Lecture 01

Jens Meiler

VANDERBILT UNIVERSITY
Center for Structural Biology and Institute of Chemical Biology
Departments of Chemistry, Pharmacology, and Biomedical Informatics
Central Dogma of Evolution

DNA  ➔  Function
Central Dogma of Structural Biology

DNA → Function

RNA → Transcription

Sequence → Folding

Structure
Given a protein’s AA sequence, what is its 3-dimensional fold, and how does it get there?

Assume 100 conformations for each amino acid in a 100 amino acid protein ⇒ 10^{200} possible conformations!

Cyrus Levinthal’s paradox of protein folding, 1968.

Given a protein fold, which primary sequence(s) fold into it?

Assume a total of 100 conformations for all 20 natural occurring amino acids side chains in a 100 amino acid protein ⇒ 10^{200} possible conformations!

Earth is less than 10^{10} years old.
More than 50% of our Therapeutics Target Membrane Proteins
Central Dogma of Molecular Biology
DNA $\Rightarrow$ mRNA $\Rightarrow$ Protein
Twenty Genetically Encoded Amino Acids
Hydrophobic Amino Acids

Group A: Nonpolar Amino Acids (Hydrophobic)

Glycine (Gly)
Alanine (Ala)
Valine (Val)
Leucine (Leu)
Isoleucine (Ile)
Methionine (Met)
Phenylalanine (Phe)
Tryptophan (Trp)
Proline (Pro)
Hydrophylic Amino Acids

Group B: Polar, Uncharged Amino Acids (Hydrophilic)

- Serine (Ser)
- Threonine (Thr)
- Cysteine (Cys)
- Tyrosine (Tyr)
- Asparagine (Asn)
- Glutamine (Gln)

Group C: Polar, Charged Amino Acids (Hydrophilic)

- Aspartate (Asp)
- Glutamate (Glu)
- Lysine (Lys)
- Arginine (Arg)
- Histidine (His)
Hydrogen Bonding Capabilities of Amino Acids
Properties of Amino Acids in Numbers

