Protein physics: potential functions

Betty Cheng
BioInformatics Centre
National University of Singapore

What mathematical representations do we need?

- Can we visualize the protein structure?
- Can we duplicate protein motion?
- Can we evaluate protein folding?
- Can we predict protein structure?

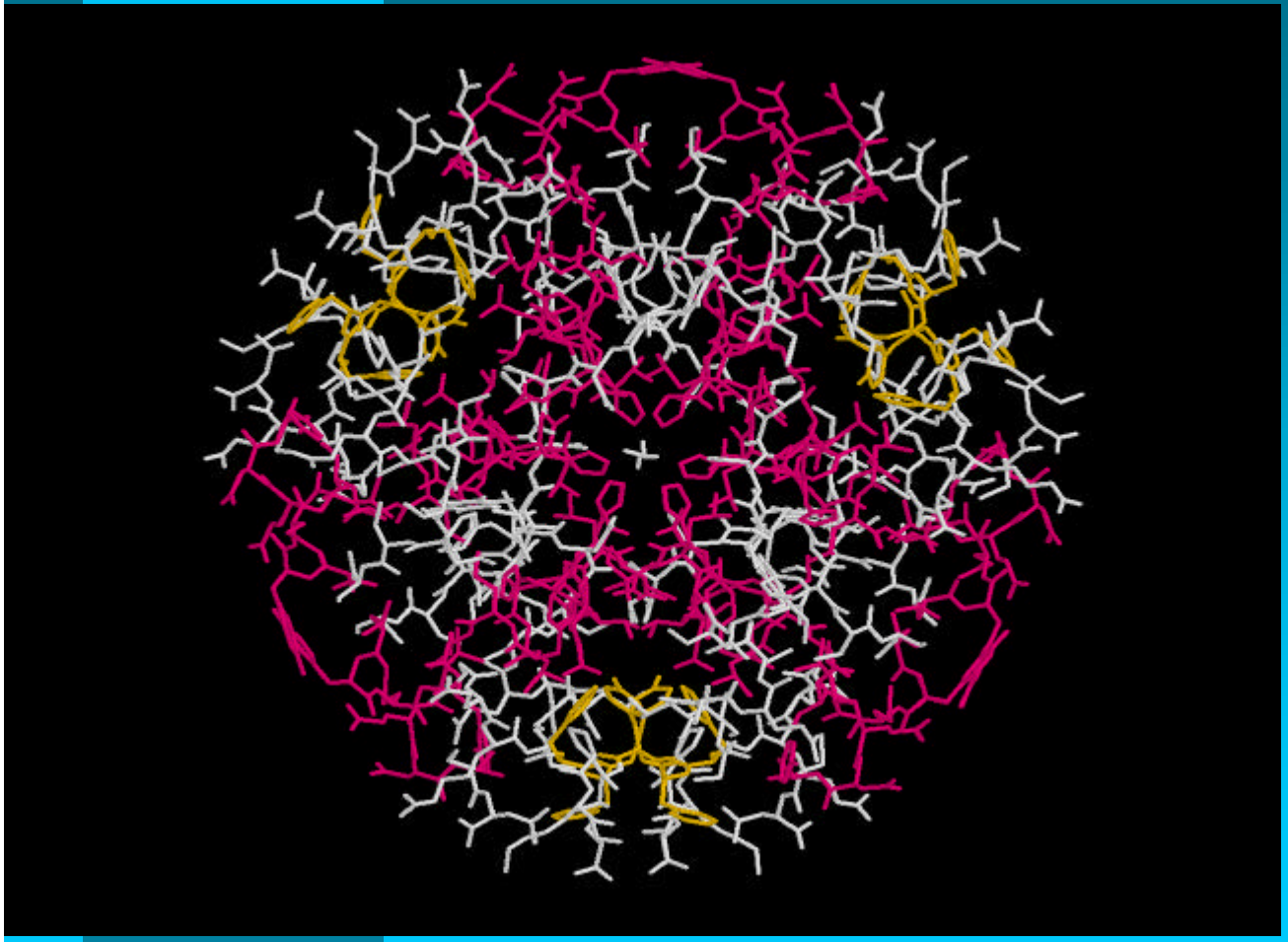
Representing reality in a computer

- In order to mimic a protein and the way it behaves in a computer, we need the following:
  - A coordinate system
  - Protein measurements, bond lengths, bond angles, atom types, atomic radii, etc…
  - Mathematical representations of the physical forces that drive protein behavior

Looking at Protein Structure

- Regular geometry/secondary structure
  - Alpha helices, beta sheets, loops
- Protein shape
  - Molecular volume, molecular surface
- Internal protein interactions
  - Hydrogen bonds
  - Internal contacts
- Different representations are better suited to display this information visually

