Simulations of peptide folding: structures, dynamics, pathways

Krzysztof Kuczera

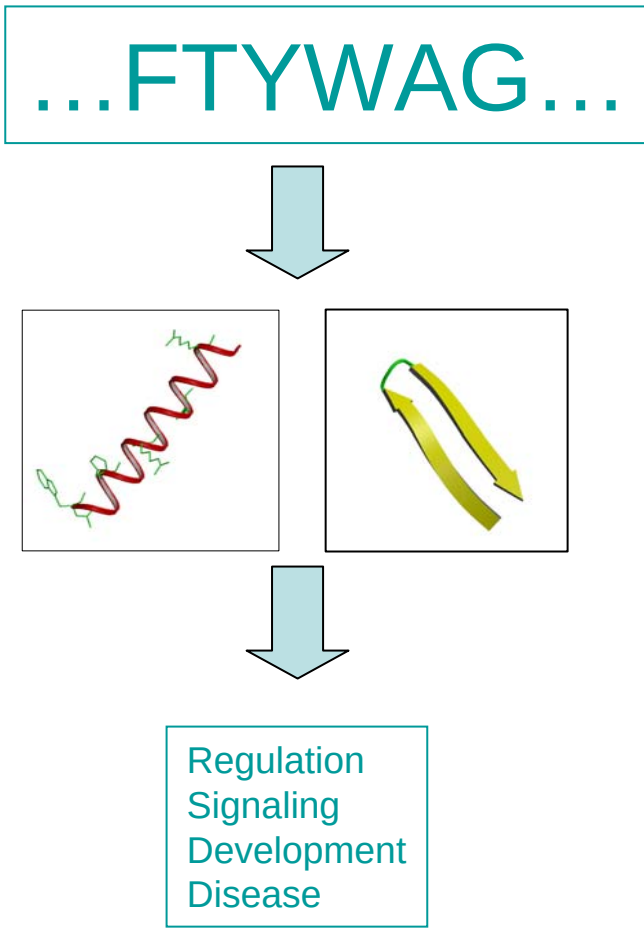
Departments of Chemistry and Molecular Biosciences,
University of Kansas, Lawrence, KS 66045

Torun-Warszawa, Maj 2010

Peptide dynamics: Significance

- Peptides = biologically active structure-forming molecules
- Peptides = small size allows study of sequence – structure – function relations
- Peptides = flexible, dynamic systems motions on ps – µs time scale experiment/simulation overlap
- Peptides = building blocks of proteins

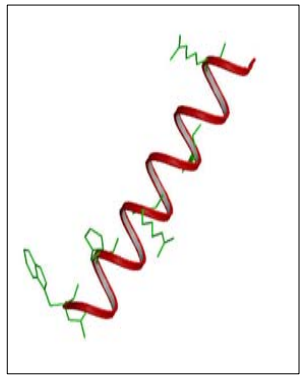
  →understanding of fundamental biological processes



Peptide Folding Simulations

GOALS:

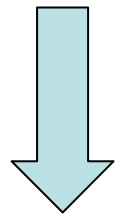
- -Predict process: populations, rates, paths
- -Verify methods: algorithms and force fields
- -Complement experimental data
- -Understanding→design materials, drugs



EXPERIMENTAL data: typically

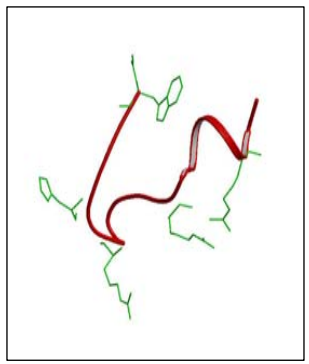
- -structure and population of folded state
- -folding and unfolding rates (T)
- - rarely: “nucleation rate”

$\tau \approx 300$ ns
50% α
@300 K



Unique ROLE for simulations: microscopic

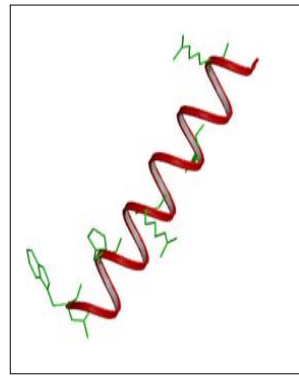
- -Information on pathways
- -Information on unfolded state(s)
- -Dynamics $\perp$ to reaction coordinate



Folding Simulation Methods

Fast processes: ($\tau \approx$ 10-100 ns)

Direct molecular dynamics (MD) gives complete description

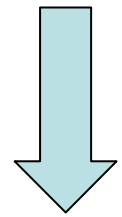


Slow processes:

Populations:

Enhanced sampling methods

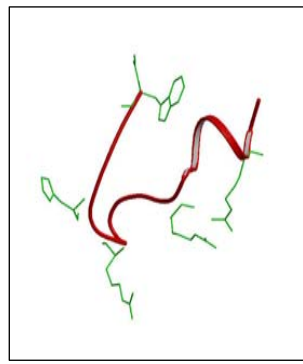
-e.g. replica-exchange MD



Kinetics:

Path sampling

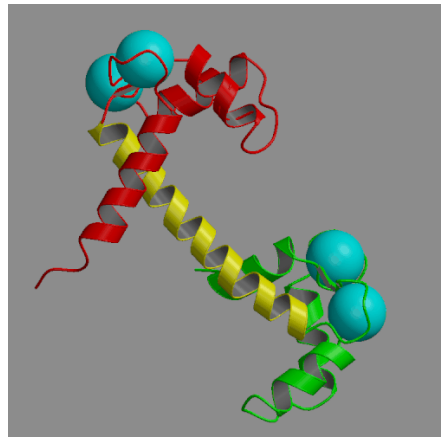
-e.g. milestoning



Limitations:

Force field accuracy, system size

MOLECULAR DYNAMICS SIMULATIONS



- Model system of N atoms
- Introduce potential energy U(x,y,z)
- Calculate force acting on each atom
- Solve Newton's equations of motion
- Generate a trajectory for each atom $x_i(t)$
- Analyze structure, motions and interactions
- Relate to experimental observations

Newton's 2nd Law

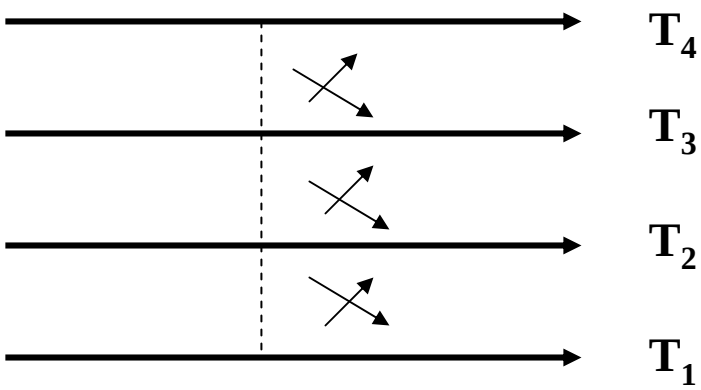
$$m_i \frac{d^2 \vec{r}_i}{dt^2} = \vec{F}_i = \nabla_i U$$

Verlet algorithm

$$x(t+\Delta t) = 2x(t) - x(t-\Delta t)$$

$$v(t) = \frac{x(t+\Delta t) - x(t-\Delta t)}{2\Delta t}$$

Replica-exchange molecular dynamics



Propagate independent trajectories at temperatures

$T_1 < T_2 < T_3 < \ldots$

Stop and compare energies

Exchange between neighbors

$$w(i \to j) = 1 \qquad \Delta \le 0$$

$$w(i \to j) = e^{-\Delta} \qquad \Delta > 0$$

$$\Delta = (\beta_j - \beta_i)(E_i - E_j)$$

$$\beta_i = \frac{1}{kT_i}$$

Advantages:

+ accelerated sampling @ low T
+ Boltzmann distributions @ all T
+ Minimal process communication
+ Property sampling as f(T)

WH5: Fastest Folding α-helix

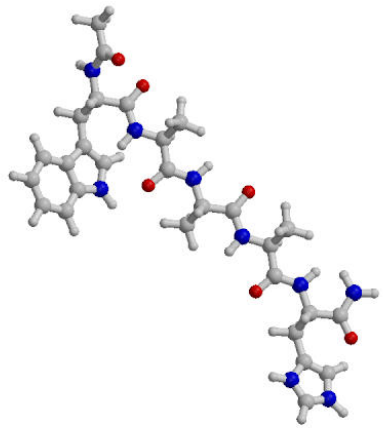
Experimental at 300 K:

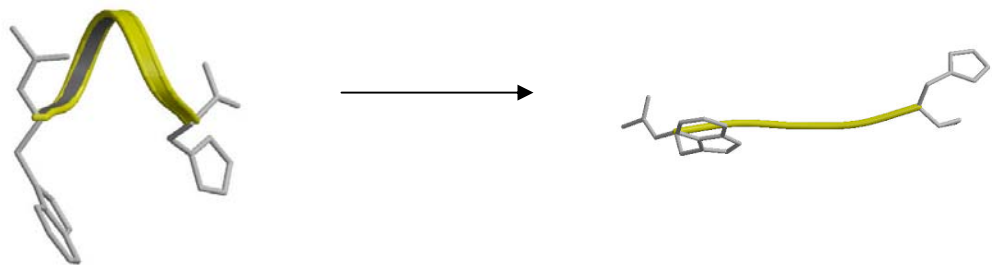
CD spectroscopy:
% α = 20-25 %

Fluorescence T-jump:
Relaxation τ_1 = 5.3 ± 1.9 ns
τ_2 = 0.85 ± 0.3 ns

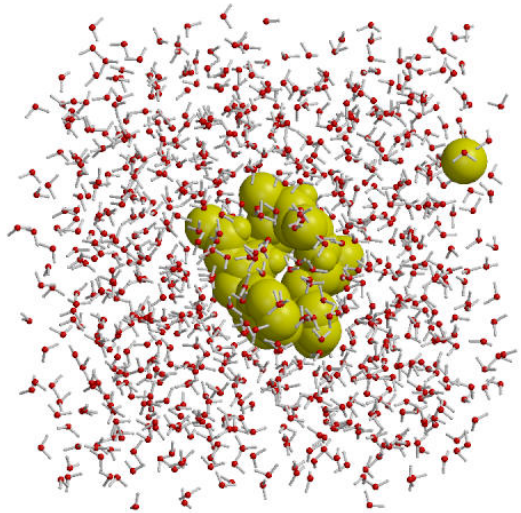
Gouri S. Jas, Baylor University
Angewandte Chem. (2009) **48**:5628

