



蛋白质与蛋白质组学
Protein & Proteomics



第2章 蛋白质的高级结构

生命科学学院 刘峰

liufeng23@mail.sysu.edu.cn

生命科学楼1号楼324室

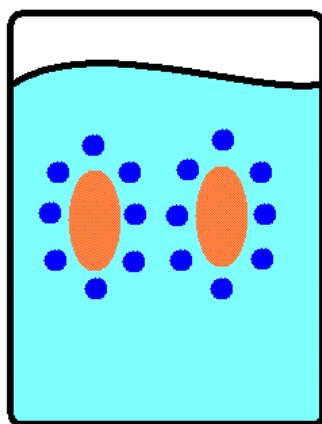
学习内容

- ▶ 第1节 蛋白质的稳定性及其动力学特性
- ▶ 第2节 三级结构：模体和结构域
- ▶ 第3节 四级结构：蛋白质寡聚体的形成
- ▶ 第4节 蛋白质高级结构的获取方法



第1节 蛋白质的稳定性及其动力学特性

$$\Delta G = \Delta H - T\Delta S$$

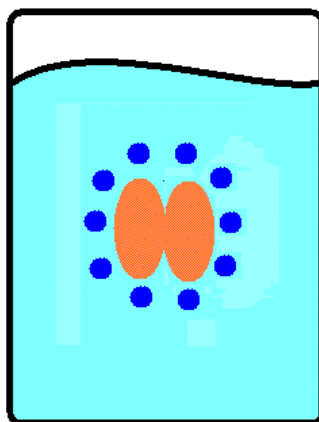


No Interactions

$\Delta H = \text{Negative}$

$\Delta S = \text{Negative}$

$\Delta G = \text{Positive}$



Hydrophobic Interaction

$\Delta H = \text{Positive}$

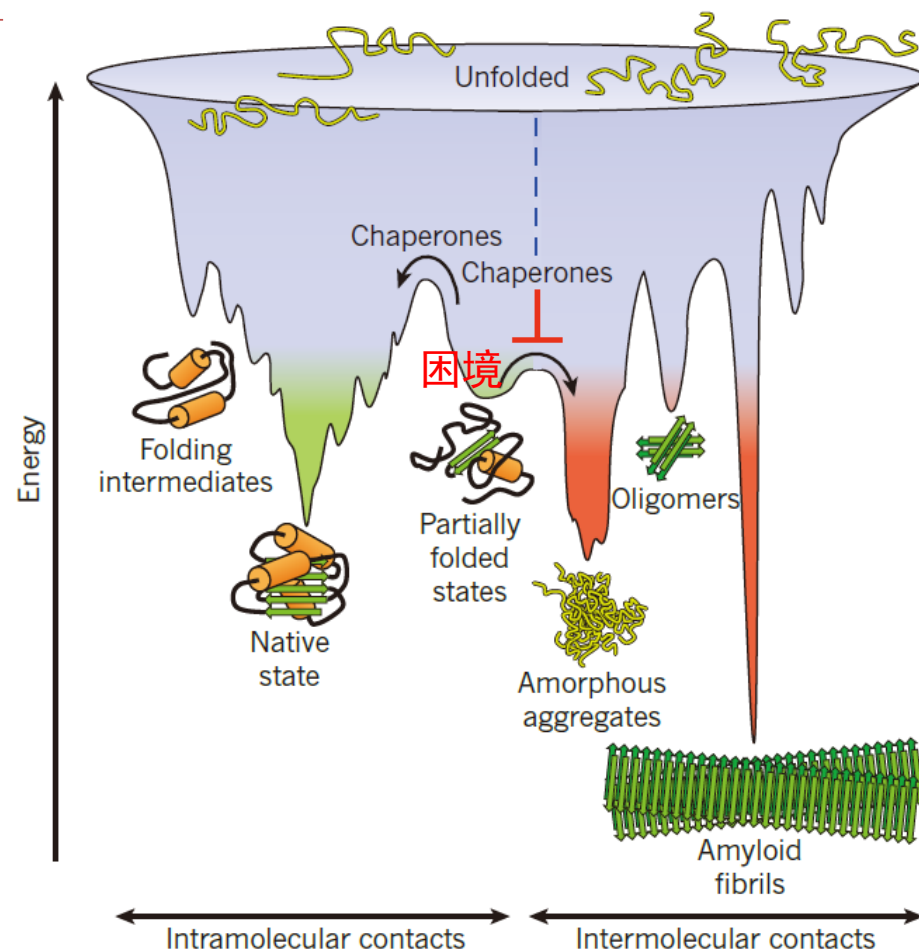
$\Delta S = \text{Positive}$

$\Delta G = \text{Negative}$

- ▶ 疏水核心（hydrophobic core）是已折叠蛋白质稳定性的关键因素。
- ▶ 疏水侧链进入蛋白中心，导致水分子熵增。
- ▶ 极性基团在疏水核心中通过氢键和盐桥稳定蛋白质。
- ▶ 螺旋和片层构成蛋白质的核心。

蛋白质折叠过程

- ▶ 安芬森规则
 - ▶ 一级序列决定高级结构
- ▶ 蛋白质的天然折叠并不总是其能量最低的构象
- ▶ 蛋白质延有限路径进行折叠
 - ▶ 100个aa长度的蛋白质经三种拉式角的完全取样，具有 3^{198} 种构象。
- ▶ “伴翻译折叠”
 - ▶ 一级序列也编码了折叠过程



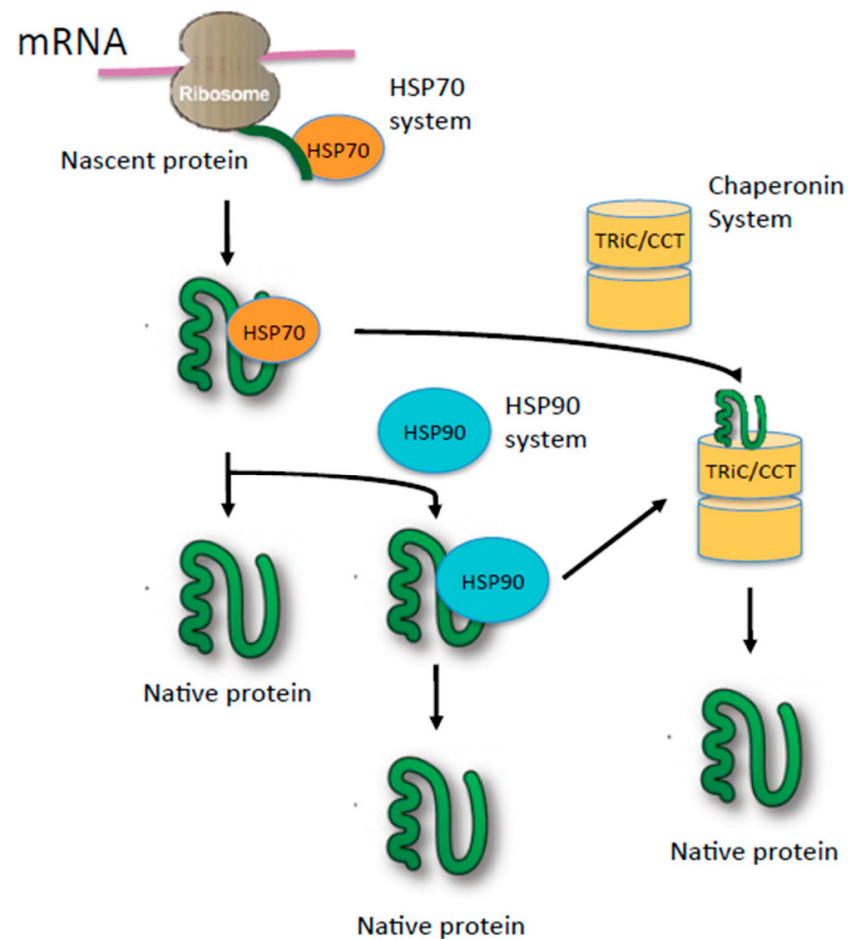
Hart et al., *Nature*, 2011.

分子伴侣 (chaperone)

- ▶ 分子伴侣对折叠类型没有特异性
- ▶ 提高折叠效率
 - ▶ 阻止蛋白聚集
 - ▶ 助力重新折叠



寻梦旅伴 2019 疯狂监护人 2011



二级结构堆积形成三级和四级结构

- ▶ 蛋白质形成天然的构象需要在二级结构形成后进行**优化堆积**
- ▶ 二级结构的堆积影响**蛋白稳定性**
- ▶ 二级结构堆积形式
 - ▶ 简单堆积（**单一**的二级结构类型），如纤维状蛋白（fibrous proteins）
 - ▶ 复杂堆积（**多种**二级结构类型协同作用），如球状蛋白（globular proteins）



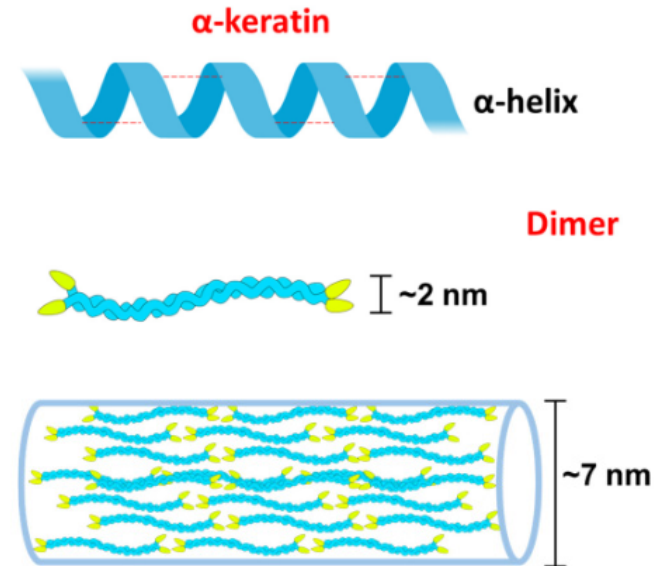
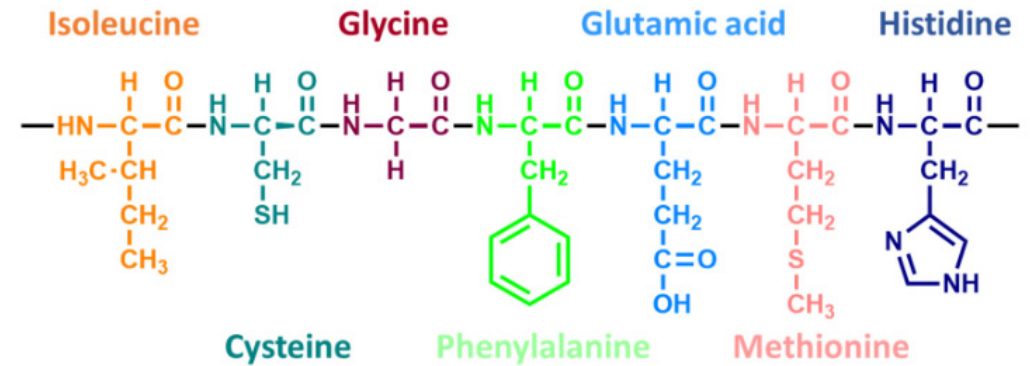
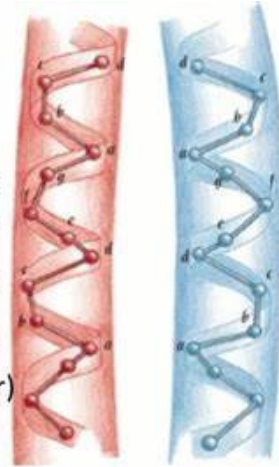
纤维状蛋白具有重复序列 (二级结构简单堆积)