Protein Views: Insulin



- The “wireframe” view of a protein.
- Every atom is shown with a line drawn representing the covalent bonds between atoms
- Best view for looking at details of internal interactions

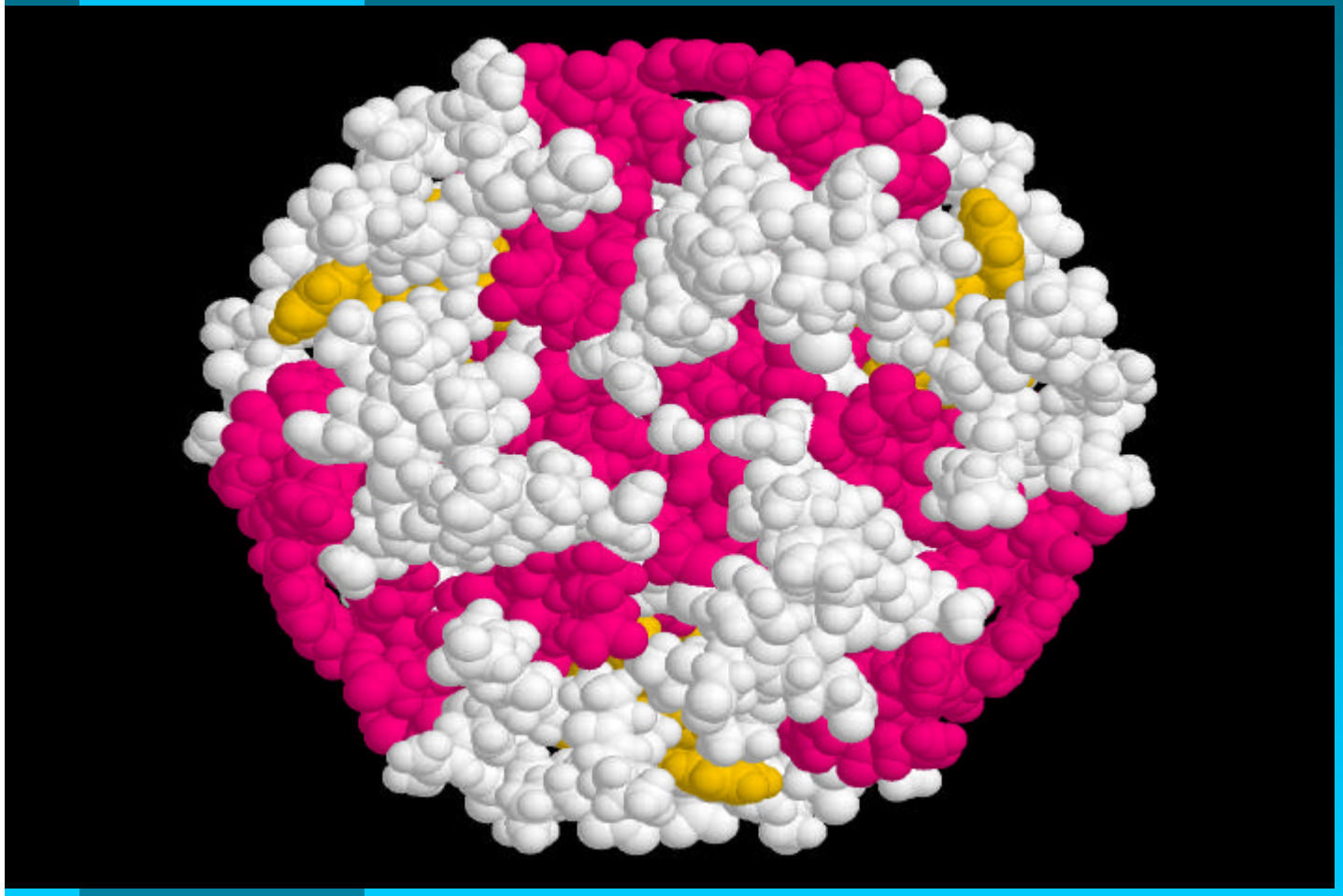
This picture was made using the Rasmol program

Protein Views: Insulin

- This is the “cartoon” view of insulin.
- Alpha helices are shown in red. Beta sheets are shown in yellow, the arrow in the sheet shows the direction of the beta sheet (from N-terminal to C-terminal).

Picture produced by Rasmol

Protein Views: Insulin



- The “spacefill” view of the insulin hexamer
- This view is well-suited to display the shape and surface details of the protein. Alpha helices (red), Beta sheets (yellow).