Sequence: 5 aa
Ac-Trp-Ala-Ala-Ala-His^+-NH_2

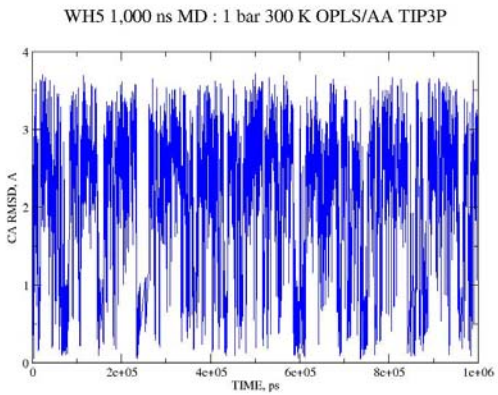


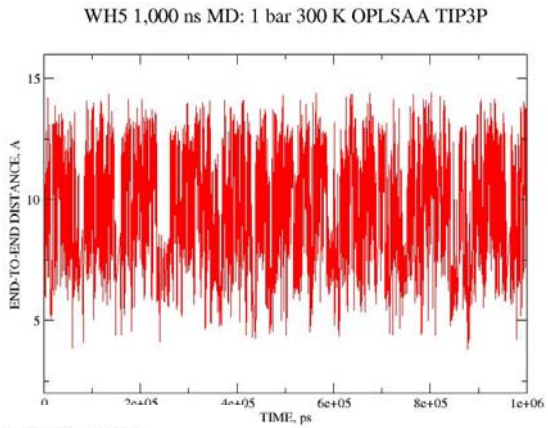


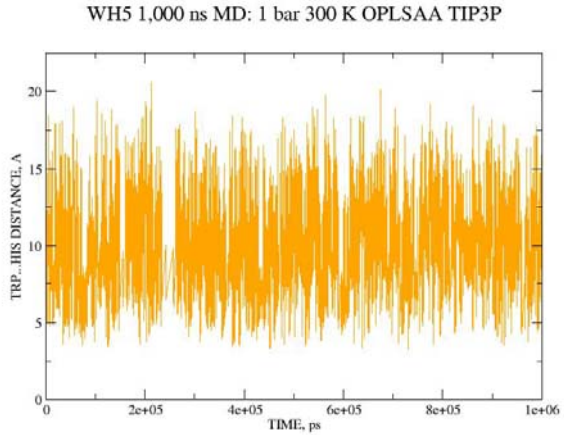
WH5 : Global MD

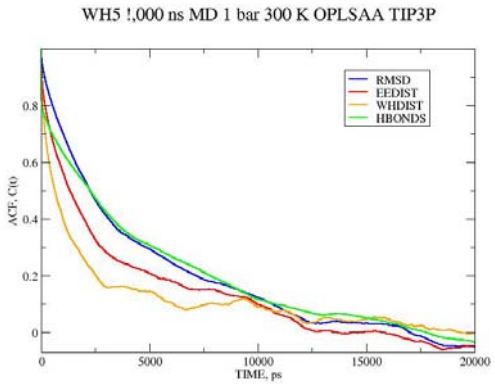


MD: 1,000 ns NPT at 300 K, 1 bar with GROMACS program and several protein force fields, ≈1000 waters, 1 Cl^-
960 ns with CHARMM program and CHARMM ff



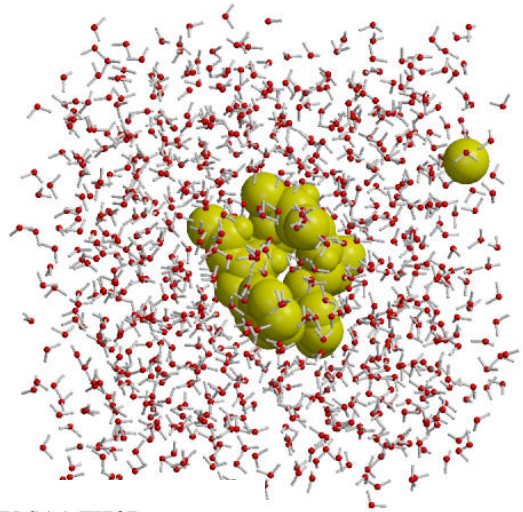




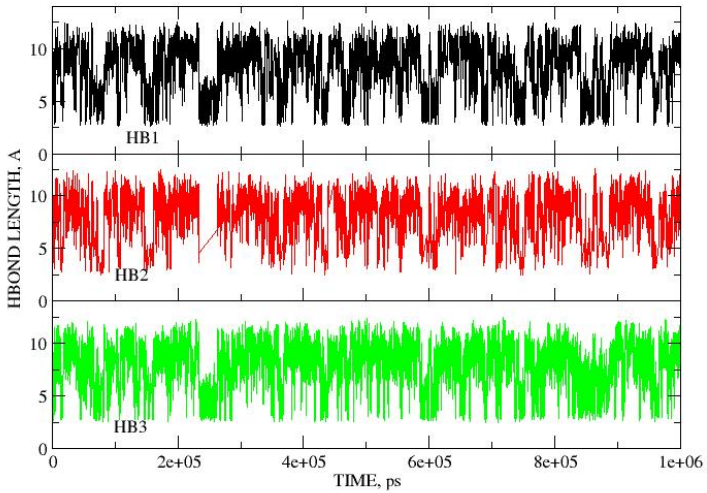


Sample OPLS/AA results

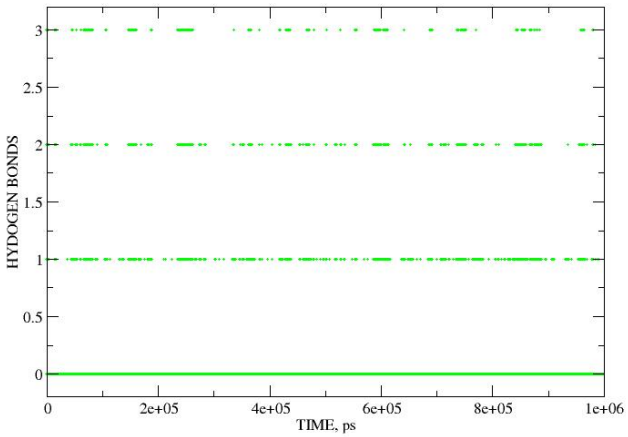
WH5 : Local MD



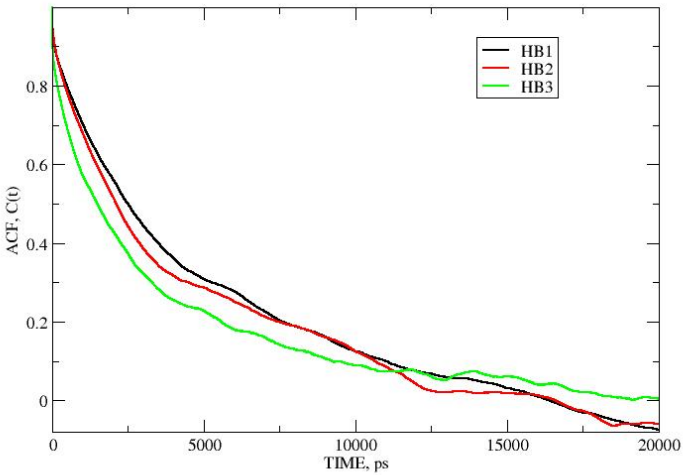
WH5 1,000 ns MD: 1 bar 300 K OPLSAA TIP3P



WH5 1,000 ns MD: 1 bar 300 K OPLSAA TIP3P



WH5 1,000 ns MD: 1 bar 300 K OPLSAA TIP3P



Sample OPLS/AA results

WH5: helix populations and kinetics

Force Field	τ_{fold} ns	τ_{unf} ns	τ_r ns	τ_{nuc} ns	% α HB	% α PP
OPLS/AA	23.	4.1	3.6	0.6	13	11
CHARMM	20.	9.7	6.5	1.0	23	21
G43A1	87.	0.8	0.8	0.1	2	8
G53A6	500.	0.4	0.4	0.05	0.4	3
AMBER03	7.1	8.0	3.8	0.4	31	27
AMBER99P	0.4	9.3	0.4	0.1	64	49
AMBER99SB	44.	3.4	3.1	0.3	6	7
AMBERGS	3.5	233.	3.5	0.1	84	65

Experiment:
%α = 20-25%

Relaxations:
5.3 and 0.8 ns

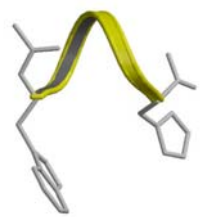
Folding:

$\tau_{fold} \approx$ 30 ns

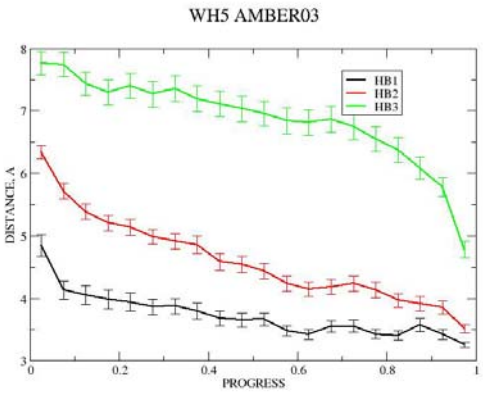
$\tau_{unf} \approx$ 6 ns

Amazing agreement:
Most force field predictions are within a factor of 10 of experimental data!

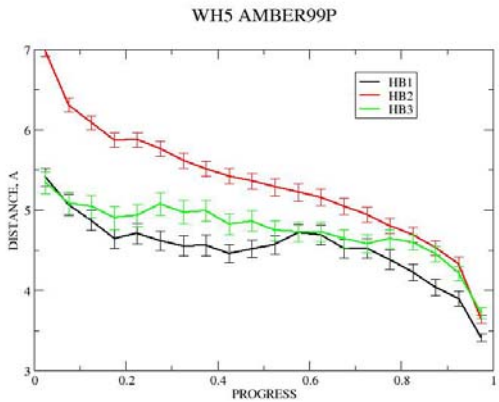
Corresponding $\Delta E \approx 1$ kcal/mol at 300 K



