Polypeptide Folding Using Monte Carlo Sampling, Concerted Rotation, and Continuum Solvation

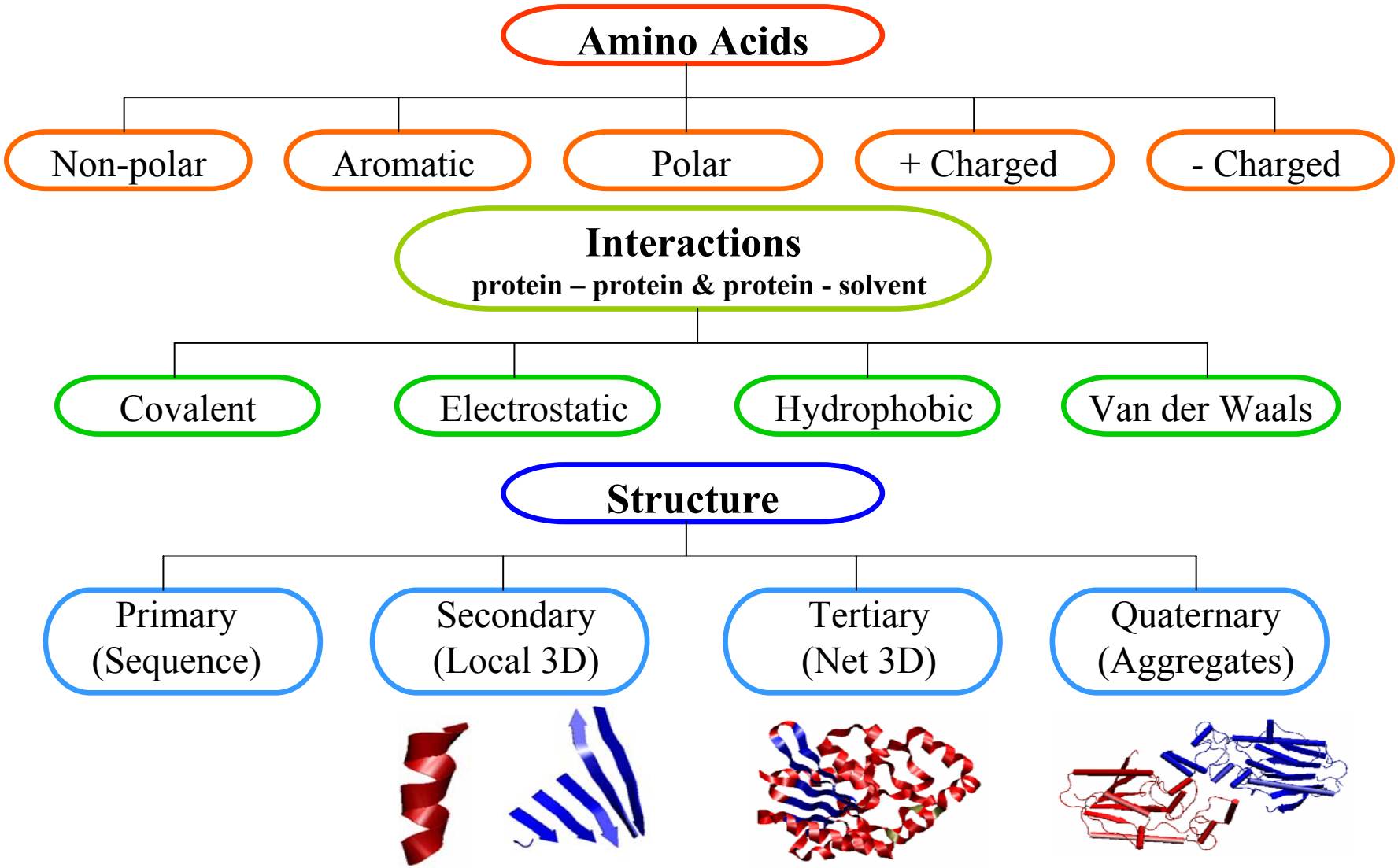
Jakob P. Ulmschneider and William L. Jorgensen
J.A.C.S. **2004**, *126*, 1849-1857

