The MARTINI coarse-grained force field for proteins

Luca Monticelli INSERM, France

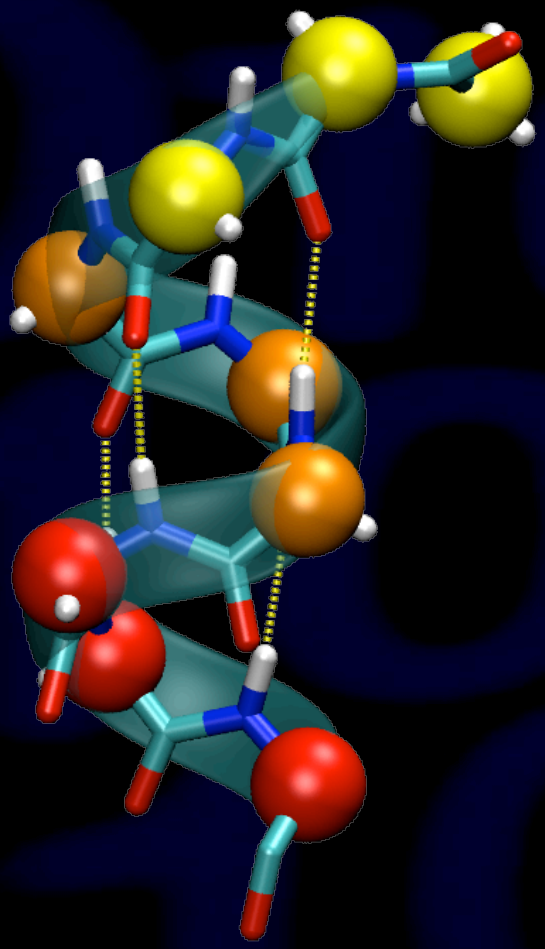
Outline

1. Mapping and general features
2. Parameterization of bonded and non-bonded interactions
3. Validation
4. Limitations of the model
5. Applications

Connecting MD simulations and the real BIO world - sources of errors

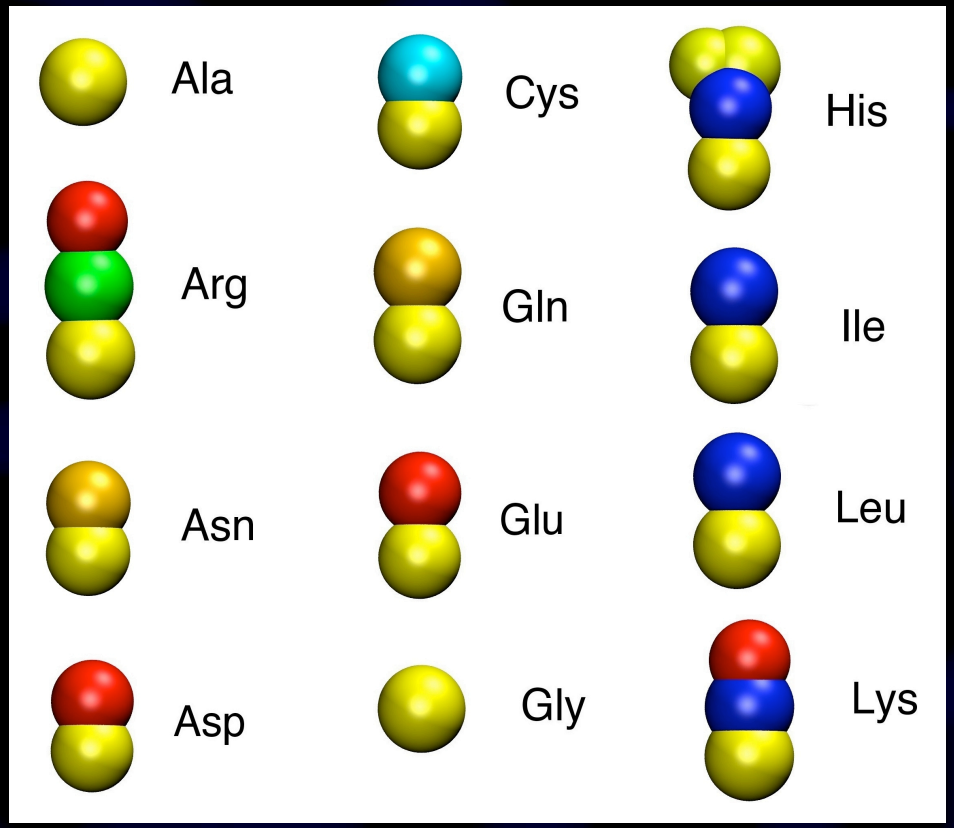
1. Living systems have complex composition
2. You need an accurate force field
3. You need to sample an equilibrium
4. Living systems are not in equilibrium...

