

# Constant pH molecular dynamics

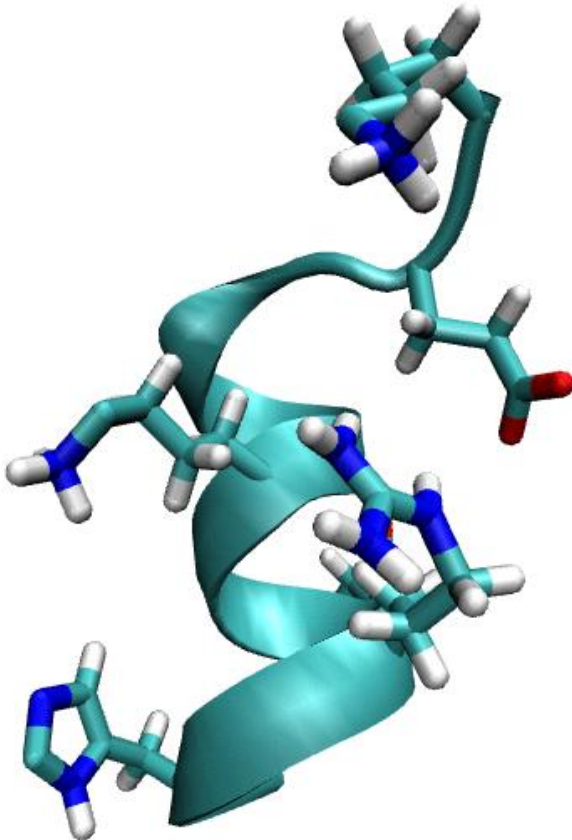
Jana K. Shen

Department of Chemistry and Biochemistry

University of Oklahoma

MMTSB/CTBP Workshop 2009

# Traditional molecular dynamics



Protonation states of titratable residues

- assigned by comparing model  $pK_a$ 's with the simulation pH
- kept fixed during simulation
- not realistic because  $pK_a$ 's in the protein can be different than model  $pK_a$ 's
- coexistence of both protonated and deprotonated states is neglected

## Example

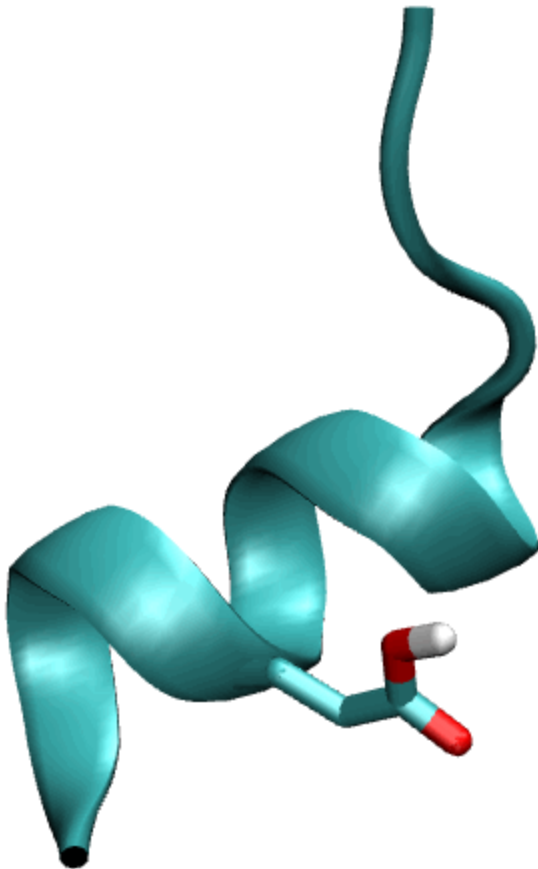
Protonation states in C-peptide at pH 7:

Glu: deprotonated

His: singly deprotonated on  $N\delta$  or  $N\epsilon$

Lys and Arg: protonated

# Constant pH molecular dynamics (PHMD)



Example  
CPHMD of C-peptide  
(only Glu-9 is titrating)

## Basic idea

- Let solution pH dictate protonation states
- Combine sampling of conformational and protonation states

## Discrete PHMD

- Periodically stop MD and perform Monte-Carlo sampling of protonation states

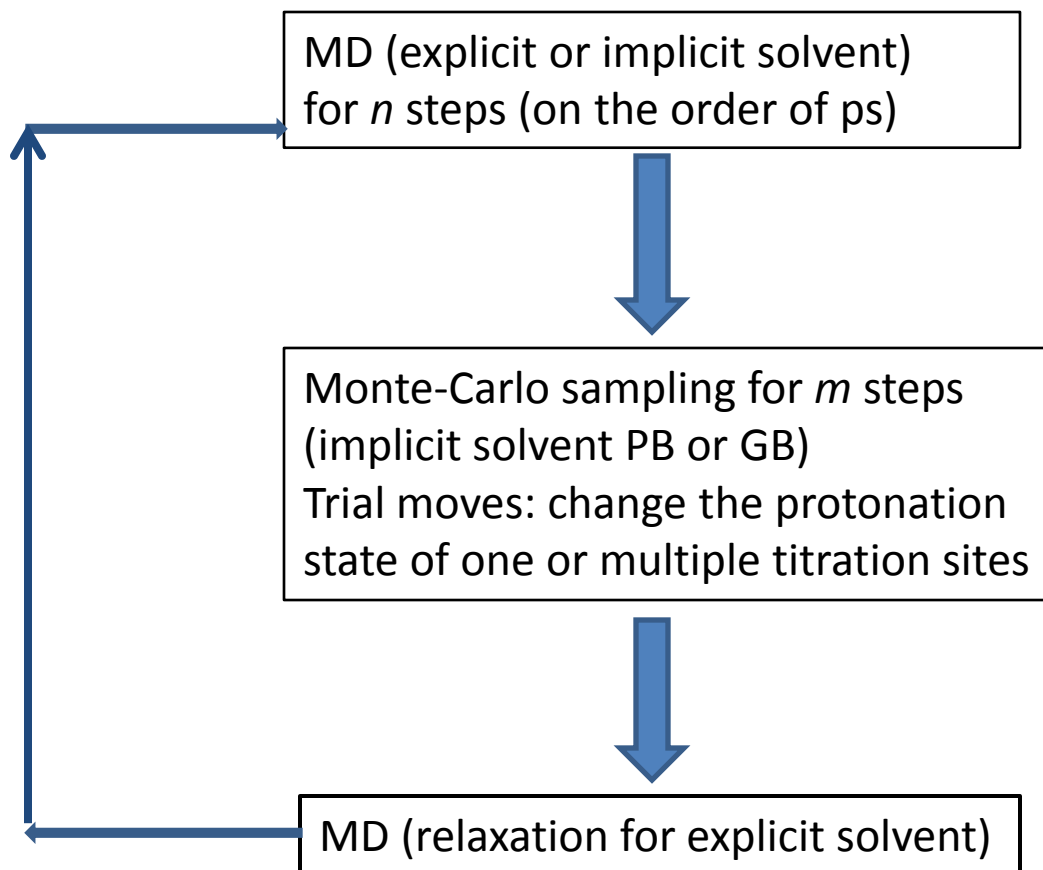
## Continuous PHMD

- Molecular dynamics sampling of spatial and continuous titration coordinates

## Assumptions

- Source of protons not considered
- Kinetics of protonation not considered

# Discrete PHMD: combine MD & MC sampling



Baptista et al, J Chem Phys 2002.

Mongan et al, J Comput Chem 2004.

# Pros and cons of discrete PHMD

Conceptually simple

Easy to implement

Flexibility in the treatment of solvent

- explicit solvent, PB or GB

Protonation states are physical

- proton charge turned on/off

Discontinuous energy and force

- may lead to instability due to large energy change, e.g. charging multiple sites
- However, this kind of moves are most likely rejected.

