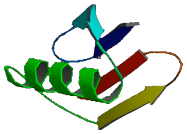
Replica Exchange Sampling

Jianhan Chen
Kansas State University

The Protein Folding Problem

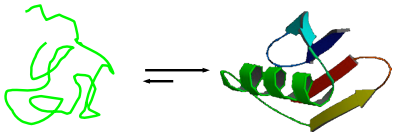
- How does the primary Sequence specify the native fold?
  (Afinsin, Science, 1973)

MTYKLILNGK TLKGETTTEA
VDAATAEKVF KQYANDNGVD
GEWTYDDATK TFTVTE



- Levinthal's Paradox: how protein folds this fast?

τ_{fold} < 1 second

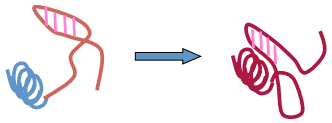


10 conformations per residue,
10^{-11} (10 ps) per state

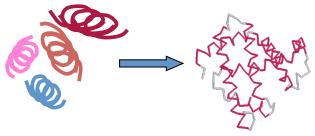
10^{60} states for a 60-residue protein,
$>10^{40}$ years for random search !

How proteins fold?

- "Classical" understandings
  - Diffusion-collision
  - Hierarchical folding
  - Assembly of foldons

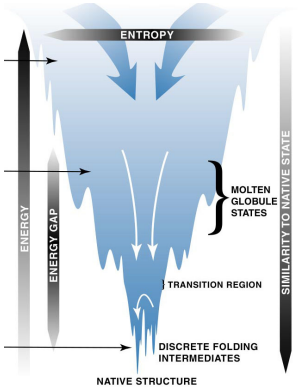


Diffusion and collision
Karplus and Weaver, *Biopolymers.*, 18, 1421 ('77).



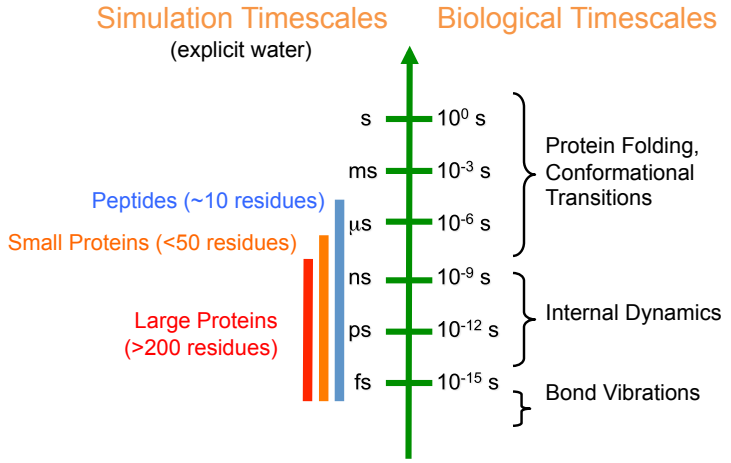
Folding via modular assembly
Ptitsyn and Rashin, *Biophys. Chem.*, 3, 1 ('75).

The Energy Landscape Theory
(arguably the prevailing theory)

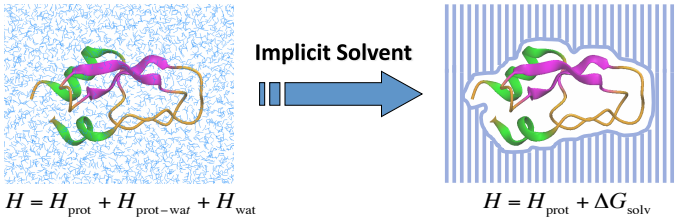


Wolynes et al., *Science* (1995).

Gap in Timescales

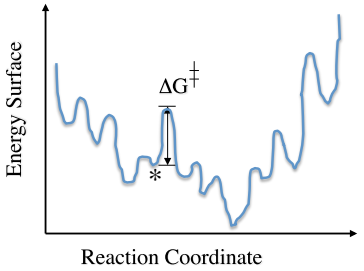


Implicit Treatment of Solvation



- Mean influence of water captured by the solvation free energy
- Very efficient with the generalize Born (GB) approach
  - GBSW (generalized Born with a smooth switching): about 30 times faster than comparable explicit solvent simulations (Im et al., JCC 2003)
- Offers a good balance between accuracy and efficiency

Barriers, Temperature and Timescales



$$\tau = \tau_0 \exp(\Delta G^{\ddagger}/kT)$$

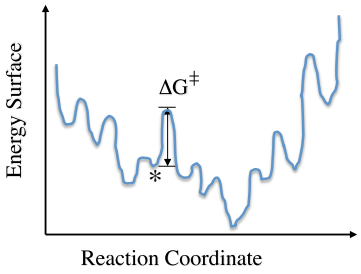
$\tau_0 \sim 10^{-12}$ s ~ ps

T = 300 K

$\Delta G^{\ddagger}$: 1 kcal/mol, $\tau \sim 5$ ps
5 kcal/mol, $\tau \sim 4$ ns
10 kcal/mol, $\tau \sim 17$ μs

Protein energy landscape is highly complex and rugged with numerous local minima.