Picture produced by Rasmol

Protein Views: Insulin

This is the “backbone” view. Rods have been drawn from each C-alpha carbon

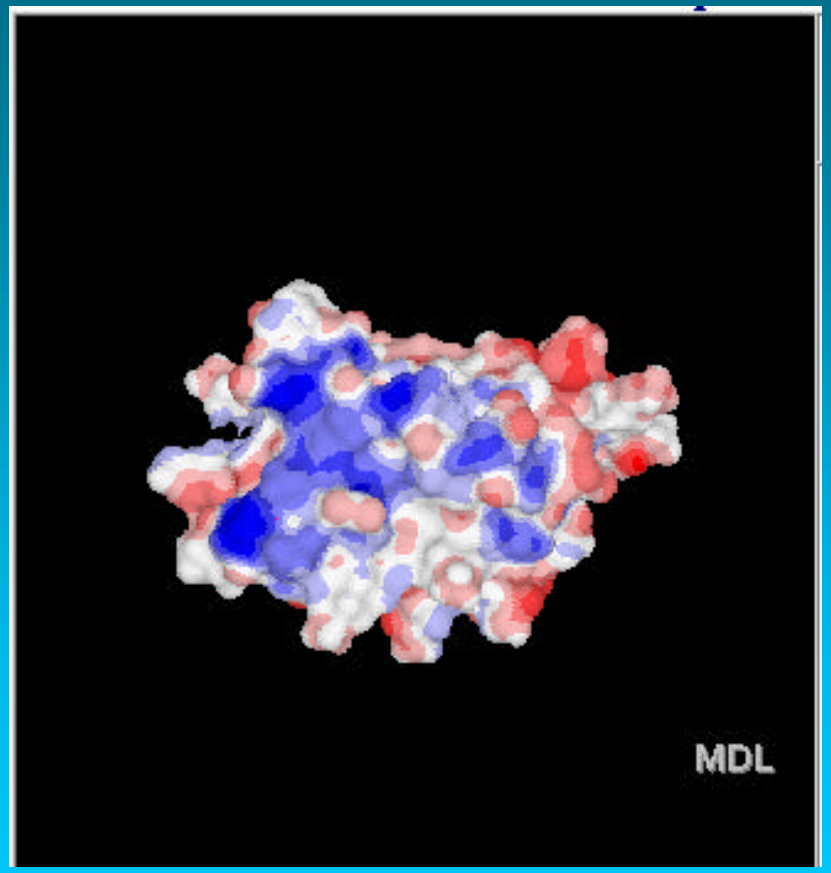
This is a good representation for an overall view of the protein

Picture made with Rasmol

View of an electrostatic surface

- Binding domain of Streptococcal protein G

- Electrostatic surface:
  - shows the location of the charges on the molecular surface of the DNA.
- Alternatively, the surface could be colored according to the lipophilicity or hydrophobicity of the amino acids.
- This type of representation is extremely helpful for analysis of intermolecular interactions, such as, drug binding.



(Picture produced by CHIME from MDLI)

Visualization

- Need to define coordinate system for the proteins.
  - For visualization, cartesian coordinate system is usually adequate
- Need ways to calculate protein representations such as the molecular surfaces which are needed for the electrostatic surfaces
  - This is beyond the scope of this lecture.
- Need functions to calculate the electrostatic surface, lipophilic or hydrophobic surface (and more?).

Protein physics: potential functions

Betty Cheng
BioInformatics Centre
National University of Singapore

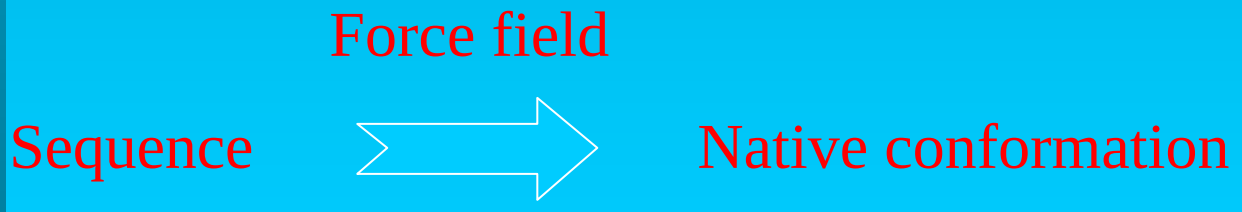
Mathematical representations needed!

- **Can we visualize the protein structure?**
- **Can we duplicate protein motion?**
- **Can we evaluate protein folding?**
- **Can we predict protein structure?**

“Ab Initio” Protein Structure Prediction

Goal: Determine protein structure from amino acid sequence information alone

Theoretical Foundation: Anfinsen’s thermodynamic hypothesis, the native conformation of a protein is the conformation with the *lowest free energy*.



