

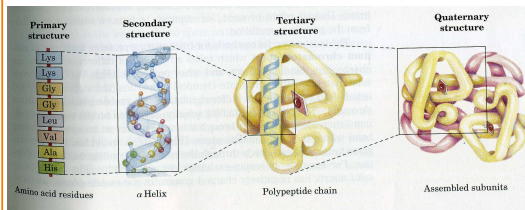
Fundamentals of Protein Structure

Yu (Julie) Chen and
Thomas Funkhouser
Princeton University
CS597A, Fall 2005

Outline

- Protein structure
 - Primary
 - Secondary
 - Tertiary
 - Quaternary
- Protein folding/binding
 - Forces and factors

Levels of Protein Structure



Lehninger Principles of Biochemistry (3rd edition)
David L. Nelson, Michael M. Cox

Outline

- Protein structure
 - ~~Primary~~
 - Secondary
 - Tertiary
 - Quaternary
- Protein folding/binding
 - Forces and factors

Primary Structure

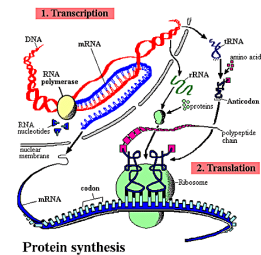
DNA

```
GGGGCTAAGGGGGTGGGGCTTCGCCCCCGCCGCTAIAAGGGGGCCCGCGGCTCCGTGCGCTTGCAGACTTGGCT
GCAGCCCTGCTGCTTGGCCCGCTCGCTCCCGCATGGCTCCAGGAGTCTTCTGGTGGTGGCACTGGAGATGAAAG
GGAGAGAGAGAGCTTGGCCAGCTGCTCCAGACGCTGATGGCCGAGACTCTCGCCAGACCCAGGTGGTGGGGAG
CCCTTCAATCTACCTGATTTGGCCGAGAGCTGATGCAAGATTTGAGGTTGCGAGCAGAACTGTTGCAAGCTGAC
CGAAGGTCCTTACAGGAGAGATCAGCCGAGCAATGATCAAAAGATTTGAGCTGCATGGGTGATCTGGGCCACTGAG
AGCGGAGGATTTTTGGGAGTCTGATGAGTTGATGGCCAGAAAGTGGCTCATGCTCTGCTGAAGGC . [Strauss5]
```

Sequence of Nucleic Acids

Primary Structure

Transcription and translation (DNA→Protein)



<http://www.accessexcellence.org>

Primary Structure



Transcription and translation (DNA→Protein)

First Position	Second Position				Third Position
	U(T)	C	A	G	
U(T)	Phe	Ser	Tyr	Cys	U(T)
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP	A
C	Leu	Pro	His	Arg	U(T)
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U(T)
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U(T)
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Primary Structure



Transcription and translation (DNA→Protein)

Alanine	Ala	A
Cysteine	Cys	C
Aspartic Acid	Asp	D
Glutamic Acid	Glu	E
Phenylalanine	Phe	F
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Lysine	Lys	K
Leucine	Leu	L
Methionine	Met	M
Asparagine	Asn	N
Proline	Pro	P
Glutamine	Gln	Q
Arginine	Arg	R
Serine	Ser	S
Threonine	Thr	T
Valine	Val	V
Tryptophan	Trp	W
Tyrosine	Tyr	Y

Short-hand Names for Amino Acids

Primary Structure



Transcription and translation (DNA→Protein)

```
GGGGCTAACGGGGGTGGGGCTTCGCGCCGCCCGCGCTAIAAGCCGGCGCGCGCTCCGCTGCGCTTCGCGACTTGGCT
GGCCGCTGCTGCTTCCGCGCCCTCCGCTCCGCGAGGCTTCTGCGGGCGCACTGGAGAGTGAAGCG
CGCGCAGAGAGACTTGGCGAGCTTCCGCGAGCTGATGGCGAGACTCTCGCGCAGCAGAGGTGTTTGGCGAG
CCCTTCAATCTACCTTGATTTGCCCGCGAGACTGATGCAAAAGATGAGGTTGCGAGCAGCAAAACTGTGCAAACTAC
CGAAAGGTCTTACAGGAGAGATCAGCCGACAGTATCAAAAGATATGAGGTGCGATGGGTATCTCGGCGCAGCTAG
AGCGGAGCATGTTTTGGGAGGCTGATGAGTTGATGGCGAAGGTGGCTCATGCTCTGCTGAGGCG...
```

