Computational Biology & Computational Medicine

Homayoun Valafar

CSCE 190

Outline

- Why proteins?
- What are proteins?
- How do we compute them?
- How do we use computational approaches?

Why Proteins?

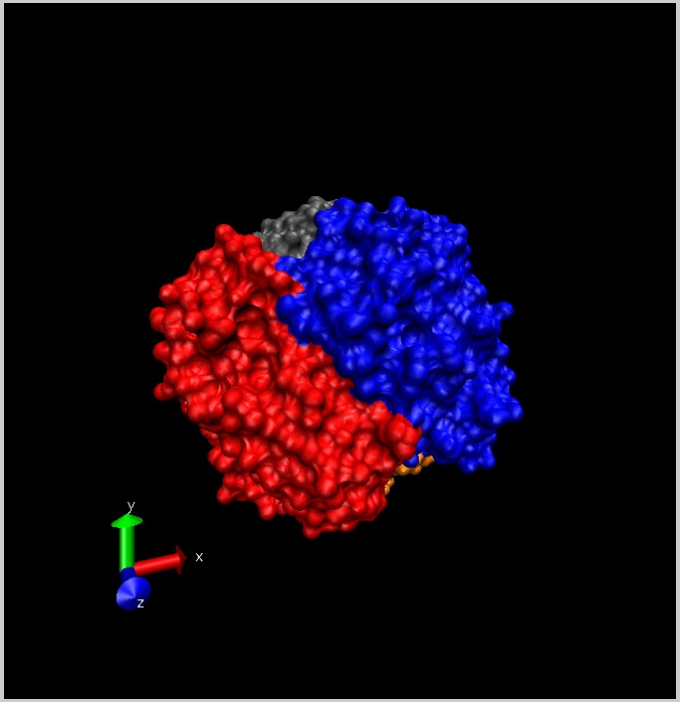
Molecular basis of any/all diseases can be traced to proteins

Alzheimer's Disease

- Alzheimer's disease (AD), first described by Alois Alzheimer in 1906
- Causes the gradual loss of memory and general cognitive decline
- AD is the most common form of dementia in the elderly
- Current estimation suggests that AD affects nearly 2% of the population and projects an increase of three fold within the next 50 years (http://www.alz.org)
- The amyloid hypothesis, proposed in 1992, suggested the accumulation of Aβ peptides is the primary cause of AD-related pathogenesis
- BACE1 was identified independently by several groups in 1999, and is speculated to be involved in production of Aβ peptides

Sickle Cell Anemia

- A genetic blood disorder
- Characterized by red blood cells that assume sickle shape
- The sickling occurs because of a mutation in the hemoglobin gene

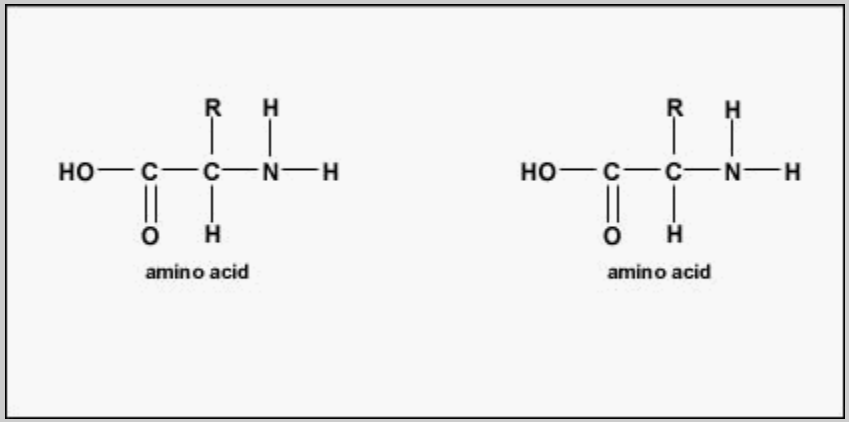


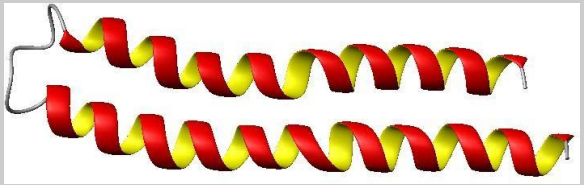
What are Proteins?