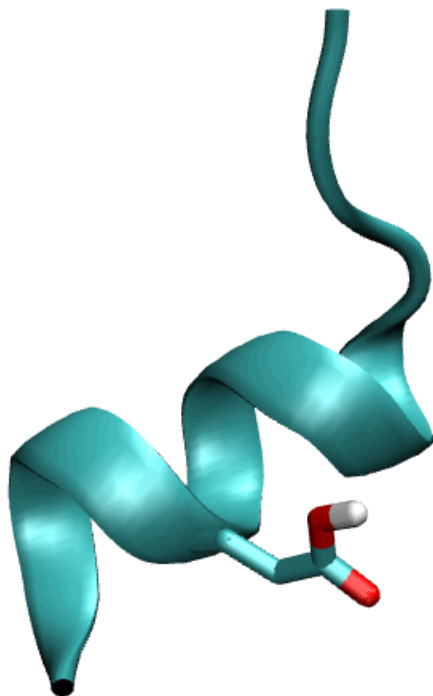
Convergence for titration of multiple-sites may be slow

- Because most likely titration occurs one at a time (see reasons above)

Getting stuck in one prot. state

- MC problem: moves associated with large energy changes are seldom accepted
- This problem may become severe when protonation event is coupled to a large conformational transition

# Continuous PHMD (CPHMD)



- Titration coordinates

$$\lambda_i = \sin^2 \theta_i$$

- Protonation states

$$\lambda_i = \begin{cases} \approx 1 & \text{unprotonated} \\ \approx 0 & \text{protonated} \end{cases}$$

- Propagation of titration coordinates

$$f_i = m_i d^2 \theta_i / dt^2$$

$$H^{\text{extend}}(\{r_j\}, \{\theta_i\}) = H^{\text{hybrid}}(\{r_j\}, \{\theta_i\}) + \sum_i \frac{m_i}{2} \dot{\theta}_i^2 + U^*(\{\theta_i\})$$

# Coupling of titration and conformational dynamics

- Hybrid Hamiltonian

$$H^{\text{hybrid}} = U^{\text{Coulomb}}(\{r_j\}, \{\theta_i\}) + U^{\text{vdW}}(\{r_j\}, \{\theta_i\}) + U^{\text{GB}}(\{r_j\}, \{\theta_i\})$$

- Linear interpolation for charge states

$$q_{i,\alpha} = \lambda_i q_{i,\alpha}^{\text{unprot}} + (1 - \lambda_i) q_{i,\alpha}^{\text{prot}}$$

- Linear interpolation for van der Waals interactions

$$\tilde{U}^{\text{vdW}} = \begin{cases} (1 - \lambda_i) U^{\text{vdW}}(i, j) & i = \text{titrating residue} \\ (1 - \lambda_i)(1 - \lambda_j) U^{\text{vdW}}(i, j) & i, j \text{ both titrating residues} \end{cases}$$

# Biasing potentials

- Biasing potential

$$U^{\text{bias}} = -U^{\text{model}} + U^{\text{pH}} - U^{\text{barrier}}$$

- PMF along the titrating coordinate for the model compound

$$U^{\text{model}} = A(\lambda_i - B)^2 \quad \text{Quadratic form due to the pair-wise GB energy form}$$

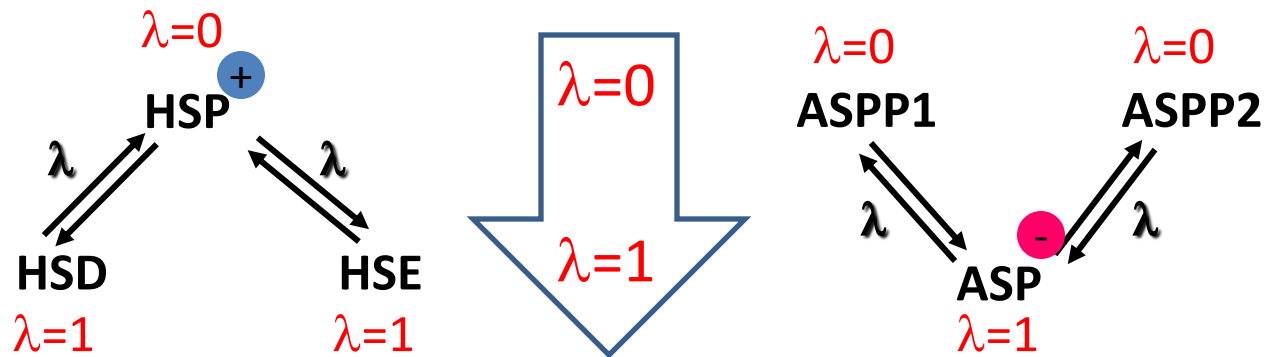
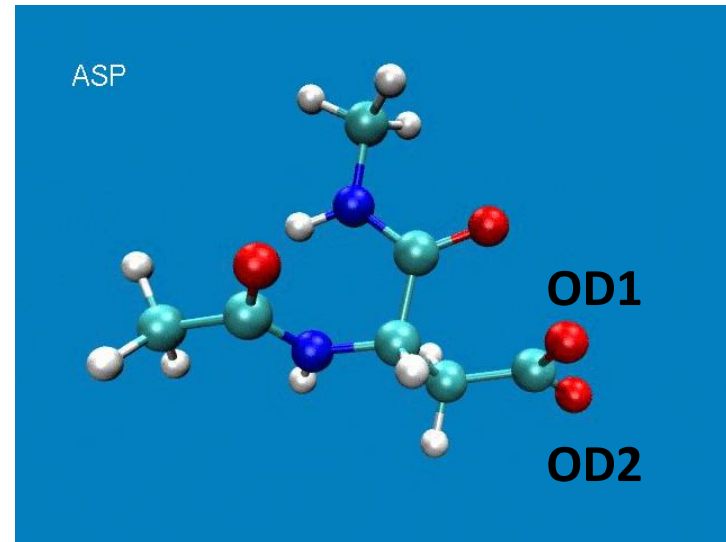
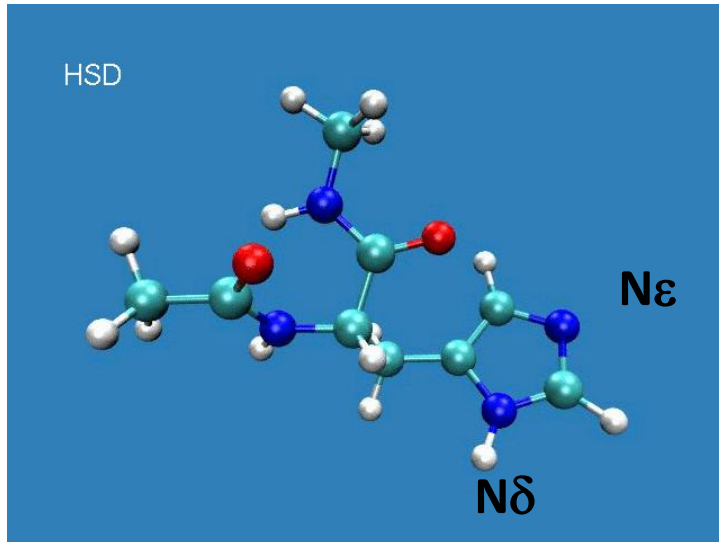
- pH dependence of deprotonation free energy

