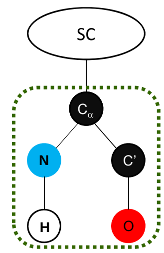
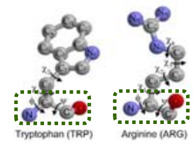
Study of Protein Motion

1

Protein

Long sequence of amino-acids (dozens to thousands)

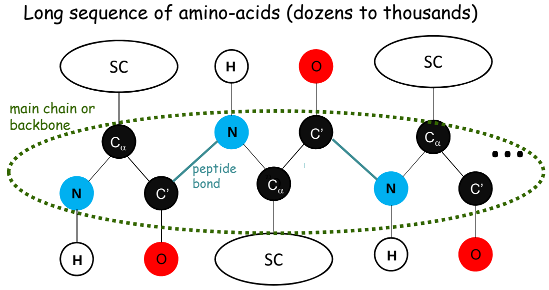




2

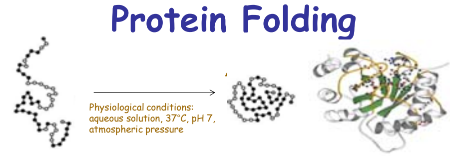
Protein

Long sequence of amino-acids (dozens to thousands)

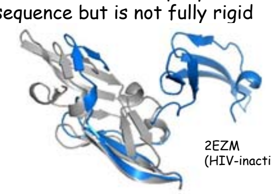


3

Protein Folding



The folded structure is uniquely determined by the protein sequence but is not fully rigid



2EZM
(HIV-inactivating protein)

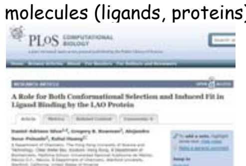
4

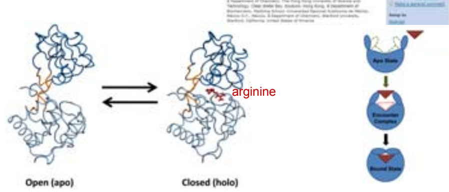
Flexibility is necessary …

… for a protein to achieve its biochemical functions by binding against other molecules (ligands, proteins)

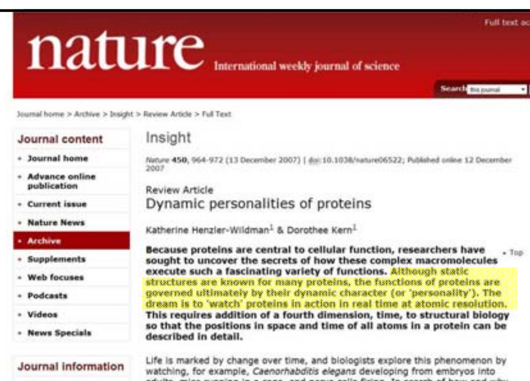
Binding models:

- Induced fit model
- Conformational selection model





5



nature International weekly journal of science

Full text acc

Search this journal

Journal home > Archive > Insight > Review Article > Full Text

Journal content
- Journal home
- Advance online publication
- Current issue
- Nature News
- Archive
- Supplements
- Web focuses
- Podcasts
- Videos
- News Specials

Journal information

Insight

Nature **450**, 964-972 (13 December 2007) | doi:10.1038/nature06522; Published online 12 December 2007

Review Article

Dynamic personalities of proteins

Katherine Henzler-Wildman[1] & Dorothee Kern[1]

Because proteins are central to cellular function, researchers have sought to uncover the secrets of how these complex macromolecules execute such a fascinating variety of functions. Although static structures are known for many proteins, the functions of proteins are governed ultimately by their dynamic character (or 'personality'). The dream is to 'watch' proteins in action in real time at atomic resolution. This requires addition of a fourth dimension, time, to structural biology so that the positions in space and time of all atoms in a protein can be described in detail.

• Top

Life is marked by change over time, and biologists explore this phenomenon by watching, for example, *Caenorhabditis elegans* developing from embryos into adults, mice running in a cage, and nerve cells firing. In search of how and why.

