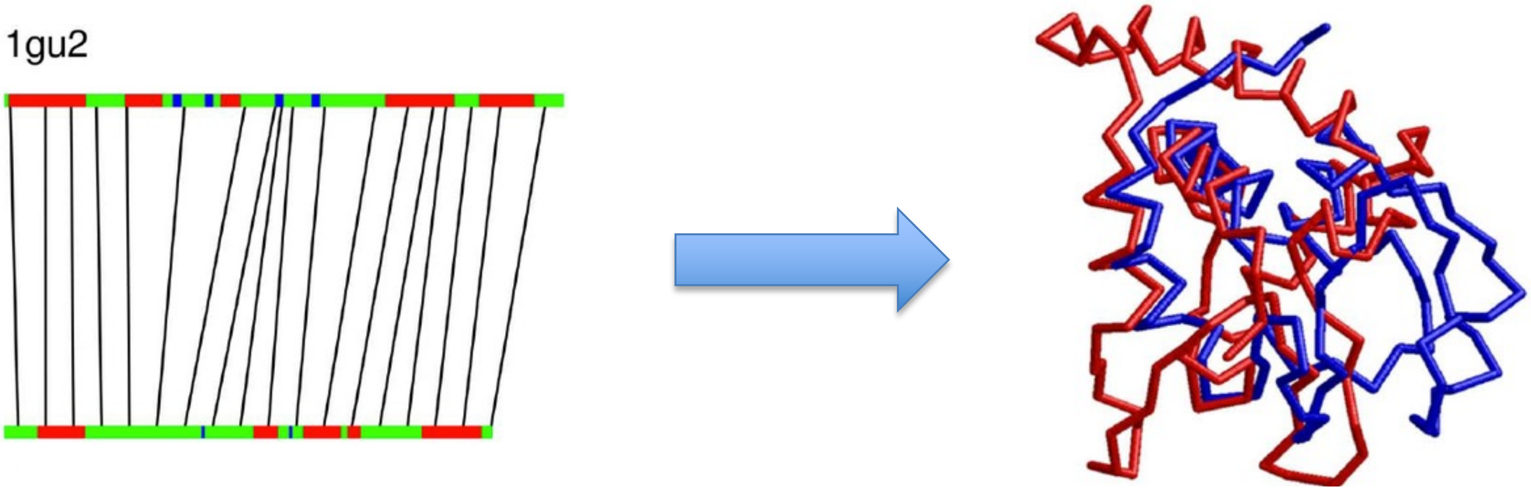
Protein Structure Alignment

Prof. Sorin Istrail
May 2024

Finding “Truth” in Molecular Mess

Introduction



- Until now we have talked about **2D sequence alignment**
- In general, the concept of alignment is a powerful technique for analyzing a variety of data
- Both DNA and proteins have **3D structure** that can be analyzed.

Introduction

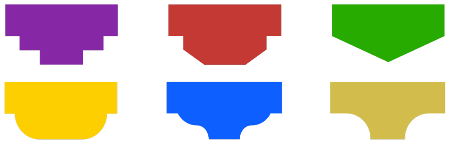
- In this lecture we will focus on **protein structure alignment**
- This will include:
  - **Motivation** for doing structure alignment
  - **Biology** of protein structure
  - **Examples** of alignment

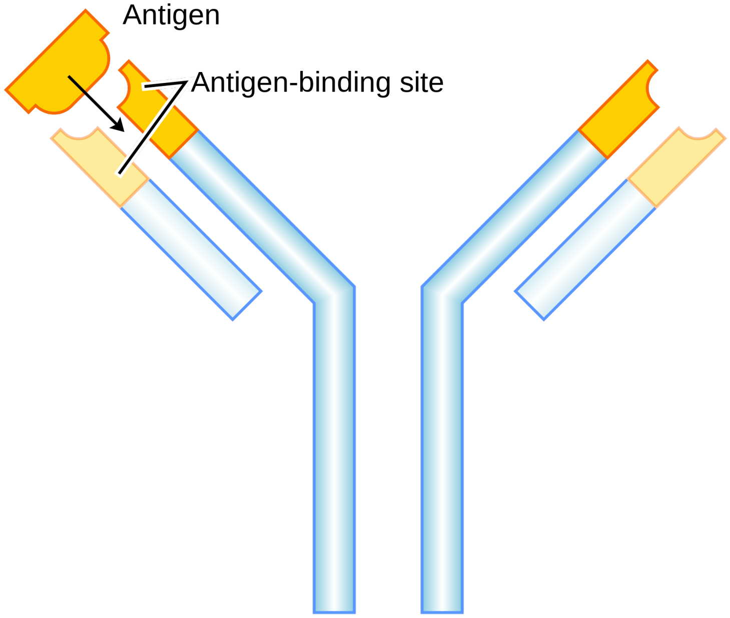
Motivation for Protein Structure Alignment

Protein **structure** determines **function**

- Sequence similarity isn't the whole picture
- Showing that two proteins are folded similarly has real biological significance ("truth")
- Ex. antibodies are proteins that recognize antigens based on shape