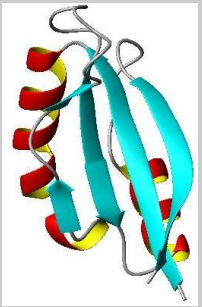
One of the major cellular macromolecules

Protein

- Proteins are functional elements of a cell
- Proteins are made of 20 “amino acid” subunits
- Proteins fold to create a their own characteristic fold (three dimensional shape)



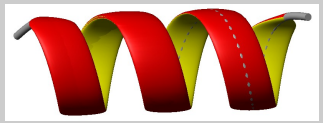


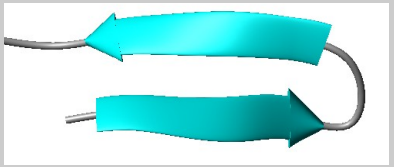
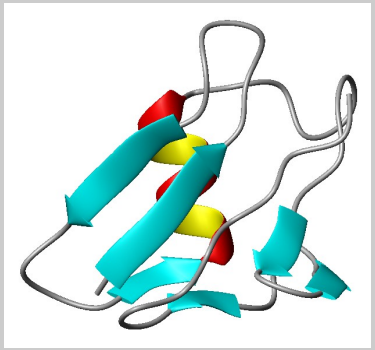


Protein Structure Hierarchy

- Primary sequence (1º):
  - linear order of connected amino acids.

    ALA-GLY-LYS-PRO-…
- Secondary Structure (2º):
  - Internal stable segments.

    α–Helix 

    β–Sheet 
- Tertiary Structure (3º): 

Why Protein Structure?

- Proteins provide metabolic and mechanical support for biological organisms.
- Structure gives rise to function.
- Structure is necessary (not sufficient) for function.
- Proteins are of special interest due to their therapeutic potential (why not DNA)?

How do we compute them?

Through the simulation of physical forces...