MARTINI proteins



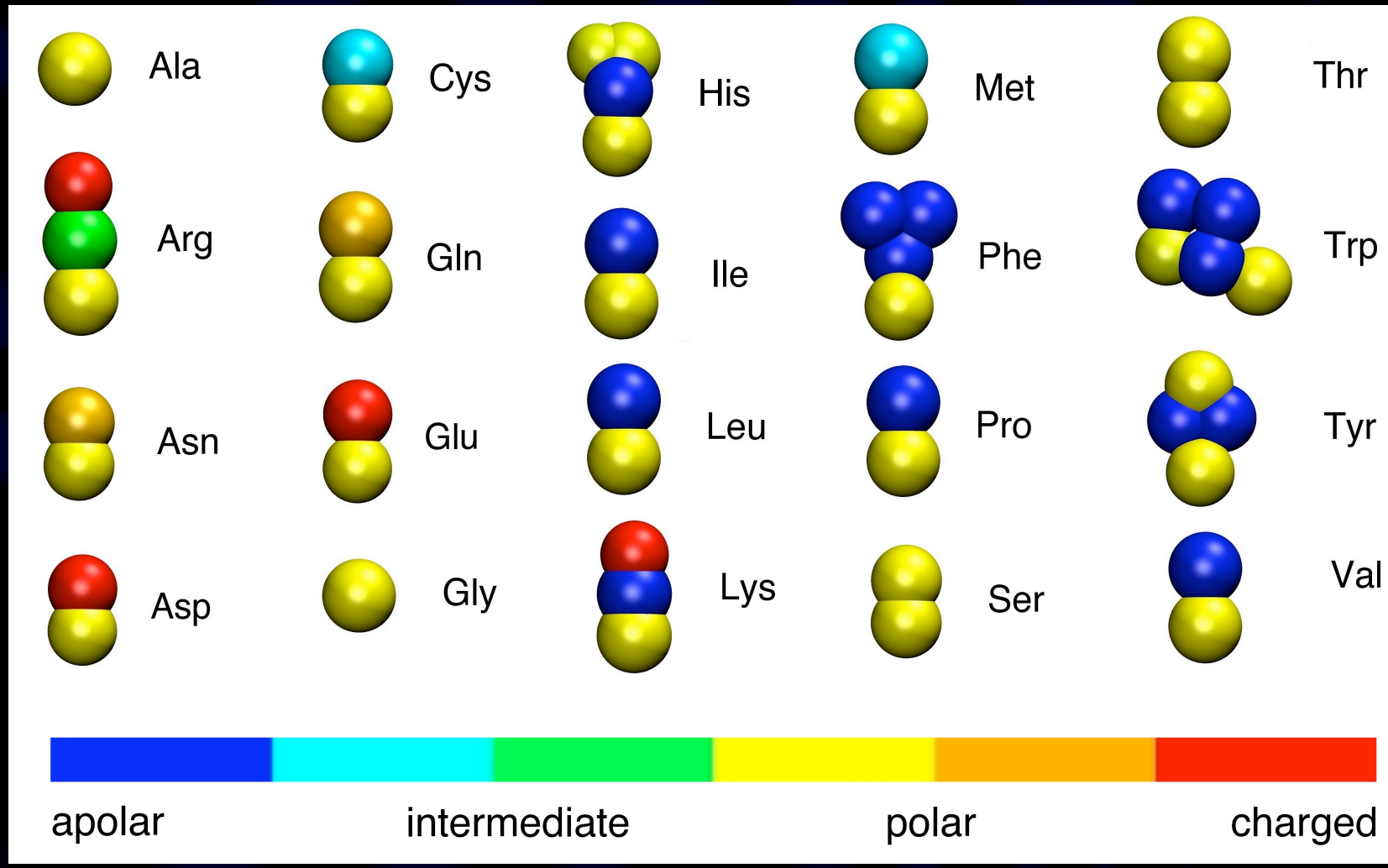
- Fully compatible with MARTINI - lipids etc
- Secondary structure of proteins is imposed, not predicted (no protein folding!)
- Range of possible applications:
  - interaction of proteins with other bio-molecules
  - changes in tertiary structure

MARTINI proteins - mapping

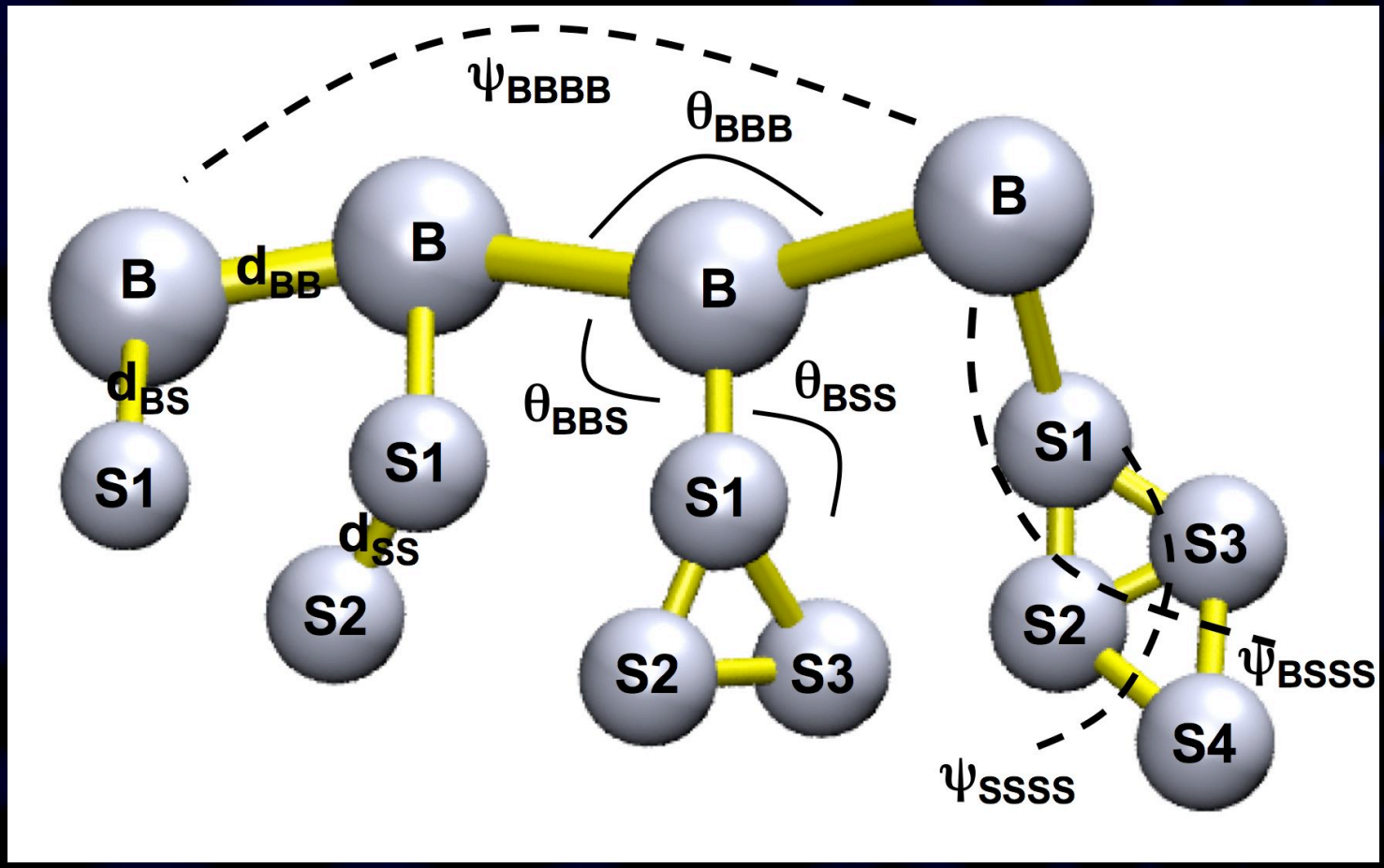


- Usual 4:1 mapping for backbone atoms and most side chains (mass: 72 amu)
- 2:1 mapping in ring structures (H, F, Y, W; mass: 45 amu)
- Same CG bead types as for lipids