$$U^{\text{pH}} = 2.3RT\lambda_i(\text{p}K_a^{\text{model}}(i) - \text{pH})$$

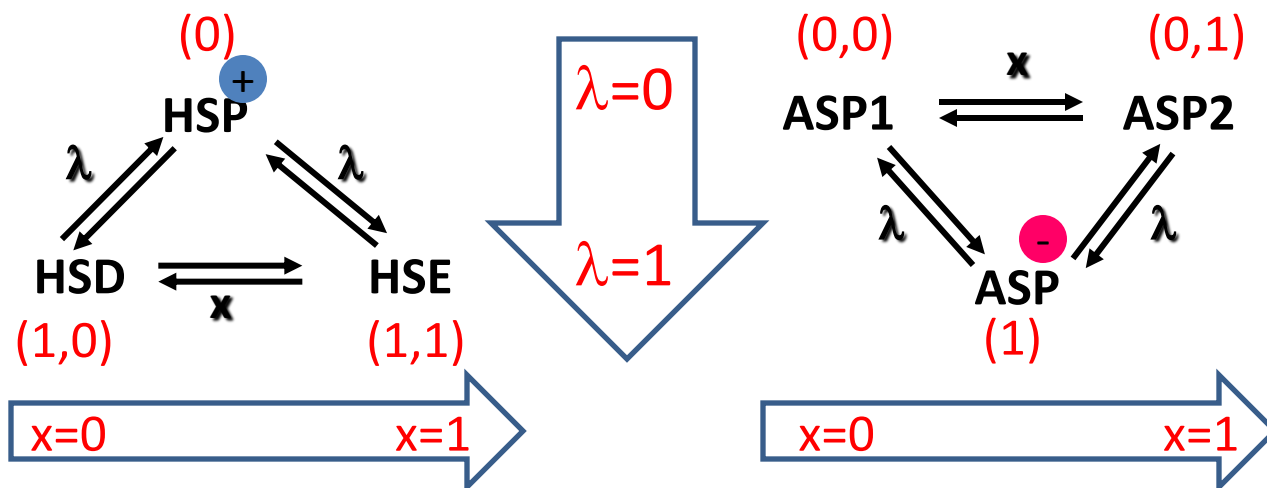
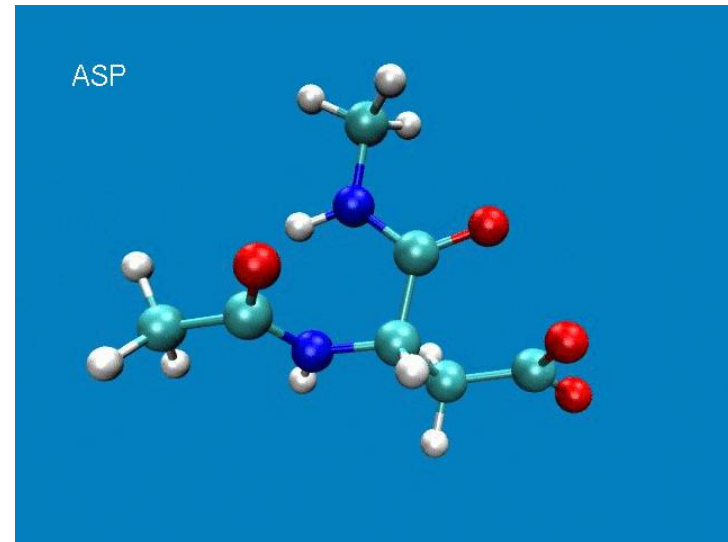
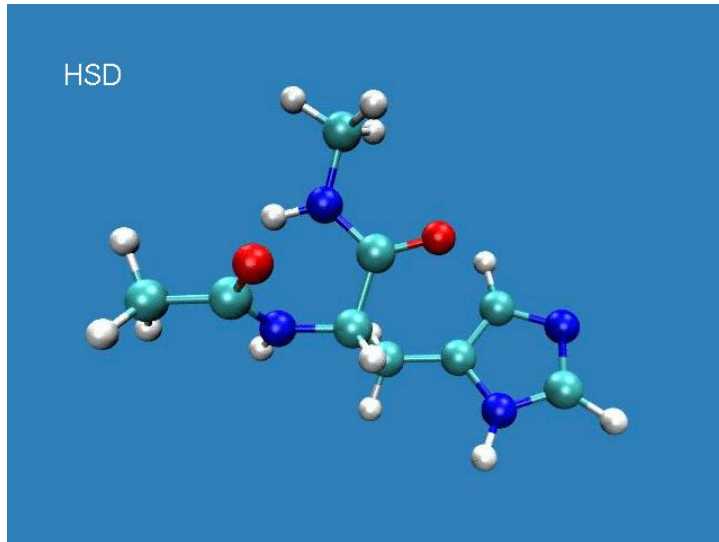
- Harmonic potential for suppressing the population of mixed states,  $0.1 < \lambda_i < 0.9$

$$U^{\text{barrier}} = 4\beta_i\left(\lambda_i - \frac{1}{2}\right)^2$$

# CPHMD: double-site titrations



# CPHMD: double-site titrations



# CPHMD: two-dimensional $\lambda$ dynamics

- Recast all interpolation formulae and biasing potential terms.
- The model PMF for titration is a polynomial function quadratic in both  $\lambda$  and  $x$ , and is subjected to a set of boundary conditions

$$U^{\text{model}}(\lambda_i, x_i) = a_0 \lambda_i^2 x_i^2 + a_1 \lambda_i^2 x_i + a_2 \lambda_i x_i^2 + a_3 \lambda_i x_i \\ + a_4 \lambda_i^2 + a_5 x_i^2 + a_6 \lambda_i + a_7 x_i$$

- Boundary conditions for HIS titration

$$U^{\text{model}}(\lambda_i, 1) = A_1 (\lambda_i - B_1)^2$$

$$U^{\text{model}}(\lambda_i, 0) = A_0 (\lambda_i - B_0)^2$$

$$U^{\text{model}}(1, x_i) = A_{10} (x_i - B_{10})^2$$

$$U^{\text{model}}(0, x_i) = 0$$

# CPHMD: two-dimensional $\lambda$ dynamics

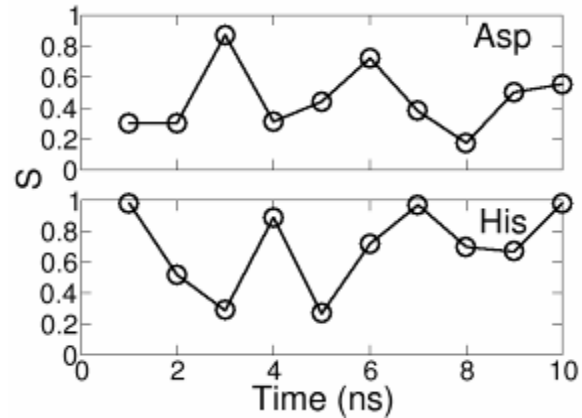
- Boundary conditions for titration of carboxyl groups

$$U^{\text{mod}}(\lambda_i, x_i) = \begin{cases} A_1(\lambda_i - B_1)^2, & x_i = 1 \\ A_0(\lambda_i - B_0)^2, & x_i = 0 \\ 0, & \lambda_i = 1 \\ A_{10}(x_i - B_{10})^2, & \lambda_i = 0 \end{cases}$$