Antigens



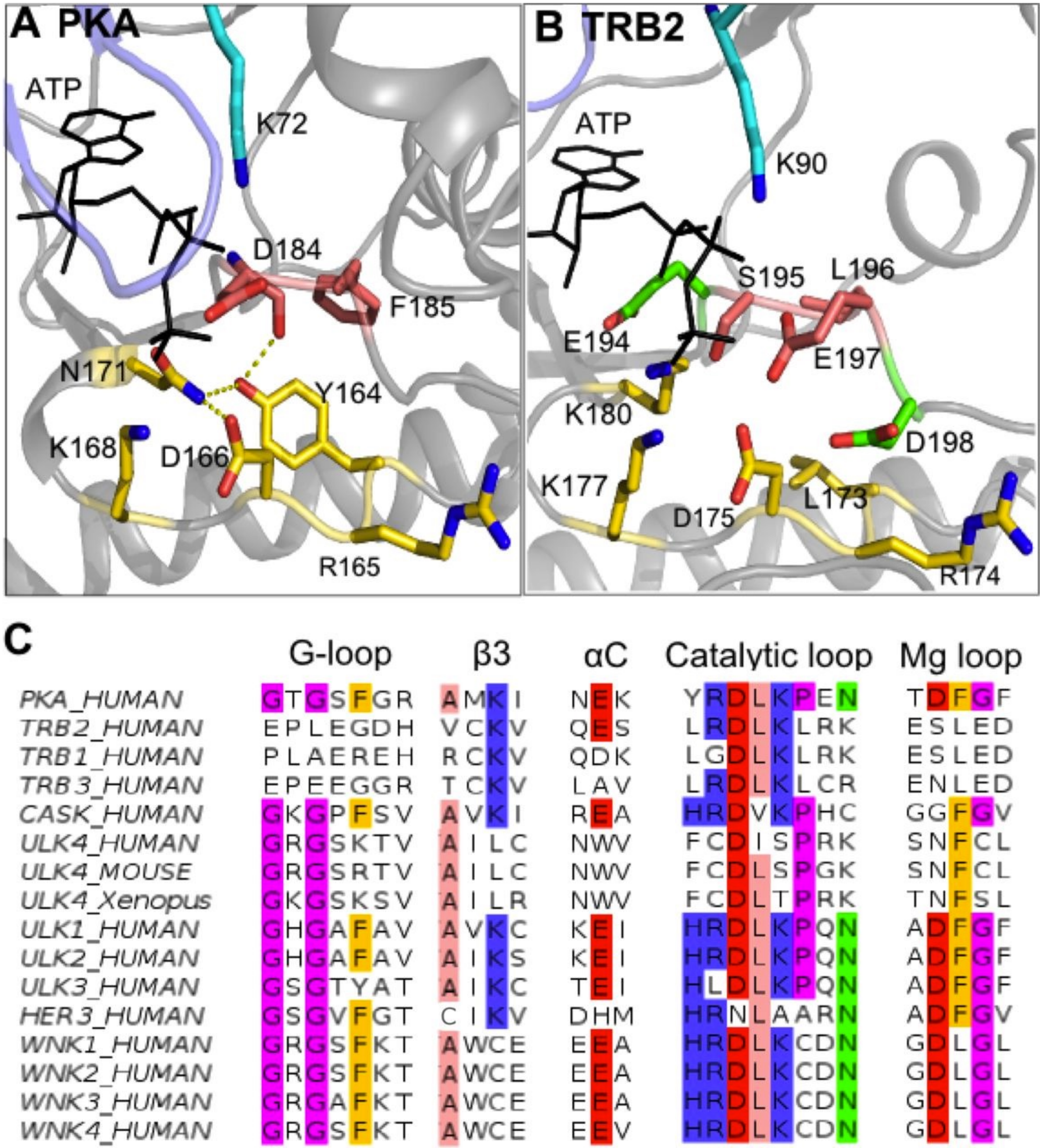


Antibody

Motivation for Protein Structure Alignment

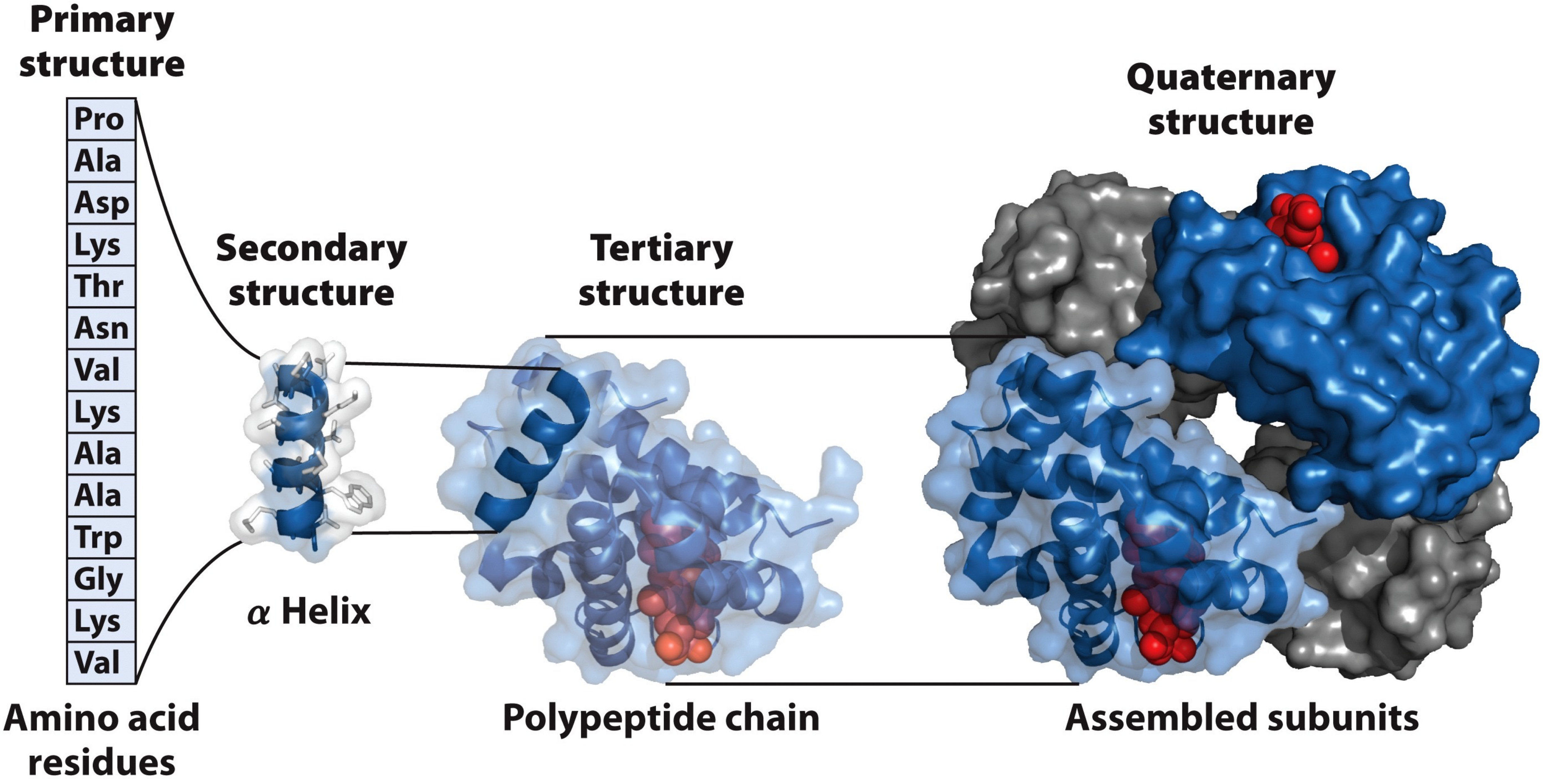
Evolutionary Argument

- Secondary and tertiary structures are more conserved during evolution than the specific amino acid sequence (*Branden and Tooze*)



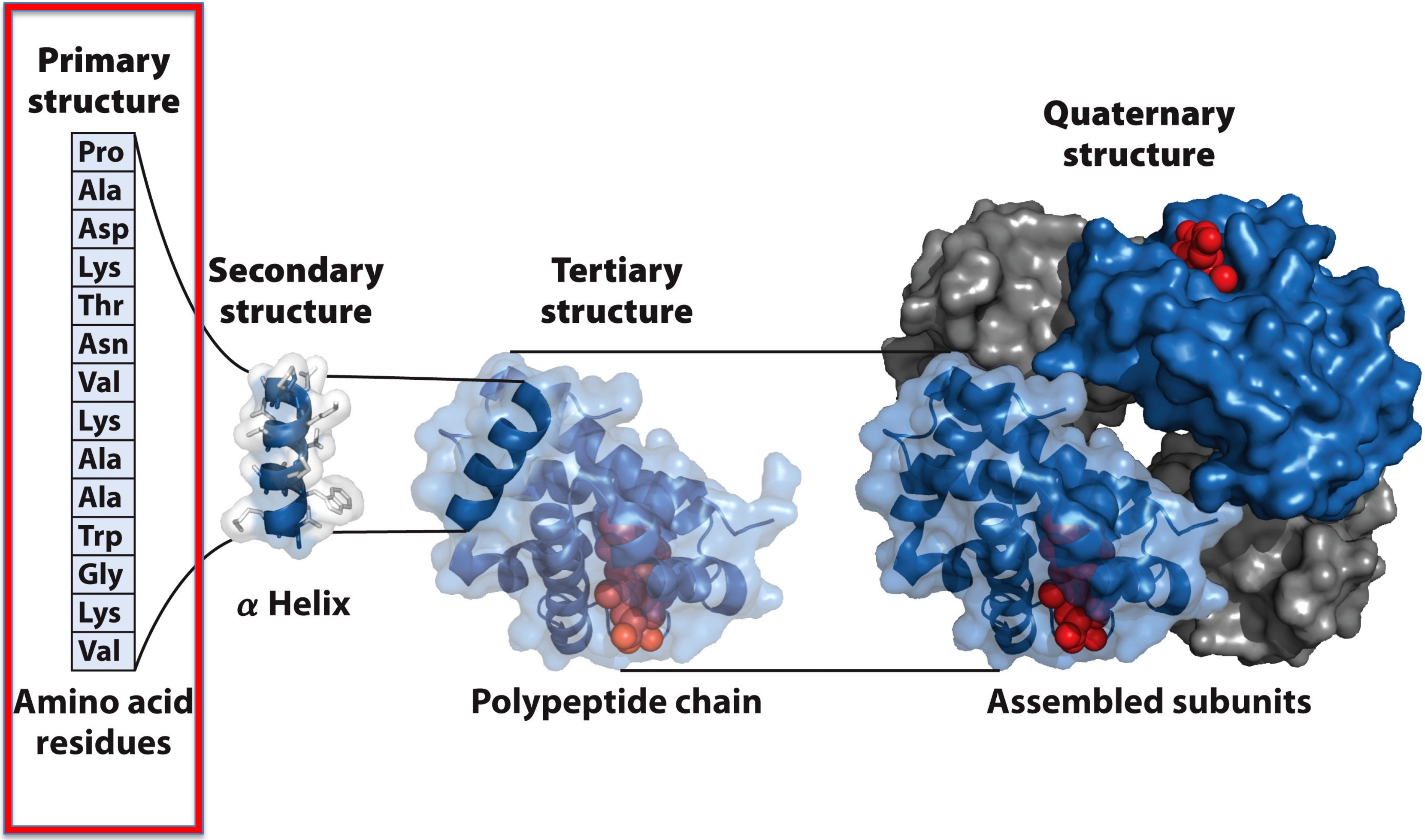
	G-loop	β3	αC	Catalytic loop	Mg loop
PKA_HUMAN	GTGSFGR	AMKI	NEK	YRDLKPEN	TDFGF
TRB2_HUMAN	EPLEGDH	VCKV	QES	LRDLKLRK	ESLED
TRB1_HUMAN	PLAEREH	RCKV	QDK	LGDLKLRK	ESLED
TRB3_HUMAN	EPEEGGR	TCKV	LAV	LRDLKLCR	ENLED
CASK_HUMAN	GKGPFSV	AVKI	REA	HRDVKPHC	GGFGV
ULK4_HUMAN	GRGSKTV	AILC	NWV	FCDISPRK	SNFCL
ULK4_MOUSE	GRGSRTV	AILC	NWV	FCDLSPGK	SNFCL
ULK4_Xenopus	GKGSKSV	AILR	NWV	FCDLTPRK	TNFSL
ULK1_HUMAN	GHGAFAV	AVKC	KEI	HRDLKPQN	ADFGF
ULK2_HUMAN	GHGAFAV	AIKS	KEI	HRDLKPQN	ADFGF
ULK3_HUMAN	GSGTYAT	AIKC	TEI	HLDLKPQN	ADFGF
HER3_HUMAN	GSGVFGT	CIKV	DHM	HRNLAARN	ADFGV
WNK1_HUMAN	GRGSFKT	AWCE	EEA	HRDLKCDN	GDLGL
WNK2_HUMAN	GRGSFKT	AWCE	EEA	HRDLKCDN	GDLGL
WNK3_HUMAN	GRGAFKT	AWCE	EEA	HRDLKCDN	GDLGL
WNK4_HUMAN	GRGSFKT	AWCE	EEV	HRDLKCDN	GDLGL