### Table 1 Amino acid parameter sets

<table>
<thead>
<tr>
<th>Name</th>
<th>$\Xi^a$</th>
<th>$\alpha^b$</th>
<th>$\nu_v^c$</th>
<th>$\pi^d$</th>
<th>$I^e$</th>
<th>$\alpha^f$</th>
<th>$\beta^g$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALA</td>
<td>1.28</td>
<td>0.05</td>
<td>1.00</td>
<td>0.31</td>
<td>6.11</td>
<td>0.42</td>
<td>0.23</td>
</tr>
<tr>
<td>GLY</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>6.07</td>
<td>0.13</td>
<td>0.15</td>
</tr>
<tr>
<td>VAL</td>
<td>3.67</td>
<td>0.14</td>
<td>3.00</td>
<td>1.22</td>
<td>6.02</td>
<td>0.27</td>
<td>0.49</td>
</tr>
<tr>
<td>LEU</td>
<td>2.59</td>
<td>0.19</td>
<td>4.00</td>
<td>1.70</td>
<td>6.04</td>
<td>0.39</td>
<td>0.31</td>
</tr>
<tr>
<td>ILE</td>
<td>4.19</td>
<td>0.19</td>
<td>4.00</td>
<td>1.80</td>
<td>6.04</td>
<td>0.30</td>
<td>0.45</td>
</tr>
<tr>
<td>PHE</td>
<td>2.94</td>
<td>0.29</td>
<td>5.89</td>
<td>1.79</td>
<td>5.67</td>
<td>0.30</td>
<td>0.38</td>
</tr>
<tr>
<td>TYR</td>
<td>2.94</td>
<td>0.30</td>
<td>6.47</td>
<td>0.96</td>
<td>5.66</td>
<td>0.25</td>
<td>0.41</td>
</tr>
<tr>
<td>TRP</td>
<td>3.21</td>
<td>0.41</td>
<td>8.08</td>
<td>2.25</td>
<td>5.94</td>
<td>0.32</td>
<td>0.42</td>
</tr>
<tr>
<td>THR</td>
<td>3.03</td>
<td>0.11</td>
<td>2.60</td>
<td>0.26</td>
<td>5.60</td>
<td>0.21</td>
<td>0.36</td>
</tr>
<tr>
<td>SER</td>
<td>1.31</td>
<td>0.06</td>
<td>1.60</td>
<td>-0.04</td>
<td>5.70</td>
<td>0.20</td>
<td>0.28</td>
</tr>
<tr>
<td>ARG</td>
<td>2.34</td>
<td>0.29</td>
<td>6.13</td>
<td>-1.01</td>
<td>10.74</td>
<td>0.36</td>
<td>0.25</td>
</tr>
<tr>
<td>LYS</td>
<td>1.89</td>
<td>0.22</td>
<td>4.77</td>
<td>-0.99</td>
<td>9.99</td>
<td>0.32</td>
<td>0.27</td>
</tr>
<tr>
<td>HIS</td>
<td>2.99</td>
<td>0.23</td>
<td>4.66</td>
<td>0.13</td>
<td>7.69</td>
<td>0.27</td>
<td>0.30</td>
</tr>
<tr>
<td>ASP</td>
<td>1.60</td>
<td>0.11</td>
<td>2.78</td>
<td>-0.77</td>
<td>2.95</td>
<td>0.25</td>
<td>0.20</td>
</tr>
<tr>
<td>GLU</td>
<td>1.56</td>
<td>0.15</td>
<td>3.78</td>
<td>-0.64</td>
<td>3.09</td>
<td>0.42</td>
<td>0.21</td>
</tr>
<tr>
<td>ASN</td>
<td>1.60</td>
<td>0.13</td>
<td>2.95</td>
<td>-0.60</td>
<td>6.52</td>
<td>0.21</td>
<td>0.22</td>
</tr>
<tr>
<td>GLN</td>
<td>1.56</td>
<td>0.18</td>
<td>3.95</td>
<td>-0.22</td>
<td>5.65</td>
<td>0.36</td>
<td>0.25</td>
</tr>
<tr>
<td>MET</td>
<td>2.35</td>
<td>0.22</td>
<td>4.43</td>
<td>1.23</td>
<td>5.71</td>
<td>0.38</td>
<td>0.32</td>
</tr>
<tr>
<td>PRO</td>
<td>2.67</td>
<td>0.00</td>
<td>2.72</td>
<td>0.72</td>
<td>6.80</td>
<td>0.13</td>
<td>0.34</td>
</tr>
<tr>
<td>CYS</td>
<td>1.77</td>
<td>0.13</td>
<td>2.43</td>
<td>1.54</td>
<td>6.35</td>
<td>0.17</td>
<td>0.41</td>
</tr>
</tbody>
</table>

*a Steric parameter (graph shape index)  
*b Polarizability  
*c Volume (normalized van der Waals volume)  
*d Hydrophobicity  
*e Isoelectric point  
*f Helix probability  
*g Sheet probability
Chemical Character of Amino Acid Side Chain

Amino Acids

A alanine (ala)
R arginine (arg)
N asparagine (asn)
D aspartic acid (asp)
C cysteine (cys)
Q glutamine (gln)
E glutamic acid (glu)
G glycine (gly)
H histidine (his)
I isoleucine (ile)
L leucine (leu)
K lysine (lys)
M methionine (met)
F phenyalanine (phe)
P proline (pro)
S serine (ser)
T threonine (thr)
W tryptophan (trp)
Y tyrosine (tyr)
Stereochemistry of amino acids and Planarity of peptide bond
Backbone Degrees of Freedom
Sidechain Degrees of Freedom

Asp
Peptide Bond Formation and Backbone Degrees of Freedom
Protein Folding
Protein Folding

Protein folding and misfolding, Christopher M. Dobson, Nature 426, 884-890
Protein structure depends on amino acid sequence and interactions.