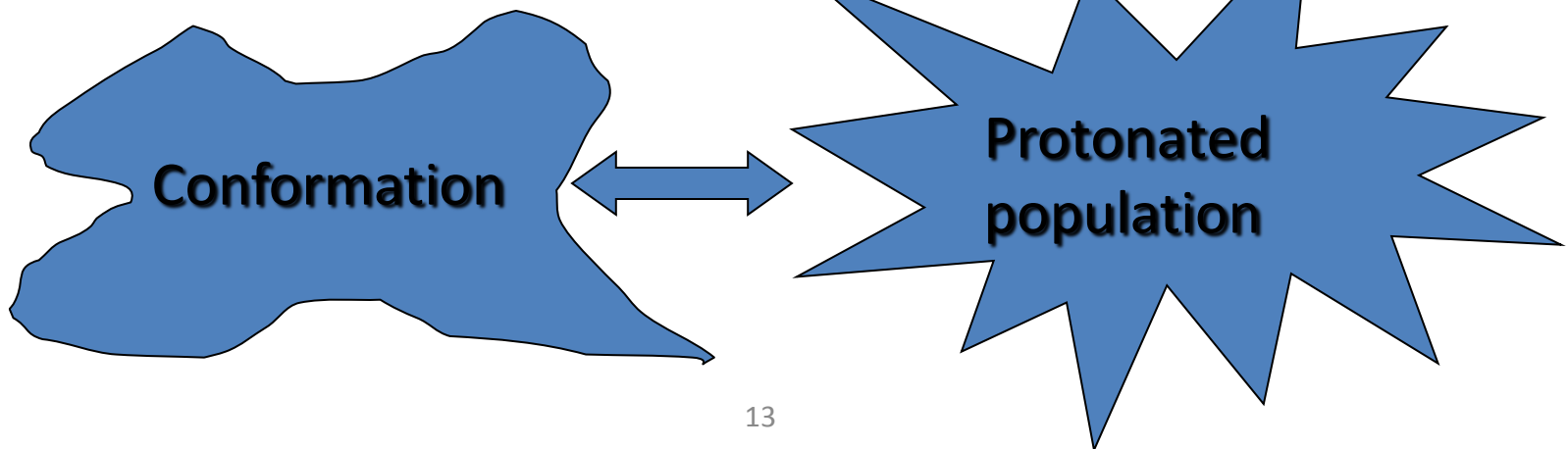
- Observing the boundary conditions, we can devise a procedure using thermodynamic integration at different  $\lambda$  and  $x$  values to find all parameters in the model PMF function (see tutorial)
- Two-dimensional  $\lambda$  dynamics is generic, e.g., it can be applied to calculate free energies of competitive binding

# CPHMD: convergence problem

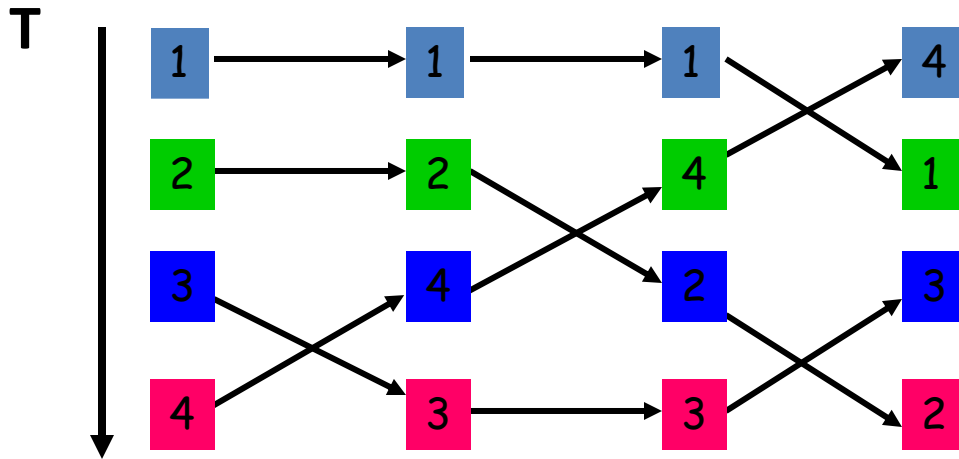
- Large random errors in sampling protonation states



- Poor fitting to the Henderson-Hasselbalch curve



# CPHMD: enhance conformational sampling



- Independent replicas running at different temperature
- Periodic swap both conformational and protonation states

⇒ Accelerate crossing of local energy barriers



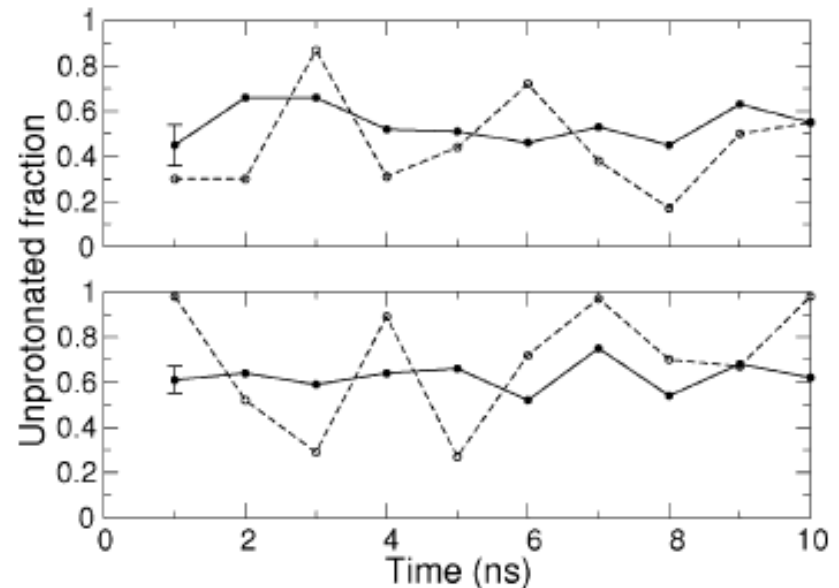
# REX-CPHMD: reduce random errors

- Better protonation state sampling with the same amount of time

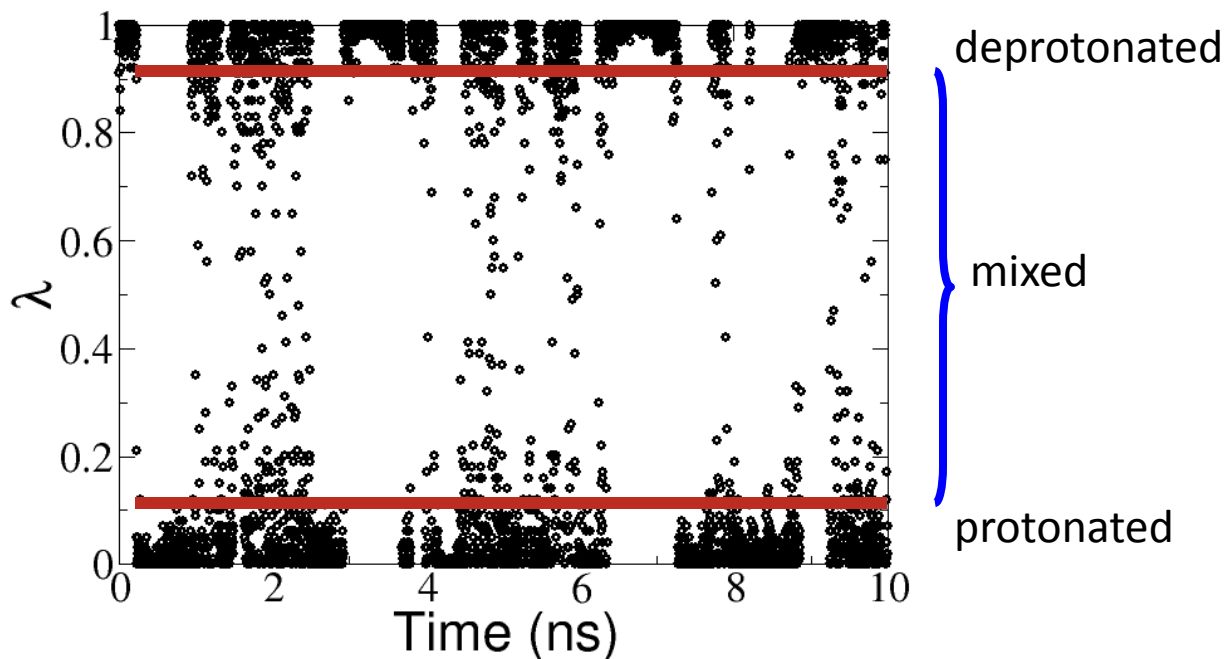
Res	RMS REX*	RMS Single T*
Asp	0.16	0.5
His	0.12	0.4