Biology: Hierarchy of Protein Organization



Dr. Gerwald Jogl (2016)
Introductory Biochemistry

Biology: Hierarchy of Protein Organization



Dr. Gerwald Jogl (2016)
Introductory Biochemistry

Primary structure of proteins:
protein sequencing

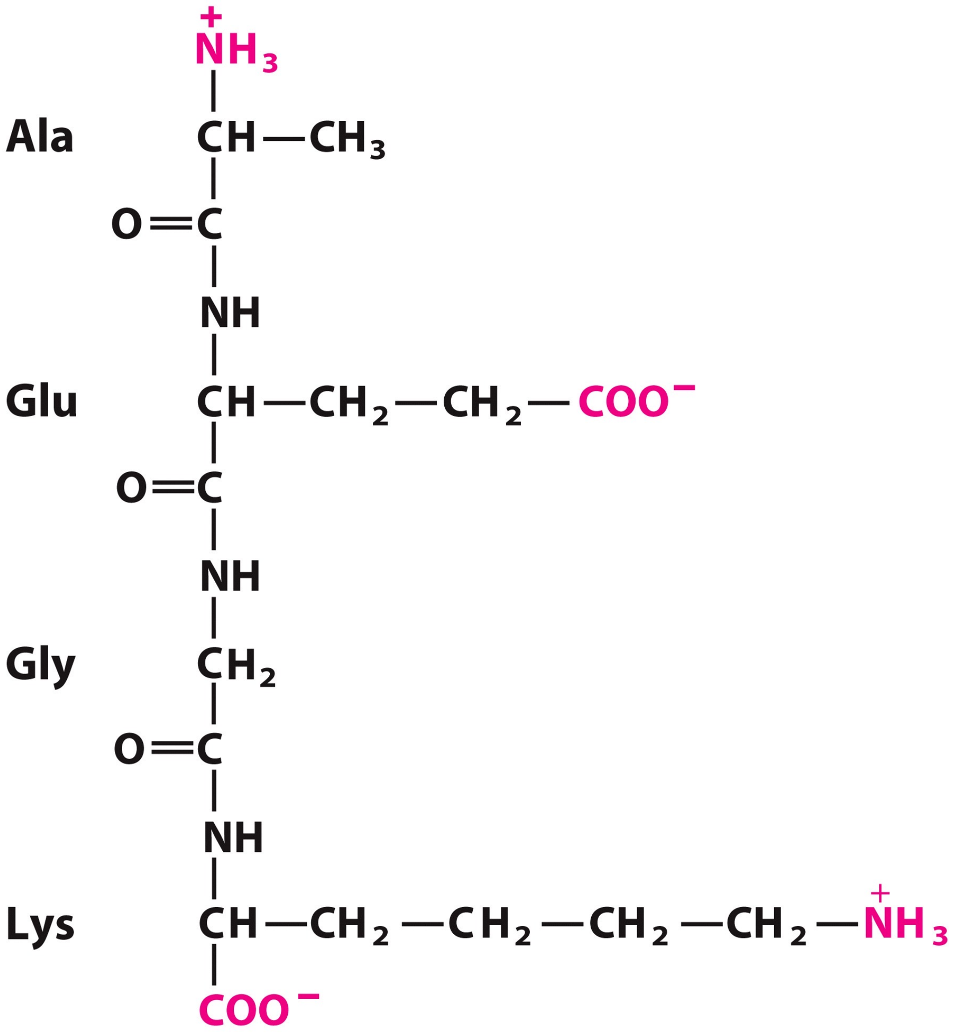


Figure 3-15
Lehninger Principles of Biochemistry, Sixth Edition

Protein sequence alignment in two dimensions

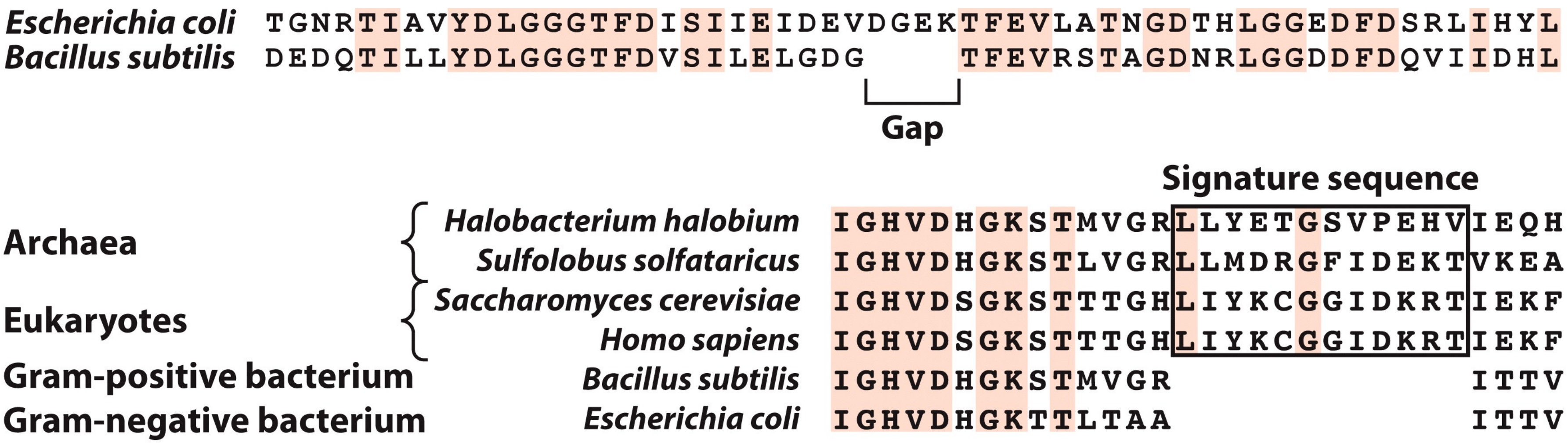
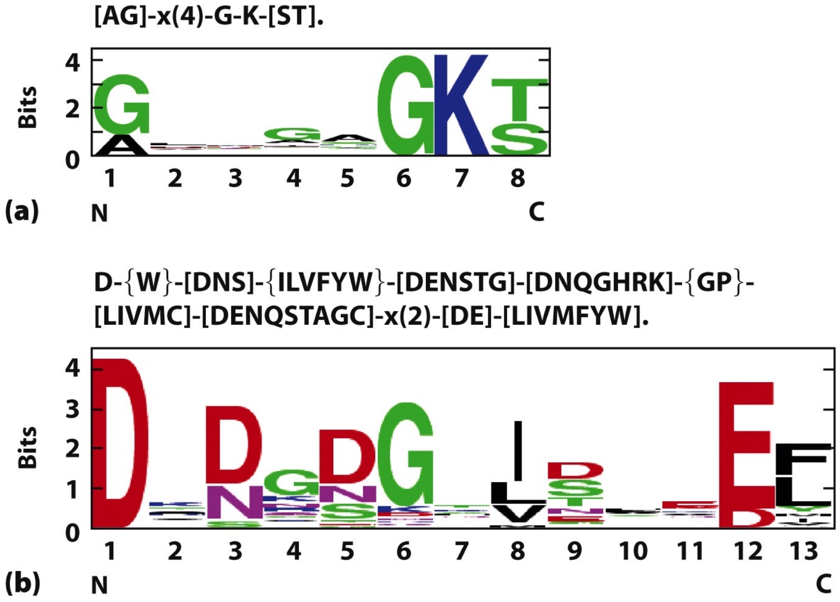


Figure 3-34
Lehninger Principles of Biochemistry, Sixth Edition

Consensus sequences reflect the most common amino acid at each position.

Sequence logos provide a graphical representation of a multiple sequence alignment.



