhhcchhhhhhhhhhhcceeeeeeccccceeeeeeccc
HNN	ccccceeeeeeechhhhhhhhhhhhhhhhhhhhccccceeeeeeccccceeeeeeccc
GOR IV	ccccceeeeeeechhhhhhhhhhhhhhhhhhhhccceeeeeeccccceeeeeeccc
nnPredict	ccccchheeeeeeechhhhhhhcchhhhhhhhhhhccceeeeeeccccchhhccccc
SOPMA	hhhtceeeeechhhhhhhhhhhtchhhhhhhhhhccccceeeeettceeeeeeeccc
PREDATOR	ccccccceeeeeechhhhhhccccchhhhhhhhhccceeeeeeccccceeeeeeccc
DPM	ctctcchhhhttcchhhhhhhhhhehhhehhehehheeeeeeetcchhhheeeccc

图 11-15　用 13 种软件预测斑马鱼视黄醇结合蛋白部分序列的二级结构

斑马鱼视黄醇结合蛋白质序列及其二级结构实验测定结果来自 PDB 数据库，编号为 1kqw。图中“h”表示 α 螺旋，“e”表示 β 折叠，“c”表示无规则卷曲，“t”表示 β 转角。图中根据准确率由高到低粗略地由上至下给这 13 种方法排位置。不过这个准确率是仅对本图中序列来确定的，当用其他序列时结果可能会不同。

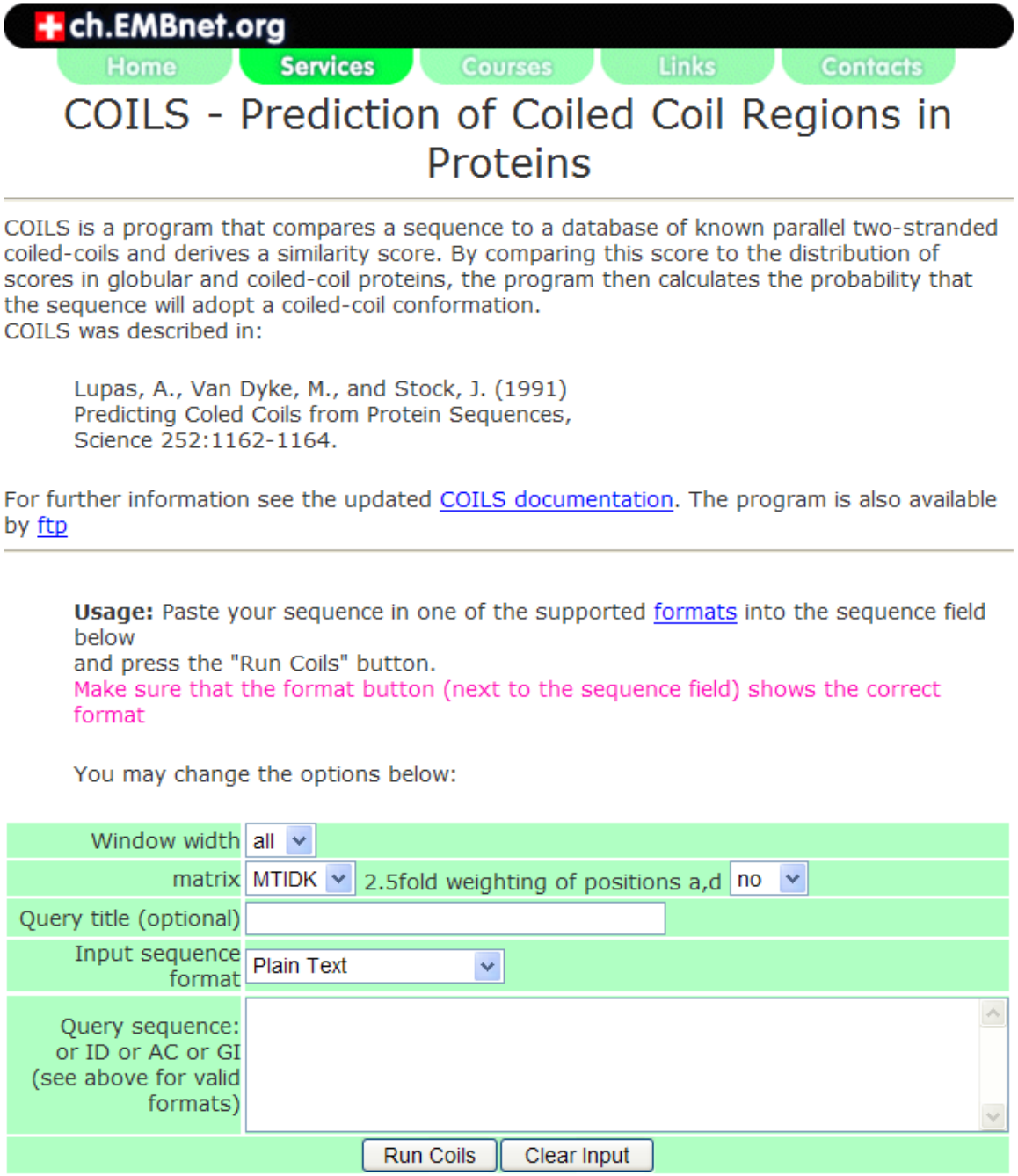
➢卷曲螺旋预测软件

- **卷曲螺旋：** 由两股或两股以上α-螺旋相互缠绕而形成的超螺旋结构的总称。该结构存在于多种天然蛋白质中，如转录因子、骨架蛋白、动力蛋白、膜蛋白、酶等。七肽重复区是典型的卷曲螺旋结构类型之一。

  a - b - c - d - e – f – g (a、d为非极性疏水氨基酸，位于卷曲螺旋的内侧；e、g为极性带电荷氨基酸，位于疏水核心的外侧)

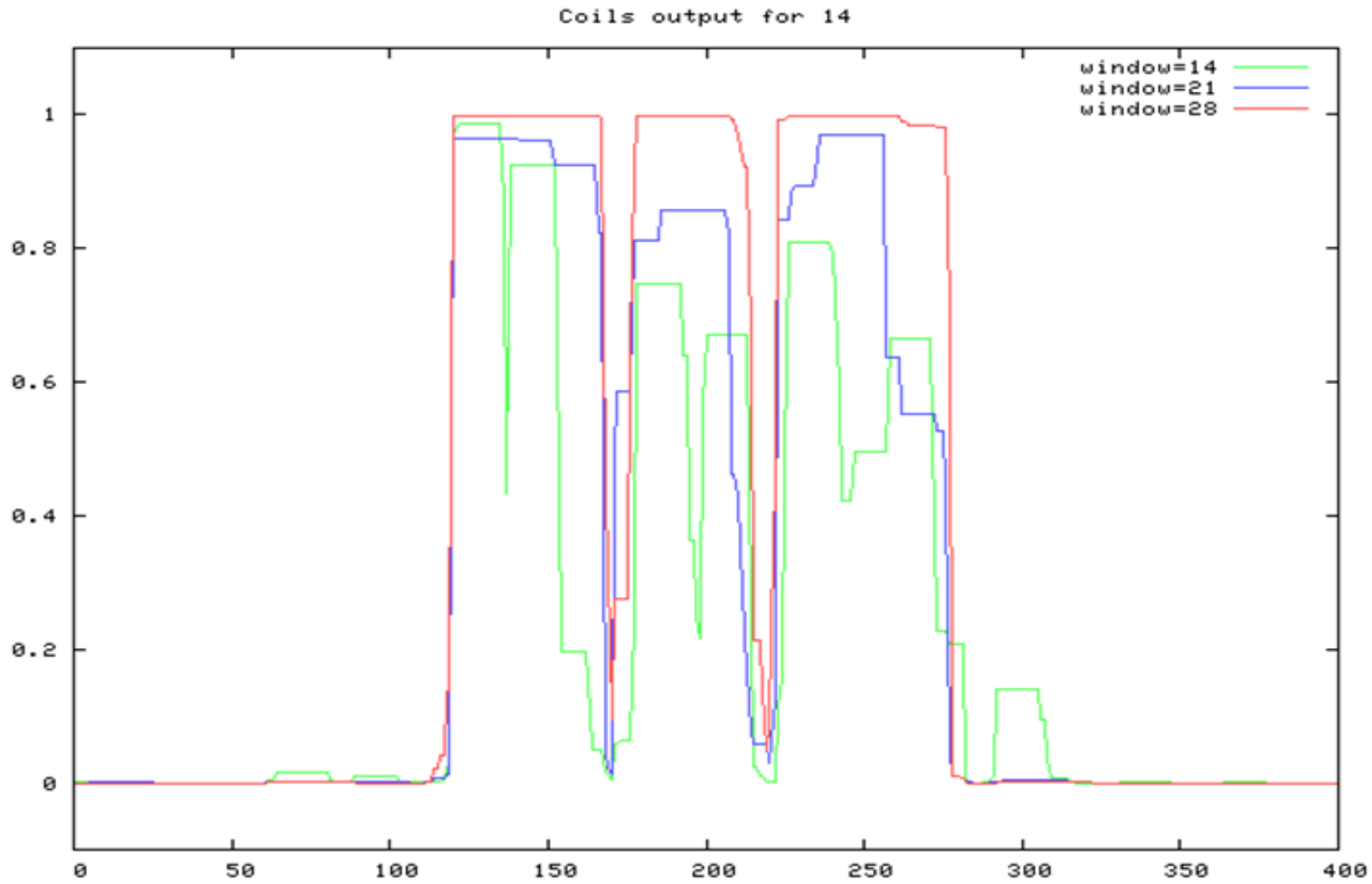
- **COILS软件：** 其原理是将输入序列提交到已知包含卷曲螺旋蛋白结构的数据库中进行搜索，同时与包含球状蛋白序列的PDB库进行比较，根据两个数据库分析的情况算出目的序列形成卷曲螺旋的概率

网址：http://www.ch.embnet.org/software/COILS_form.html



- Window width为7的倍数；
- 打分矩阵权重a和d的分值可以提高2.5倍，加权后和未加权分值的结果做比较，如果其概率值相差20%-30%，表明该高分值片段是由氨基酸偏好所引起的假阳性。

举例：GO_HUMAN


Coils output for 14
window=14
window=21
window=28
1
0.8
0.6
0.4
0.2
0
0
50
100
150
200
250
300
350
400

➢跨膜区结构预测软件

- **跨膜蛋白**：膜蛋白不溶于水，不容易生长晶体，很难确定其结构。蛋白质序列含有跨膜区，提示它可能作为膜受体起作用，也可能是定位在膜上的锚定蛋白或离子通道蛋白。对膜蛋白的跨膜螺旋进行预测是生物信息学的重要应用。其预测算法都是基于统计学模型或神经网络，综合不同的软件预测结果并结合疏水性图，可以获得较好的预测结果，准确率可达80%-95%。
- **常见的跨膜区分析在线软件**：