\*REX: 4 replicas, 1 ns each

\*Single temperature: 4 ns



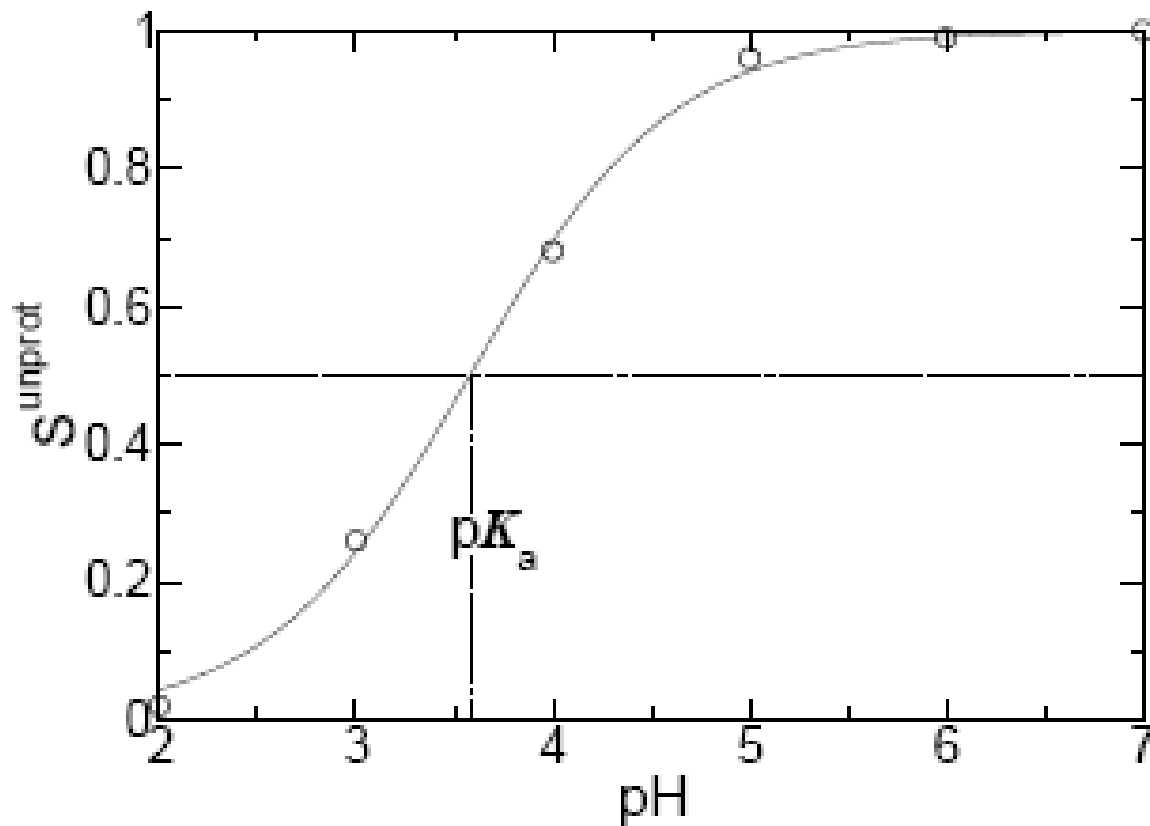
# CPHMD: how to compute $pK_a$ values



Compute  $pK_a$  by fitting **unprotonated fractions**,  $S$ , at different pH, to the generalized Henderson-Hasselbach (Hill) equation:

$$S = \frac{\rho^{unprot}}{\rho^{unprot} + \rho^{prot}} = \frac{1}{1 + 10^{n(pK_a - pH)}}$$

# CPHMD: how to compute $pK_a$ values



$$S = \frac{1}{1 + 10^{n(pK_a - pH)}}$$

# Pros and cons of CPHMD

- Continuous energy and force
- May converge faster than DPHMD for multi-site titrations
- Accuracy can be systematically improved
- Well suited for studying pH-coupled conformational dynamics
- Has been validated using both  $pK_a$  calculations and pH-dependent conformational dynamics
- Not easy to implement
- Has unphysical states (intermediate lambda values): but populations can be suppressed to be low
- Non-trivial to incorporate explicit solvent

# REX-CPHMD: accuracy of pK<sub>a</sub> predictions

Calculated and experimental pK<sub>a</sub>'s for RNase A

Residue	CPHMD		Expt	
	pK <sub>a</sub>	ΔpK <sub>a</sub>	pK <sub>a</sub>	ΔpK <sub>a</sub>
Glu-2	3.6	-0.4	2.6	-1.8
Glu-9	3.8	-0.6	-	-
His-12	5.8	-0.8	6.0	-0.6
Asp-14	3.4	-0.6	1.8	-2.2
Asp-38	3.0	-1.0	2.1	-1.9
His-48	4.9	-1.7	6.1	-0.5
Glu-49	3.3	-1.1	4.3	-0.1
Asp-53	4.0	0.0	3.7	-0.3
Asp-83	3.2	-0.7	3.3	-0.7
Glu-86	4.5	0.1	4.0	-0.4
His-105	6.4	-0.2	6.5	-0.1
Glu-111	3.4	-1.0	-	-
His-119	5.6	-1.0	6.5	-0.1
Asp-121	2.7	-1.3	3.0	-1.0
<b>RMSD</b>	0.8			

# Prediction of pK<sub>a</sub>'s of deeply buried groups

Calculated and experimental pK<sub>a</sub>'s for Staph nuclease

PDB	Protein	Residue	CPHMD	Expt	MCCE	PROPKA
-	L38D/ $\Delta$ +PHS	Asp-38	6.6	7.2	8.7	5.0
3D6C	L38E/ $\Delta$ +PHS	Glu-38	6.9	7.0	8.5	5.3
2RKS	L38K/ $\Delta$ +PHS	Lys-38	9.3	10.4	7.3	9.7
2SNM	V66K/WT	Lys-66	7.5	$\leq 6.4$	-	-
-	V66K/PHS	Lys-66	6.9	6.35	-	-
-	V66K/ $\Delta$ +PHS	Lys-66	7.0	5.8	-	-

# Prediction of pK<sub>a</sub>'s of highly flexible proteins

Calculated and experimental pKa's for alpha-lactalbumin

Residue	CPHMD	PROPKA	Expt
Glu-7	3.2	3.1	2.9
His-15	6.2	6.4	5.7
Asp-18	3.3	3.7	2.7
Glu-35	5.5	5.8	6.2
Asp-48	3.5	1.4	2.5
Asp-52	4.7	4.8	3.7
Asp-66	1.9	0.5	2.0
Asp-87	2.7	2.2	2.1
Asp-101	4.0	4.0	4.1
Asp-119	2.5	3.4	3.2
<b>RMSD</b>	0.6	0.8	

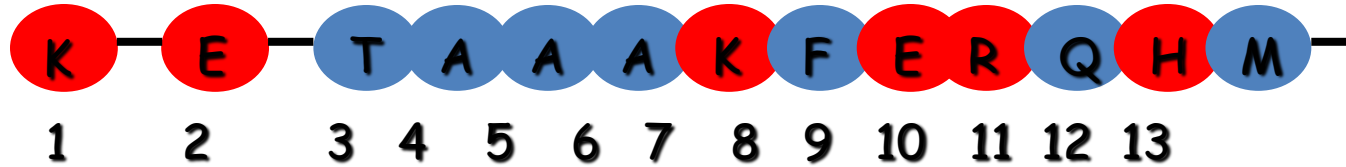
# Errors in CPHMD

- Major source of error: underlying implicit solvent model (GB)
    - Van der Waals surface used in GB underestimates Born radii
- ⇒ Too much self solvation