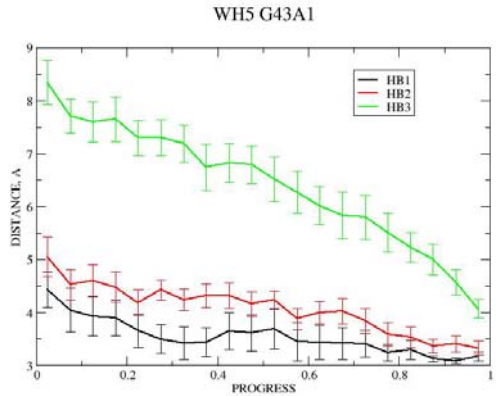
Folding of WH5: pathways



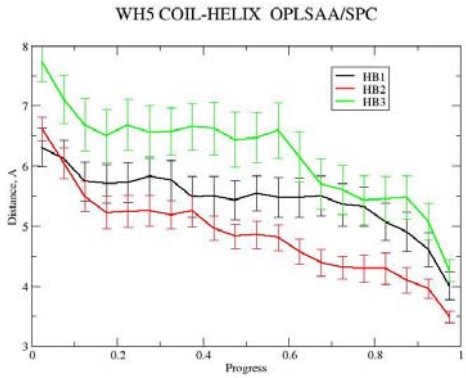
AMBER03, AMBERGS: 1-2-3



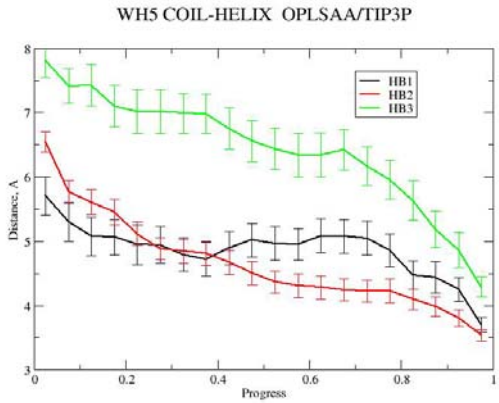
AMBER99P: (1+3)-2



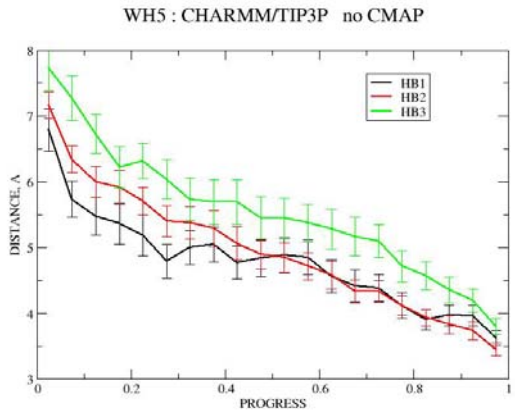
G43A1, AMBER99SB: (1+2)-3



OPLS/AA(SPC): 2-1-3

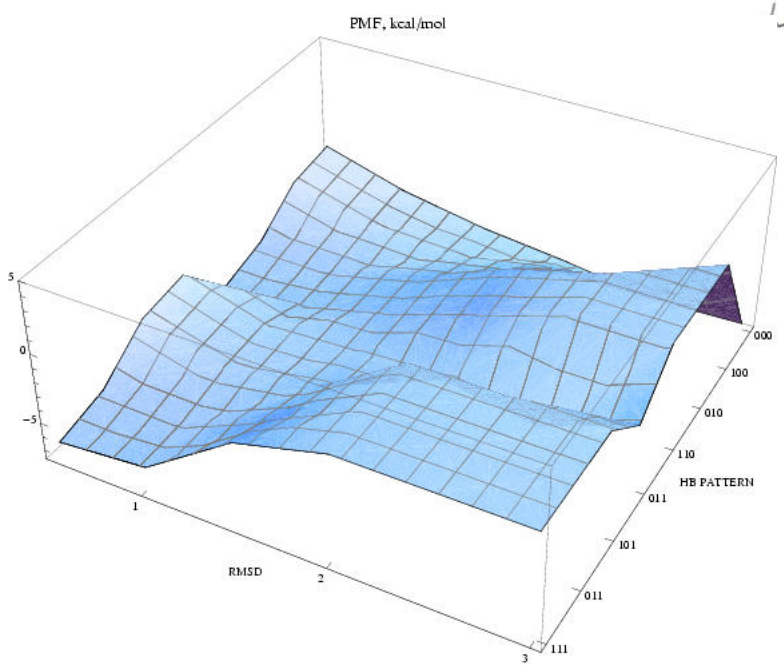


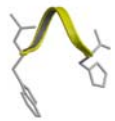
OPLS/AA(TIP3P): 2-1-3
or (1+2)-3

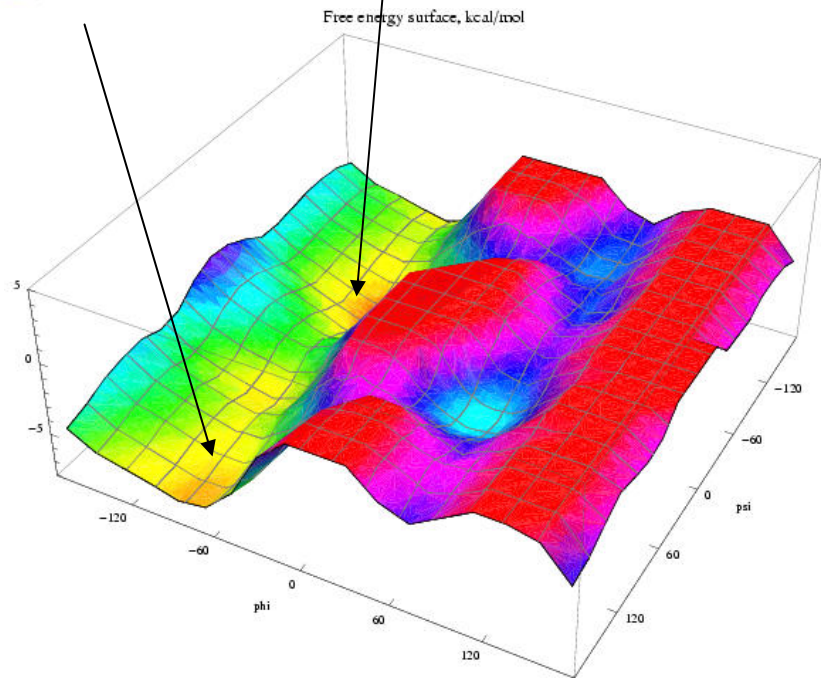


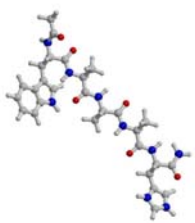
CHARMM: 1+2+3
or (1+2)-3

WH5 in OPLS/AA: conformations

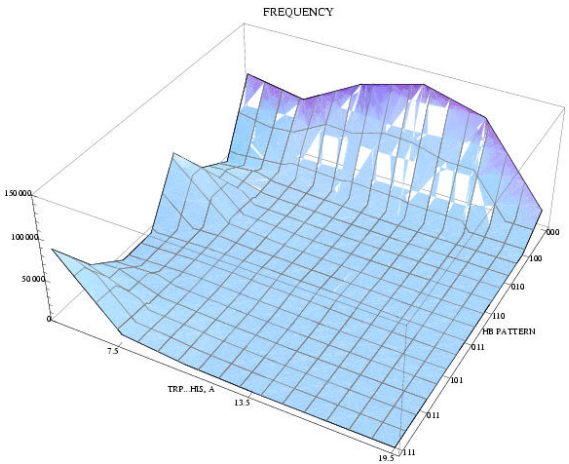


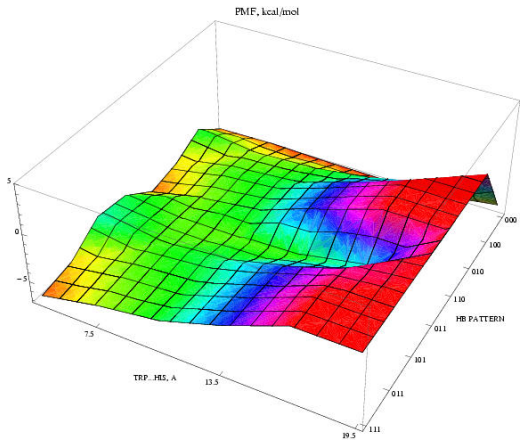






WH5: Trp…His distance (CHARMM)



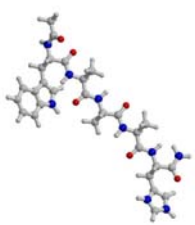


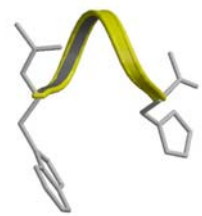
Correlations:

R(W…H) - RMSD: r = 0.55

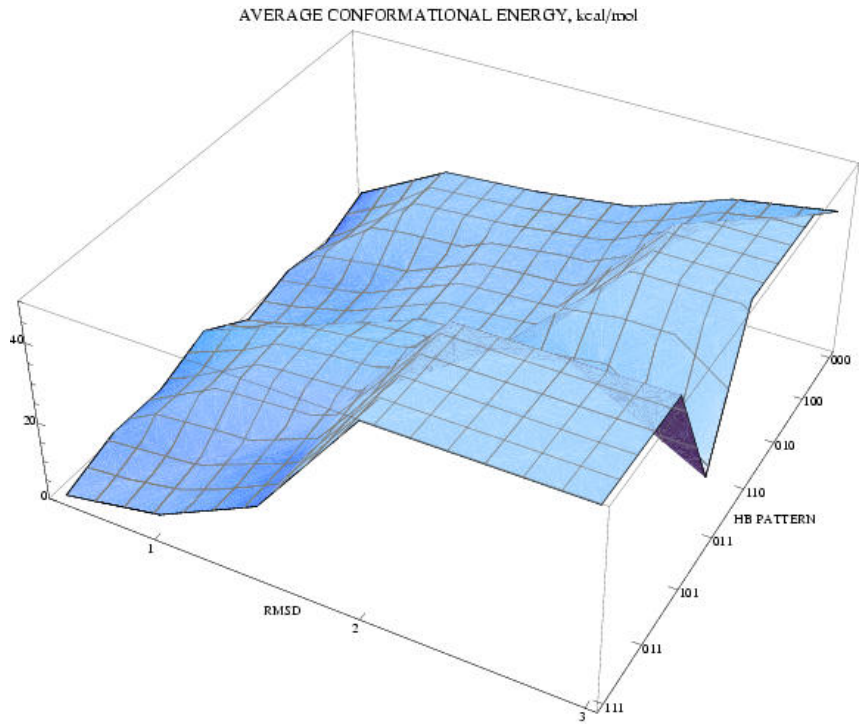
R(W…H) - HB1, HB2, HB3 : r = 0.43, 0.59, 0.35

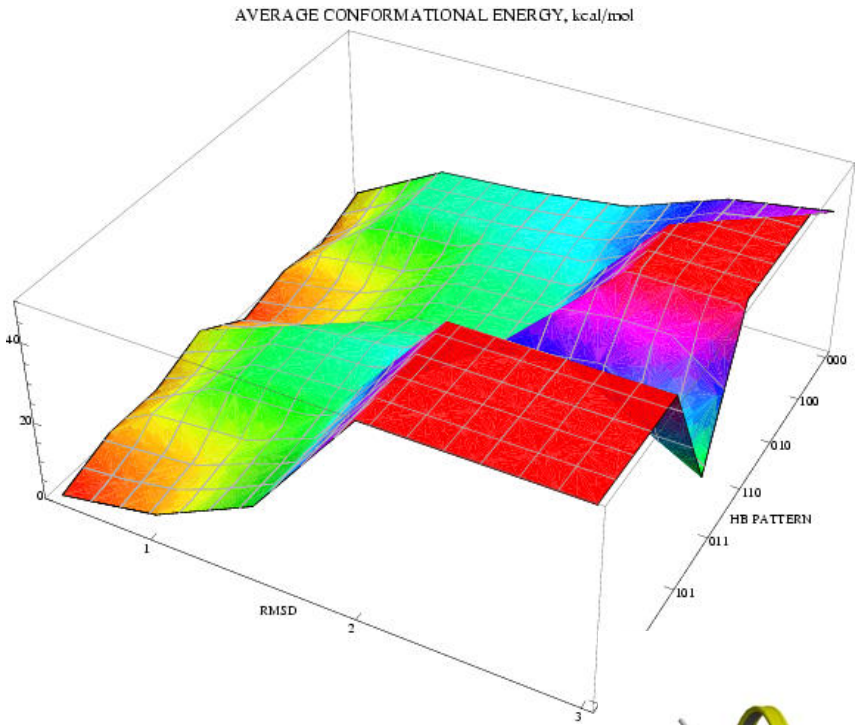
Close Trp…His contact is correlated with
global RMSD from helix & HB2 formation