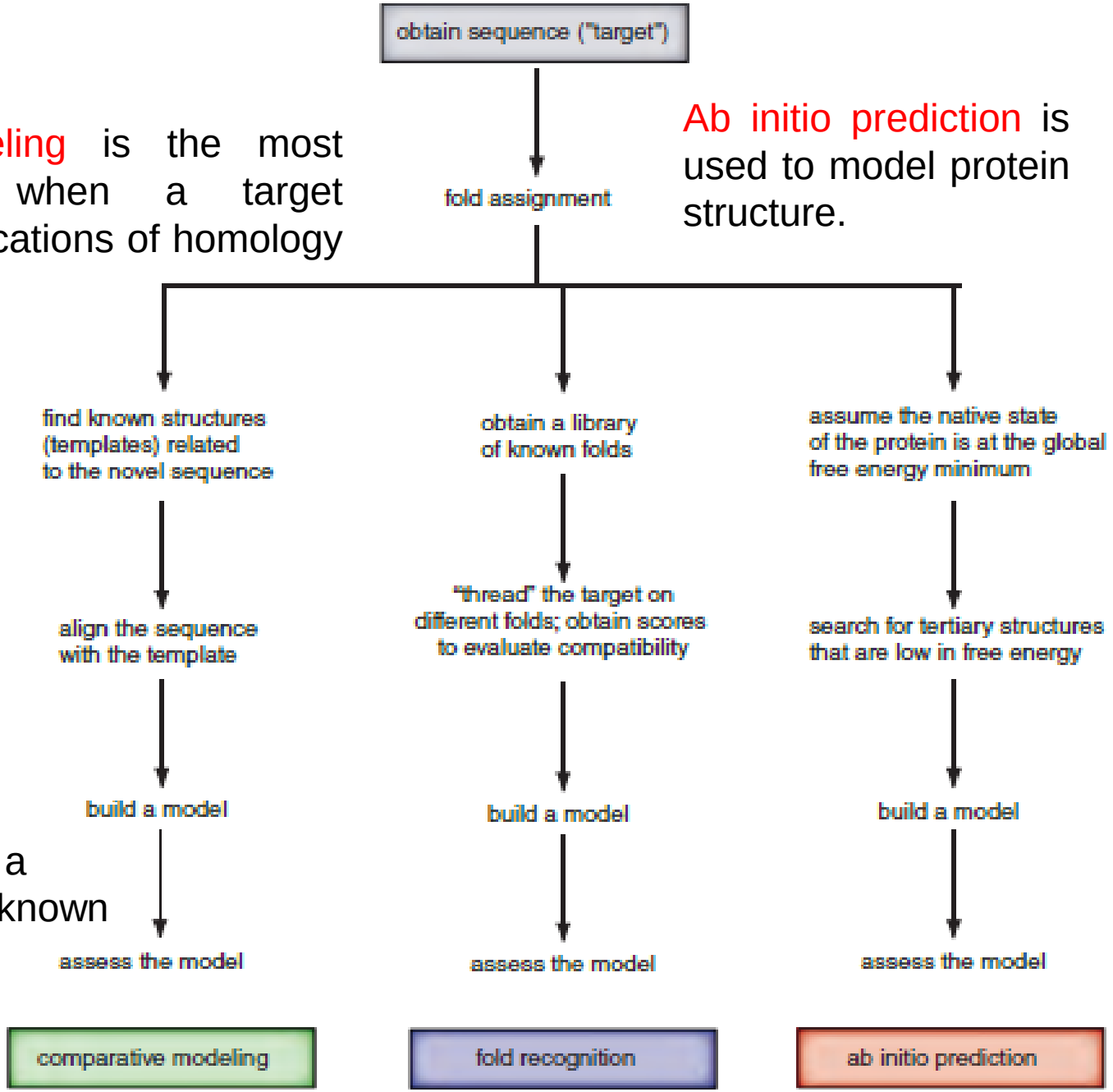
软件	网址	说明
TMHMM	https://services.healthtech.dtu.dk/service.php?TMHMM-2.0	判定某蛋白是否为膜蛋白
SPLIT	http://split.pmfst.hr/split/4/	准确预测跨膜蛋白的跨膜片段
HMMTOP	http://www.enzim.hu/hmmtop/html/submit.html	用于预测蛋白质的跨膜螺旋和拓扑结构

⑥ 蛋白质三级结构预测算法

- 目前蛋白质三级结构预测方法一般归为同源模建(Comparative modeling，CM)、折叠识别如线串法(Threading)、从头预测(Ab initio)三大类。
- 深度学习技术与以上经典方法的结合提升了蛋白结构预测的准确性和速度，但是对于没有同源性蛋白结构的预测，仍然存在巨大的挑战。

Comparative modeling is the most powerful approach when a target sequence has any indications of homology with a known structure.

Ab initio prediction is used to model protein structure.



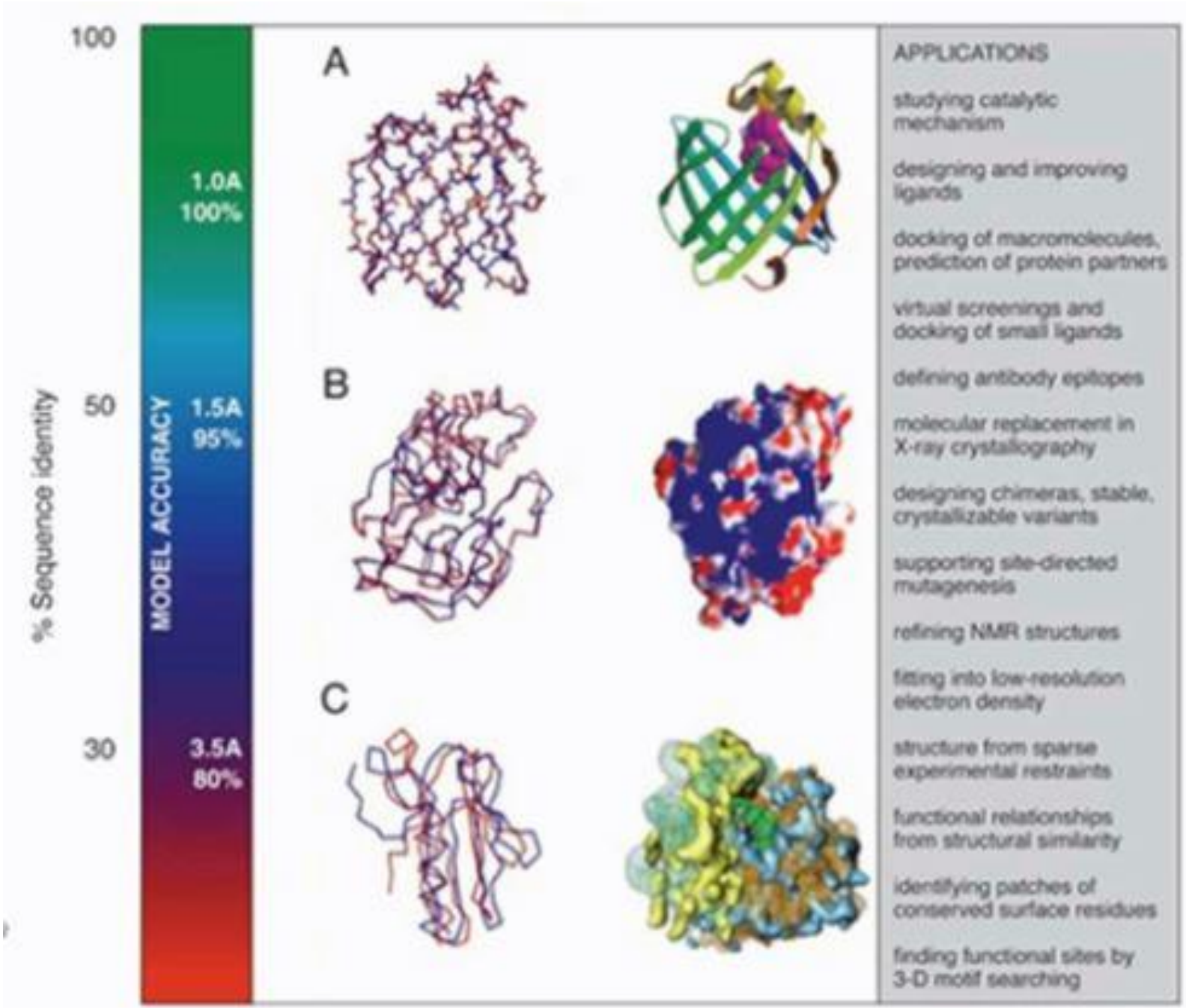
Threading is used to compare segments of a protein to a library of known folds.