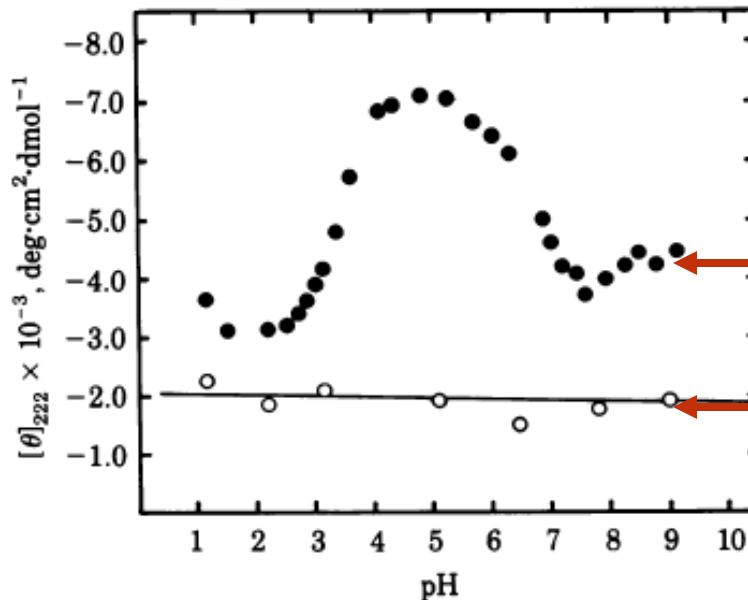
$$\Delta G = \Delta G_{self\ solvation} + \Delta G_{screening\ Coumb}$$

- Other potential errors:
  - Larger errors for lysines and argines may be due to lack of multi-site titration
  - Force field bias: seen in fragment peptides but so much in native-state simulations (small does not make it easier)
  - Polarization effects neglected

# C-peptide of ribonuclease A



- Experimentally well characterized folding behavior
- pH-dependent partial helix formation

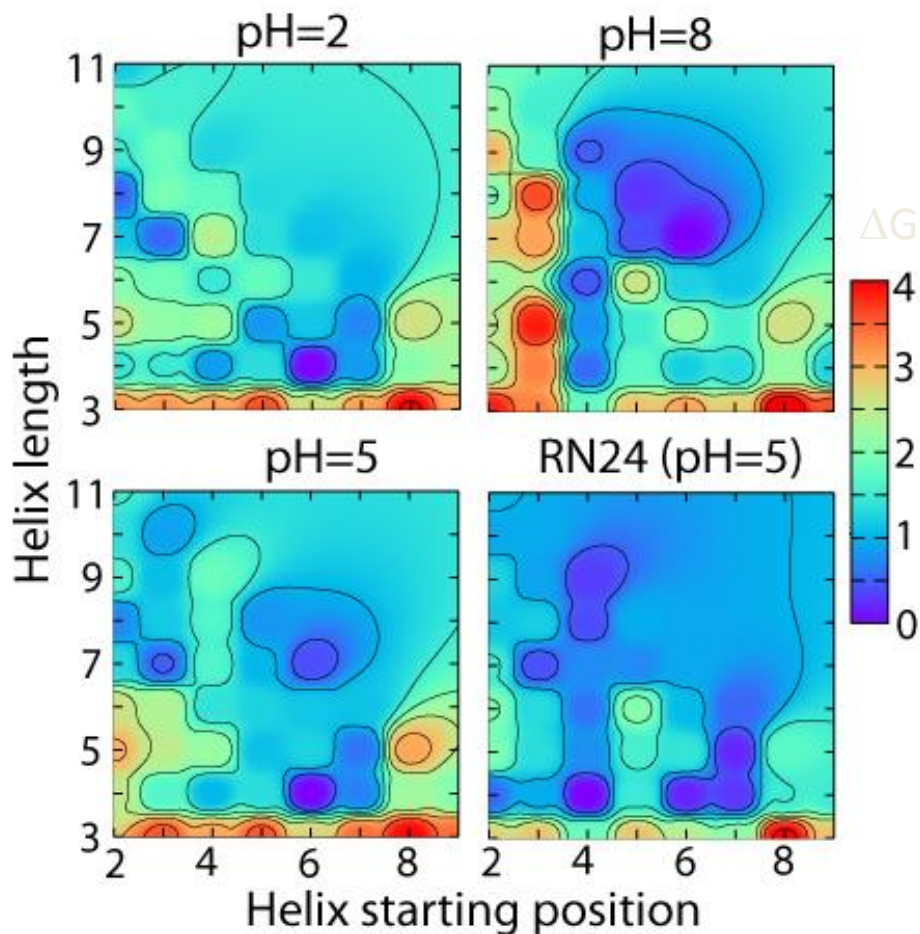
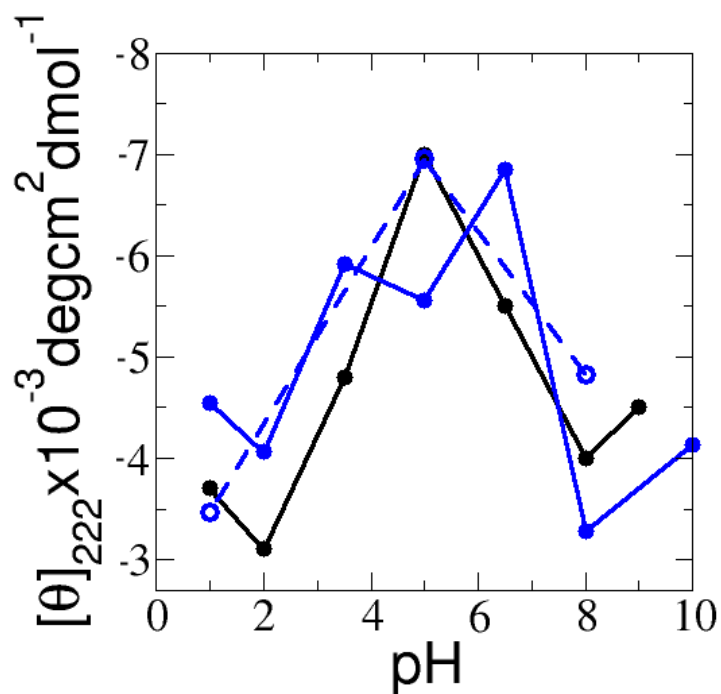