Box 3-2 figure 1
Lehninger Principles of Biochemistry, Sixth Edition

Evolutionary trees based on sequence analysis of protein families

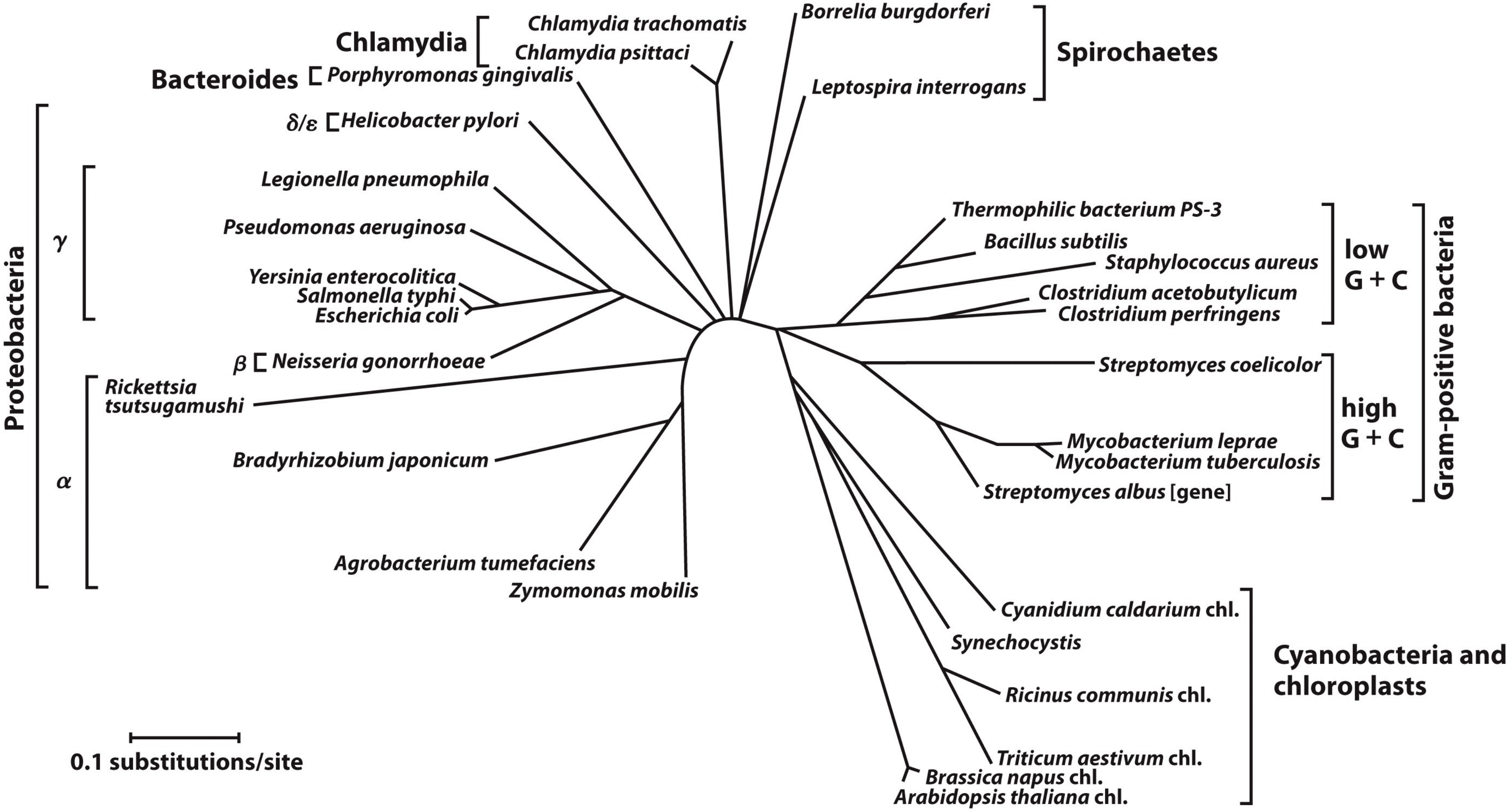
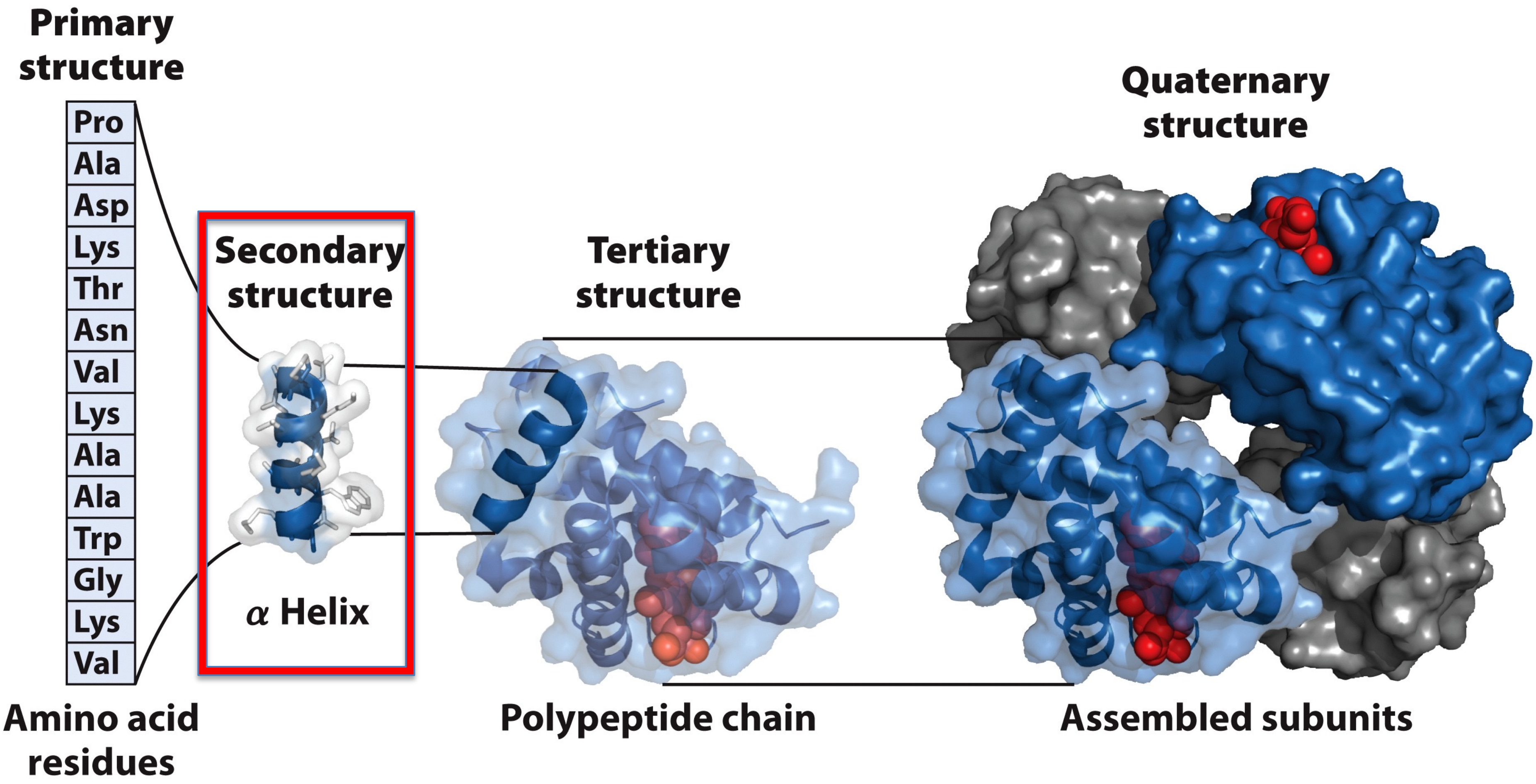