Barriers, Temperature and Timescales



$$\tau = \tau_0 \exp(\Delta G^{\ddagger}/kT)$$

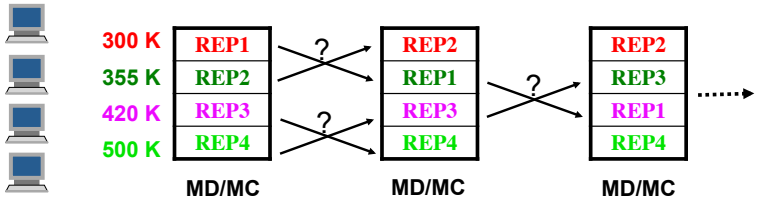
$\tau_0 \sim 10^{-12}$ s ~ ps

T = 600 K

$\Delta G^{\ddagger}$: 1 kcal/mol, $\tau \sim 2$ ps 2x
5 kcal/mol, $\tau \sim 65$ ps 60x
10 kcal/mol, $\tau \sim 4$ ns 4000x

- Increase temperature enhances barrier crossing.
- The enhancement factor depends exponentially on the barrier height.

Replica Exchange (REX) Sampling

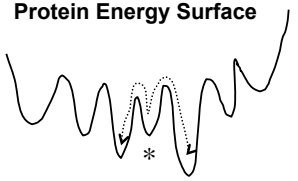


Exchange criteria

$$P_{i \leftrightarrow j} = \begin{cases} 1 & \Delta \le 0 \\ \exp(-\Delta) & \Delta > 0 \end{cases}$$

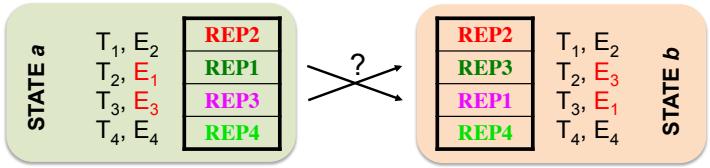
$\Delta = (E_i - E_j) \bullet (1/kT_j - 1/kT_i)$

Protein Energy Surface



Sugita and Okamoto, *CPL* (1999)

REX Exchange Criteria



$$W_a = e^{-\beta_1 E_2} e^{-\beta_2 E_1} e^{-\beta_3 E_3} e^{-\beta_4 E_4} \qquad W_b = e^{-\beta_1 E_2} e^{-\beta_2 E_3} e^{-\beta_3 E_1} e^{-\beta_4 E_4}$$

Detailed Balance requires: $W_a P_{a\to b} = W_b P_{b\to a}$

$$\frac{P_{a\to b}}{P_{b\to a}} = \frac{W_b}{W_a} = \frac{e^{-\beta_1 E_2} e^{-\beta_2 E_1} e^{-\beta_3 E_3} e^{-\beta_4 E_4}}{e^{-\beta_1 E_2} e^{-\beta_2 E_3} e^{-\beta_3 E_1} e^{-\beta_4 E_4}} = e^{-(\beta_2-\beta_3)(E_1-E_3)} = e^{-\Delta}$$

- Proper thermodynamic ensembles at all temperatures.
- No fundamental requirement on the exchange scheme.
- Anything (temperature, Hamiltonian, etc) can be exchanged.