6

Experimental Methods

Protein Data Bank (PDB):
Repository of folded structures of **74,732** proteins (August 2011)

- **X-ray crystallography** (65,195 entries)
  → high resolution, but one or few conformations
- **NMR spectroscopy** (9,014 entries)
  → multiple conformations, but for small proteins
- **Cryo-electron microscopy** (373 entries)
  → multiple conformations, but low resolution

7

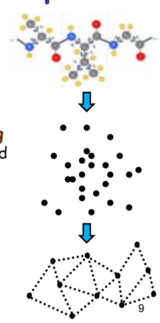
Energy-Based Computer Simulation

- Molecular Dynamics simulation
- Monte Carlo simulation
- Others:
  - coarse-grained force fields,
  - multi-scale modeling,
  - replica exchange,
  - normal mode analysis,
  - elastic network models, …
- **Advantages:** produce time-dependent information at atomic resolution
- **Drawbacks:** huge running times, enormous amount of data, delicate setup

8

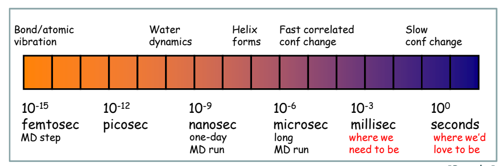
Application of Robotics and Motion Planning Techniques

1. **Kinematic/Geometric models of proteins**
   Develop algorithm-friendly models that directly encode dominant energy terms.
2. **Kino-Geometric Conformational Sampling**
   Use these models to sample conformations and represent a protein's folded state by a cloud of points.
3. **Graph-Based Models of Protein Motion**
   Transform a cloud representation into a probabilistic roadmap representing protein kinetics.

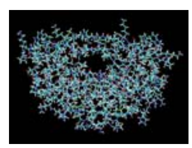


9

Timescales of Protein Motion



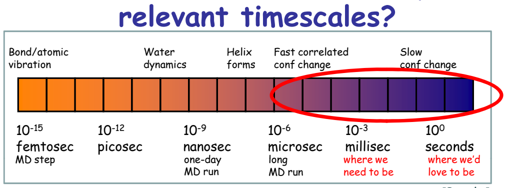
[Pande]



HIV-1 protease
[courtesy L.E. Kavraki]

10

How can one access directly to relevant timescales?



[Pande]

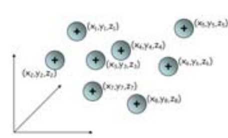
- Simulating a protein over a nanosecond timescale is like simulating human locomotion over a tiny fraction of a footstep, or like trying to understand how to reach the Moon by jumping 1.5 feet in the air.

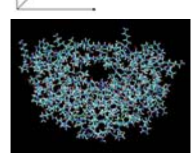
11

Kinematic Model #1: Collection of independent atoms

[Quick reminder: Kinematics studies the motion of objects without consideration of the forces that cause the motion.]

- Atoms can move independently, i.e., all constraints between atoms are eventually represented in a **force field** function
- A conformation is defined by $3 \times n$ parameters (the coordinates of the atom centers)
- All motion frequencies can be simulated

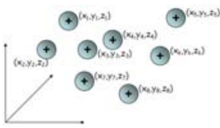




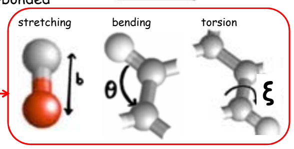
12

Kinematic Model #1: Collection of independent atoms

- Atoms can move independently, i.e., all constraints between atoms are eventually represented in a **force field** function

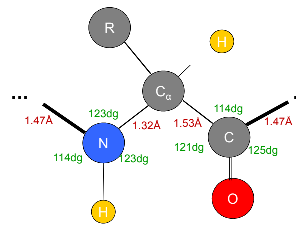


$F = F_{bonded} + F_{non\text{-}bonded}$



13

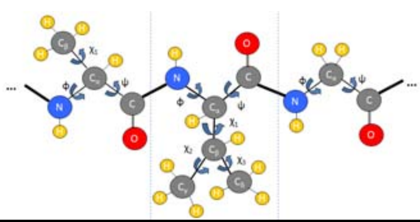
Over picosecond timescales bond lengths and angles average to constants



14

Kinematic Model #2: Linkage of connected atoms

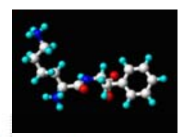
- Bonded atoms are connected by links of fixed lengths
- The only **degrees of freedom** are the dihedral angles around the simple bonds