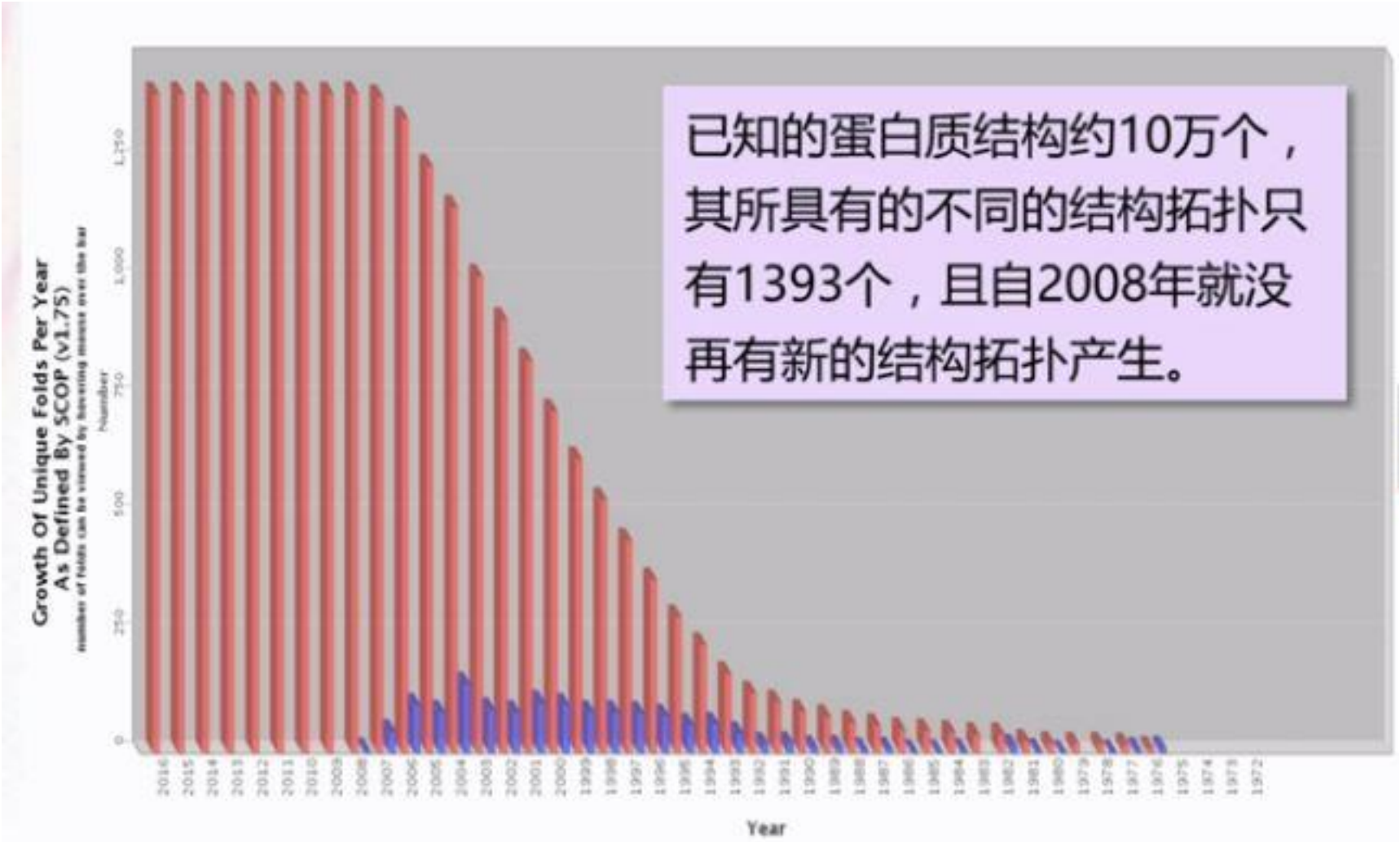
➢同源模建

预测未知结构的蛋白质的三级结构时，通过序列比对分析找到已知结构的同源蛋白质，再以同源蛋白质的结构为模板，构建待测蛋白质的三级结构。

其步骤如下：

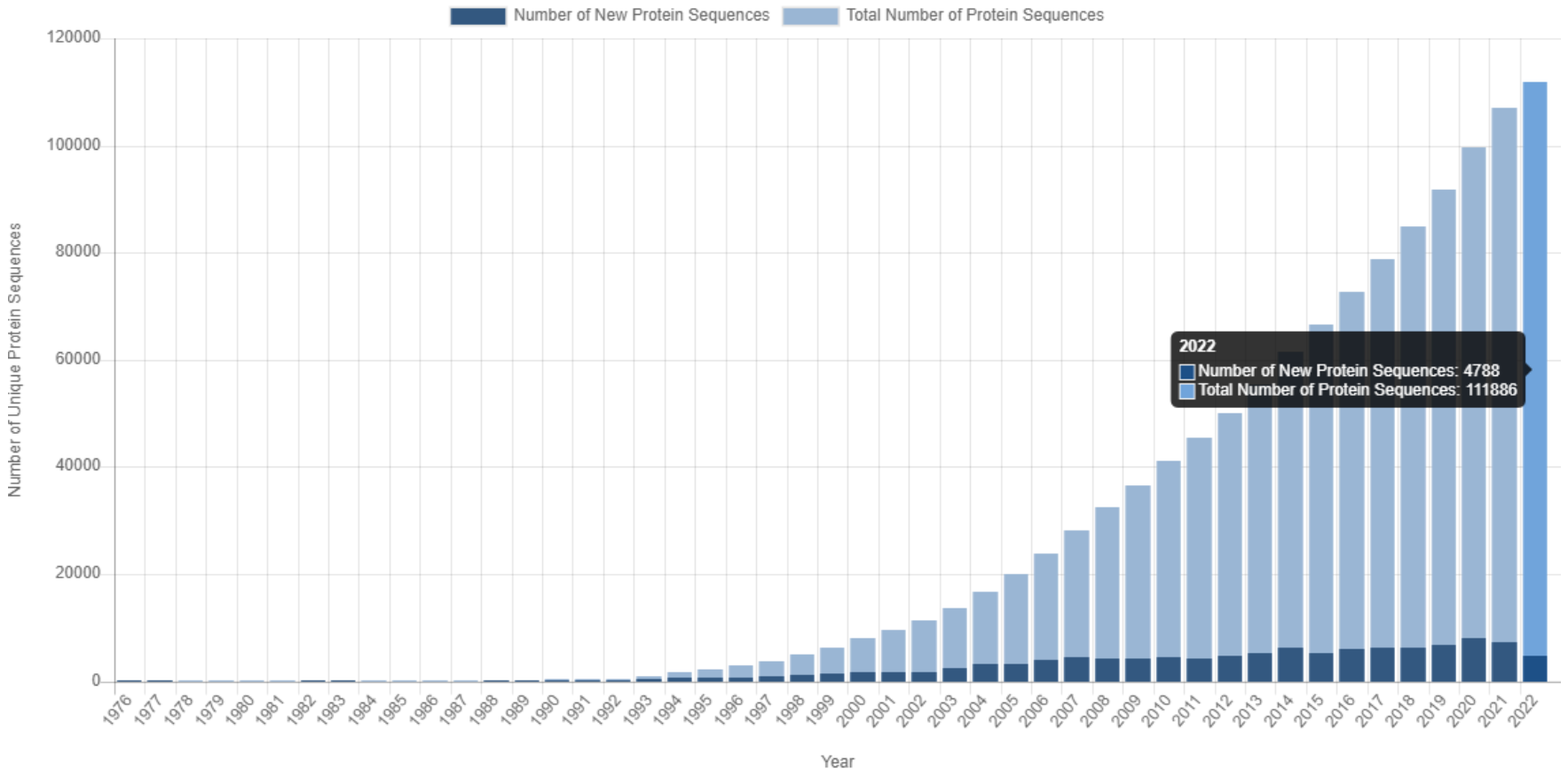
- 定模板：确定与待测蛋白质同源的并已知三级结构的蛋白质
- 序列比对：通过序列比对找到序列保守的部分
- 建骨架：仿照同源蛋白质的结构建立待测蛋白质的结构骨架
- 建环：将序列匹配的骨架片段连接起来，与模板不匹配的序列用来构建环区
- 建侧链
- 评估模型


100
50
30
% Sequence identity
MODEL ACCURACY
1.0A
100%
1.5A
95%
3.5A
80%
A
B
C
APPLICATIONS
studying catalytic mechanism
designing and improving ligands
docking of macromolecules, prediction of protein partners
virtual screenings and docking of small ligands
defining antibody epitopes
molecular replacement in X-ray crystallography
designing chimeras, stable, crystallizable variants
supporting site-directed mutagenesis
refining NMR structures
fitting into low-resolution electron density
structure from sparse experimental restraints
functional relationships from structural similarity
identifying patches of conserved surface residues
finding functional sites by 3-D motif searching



图片来源：http://www.rcsb.org/pdb/statistics/contentGrowthChart.do?content=fold-scop

海量的蛋白质序列


Number of New Protein Sequences
Total Number of Protein Sequences
Number of Unique Protein Sequences
120000
100000
80000
60000
40000
20000
0
2022
Number of New Protein Sequences: 4788
Total Number of Protein Sequences: 111886
1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022
Year

➢线串法，Threading

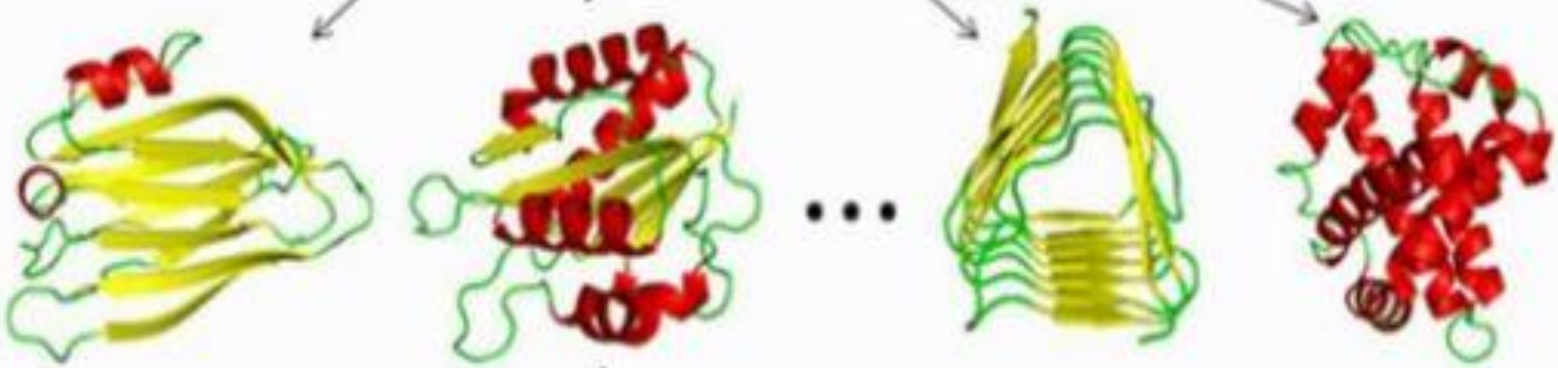
用一个已知结构的模板与待测蛋白质序列进行匹配分析，这过程需要用一些计分函数进行评估，计分函数中包括序列相似性、结构相似性、不同距离氨基酸残基的相互作用等因素，最后根据计分函数的评估结果确定待测蛋白质序列（或其中的片段）的结构是否属于某种模板结构（或其中的部分），折叠识别是线串法的重要方式。**一般在待测序列搜索不到相似性高于25%的序列时才使用**，否则可使用准确性相对较高的同源模建方法。

Target sequence

LKADSSTATSTIQKALNNCDQGKAVRLSGVSLLIDKGVTLRAVNNAKSFENAPSSCGVVDKNG.....

Threading

Template library



Best template (energy function)

Predicted model

➢从头预测法 (Ab initio)

要获得一个实用模型就必须考虑以下两个因素：

- **蛋白质结构的计分方法:** 有些计分方法只考虑蛋白质的疏水作用而建立的简化模型，如HP模型。
- **结构空间的搜索方法：** 当蛋白质序列较短时，可用穷举法搜索所有的空间，经计分方法评价后找出最优结构。

从头预测法有3个重要要素：能量函数、构象搜索、模型选择

HP模型

该模型规定氨基酸残基只分亲水氨基酸（用P表示）与疏水氨基酸残基（用H表示）两类，可用小方块或小圆球表示氨基酸残基。

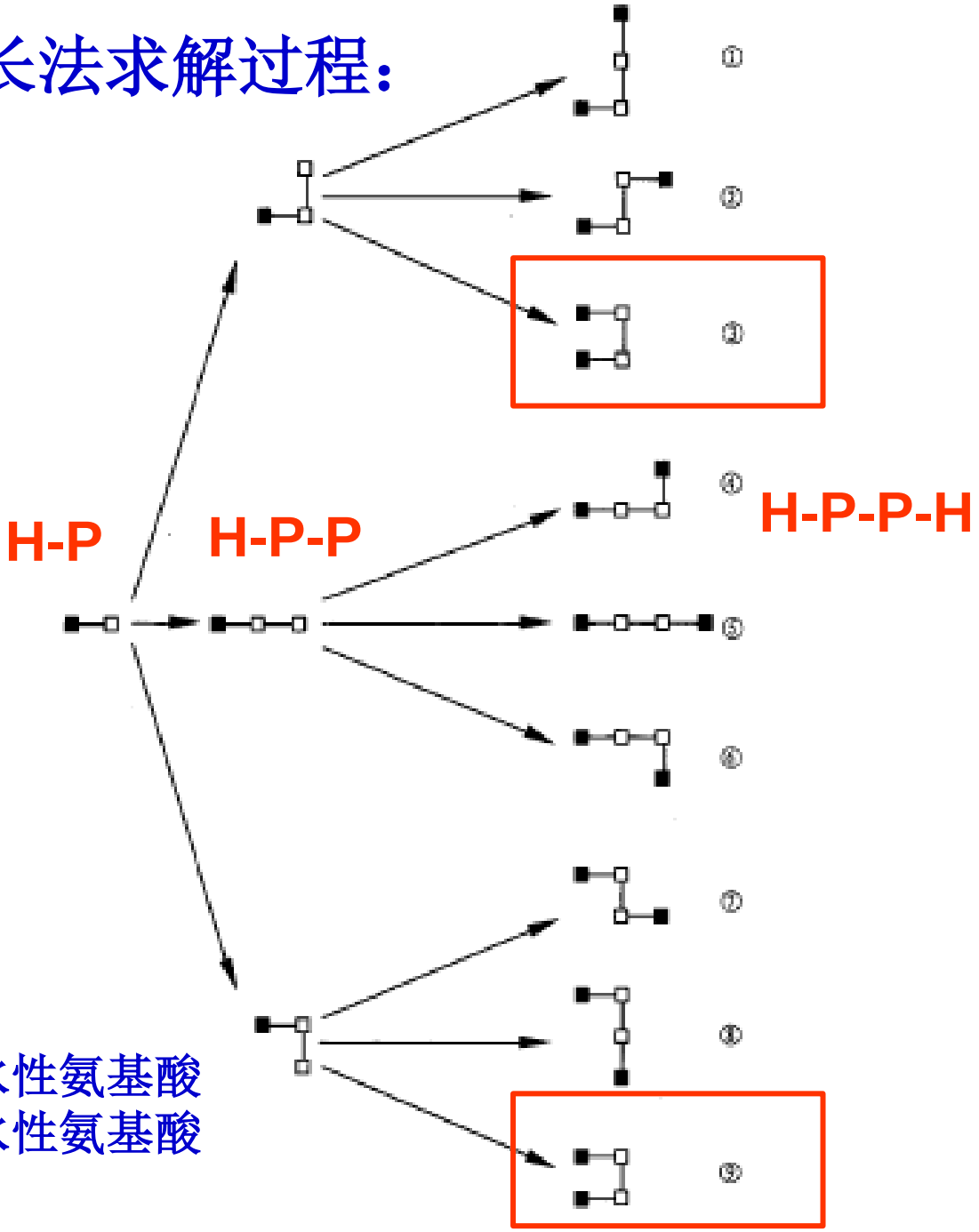
- **规则**：蛋白质序列中每个残基须放在平面整数坐标系中，序列中相邻残基在坐标系中也相邻，距离为1个刻度单位，整数坐标系中每一个格只能放1个氨基酸残基
- **计分方法**: 序列不相邻而平面空间中相邻的一对疏水氨基酸的能量为-1，其他为0，也就是该模型的最优结构应该是把疏水氨基酸残基尽量聚在一起。
- **在蛋白质序列较少时，可用穷举法找出HP模型的最优结构；蛋白质序列较长时，可用遗传算法寻找最优结构。**