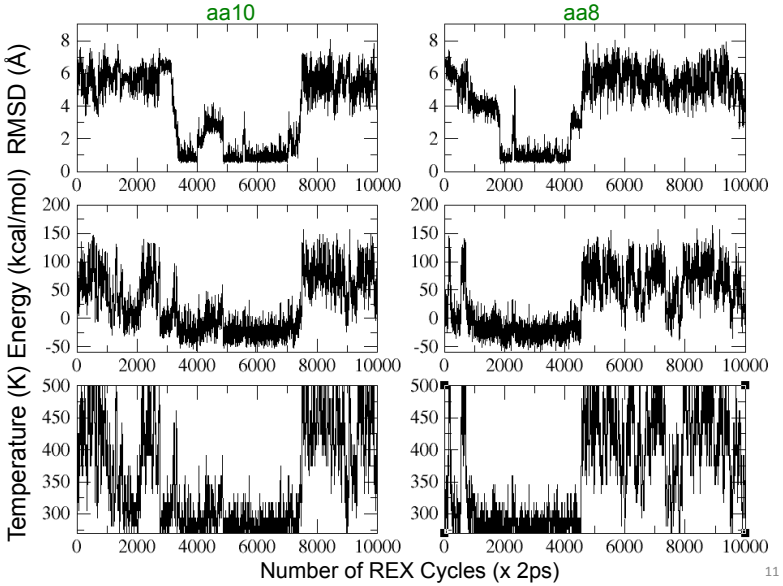
REX using MMTSB Tool Set

- A set of `*rex.pl` scripts (`aarex.pl, gorex.pl`, etc)
- Offers great flexibility in simulation setup (CHARMM vs Amber, choice of force fields, custom setups, etc), machine architecture (shared memory supercomputer, loosely coupled Linux cluster, or any computers linked by Ethernet), job monitoring, post simulation analysis (quick look with `rexinfo.pl`, extracting ensemble, clustering, WHAM etc)

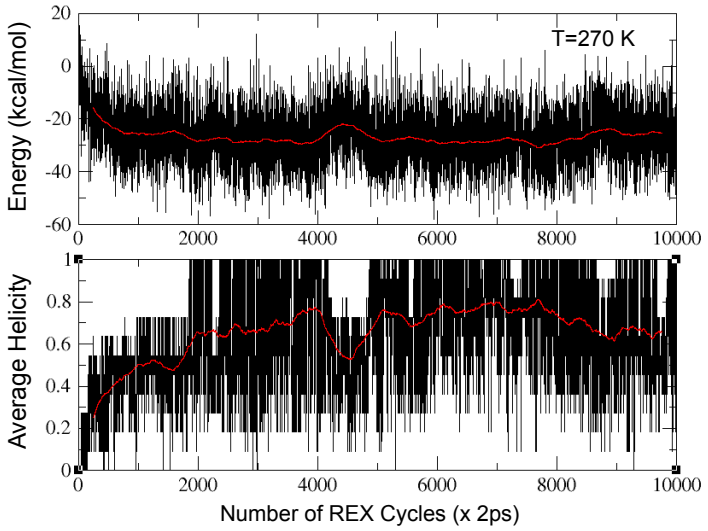
```
aarex.pl –n 100 -mdpar dynsteps=100,param=22,gb,nocut \
–par equilruns=2,natpdb=1vii.exp.pdb -temp 4:298:400 \
1vii.{1,2,3,4}.pdb
```

Runs 100 replica exchange MD simulation cycles with four exponentially spaced temperature windows from 298 to 400K. The native PDB structure is given as reference for calculation RMSD values. MD parameters are set to run 100 steps for each cycle, use CHARMM22 parameters with GB implicit solvent. Initial conformations are taken from the files 1vii.?.pdb.

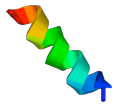
REX-MD Simulation of Helix $(AAQAA)_3$

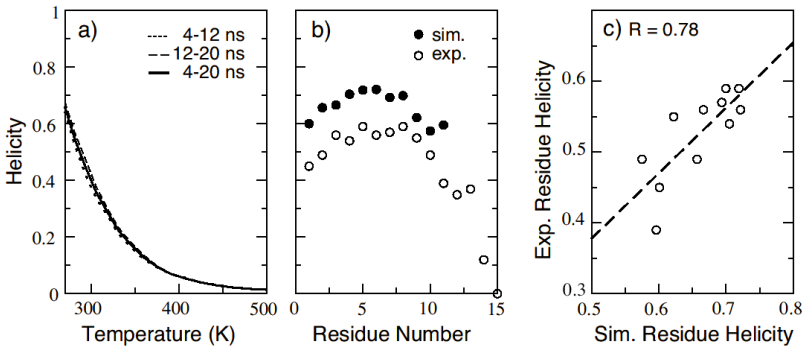


REX Simulation of $(AAQAA)_3$



Helicity of $(AAQAA)_3$

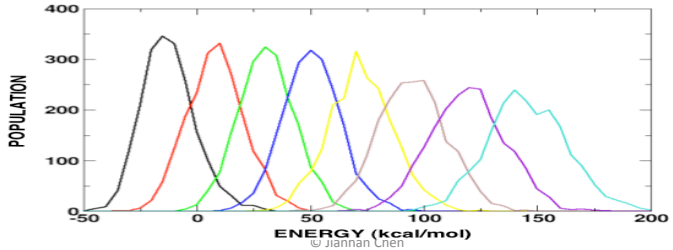




Simulation conditions: CHARMM22/CMAP^{GBSW} with GBSW implicit solvent; total of 20 ns REX-MD with 16 replicas spanning 270-550 K from fully extended conformation; helicity computed from averaged 1-4 backbone hydrogen-bonding frequency using a heavy atom distance cutoff of 3.6 Å;

Optimal Setup of REX

- Temperature range and distribution
  - Exponentially distributed for systems with constant heat capacity
  - Lowest temperature: often the temperature of interest (e.g, 300 K)
  - Highest temperature: less obvious, but a critical parameter
- Number of replicas
  - Depends on the system size: ~ sqrt(*N*), and the temperature range
  - Judged by the exchange acceptance ratio (optimal 0.2 – 0.3?)
  - Implicit solvent is particularly suitable for REX!