"Ab Initio" Protein Structure Prediction

- It is known that not all functional proteins are in the lowest free energy state
  - Example: prions (the diseased proteins from mad cow disease) are functional in a disordered state but the disease state is the energetically more stable beta sheet conformation.
- This method can only work on proteins that are not kinetically trapped.
  - Any protein that folds *independently* is probably in the lowest free energy state

Select Sequence

Force field

Native conformation

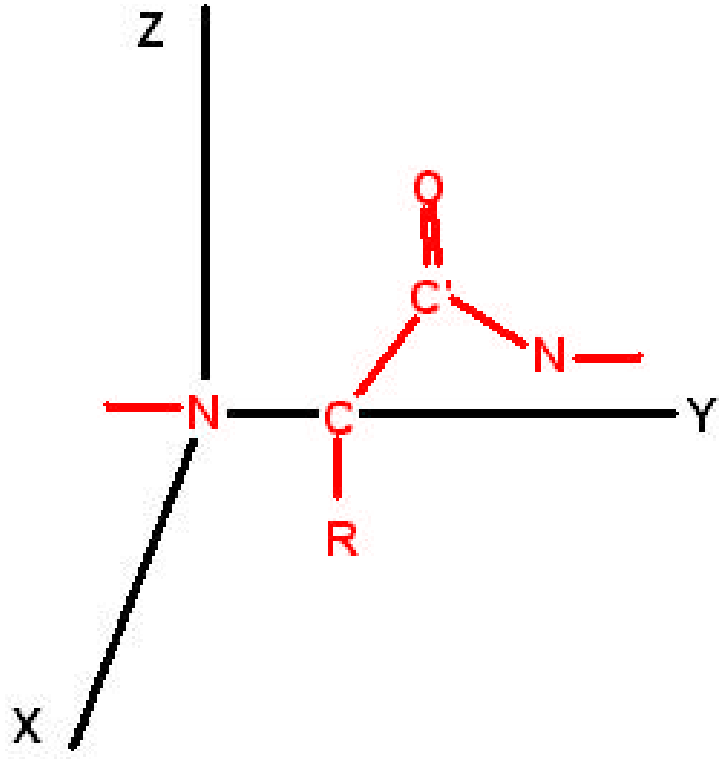
Virtual Proteins

- How do we calculate the free energy of a protein?
  - First approximation: Proteins are usually in aqueous solution. We will ignore the presence of the water for now
  - How do we calculate the energy of a protein?