举例

已知一氨基酸序列有4个残基，第一及第四个残基是疏水氨基酸残基，第二和第三个残基是亲水氨基酸残基。请用HP模型预测该序列在二维整数坐标系中的最优结构。

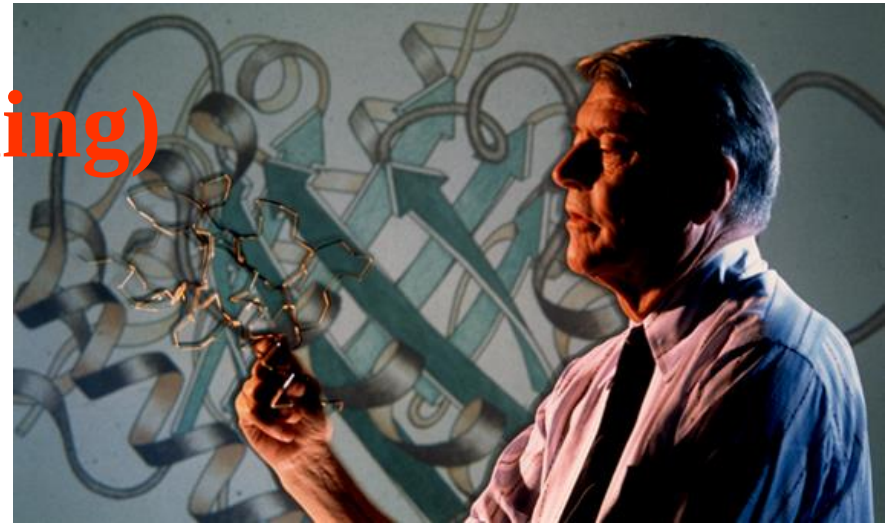
H-P-P-H

“HPPH”的逐点生长法求解过程：

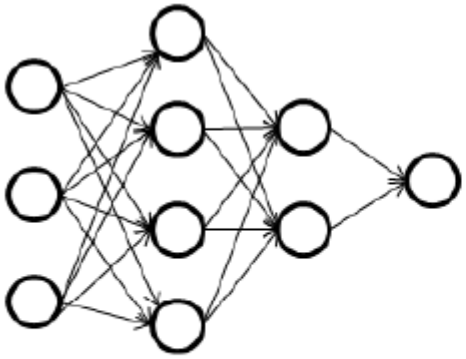


实心框代表疏水性氨基酸
空心框代表亲水性氨基酸

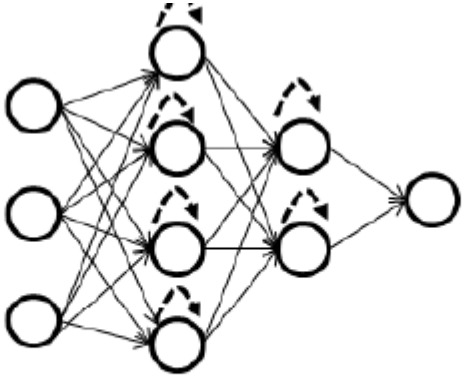
➢深度学习 (Deep learning)