© Jianhan Chen

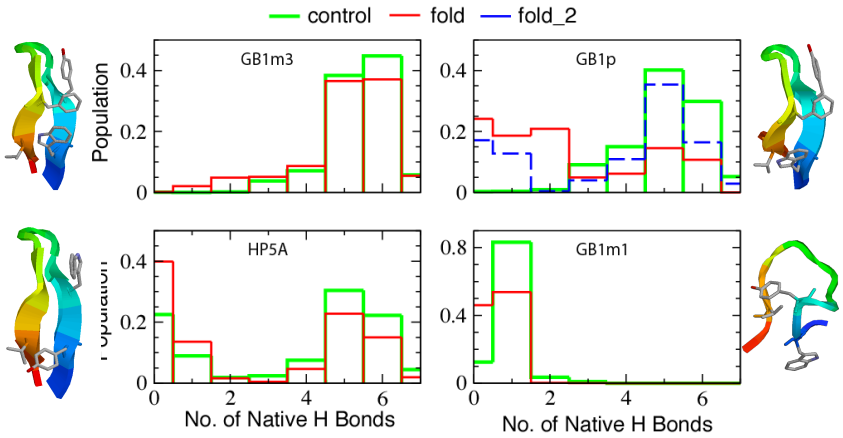
Optimal Setup of REX

- Temperature range and distribution
  - Exponentially distributed for systems with constant heat capacity
  - Lowest temperature: often the temperature of interest (e.g, 300 K)
  - Highest temperature: less obvious, but a critical parameter
- Number of replicas
  - Depends on the system size: ~ sqrt(*N*), and the temperature range
  - Judged by the exchange acceptance ratio (optimal 0.2 – 0.3?)
  - Implicit solvent is particularly suitable for REX!
- Exchange attempt frequency
  - fast but not too fast (e.g., 1-2 ps)
- Length of simulation
  - **Convergence** can be tricky to judge: plateau ≠ convergence!
  - Depends on the target properties: folding takes very very very long
  - Multiple independent simulations often helpful in testing convergence

Simulations are rarely too long!

© Jianhan Chen

Conformational Equilibria of β-Hairpins

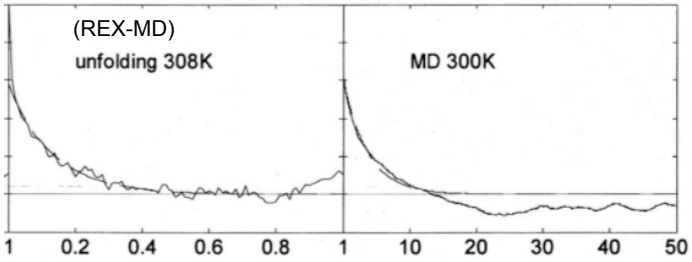


Simulation details: CHARMM22/CMAP^{GBSW}/GBSW; 30-50 ns REX; 16 replicas at 270 - 550 K (fold) or 270-400 K (fold_2); population computed from final 10 ns; a heavy atom cutoff of 3.6 Å used for h-bond.

Does REX Really Work?

(Peptide simulation benchmarks)

Helicity auto-correlation functions (Zhang et al, JCP, 2005)



System: Fs-21, $A_5(AAARA)_3$, peptide in GB implicit solvent

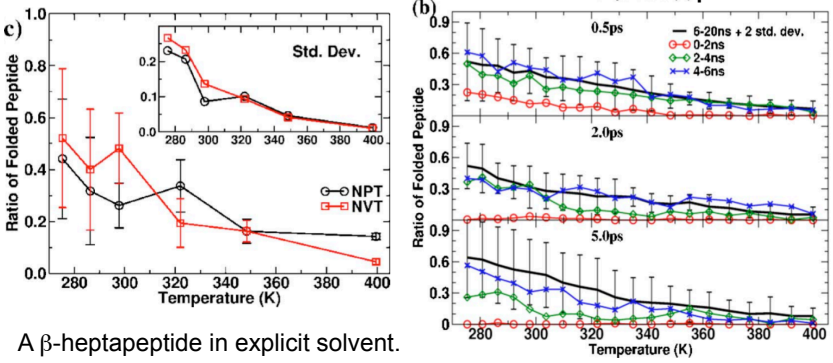
Observation: enhancement of 35X at 300K and 72X at 270K.