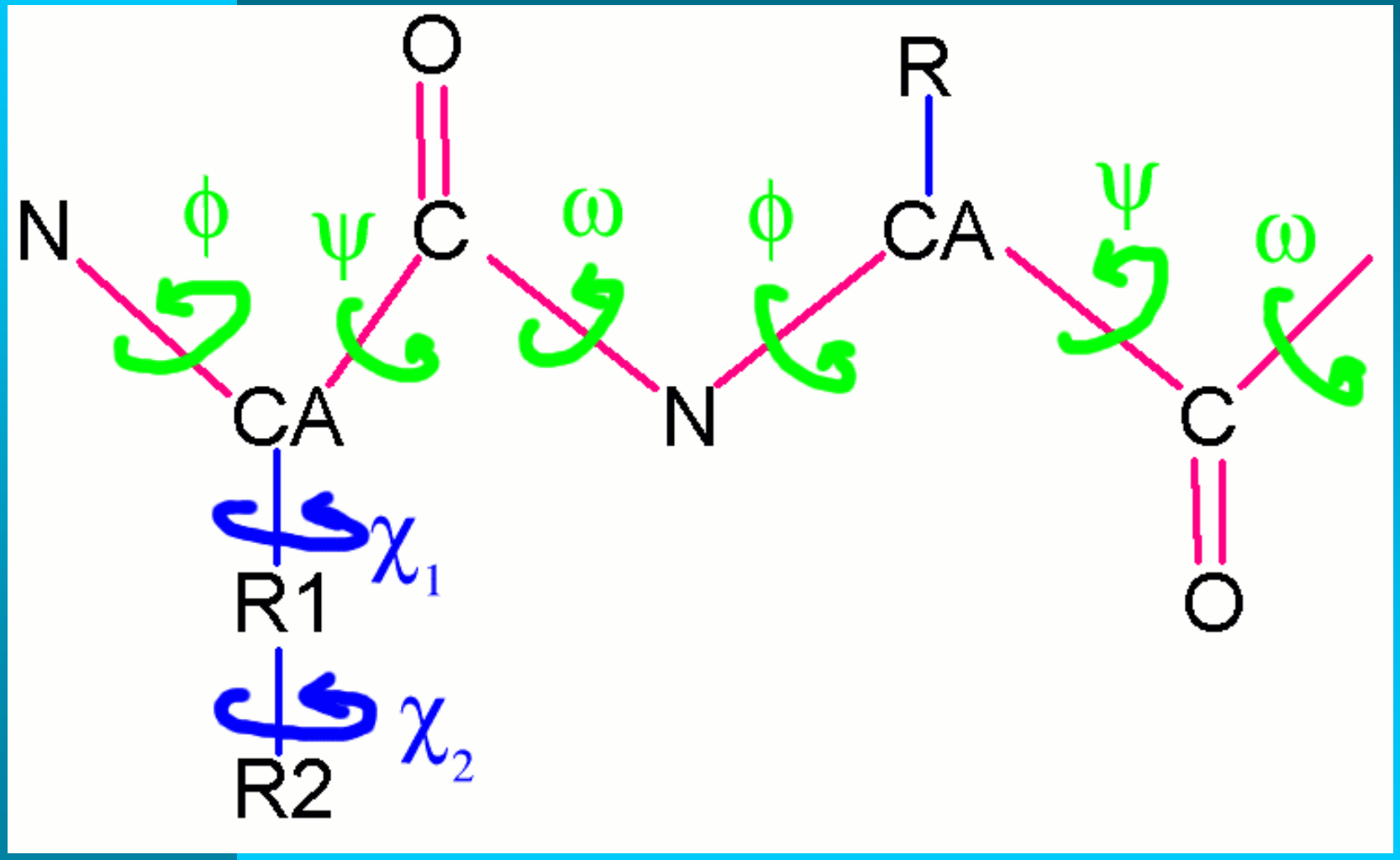
Define a coordinate system

- In order to calculate forces, we need to know the coordinates of every atom in the protein.
- Two possible choices:
  - Cartesian coordinates
  - Internal coordinates

Cartesian coordinates



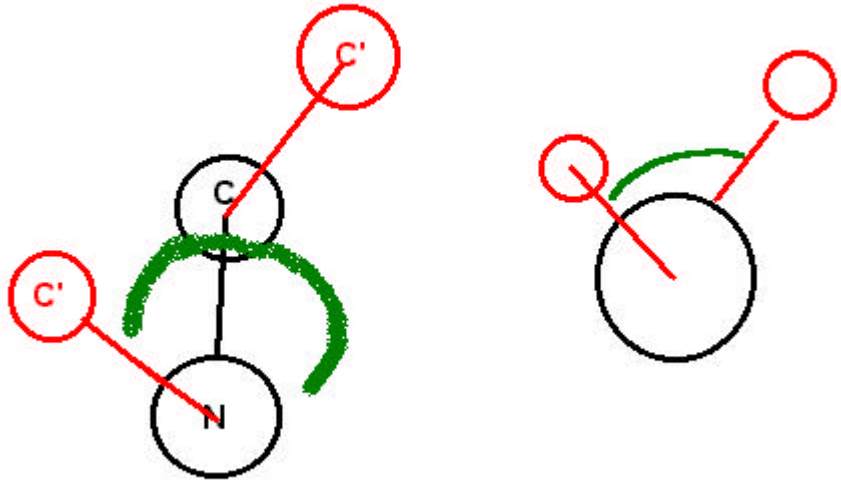
Covalent structure of proteins- Primary structure



ϕ, ψ, ω are the backbone dihedral angles; χ_1, and χ_2 are the sidechain dihedral angles

Internal coordinates

- Four points are required to define a dihedral angle
- The diagram below shows the definition of the **f** (phi) dihedral angle
- Assume standard bond lengths and bond angles
- Advantages of internal coordinates – only N variables for N atoms instead of 3N (cartesian); fewer variables=faster computations



phi (in green)

Protein physics: potential functions

Betty Cheng
BioInformatics Centre
National University of Singapore

Energy of a molecule

- From quantum mechanics: Schrödinger's Equation

$$H\Psi = E\Psi$$

- H is the Hamiltonian (an *operator*)
- Ψ is the wavefunction or eigenfunction
- E is the energy of the molecule (a constant)

Schrödinger's Equation

- The Hamiltonian H=

$$\{\frac{-h^2}{8\pi^2 M}(\nabla_n + \nabla_e) + \frac{-Ze^2}{r}\}$$

- Υ is any mathematical function that solves this equation!

- $\Upsilon(r_1, r_2, r_3, \ldots, r_n)$

- What is the problem?

Not enough CPU!