Figure 3-35
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

Biology: Hierarchy of Protein Organization



Primary structure

Pro
Ala
Asp
Lys
Thr
Asn
Val
Lys
Ala
Ala
Trp
Gly
Lys
Val

Amino acid residues

Secondary structure

α Helix

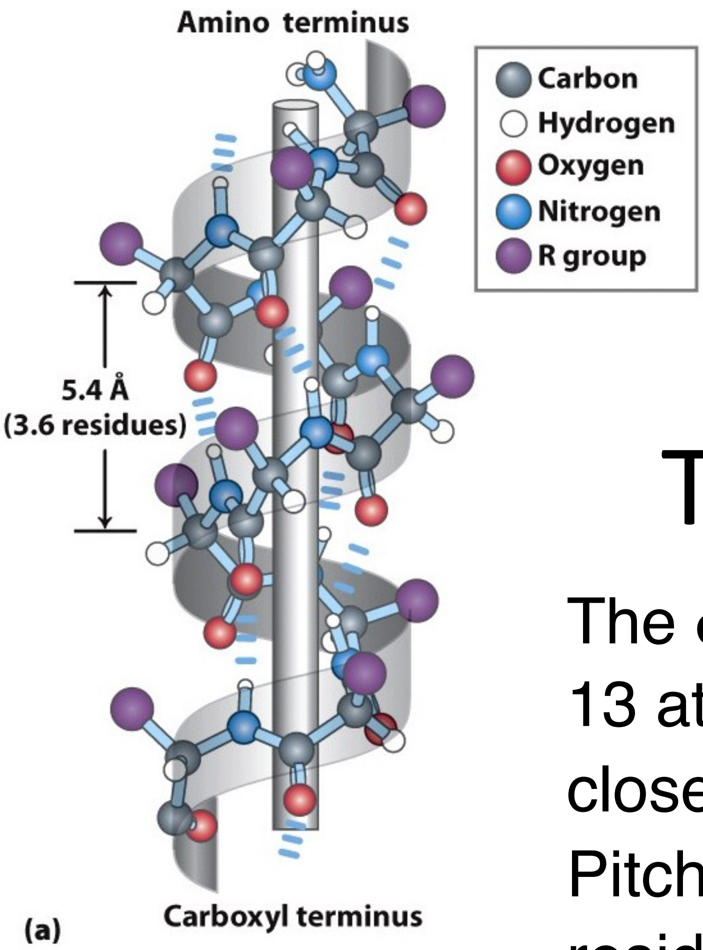
Tertiary structure

Polypeptide chain

Quaternary structure

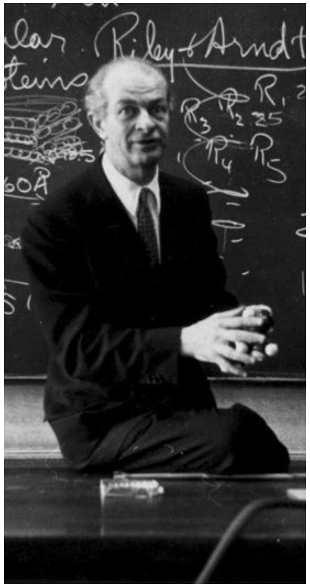
Assembled subunits

Dr. Gerwald Jogl (2016)
Introductory Biochemistry



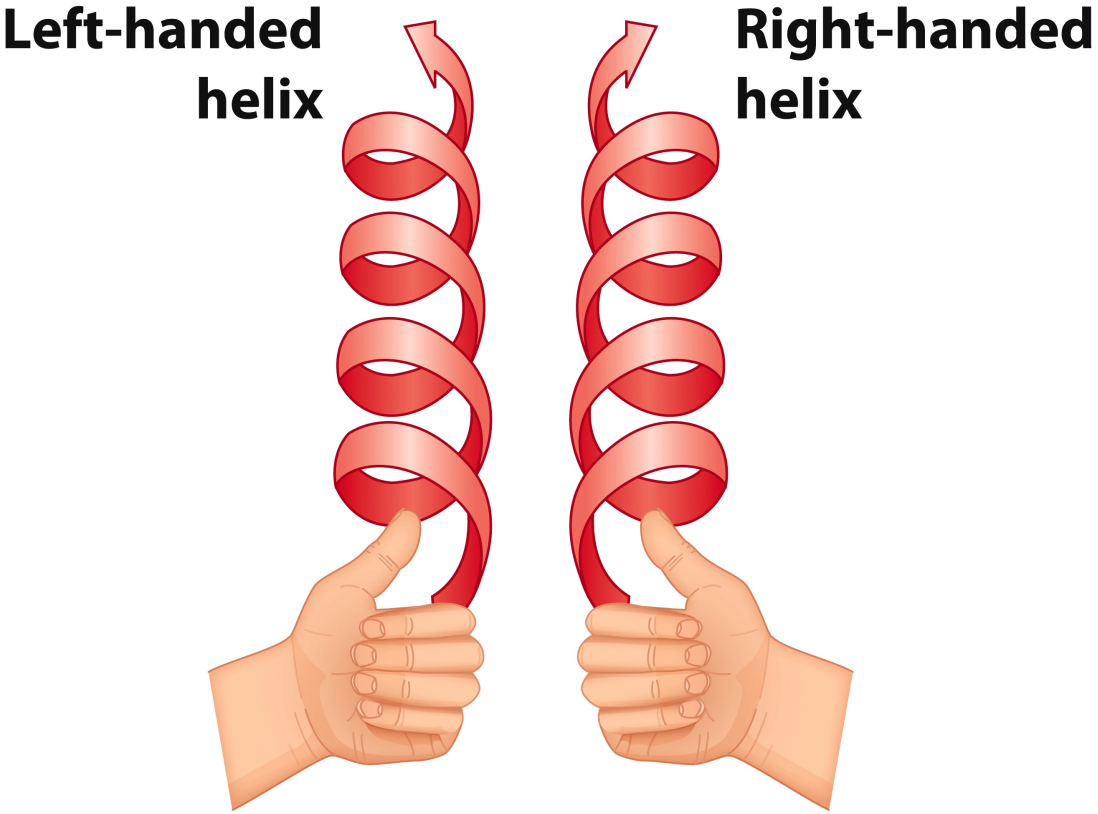
The α-Helix

The α-helix is a **3.6_{13}** helix.
13 atoms form the ring that is closed by a hydrogen bond.
Pitch: 5.4 Å per turn (3.6 residues per turn)



Linus Pauling

Both **Left- and Right-handed α-helices** are possible
However, most α-helices found in proteins are **right handed**