Enhancement is even larger if considering the full temperature range. Also note some caveat with auto-correlation analysis.

Does REX Really Work?

Standard deviations of folded fraction: (Periole and Mark, JCP, 2007)

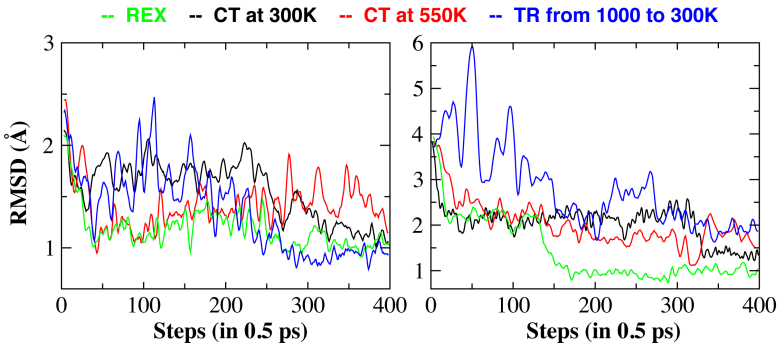
800 ns MD runs | **REX-MD with 20 replicas**



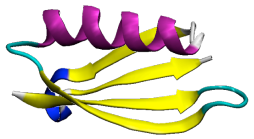
A β-heptapeptide in explicit solvent.

Observation: approximately 10X more efficient (than a single 800 ns conventional MD at 275 K); only about 2-4X at 300 K. If the goal is to determine the temperature dependence, >30X.

Does REX Really Work?



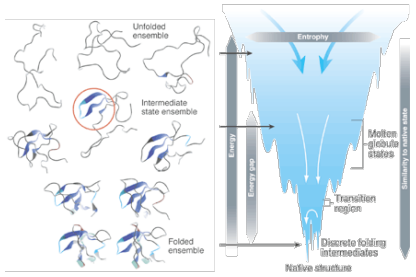
Refinement of NMR Structures: protein G B1 in GBSW implicit solvent. REX traces from single replicas are shown.



Chen et al., JACS (2004)

How to Quantify "Sampling"

- Auto-correlation analysis
- Statistical tests
- Conformational space analysis
- Thermodynamics properties
- Transition events
- Effective sample size through structural distribution analysis (Zuckerman, JCP 2007)

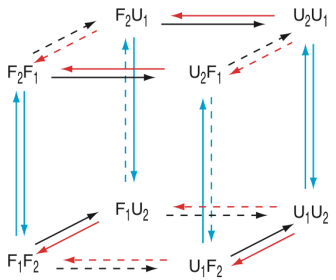


Brooks et al, Science (2001)

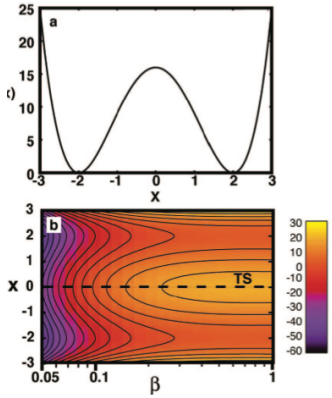
- All these analysis limited by "sampling" (i.e., knowledge of the true conformational space)!

Does REX Really Work?

- Theoretical considerations based on simple kinetic model systems: enhancement possible as long as there is a positive heat of activation.



Zhang et al, (2007, 2008)

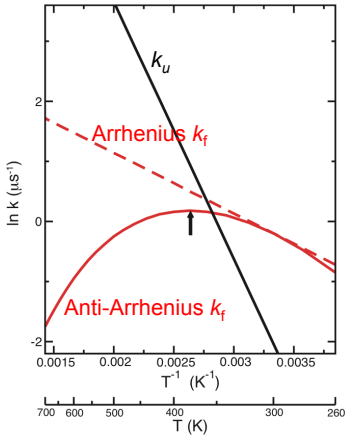


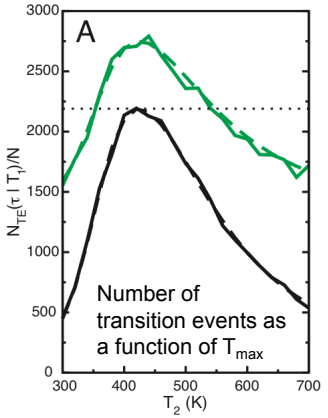
Also see: Zuckerman (2006)

Nymeyer (2008)

Importance of the Maximum Temperature

- Anti-Arrhenius behavior of protein folding: heat of activation disappears or even becomes negative at high temperature.





Zhang et al., PNAS (2007).