C-peptide

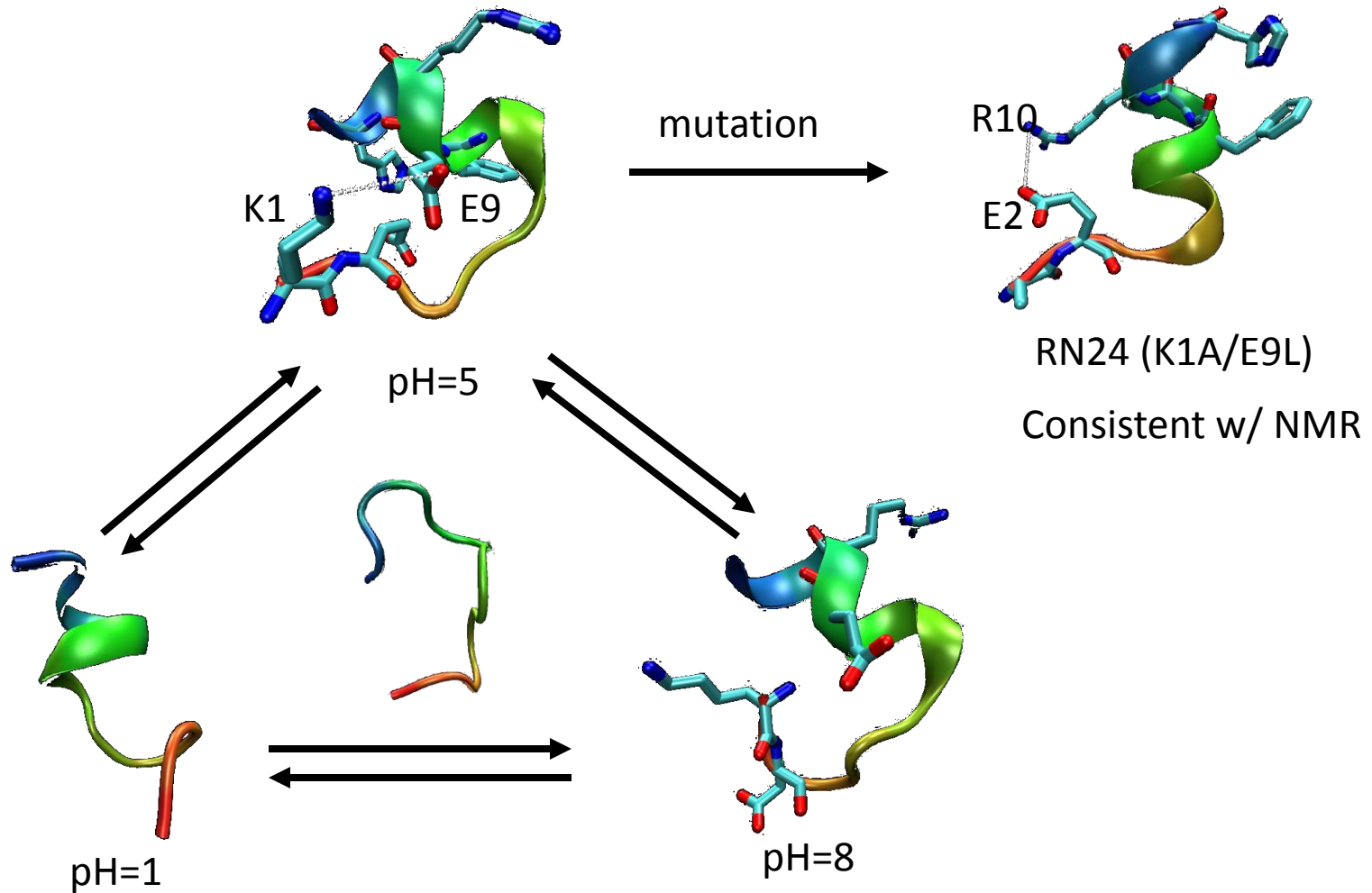
random coil

$$[\theta]_{222} \propto \text{helix content}$$

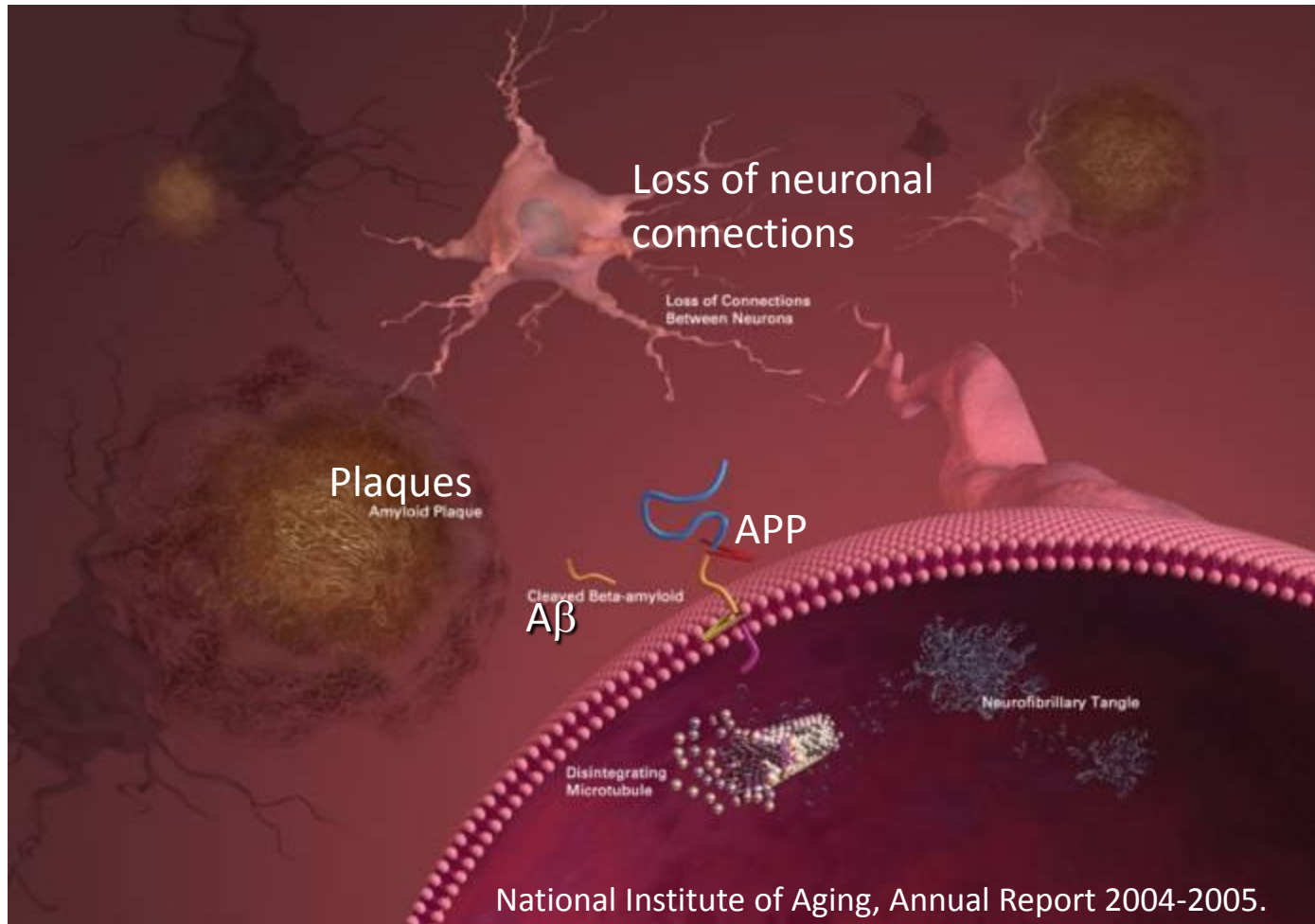
# REX-CPHMD folding simulations



# pH-dependent conformational equilibrium

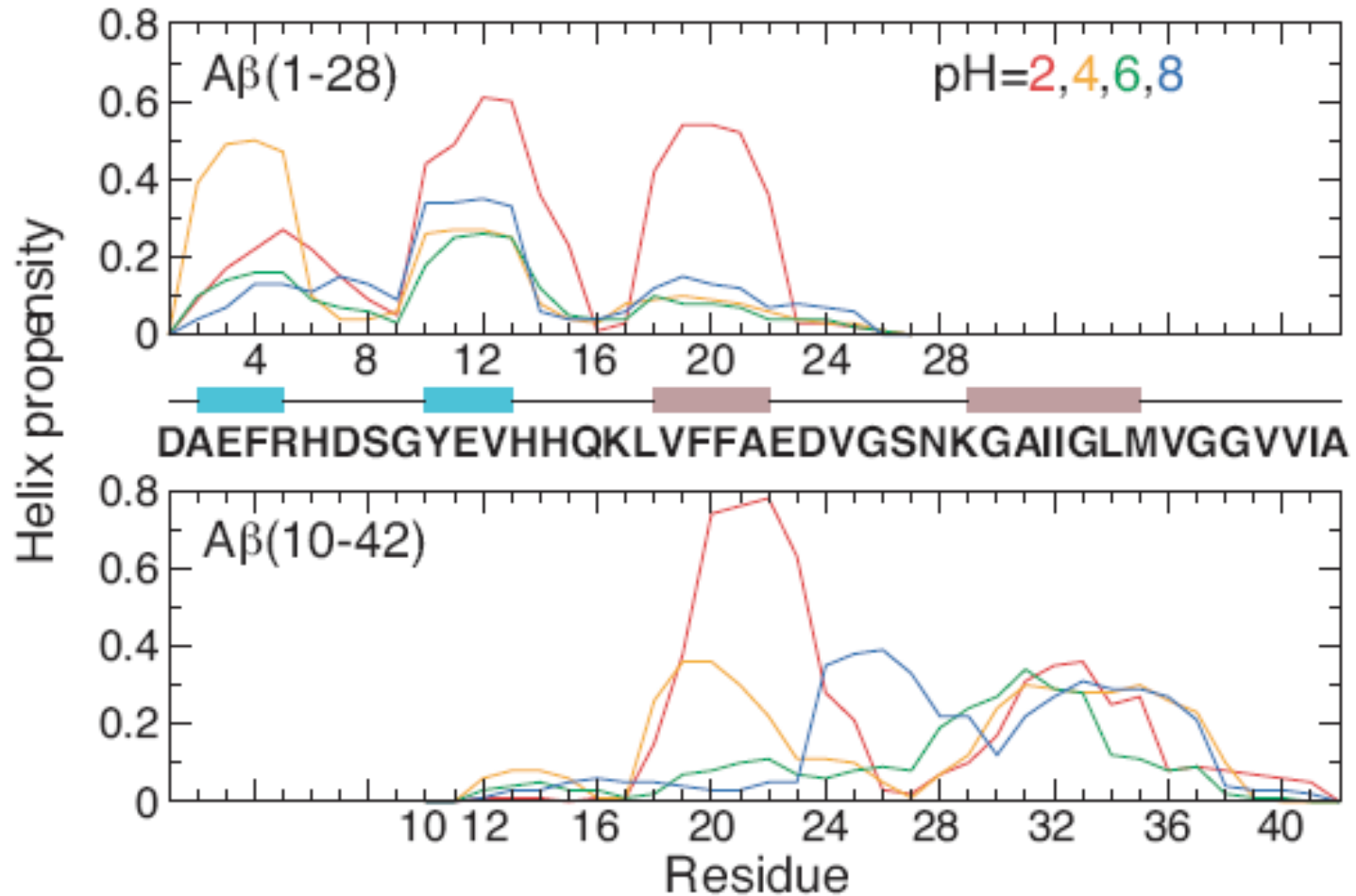


# $\beta$ Amyloid peptide from Alzheimer's disease

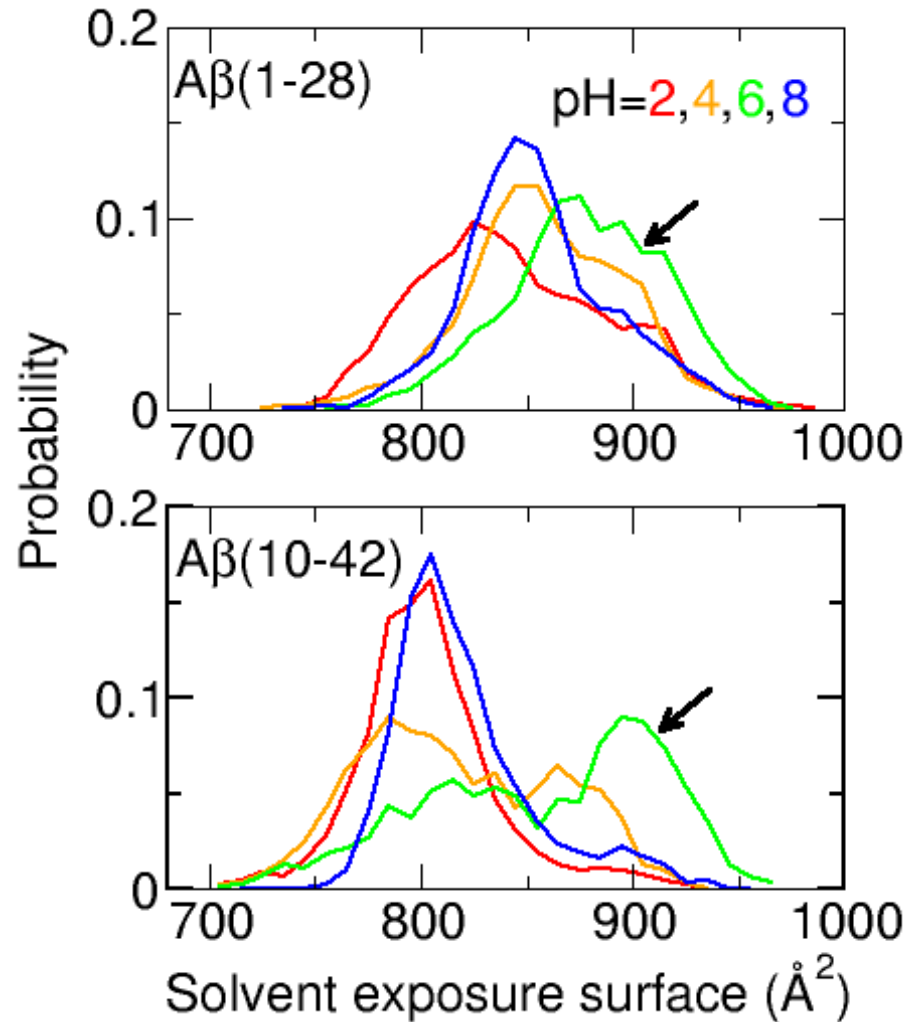