Sequence of Nucleic Acids



```
APRKFYVGGNWKMGDKKSLGELIHLNLNGAKLSADTEVCGAPSIYLDFA
RQKLDKIGVAAQNCYKVPKGAFTGEISPAKIDGAAWVILGHSERRHVFGE
SDELIGQVAHALAEGLVIAICIGEKLDEREAGITEKVVFEQTKAIADNVKDW
SKVVLAYEPVWAIGTGKTATPQQAQEVHEKLRGLKSHVSDAVAQSTRI
IYGGSVTGGNCKELASQHDVDFLVGGASLKPEFVDIINAKH
```

Sequence of Amino Acids

[Straus85]

Primary Structure



Transcription and translation (DNA→Protein)

```
GGGGCTAACGGGGGTGGGGCTTCGCGCCGCCCGCGCTAIAAGCCGGCGCGCGCTCCGCTGCGCTTCGCGACTTGGCT
GGCCGCTGCTGCTTCCGCGCCCTCCGCTCCGCGAGGCTTCTGCGGGCGCACTGGAGAGTGAAGCG
CGCGCAGAGAGACTTGGCGAGCTTCCGCGAGCTGATGGCGAGACTCTCGCGCAGCAGAGGTGTTTGGCGAG
CCCTTCAATCTACCTTGATTTGCCCGCGAGACTGATGCAAAAGATGAGGTTGCGAGCAGCAAAACTGTGCAAACTAC
CGAAAGGTCTTACAGGAGAGATCAGCCGACAGTATCAAAAGATATGAGGTGCGATGGGTATCTCGGCGCAGCTAG
AGCGGAGCATGTTTTGGGAGGCTGATGAGTTGATGGCGAAGGTGGCTCATGCTCTGCTGAGGCG...
```

Sequence of Nucleic Acids



```
APRKFYVGGNWKMGDKKSLGELIHLNLNGAKLSADTEVCGAPSIYLDFA
RQKLDKIGVAAQNCYKVPKGAFTGEISPAKIDGAAWVILGHSERRHVFGE
SDELIGQVAHALAEGLVIAICIGEKLDEREAGITEKVVFEQTKAIADNVKDW
SKVVLAYEPVWAIGTGKTATPQQAQEVHEKLRGLKSHVSDAVAQSTRI
IYGGSVTGGNCKELASQHDVDFLVGGASLKPEFVDIINAKH
```

Sequence of Amino Acids

[Straus85]

Primary Structure



Transcription and translation (DNA→Protein)

```
GGGGCTAACGGGGGTGGGGCTTCGCGCCGCCCGCGCTAIAAGCCGGCGCGCGCTCCGCTGCGCTTCGCGACTTGGCT
GGCCGCTGCTGCTTCCGCGCCCTCCGCTCCGCGAGGCTTCTGCGGGCGCACTGGAGAGTGAAGCG
CGCGCAGAGAGACTTGGCGAGCTTCCGCGAGCTGATGGCGAGACTCTCGCGCAGCAGAGGTGTTTGGCGAG
CCCTTCAATCTACCTTGATTTGCCCGCGAGACTGATGCAAAAGATGAGGTTGCGAGCAGCAAAACTGTGCAAACTAC
CGAAAGGTCTTACAGGAGAGATCAGCCGACAGTATCAAAAGATATGAGGTGCGATGGGTATCTCGGCGCAGCTAG
AGCGGAGCATGTTTTGGGAGGCTGATGAGTTGATGGCGAAGGTGGCTCATGCTCTGCTGAGGCG...
```

Sequence of Nucleic Acids



```
APRKFYVGGNWKMGDKKSLGELIHLNLNGAKLSADTEVCGAPSIYLDFA
RQKLDKIGVAAQNCYKVPKGAFTGEISPAKIDGAAWVILGHSERRHVFGE
SDELIGQVAHALAEGLVIAICIGEKLDEREAGITEKVVFEQTKAIADNVKDW
SKVVLAYEPVWAIGTGKTATPQQAQEVHEKLRGLKSHVSDAVAQSTRI
IYGGSVTGGNCKELASQHDVDFLVGGASLKPEFVDIINAKH
```

Sequence of Amino Acids

[Straus85]