Box 4-1
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

Another major **secondary structure** is the β-**sheet**

There are **two types** of β-sheet

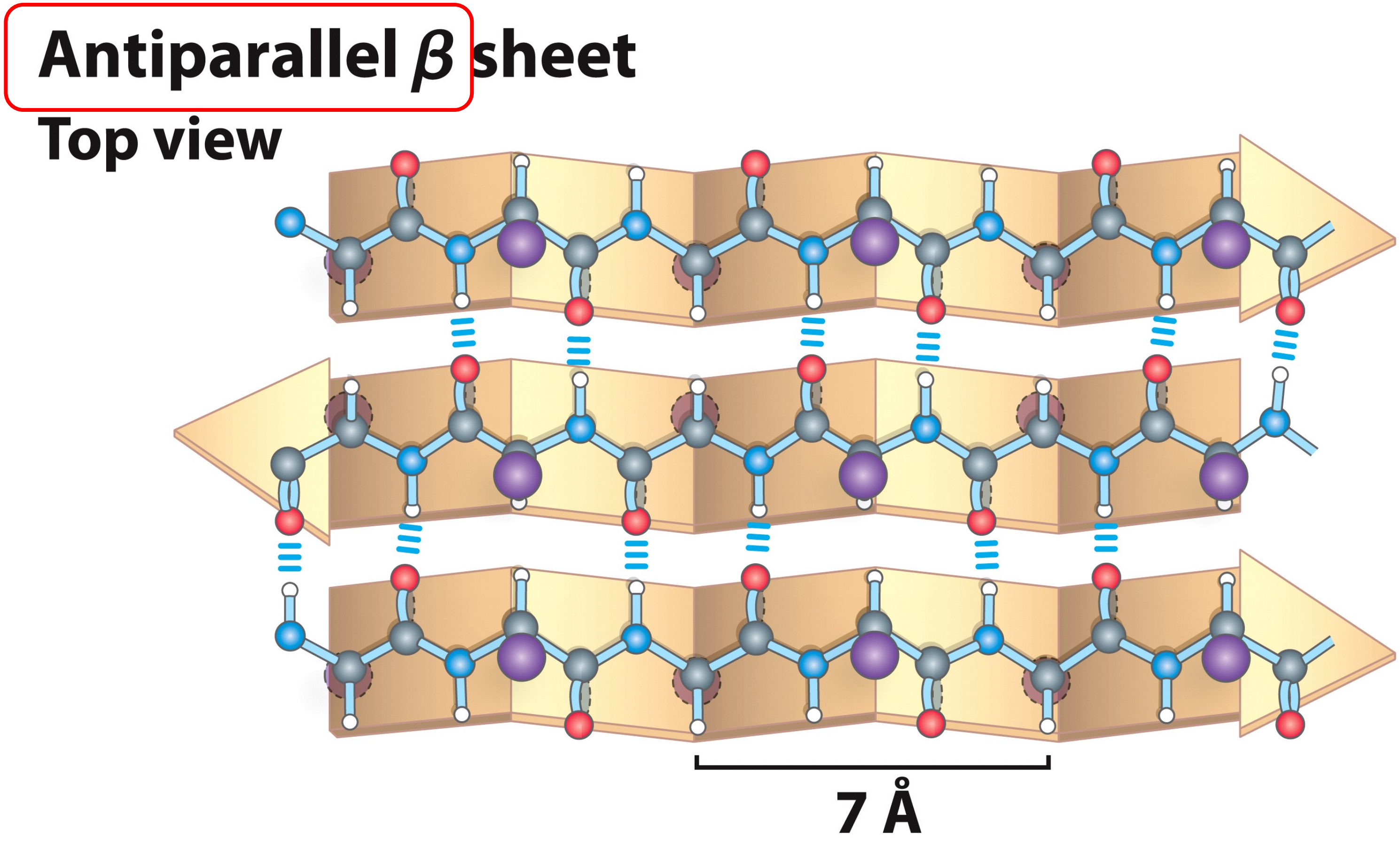


Figure 4-6b
Lehninger Principles of Biochemistry, Sixth Edition

Repeat distance = 7.0 Å

There are **two types** of β-sheet

Parallel β sheet

Top view

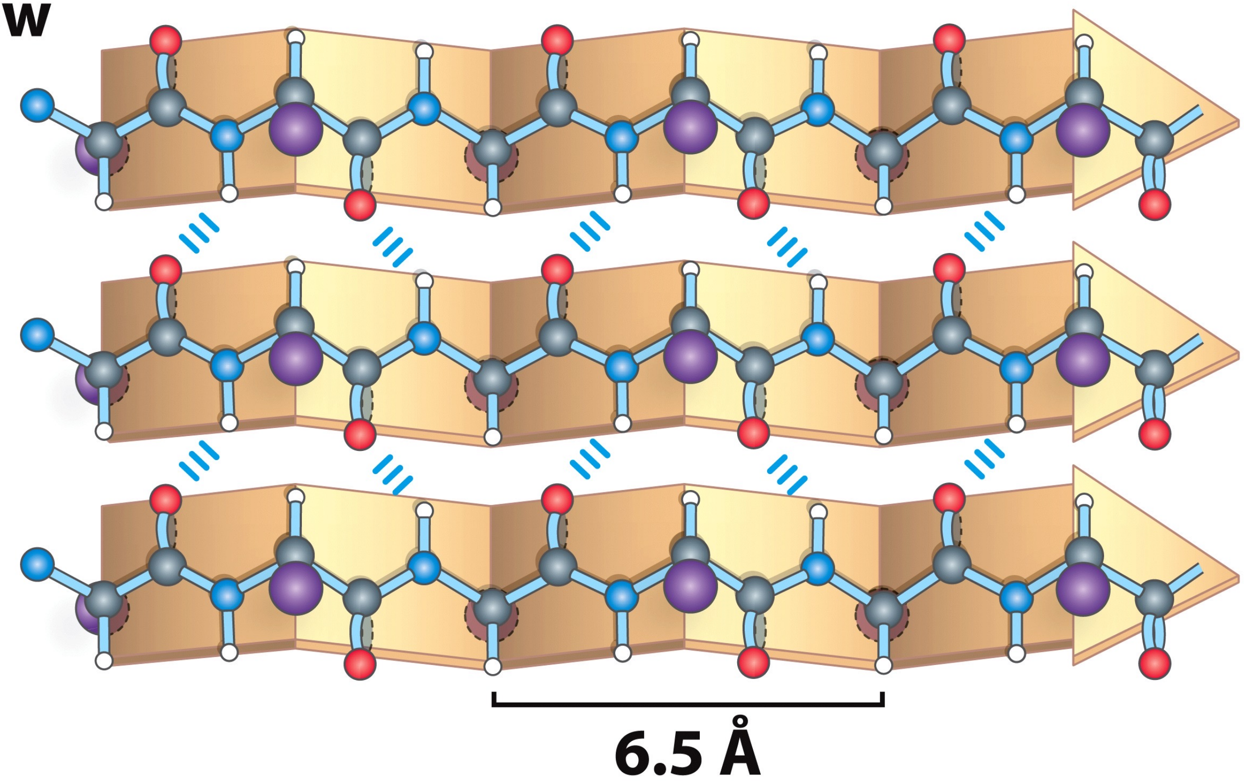
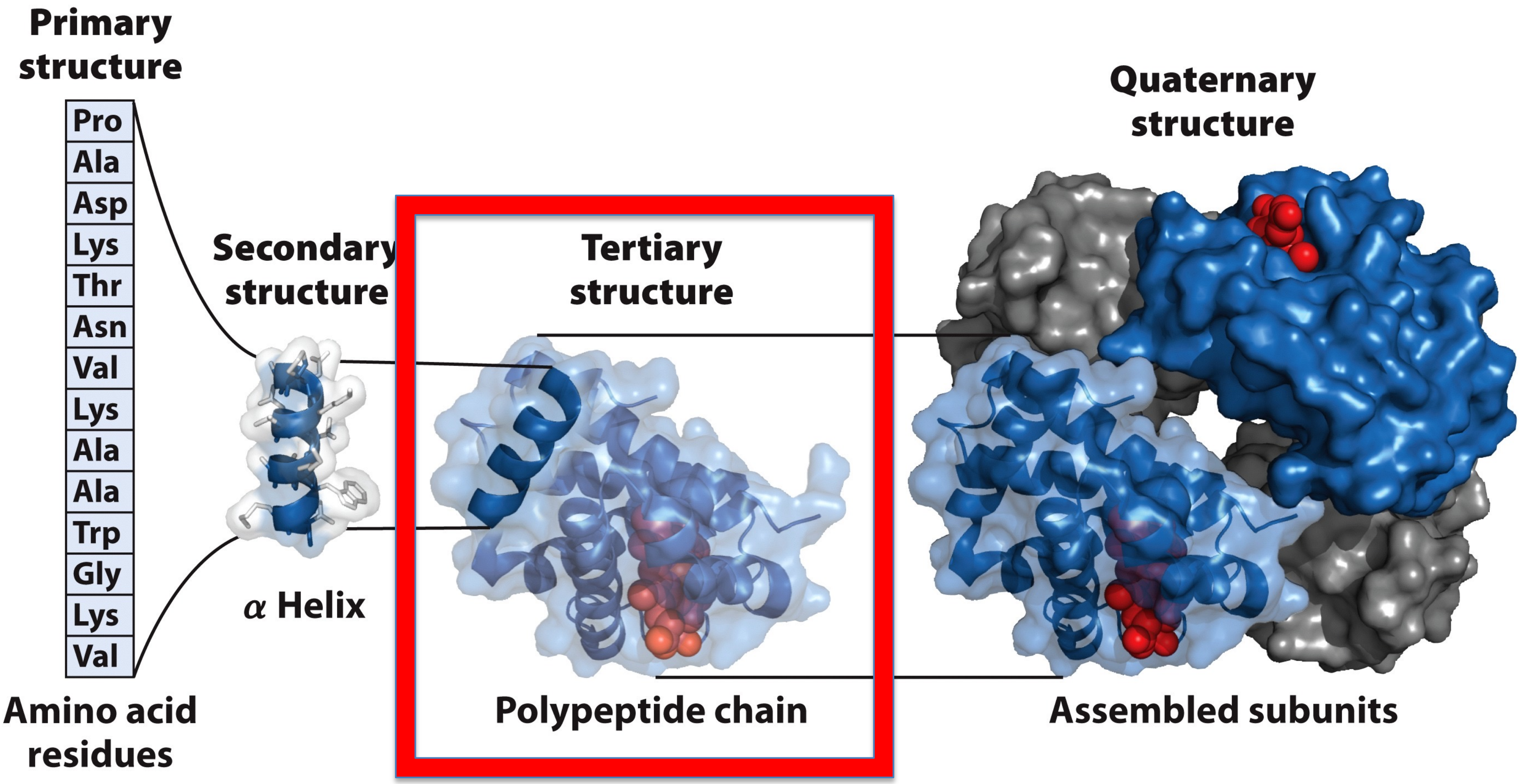


Figure 4-6c
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

Repeat distance = 6.5 Å

Biology: Hierarchy of Protein Organization



Dr. Gerwald Jogl (2016)
Introductory Biochemistry

Tertiary Structure

Secondary structure elements such as α-helices, β-sheets, and β-turns are arranged into supersecondary structures (also called **motifs** or **folds**).