MARTINI proteins - mapping



MARTINI proteins - bonded interactions



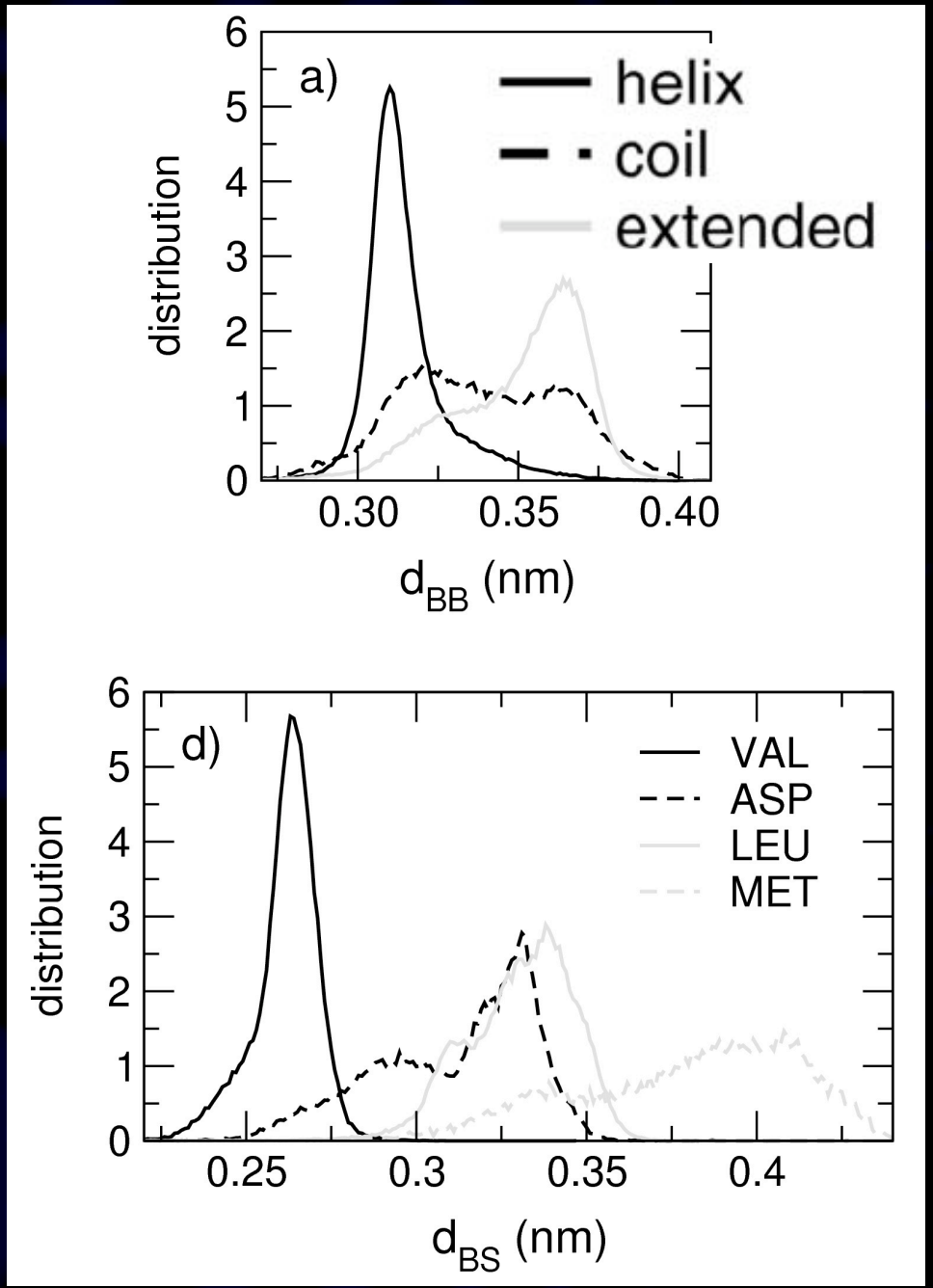
Parameterization of bonded interactions: guidelines

$$V(\mathbf{r}^N) = \sum_{bonds} \frac{K_{b,i}}{2}(d_i - d_{i,0})^2 + \sum_{angles} \frac{K_{a,i}}{2}(\theta_i - \theta_{i,0})^2 + \sum_{torsions} K_{d,i}(1 + \cos(n\psi_i - \psi_{d,i})) + \sum_{impropers} K_{id}(\psi_i - \psi_{id,i})^2$$

1. Initial guess:
   - fit distributions from the PDB
2. Adjustments:
   - Simplicity of the model (simple function types)
   - Stability and speed

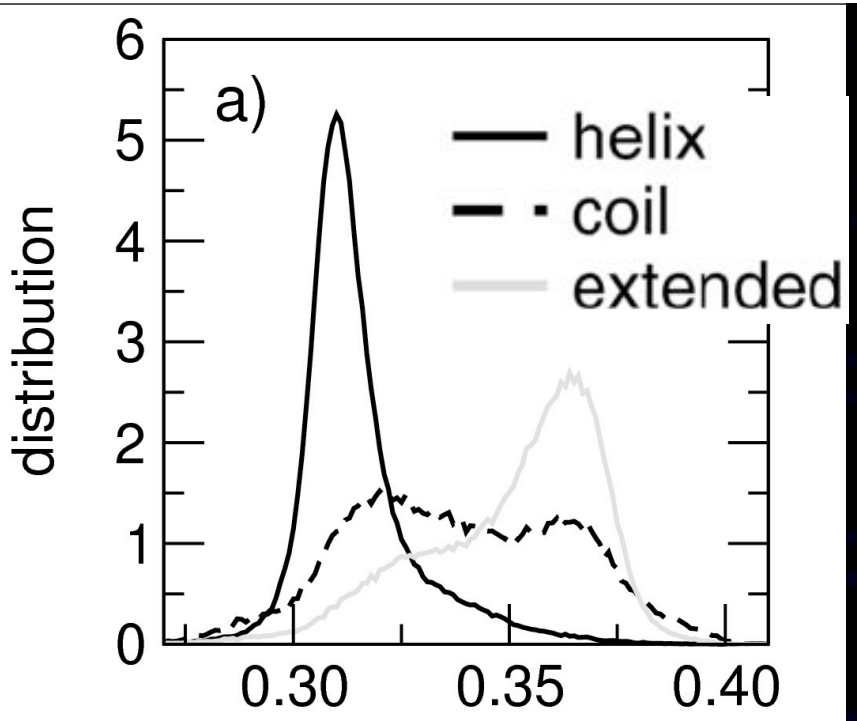
Bonds: fitting to PDB data

1. Take 2000 structures from the PDB
2. Convert them into CG structures
3. Calculate (pseudo) bond distributions
4. Fit with appropriate parameters



Bonds: some simplifications

- Harmonic functions
- d_{BB} = 0.35 nm for all amino acid pairs and for all secondary structures
- Force constants reproduce the width of the distributions (≥ 400)



backbone	d_{BB} (nm)	K_{BB} (kJ nm^{-2} mol^{-1})
helix	0.35	1250
coil	0.35	200
extended	0.35	1250
turn	0.35	500
bend	0.35	400