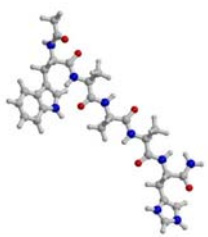


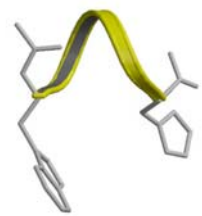


WH5: conformational energy (CHARMM)









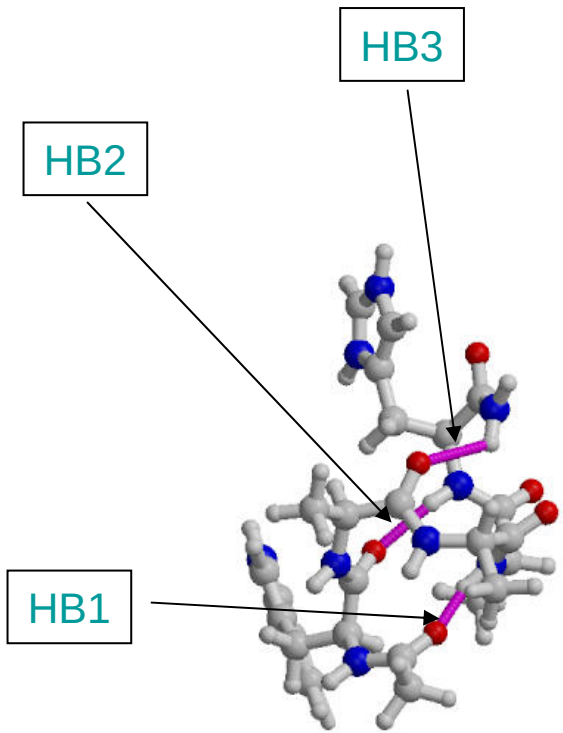
WH5 hydrogen bond dynamics

Force Field	HB1			HB2			HB3		
	τ_f	τ_u	τ_r	τ_f	τ_u	τ_r	τ_f	τ_u	τ_r
AMBER03[a]	261	355	150	784	911	**421**	147	50	37
AMBER99P[a]	39	366	35	102	1066	**93**	36	228	31
AMBER99SB[a]	815	110	97	3258	351	**317**	637	53	49
AMBERGS[a]	24	853	23	148	30354	**147**	43	392	39
G43A1[a]	2278	126	119	1384	88	**83**	3246	45	44
G53A6[a]	2440	44	43	5460	45	**45**	4274	23	23
OPLS/AA[a]	1066	218	152	2768	623	**508**	840	148	126
OPLS/AA[b]	723	187	149	2699	762	**594**	496	110	90
CHARMM[b]	478	236	158	3160	1596	**1060**	218	122	78

[a]With SPC water [b]With TIP3P water

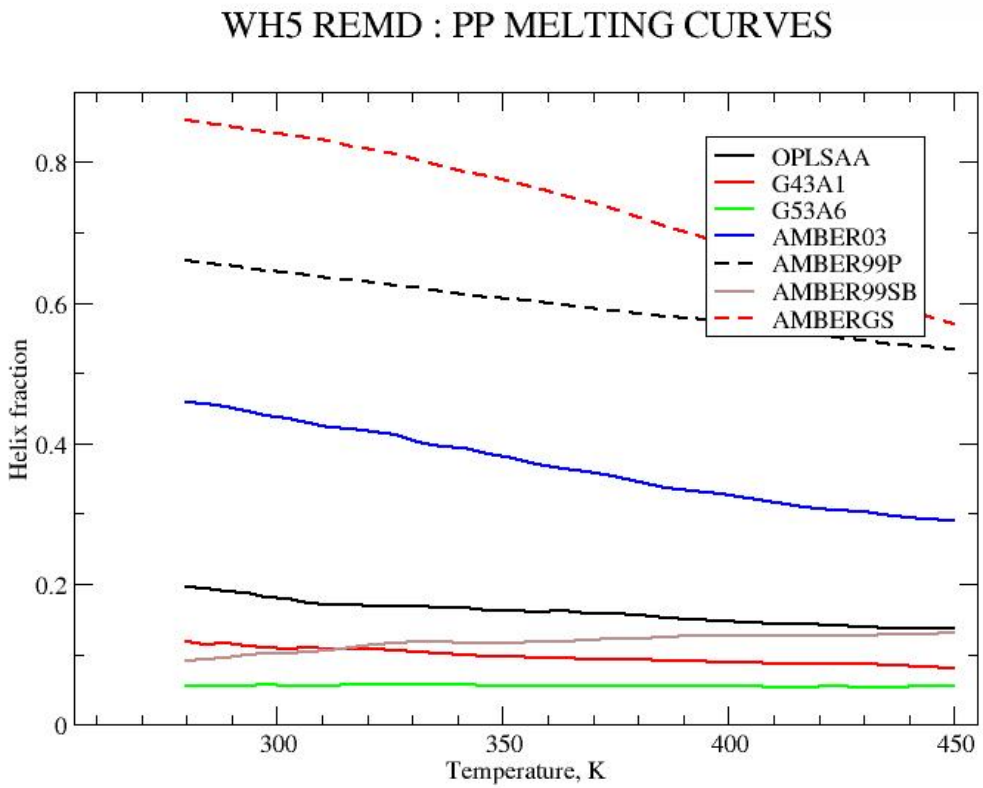
H-bond dynamics time constants in ps.

Relaxation of central hydrogen bond HB2
is in the 0.1-1.0 ns range for most studied FF.



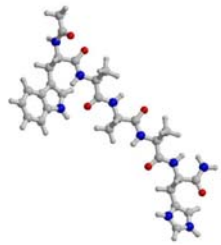
REMD of WH5

- At 300 K REMD=MD
- OPLS/AA, AMBER03, AMBER99P and CHARMM22 give excellent helicity predictions at 300 K
- Helix persistence exaggerated
- AMBER99SB – anti-melting

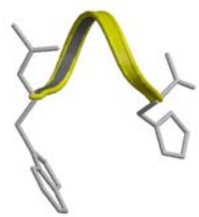


REMD simulations: 32 replicas, 280-450 K, 30 Å cubic box with ca. 1000 waters, 100 ns NPT trajectory with GROMACS

WH5: CONCLUSIONS



- Most of the popular force fields applied here give reasonable predictions for WH5 helicity and kinetic rate constants
- Simulations suggest that the experimentally observed 5 ns process corresponds to helix folding, while the 1 ns process corresponds to helix nucleation – most probably formation of first two hydrogen bonds
- Force fields differ in details of the predicted folding pathway; a majority suggest a “zipper” model, with folding initiated at the N-terminus and progressing consecutively to C-terminus
- The formations of the three hydrogen bonds exhibit significant correlation
- Trp…His sidechain interactions play an important role in structure stabilization
- Force field accuracy is the limiting factor for current biomolecular computer simulations



Acknowledgments

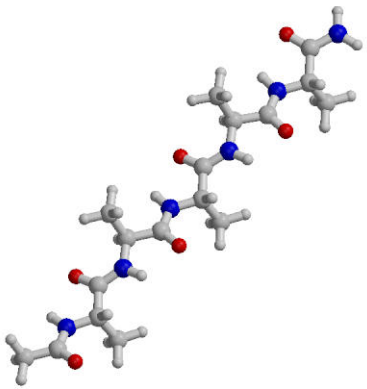
Experiments:
Gouri Jas, Baylor University.

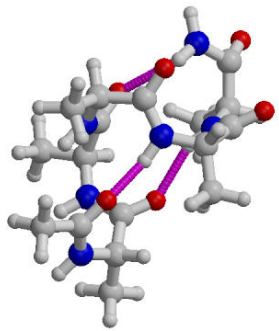
Computer time and technical assistance:
Academic and Research Computing Services,
Baylor University

Funding:
Baylor University internal funds
Big XII Fellowship from University of Kansas

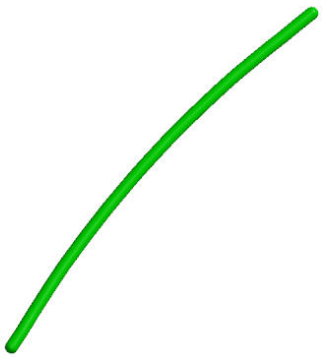
The story of Ala_5

Sequence: 5 aa
Ac-Ala-Ala-Ala-Ala-Ala-NH_2





?



