COMP 564: Introduction to protein structure prediction

Jérôme Waldispühl
School of Computer Science
McGill University
Folding problem

\[ N \text{ états} \sim 10^n \]
\[ n = 100-300 \]

Levinthal paradox
Amino acids: The simple ones

Glycine (Gly, G)

Alanine (Ala, A)

Figure 2-7
Biochemistry, Sixth Edition
© 2007 W. H. Freeman and Company
Amino acids: Aliphatics

Valine (Val, V)
Leucine (Leu, L)
Isoleucine (Ile, I)
Methionine (Met, M)

Figure 2-8
Biochemistry, Sixth Edition
© 2007 W.H. Freeman and Company
Amino acids: Cyclic and Sulfhydryl

Proline (Pro, P)

Cysteine (Cys, C)
Amino acids: Aromatics

Phenylalanine (Phe, F)
Tyrosine (Tyr, Y)
Tryptophan (Trp, W)

Figure 2-10
Biochemistry, Sixth Edition
© 2007 W.H. Freeman and Company
Amino acids: Aliphatic hydroxyl

Serine (Ser, S)

Threonine (Thr, T)
Amino acids: Carboxamides & Carboxylates
Amino acids: Basics

Lysine (Lys, K)

Arginine (Arg, R)

Histidine (His, H)

Figure 2-14
Biochemistry, Sixth Edition
© 2007 W. H. Freeman and Company
Histidine ionisation

Figure 2-15
Biochemistry, Sixth Edition
© 2007 W.H. Freeman and Company
Primary structure

A peptide bond assemble two amino acids together:

\[
\begin{align*}
+&\text{H}_3\text{N}\text{C}\text{C}=\text{O}^- + +\text{H}_3\text{N}\text{C}\text{C}=\text{O}^- &\rightarrow &+\text{H}_3\text{N}\text{C}\text{C}=\text{C}\text{N}=\text{C}=\text{O}^- + \text{H}_2\text{O}
\end{align*}
\]

Peptide bond

A chain is obtained through the concatenation of several amino acids:
Peptide bond is pH dependent

Figure 2-6
Biochemistry, Sixth Edition
© 2007 W.H. Freeman and Company
Peptide bond features (1)

Peptide bonds lies on a plane

Bond lengths
The chain has 2 degrees of liberty given by the dihedral angles $\Phi$ and $\Psi$. The geometry of the chain can be characterized through $\Phi$ and $\Psi$. 
Peptide bond features (3)

Cis/trans isomers of the peptide group

Trans configuration is preferred versus Cis (ratio $\sim$1000:1)

An exception is the Proline with a preference ratio of $\sim$3:1
Ramachandran diagram gives the values which can be adopted by $\Phi$ and $\Psi$
The side chains also have flexible torsion angles
The preferred side-chains conformations are called “rotamers”

Example: Asparagine

Energy (chi1,chi2)

-2.5
-4.3
-3.3
4.5

1 kcal/mole between levels

Typical conformations experimentally observed
• conformations observed by simulation
In helices and sheets, polar groups are involved into hydrogen bonds

α helix

β-sheet

3.6 residues per turn

Pseudo-periodicity of 2
3.6 residues per turn, H-bond between residue n and n+4
Although other (rare) helices are observed: π-helices, 3.10-helices...
**β-sheets**

**β-strand (elementary blocks):**

β-strands are assembled into (parallel, anti-parallel) β-sheets.
β-sheets

Anti-parallel β-sheets

Parallel β-sheets
β-sheets

Various shapes of β structures

Twisted β-sheets

β-barrel
β-sheets

- Antiparallel beta-sheet
- Parallel beta-sheet
- Mixed beta-sheet

The different types of beta-sheet. Dashed lines indicate main chain hydrogen bonds.

Figure 2-34
Biochemistry, Sixth Edition
© 2007 W. H. Freeman and Company
Loops

~ 1/3 of amino acids
Super-secondary & Tertiary structure

Secondary structure elements can be assembled into super-secondary motifs.

The tertiary structure is the set of 3D coordinates of atoms of a single amino acid chain.
A protein can be composed of multiple chains with interacting subunits.
Protein can interact with molecules
Example: Hemoglobin

An Heme (iron + organic ring) binds to the protein, and allow the capture of oxygen atoms.
Disulfide bond

Two cysteines can interact and create a disulfide bond.
The tertiary structure is globular, with a preference for polar residues on its surface but rather apolar in its interior.
Membrane proteins are an exception

~ 30% of human genome, ~ 50% of antibiotics
Proteins folds into a native structure
Overview of the methods used to predict the protein structure

Several issue must be addressed first:

• Which degree of definition?
• What's the length of the sequence?
• Which representation/modeling suits the best?
• Should we simulate the folding or predict the structure?
• Do we want a single prediction or a set of candidates?
• Machine learning approach or physical model?
Molecular Dynamics
HP lattice model
Hidden Markov models
(and other machine learning approaches)
Structural template methods