Protein Modeling

Examples of Protein Modeling

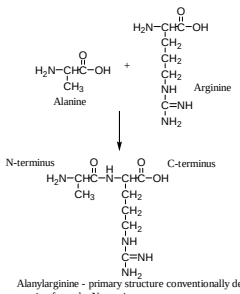
- Visualization
  - Examination of an experimental structure to gain insight about a research question
- Dynamics
  - To examine the dynamics of protein structures
  - To examine binding free energy differences of ligands
- Docking
  - To explore fit of a small molecule against a protein
- Computational Model Development
  - Protein structure prediction from sequence
  - Homology modeling

Protein Structure Description

Protein Structure can be Described at Several Levels

- Primary
  - Linear sequence of amino acids in protein chain
- Secondary
  - Three-dimensional local conformation
- Tertiary
  - Overall fold of an entire protein chain
- Quaternary
  - Overall shape of a multi-chain protein

Primary Structure



Alanylarginine - primary structure conventionally described starting from the N-terminus

Protein Sequence Sources

- GeneBank: www.ncbi.nlm.nih.gov/Entrez/
- Protein Databank: www.rcsb.org (not limited to primary structure)
- Swiss-Prot: www.expasy.ch/sprot/

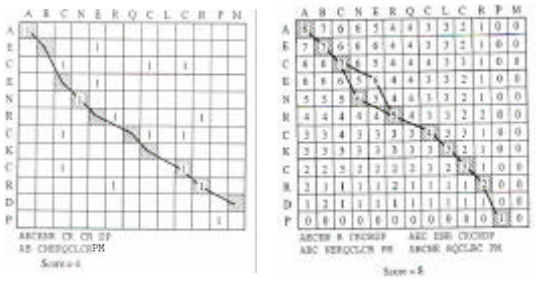
Importing Sequences to MOE

- Option I
  - Copy the sequence into a text file
  - Insert a line prior to the sequence that starts with >, followed by a name for the sequence
  - Save the text file with a .fasta extension
- Option II
  - Display the file as fasta in the source database
  - Save with a .fasta extension
- Open the file in MOE, and view the sequence from the sequence editor

Class Exercise I

- Use one of the protein databases to locate several related protein sequences (ask for suggestions if you aren't currently interested in any proteins) – One guarantee of relation is to find proteins with the same name from different species
- Import them into MOE
- Make sure you know how to select residues (single residues, continuous stretches of residues, and scattered residues)

Sequence Alignment



Class Exercise II

- Align the sequences that you have imported into MOE
- From the sequence window choose
  Homology->Align
- Open the commands window to see the pairwise residue identities for your aligned sequences

Protein Secondary Structure

- Examples
  - Alpha Helix
  - Beta sheet
  - Beta turn
- Major stabilizing contributions
  - Hydrogen bonding
  - Relief of steric crowding

Useful MOE Tools

- Sequence Window
  - Display menu allows you to highlight actual secondary structures (red=helix, yellow=sheet)
  - Display menu allows you to highlight hydrogen bonding (generally only for the backbone)
- Main Window
  - Render->Draw menu allows you to show hydrogen bonds and protein ribbon diagram

Notes on Experimental Structures

- Experimental protein structures are determined mainly by two methods:
  - NMR
    - Often by ^{1}H-^{1}H NOE enhancements
    - Structures are then modeled to be consistent with the distance data derived from the spectra
  - X-ray crystallography
    - X-ray diffraction patterns are dependent on electron density
    - Hydrogen atoms have negligible electron density and are not present in x-ray structures
    - The O and N of a terminal amide have similar electron density and are often placed on the basis of expected hydrogen bonding

Class Exercise III

- Download a protein structure from the Protein Databank and add hydrogens (Edit->Add Hydrogens)
- Isolate an alpha-helical secondary structure
  - Examine hydrogen-bonding in that region
  - Examine spacing of the amino acid sidechains
- Isolate a beta-sheet secondary structure (you will need non-contiguous parts of the sequence) and examine similarly