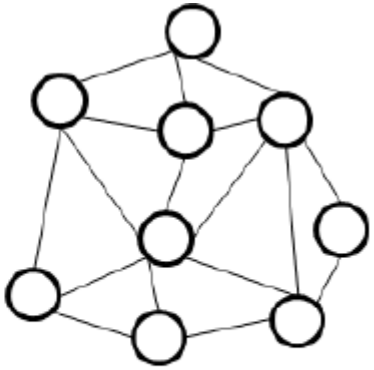
神经网络



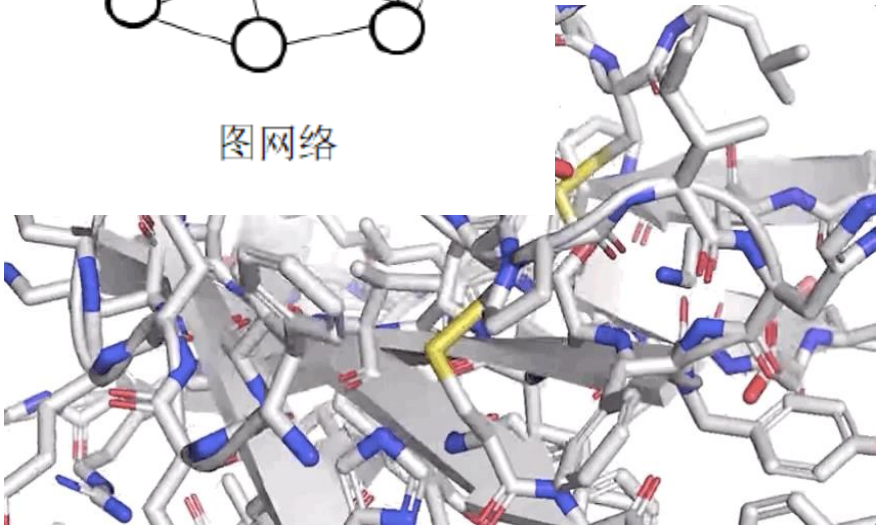
前馈网络



反馈网络

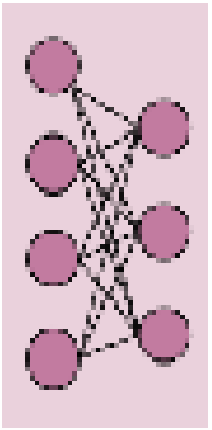


图网络



卷积神经网络（Convolutional Neural Network, CNN）

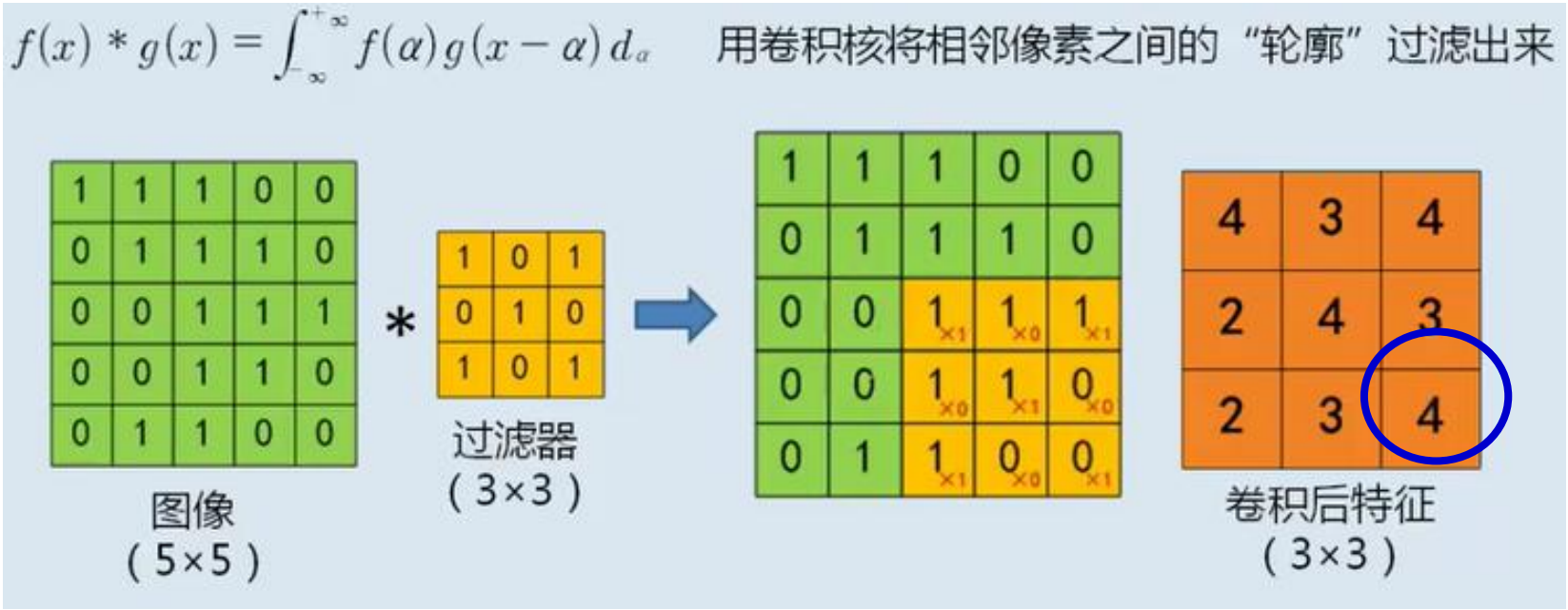
- 卷积核大小



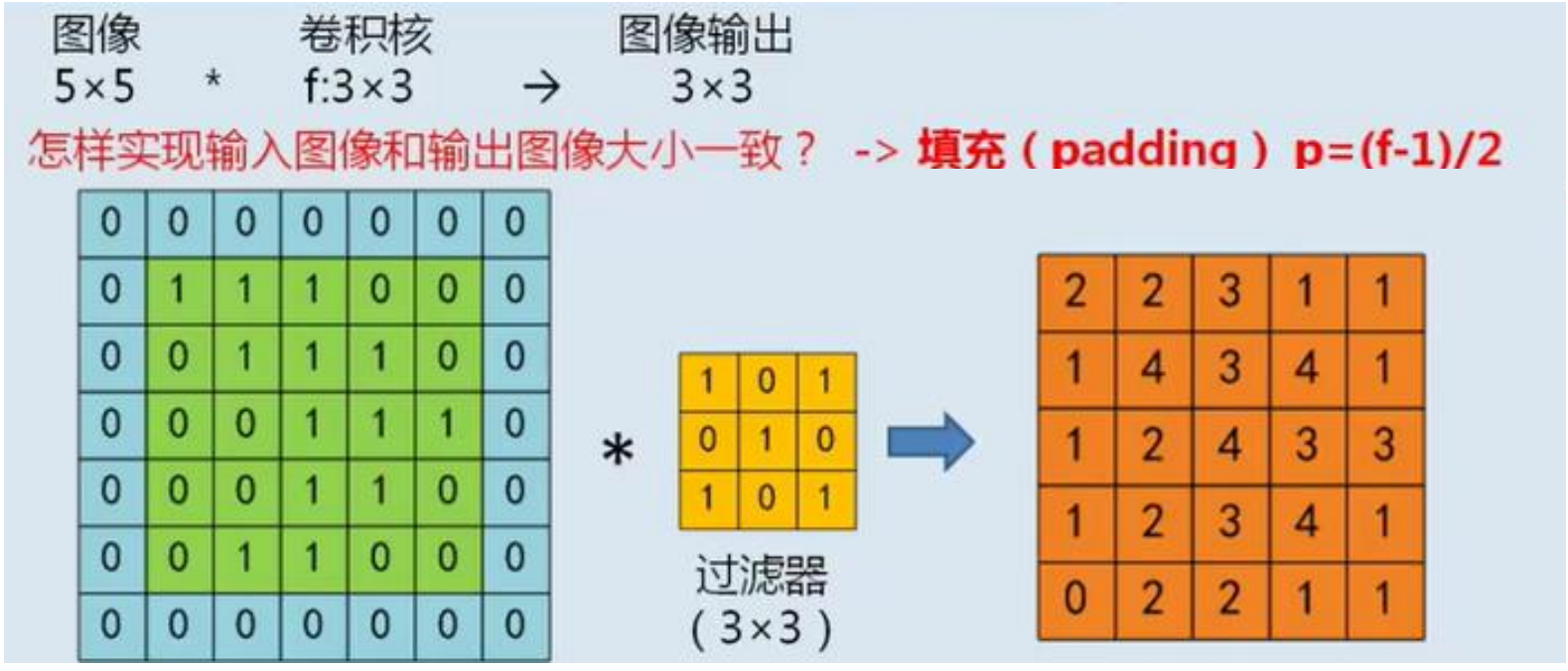
不是全连接的方式！

$$f(x) * g(x) = \int_{-\infty}^{+\infty} f(\alpha) g(x-\alpha)\, d_\alpha$$

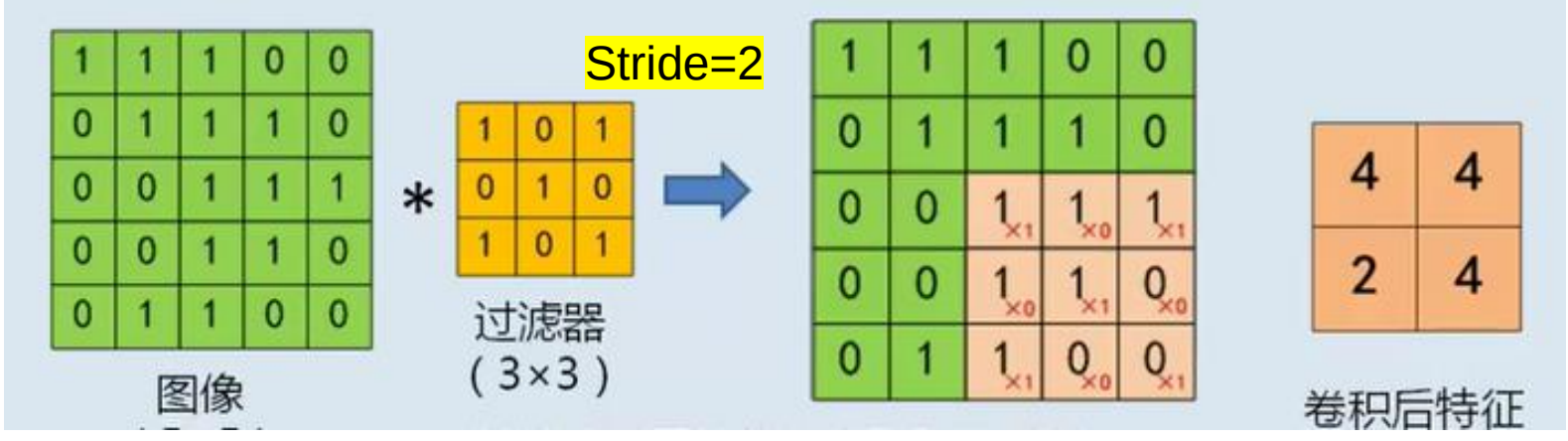
用卷积核将相邻像素之间的“轮廓”过滤出来



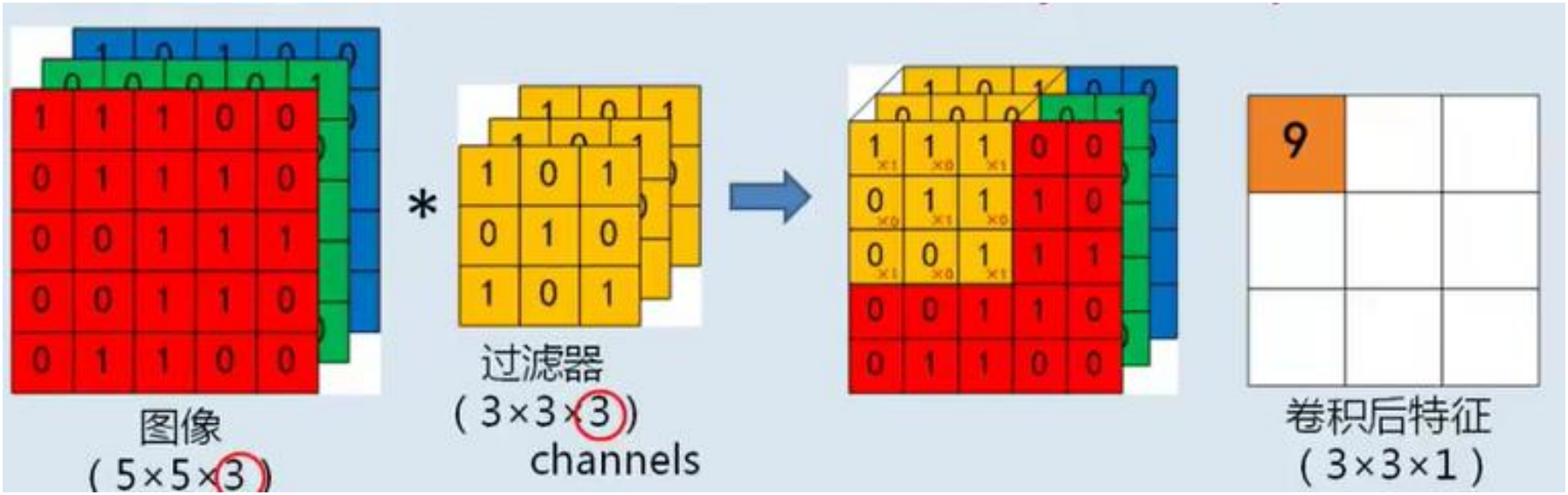
- 填充（Padding）



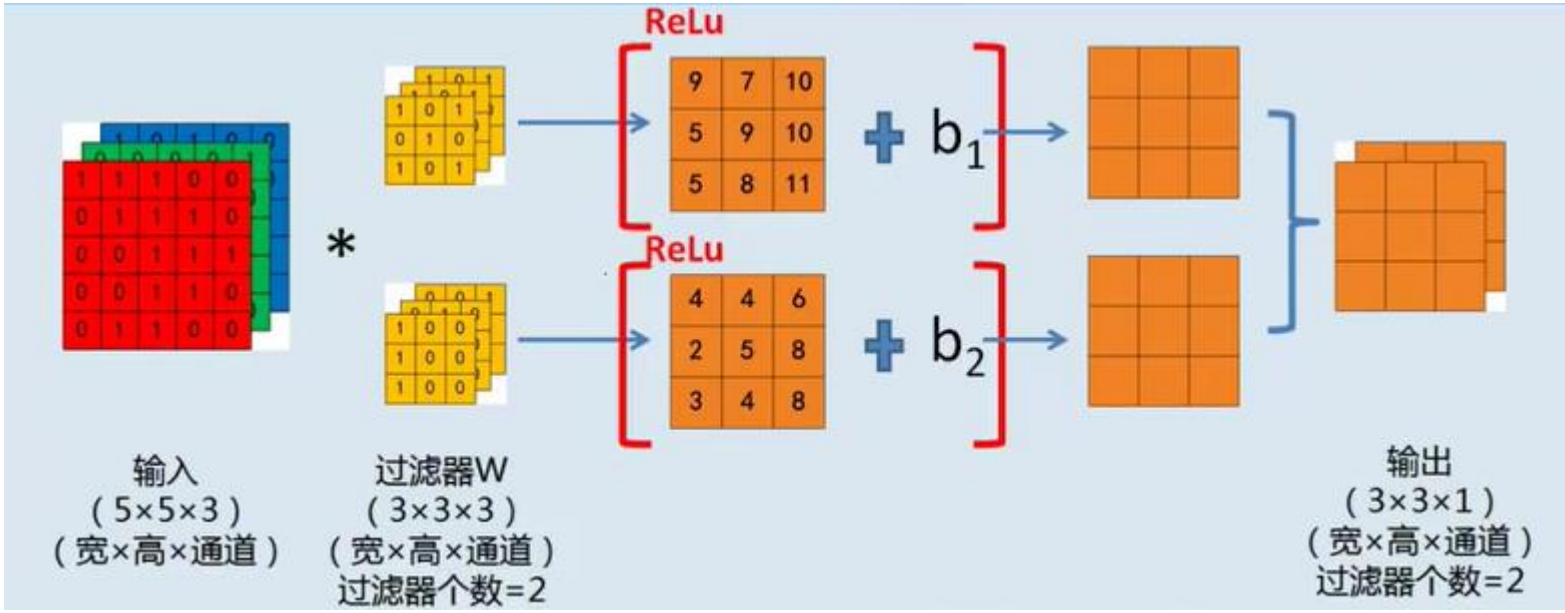
- 步长（Stride）



- 3个通道时，卷积核由1个变为相同的3个

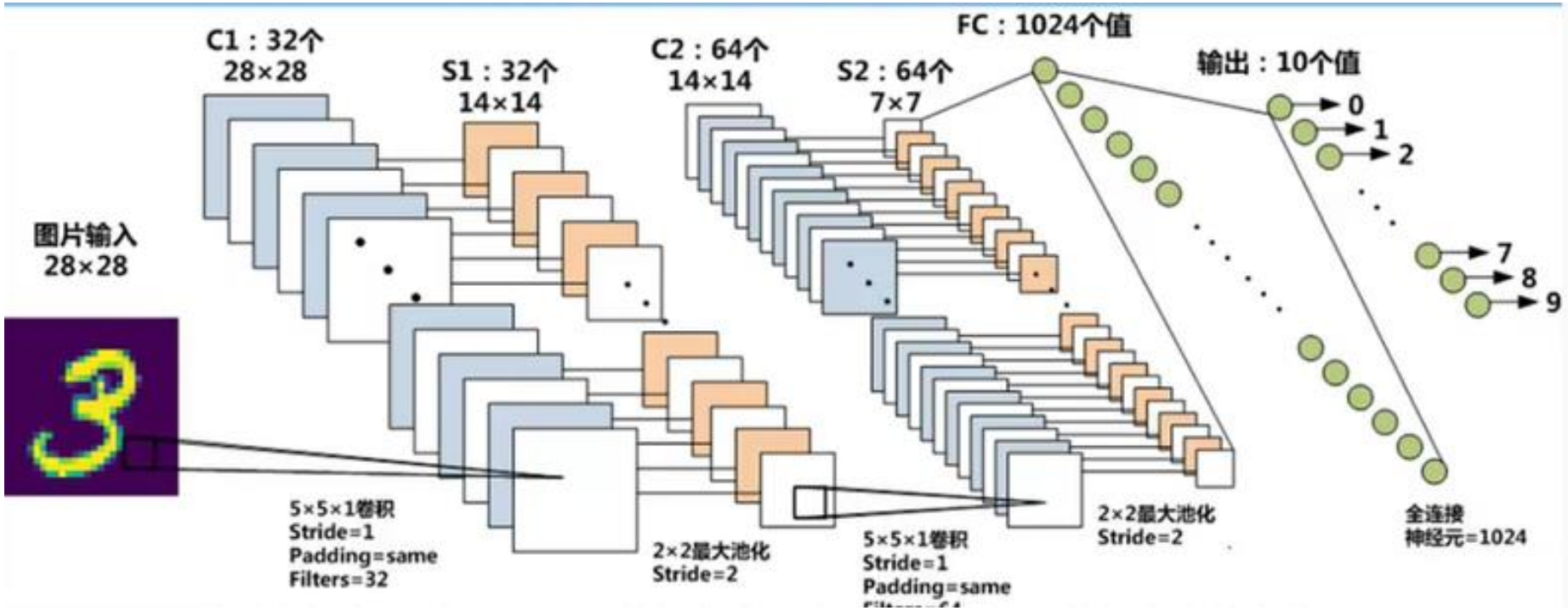


- Relu（激活函数）

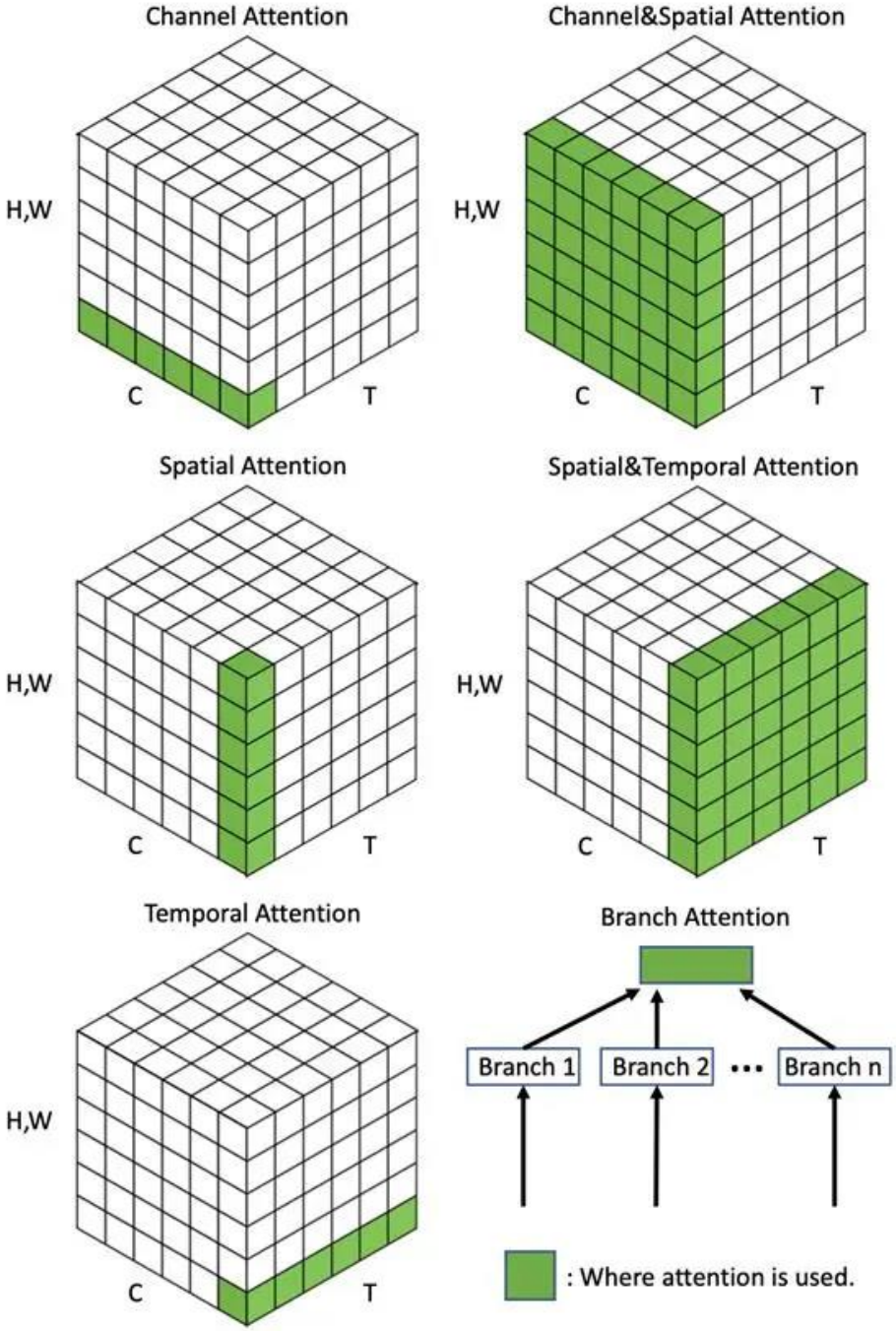


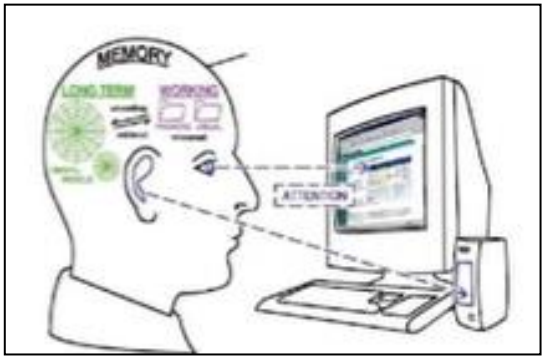
- 池化（Pooling，降维）

一个典型的多层CNN模型



变为全连接方式





Attention 注意力机制

⑦ 蛋白质三级结构预测软件

- **瑞士蛋白质专家分析系统 ExPASy（Expert Protein Analysis System）**
- **三级结构预测相关软件及下载地址：**

•SwissModel（同源模建法）

http:// swissmodel.expasy.org

•Phyre2（线串法）

http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index

I-TASSER（线串法）

https://zhanggroup.org/I-TASSER/

http://www.expasy.org/

Live Search

(F) 编辑(E) 查看(V) 收藏夹(A) 工具(T) 帮助(H)

收藏夹 建议网站 免费 Hotmail 获取更多加载项

PASy Proteomics Server

主页(M)

Search ExPASy web site for Go Clear

Swiss Institute of Bioinformatics

ExPASy Proteomics Server

Databases Tools Services Mirrors About Contact

re here: ExPASy CH

Notice: Due to maintenance work, some ExPASy services (including HAMAP, ScanProsite, ProRule, Swiss-2DPAGE, World-2DPAGE) will be inaccessible Wednesday November 17, 2010 from 7am to 8am, GMT+1.

e ExPASy (**Ex**pert **P**rotein **A**nalysis **Sy**stem) proteomics server of the Swiss Institute of Bioinformatics (SIB) is dedicated t
e analysis of protein sequences and structures as well as 2-D PAGE (Disclaimer / References / Linking to ExPASy).

Databases

niProtKB, PROSITE, HAMAP, SwissVar,
ralZone, SWISS-MODEL Repository,
WISS-2DPAGE, World-2DPAGE
epository, MIAPEGelDB, ENZYME,
lycoSuiteDB, SugarBind, UniPathway
etails] [full list]

Tools & Software

Proteomics tools, Blast, ScanProsite,
Melanie, MSight, Make2D-DB, SWISS-
MODEL, Swiss-PdbViewer,
SwissDock, SwissParam, QuickMod
[full list]

Latest News

Temporary inaccessibility of some ExPASy services - November 17, 2010