Bonds: more simplifications

When the force constant was > 7500 we turned it into a constraint

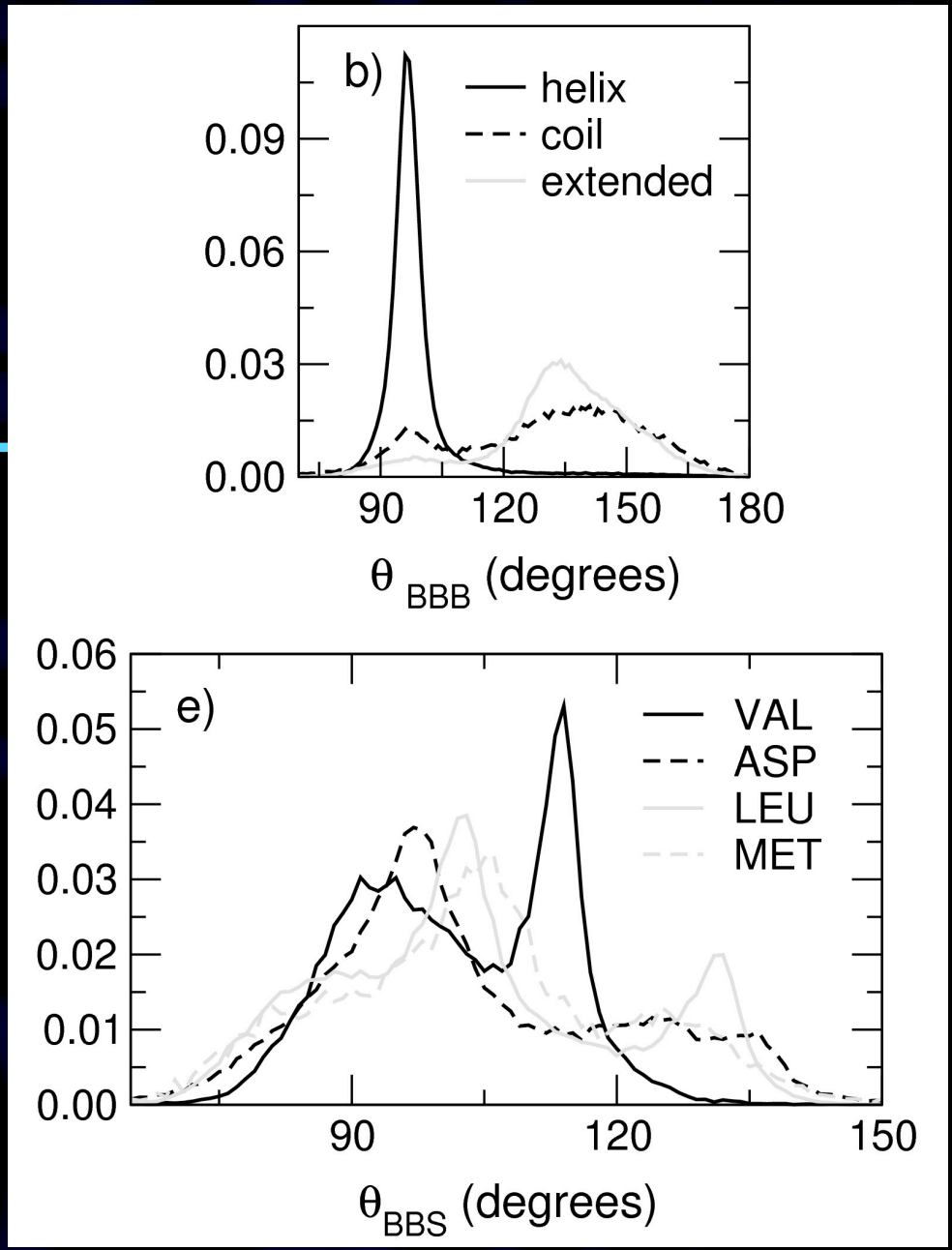
(avoiding fast vibrations allows for longer time step and improves stability)

Table 4. Equilibrium Bond Length and Force Constants for Each Amino Acid Side Chain

side chain	d (nm)	K (kJ nm^{-2} mol^{-1})
Leu	0.33	7500
Ile	0.31	constraint
Val	0.265	constraint
Pro	0.30	7500
Met	0.40	2500
Cys	0.31	7500
Ser	0.25	7500
Thr	0.26	constraint
Asn	0.32	5000
Gln	0.4	5000
Asp	0.32	7500
Glu	0.4	5000
Arg d_{BS}	0.33	5000
Arg d_{SS}	0.34	5000
Lys d_{BS}	0.33	5000
Lys d_{SS}	0.28	5000
His d_{BS}	0.32	7500
His d_{SS}	0.27	constraint
Phe d_{BS}	0.31	7500
Phe d_{SS}	0.27	constraint
Tyr d_{BS}	0.32	5000
Tyr d_{SS}	0.27	constraint
Trp d_{BS}	0.3	5000
Trp d_{SS}	0.27	constraint
Cys−Cys d_{S-S}	0.39	5000

Angles: fitting to PDB data

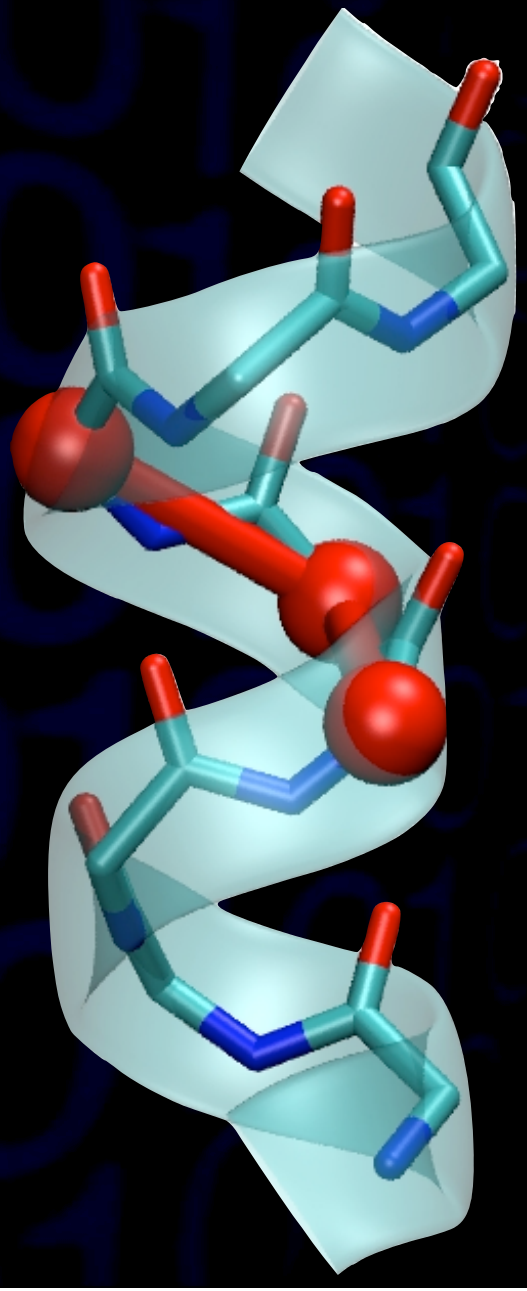
1. Take 2000 structures from the PDB
2. Map them onto CG sites
3. Calculate (pseudo) bond distributions
4. Fit with appropriate parameters