Primary Structure



Transcription and translation (DNA→Protein)

```
GGGGCTAACGGGGGTGGGGCTTCGCGCCGCCCGCGCTAIAAGCCGGCGCGCGCTCCGCTGCGCTTCGCGACTTGGCT
GGCCGCTGCTGCTTCCGCGCCCTCCGCTCCGCGAGGCTTCTGCGGGCGCACTGGAGAGTGAAGCG
CGCGCAGAGAGACTTGGCGAGCTTCCGCGAGCTGATGGCGAGACTCTCGCGCAGCAGAGGTGTTTGGCGAG
CCCTTCAATCTACCTTGATTTGCCCGCGAGACTGATGCAAAAGATGAGGTTGCGAGCAGCAAAACTGTGCAAACTAC
CGAAAGGTCTTACAGGAGAGATCAGCCGACAGTATCAAAAGATATGAGGTGCGATGGGTATCTCGGCGCAGCTAG
AGCGGAGCATGTTTTGGGAGGCTGATGAGTTGATGGCGAAGGTGGCTCATGCTCTGCTGAGGCG...
```

Sequence of Nucleic Acids



```
APRKFYVGGNWKMGDKKSLGELIHLNLNGAKLSADTEVCGAPSIYLDFA
RQKLDKIGVAAQNCYKVPKGAFTGEISPAKIDGAAWVILGHSERRHVFGE
SDELIGQVAHALAEGLVIAICIGEKLDEREAGITEKVVFEQTKAIADNVKDW
SKVVLAYEPVWAIGTGKTATPQQAQEVHEKLRGLKSHVSDAVAQSTRI
IYGGSVTGGNCKELASQHDVDFLVGGASLKPEFVDIINAKH
```

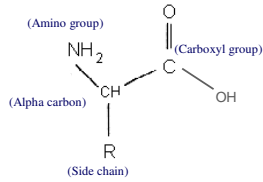
Sequence of Amino Acids

[Straus85]

Primary Structure



Amino acid:

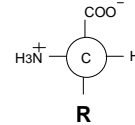
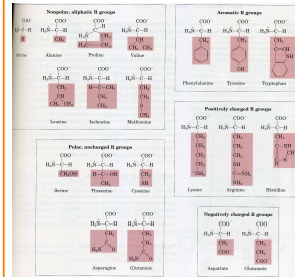


[<http://www.cryst.bbk.ac.uk>]

Primary Structure



Twenty amino acids:

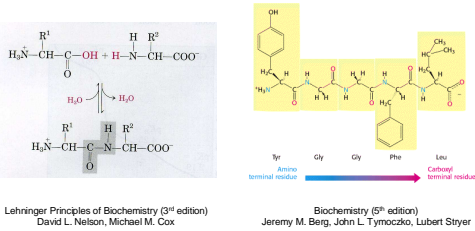


Lehninger Principles of Biochemistry (3rd edition)
David L. Nelson, Michael M. Cox

Primary Structure



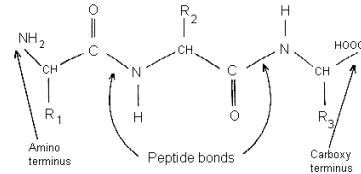
Amino acids are linked by peptide bonds



Primary Structure



Polypeptide chain:



APRKFFVGGNWKMGDKKSLGELIHLTLNGAKLSADTEVVCGAPSIYLFDFAR
QKLDKIGVAAQNCYKVPKGAFTGEISPAKIDGAAWVILGHSERRHVFGE
DELIGQKVAHALAEGLGVIACIGEKLDEREAGITEKVVFEQTKAIADNVKDW
SKVVLAYEPVVAIGTGKTATPQQAQEVHEKLRGLWLSHVSDAVAQSTRI
IYGGVTTGGNCKELASQHDVDFGLVGGASLKPFEVDIINAKH

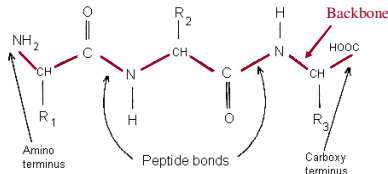
Sequence of Amino Acids

[<http://www.cryst.bbk.ac.uk>]

Primary Structure



Polypeptide chain:



APRKFFVGGNWKMGDKKSLGELIHLTLNGAKLSADTEVVCGAPSIYLFDFAR
QKLDKIGVAAQNCYKVPKGAFTGEISPAKIDGAAWVILGHSERRHVFGE
DELIGQKVAHALAEGLGVIACIGEKLDEREAGITEKVVFEQTKAIADNVKDW
SKVVLAYEPVVAIGTGKTATPQQAQEVHEKLRGLWLSHVSDAVAQSTRI
IYGGVTTGGNCKELASQHDVDFGLVGGASLKPFEVDIINAKH

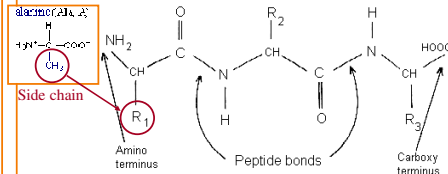
Sequence of Amino Acids

[<http://www.cryst.bbk.ac.uk>]