Alanine-based peptide folding simulations

- **Replica exchange** simulations by Garcia et al. showed exaggerated helix stability in AMBER99
  $\rightarrow$modified potential AMBER99GS

- MD **simulation** of α -helix folding kinetics by Pande also suggested the need for modified (φ,ψ) potenial
  $\rightarrow$modified potential AMBER99P

- Hummer proposed that most **popular force fields over-stabilize** the α–helix structure in short Ala-based peptides [Best et al. *Biophys.J.* **95:**L07 (2008)]

- Based on **NMR measurements of J** couplings in Ala_n
  [Graf et al., *J.Am.Chem. Soc.**129**:1179* (2007)]

New experimental data on ac-Ala_5-NH_2

- CD of Ac-Ala_5-NH_2 over 266-363 K

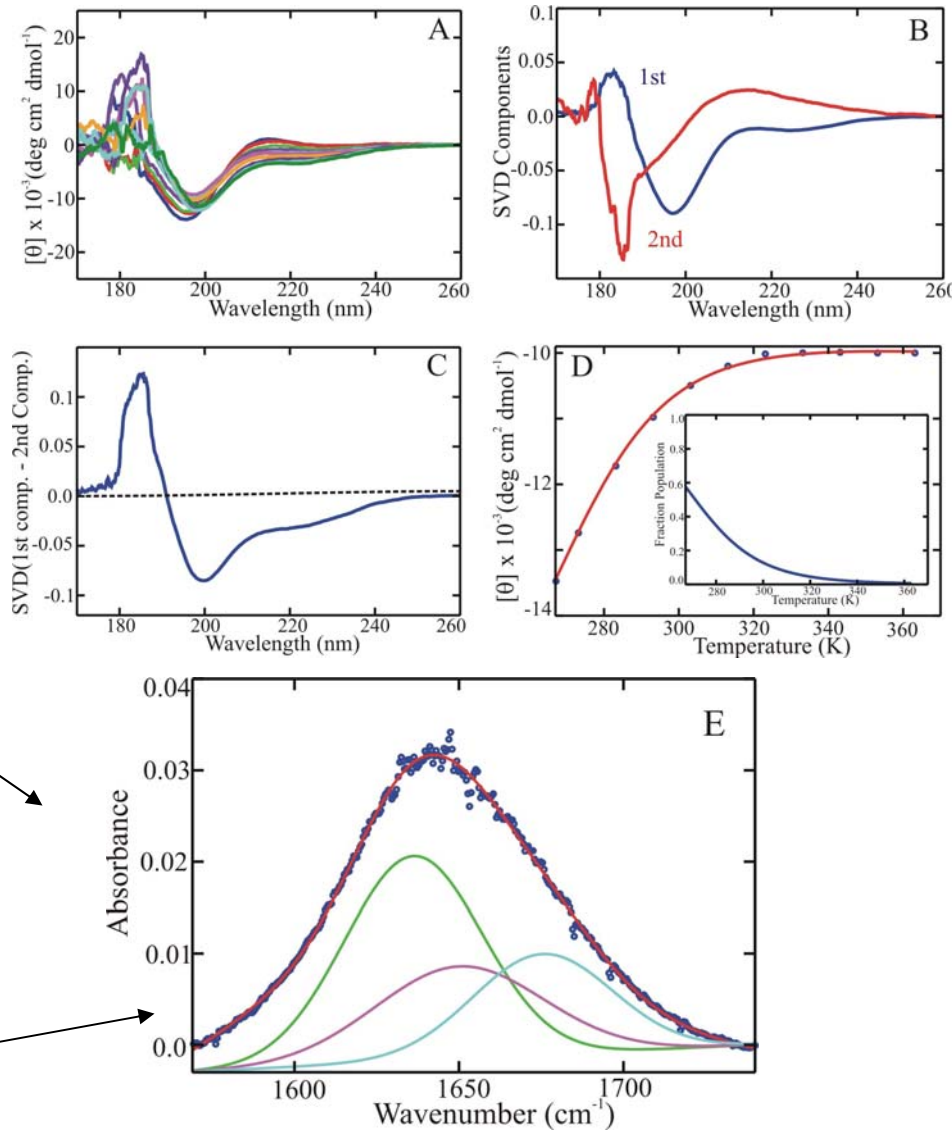
→ melting transition with T_m = 271 K ΔH = 9.5 kcal/mol

→ **13 ± 2** % helix @300 K

- FTIR measurement of amide I peak:

→ **26 ± 5** % helix @293 K

- New experimental data support population of α-helix @ low temperature



[Hegefeld, DeLeon, Kuczera & Jas – submitted]

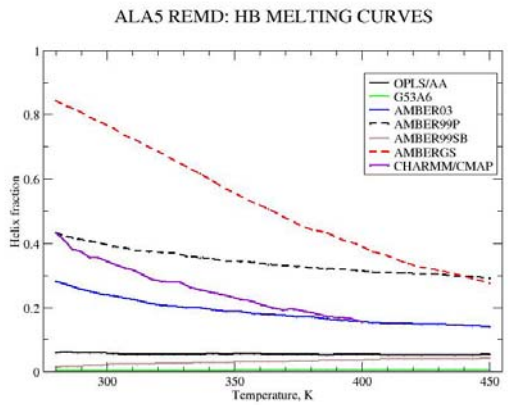
Green : β

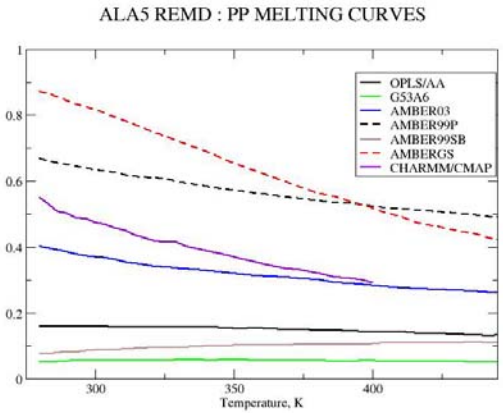
Magenta: α

Cyan: turn

Folding of Ac-Ala_5-NH_2 : REMD

- G43A1, G53A6 and AMBER99SB **underestimate** helicity
- OPLS/AA & AMBER03 closest to **new data** @ room T
- AMBER99P, AMBERGS, CHARMM22/CMAP **over-stabilize** helix
- REMD: melting **not modeled well** by most of the studied potentials
- Deviations from experiment
  ≈2-3 kcal/mol energy @300 K
  for all studied force fields



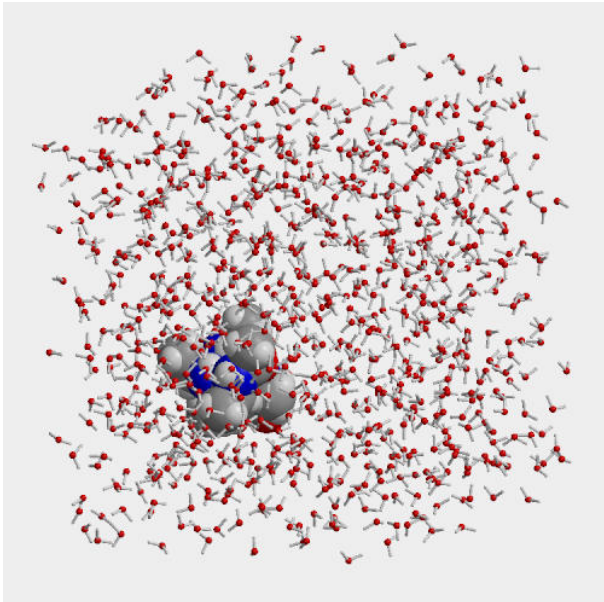


REMD simulations: 32 replicas, 280-450 K, 30 Å cubic box with ca. 1000 waters, 100 ns NPT trajectory with GROMACS, for all except CHARMM potential
CHARMM REMD: 40 ns in 37 Å bcc cell.

Ac-Ala5-NH_2 MD

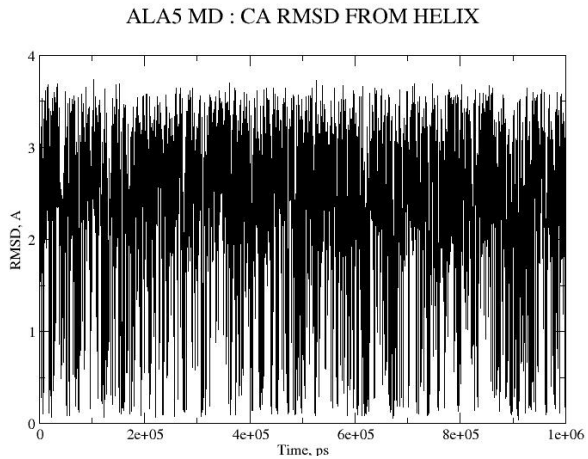
MD: 1,000 ns NPT MD at 1 atm, 300 K with GROMACS
several popular force fields, ca. 1000 waters

400 ns NPT MD at 1 atm and 300 K with CHARMM/CMAP



Folding of Ac-Ala_5-NH_2 : kinetics from MD

	τ_{fold} ns	τ_{unf} ns	τ_r ns	τ_{nuc} ns	% α HB	% α PP
OPLS/AA	7.2	0.6	0.6	0.1	5	9
CHARMM	6.1	5.5	2.9	0.2	37	40
G43A1	12.0	0.4	0.3	0.07	2	8
G53A6	170.	0.25	0.25	0.02	0.4	4
AMBER03	3.9	2.5	1.5	0.2	23	24
AMBER99P	0.2	16.2	0.3	0.04	39	42
AMBER99SB	4.4	0.5	0.5	0.1	2	4
AMBERGS	1.6	9.8	2.0	0.3	71	60

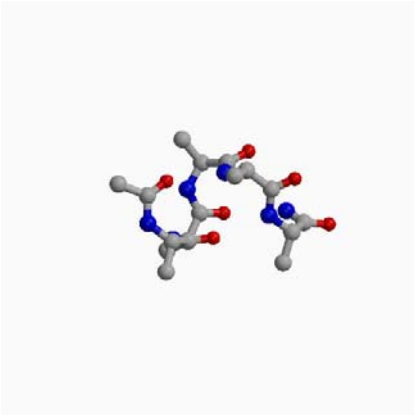


1. Predicted kinetic and equilibrium parameters span 2-3 orders of magnitude; helicities agree with exp. data
2. Helix content tends to be lower and kinetics faster compared to WH5 - consistent with W/H effects.

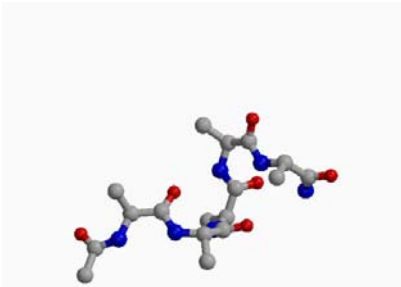
MD: 1,000 ns NPT MD at 1 atm, 300 K with GROMACS