Protein Tertiary Structure

- Covalent stabilization:
  - Disulfide bond formation
- Non-covalent stabilization:
  - Hydrophobic sequestration from water (entropy driven)
  - Salt bridge formation (enthalpy driven)
  - Hydrogen bonding (enthalpy driven)

pH, pKa and Ionizability

Electrostatic (charge) interactions are an important feature of protein structure and activity - models cannot ignore ionizability and charge

- Ammonium pKa ~9, thus will remain cationic in water - this generalizes to amine groups in proteins
- Carboxylic acid pKa's ~5, thus will lose their protons in water - this generalizes to carboxylic acid groups in proteins
- Check out the section on amino acids in your text for other groups to be aware of

Class Exercise IV

- Show all atoms of your protein as a space-filling model
- Select all atoms that are part of hydrophobic residues
- Visually decide how well they are screened away from the solvent

Class Exercise V

- Isolate all cysteine residues in your protein
- Are any involved in disulfide linkages?
- Isolate all cationic and anionic residues in your protein
- Are any near enough in space to be participating in salt bridges?

Crystallographic Structures

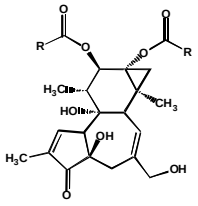
- Crystallography does not identify hydrogen positions - they must be added
- Ionization of standard residues will be handled automatically (groups with pKa's near 7, like histidine, should be manually checked)
- Residues may be unresolved (missing)
- No partial charges included, must be assigned
- Non-standard residues may be incorrectly assigned atom types
- Resolution and crystal packing effects contribute to the fact that the structures are NOT energy minimized in your forcefield!

Example: Missing Residues

- Download PDB entry 1f88
- Identify regions in both chains that are missing amino acids
  - Chain 1
    - residue 235 is not attached to residue 236
  - Chain 2
    - residue 142 is not attached to residue 143
    - Residue 221 is not attached to residue 222

Example: Incorrect Atom Types

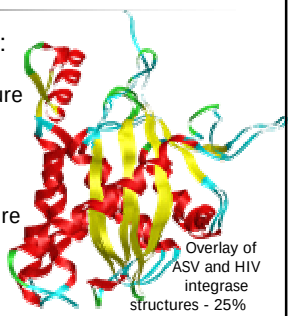
- Download PDB entry 1ptr
- Add hydrogens
- Isolate the ligand and draw its structure



Homology Modeling

- Founding Assumption:

homologous primary structure
AND
homologous function
INDICATE
homologous tertiary structure



Overlay of ASV and HIV integrase structures - 25% sequence homology

Homology Modeling Needs

- Alignment of:
  - Template sequence with known structure
  - Target sequence with known sequence
- Knowledge about the function of both proteins
  - Knowing residues critical for function allows examination of homology in those regions – should be higher than overall homology
  - Greater functional homology indicates likelihood that proteins have a common ancestor

Homology Modeling Method

- For regions of identical length:
  - Protein target backbone is taken from template
  - Identical target sidechains are taken from template
  - Nonidentical target sidechains are derived from library
- For target sequence insertions (indels)
  - High resolution structures from PDB are scanned to find those that have regions that superpose on the anchor residues on either side

Class Exercise VI

- Search the PDB for one of the proteins in your set from exercises I and II
- Use this protein as your template structure, and any of the other sequences as your target
  - (If you find two in the pdb, use one as the template and the other as the target -> you can compare your modeled structure with the experimental structure when you are done)
- Build a homology model (Homology ->Homology Model)

Evaluating Protein Models

- Final structures after minimization may have structural features inconsistent with known protein structural characteristics
  - Examples
    - Cis-amide bonds at locations other than proline
    - Incorrect stereochemistry at alpha or beta carbons
    - Strained bond angles
    - Dihedral angles that match unpopulated regions of a Ramachandran plot
    - Steric bumps
  - Evaluation tools accessible in the sequence window
    - Measure->Protein Report
    - Measure->Protein Contacts

Class Exercise VII

- Evaluate your protein model and identify
  - Steric problems
  - Residues connected by cis-amide bonds
  - Dihedrals outside the expected ranges
  - Reversed stereochemistry at alpha carbons

Reading

- First Edition
  - Section 8.13
- Second Edition
  - Section 9.10
  - Chapter 10