Primary Structure



Polypeptide chain:



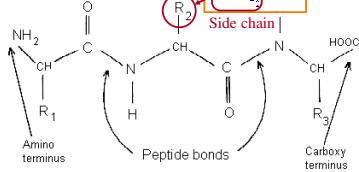
APRKFFVGGNWKMGDKKSLGELIHLTLNGAKLSADTEVVCGAPSIYLFDFAR
QKLDKIGVAAQNCYKVPKGAFTGEISPAKIDGAAWVILGHSERRHVFGE
DELIGQKVAHALAEGLGVIACIGEKLDEREAGITEKVVFEQTKAIADNVKDW
SKVVLAYEPVVAIGTGKTATPQQAQEVHEKLRGLWLSHVSDAVAQSTRI
IYGGVTTGGNCKELASQHDVDFGLVGGASLKPFEVDIINAKH

Sequence of Amino Acids

[<http://www.cryst.bbk.ac.uk>]

Primary Structure

Polypeptide chain:



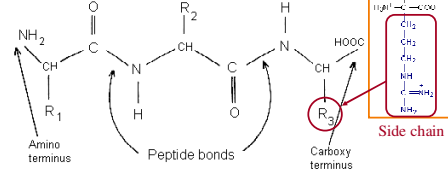
APRKFYVGGNWKMGDKKSLGELIHLNLGAKLSADTEVCGAPSIYLDFA
 RKLDAKIGVAAQNCYKVPKGAFTGEISPAMIKDIGAAWVILGHSERRHVFGE
 DELIGQVAHALAEGLVACIGEKLDEREAGITEKVVEFQTKAIADNVKDW
 SKVVLAYEPVVAIGTKTATPQQAQEVHEKLRGLKSHVSDAVAQSTRI
 IYGGVTVGGNCKELASQHDVDFLVGGASLKPFDIINAKH

Sequence of Amino Acids

[http://www.cryst.bbk.ac.uk]

Primary Structure

Polypeptide chain:



APRKFYVGGNWKMGDKKSLGELIHLNLGAKLSADTEVCGAPSIYLDFA
 RKLDAKIGVAAQNCYKVPKGAFTGEISPAMIKDIGAAWVILGHSERRHVFGE
 DELIGQVAHALAEGLVACIGEKLDEREAGITEKVVEFQTKAIADNVKDW
 SKVVLAYEPVVAIGTKTATPQQAQEVHEKLRGLKSHVSDAVAQSTRI
 IYGGVTVGGNCKELASQHDVDFLVGGASLKPFDIINAKH

Sequence of Amino Acids

[http://www.cryst.bbk.ac.uk]

Outline

Protein structure

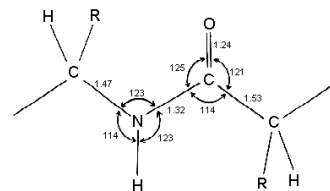
- Primary
- Secondary
- Tertiary
- Quaternary

Protein folding/binding

- Forces and factors

Primary Structure

Peptide bond:

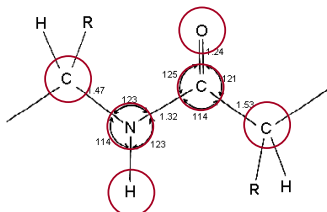


Most bond angles are constrained

[http://www.cryst.bbk.ac.uk]

Primary Structure

Peptide bond:

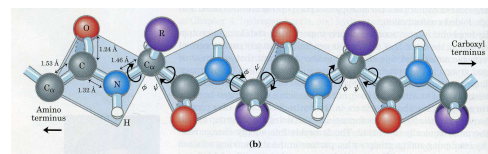


Most atoms in peptide bond are co-planar

[http://www.cryst.bbk.ac.uk]

Primary Structure

N-C α and C α -C bonds can rotate

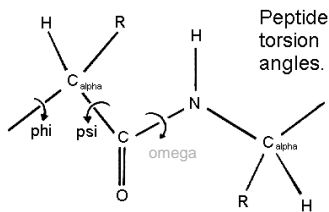


Lehninger Principles of Biochemistry (3rd edition)
 David L. Nelson, Michael M. Cox

Secondary Structure



Peptide bond:



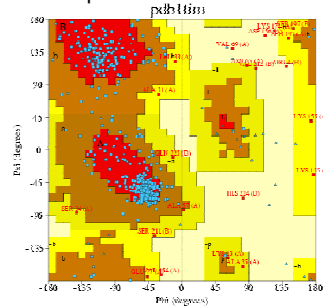
Peptides have at most two free torsion angles
(omega is 180° (trans) or 0° cis)

[<http://www.cryst.bbk.ac.uk>]

Secondary Structure



Ramachandran plot for 1tim:



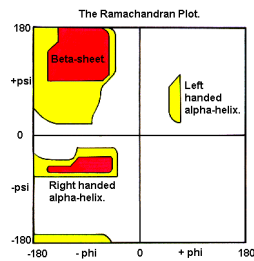
[PDBSUM]

Secondary Structure



Some repeating sequences of torsion angles are very stable

- Alpha helix
- Beta sheet



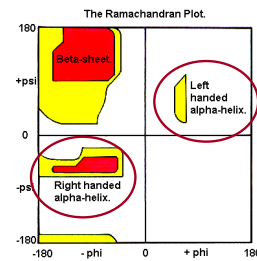
[<http://www.cryst.bbk.ac.uk>]

Secondary Structure



Some repeating sequences of torsion angles are very stable

- Alpha helix
- Beta sheet



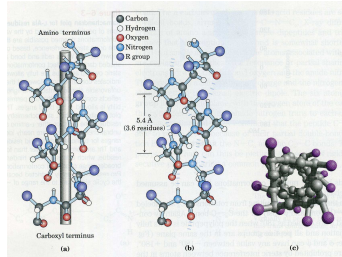
[<http://www.cryst.bbk.ac.uk>]

Secondary Structure



Some repeating sequences of torsion angles are very stable

- Alpha helix
- Beta sheet

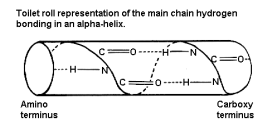


Secondary Structure

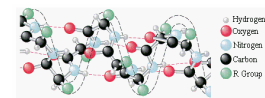


Some repeating sequences of torsion angles are very stable

- Alpha helix
- Beta sheet



[<http://www.cryst.bbk.ac.uk>]

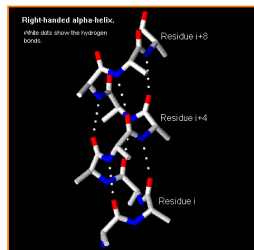


Secondary Structure



Some repeating sequences of torsion angles are very stable

- Alpha helix
- Beta sheet



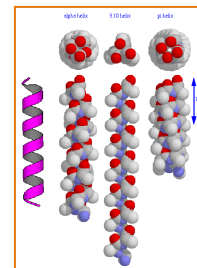
[<http://www.cryst.bbk.ac.uk>]

Secondary Structure



Some repeating sequences of torsion angles are very stable

- Alpha helix
- Beta sheet



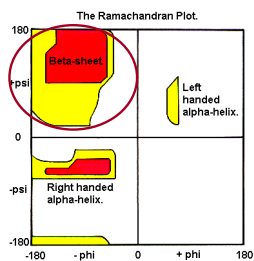
[<http://www.cryst.bbk.ac.uk>]

Secondary Structure



Some repeating sequences of torsion angles are very stable

- Alpha helix
- Beta sheet



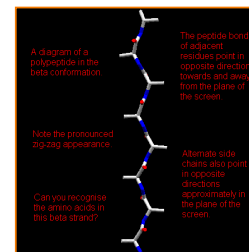
[<http://www.cryst.bbk.ac.uk>]

Secondary Structure



Some repeating sequences of torsion angles are very stable

- Alpha helix
- Beta sheet



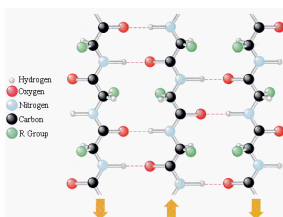
[<http://www.cryst.bbk.ac.uk>]

Secondary Structure



Some repeating sequences of torsion angles are very stable

- Alpha helix
- Beta sheet



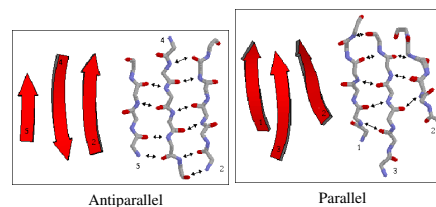
[chemed.chem.purdue.edu]