Sample OPLS/AA results

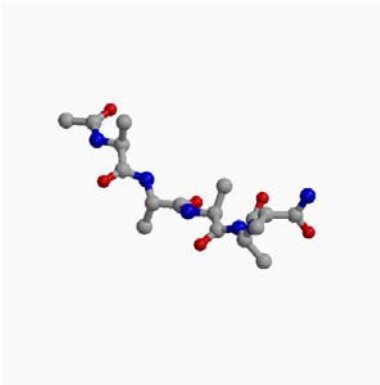
Folding of Ac-Ala_5-NH_2 : structures



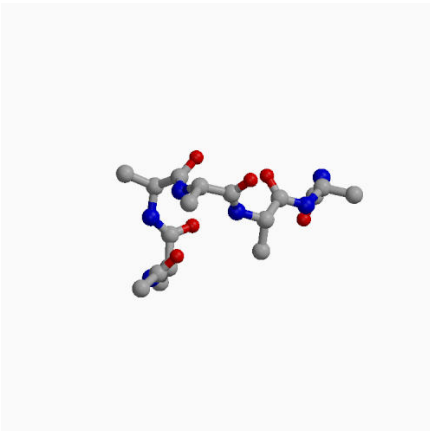
111



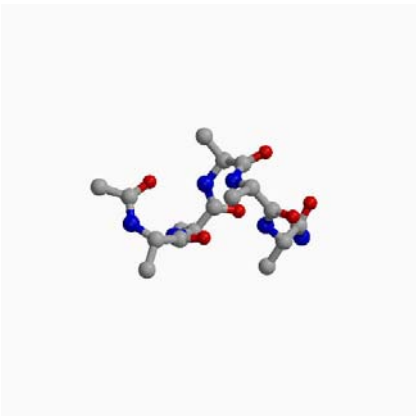
001



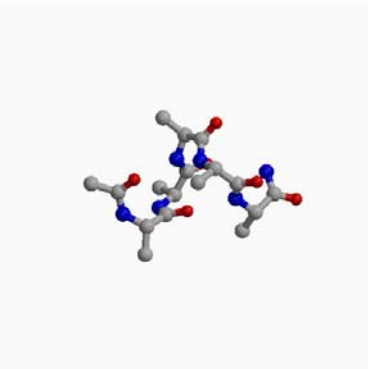
000



100

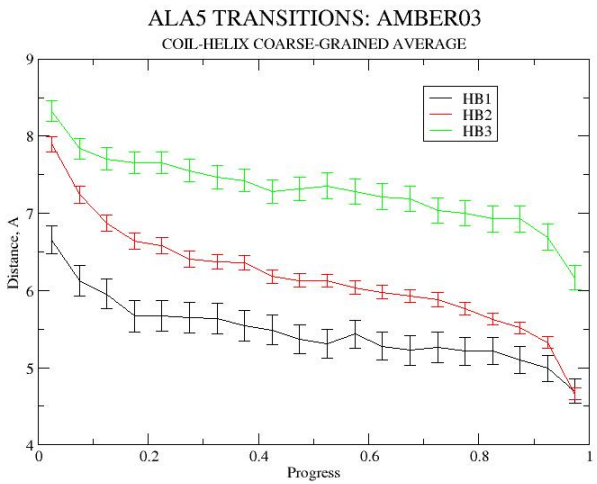


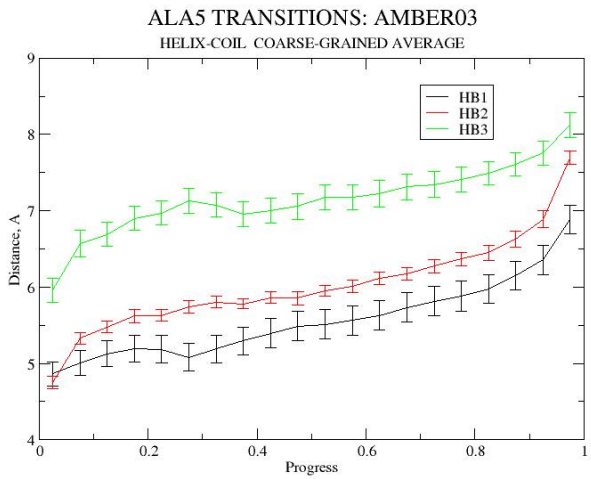
011



010

Folding of Ac-Ala_5-NH_2 : pathways





Pathway results - analogous to WH5:
- FF dependent details
- most FF predict initiation at N-terminus

Conclusions

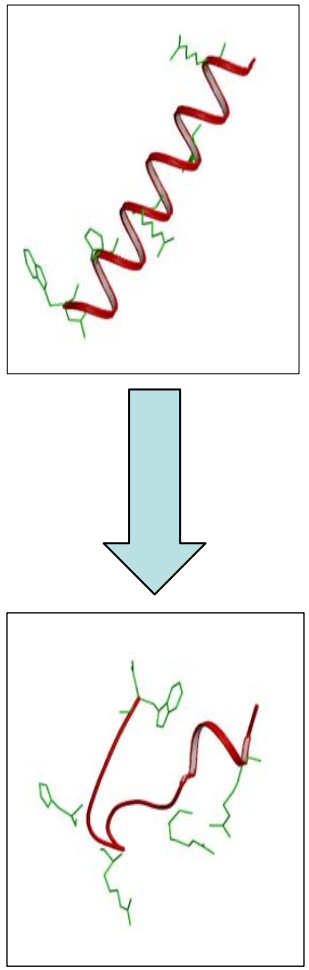
- Helix content for most popular models is in good agreement with **new experimental data**
- Calculated folding, unfolding and nucleation rates tend to be **faster than those for WH5**
- Most ff predict that helical hydrogen bond formation is **cooperative**
- Helix-coil transition paths vary with model; most studied models predict a **zipper-like mechanism**, with unfolding initiated at C-terminus and folding initiated at N-terminus.
- We have achieved full sampling of conformations and dynamics for modest size systems; results are now primarily **limited by force field accuracy**
- **More and better experimental data** are also needed to calibrate molecular models

Acknowledgments:
Experiments: Gouri Jas, Baylor University.

The computer time and technical assistance from the Academic and Research Computing Services at Baylor University are gratefully acknowledged.

Funding: Baylor University and Big XII Fellowship from University of Kansas

Future: WH21



Significantly more complex system

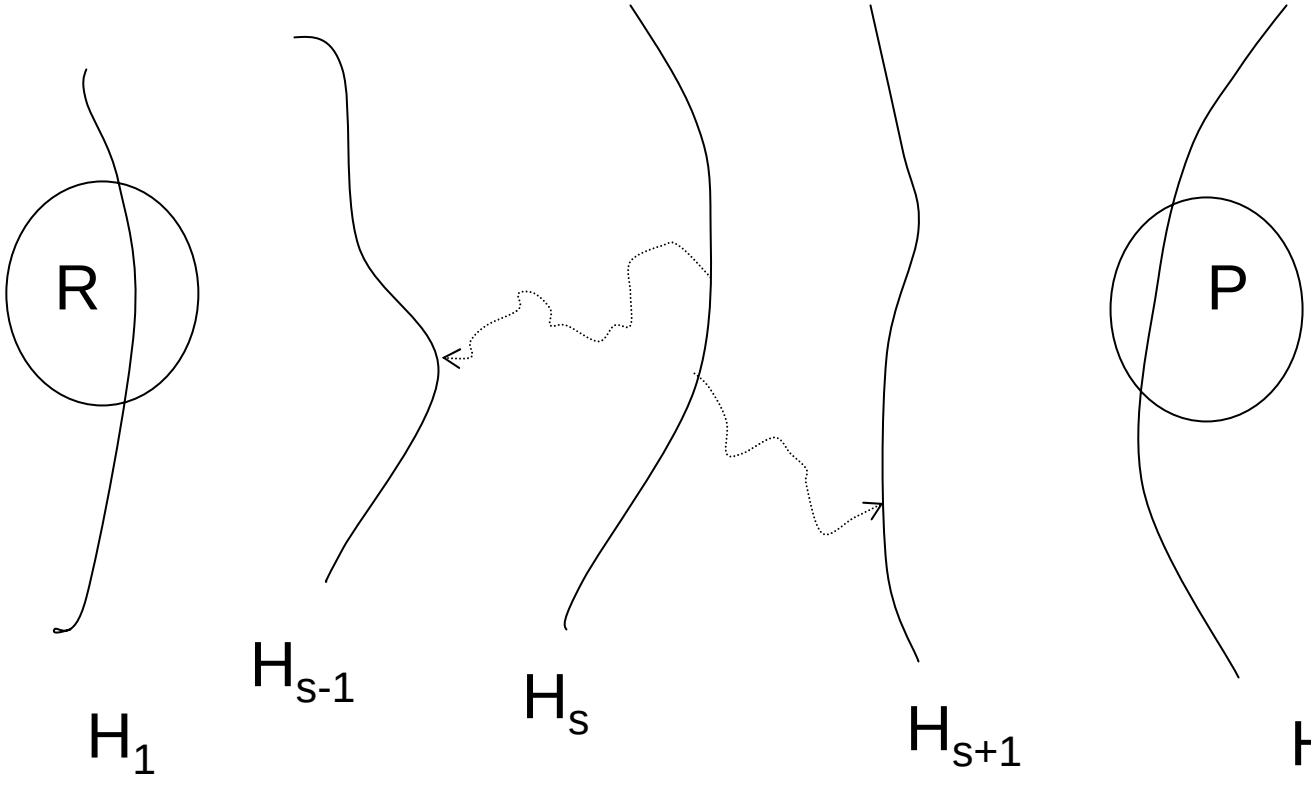
- 21 residues
- 19 hydrogen bonds
- 50 % helix and τ = 300 ns at 300 K

Sequence: 21 aa

Ac-Trp-Ala-Ala-Ala-His^+-Ala-Ala-Ala-Arg-Ala-Ala-Ala-Ala-Arg-Ala-Ala-Ala-Ala-Arg-Ala-Ala-NH_2

Milestoning: principles

West, Elber & Shalloway,
J.Chem.Phys. **126**:145104 (2007)



Partition configuration space x with hyperplanes H_s, s=1,…,M

Determine local kinetics: $\langle \tau_s^+ \rangle$ s $\rightarrow$ s+1 and $\langle \tau_s^- \rangle$ s $\rightarrow$ s-1

Recover global kinetics R $\rightarrow$ P

Milestoning: from local to global kinetics

In each hyperplane s, s=1,…,M

-Generate N trajectories (e.g. N=100)

-Record forward and backward termination:

N_s^+ , $\langle \tau_s^+ \rangle$ and N_s^- , $\langle \tau_s^- \rangle$

Global solution:

-Build kinetic matrix = asymmetric random walk

-Obtain stationary state solution with boundary conditions at s=1 and s=M

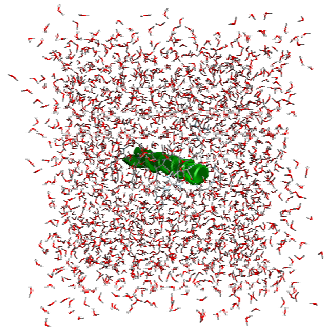
Result:

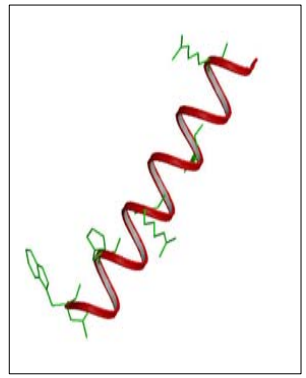
-Global forward and backward rate for whole process