Due to maintenance work, some ExPASy services (including HAMAP, ScanProsite, ProRule, Swiss-2DPAGE or World-2DPAGE) will be inaccessible **Wednesday November 17, 2010 from 7am to 8am, GMT+1**.

SWISS-MODEL

Modelling
- myWorkspace
- Automated Mode
- Alignment Mode
- Project Mode

Tools
- Template Identification
- Domain Annotation
- Structure Assessment
- Template Library

Repository
- Search by Sequence
- Search by AC
- Search by full text

SWISS-MODEL is a fully automated protein structure homology-modeling server, accessible via the ExPASy web server, or from the program DeepView (Swiss Pdb-Viewer). The purpose of this server is to make Protein Modelling accessible to all biochemists and molecular biologists WorldWide.

What's new?

- SWISS-MODEL is running on new hardware with better performance
- Find more news on **SWISS-MODEL Blog**

SWISS-MODEL Team

Torsten Schwede:	Project Leader
Florian Kiefer:	SWISS-MODEL Repository
Lorenza Bordoli:	Method Development and user support
Konstantin Arnold:	SWISS-MODEL Workspace

References:
When you publish or report results using SWISS-MODEL, please cite the relevant publications:

- Arnold K., Bordoli L., Kopp J., and Schwede T. (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics, 22,195-201.
- Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T (2009). The SWISS-MODEL Repository and associated resources. Nucleic Acids Research. 37, D387-D392.
- Peitsch, M. C. (1995) Protein modeling by E-mail Bio/Technology 13: 658-660.

PHYRE2 Protein Fold Recognition Server - Internet Explorer, optimized for Bing and MSN

http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index

File Edit View Favorites Tools Help

Favorites | Web Slice Gallery | Suggested Sites | Biology - An Open Acc...

Zimbra: ... | Proteins ... | Macromole... | ScanProsite | Scratch P... | Chimera-R... | PredictPr... | PHYRE2...

Page | Safety | Tools

Standard Mode | Login for job manager, batch processing, Phyre alarm and other advanced options

Retrieve Phyre Job Id [ ] Fetch

Phyre2

Protein Homology/analogY Recognition Engine V 2.0

Subscribe to Phyre at Google Groups

Email: [ ] Subscribe

Visit Phyre at Google Groups

Follow @Phyre2server

Feedback welcome

E-mail Address	[ ]
Optional Job description	[ ]
Amino Acid Sequence	[ ]
	Or try the sequence finder (NEW!)
Modelling Mode	Normal ◉ Intensive ○
	Phyre Search Reset