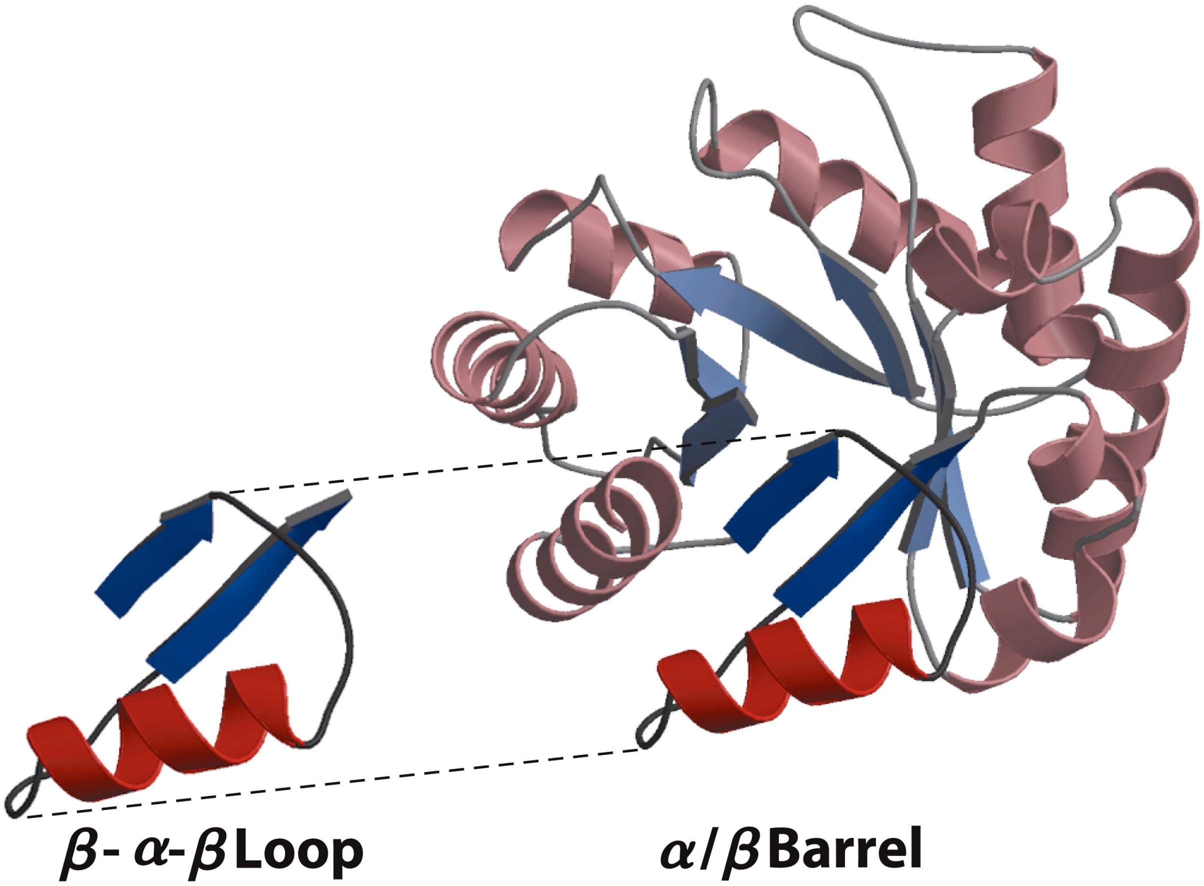
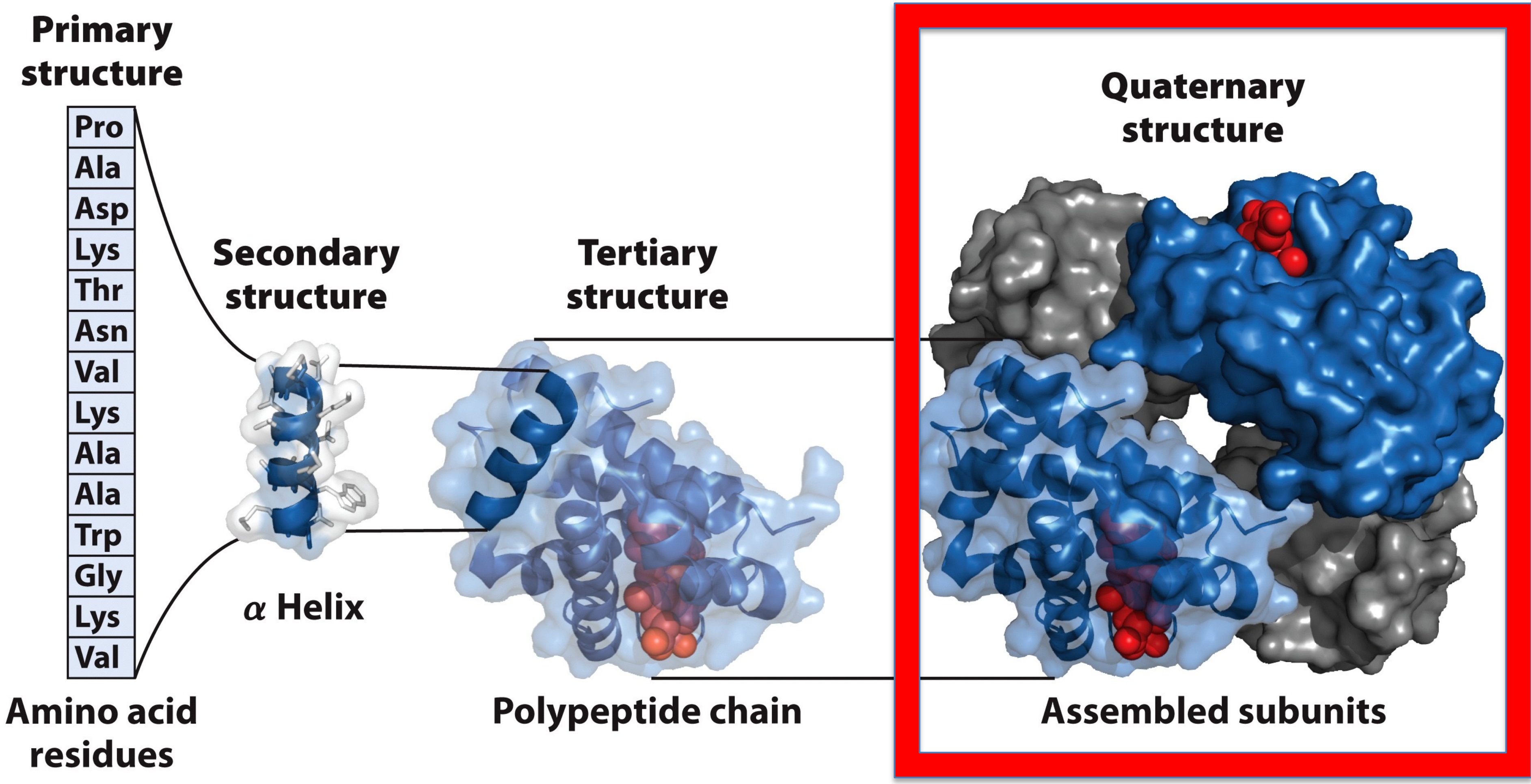


Figure 4-21
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

Biology: Hierarchy of Protein Organization



Primary structure

Pro
Ala
Asp
Lys
Thr
Asn
Val
Lys
Ala
Ala
Trp
Gly
Lys
Val

Amino acid residues

Secondary structure

α Helix

Tertiary structure

Polypeptide chain

Quaternary structure

Assembled subunits

Dr. Gerwald Jogl (2016)
Introductory Biochemistry

Quarternary Structure

Although there is an infinity of possible protein motifs, the actual number found among the proteins whose structure have been determined are surprisingly few (~ 1000 fold entries in the SCOP database)

Larger polypeptides often are composed of **domains**, somewhat independent regions linked by less well defined connecting bridges.