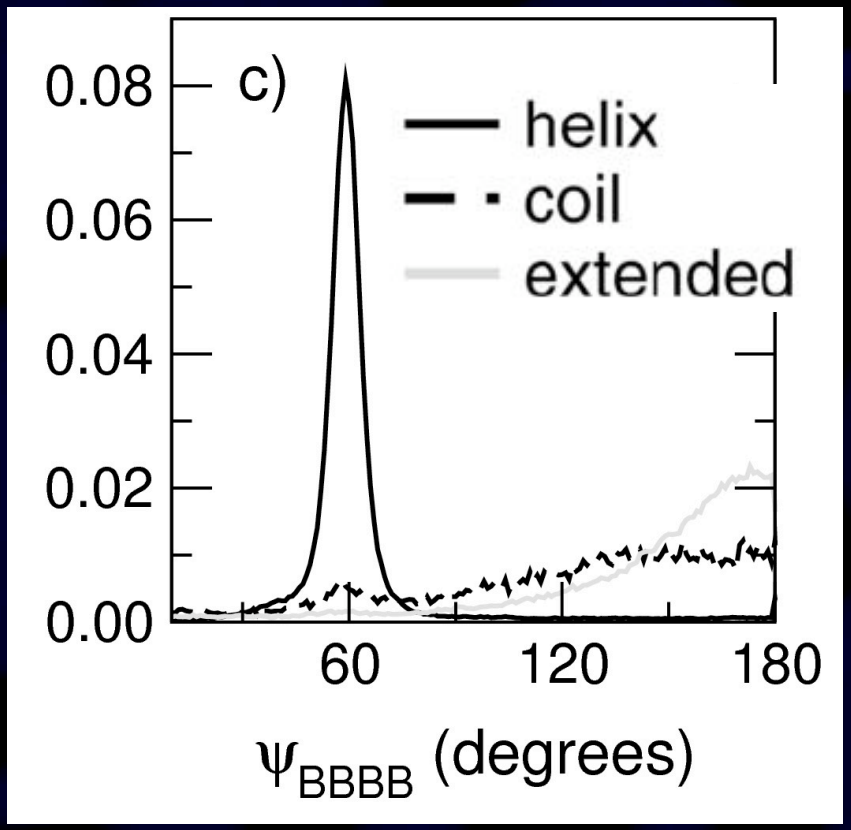
Angles: simplifications

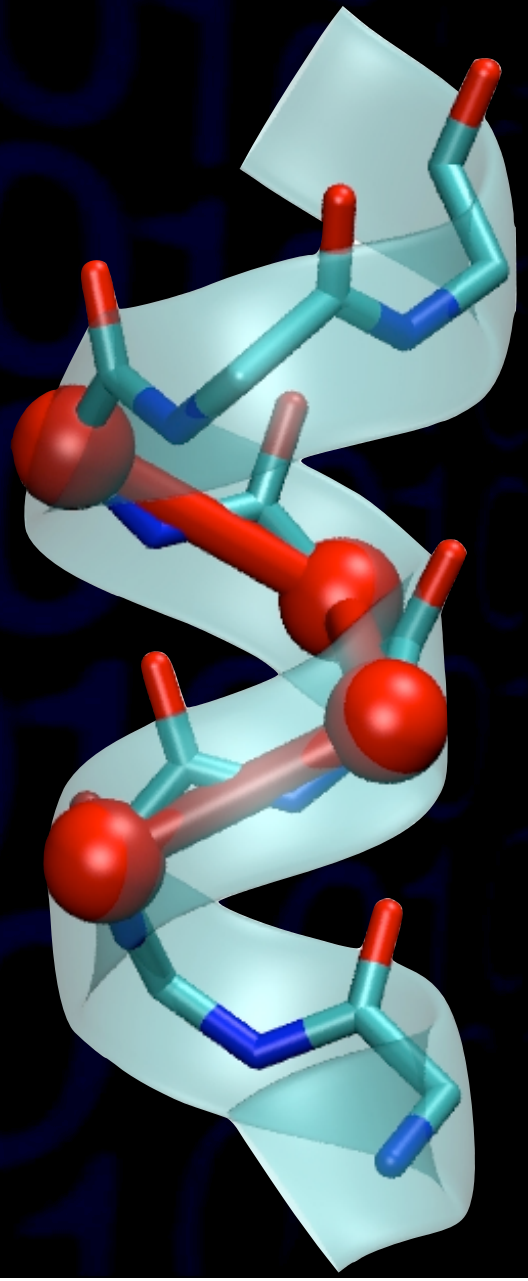
θ_{BBB} and K_{BBB} depend on the secondary structure but are independent of the amino acid sequence.

backbone	θ_{BBB} (deg)	K_{BBB} (kJ mol^{-1})
helix	96[a]	700
coil	127	25
extended	134	25
turn	100	25
bend	130	25



Dihedrals: helices (and sheets?)





Non-bonded interactions: the potential energy functions

Both electrostatics and LJ: shifted potential

Coulomb:
only for ions!