- Schrödinger's equation would give us the positions of the nuclei, distribution of the electron density, polarization effects, everything… but…
- At the current time, we can only solve Schrödinger's equation for smaller molecules.
- Use classical physics to approximate results from quantum mechanics to obtain faster estimates of the energy

Common potentials for protein structure work

- **ECEPP: Empirical Conformational Energy Program for Peptides (Scheraga and co-workers)**
- **CHARMM: Chemistry at Harvard Macromolecular Mechanics (Karplus and co-workers)**
- **AMBER: Assisted Model Building with Energy Refinement (Kollman and co-workers)**
- **OPLS: Optimized Potentials for Liquid Simulations (Jorgensen and co-workers)**

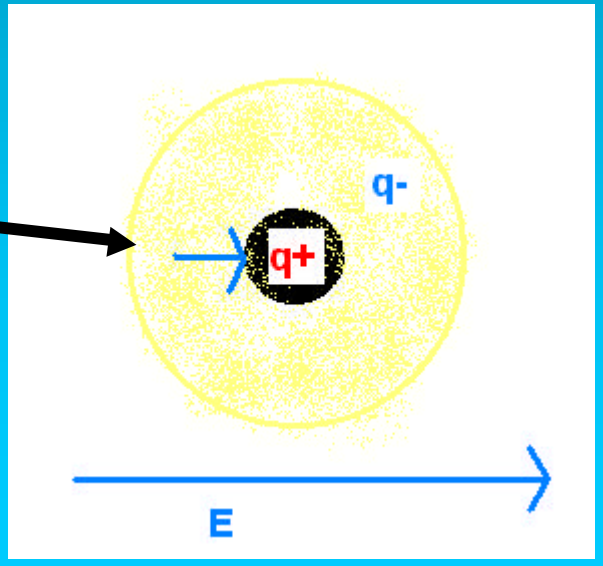
Semi-classical force fields

- **Assumptions**
  - Pairwise atomic interactions
  - The internal forces within the protein can be approximated by pairwise interactions (interactions between two atoms, i and j)
  - Summing over all pairwise interactions will give us an estimate of the internal protein energy
  - Many body forces will be neglected
  - How do we deal with the dielectric constant for water?
  - United atoms are occasionally used. Hydrogen is commonly included implicitly rather than explicitly.

Definition of dielectric

- When matter is exposed to an electric field, some materials act as insulators.
- Physically, the electron density in these atoms is *polarized* by the electric field.
- Below the effect of an electric field on a neutrally charged atom is shown: the nucleus has a charge of +q and the electron cloud, a charge of –q. The small blue arrow shows the direction of internal charge movement once the electric field, E, is applied.

- Induced dipole moment of an atom due to an external electric field.



Bulk matter

- For bulk matter, the effect of this internal induced polarization is known as the dielectric constant.
  - The dielectric constant for vacuum is 1
  - The dielectric constant for water is about 80
  - The dielectric constant for proteins is about 4.
- Dielectric constants are an approximation for bulk matter only. For molecular-level work, must use dipole moments instead.

Dielectric constant

- In order to calculate the potential for an electric charge:

$$V = \left(\frac{ze}{4\pi\varepsilon_0 K_R}\right)\frac{1}{r}$$

- K_r is the dielectric constant (or relative permittivity) and the value depends on the material

Protein physics: potential functions

Betty Cheng
BioInformatics Centre
National University of Singapore

ECEPP (Empirical Conformational Energy Program for Peptides)

$$E_{ecepp} = \frac{Cq_i q_j}{Dr_{ij}}$$

$$+ F\frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}}$$

$$+ \frac{A'_{HX}}{r_{HX}^{12}} - \frac{B_{HX}}{r_{HX}^{10}}$$

$$+ \frac{U_0}{2}(1 \pm \cos n\theta)$$

- **Electrostatic (the dielectric constant is hidden inside the parameter D, for ECEPP the value is 2)**
- **Non-bonded (van der Waals)**
- **hydrogen bond**
- **torsional energy**

Unusual features of the ECEPP force field

- **Atomic coordinates are expressed in dihedral angles NOT cartesian coordinates**
  - **Fix the atomic geometry to "standard" values found from library of protein structures => fixed bond lengths, fixed bond angles.**
  - **Requires more care with minimization procedure than other force fields due to fixed geometry**
- **Minimum number of variables for this problem**
  - **Least expensive to evaluate**

Electrostatics (non-bonded)