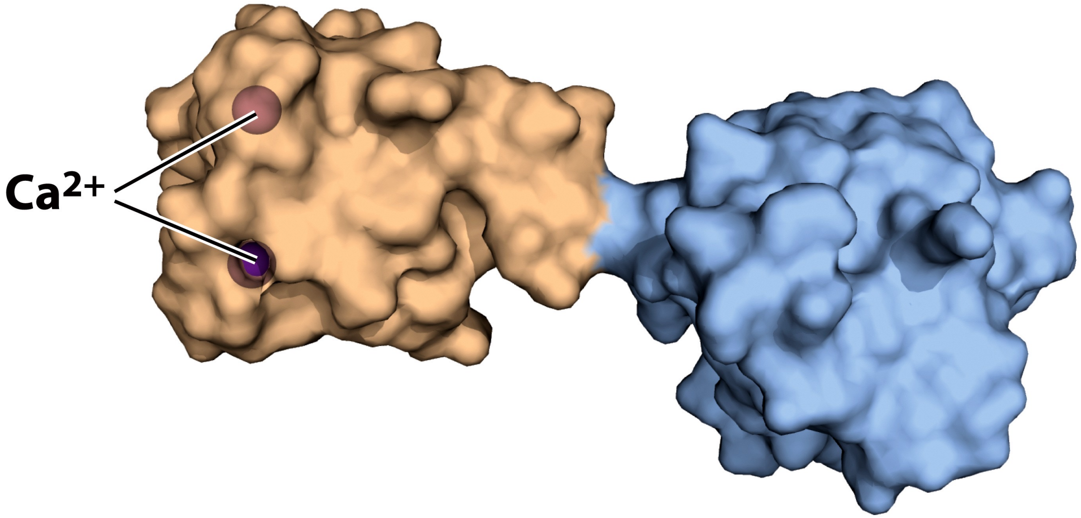
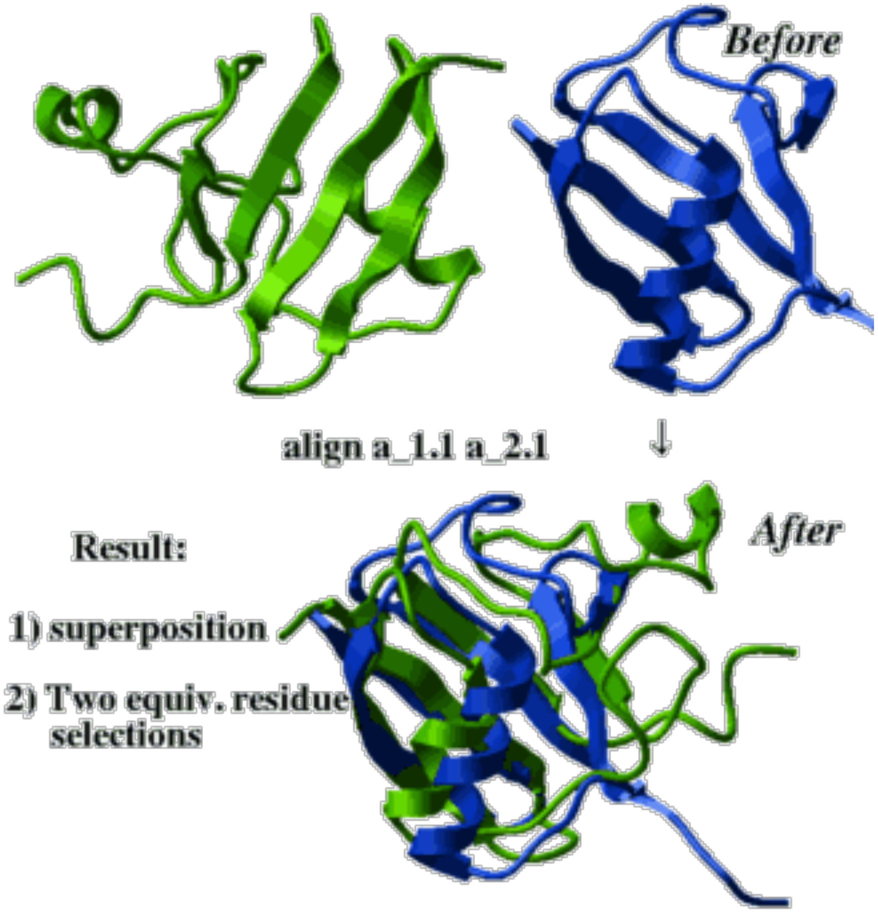
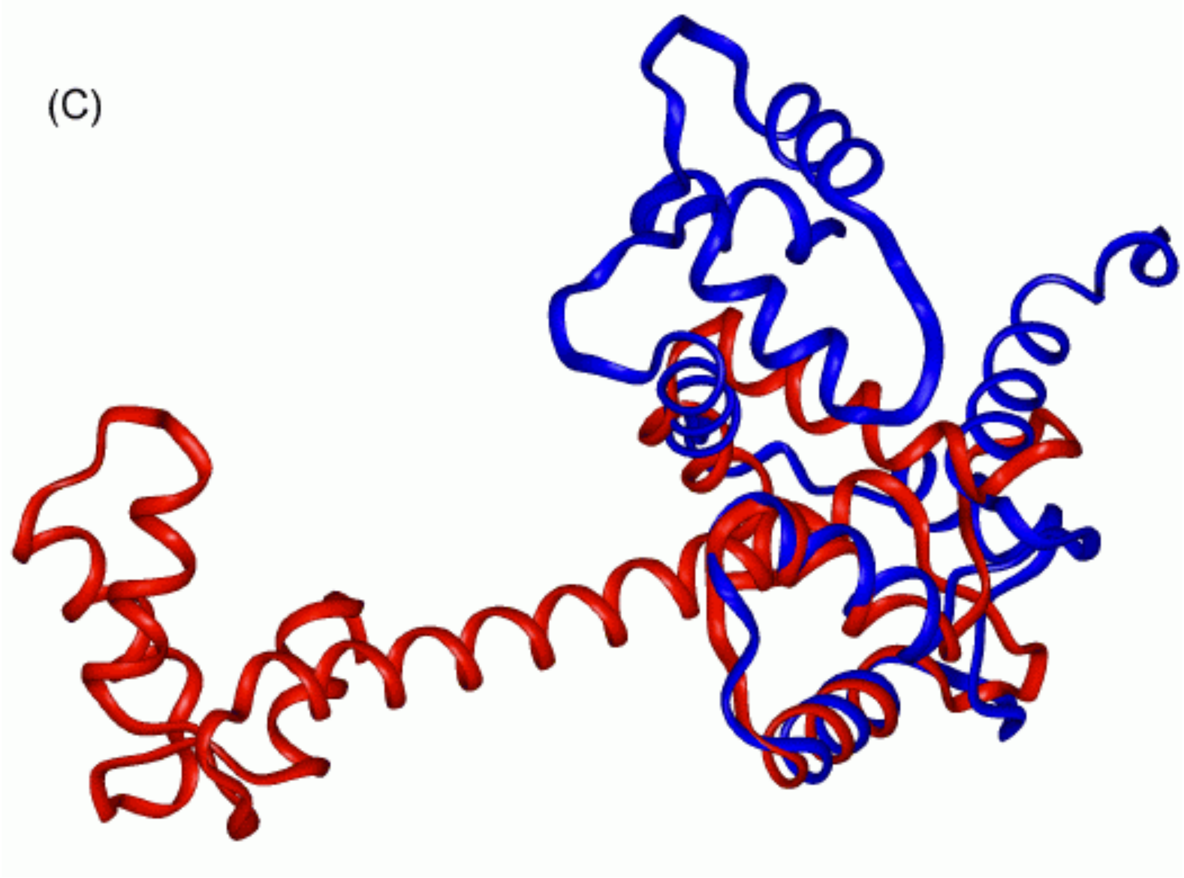


Figure 4-19
Lehninger Principles of Biochemistry, Sixth Edition

Protein Structure Alignment

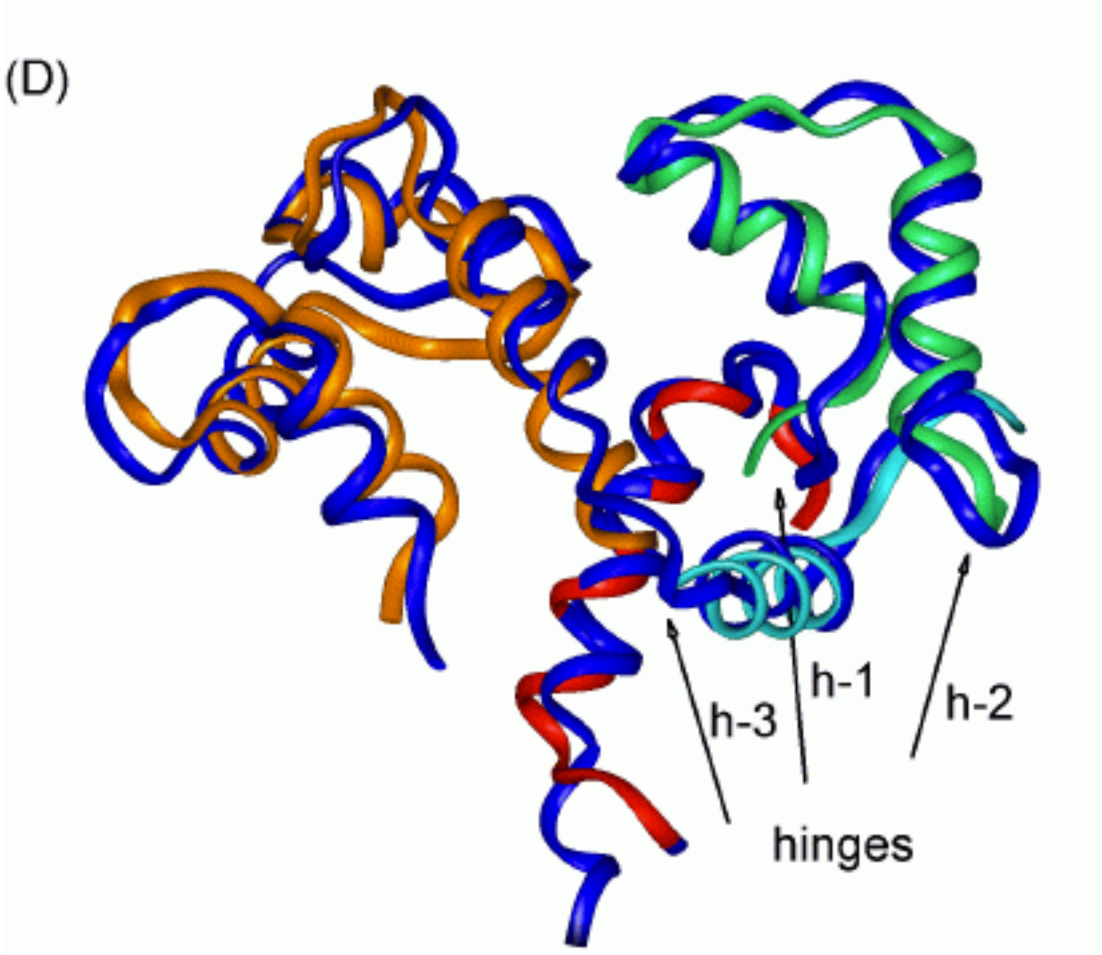
- We want to **superimpose** two proteins and show **structural similarity**
- Common structural elements (ex. alpha helices) help guide the search for an **optimal structure alignment** based on a **distance metric**
- Recall: structure determines function, so these alignments have real biological implications





Unaligned

We know these two proteins have similar primary structure based on sequence alignment techniques. This tells us nothing about their **function** or **biological relationship**.



Aligned

We notice that alpha helices line up and the two proteins **have similar overall structure**. We now have some real biological evidence that the two proteins are somehow related in function.

Some Methods and Tools

Methods

- DALI (Holm and Sander)
- STRUCTURAL (Gerstein and Levitt)
- VAST (Gibrat)
- MINAREA (Falicov and Cohen)

Tools

- TM Align
- RCSB Protein Comparison Tool
- PROMALS3D
- Overview of Tools from UCSF