Presented by
Laura L. Thomas and Andrew Kohlway
BBSI Journal Club Meeting
June 3, 2004

Introduction
Protein Folding

- Protein structure is intimately related to function
- Protein misfolding is associated with diseases such as Alzheimer's and Cystic Fibrosis
- Computational studies can give additional insight into experimental studies
- Biological relevance of crystal structures
- Novel protein structures may serve roles in nanotech applications

Protein Structure



In This Study…

- The folding of three polypeptides which adopt β-hairpin and α-helical secondary structures were simulated using Monte Carlo sampling
- An implicit solvent model (GBSA) was implemented, while the proteins were represented with an all atom model (OPLS-AA)
- A concerted rotation algorithm was applied in order to increase sampling efficiency

Methods

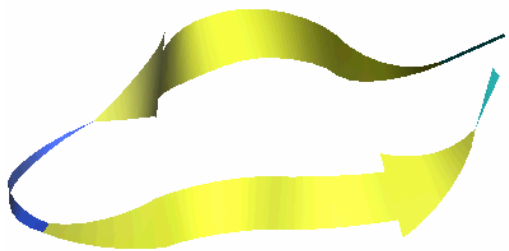
- Systems
- OPLS-AA Force Field
- GBSA Continuum Solvent Model
- Monte Carlo Sampling
- Concerted Rotation Algorithm
- Simulation Protocol

Peptide Systems Under Investigation

- U(1-17)T9D
  - MQIFVKTLDGKTITLEV pdb 1e0q
  - Derived from ubiquitin
  - β-hairpin

- α_1
  - ELLKKLLEELKG pdb 1al1
  - De novo
  - α-helical

- trpzip2
  - SWTWENGKWTWK pdb 1le1
  - De novo
  - β-hairpin



OPLS-AA Force Field

- **O**ptimized **P**otentials for **L**iquid **S**imulations (**A**ll **A**tom)

- Intramolecular terms
  - Bond Stretching
  - Angle Bending
  - Torsional Angle

- Non-bonded terms
  - Coulombic
  - Lennard-Jones
  - 1,4-interactions

$$E_{tot} = E_{bond} + E_{angle} + E_{dih} + E_{nb}$$

$$E_{bond} = \sum_i k_{b,i}\left(r_i - r_{0,i}\right)^2$$

$$E_{angle} = \sum_i k_{\theta,i}\left(\theta_i - \theta_{0,i}\right)^2$$

$$E_{dih} = \sum_i \left\{ \frac{V_{1,i}\left(1+\cos\varphi_i\right)}{2} + \frac{V_{2,i}\left(1-\cos 2\varphi_i\right)}{2} + \frac{V_{3,i}\left(1+\cos 3\varphi_i\right)}{2} \right\}$$

$$E_{nb} = \sum_i \sum_{j>i} \left\{ \frac{q_i q_j e^2}{r_{ij}} + 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right] \right\}$$

GBSA Continuum Solvation

- Implicit model where solvent is treated as a statistical continuum
- Solvation free energy consists of a solvent-solvent cavity term, a solute-solvent van der Waals term, and a solute-solvent electrostatic polarization term

$$G_{sol} = G_{cav} + G_{vdW} + G_{pol}$$

- G_{sol} is linearly related to SASA for saturated hydrocarbons in water

$$G_{cav} + G_{vdW} = \sum \sigma_k SA_k$$

- G_{pol} (The Generalized Born Equation) is derived from Coulomb's Law in a dielectric and the Born Equation

$$G_{pol} = -166\left(1 - \frac{1}{\varepsilon}\right)\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{q_i q_j}{f_{GB}}$$

- f_{GB} is a function of the internuclear separation and atomic radii

Monte Carlo Sampling

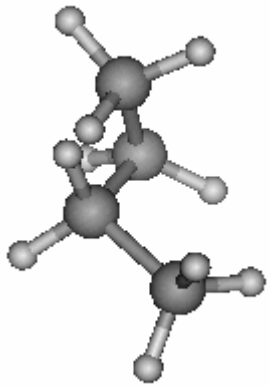
- Classical statistical mechanics method based on the Metropolis Algorithm
- Random configurations of a molecule are generated

$$x_{new} = x_{old} + \left(2r_n - 1\right)\delta r_{\max}$$

- The energy of the system is computed
  - If $E_{new} < E_{old}$
    - Accept new configuration
  - If $E_{new} > E_{old}$
    - Compute the Boltzmann Factor $e^{\left(-\Delta V(r^N)/k_bT\right)}$
    - If BF > r_n
      - Accept new configuration
    - If BF < r_n
      - Reject new configuration