Secondary Structure



Some repeating sequences of torsion angles are very stable

- Alpha helix
- Beta sheet



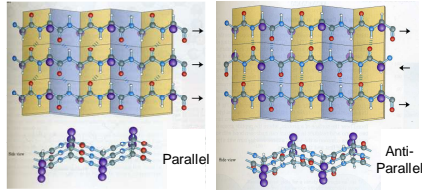
[<http://www.cryst.bbk.ac.uk>]

Secondary Structure



Some repeating sequences of torsion angles are very stable

- Alpha helix
- Beta sheet



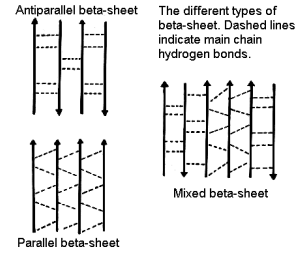
Lehninger Principles of Biochemistry (3rd edition)
David L. Nelson, Michael M. Cox

Secondary Structure



Some repeating sequences of torsion angles are very stable

- Alpha helix
- Beta sheet

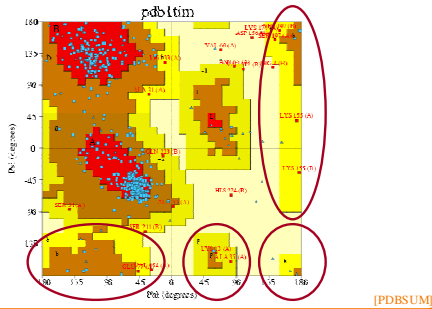


[<http://www.cryst.bbk.ac.uk>]

Secondary Structure



Others form loops, turns etc.

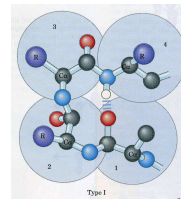


[PDBSUM]

Secondary Structure



Others form loops, turns etc.

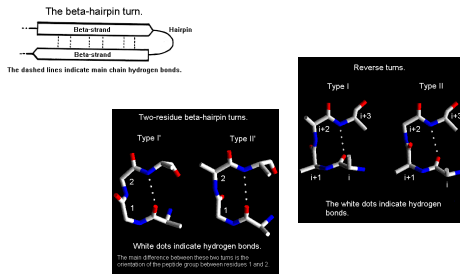


Lehninger Principles of Biochemistry (3rd edition)
David L. Nelson, Michael M. Cox

Secondary Structure

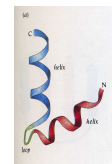


Others form loops, turns etc.

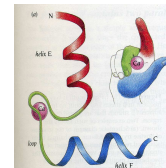


[<http://www.cryst.bbk.ac.uk>]

Supersecondary structure / motifs



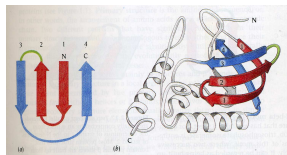
Helix – loop – helix



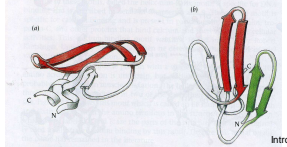
Helix – loop – helix
Calcium-binding motif

Introduction to protein structure (2nd edition)
Carl Branden, John Tooze

Supersecondary structure / motifs



Geek Key motif



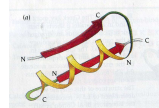
Hairpin motif

Introduction to protein structure (2nd edition)
Carl Branden, John Tooze

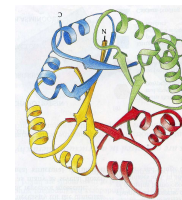
Supersecondary structure / motifs

Four classes:

- All α
- All β
- α/β
- $\alpha+\beta$



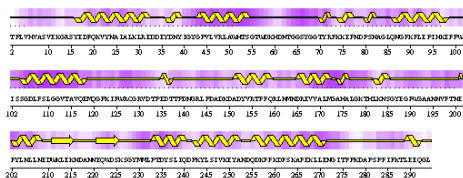
β - α - β motif



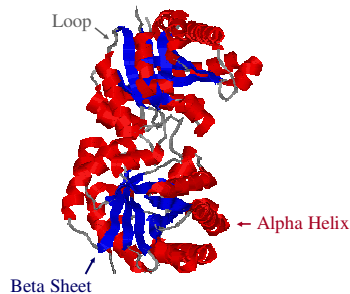
Introduction to protein structure (2nd edition)
Carl Branden, John Tooze

Secondary Structure Visualization

Secondary structure pdb2cyp



Secondary Structure Visualization



Itim
[Jena]

Outline

Protein structure

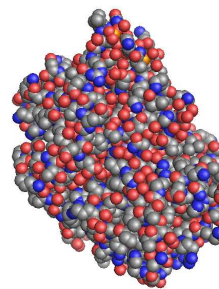
- Primary
- Secondary
- Tertiary
- Quaternary