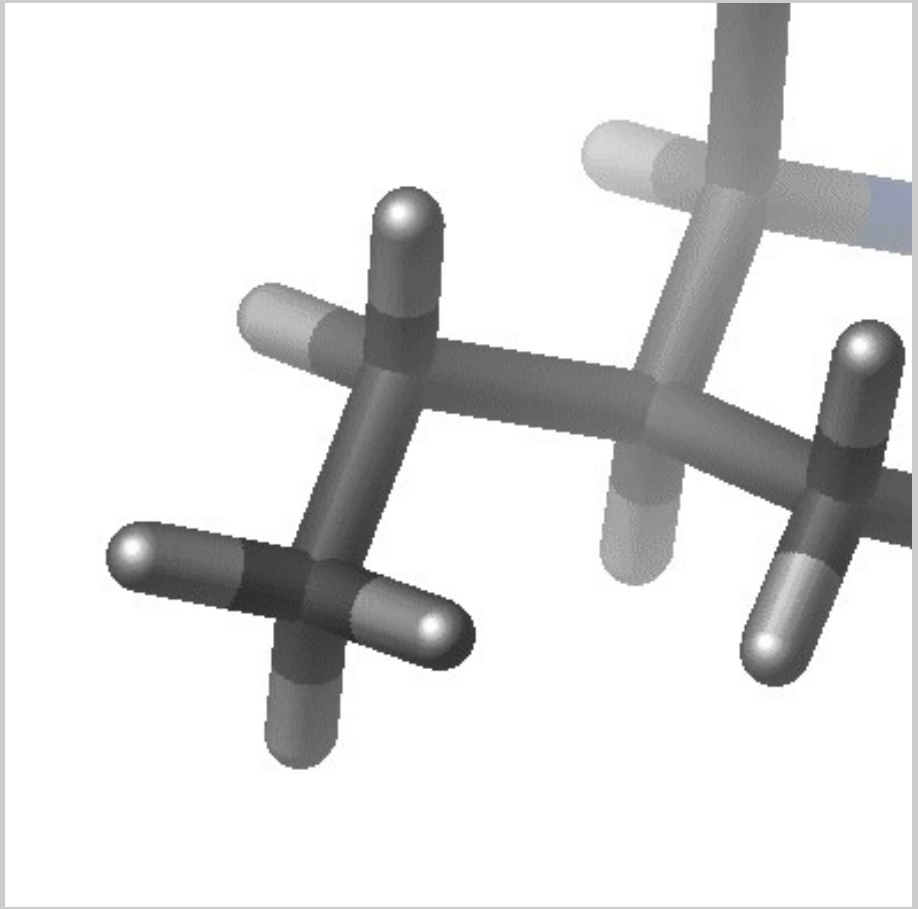
From Sequence to Structure

- Does primary sequence lead to functional structure?
  - Isolate functional protein
  - Denature using urea or high temperature
  - Confirm loss of function
  - Reinstate folding conditions (remove urea, lower temp.)
  - Confirm gain of function
- In general protein sequence leads to functional structure
- Simulation of physical forces should allow computational folding of proteins
  - Levitt, M. and A. Warshel, *Computer simulation of protein folding. Nature, 1975. 253: p. 694-698.*

Potential Energy of Bond Lengths

- The bond length a pair of atoms is known empirically
- Bond lengths should not exceed the expected values
- Defined by two atoms

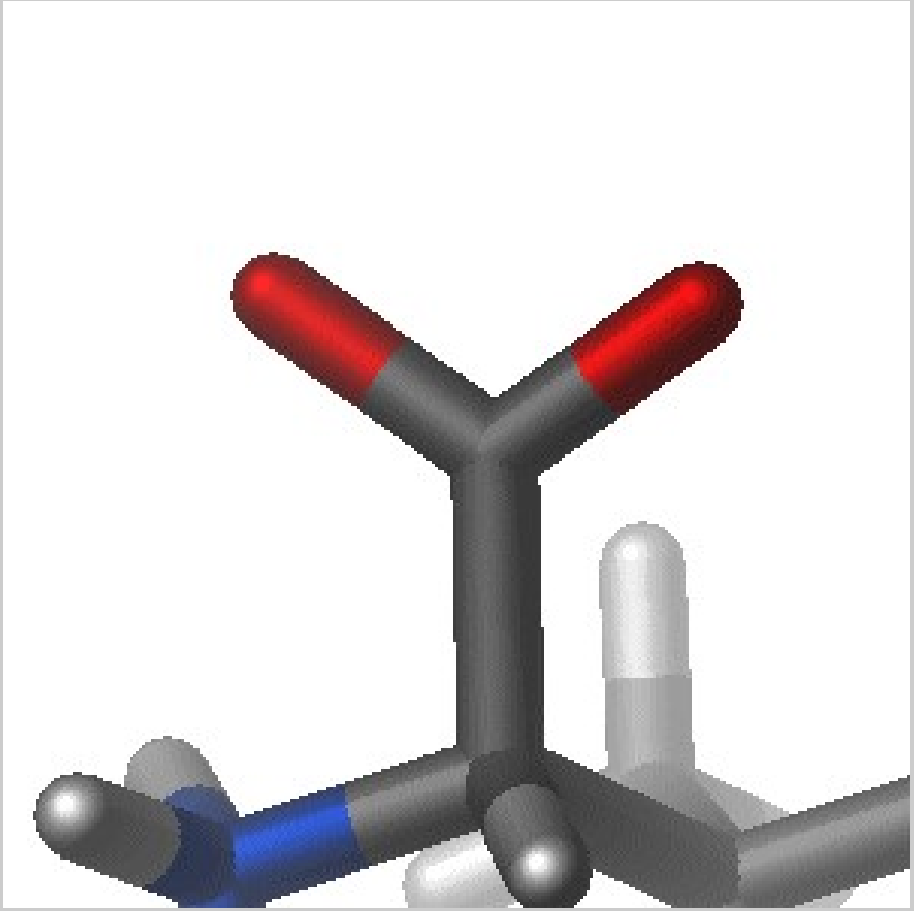
$$E_{Bond} = \sum_{bond} k_b (r - r_0)^2$$