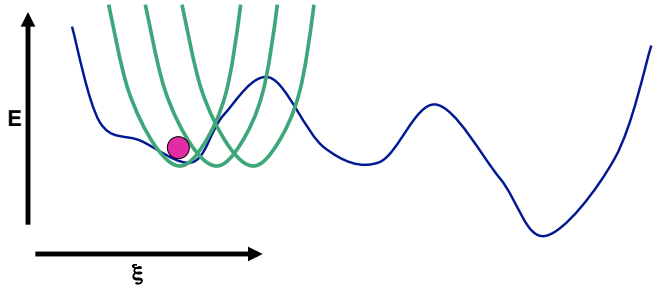
REX Extensions and Variants (REX zoo?)

- No fundamental requirement on the exchange scheme.
- Anything (temperature, Hamiltonian, etc) can be exchanged
- Goals are generally to
  - **Enhance temperature random walk** (exchange efficiency): reduce the number of exchanging degrees of freedom (solute tempering, local REX, REX essential dynamics, etc), global reassignment (TIGER & TIGER2), dynamic scaling, …
  - **Enhance basin hopping**: Hamiltonian REX (like any accelerated MD, it is very tricky to find proper scheme for modifying the potential; there are many versions of this type), multiscale REX (resolution exchange)
  - **Combine with biased sampling**: 2D REX (mostly for Go-modeling)
  - **Others**: serial REX, multiplex REX (no synchronization requirement), …

Simple temperature REX still appears to be the most applicable in general. There is no substitution for serious sampling?

Umbrella Sampling

Apply series of overlapping harmonic functions (umbrellas):

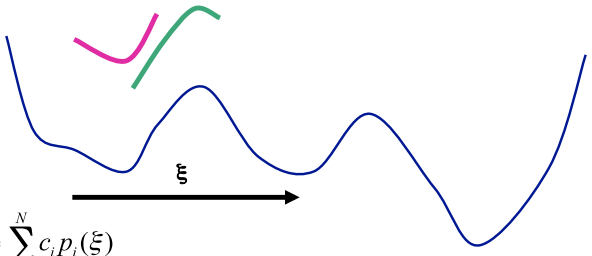


Need to know good reaction coordinate to find reasonable path.

Poorly chosen reaction coordinate might result in high barriers.

Weighted Histogram Analysis Method (WHAM)

Multiple umbrellas result in free energy profiles that are determined only within a constant C_i. They need to be stitched together to get profile for entire path!



$$p(\xi) = \sum_{i=1}^{N} c_i p_i(\xi)$$

$$= \sum_{i=1}^{N} c_i \, e^{\beta(U_{umbrella,i}(\xi) - C_i)} \, p_{biased,i}(\xi)$$

c_i's are determined from minimizing statistical variance at each histogram bin

→ $c_i = f(C_i)$ but $C_i = f(p(\xi))$ need to solve iteratively

Analysis of REX Results

- Generally concerns thermodynamics
  - Methods exist for recovering some kinetic information (e.g., see Buchete and Hummer, 2008)
- Direct histogram (free energy) and clustering analysis of structural ensembles at temperatures of interest
  - Combine with transition state analysis
- T-WHAM (temperature weighted histogram analysis method)
  - Combine sampling from all temperatures
  - 2D or higher dimension histograms require large number of snapshots
  - More useful for small systems with moderate range of energy fluctuation (small peptides and/or reduced models): do not even attempt for explicit solvent REX
  - Implemented in MMTSB (see tutorial)

Gallicchio et al, JPCB (2005); Chodera et al, JCTC (2007)

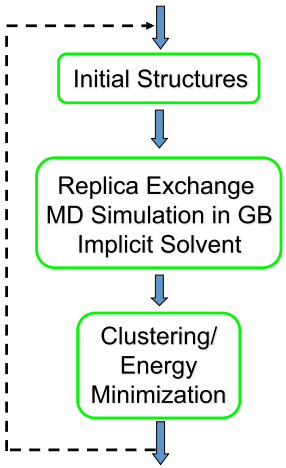
© Jianhan Chen





REX Application Example 1:

REX/GB Refinement of NMR Structures

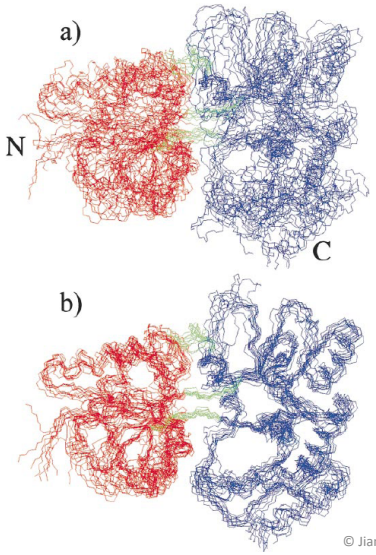


- from experiment or prediction
- in present of restraints, derived either from experiments or prediction
- weakly imposed for stability and flexibility
- hierarchical clustering
- rank by population of clusters
- repeat for recursive restraint assignments or for convergence

Chen, Im, Brooks, *JACS* (2004).