▶ **α -keratin**: coiled-coil of α -helices

The heptad repeat

	a	b	c	d	e	f	g
1	Met	Lys	Gln	Leu	Glu	Asp	Lys
8	Val	Glu	Glu	Leu	Leu	Ser	Lys
15	Asn	Tyr	His	Leu	Glu	Asn	Glu
22	Val	Ala	Arg	Leu	Lys	Lys	Leu

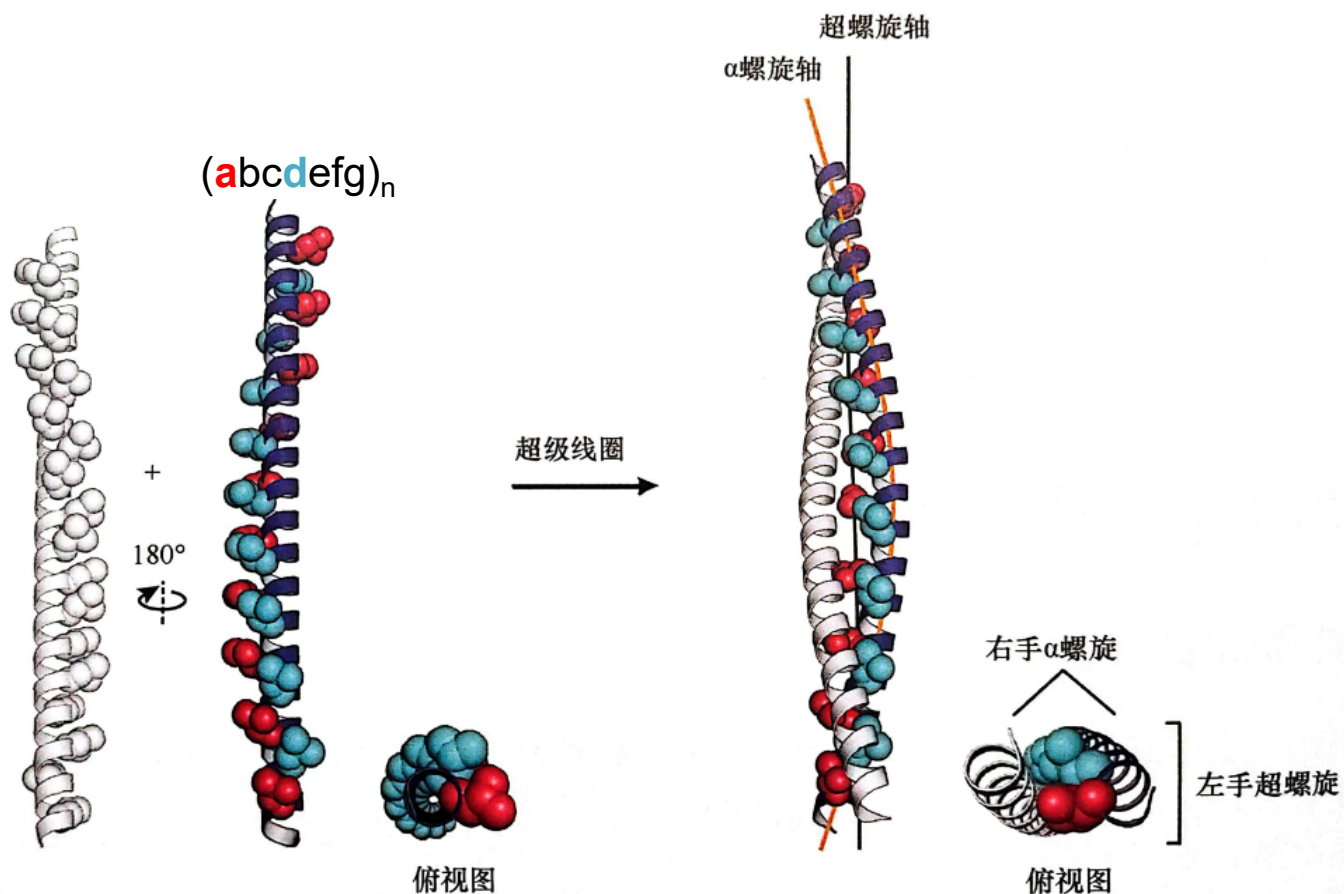
- d: Very often Leu (hence leucine zipper)
- a: often hydrophobic
- e, g: often charged
- b,c,f: charged or polar
- The above preferences are strong enough to be predicted from sequence



▶ Tahoun et al., *Biology*, 2022.

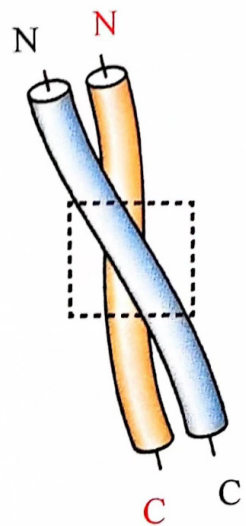
螺旋卷曲 (coiled-coil) 是蛋白质寡聚体中由 α -螺旋间形成的

- ▶ 大多由七肽重复序列组成
- ▶ α -螺旋周期3.6 aa, 使得相邻的a和d残基不能延螺旋一侧排列
- ▶ 为了补偿右手 α -螺旋中a和d残基在系统上的漂移, 超螺旋形成左手螺旋。
- ▶ 亲手验证一下

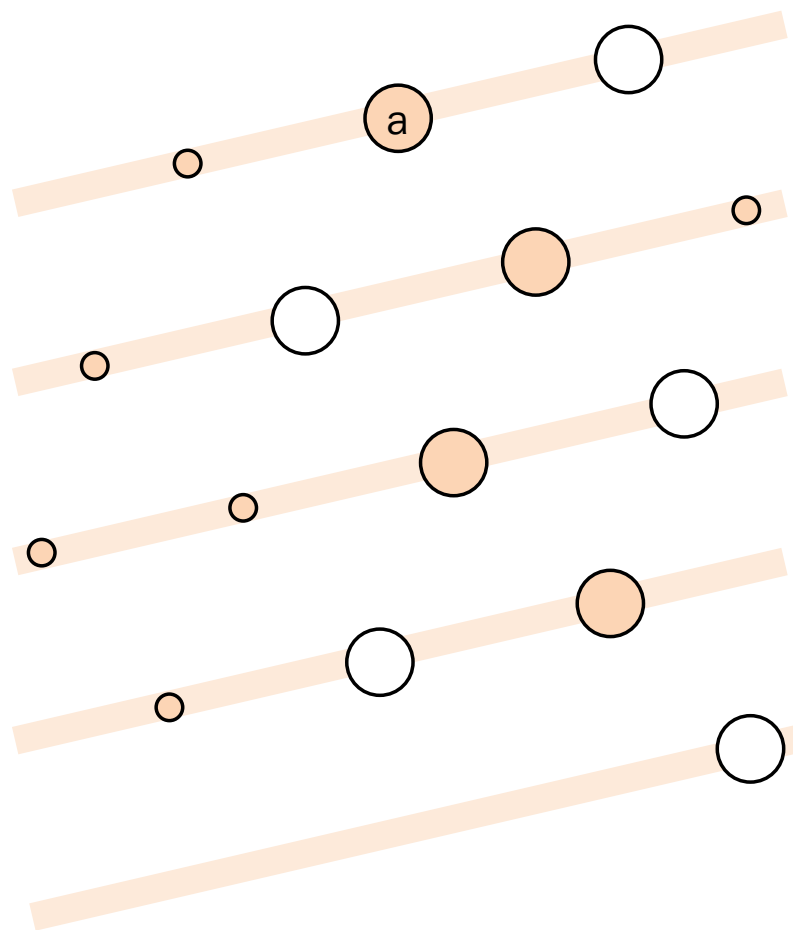
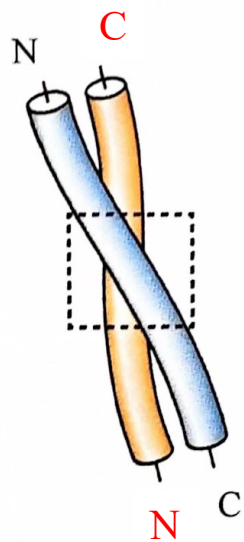


做一做：亲手制作模型并回答以下问题？

- ▶ 右手 α -螺旋形成的**平行**和**反平行**超螺旋分别是**左手**还是**右手**螺旋？

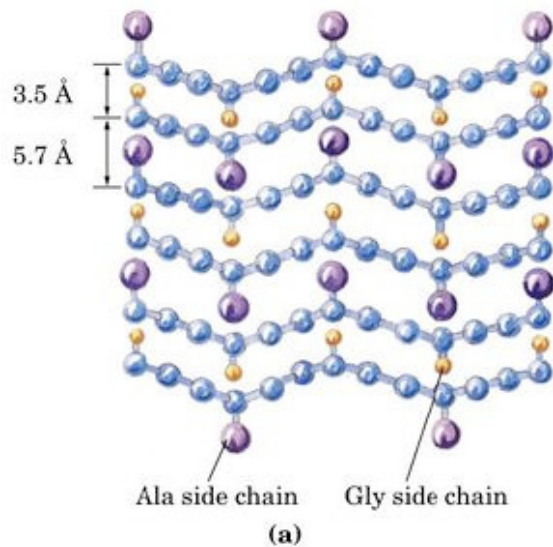


or

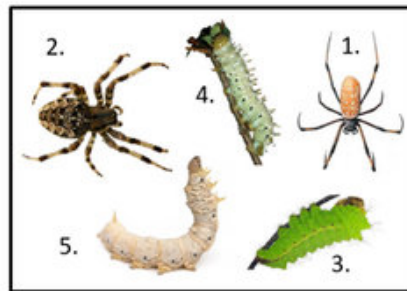


Fibrous proteins: repetitive structures

- ▶ **β -keratin**: stacked β sheets. Alternating Ala-Gly allows close packing of Ala's methyl group in silk.