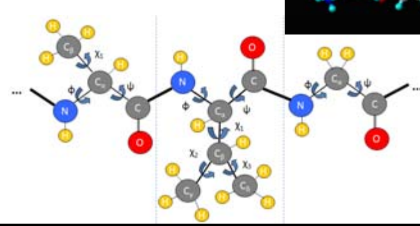


15

Kinematic Model #2: Linkage of connected atoms

- Bonded atoms are connected by links of fixed lengths
- The only **degrees of freedom** are the dihedral angles around the simple bonds

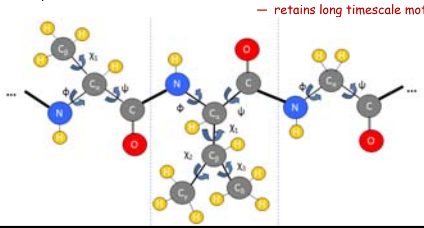




16

Kinematic Model #2: Linkage of connected atoms

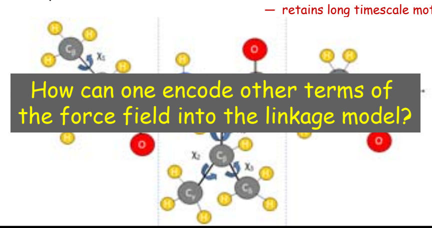
- Bonded atoms are connected by links of fixed lengths
- The only **degrees of freedom** are the dihedral angles around the simple bonds
- Free vibrations of the atoms can no longer be generated
- The linkage model
  - encodes terms of the force field,
  - filters out free atomic vibrations,
  - retains long timescale motions



17

Kinematic Model #2: Linkage of connected atoms

- Bonded atoms are connected by links of fixed lengths
- The only **degrees of freedom** are the dihedral angles around the simple bonds
- Free vibrations of the atoms can no longer be generated
- The linkage model
  - encodes terms of the force field,
  - filters out free atomic vibrations,
  - retains long timescale motions

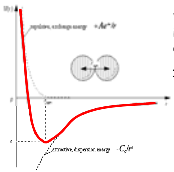


How can one encode other terms of the force field into the linkage model?

18

Van der Waals Forces

- $F_{non\text{-}bonded} = F_{van\ der\ Waals} + F_{Coulomb}$

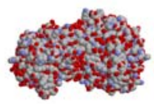


Van der Waals forces between two atoms result from induced polarization effect(formation of electric dipoles). They are weak, except at close range.

12-6 Lennard-Jones potential:

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

- In a folded conformation, atoms are densely packed against one another. Small perturbation can result into large repulsive vdW terms.

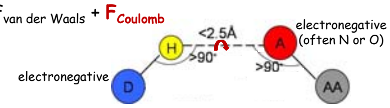


→ Atoms are modeled as **hard spheres** with radii ≈ α×vdWradii, where α = 0.7 to 0.8
+ **no two hard spheres are allowed to overlap** (volume exclusion constraint)

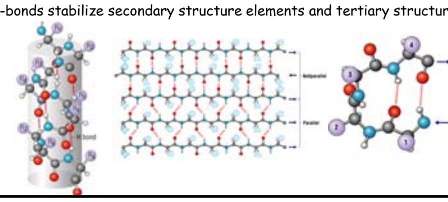
19

Hydrogen Bonds

- $F_{non\text{-}bonded} = F_{van\ der\ Waals} + F_{Coulomb}$



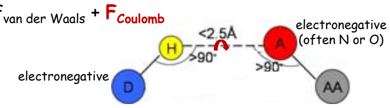
H-bonds stabilize secondary structure elements and tertiary structure



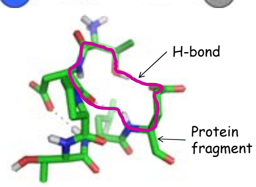
20

Hydrogen Bonds

- $F_{non\text{-}bonded} = F_{van\ der\ Waals} + F_{Coulomb}$



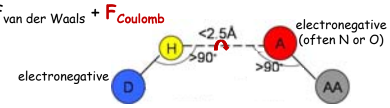
H-bonds rigidify portions of the protein and create closed cycles in linkage model



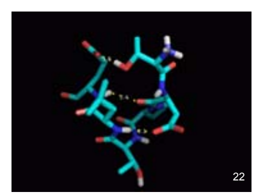
21

Hydrogen Bonds

- $F_{non\text{-}bonded} = F_{van\ der\ Waals} + F_{Coulomb}$



H-bonds rigidify portions of the protein and create closed cycles in linkage model