Potential Energy of Bond Angles

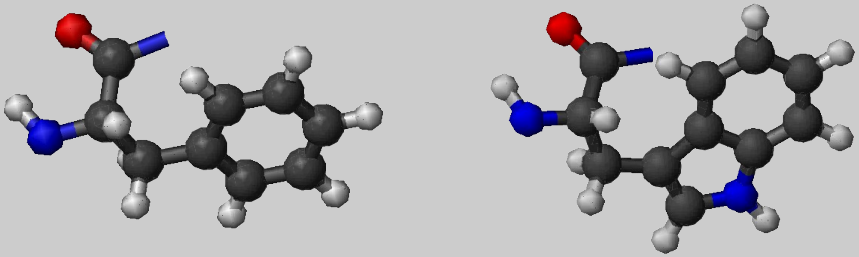
- Bond angles should not deviate from the known quantities
- Involves three atoms

$$E_{Angles} = \sum_{angles} k_\theta (\theta - \theta_0)^2$$



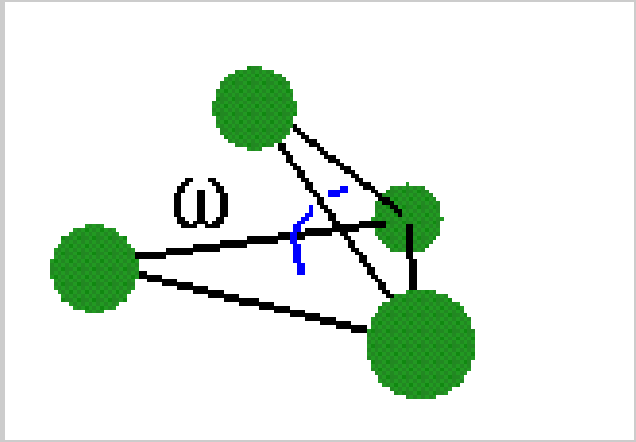
P.E. of Improper Dihedrals

- Improper dihedrals represent the planarity of a group of atoms
  - Peptide plane
  - Aromatic side chains: phenylalanine, tryptophan, tyrosine, histidine



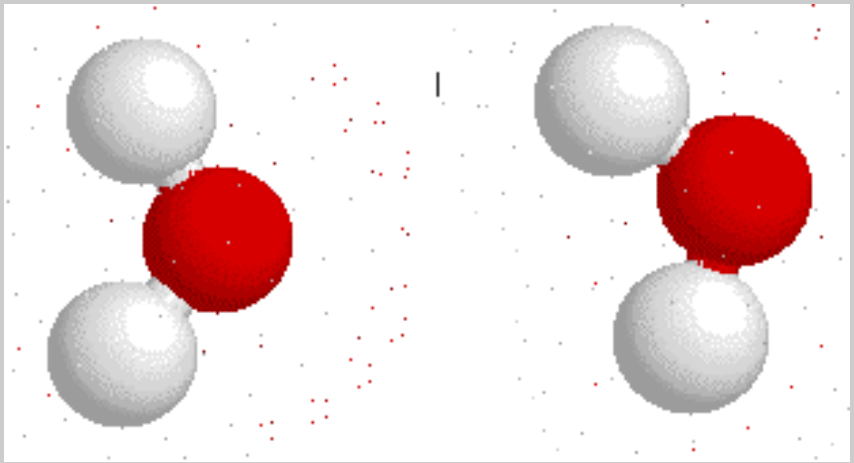
- Four atoms are required for this measure