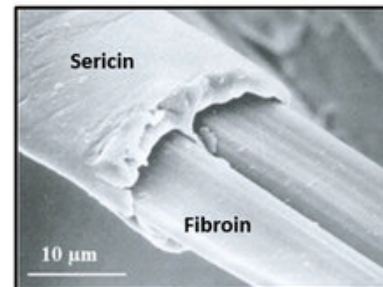
A. Popular silk sources



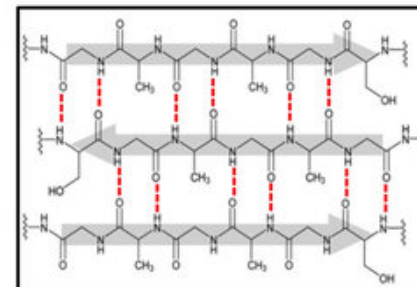
B. *B. mori* silkworm and cocoons



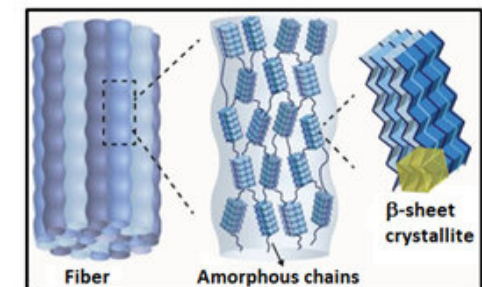
C. Composition of silk protein



D. Amino acid sequence

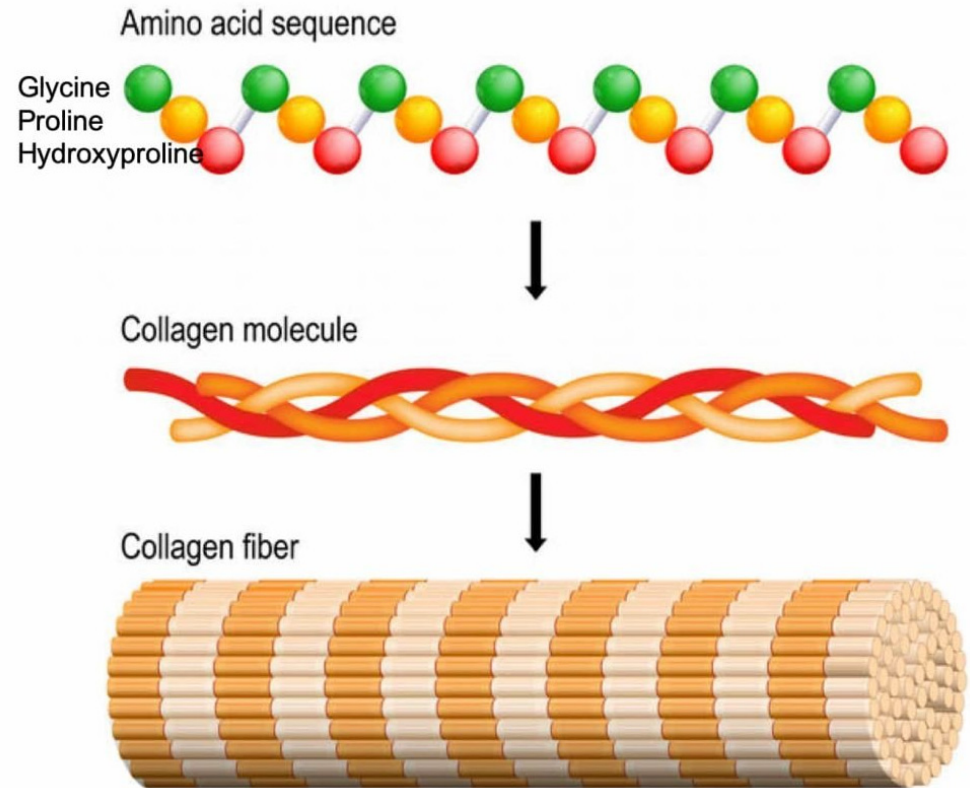
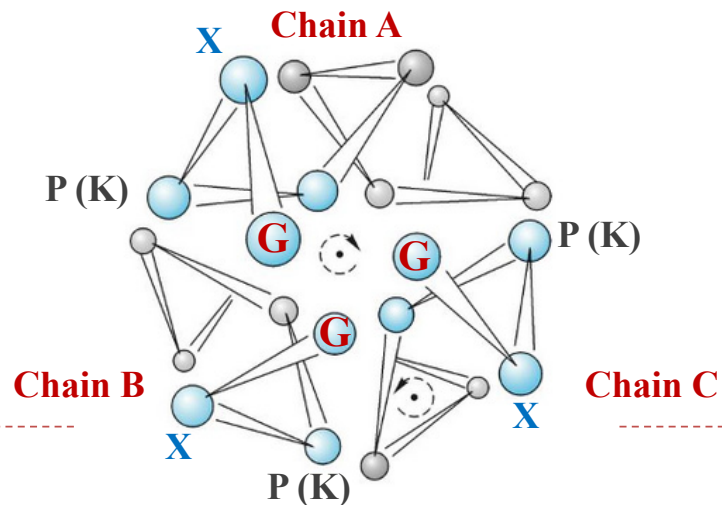


E. Semi-crystalline network

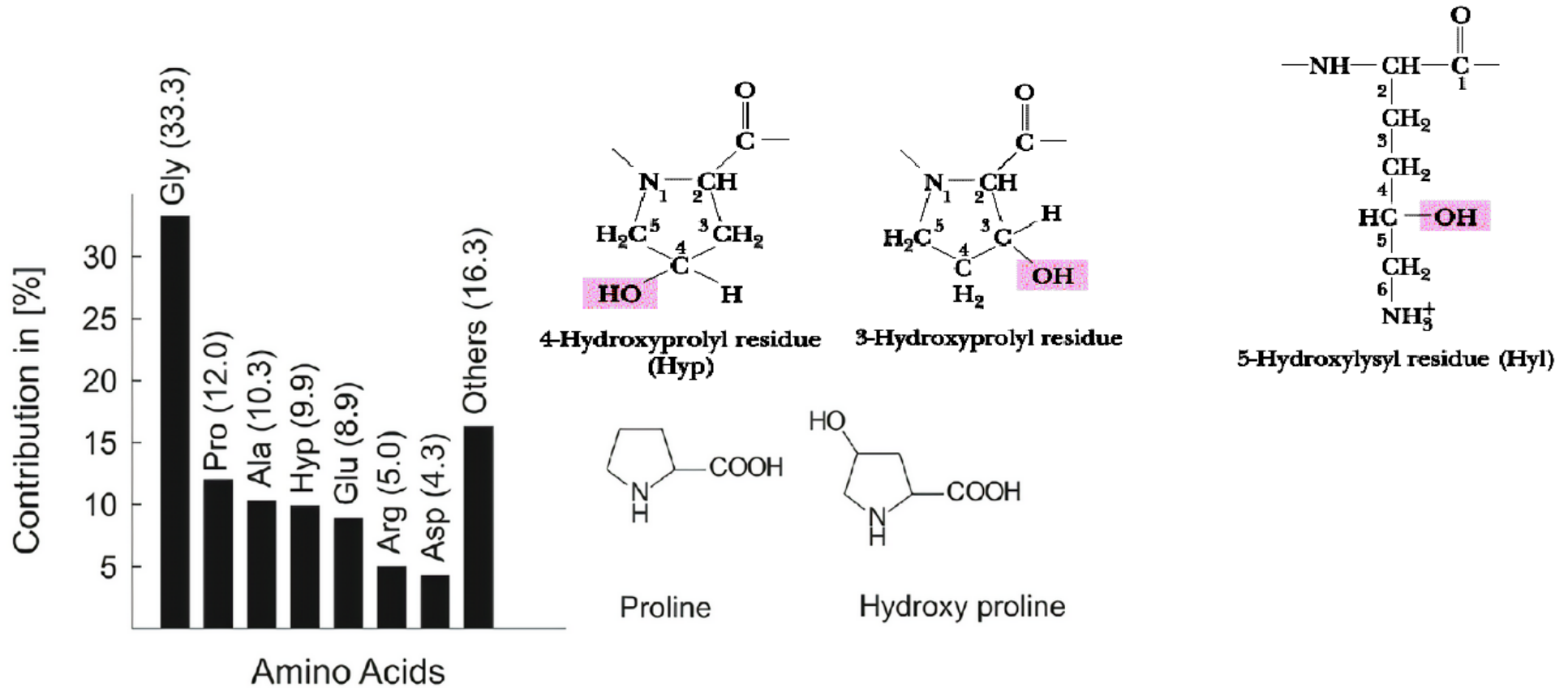


Fibrous proteins: repetitive structures

- ▶ **Collagen:** Usual sequence is repeats of Gly-X-Pro (HyPro). Three **left-handed** helices wrap around each other in a **right-handed** spiral.
- ▶ The glycines are at core of bundle (no room for any side chain). Makes a very rigid structure.

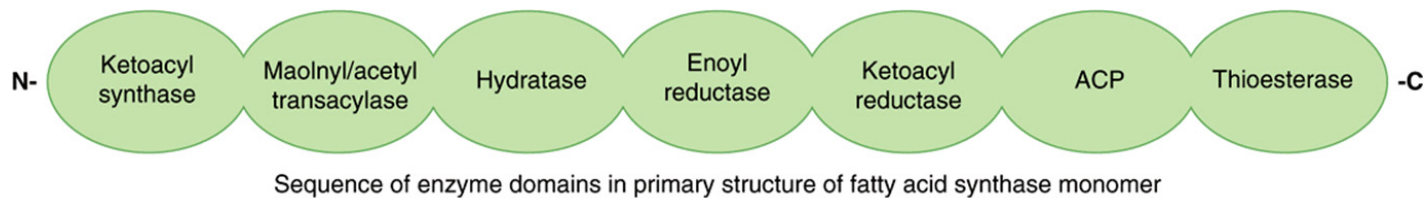


胶原中的罕见氨基酸 (Novel amino acids in collagen)

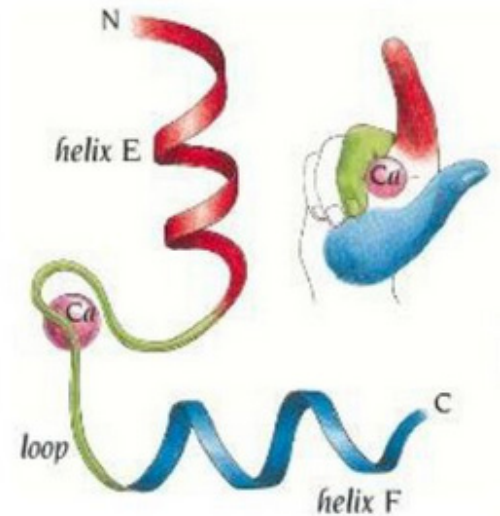


球状蛋白质的折叠（三级结构）

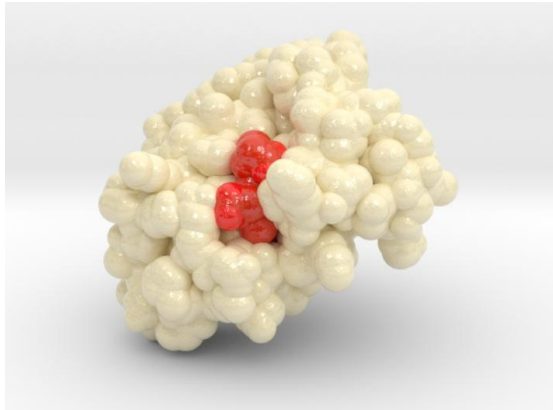
- ▶ 结构域（domain）：具有单一疏水核心的折叠单元
- ▶ 拓扑结构和模体（motif）：二级结构原件及其连接方式



脂肪酸合酶具有7个结构（功能）域

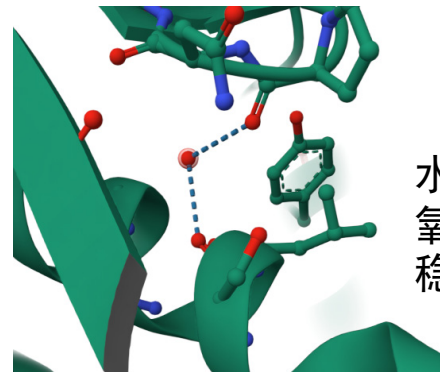


球状蛋白质中的空洞（非极性空腔）



肌红蛋白的3D打印模型

- ▶ 极性空腔会有水分子填充
 - ▶ 水分子在蛋白折叠过程中被**困**于其中，可以稳定氢键
 - ▶ 蛋白内部和外部水分子快速交换，好像蛋白质分子进行“**呼吸**”
- ▶ 疏水空洞有利于蛋白质结合其他分子或**快速变构**



水分子与主链上
氧原子形成氢键
稳定蛋白结构

Q: 请填入“高”、“中”、“低”

	嗜热酶	嗜温酶	嗜冷酶
高温下的稳定性	?	?	?
低温下的酶活力	?	?	?