Protein folding/binding

- Forces and factors

Tertiary Structure

Arrangement of atoms:

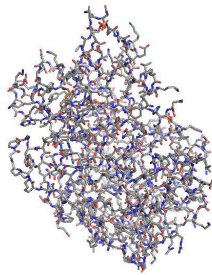


latp
[pymol]

Tertiary Structure



How protein folds:

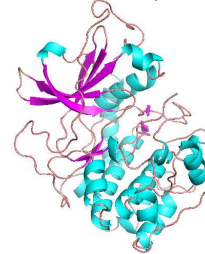


1atp
[pymol]

Tertiary Structure



Arrangement of secondary structures – plus amino acid side chains (not shown)



1atp
[pymol]

Outline



Protein structure

- Primary
- Secondary
- Tertiary
- Quaternary

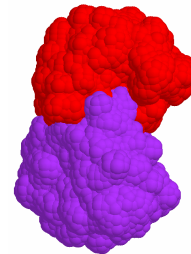
Protein folding/binding

- Forces and factors

Quaternary Structure



How multiple chains/proteins form a complex:

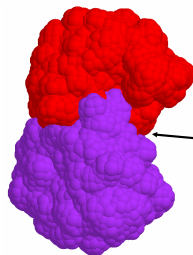


1tim
[Rasurf]

Quaternary Structure



How multiple chains/proteins form a complex:



Active binding site
may be at interface
between two chains

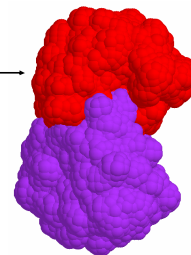
1tim
[Rasurf]

Quaternary Structure



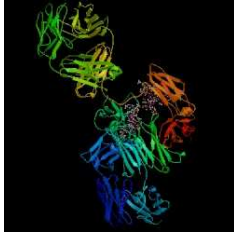
How multiple chains/proteins form a complex:

Chain may take
a new (active)
conformation —→
when bound
to another



1tim
[Rasurf]

Quaternary structure

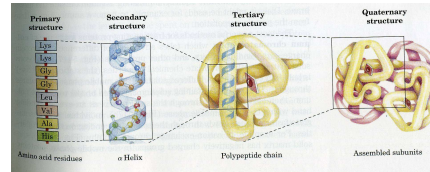


Igg2A Intact Antibody-Mab231; Chain: A, B, C, D
1IGT
L.J.Harris, S.B.Larson, K.W.Hasel, A.Mcpherson

Protein Structure Level Summary

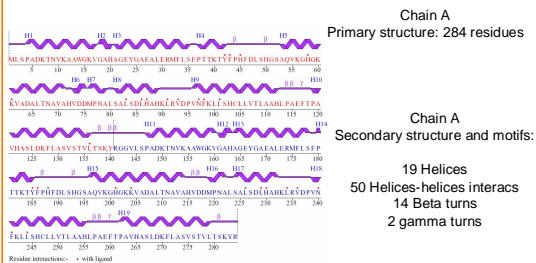
Protein structure description

- Primary ← amino acid sequence
- Secondary ← local fold pattern of small subsequence
- Tertiary ← fold of entire protein chain
- Quaternary ← complex of multiple chains



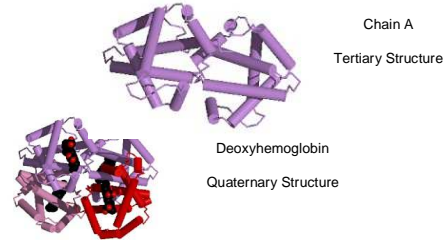
Lehninger Principles of Biochemistry (3rd edition)
David L. Nelson, Michael M. Cox

Example: Hemoglobin



Deoxyhemoglobin (alpha chain). Chain: a. Engineered: yes. Mutation: yes.
Deoxyhemoglobin (beta chain). Chain: b, d. Engineered: yes. Mutation: yes
1C7D
E.A.Brucker

Example: Hemoglobin



Deoxyhemoglobin (alpha chain). Chain: a. Engineered: yes. Mutation: yes.
Deoxyhemoglobin (beta chain). Chain: b, d. Engineered: yes. Mutation: yes
1C7D
E.A.Brucker

Example: Hemoglobin



Deoxyhemoglobin (alpha chain). Chain: a. Engineered: yes. Mutation: yes.
Deoxyhemoglobin (beta chain). Chain: b, d. Engineered: yes. Mutation: yes
1C7D
E.A.Brucker