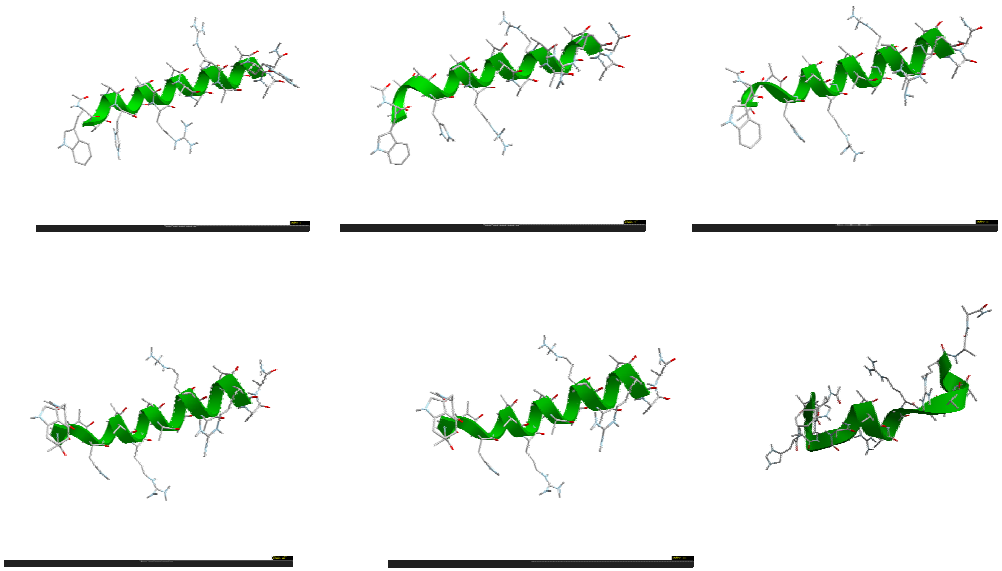
-P(s) – free energy profile or PMF

WH21 : Helix unfolding kinetics

130 milestones
13,000 trajectories
≈1μs total simulation time



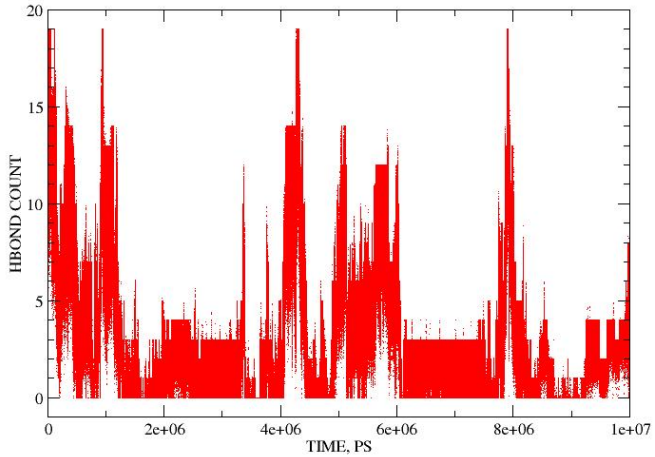


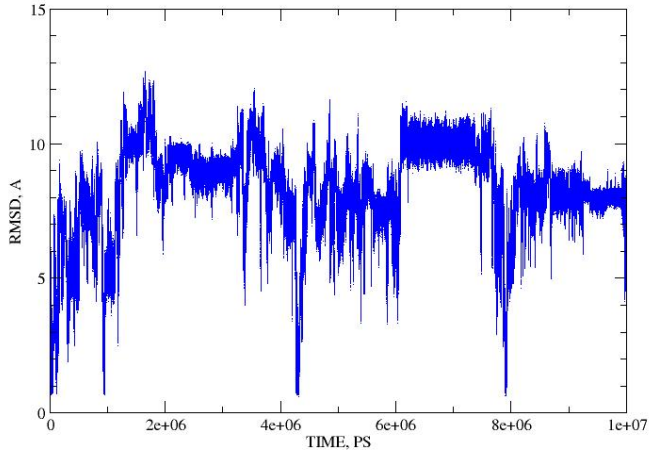


WH21 Milestoning	Mean first passage time	
	Unfolding	Elementary step
Path 1	280 ns	455 ps
Path 2	7 μs	1.6 ns
Path 3	86 μs	8.9 ns

Kuczera, Jas & Elber, *J. Phys.Chem. A* **113**:7461-7473 (2009)

WH21 MD: 10 μs with OPLS/AA-SPC





MD production: 100 ns/day with GROMACS on 36 CPUs

Properties: %$\alpha \approx 20\%$ relaxation $\tau \approx 100$ ns

3 folding/unfolding transitions sampled

MD is slow but steady …

Milestoning requires input path …

Force fields are not parameterized on μs-length simulations …

Milestoning: P,Q and K

- Termination time distributions:
  $s \to s+1$: $K^{+}_{s}(\tau)$ and $s \to s-1$: $K^{-}_{s}(\tau)$

- $P_s(t)$ – probability that system is between milestones s-1 and s+1 at t

- $Q_s(t)$ – probability that system transitions to s at time t.

- Probability balance:

$$\int_0^\infty K_s(\tau)d\tau = 1$$

$$K_s = K_s^+ + K_s^-$$

$$K_1^- = K_M^+ = 0$$

West, Elber & Shalloway, *J.Chem.Phys.* **126**:145104 (2007)

$$P_s(t) = \int_0^t \left[1 - \int_0^{t-t'} K_s(\tau)d\tau\right] Q_s(t')dt'$$

$$Q_s(t) = \eta_s \delta(t - 0^+) + \int_0^t \left[K^-_{s+1}(t-t')Q_{s+1}(t') + K^+_{s-1}(t-t')Q_{s-1}(t')\right]dt'$$

Milestoning: solving P-Q equations

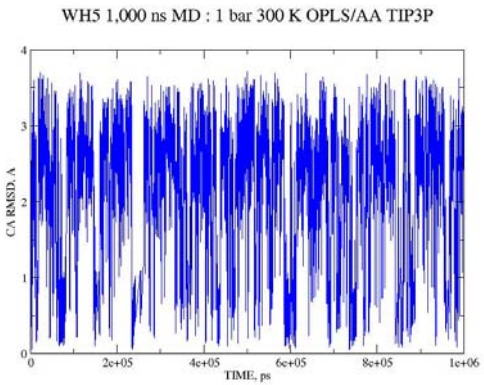
Kinetic matrix K

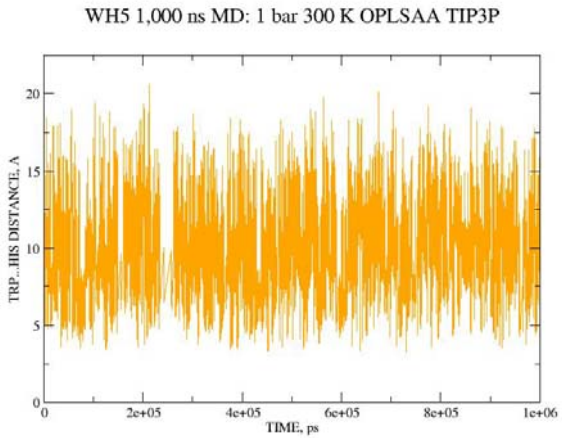
$$K_{s,s-1} = K^{+}_{s-1}(\tau)$$
$$K_{s,s} = -K^{-}_{s}(\tau) - K^{+}_{s}(\tau)$$
$$K_{s,s+1} = K^{-}_{s+1}(\tau)$$
$$K_D = diag(K)$$
$$K_d = I - K_D$$
$$\langle K \rangle \equiv \int_0^\infty K(\tau) d\tau$$

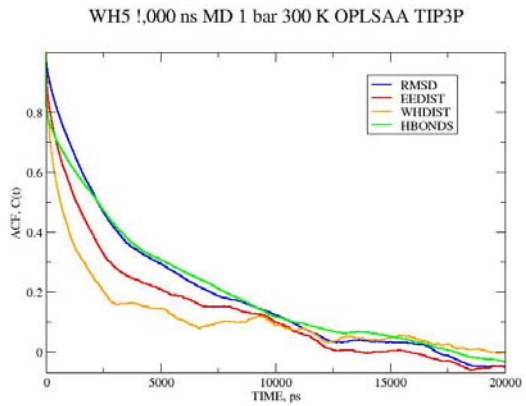
Global Solution

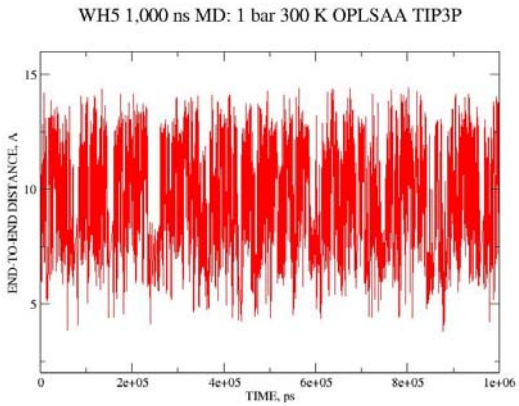
$$\langle K \rangle \hat{Q}^{eq} = \hat{0}$$
$$\hat{P}^{eq} = -\langle \tau K_D \rangle \hat{Q}^{eq}$$
$$\langle\langle \tau \rangle\rangle = \hat{\varepsilon}_i \cdot \left[I - \langle K_d \rangle \right]^{-1} \langle \tau K_d \rangle \left[I - \langle K_d \rangle \right]^{-1} \cdot \hat{\varepsilon}_f$$