	嗜热酶	嗜温酶	嗜冷酶
高温下的稳定性	高	中	低
低温下的酶活力	低	中	高



第2节 三级结构：模体和结构域

▶ 模体 (motif)

- ▶ 结构模体 (structural motif) : 二级结构的排列方式构成的**蛋白质折叠类型**, 又称为**超二级结构**
 - ▶ β 类、 $\alpha\beta$ 混合类、 α 类
- ▶ 功能模体 (functional motif) : 具有**特定功能的一级序列**, 又称为**基序**

▶ 结构域 (domain)

- ▶ 结构域和功能域基本可以互换使用

Structural Motifs

Combination of multiple secondary structures
(模体)

Functional Motifs

A short primary protein sequence for specific function
(基序)



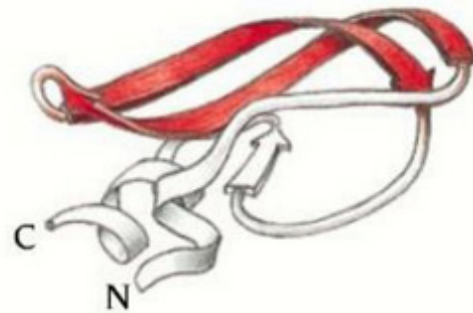
Structural motifs (units) in globular proteins

▶ **β hairpin**

- ▶ 2 adjacent antiparallel β -strands connected by a β -turn.
- ▶ Found in enzyme or receptor inhibitor.

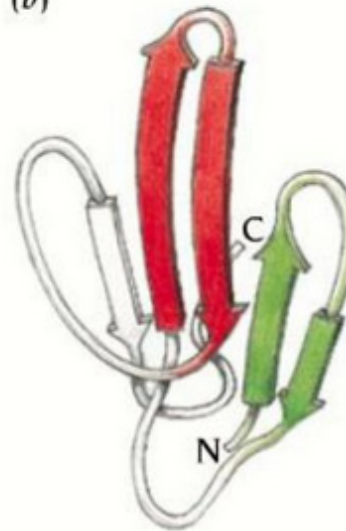
β hairpin turn or β -turn ?

(a)



Bovine trypsin inhibitor

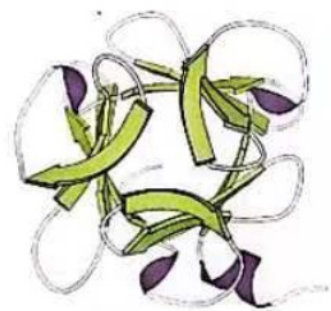
(b)



Snake venom-
erabutoxin



其他 β 类结构模体



β 三叶草 (1ybiA01)



β 正交 β -棱柱 (1b2pA00)



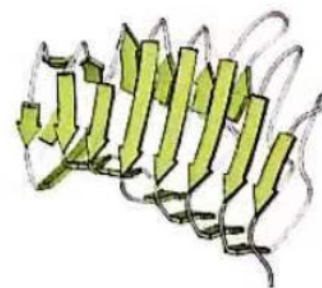
β 平行 β -棱柱 (1ouwA00)



三层 β 三明治 (1tg7A02)



β 螺旋桨 (1k3iA02)



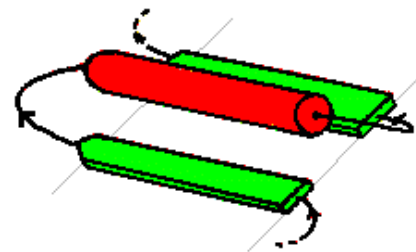
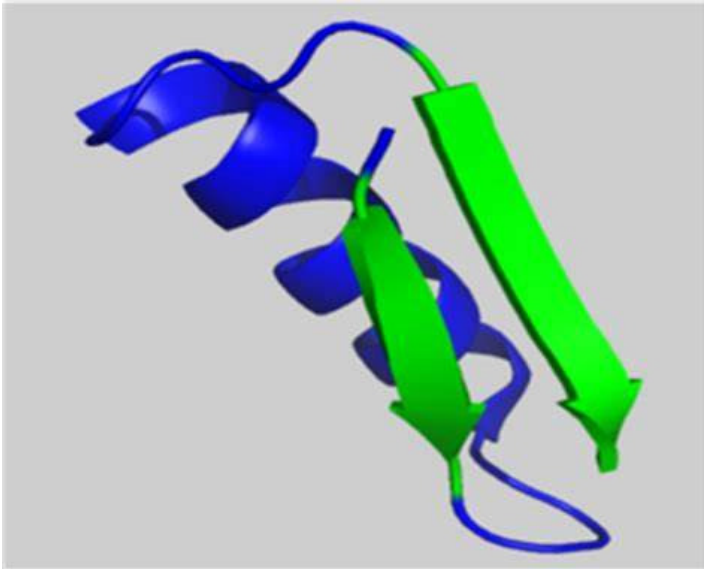
β 螺线管 (1ee6A00)



Structural motifs (units) in globular proteins

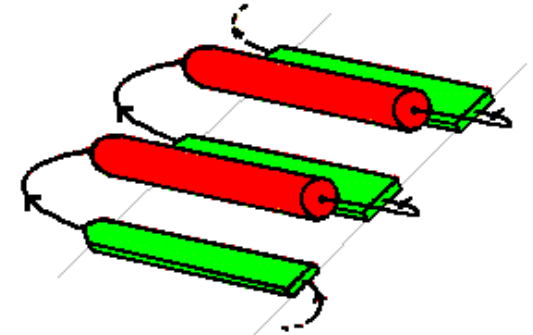
▶ β - α - β loop

- ▶ An α helix serves as the connection between 2 **parallel** β -strands.



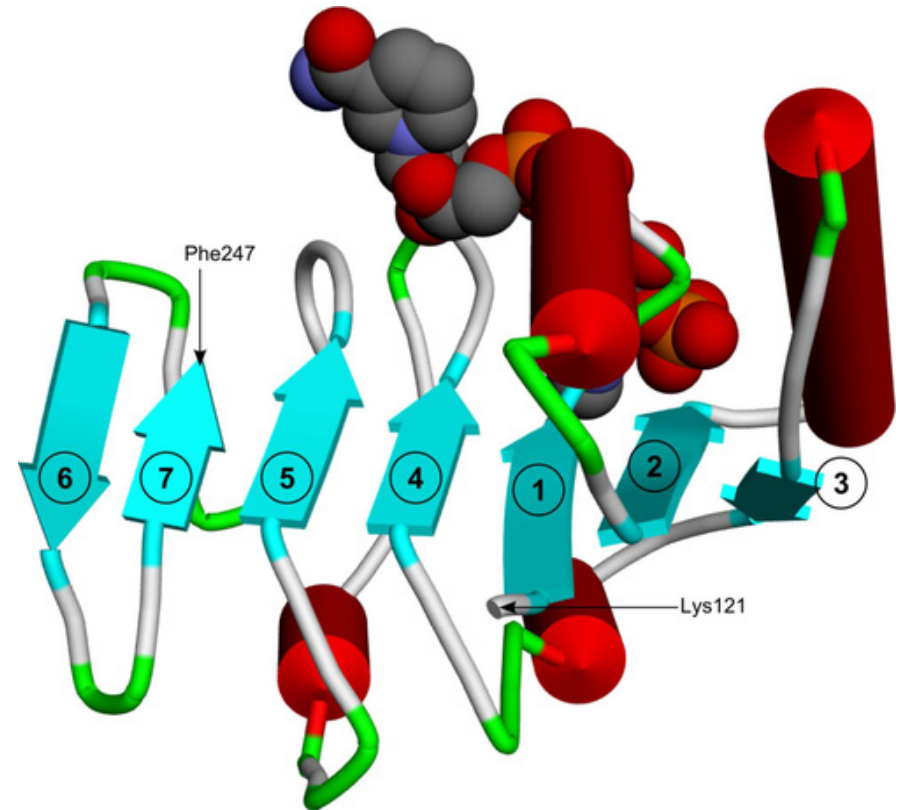
The right-handed beta-alpha-beta unit. The helix lies above the plane of the strands.

The Rossman fold



A Rossmann fold of the NADP binding site

- ▶ Rossmann fold is composed of two successively joint β - α - β loop.
- ▶ Binds with coenzyme I or II.
- ▶ Seen in dehydrogenase which requires coenzyme I or II.
- ▶ Strand 6 is in **antiparallel orientation** to the other six strands.
- ▶ Note that the helix between strand 1 and 2 is in contact with the phosphate group of the NADP.



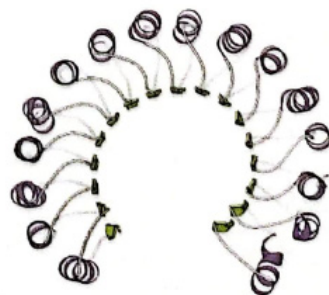
Methylene-tetrahydromethanopterin dehydrogenase
(PDB ID: 1LUA)



其他 $\alpha\beta$ 混合类结构模体



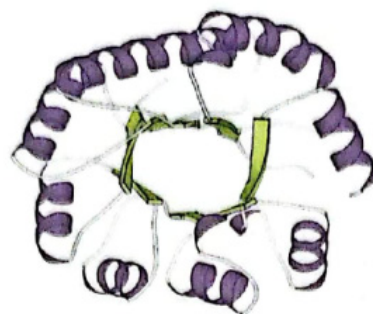
$\alpha\beta$ 盒 (1plq000)



$\alpha\beta$ 马蹄 (2bexA00)

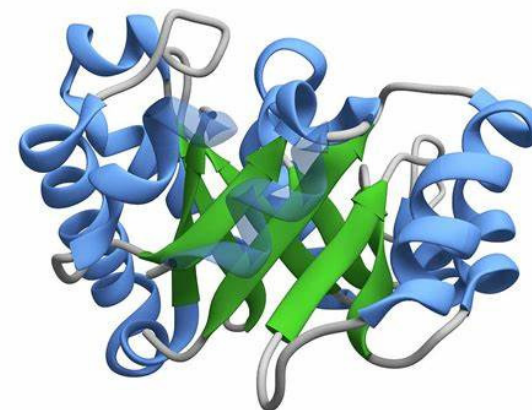
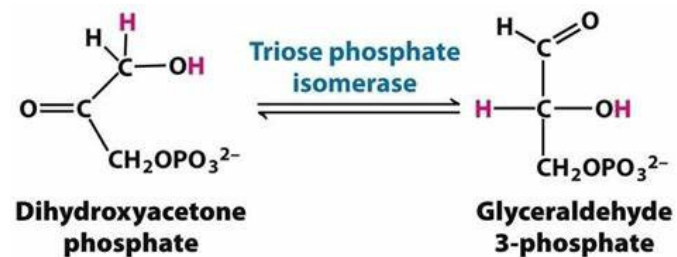