© Jianhan Chen

Refinement of Maltose-Binding Protein (MBP)



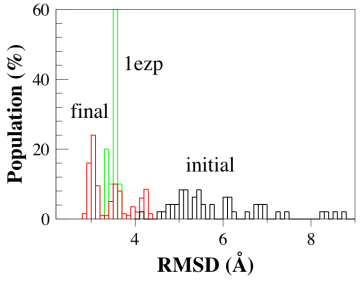
- 370 residues, 42 kDa
- 1943 NOE
- 45 hydrogen bonding
- 555 dihedral angle restraints.
- Average backbone RMSD to X-ray structure is 5.5 Å (a).
- Improved to 3.3 Å with 940 additional dipolar coupling based restraints (b).

Mueller *et al*., JMB **300**, 197 (2000)

© Jianhan Chen

REX/GB Refinement Results

- All NOE and dihedral angle restraints were used.
- 48 replicas were simulated at 300 to 800 K until "converged".
- Total of 1.0 ns REX/GB

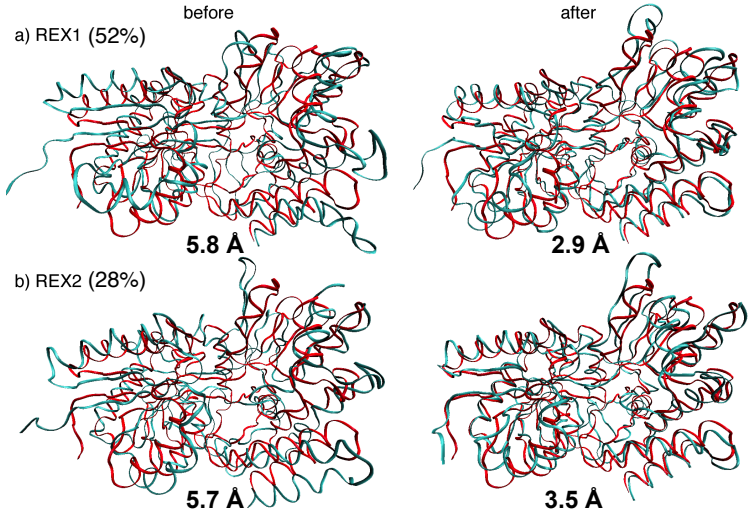


	Initial	Final
RMSD to X-ray (Å) [a]		
Global	4.3±4.1	2.3±2.6
N-domain	2.5±2.1	2.2±1.4
C-domain	3.0±3.2	2.0±1.9
ϕ/ψ space: residues (%)		
Most favored	72.2	84.3
Additionally allowed	22.8	13.3
Generously allowed	3.8	1.6
Disallowed	1.2	0.8
Violation statistics		
RMSD of NOEs (Å)	0.0047	0.014
NOE violations (> 0.2Å)	2.85	4.42
RMSD of angles (in degrees)	0.53	6.25

[a] Backbone RMSD with respect to PDB: 1dmb shown. Global: residues 6-235 and 241-370; N-domain: 6-109 and 264:309; C-domain: 114-235, 241-258 and 316-370.

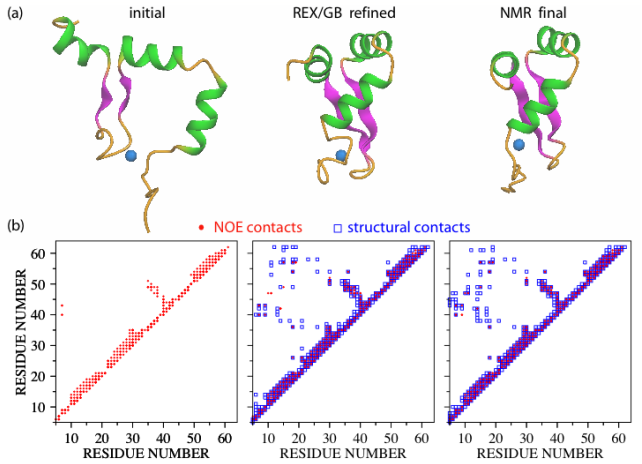
© Jianhan Chen

Representative Structures: MBP



RMSD values: from X-ray (PDB:1dm); backbone atoms of residues 6-235 and 241-370

REX/GB Refinement can be used to obtain native-like models from limited NMR data.