ε_{rel} = 15

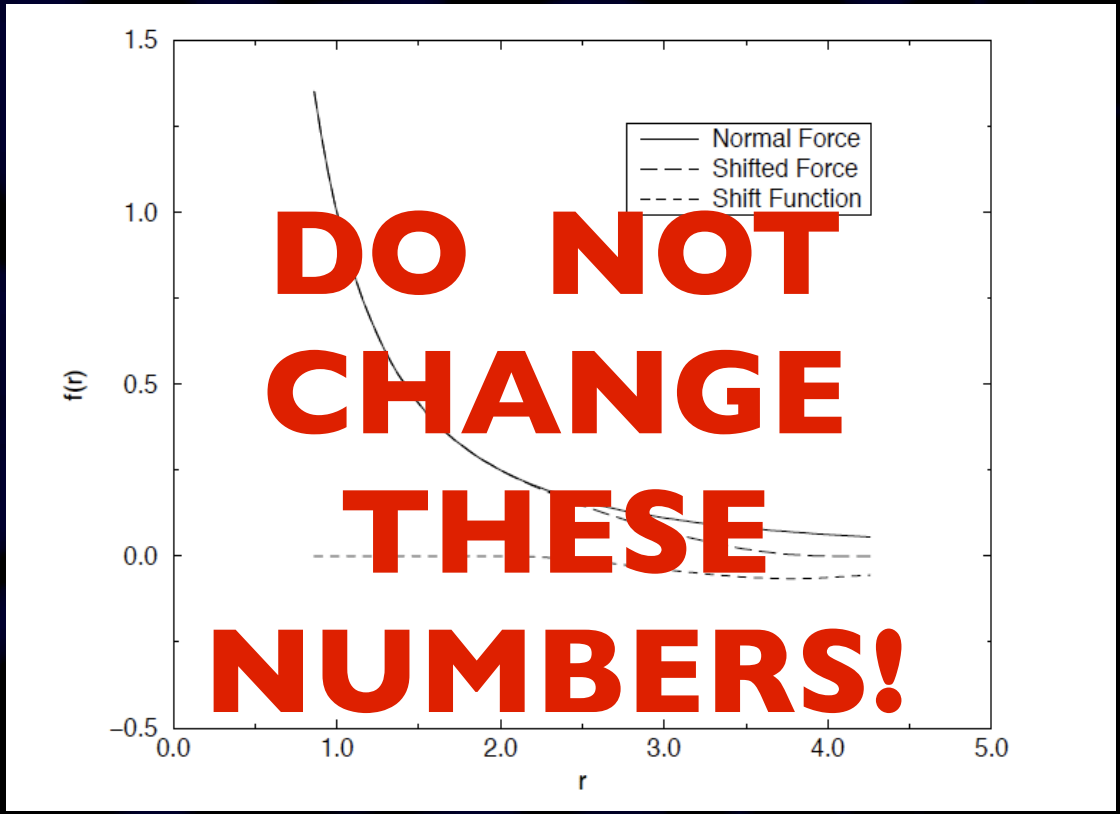
r_{cut} = 1.2 nm

r_{shift} = 0.0 nm

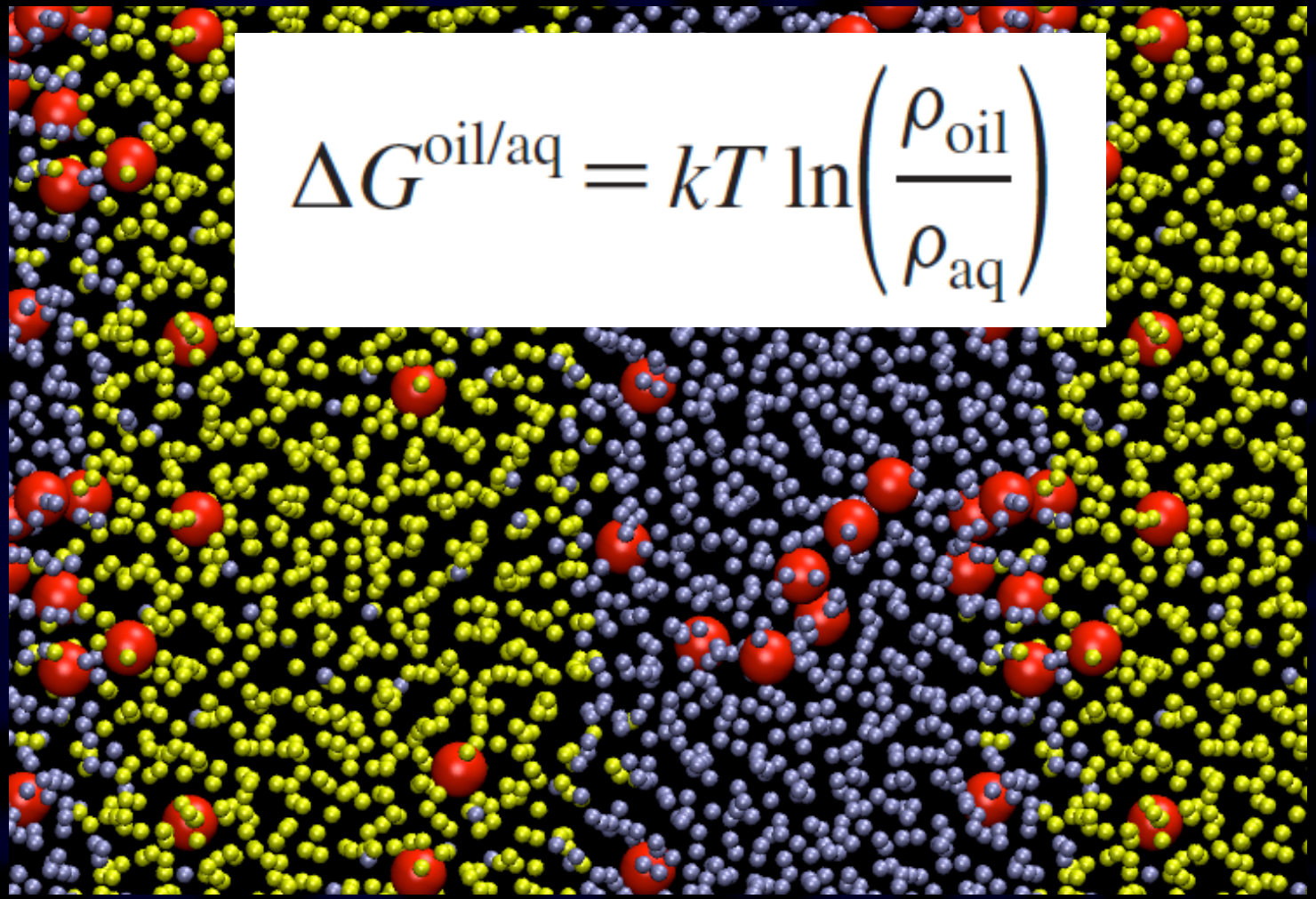
Lennard-Jones

r_{cut} = 1.2 nm

r_{shift} = 0.9 nm



L-J: fitting partitioning data



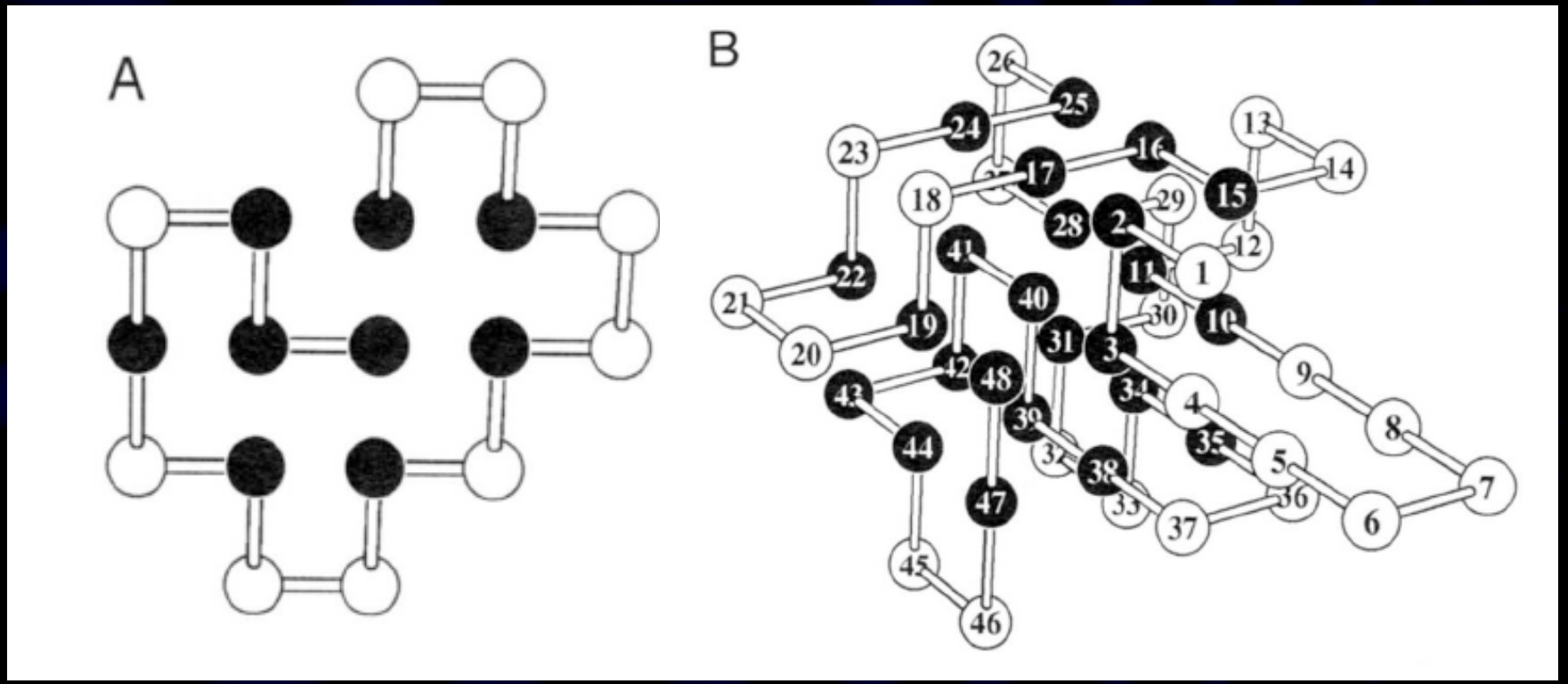
$$\Delta G^{\text{oil/aq}} = kT \ln\left(\frac{\rho_{\text{oil}}}{\rho_{\text{aq}}}\right)$$