Concerted Rotation with Bond Angles

- Benefits of MC over MD
  - Energy derivatives are not needed
  - Exploits implicit solvent by enabling large conformational changes to cross over energy barriers
- However…
  - MC requires nontrivial move sets
  - Variation in backbone degrees of freedom leads to global conformational changes
- CRA
  - Change consecutive backbone torsion angles while leaving others unchanged
  - Smaller number of non-bonded terms to evaluate
  - Includes flexible angles, an improvement over earlier models
  - Includes a Gaussian bias in order to increase sampling efficiency



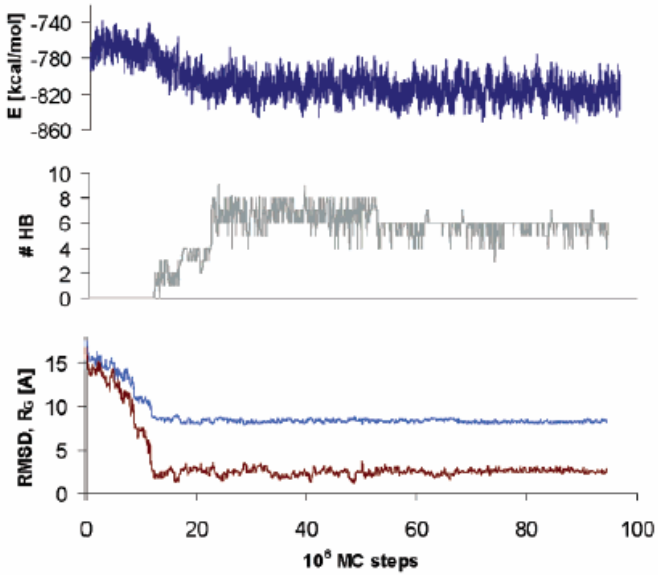
Simulation Set-up Protocol

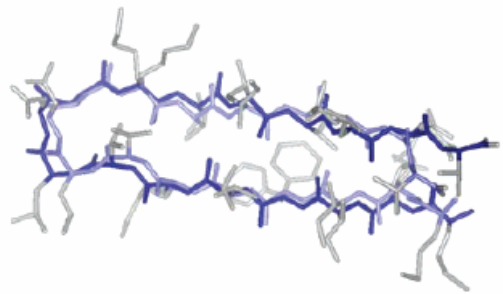
- Modified MCPRO package
- Temperatures 30 – 50 ºC
- No cutoffs for the non-bonded interactions
- Electrostatic energy was recomputed for every MC configuration
- The non-polar contribution to the solvation free energy was considered to be proportional to the SASA, which was updated every 30 MC configurations
- Backbone moves were made every 4 MC configurations, with the remainder being simple side chain moves
- The simulated structures were compared to NMR structures

Results

β-Hairpin U(1-17)T9D

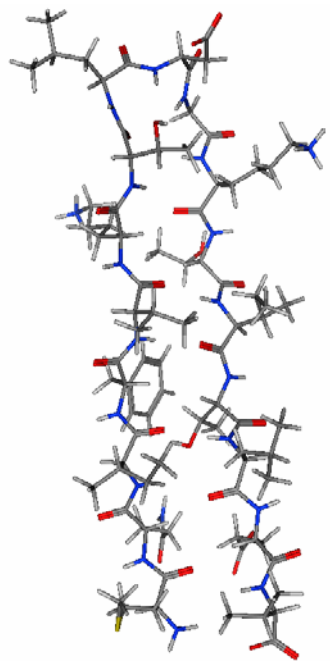
- Eight MC runs at 30 ºC
- Started from completely extended conformations
- All runs stable β-hairpin conformations within 20 – 80 million MC configurations
- Rapid relaxation into compact state with twist which lasts from a few to 70 million configurations
- Turn formed (2-10M)
- Side chain reorientation and formation of interstrand H-bonds