- Extracellular plaques consist of fibrils of  $\beta$ -amyloid peptides ( $A\beta$ ).
- $A\beta$  is a cleavage product of APP.

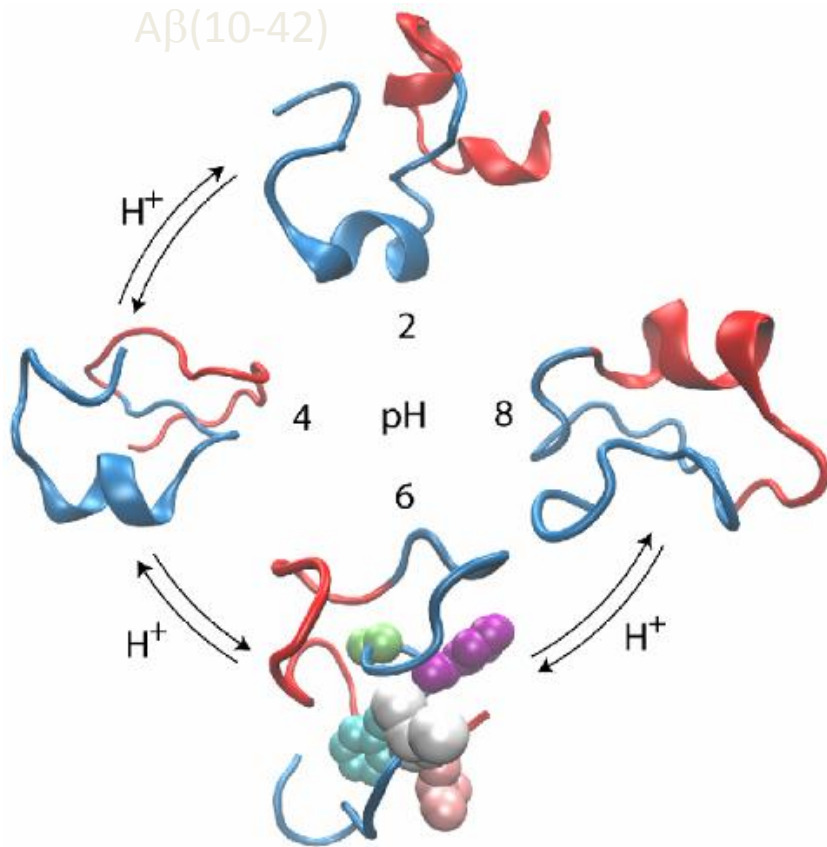
# pH-dependent nascent helix segments



# Solvent exposure of CHC residues



# pH-modulated folding of A $\beta$



Blue: 10-28; red: 29-42; spheres: CHC