770449 submissions since Feb 14 2011

Done

Internet

⑧ 蛋白质结构综合分析软件

Chimera

网址： **http://plato.cgl.ucsf.edu/chimera/**

⑨ 蛋白质结构预测评价

- CASP
- EVA

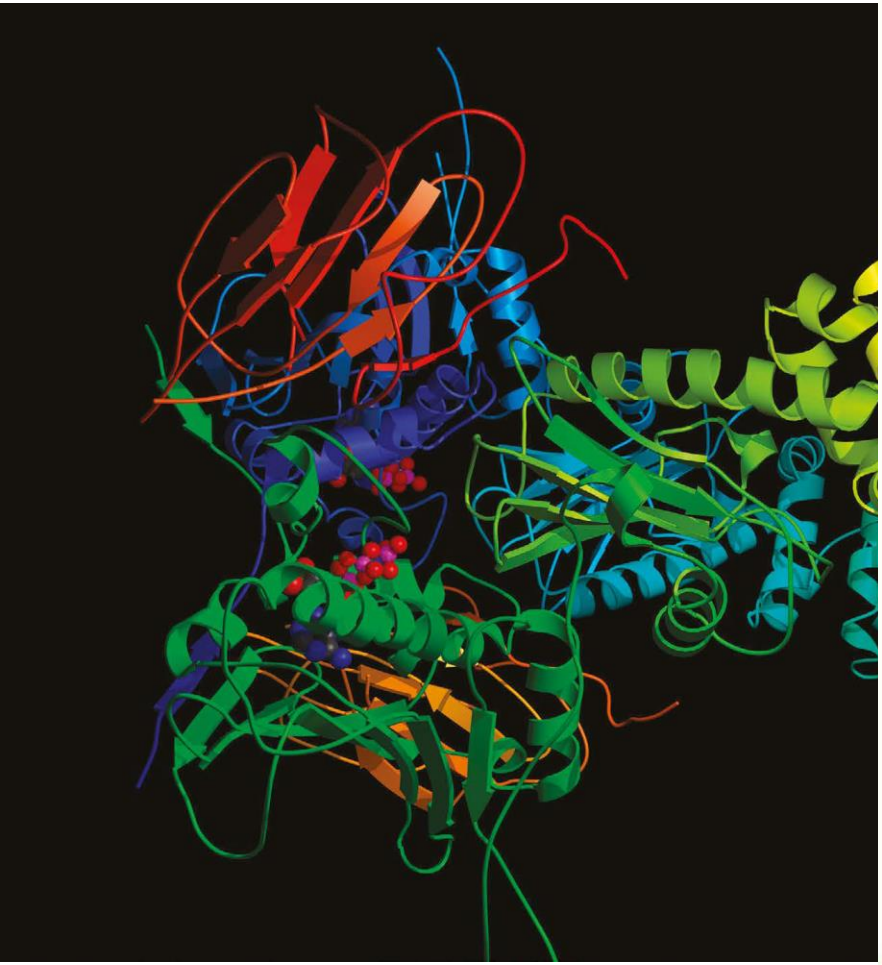
➢三态残基准确率Q_3

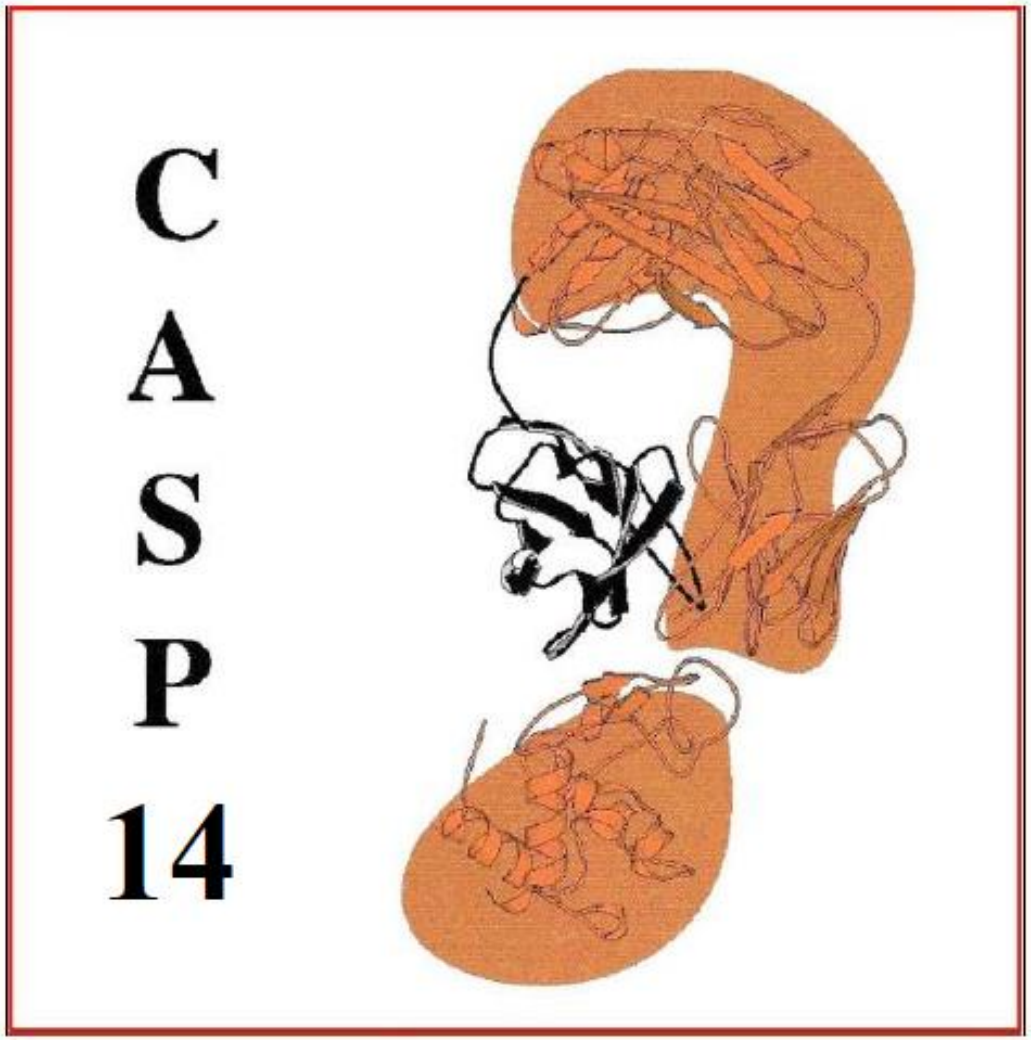
$$Q_3=\frac{P\alpha+P\beta+P_C}{N}\times 100\%$$

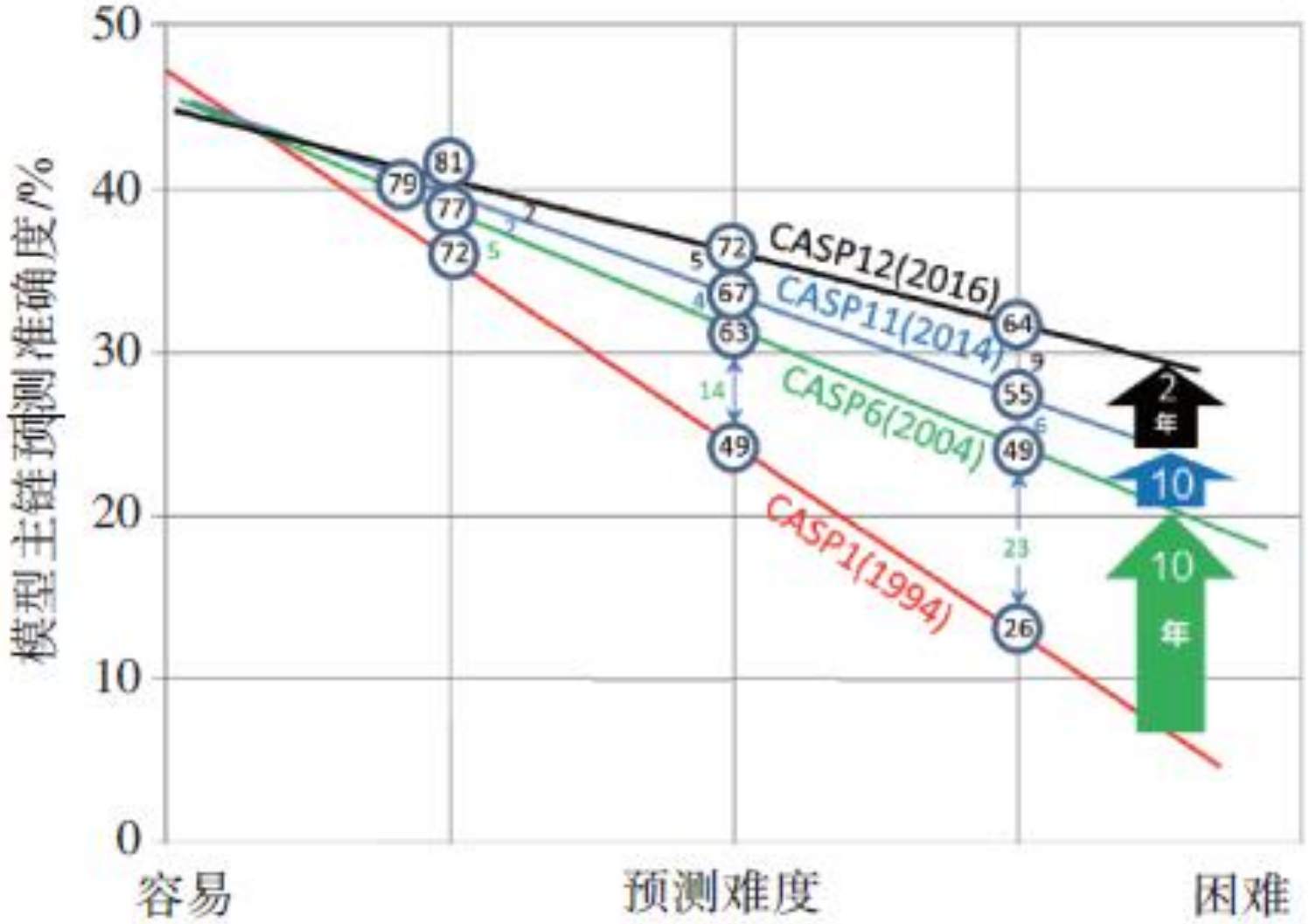
➢残基整体准确率

➢Matthews校正系数

Critical Assessment of protein Structure Prediction, CASP

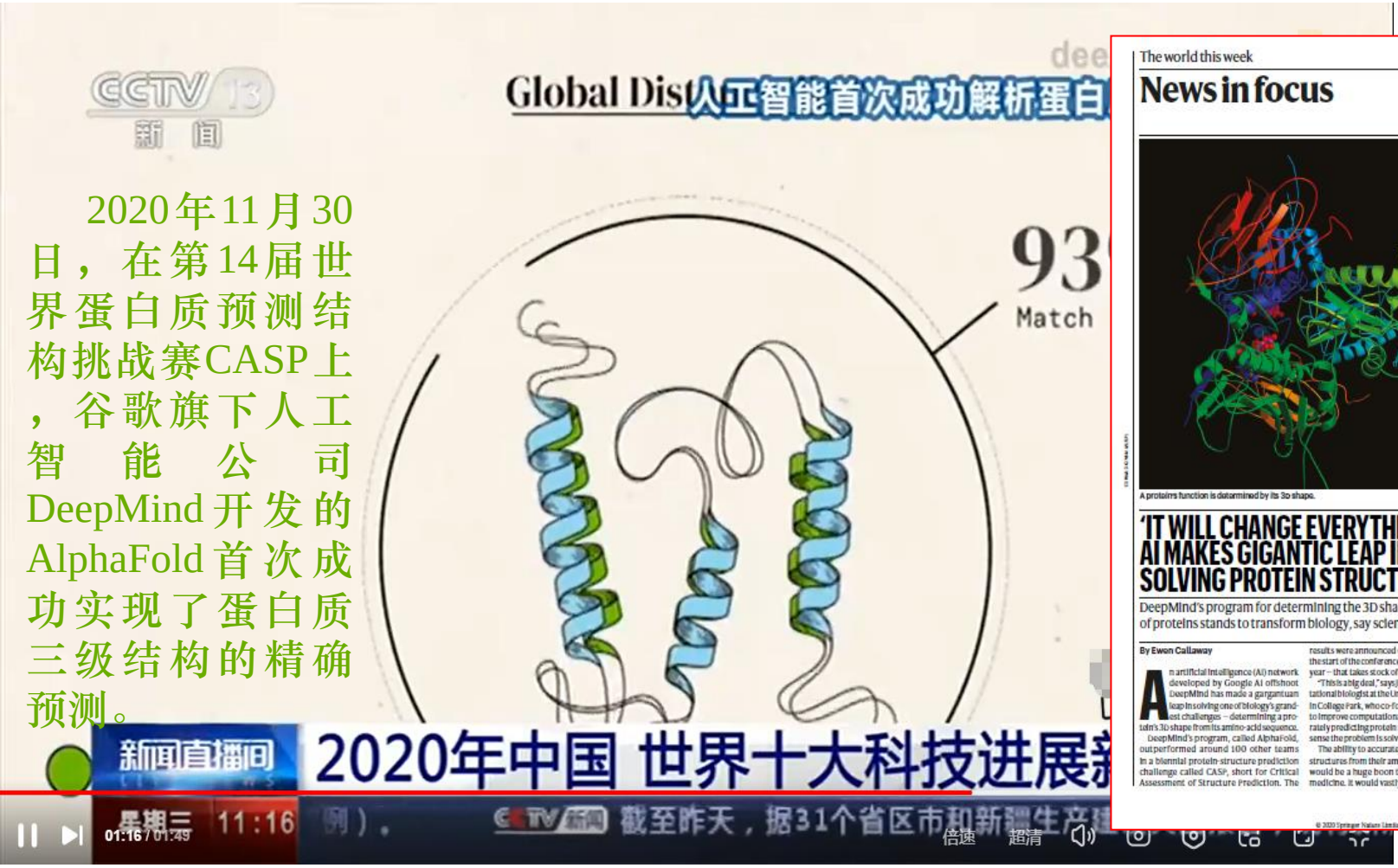






CASP 中模型主链预测准确度比较

2020年11月30日，在第14届世界蛋白质预测结构挑战赛CASP上，谷歌旗下人工智能公司DeepMind开发的AlphaFold首次成功实现了蛋白质三级结构的精确预测。



The world this week

News in focus

A protein's function is determined by its 3D shape.

'IT WILL CHANGE EVERYTHING': AI MAKES GIGANTIC LEAP IN SOLVING PROTEIN STRUCTURES

DeepMind's program for determining the 3D shapes of proteins stands to transform biology, say scientists.

By Ewen Callaway

An artificial intelligence (AI) network developed by Google AI offshoot DeepMind has made a gargantuan leap in solving one of biology's grandest challenges – determining a protein's 3D shape from its amino-acid sequence.

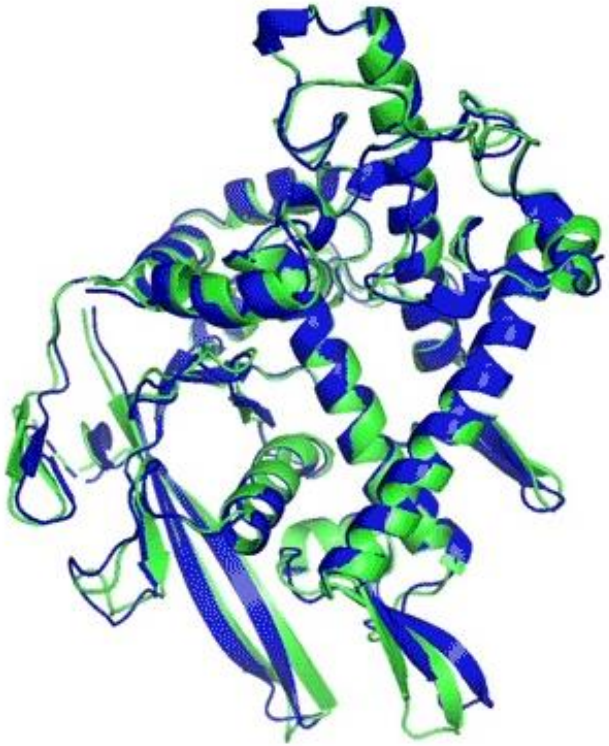
DeepMind's program, called AlphaFold, outperformed around 100 other teams in a biennial protein-structure prediction challenge called CASP, short for Critical Assessment of Structure Prediction. The results were announced on 30 November, at the start of the conference – held virtually this year – that takes stock of the exercise.