$\alpha\beta$ 卷 (laarA00)



$\alpha\beta$ 桶 (7odcA02)

TIM桶

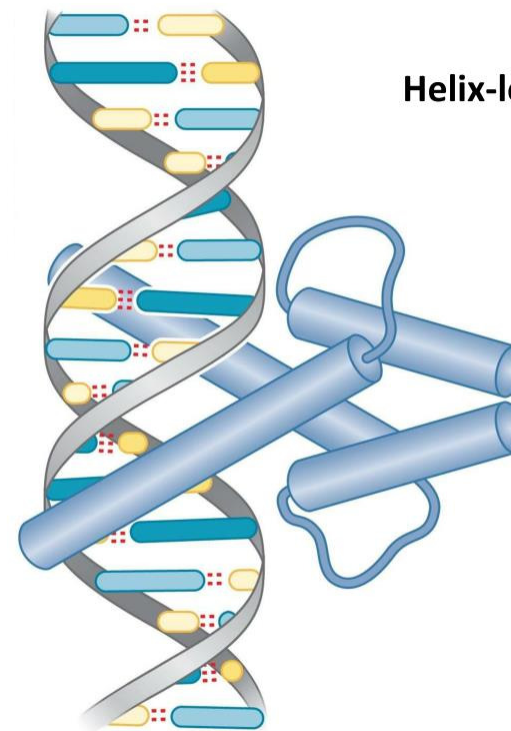
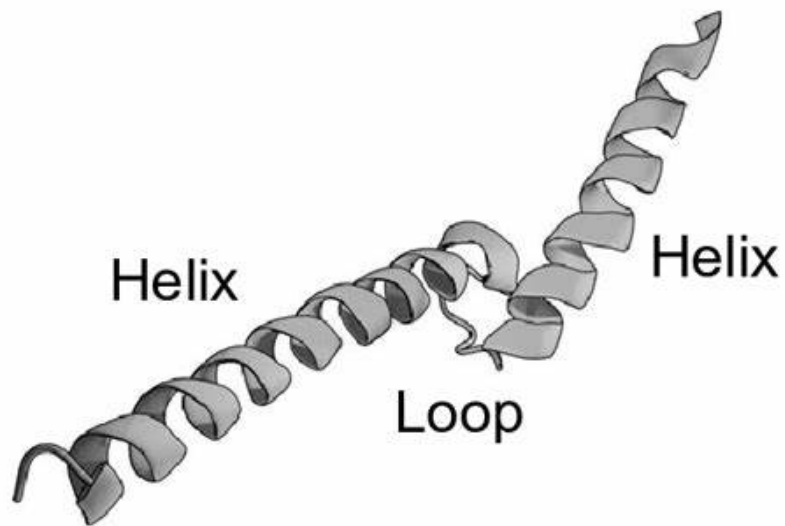


磷酸甘油醛异构酶 (α/β)₈
(triose phosphate isomerase)

Structural motifs (units) in globular proteins

▶ **Helix-loop-helix**

- ▶ 2 α -helices joined by bridge (When arranged at $\sim 90^\circ$, can be called an " α - α corner.")

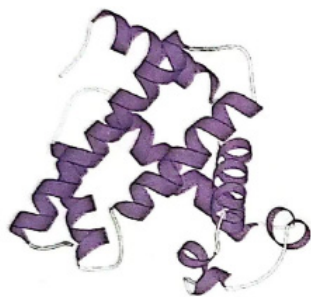


**Basic Region
Helix-loop-Helix domain (bHLH)**

Example: MyoD



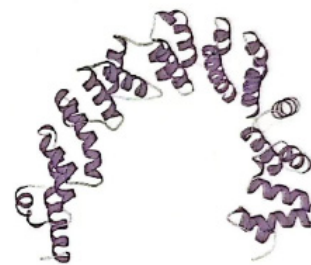
其他 α 类结构模体



α 正交束 (1mbn000)



α 上下束 (1e85A00)



α 马蹄 (1qsaA01)



α 螺线管 (1pprM01)



$\alpha\alpha$ 桶 (1fce001)



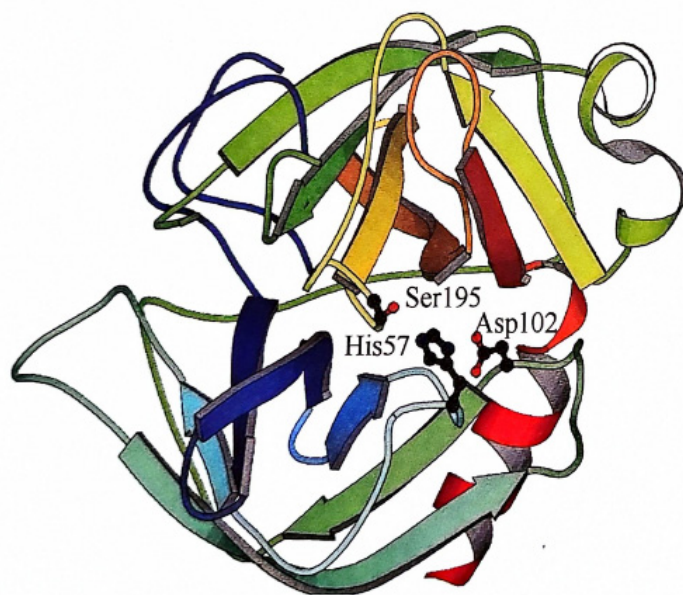
蛋白质折叠类型分类与蛋白质的功能分类无关

- ▶ 已知的有1000种以上的折叠类型
- ▶ 具有同一种类型折叠的蛋白往往功能差异很大
- ▶ 非同源蛋白可能趋同演化出类似的折叠类型

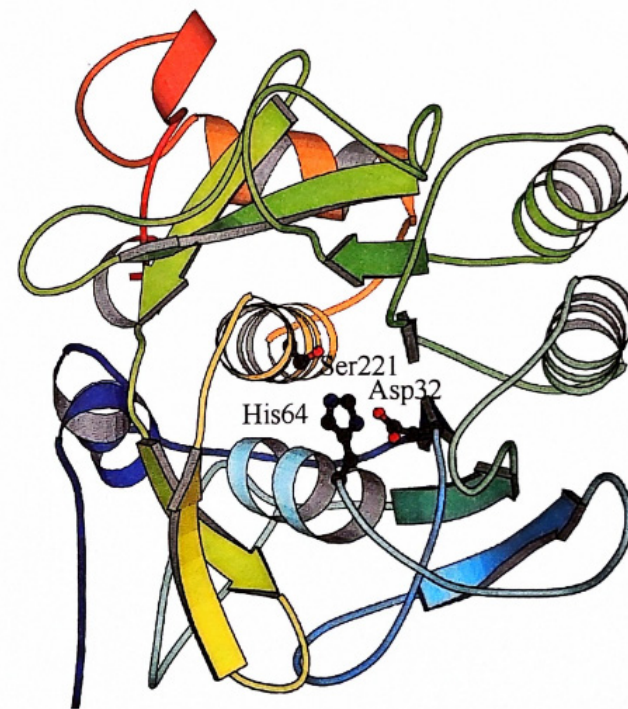
名称	折叠类型	举例
α/β 双缠绕	主要为平行片层，螺旋分列两边	Ras 折叠、枯草芽孢杆菌蛋白酶、腺苷酸激酶
TIM 桶	8 条 β 链封闭圆柱，间以螺旋相连	磷酸甘油醛异构酶、乙醇酸氧化酶、醛缩酶
劈裂 α/β 三明治	反平行片层，螺旋位于一侧	4Fe-4S 铁氧还蛋白、酰基磷酸酶、RNA 结合蛋白
免疫球蛋白	β 三明治	免疫球蛋白、受体结构域、超氧化物歧化酶
α 上下折叠	四对反平行螺旋	血红素、TMV 病毒衣壳蛋白
球蛋白	两层非平行螺旋	血红蛋白、藻青蛋白
瑞士果冻卷	β 三明治	肿瘤坏死因子、病毒衣壳蛋白、伴刀豆球蛋白
三叶形	三片层形成的圆柱	白细胞介素、蓖麻毒蛋白
泛素 α/β 卷	小片层，螺旋位于一侧	泛素、2Fe-2S 铁氧还蛋白

实例：丝氨酸蛋白酶的趋同演化

- ▶ 没有共同的起源和折叠类型，但拥有几乎完全一样的折叠类型和催化机制



糜蛋白酶 (4CHA)



枯草芽孢杆菌蛋白酶 (1GCI)



功能模体 (functional motifs) , 基序

- ▶ 具有**特定功能**的一级序列
- ▶ 比如作为某些激酶底物的磷酸化保守序列
 - ▶ Thr-Pro-Tyr motif for JNK
 - ▶ Thr-Gly-Tyr motif for p38
- ▶ 通过同源蛋白的多序列比对, 可以发现潜在的功能模体

<i>BASU</i>	53	QT	AG	RGG	MS	RV	WHS	QEG	NG	VW	MS	L	I	LR	PD	I	PL	QK	TP	QL
<i>CrBPL*</i>	32	QTR	GR	GG	LGR	DW	VD	EP	GS	AL	LIS	I	VL	RP	AVE	VSR	LP	QL		
<i>BioID2</i>	34	QTK	GR	GG	LGR	KW	LS	QEG	-	GL	YFS	F	LL	NP	KE	FEN	L	L	QL	P



结构域 (domain)

- ▶ The stable unit of protein folding and tertiary structure.
- ▶ 蛋白质的**独立折叠单元**，拥有各自的疏水核心
- ▶ 通常在100-200 aa左右
- ▶ 通过基因的**复制**和**融合**形成
- ▶ 大多数情况下蛋白质的各部分之间没有明显界限，因此对结构域的划分具有一定的**主观性**
- ▶ 结构域往往具有独自的**功能**和**相互作用**



从折叠的方式对结构域类型进行分类

- ▶ 1) All α domains
 - ▶ repeat of helix-loop-helix motif
- ▶ 2) All β domains
 - ▶ repeat of β hairpin gives anti-parallel β structures
- ▶ 3 α/β domains
 - ▶ repeat of β - α - β motif to give parallel β structures
- ▶ 4) $\alpha+\beta$ domains
 - ▶ The α -helices and β -sheets are relatively separate.