Outline

Protein structure

- Primary
- Secondary
- Tertiary
- Quaternary

Protein folding/binding

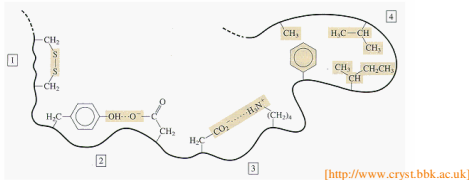
Ø Forces and factors

Folding/Binding



Factors determining tertiary/quaternary structure:

1. Disulfide linkages
2. Hydrogen bonding
3. Electrostatic interactions
4. Hydrophobic interactions
5. Van der Waals forces



Folding/Binding



Important properties of amino acids:

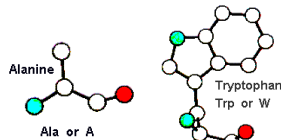
- Size
- Charge
- Polarity
- Aromaticity
- Hydrophobicity
- Conformational constraints

Folding/Binding



Important properties of amino acids:

- Size
- Charge
- Polarity
- Aromaticity
- Hydrophobicity
- Conformational constraints

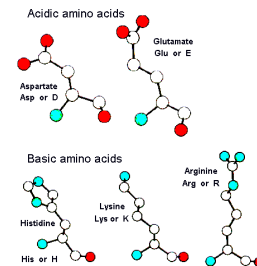


Folding/Binding



Important properties of amino acids:

- Size
- Charge
- Polarity
- Aromaticity
- Hydrophobicity
- Conformational constraints

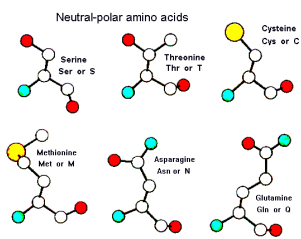


Folding/Binding



Important properties of amino acids:

- Size
- Charge
- Polarity
- Aromaticity
- Hydrophobicity
- Conformational constraints

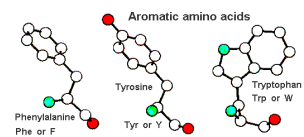


Folding/Binding



Important properties of amino acids:

- Size
- Charge
- Polarity
- Aromaticity
- Hydrophobicity
- Conformational constraints

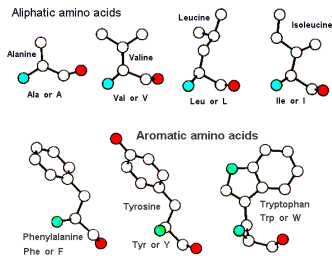


Folding/Binding



Important properties of amino acids:

- Size
- Charge
- Polarity
- Aromaticity
- Hydrophobicity
- Conformational constraints



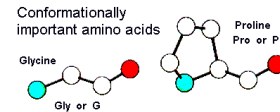
[<http://www.cryst.bbk.ac.uk>]

Folding/Binding



Important properties of amino acids:

- Size
- Charge
- Polarity
- Aromaticity
- Hydrophobicity
- Conformational constraints



[<http://www.cryst.bbk.ac.uk>]

Summary

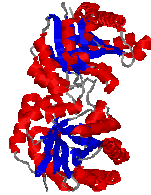


Protein structure description

- Primary ← amino acid sequence
- Secondary ← local fold pattern of small subsequence
- Tertiary ← fold of entire protein chain
- Quaternary ← complex of multiple chains

Protein folding/binding

- Disulfide linkages
- Hydrogen bonding
- Electrostatic interactions
- Hydrophobic interactions
- Van der Waals forces



Itim
[Jens]

References



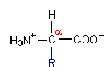
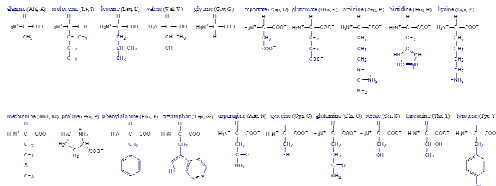
Information and figures were taken from:

- *Introduction to protein structure* (2nd edition)
Carl Branden, John Tooze
- *Lehninger Principles of Biochemistry* (3rd edition)
David L. Nelson, Michael M. Cox
- *Biochemistry* (5th edition)
Jeremy M. Berg, John L. Tymoczko, Lubert Stryer
- <http://www.cs.cryst.bbk.ac.uk>
- <http://www.accessexcellence.org>
- <http://chemmed.chem.purdue.edu>

Primary Structure



Twenty amino acids:



[<http://www.cryst.bbk.ac.uk>]