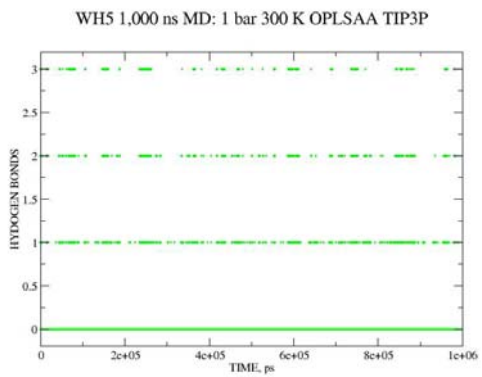
WH5 FiGURES


WH5 1,000 ns MD : 1 bar 300 K OPLS/AA TIP3P
CA RMSD, A
TIME, ps

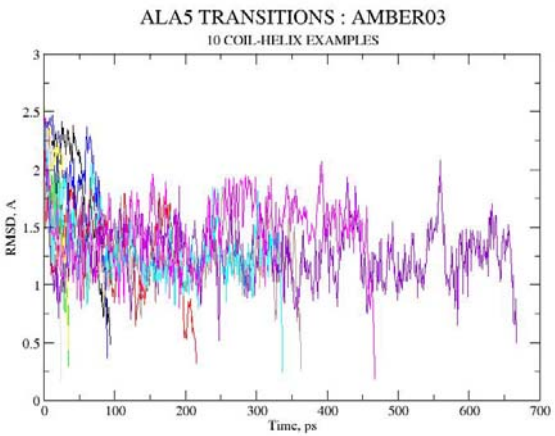

WH5 1,000 ns MD: 1 bar 300 K OPLSAA TIP3P
TRP...HIS DISTANCE, A
TIME, ps

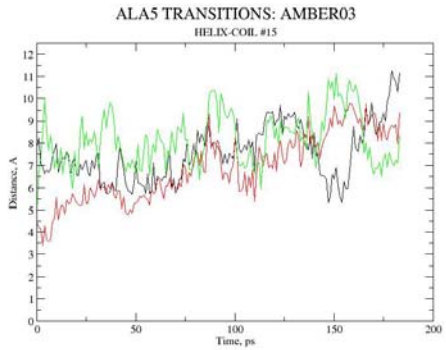

WH5 !,000 ns MD 1 bar 300 K OPLSAA TIP3P
RMSD
EEDIST
WHDIST
HBONDS
ACF, C(t)
TIME, ps

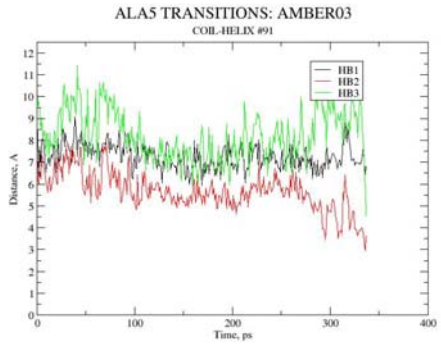

WH5 1,000 ns MD: 1 bar 300 K OPLSAA TIP3P
END-TO-END DISTANCE, A
TIME, ps

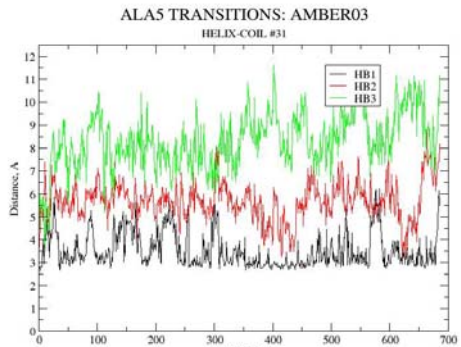

WH5 1,000 ns MD: 1 bar 300 K OPLSAA TIP3P
HYDOGEN BONDS
TIME, ps

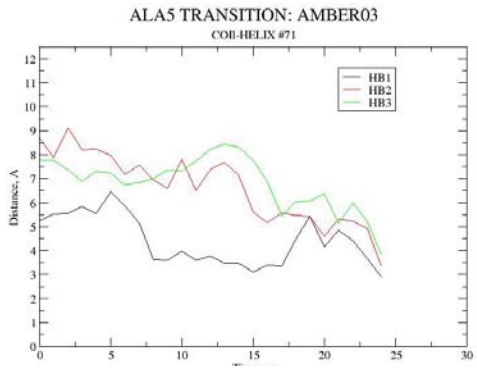
Folding of Ac-Ala_5-NH_2 : pathways

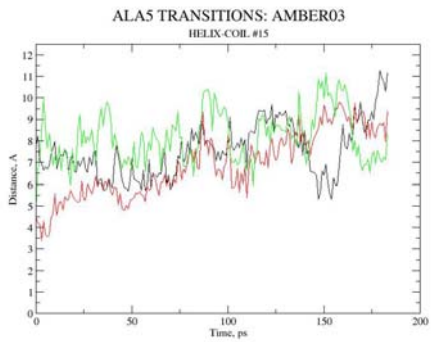


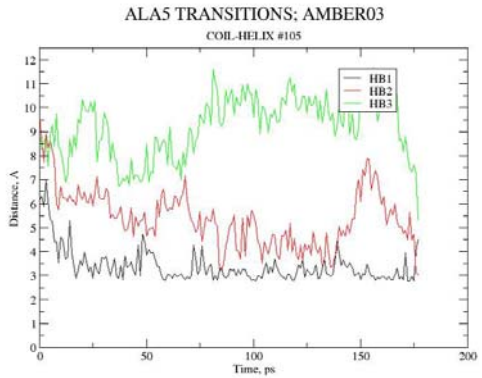




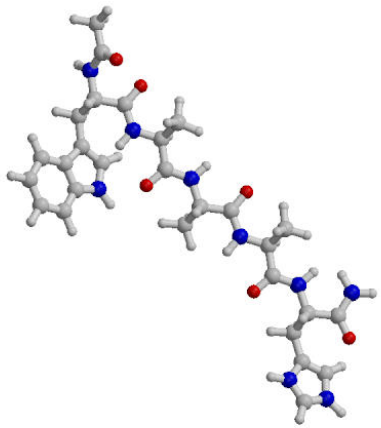


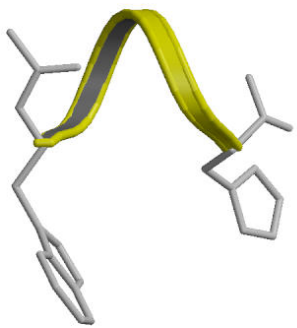


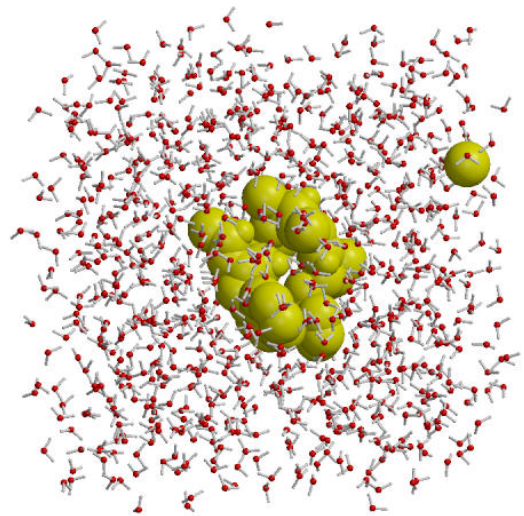




Transitions vary in
- duration time
- path details



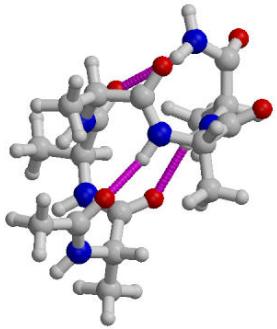


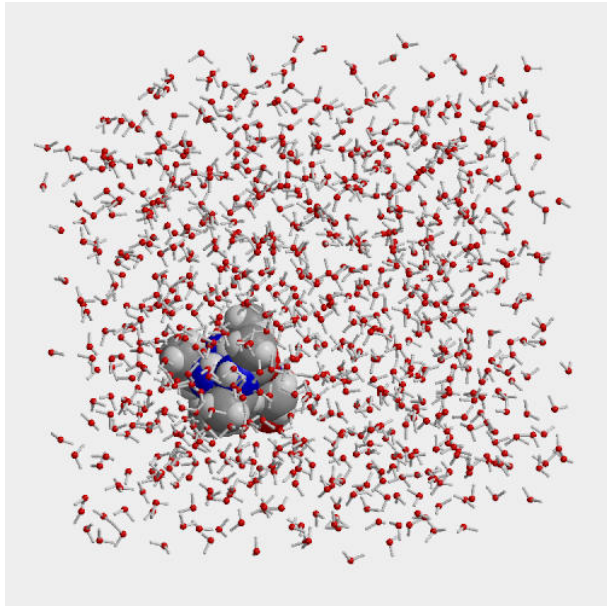


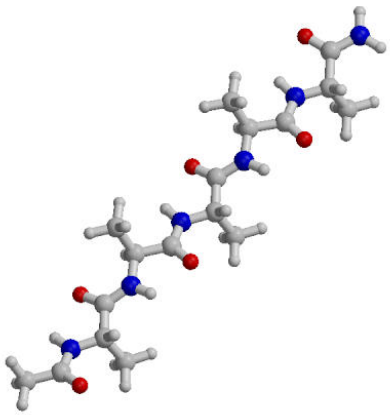
Folding of Ac-Ala_5-NH_2 : patterns

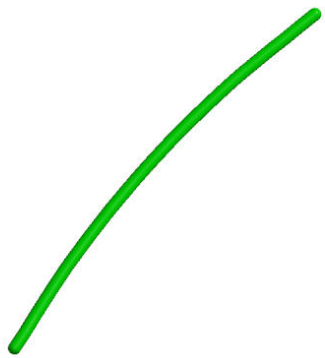
State	OPLS/AA		G43A1		AMBER03		AMBER99P		AMBER99SB		AMBERGS	
	Frac	Coop	Frac	Coop	Frac	Coop	Frac	Coop	Frac	Coop	Frac	Coop
000	0.869	1.2	0.954	1.0	0.592	1.3	0.302	1.4	0.958	1.0	0.155	6.3
100	0.042	0.8	0.021	0.7	0.098	0.6	0.151	0.8	0.018	0.8	0.040	0.7
010	0.018	0.4	0.011	0.5	0.057	0.4	0.095	0.6	0.013	0.7	0.028	0.4
110	0.013	4.2	0.009	14.	0.096	1.5	0.160	1.1	0.003	7.5	0.132	0.7
001	0.032	0.7	0.003	0.6	0.029	1.4	0.082	0.9	0.005	0.6	0.015	0.3
101	0.004	1.3	0.000	0.7	0.014	0.4	0.030	0.4	0.000	1.3	0.024	0.2
011	0.010	4.2	0.001	9.5	0.037	1.2	0.065	0.9	0.002	13.	0.094	1.6
111	0.013	67.	0.001	342.	0.078	6.5	0.112	2.0	0.001	310.	0.512	1.5

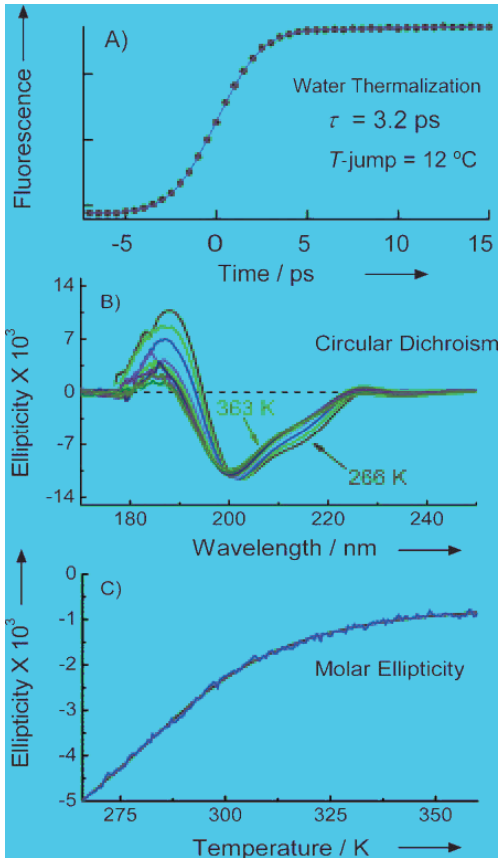
- Most FF : 000 dominant, very little 111, positive h-bond cooperativity
- Populated intermediates: involve h-bonds #1 and #2
- Unusual: AMBERGS










A)
Water Thermalization
τ = 3.2 ps
T-jump = 12 °C
Fluorescence
-5
0
5
10
15
Time / ps
B)
Circular Dichroism
363 K
266 K
Ellipticity X 10^3
14
7
0
-7
-14
180
200
220
240
Wavelength / nm
C)
Molar Ellipticity
Ellipticity X 10^3
0
-1
-2
-3
-4
-5
275
300
325
350
Temperature / K

WH5 figs