β-Hairpin U(1-17)T9D

- Lowest energy system most closely resembles the NMR structure
- The other structures were sampling local minima
- Formation of turn critical
- All simulations yielded a turn in the center of the chain
- No helical structures formed
- Transient salt bridge between amino group of K11 and carboxyl group of D9
- 7 non-native structures all very similar with a L-D-G-K turn
  - caused all sidechain interactions to be out of sync (G should be in 4th position)
- Hydrophobic core vs. H-bond centric modes of folding



α-Helical Polypeptide α_1

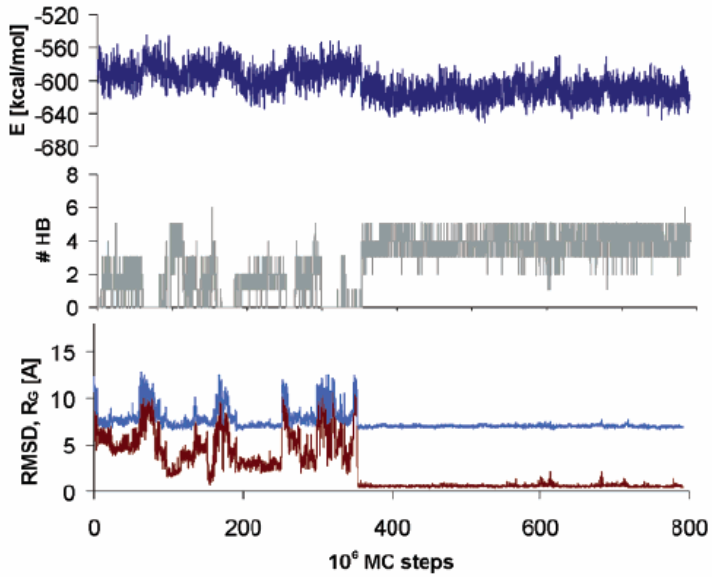
- Four MC runs at 40 ºC (350 – 800 M MC configurations)
- Started from completely extended conformations
- Forms a random coil at low concentrations
- Analysis of helical content
  - Helical = Φ and Ψ within a range of (30º, 25º) of ideal (-57º, -47º)
  - H-bonds = D-H-A distance of less than 2.8 Å and a D-H-A angle between 120º-180º
- All simulations showed that transient helical structures formed numerous times and were stable for 10 – 30 MC configurations

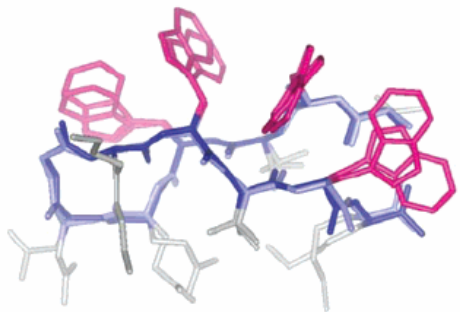
α-Helical Polypeptide α_1

- Cluster analysis used to assess secondary structural motifs encountered in simulations
  - Structures were sampled every millionth MC step
  - Main chain least squares superposition with a 1.5 Å similarity cutoff
- 606 clusters identified – 200 analyzed
- 43% of the structures were helical
  - A majority of these have only one helical turn
  - Two turn helices make up 15%
  - Complete helices make up 1%
- Indicates that fully helical $\alpha_1$is not stable on its own at this temperature
- 21% of structures were β-like
- 36% were random coils
- This is all in agreement with the NMR studies

Tryptophan Zipper trpzip2

- Eight MC runs at 50 ºC (up to 8M configurations)
- Started from completely extended conformations
- Two of the simulations reached the native state
- Immediate relaxation
- Frequent transitions between folded states (compact coil or β-hairpin)
- The system then unfolds again and repeats this process