"This is a big deal," says John Moult, a computational biologist at the University of Maryland in College Park, who co-founded CASP in 1994 to improve computational methods for accurately predicting protein structures. "In some sense the problem is solved."

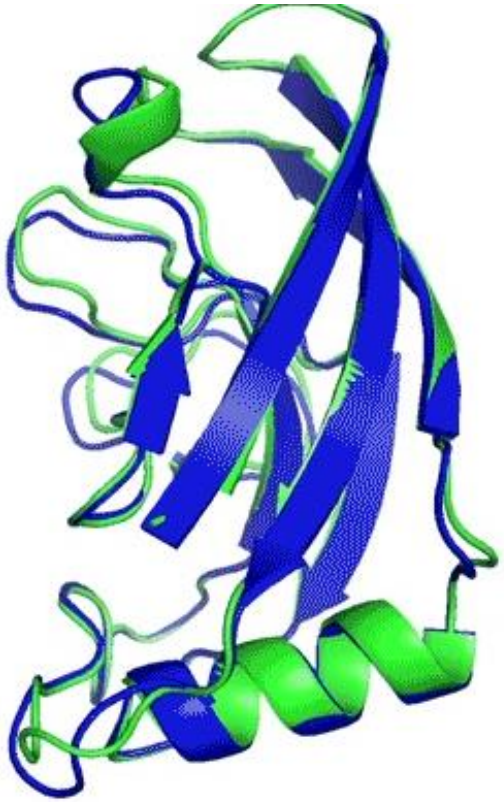
The ability to accurately predict proteins' structures from their amino-acid sequences would be a huge boon to life sciences and medicine. It would vastly accelerate efforts to understand the building blocks of cells and aid more advanced drug discovery.

AlphaFold came top of the table at the last CASP – in 2018, the first year that London-based DeepMind participated. But, this year, the outfit's deep-learning network was head-and-shoulders above other teams and, say scientists, performed so mind-bogglingly well that it could herald a revolution in biology.

"It's a game changer," says Andrei Lupas, an evolutionary biologist at the Max Planck Institute for Developmental Biology in Tübingen,



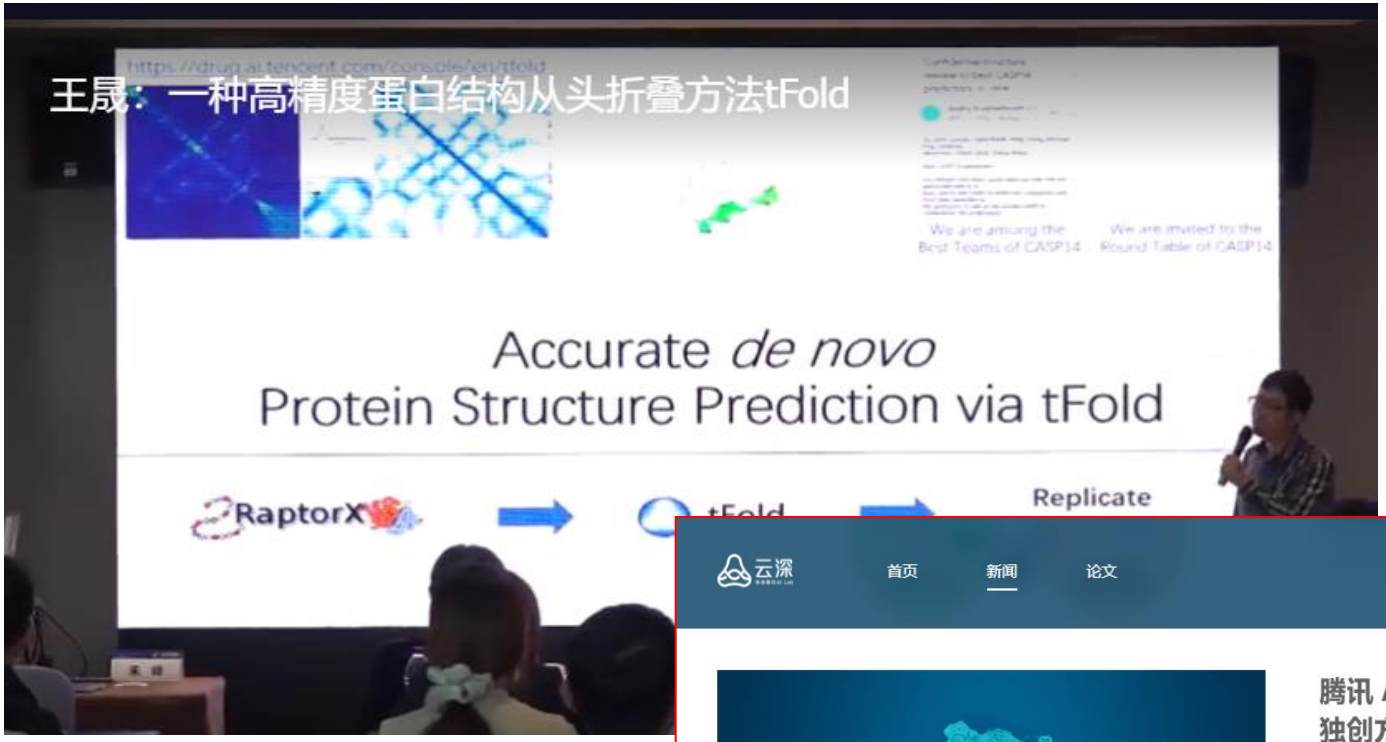
T1037 / 6vr4
90.7 GDT
(RNA polymerase domain)

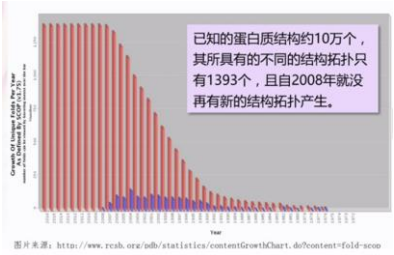


T1049 / 6y4f
93.3 GDT
(adhesin tip)

Structures of a protein that were predicted by artificial intelligence (blue) and experimentally determined (green) match almost perfectly. **DEEPMIND**

腾讯AI工具“tFold”助力蛋白结构解析

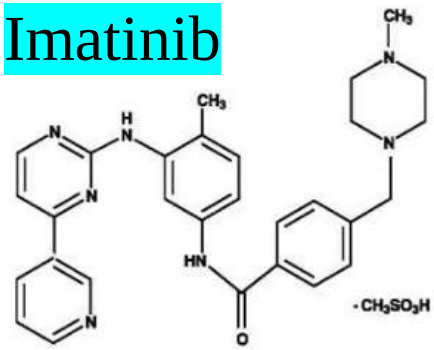




在新一代信息技术及互联网、大数据的时代背景下，我们更应**贯彻“务实、创新”**的精神，在科研突破中发挥核心作用。

伊马替尼是首个通过针对特定蛋白激酶开发的药物，也是人类第一个用于抗癌的分子靶向药。

Imatinib



电影中“格列宁”的真实原型

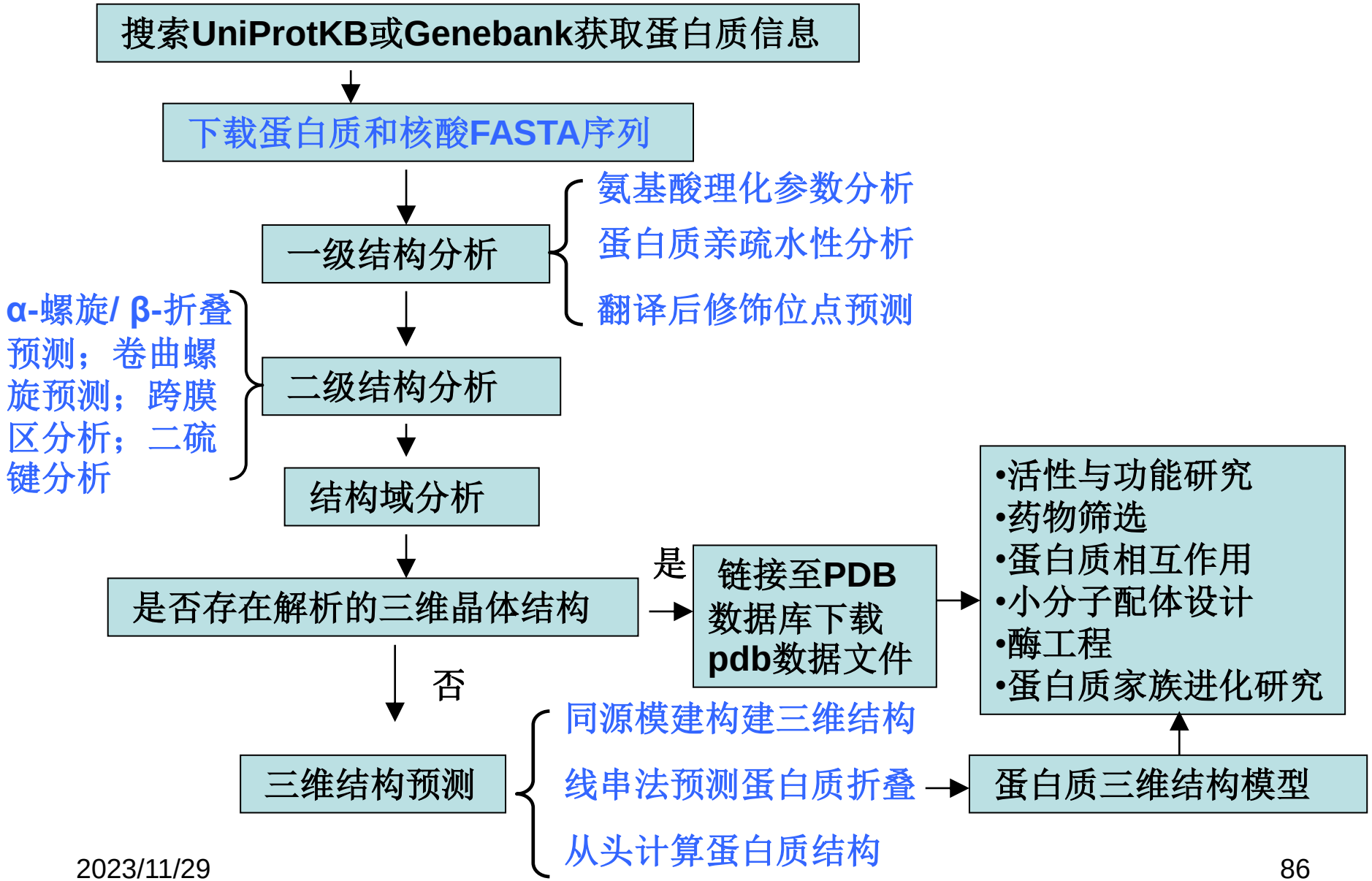
人民网 people.cn 人民网 >> 健康·生活

《我不是药神》引热议 李克强作批示

蛋白质三级结构预测对于药物设计、现代新药研发具有重要的意义。

“学以致用、科技报国”

⑩ 蛋白质结构与功能分析流程



本章作业（第6次课后作业）

对水稻瘤病毒P8 （RGDV P8）蛋白质进行如下分析：

① 疏水性分析

② 跨膜区分析

③ 通过SWISS-MODEL同源模建其三级结构