(a) Primary structure

Backbone hydrogen bonds

(b) Secondary structure

Side chain interactions

(c) Tertiary structure
Secondary Structure: Build from Backbone Hydrogen Bonds

- \( \alpha \)-Helix:
  - Periodicity = 3.6
  - Rise = 1.5Å
  - Pitch = 5.4Å

- \( \beta \)-Sheet:
  - Periodicity = 2
  - Translation = 3.4Å
  - Distance = 5.4Å
**α-Helix**

- Most abundant secondary structure
- 3.6 amino acids per turn
- Hydrogen bond formed between every fourth residue
- Average length: 10 amino acids, or 3 turns
- Varies from 5 to 40 amino acids
β-Sheet

- 5-10 amino acids in one portion of the chain with another 5-10 farther down the chain
- Interacting regions may be adjacent with a short loop, or far apart

Parallel β-sheet

Anti-Parallel β-sheet
The Ramachandran Plot

Beta-sheet

Left handed alpha-helix.

Right handed alpha-helix.

http://www.cryst.bbk.ac.uk/PPS2/course/section3/rama.html
Protein structure depends on amino acid sequence and interactions.
Important bonds for protein folding and stability

The oxidization of the sulfhydryl groups of two cystein residues (intramolecule: ribonuclease; intersubunit; dimeric protein insulin)

Weak (2-5 kcal/mol vs. covalent: 70-100 kcal/mol), but massive

Weak (3 kcal/mol), affected by pH value

Dipole molecules attract each other by van der Waals force (transient and weak: 0.1-0.2 kcal/mol)

Hydrophobic interaction, a tendency of hydrophobic groups or molecules being excluded from interact with hydrophilic environment
Tertiary structure: Assembly of Secondary Structure in Domains

**Domains**
Discrete locally folded units of tertiary structure, often containing regions of alpha helix and beta sheets packed together compactly, typically 50-350 aa in length and usually has a specific function.

**Tobacco mosaic coat protein**
- Predominantly α helix

**Immunoglobulin, V₂ domain**
- Predominantly β sheet

**Hexokinase, domain 2**
- Mixed α helix and β sheet
These coiled coils have a heptad repeat \textit{abcdefg} with nonpolar residues at position a and d and an electrostatic interaction between residues e and g.

Originally identified in the liver transcription factor C/EBP which has a Leu at every seventh position in a 28 residue segment.
Tertiary Structure: The Leucine Zipper
Tertiary Structure: Helix-Helix Interactions “Coiled Coils”

\[ \text{Helix-Helix Interactions “Coiled Coils”} \]
Tertiary Structure: Sheet-Sheet Packing

- Ideal beta sheet is planar and flat. In the context of proteins, \( \beta \) sheets have a right-handed twist possibly due to interaction between backbone and the chiral L-amino acid side chains.
Tertiary Structure: Sheet-Sheet Packing

\((-\text{Gly-Ser-Gly-Ala-Gly-Ala-})_n\)
These motifs occur quite frequently in Nature. Theoretical studies suggest that this motif may have originated from one long antiparallel structure with loops in the middle of both strands.

- **topology of the Greek key**
Tertiary Structure: Helix-Sheet Packing – The Rossman Fold

The most regular and common domain structures consist of repeating $\beta$-$\alpha$-$\beta$ units. The outer layer of the structure is composed of $\alpha$ helices packing against a central core of parallel $\beta$ sheets. These folds are called $\beta/\alpha/\beta$. This motif is always right-handed.
Tertiary Structure: Helix-Sheet Packing – The Barrel Fold

- This structural motif was first observed in the X-ray structure of triosephosphate isomerase (TIM) so it is also called a TIM fold. (Banner et al., Nature 255, 609-614 (1975).
Protein Tertiary Structure is Tied to Function
Quaternary Structure: Assembly of Domains

- Glyceraldehyde phosphate dehydrogenase:
  - domain 1 for binding the substance being metabolized
  - domain 2 for binding a cofactor
Quaternary Structure: Assembly of Domains

- Hemoglobin
Protein Sequence and Structure Data

- **Genbank**
  - ~5,000,000 sequences

- **Protein Databank**
  - ~60,000 structures
Structural Biology After the Human Genome Project

- Sequence versus Structure

![Graph showing protein sequences and structures from 1990 to 2006](Science_cover.jpg)
Structural Biology After the Human Genome Project