22

Advantages/Drawbacks of Linkage Model

- Fewer DOFs, hence smaller dimensionality of the conformational space
- Many force terms are directly encoded in representation, hence the model can't create motion that would violate these terms
- Most high-frequency motions are de facto filtered out

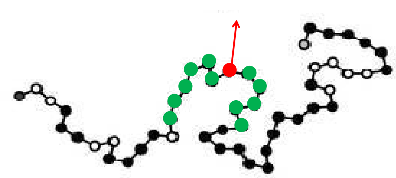
But:

- Generating kinematically valid conformations can be more difficult

23

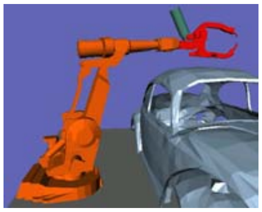
Inverse Kinematics Problem

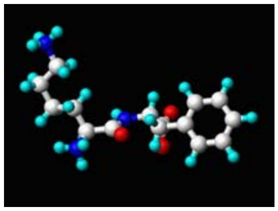
How does a change in the position of an atom affect the rest of the protein?



24

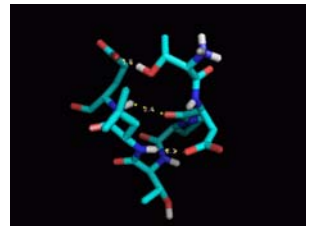
Analogy with Robotics





25

Analogy with Robotics



26

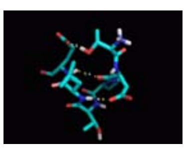
Inverse Kinematics Methods

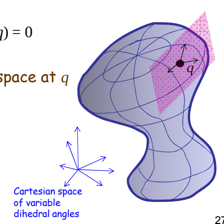
Null-space motion

How to deform a subset of a protein with more than 6 ϕ and ψ angles without breaking cycles?

SVD method:

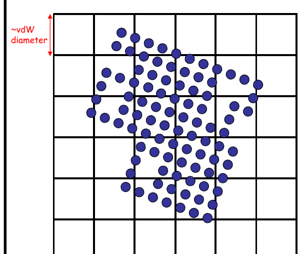
- Cycle closure constraints → $F(q) = 0$
- Differentiation → $J_F \times dq = 0$
- SVD of J_F → Basis of tangent space at q

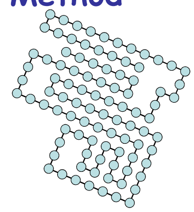




27

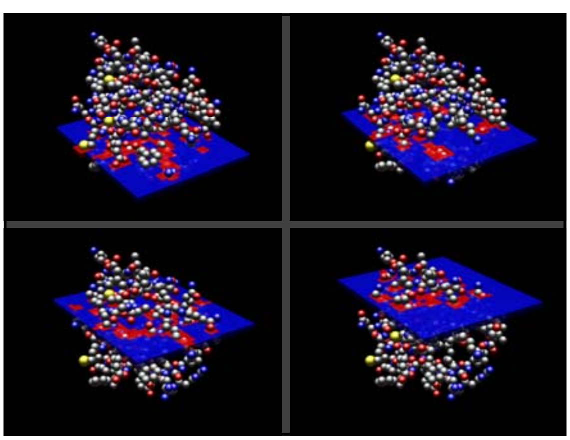
Checking Volume-Exclusion Constraint: Grid Method





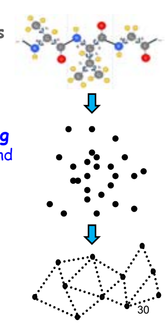
- Subdivide 3-space into cubic cells
- Compute cell that contains each atom center
- Represent grid as hashtable

28



Beyond Simulation

1. **Kinematic/Geometric models of proteins**
   Develop algorithm-friendly models that directly encode dominant energy terms.
2. **Kino-Geometric Conformational Sampling**
   Use these models to sample conformations and represent a protein's folded state by a cloud of points.
3. **Graph-Based Models of Protein Motion**
   Transform a cloud representation into a probabilstic representing protein kinetics.