$$E_{Impr} = \sum_{impropers} k_i(\omega - \omega_0)^2$$



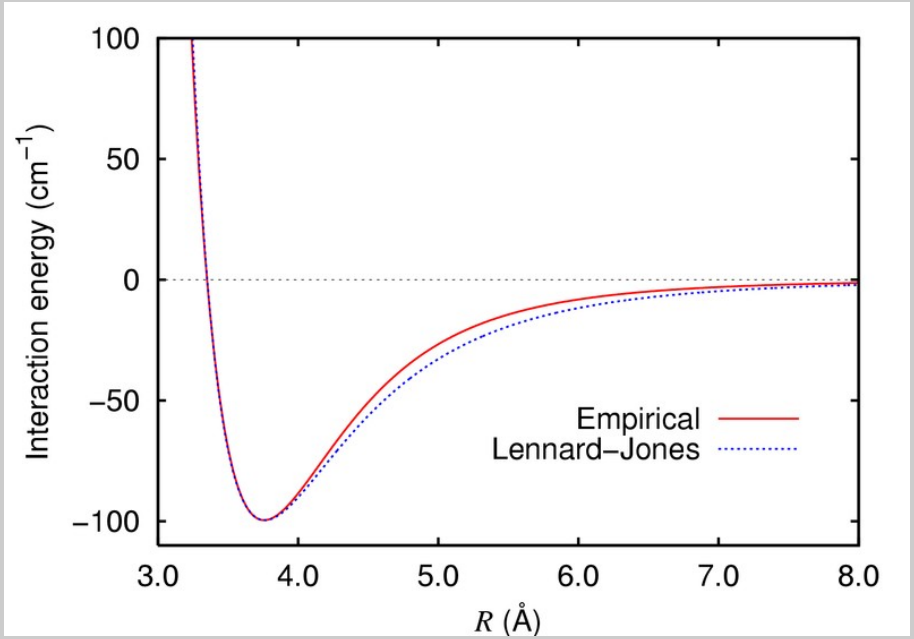
Hydrogen Bond

- Two participating atoms: donor and acceptor
- Normally between O-H or N-H
- Strongest non-bonded force



Van Der Waals Forces

- Van der Waals forces may be:
  - Attractive in long range.
  - Repulsive in short range.
- Modeled by L-J poetntial:
  - ε is the well depth
  - σ is the van der Waals radia
  - Experimentally determined!
- (6-12) L-J potential is defined as:



$$V(r)=4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12}-\left(\frac{\sigma}{r}\right)^{6}\right]$$

Empirical Energy Terms

- Conformational Energy Terms:
  - E_{BOND}: describes the covalent bond energy over all covalent bonds
  - E_{ANGL}: describes the bond angle energy over all bond angles
  - E_{DIHE}: describes the dihedral angle energy over all dihedrals
  - E_{IMPR}: describes the improper angle energies (planarity and chirality)
- Nonbonded Energy Terms:
  - E_{VDW}: describes the energy of Van Der Waals terms
  - E_{ELEC}: describes the energy of electrostatic interactions
  - E_{HBOND}: hydrogen bond interaction

Total Energy Term

- E_{Total} is the total potential energy of a conformation
- Force Field: A vector field representing the gradient of the total potential
- ω is referred to as the force constants

$$E_{Total} = \sum \left[w^p_{BOND} E_{BOND} + w^p_{ANGL} E_{ANGL} + w^p_{DIHE} E_{DIHE} + w^p_{IMPR} E_{IMPR} + w^p_{VDW} E_{VDW} + w^p_{ELEC} E_{ELEC} \right]$$

$$E_{BOND} = \sum_{bonds} k_b (r - r_0)^2$$

$$E_{ANGL} = \sum_{angles} k_\theta (\theta - \theta_0)^2$$

$$E_{DIHE} = \sum_{dihedrals} \sum_{i=1,m} \begin{cases} k_{\varphi_i}(1 + \cos(n\varphi_i + \delta_i)) & n_i > 0 \\ k_{\varphi_i}(\varphi_i - \delta_i)^2 & n_i = 0 \end{cases}$$

$$E_{IMPR} = \sum_{impropers} \sum_{i=1,m} \begin{cases} k_{\varphi_i}(1 + \cos(n\varphi_i + \delta_i)) & n_i > 0 \\ k_{\varphi_i}(\varphi_i - \delta_i)^2 & n_i = 0 \end{cases}$$

$$E_{VDW} = \sum_{vdw} \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}$$

$$E_{ELEC} = \sum_{i,j} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

Force Field

- Technically, the derivate of the potential energy
  - A vector field of forces
- Some currently existing force fields (forcefield):
  - Xplor-NIH
  - AMBER
  - CHARMm
  - MM2, MM3 and MM4
  - Sybyl
  - Etc.

Minimization of Total Energy

- Theoretically, the structure with the minimum total energy is the structure of interest.
- A number of minimization algorithms can be utilized.
  - Gradient descent
  - Monte Carlo and Simulated Annealing
  - Newton's
  - Genetic Algorithm
  - Distributed Global Optimization
  - Branch and Bound
  - …

$$E_{Total} = \sum \left[w^p_{BOND} E_{BOND} + w^p_{ANGL} E_{ANGL} + w^p_{DIHE} E_{DIHE} + w^p_{IMPR} E_{IMPR} + w^p_{VDW} E_{VDW} + w^p_{ELEC} E_{ELEC} \right]$$

How do we use computational approaches?