- Sequence versus Structure

![Graph showing the increase in protein sequences and protein structures over time.](image-url)
Membrane Proteins and Large Macromolecular Assemblies

- **Sequences:** 1.0E+06 soluble proteins, 3.0E+06 membrane proteins
- **Structures:** 41298 soluble proteins, 777 membrane proteins, 455 macromolecular assemblies
- **Folds:** 1800 soluble proteins, 54 membrane proteins, 273 macromolecular assemblies

© Jens Meiler
More than 50% of our Therapeutics Target Membrane Proteins
β Adrenergic Receptors as Model System for Drug-Receptor Interactions

- β adrenergic receptors function in sympathetic nervous system activation (epinephrine / adrenaline hormones)
- First x-ray crystal structure of a GPCR with a diffusable ligand - clinical drug carazolol.
- Carazolol (β blocker drugs) indicated for treatment of hypertension, cardiac arrhythmias, angina, glaucoma, migraine headaches, anxiety
Allosteric Modulators of mGluR<sub>5</sub> Bind to the 7TM Domain of the GPCR

Glutamate binding site
- Site 1: MPEP / DFB / CDPPB / VU-29
- Site 2: CPPHA binding site
A mGluR5 Comparative Model based on β-2-Adrenergic Receptor (2RH1)

- ROSETTA constructed mGluR5 Comparative Model for 7TM domain
- Based on crystal structure of β-2-Adrenergic Receptor (2RH1)
- Sequence identity in the TM spans is ~12%, similarity is ~35%
- From this model S806, S808, T810, E578 are hypothesized as potential interaction sites for CPPHA and will be tested experimentally
Treatment Strategies for CNS Disorders through Modulation of mGluR$_5$

- Allosteric positive modulation (activation) of mGluR$_5$ may ameliorate the symptoms of schizophrenia.

- Allosteric negative modulation of mGluR$_5$ offers a potential treatment strategy of fragile X syndrome symptoms, a CNS disorder associated with autism spectrum disorders (ASD).
150,000 compounds were tested for allosteric modulation of mGluR$_5$ measuring receptor-induced intracellular release of calcium. 1,387 (0.94%) compounds were verified as PAMs of mGluR$_5$. 345 (0.23%) compounds were verified as NAMs of mGluR$_5$.

Accelerate Discovery of Leads and Design of novel Therapeutics

- ~150,000 compounds physically available for high-throughput screening in VICB screening center
- »10,000,000 substances are readily available for purchase
- Chemical Abstracts Services knows ~50,000,000 chemicals
Relate Chemical Structure and Biological Activity

Chemical Structure

Biological Activity
Transformation-Invariant, Problem-Optimized Numerical Description

Chemical Structure

<table>
<thead>
<tr>
<th>I(s)</th>
<th>a)</th>
<th>b)</th>
<th>c)</th>
<th>d)</th>
<th>e)</th>
<th>f)</th>
</tr>
</thead>
</table>

Biological Activity

6 December 2011 © Jens Meiler
Machine Learning Calculates Activity from Numerical Description

Chemical Structure

I(s)

a) b) c) d) e) f)

Biological Activity

% Maximal Glutamate

6 December 2011 © Jens Meiler
Radial Distribution Functions describe 3D shape ...

where:

\( d_{ij} \) – distance between two atoms

\( B \) – temperature factor, here 100
Experimental Results mGluR$_5$ Negative Allosteric Modulators

749 Compounds with novel Scaffolds predicted with EC$_{50}$ < 10μM by QSAR model

12 Compounds (3.6%) were confirmed as mGlur$_5$ NAMs
Enrichment = 3.6% / 0.23% = 16

VU0240790-4
EC$_{50}$ = 75 nM

VU0360620-1
EC$_{50}$ = 124 nM
Steric Bulk Map for mGluR5 PAM
Benzoxazapine

- increase
- neutral
- decrease
Computational Biology

Physical & Theoretical Chemistry

Biochemistry

Chemical Biology

Molecular Biology

Structural Biology

Bioinformatics / Biostatistics

Computer Science / Control Theory

Mathematics