- **Problems with the electrostatic term**

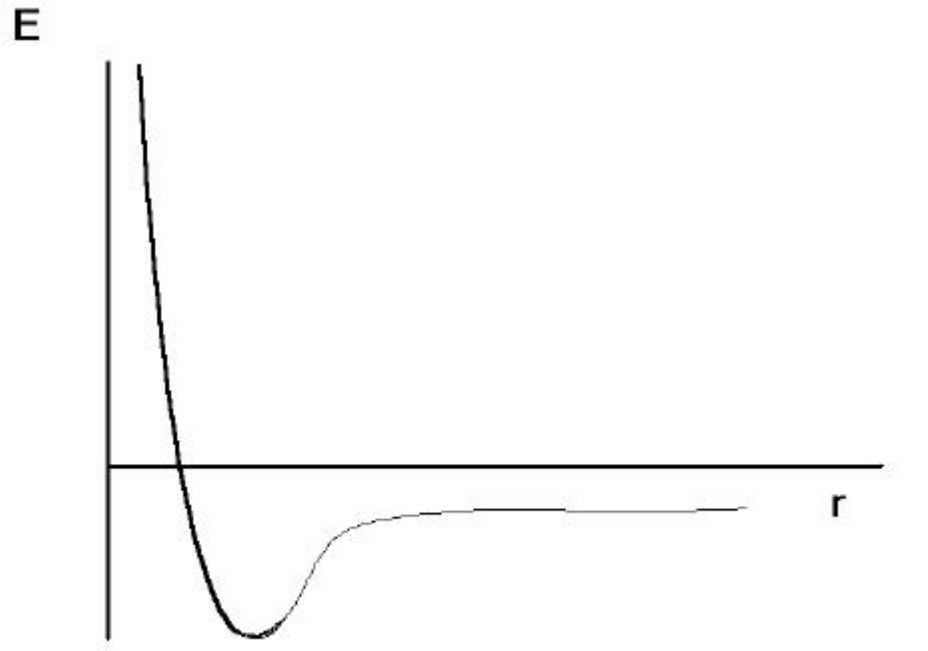
$$\frac{Cq_i q_j}{Dr_{ij}}$$

- **This term has static charges but electrons move (lose all polarization effects)**
- **Electrons can be thought of as a cloud of charge density but this equation works with *fixed number of point charges.***
- **The electrostatic approximation probably introduces the greatest error**

Van der Waals term (non-bonded)

- **Approximates the van der Waals forces**

$$F \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}}$$



- r is the distance between i and j
- At short distances, repulsion between electrons leads to high energies
- At intermediate distances, polarization leads to an attractive force

Hydrogen bond

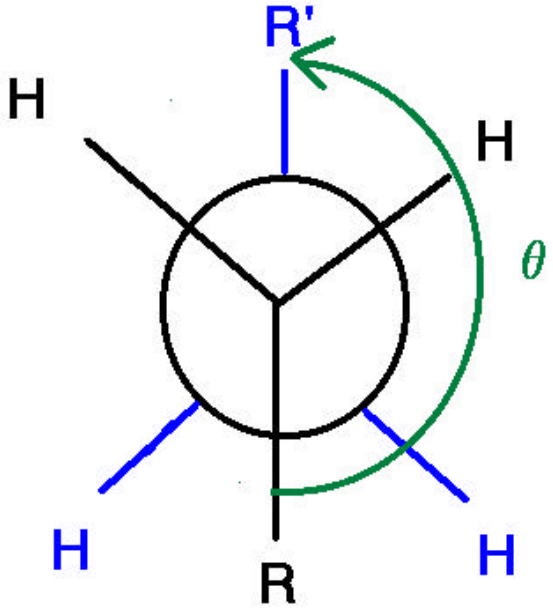
- **Unlike Charmm, the Ecepp hydrogen-bond term is not directional**