Why is partitioning important in simulations of proteins?

1. Partitioning is one of the main driving forces determining protein folding and structure

(polar residues are found mostly on water-exposed surfaces of proteins, non-polar residues in the interior or on lipid-exposed surfaces)

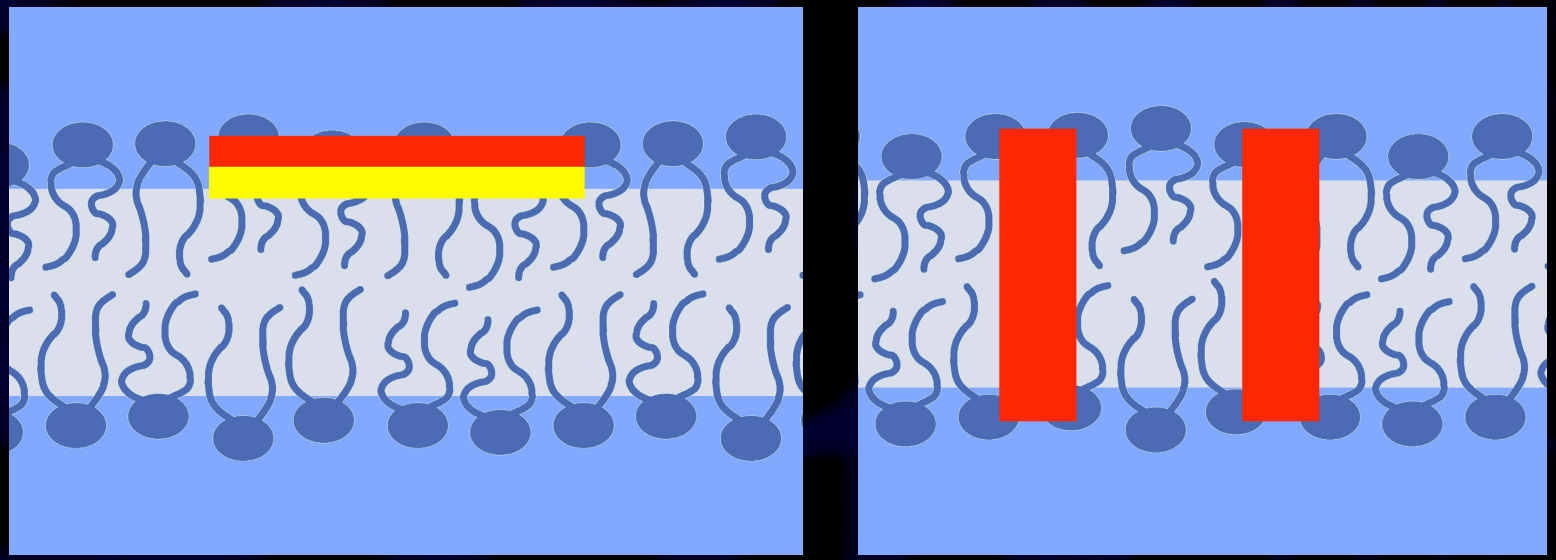
Why is partitioning important in simulations of proteins?