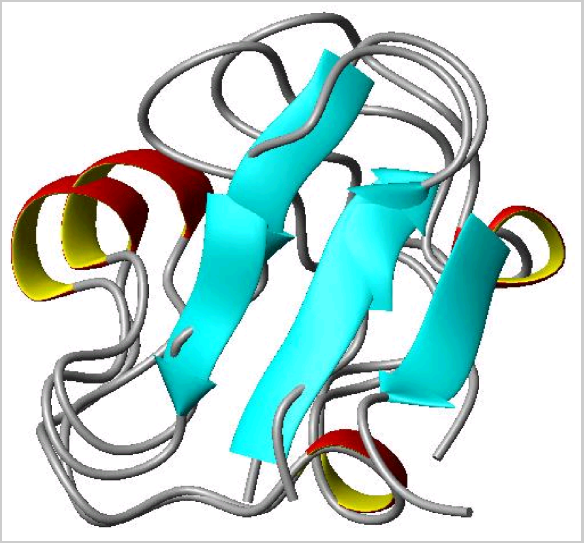
In short, we can compute you!

A significant step toward personalized medicine!

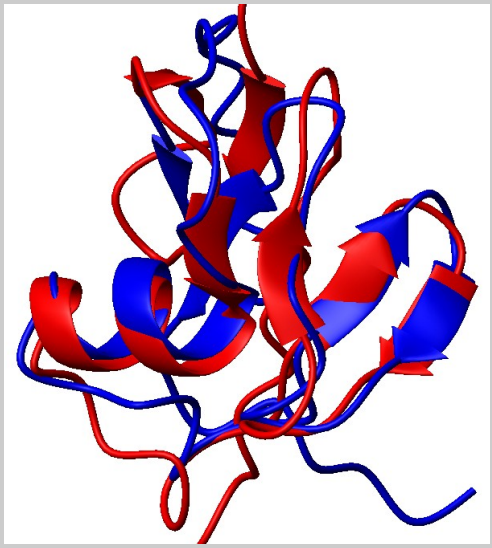
A significant step toward reengineered organisms!

Successes

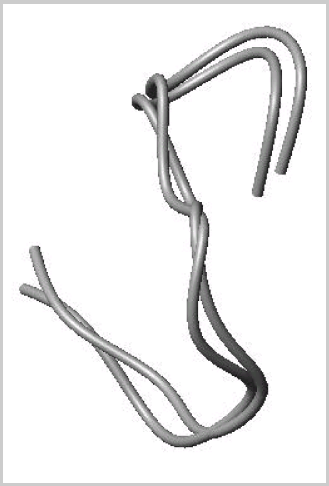
- Study of structure and dynamics of challenging proteins or complexes from minimal experimental data



- Size: 56 residues.
- Data acquisition: 1 week.
- Data analysis: 2 hours.
- Results: 1.8 Å with X-ray

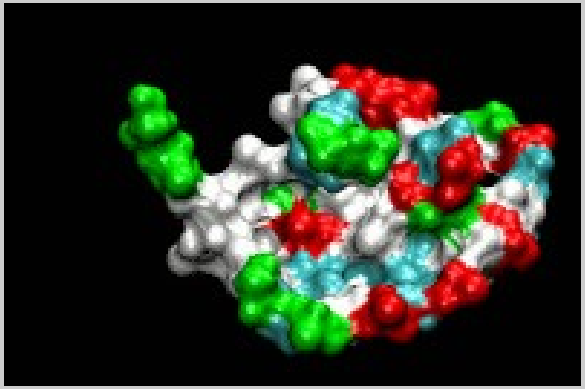


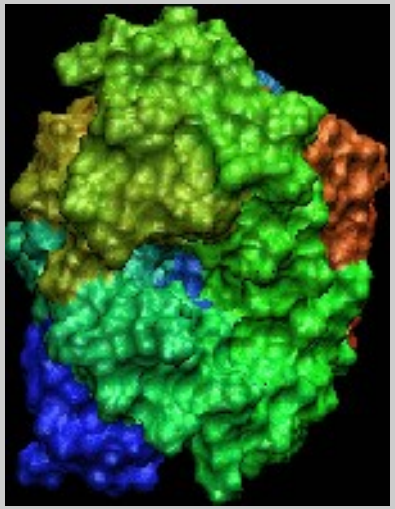
- Size: 70 residues.
- Data acquisition: 1 week.
- Data analysis: 1 Day.
- Results: C-terminal motion.