$$\frac{A'_{HX}}{r_{HX}^{12}} - \frac{B_{HX}}{r_{HX}^{10}}$$

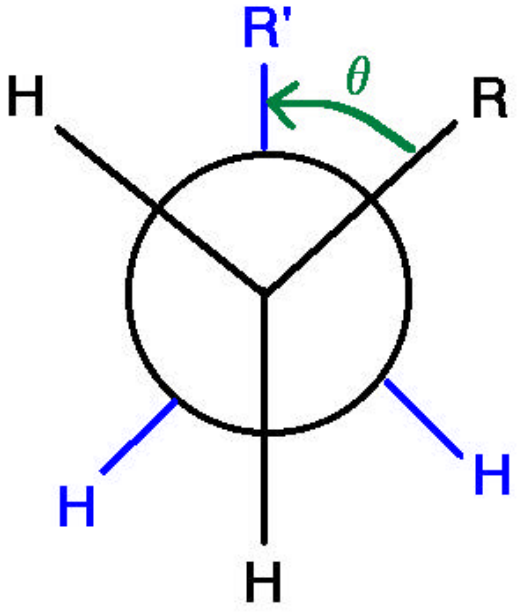
Torsional energy term

- Torsional energy
- Steric hinderance

$$\frac{U_0}{2}(1 \pm \cos n\theta)$$



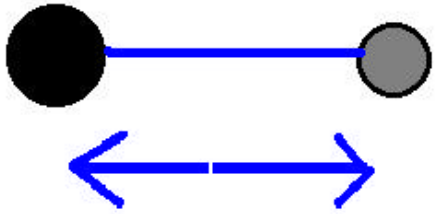
Anti or Trans

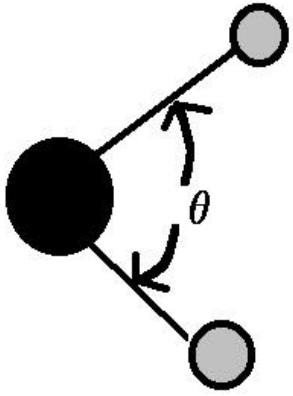


Gauche

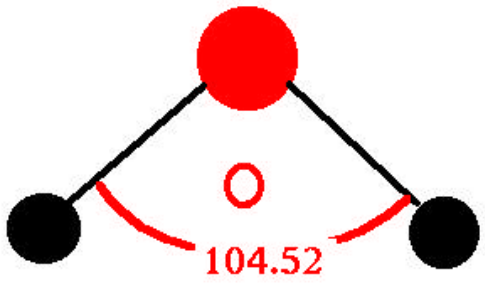
CHARMM

- extra energy terms below (in red)
- $E = E_{bonded} + E_{nonbonded}$
- $E_{bonded} = E_{stretch} + E_{angle} + E_{rotation}$





What about water?



- Add water explicitly
- Create a potential for water just like protein potential
- TIP4P

4 sites for point charges

$$V = \sum_i \sum_j \left(\frac{q_i q_j}{r_{ij}} + 4\varepsilon \frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - 4\varepsilon \frac{\sigma_{ij}^{6}}{r_{ij}^{6}} \right)$$

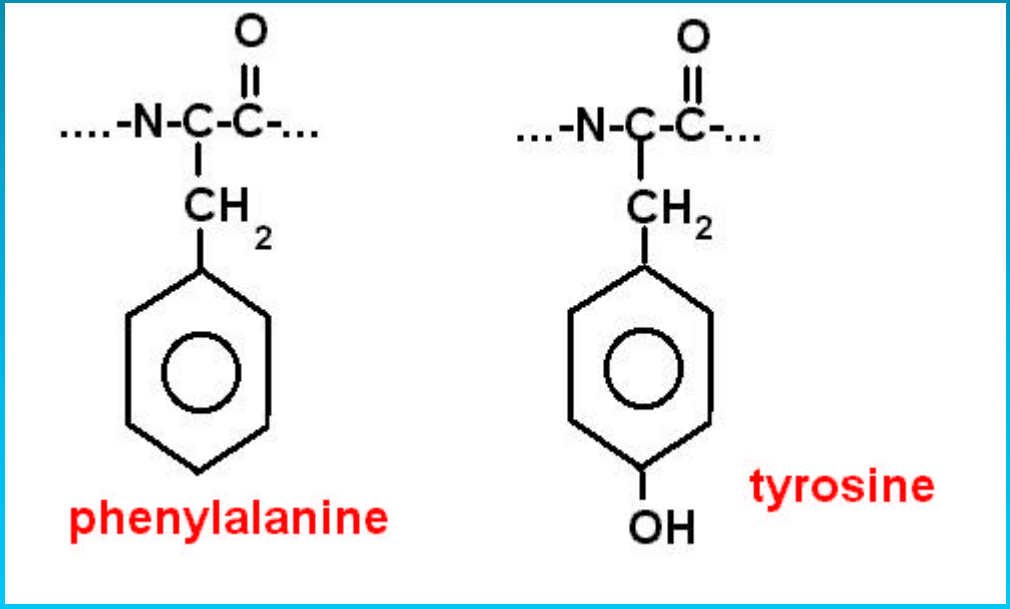
Potential Parameters

- How do we derive the parameters for these potentials?
  - Obtain experimental data
    - Crystal structures of small molecules
    - Heats of vaporization and densities for liquids
  - Do quantum mechanical calculations on small molecules to derive parameters for potentials
  - Run simulations of small molecules until simulated results match experimental results
  - Transfer parameters of chemically matching atoms from small molecule to protein
  - Example: CH_3SSCH_3
    - The parameters for the sulfurs from this small organic molecule can be used for disulfide bridges in proteins

$$F\frac{A_{ij}}{r_{ij}^{12}}-\frac{C_{ij}}{r_{ij}^{6}}$$

Drawbacks and assumptions

- Chemical potentials of peptides will be built up from small molecule parameters
- Assumed that free energies of chemical moieties are *additive*
- *Example: obtain estimate for –OH group from small molecule and add to phenylalanine energy*



- Limited experimental and more accurate quantum mechanical results suggest this approximation is reasonable

Conclusion

- Potentials in common use are far from perfect
- Many assumptions have been made to reduce computation time
- Other more sophisticated and accurate potentials are available but these are more expensive to use
- Other faster potentials with more approximations are also available
- More accurate potentials are needed