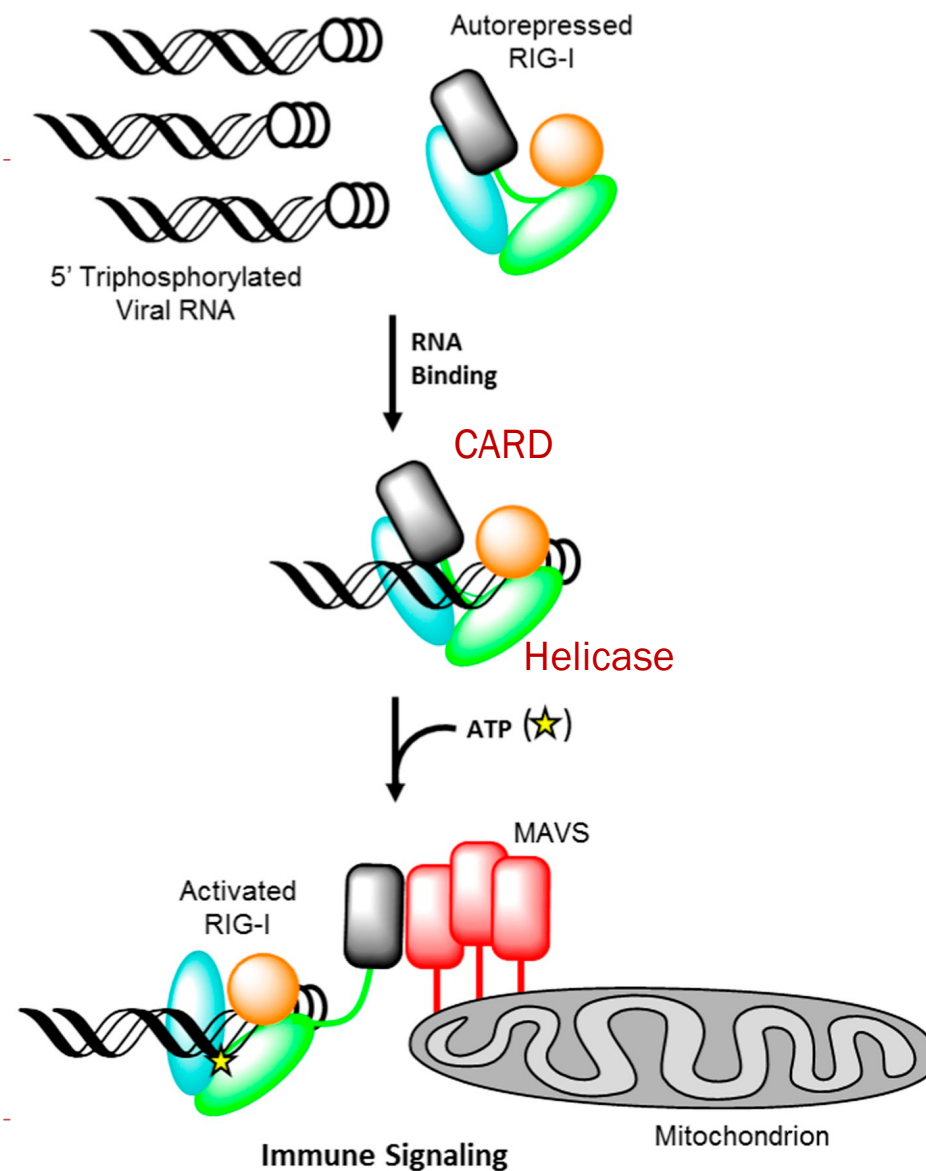
结构域存在的意义

- ▶ Separate domains allows for:
 - ▶ 1) Efficient folding
 - ▶ 2) Larger folded proteins
 - ▶ 3) Flexibility and motion
 - ▶ substrate-binding sites are often at the interface between domains
 - ▶ 4) Union of new functions

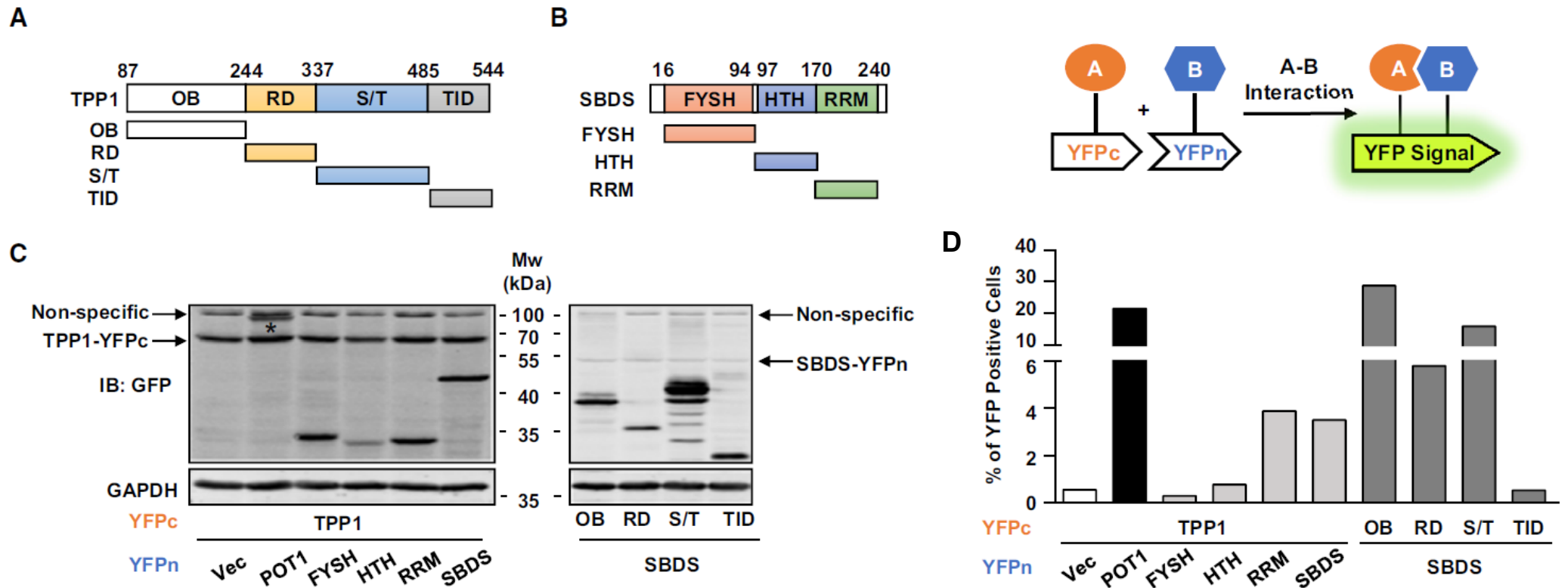


实例：胞内RNA病毒受体

- ▶ 先天免疫蛋白RIG-I在静息时通过分子内相互作用自抑制
- ▶ 当病毒RNA与Helicase结构域结合后，释放出CARD结构域激活下游信号通路



实例：SBDS和端粒体蛋白TPP1各结构域的相互作用分析



第3节 四级结构：蛋白质寡聚体的形成

- ▶ 蛋白质的**相互作用**是其形成寡聚体（四级结构）的基础
- ▶ 单体通过疏水作用等形成较为稳定的寡聚体
 - ▶ 单体相同：homo-oligomer
 - ▶ 单体不同：hetero-oligomer
- ▶ 稳定复合物的平均相互作用面积在1600平方埃（ \AA^2 ）以上



寡聚蛋白的排列方式

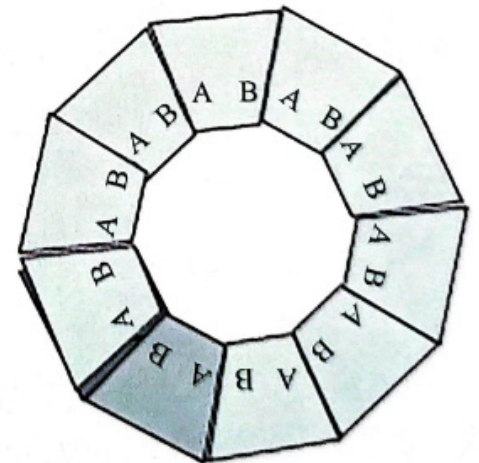
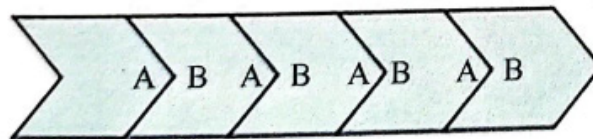
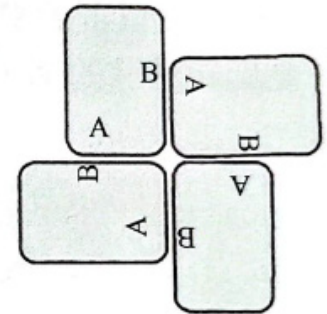
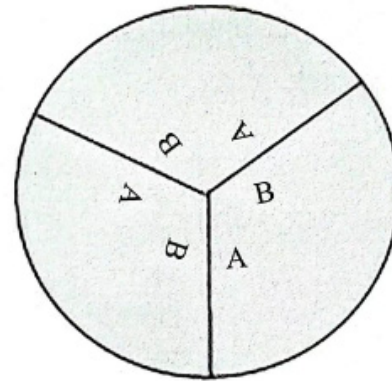
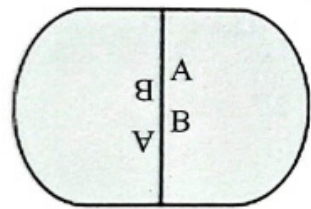
▶ 封闭排列

▶ 具有对称性

▶ 开放排列

▶ 线性

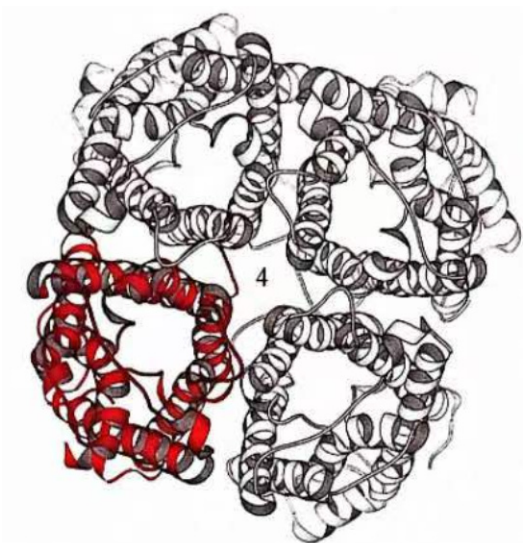
▶ 螺旋型：肌动蛋白、纤毛蛋白



周期对称聚集的蛋白质

▶ 简单对称轴多聚蛋白

- ▶ 比如水通道蛋白 (aquaporin)
具有**四重旋转轴对称**



旋转对称	举例
2	乙醇脱氢酶
3	膜孔蛋白、流感病毒凝血蛋白
4	流感病毒唾液酸苷酶、水通道蛋白
5	AB ₅ 肠毒素
6	C-藻蓝蛋白
7	GroES
8	捕光复合物 2
9	捕光复合物 2
11	Trp RNA 结合减弱蛋白 (TRAP)
16	捕光复合物 1
17	TMV 衣壳蛋白环
39	穹窿核糖核蛋白