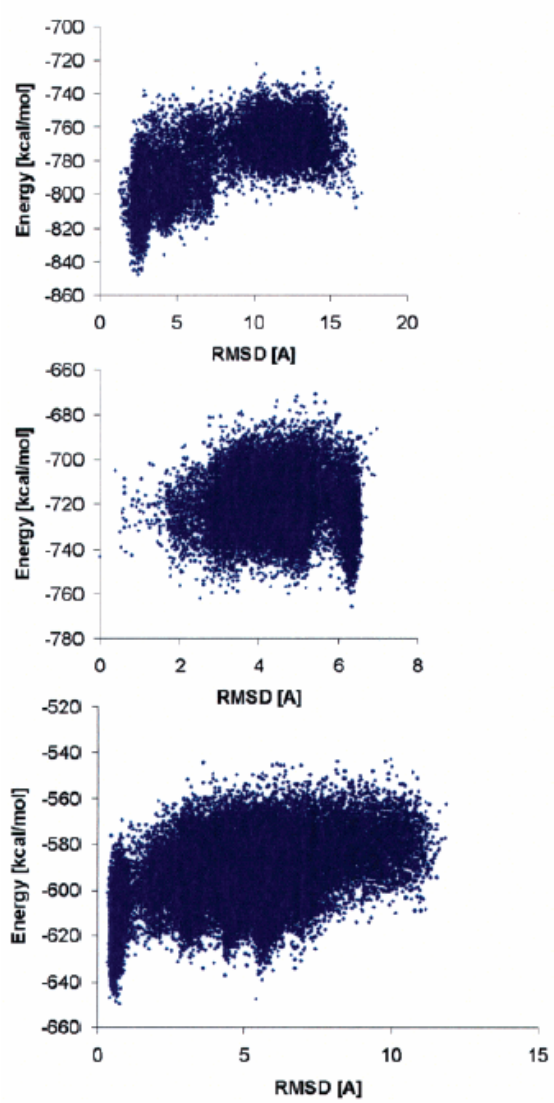


Tryptophan Zipper trpzip2

- Comparison to the larger β-hairpin U(1-17)T9D where all runs led to stable folded structures
- Trpzip2 runs were done at higher temperature and were 5-10 times longer
  - This criteria was sufficient to unfold local structures
  - Only the native state had a lifetime longer than the simulation length
  - The 6 non-native trpzip2 runs were not trapped in local minima, but rather did not yet sample their native basin
  - Random search for native state without necessary intermediates
  - All H-bonds formed simultaneously-no zipping (similar for both)

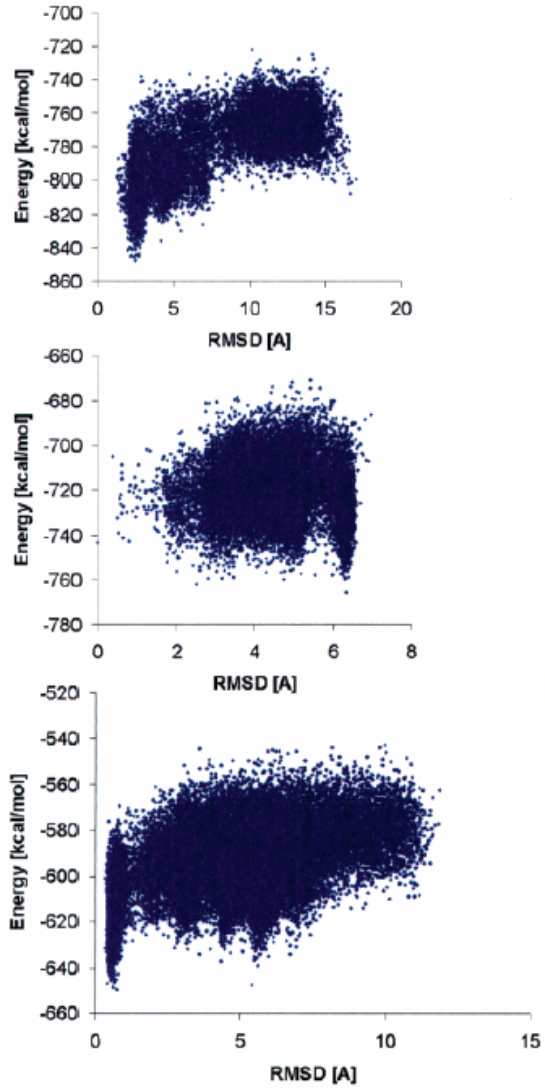
Native States

- How to distinguish between native and non-native states?
- Intervals of .1 M configurations from all runs used to create plot
- Native structures group in the bottom left corner (low energy, low RMSD)