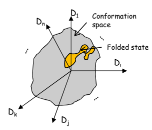


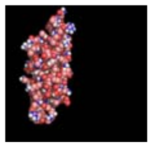
30

Kino-Geometric Conformational Sampling

Computational challenges:

- Requires satisfying often **antagonistic** constraints: kinematic and volume exclusion constraints
- Folded conformations form a relatively **tiny region** of the conformational space. How to hit this region?

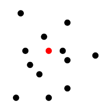




31

Kino-Geometric Conformational Sampling

- ROCK (Rigidity Optimized Conformational Kinetics) [Zavodsky et al., 2004]
- FRODA (Framework Rigidity Optimized Dynamic Algorithm) [Wells et al., 2005, Farrell et al., 2010]
- KGS (Kino-Geometric Sampling) [Yao et al. 2011]
- PEM (Protein Ensemble Method) [Shehu et al., 2006]

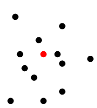


1. Initialize conformation distribution Δ to $\{q_{given}\}$
2. Iterate
   a. Pick q from Δ
   b. Deform q into new conformation q_{new}

32

Kino-Geometric Conformational Sampling

- ROCK (Rigidity Optimized Conformational Kinetics) [Zavodsky et al., 2004]
- FRODA (Framework Rigidity Optimized Dynamic Algorithm) [Wells et al., 2005, Farrell et al., 2010]
- KGS (Kino-Geometric Sampling) [Yao et al. 2011]
- PEM (Protein Ensemble Method) [Shehu et al., 2006]



1. Initialize conformation distribution Δ to $\{q_{given}\}$
2. Iterate
   a. Pick q from Δ
   b. Deform q into new conformation q_{new}

33

Kino-Geometric Conformational Sampling

- ROCK (Rigidity Optimized Conformational Kinetics) [Zavodsky et al., 2004]
- FRODA (Framework Rigidity Optimized Dynamic Algorithm) [Wells et al., 2005, Farrell et al., 2010]
- KGS (Kino-Geometric Sampling) [Yao et al. 2011]

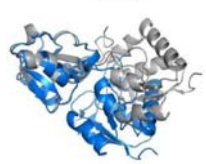
1. Select conformation q in Δ
   ROCK and FRODA: q is most recent conformation on Δ
   KGS: q is picked at random with probability inverse to sampling density
2. Select stable H-bonds in q
   ROCK and FRODA: select H-bonds with energy less than a threshold
   KGS: uses a regression tree trained from Molecular Dynamics data
3. Perform rigidity analysis in q
   ROCK, FRODA, and KGS: transform kinematic constraints into distance constraints between atoms, run Pebble Game algorithm to identify all rigid groups of atoms
4. Deform q into q_{new}
   ROCK: Perturb dihedral angles and close cycles by minimizing to zero a measure of closure violation
   FRODA: Perturb all atom positions and reform rigid groups of atoms
   KGS: Perturb dihedral angles in null space
5. Check q_{new} for volume exclusion

34

Statistics for Two Proteins

- 2EZM
  \# atoms: 992
  \# rigid groups: 503
  \# cycles: 47

- 2LAO 84
  \# atoms: 3649
  \# rigid groups: 1023
  \# cycles: 84



35

Kino-Geometric Conformational Sampling

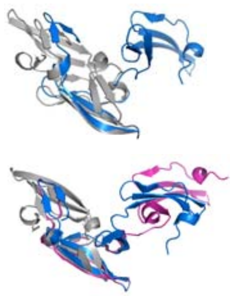
- ROCK (Rigidity Optimized Conformational Kinetics) [Zavodsky et al., 2004]
- FRODA (Framework Rigidity Optimized Dynamic Algorithm) [Wells et al., 2005, Farrell et al., 2010]
- KGS (Kino-Geometric Sampling) [Yao et al. 2011]

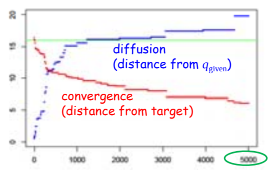
1. Select conformation q in Δ
   ROCK and FRODA: q is most recent conformation on Δ
   KGS: q is picked at random with probability inverse to sampling density
2. Select stable H-bonds in q
   ROCK and FRODA: select H-bonds with energy less than a threshold
   KGS: uses a regression tree trained from Molecular Dynamics data
3. Perform rigidity analysis in q
   ROCK, FRODA, and KGS: transform kinematic constraints into distance constraints between atoms, run Pebble Game algorithm to identify all rigid groups of atoms
4. Deform q into q_{new}
   ROCK: Perturb dihedral angles and close cycles by minimizing to zero a measure of closure violation
   FRODA: Perturb all atom positions and reform rigid groups of atoms
   KGS: Perturb dihedral angles in null space
5. Check q_{new} for volume exclusion