具有复杂多重对称轴的蛋白复合体

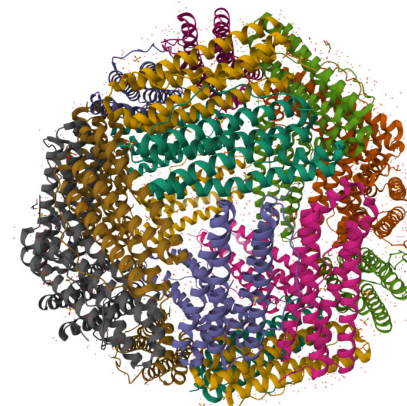
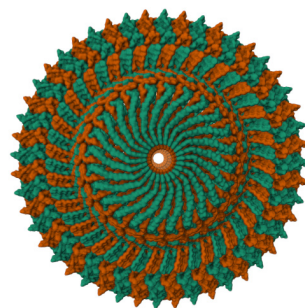
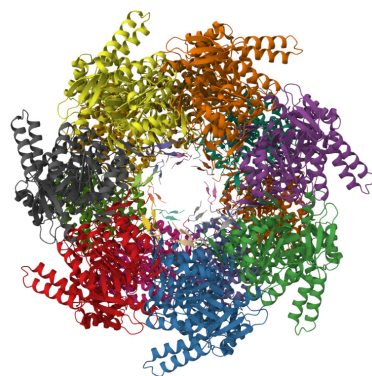
▶ 如222，表示在空间X、Y和Z方向各有一个二重对称轴

对称性类型	举例	说明
222	乳糖脱氢酶	4 个亚基
32	天冬氨酸甲氨酰转移酶	两种 12 个亚基
422	蚯蚓血红蛋白、乙醇酸氧化酶	8 个亚基
622	谷氨酰胺合成酶	12 个亚基
72	GroEL	14 个亚基
23	蛋白酶 3,4-双加氧酶	四面体对称, 12 个亚基
39/2	穹窿核糖核蛋白	78 个亚基
432	铁蛋白	八面体对称, 24 个亚基
532	正二十面体病毒	二十面体对称, 60 个或者 60 的倍数个亚基

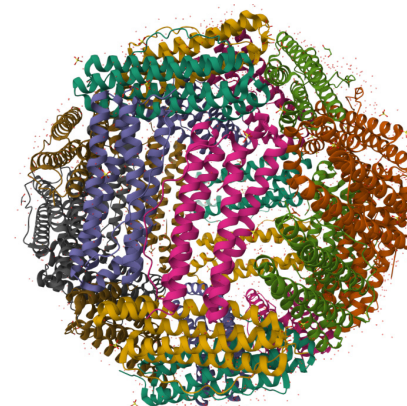
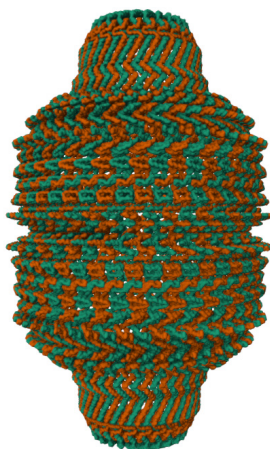
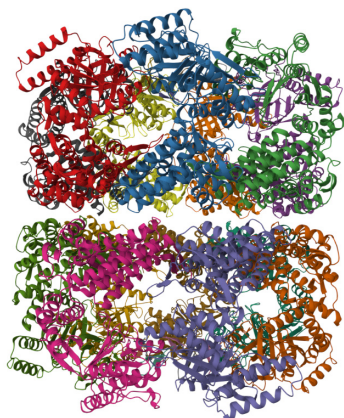


实例：GroEL (72)、穹窿体 (39/2) 和铁蛋白 (432)

俯视



侧视



GroEL (1MNF)

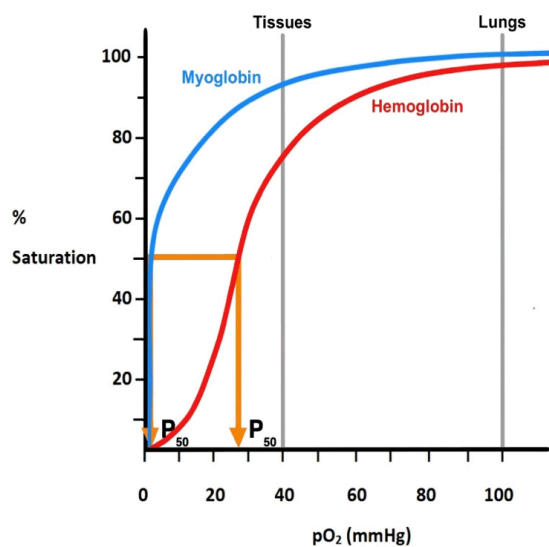
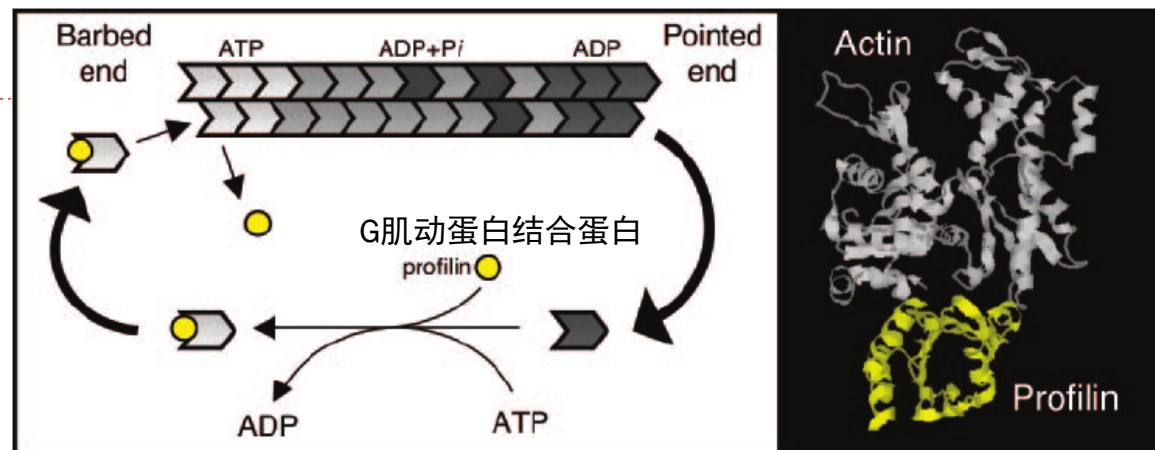
the Vault (2QZV)

Ferritin (1VLG)

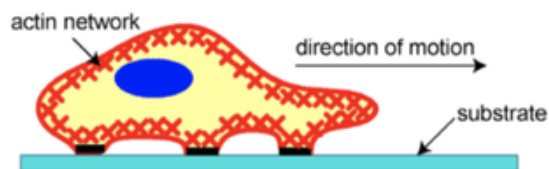
内腔直径75Å，可容下4000个铁原子

蛋白多聚化的意义

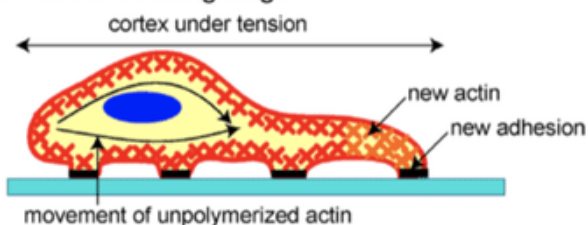
- ▶ 提高蛋白质的稳定性
- ▶ 动态调节蛋白功能
 - ▶ 别构效应（如血红蛋白）
 - ▶ 细胞骨架微管和微丝
 - ▶ 病毒衣壳蛋白的动态组装



1) Protrusion of the Leading Edge



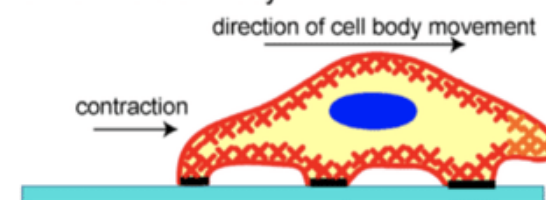
2) Adhesion at the Leading Edge



Deadhesion at the Trailing Edge

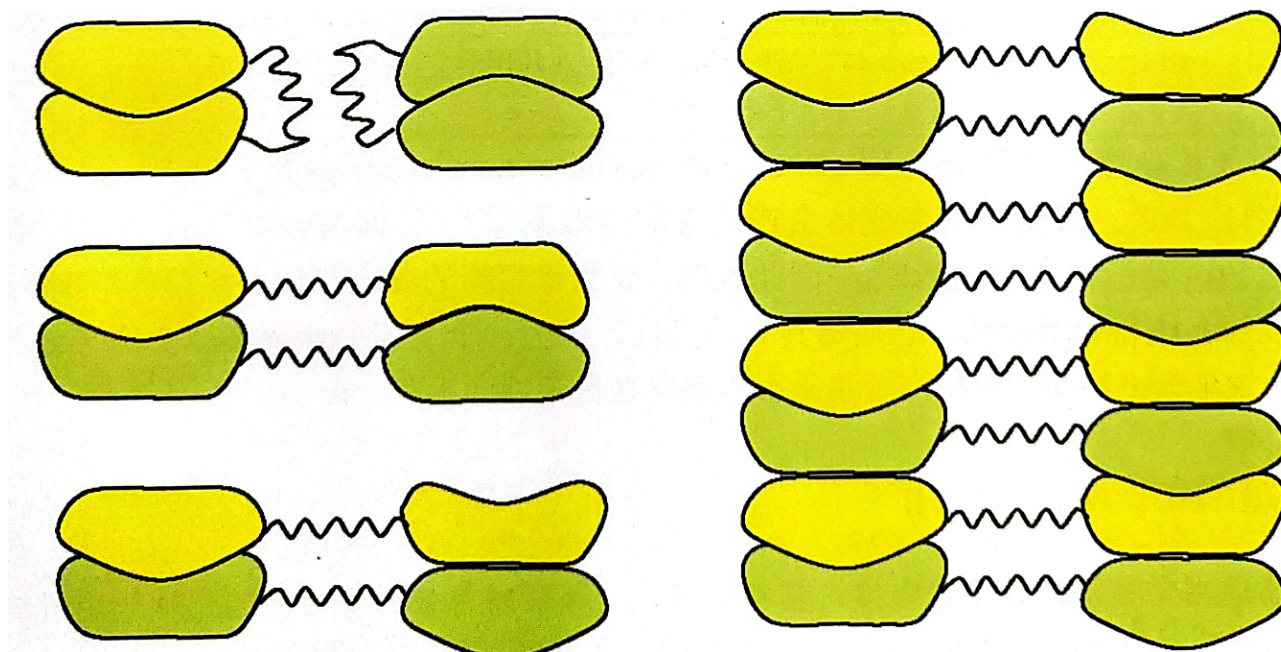


3) Movement of the Cell Body

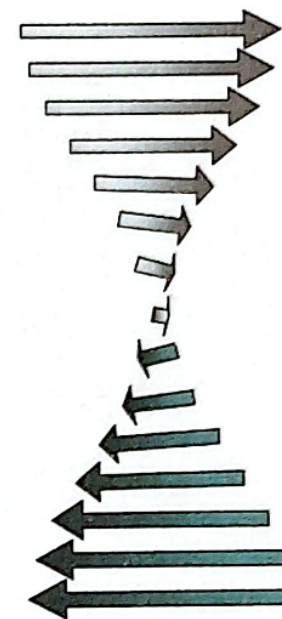


蛋白质的非天然聚集：结构域互换 (domain swap)