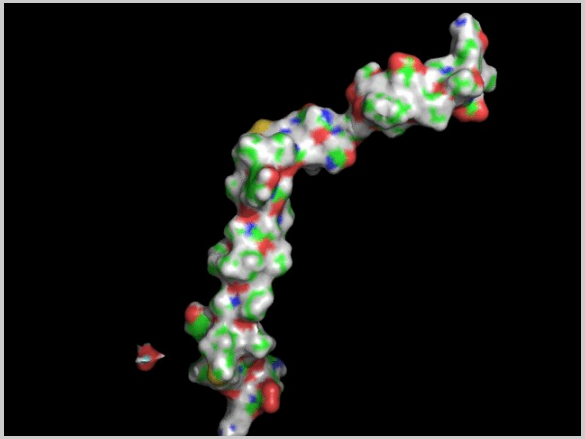


- Size: ~60 residues.
- Data acquisition: 1 week.
- Data analysis: 2 hours.
- Results: Unstructured Zn-binding region.
- ~1.5 Å with X-ray.

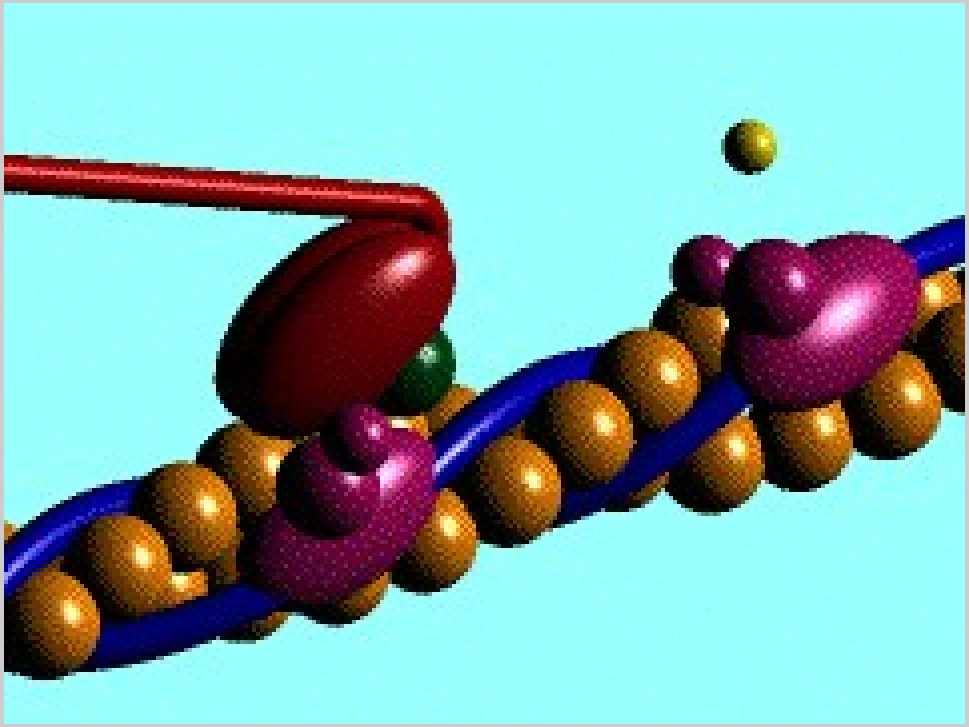
Simultaneous Structure and Dynamics







Why Motion?



Sickle Cell Anemia

- Genetic disorder cause by VAL substitution for ALA at β6
- Hydroxyurea (Hydrea) aleviates symptoms of SCA
- Hydrea has negative sideeffects
- Can we identify patients that will respond to Hydrea?
- Used
- Have demonstrated more than 90% success

Computational Medicine

- Disorder: Sickle Cell Anemia
- Treatment: Hydroxyurea
- Question: Who benefits?
- Approach: Computational models predicting patients' response to treatment.