36

KGS Sampling: 2EZM

- 992 atoms
- Given conformation in grey
- Target conformation in blue at RMSD 16Å from given conformation (hinge and twist)



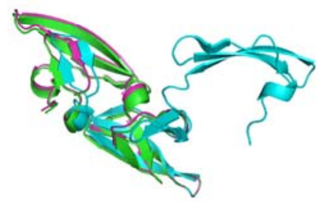


Running time on a dual quad-core 3GHz computer: 93 minutes, 1.12 seconds per sample

37

KGS Sampling: 2EZM

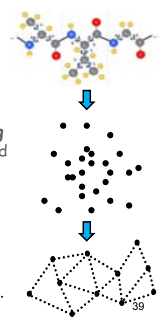
- "Path" from given conformation to sampled conformation closest to target:



38

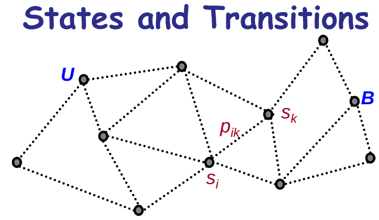
Beyond Simulation

1. **Kinematic/Geometric models of proteins**
   Develop algorithm-friendly models that directly encode dominant energy terms.
2. **Kino-Geometric Conformational Sampling**
   Use these models to sample conformations and represent a protein's folded state by a cloud of points.
3. **Graph-Based Models of Protein Motion**
   Transform a cloud representation into a motion network representing protein kinetics.



39

Graph-Based Model: States and Transitions

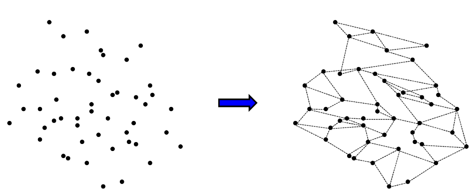


- State: conformation or conformation set
- Transition: probability of going from one state to another (Markov model)
- Compact representation of a huge collection of possible motion paths
- $P_B(s)$ = probability that protein reaches B from s before reaching U

$$P_B(s_i) = \Sigma_k p_{ik} \times P_B(s_k)$$

40

States are conformations

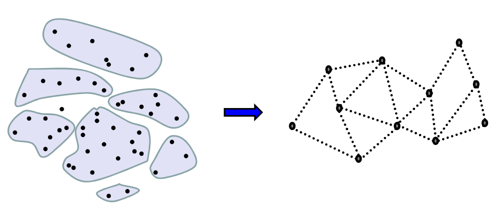


- Sampling distribution can be generated using kino-geometric methods, or by sub-sampling many short MD simulation trajectories, or by other methods

[Apaydin et al., 2003] [Singhal et al., 2004]

41

States are conformation sets



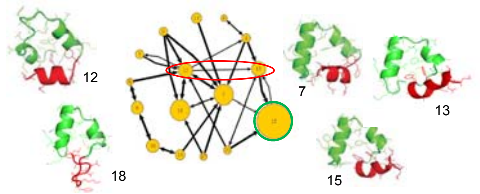
- Sampled conformations are clustered to produce a more compact model and better satisfy the Markov assumption
- Each cluster is computed to approximately match a basin of attraction of the energy landscape

[Chodera et al., 2007] [Chiang et al., 2010]

42

Applications

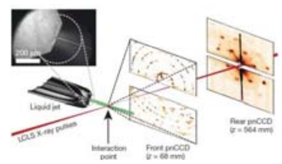
- Analysis of the dynamics of small proteins: alinine dipeptide, villin, tryptophan zipper beta hairpin

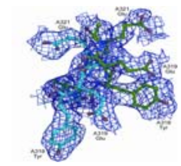


43

New Project (with SLAC)

- Interpretation of electron-density maps generated with **femtosecond X-ray protein nanocrystallography**
  [Chapmann et al., Nature, Feb 2011]





- Combination of state-of-the-art kino-geometric sampling with a (hopefully) revolutionary experimental method

44