Native States

- U(1-17)T9D Peptide – native structures seen below -820 kcal/mol
- α_1 Peptide – native structures much harder to discern:
  1) β-Turns above 6 Å
  2) Partial Helical Structures in Middle
  3) Full Helical Structures below 2 Angstroms (High in Energy: Stability obtained from interaction with other monomer units)
- Tryptophan Zipper – more like U(1-17)T9D with native structures seen below -630 kcal/ml



Conclusions

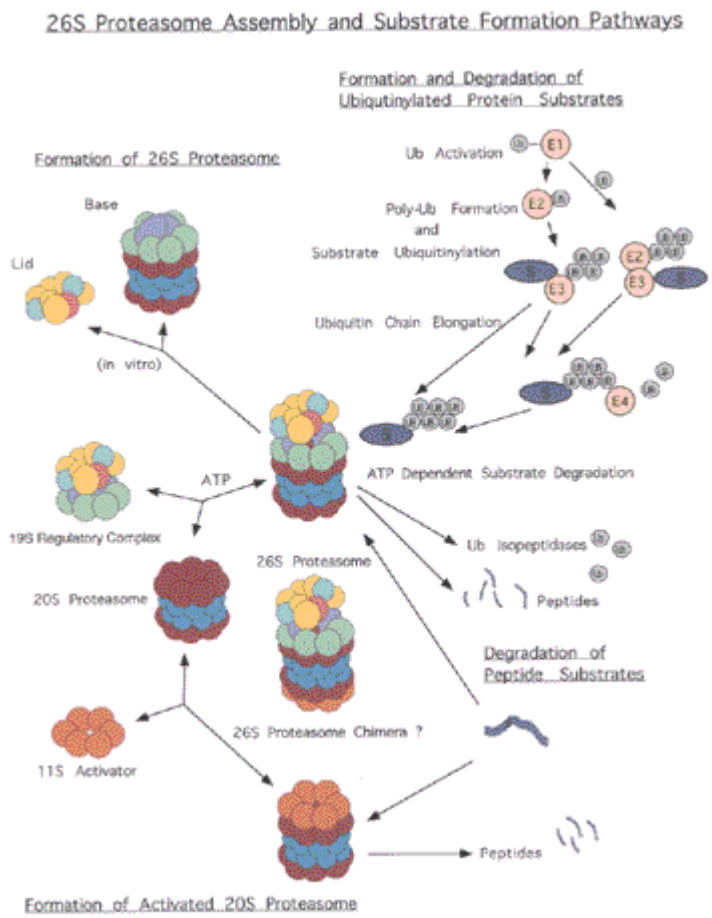
- Successful folding of peptides obtained using Monte Carlo sampling with concerted backbone rotations, OPLS-AA force field, GBSA solvation model
  - The β-hairpins both assumed native structures
  - The α-helical structure assumed random coil conformations in agreement with low concentration NMR studies
- Higher temperatures were not necessary in contrast to similar MD studies (0-50 °C)
- CRA method allowed for relatively larger conformational changes while avoiding intramolecular steric clashes

References

- Ulmschneider, J. P.; Jorgensen. *J. Am. Chem. Soc.* **2004**, *126*, 1849-1857.
- Still, W.C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127-6129.
- Ulmschneider, J. P.; Jorgensen. W.L. *J. Chem. Phys.* **2003**, *118*, 4261-4271.
- Paul von Rague Schleyer. Encyclopedia of Computational Chemistry; John Wiley & Sons Ltd, 1998.

Ubiquitin

- Ubiquitin is an evolutionarily highly conserved 76 amino acid polypeptide that is abundant in all eukaryotic cells.
- Polyubiquitination of substrates targets them for degradation by the 26S proteasome, a multiprotein complex conserved from archaebacteria to humans.



*Taken from http://www.proteasome.com/publications/ferrell/diagram1.htm