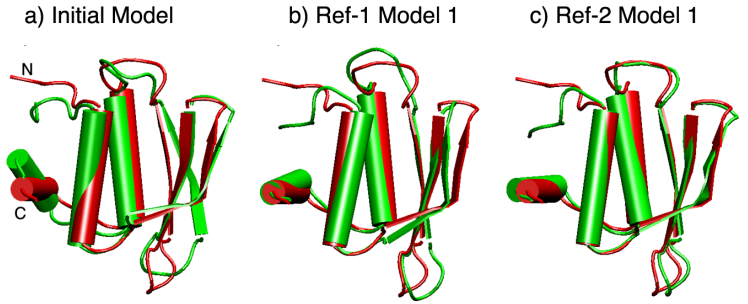
Cost: ~12h wall time using 16 Intel 2.4GHz CPUs

Chen et al., *JBNMR* (2004).

REX Application Example 2:

REX/GB Refinement of Predicted Structures

Target TMR04 from CASPR Refinement Experiment



	a) Initial Model	b) Ref-1 Model 1	c) Ref-2 Model 1
RMSD (Å)	2.18 (2.04)	1.88 (1.26)	1.57 (0.98)
GDT_TS	76.1	85.7	90.0

Chen and Brooks, Proteins (2006)

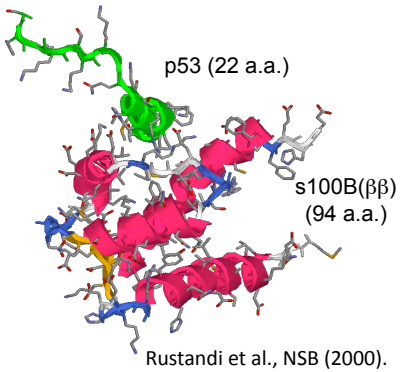
© Jianhan Chen

REX Application Example 3:

Free Energy of p53/S100B(ββ) Interaction

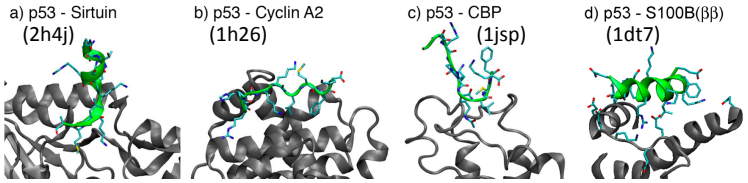
- S100B(ββ) binds p53 in a Ca^{2+}-dependent manner
- Inhibits the phosphorylation of C-terminal negative regulatory domain (residues 367-388) and prevents p53 activation
- An anti-cancer drug target

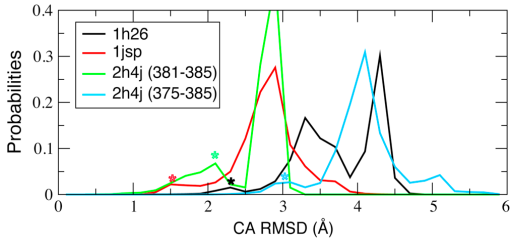
- Free p53 extreme C-terminus is an intrinsically disordered protein (IDP)
- Folds into a partial helix upon binding s100B(ββ)
- One of the few known IDPs that adopt multiple conformations when bound to different targets.
- an ideal model IDP for simulation?



Rustandi et al., NSB (2000).

Pre-existence of Fold-Like Conformations in the Free p53 Peptide

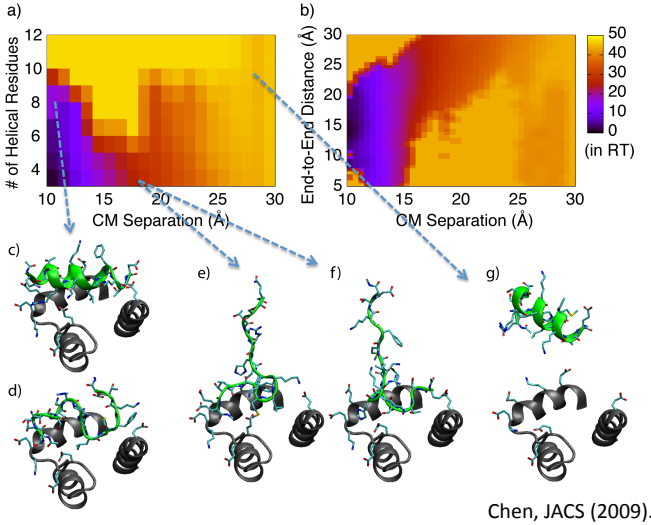




A usual suspect of conformational selection, or not?

Note: a randomized sequence does not give rise to folded-like conformations.

The p53 extreme C-terminus Binds to S100B(ββ) through Induced Folding/Fly-Casting



Chen, JACS (2009).

REX is more (or less) limited than what you might believe.

The complexity of conformational space should never be under-appreciated.

There is really no way around it.
(reduced models might help but with compromises)

The energy function is important in sampling.