- ▶ 在某些条件下，天然的分**子内互作**变为**分子间互作**，形成线性多聚体
- ▶ 往往是疾病产生的原因（如淀粉样变性疾病）



失控的结构域互换形成蛋白纤维样聚集



聚集成为平行 β 结构的蛋白

第4节 蛋白质高级结构的获取方法

▶ 实验方法

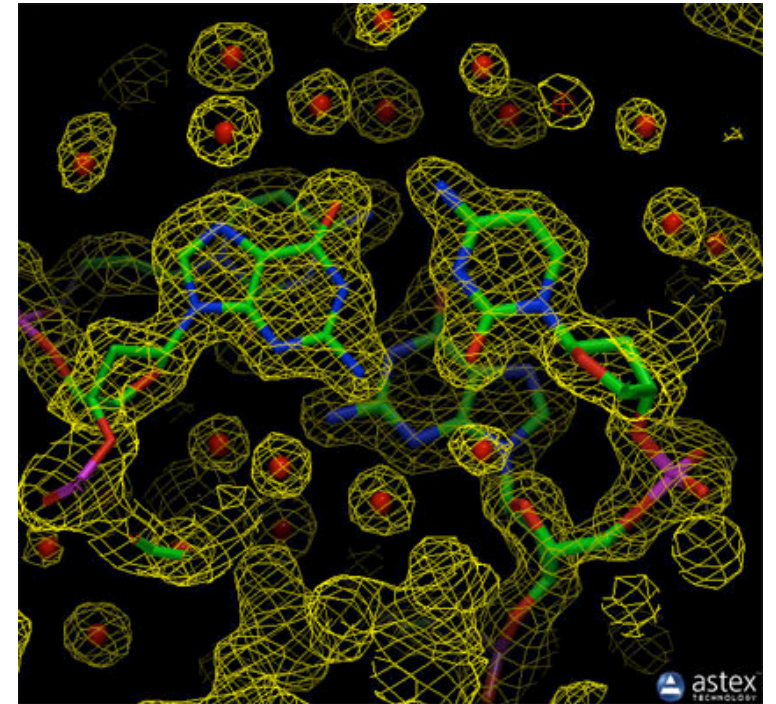
- ▶ X射线晶体学 (X-ray Crystallography)
 - ▶ 自由电子激光(free electron laser, FEL)
- ▶ 核磁共振波谱 (NMR Spectroscopy)
- ▶ 三维电镜 (3D Electron Microscopy)
- ▶ 集成建模 (Integrative Modeling)

▶ 预测方法



X射线晶体学 (X-ray Crystallography)

- ▶ 纯化蛋白、结晶、衍射、解析结构
- ▶ 优点
 - ▶ 可以提供**非常详细的原子信息**，显示蛋白质或核酸中的每个原子，以及配体、抑制剂、离子和其他结合到晶体中的分子的原子细节。
- ▶ 缺点
 - ▶ 需要**高质量晶体**
 - ▶ 蛋白质的**柔性部分**在晶体学电子密度图中通常是不可见的



Electron density from a structure of DNA
(PDB entry 196d)



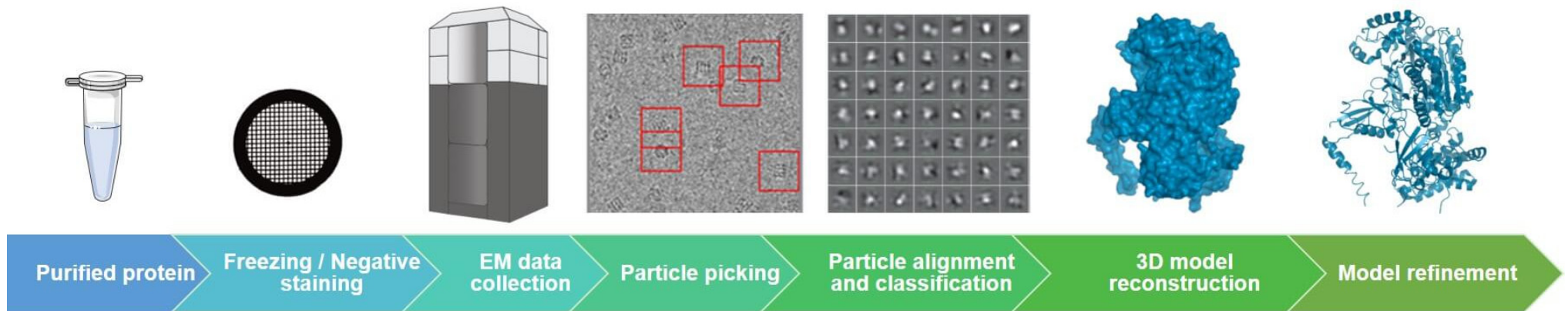
核磁共振波谱 (NMR Spectroscopy)

- ▶ 蛋白质被纯化，置于强磁场中，根据观察到的共振，以给出彼此接近的原子核位置，并分析蛋白的构象。
- ▶ 优点
 - ▶ 研究柔性蛋白质原子结构的首要方法
 - ▶ 提供了溶液中蛋白质的信息，而不是锁定在晶体中的蛋白质
- ▶ 缺点
 - ▶ 大蛋白质在NMR波谱中存在重叠峰，因此该技术目前仅限于较小蛋白。



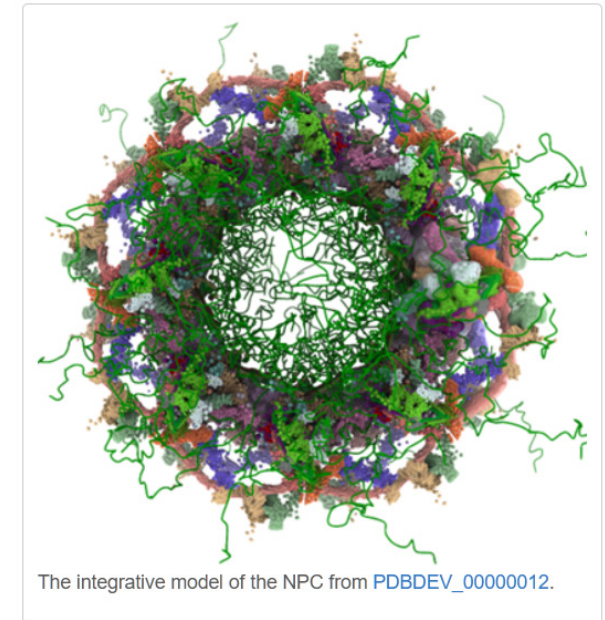
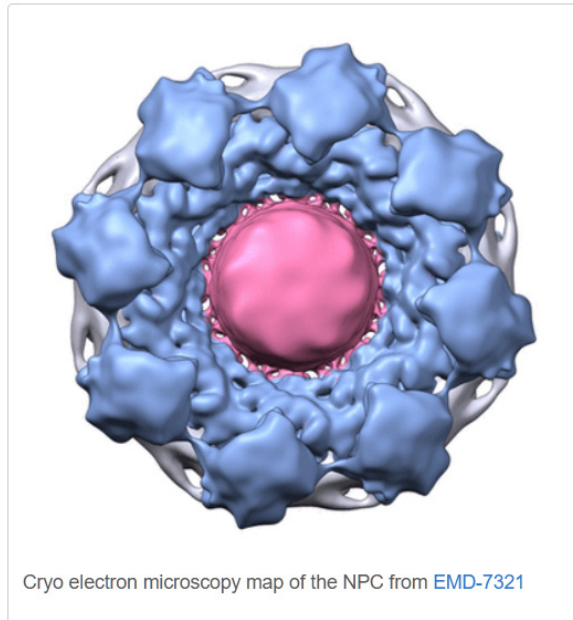
三维电镜 (3D Electron Microscopy)

- ▶ 冷冻电子断层扫描法 (Cryo-electron tomography)
 - ▶ 细胞或组织均可，样本旋转角度对电子束的透过有要求
 - ▶ <https://www.thermofisher.cn/cn/zh/home/electron-microscopy/life-sciences/cryo-tomography/workflow.html>
- ▶ 单颗粒分析成像法 (Single particle analysis, SPA)
 - ▶ 分辨率高，辐射影响小，但需纯化蛋白



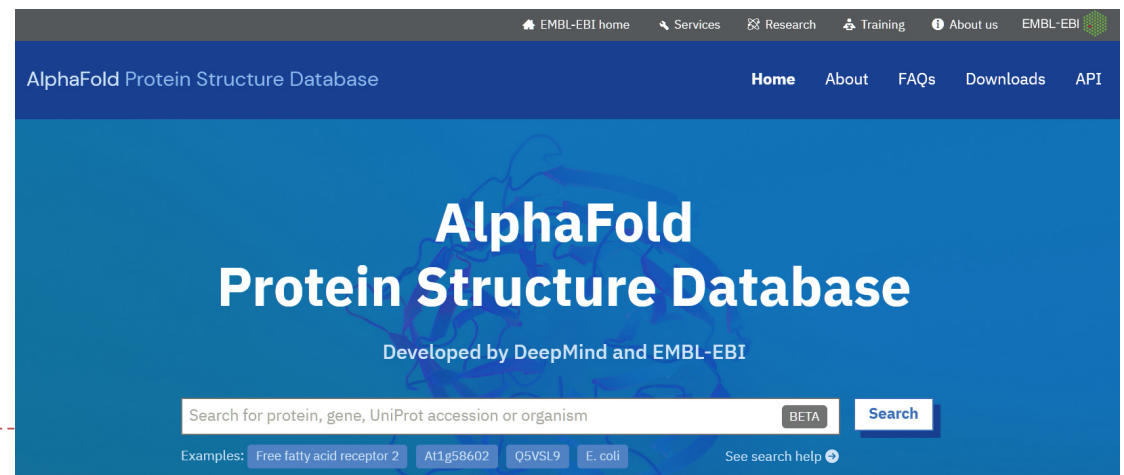
集成建模 (Integrative Modeling)

- ▶ 将来自各种方法的信息结合起来，每种方法都有助于研究系统的特定方面，以创建组装的全貌。
- ▶ 比如使用化学交联、小角散射和冷冻电镜的数据确定出芽酵母的核孔复合物（NPC）的结构。



蛋白质三维结构的预测

- ▶ Try to predict structure based on the primary sequences, using
 - ▶ known structures
 - ▶ tendencies of amino acids
 - ▶ energy minimization
- ▶ Method of the Year 2021: Protein structure prediction
- ▶ Enhancing AlphaFold-Multimer-based Protein Complex Structure Prediction with MULTICOM in CASP15



第2章 重点总结

- ▶ 第1节 蛋白质的稳定性及其动力学特性
 - ▶ 蛋白质折叠的动力与过程
 - ▶ 二级结构堆积特征
- ▶ 第2节 三级结构：模体和结构域
 - ▶ 模体与蛋白质折叠类型
 - ▶ 结构域的功能
- ▶ 第3节 四级结构：蛋白质寡聚体的形成
 - ▶ 蛋白多聚化的类型与意义
- ▶ 第4节 蛋白质高级结构的获取方法
 - ▶ 实验方法和信息学方法