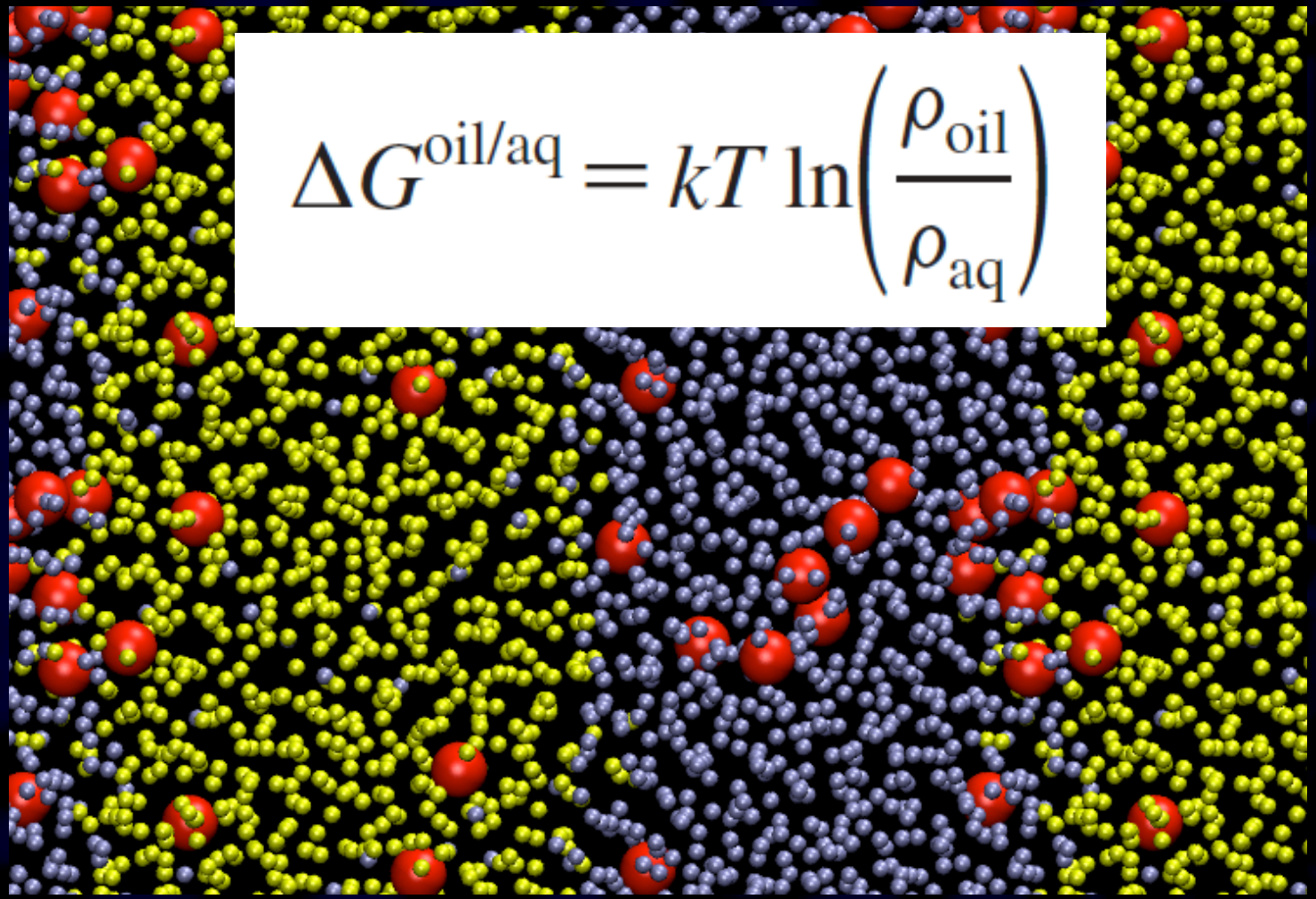
The HP model (Ken Dill & co, early 1990s)

Why is partitioning important in simulations of proteins?

2. Partitioning determines protein interaction with membranes and with other bio-molecules



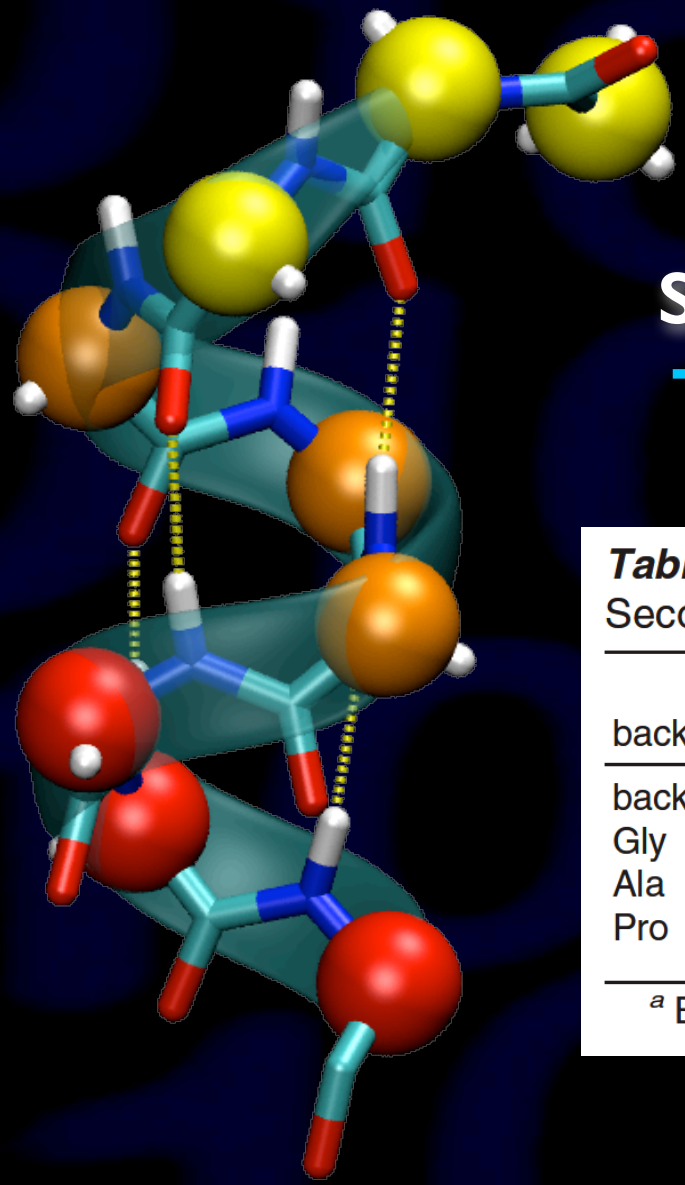
L-J: fitting partitioning data



$$\Delta G^{\text{oil/aq}} = kT \ln\left(\frac{\rho_{\text{oil}}}{\rho_{\text{aq}}}\right)$$

L-J: fitting partitioning data

Side chain	CG representation	Mapping scheme[b]	Free Energy (kJ/mol) CG	exper.
Leu	C1		22	22
Thr	P1		-11	-11
Asn	P5		<-25	-28
Arg	N0-Qd	N0: Cβ-Cγ-Cδ-Nε	< -25	--
Arg (uncharged)	N0-P4	Qd/P4: Cζ–Nω1–Nω2	-23	-25
Trp	SC4-SP1-SC4-SC4	SC4: Cβ-Cγ-Cδ2 SP1: Cδ1-Nε–Cε1 SC4: Cε2–Cζ2 SC4: Cε1–Cω	12	9



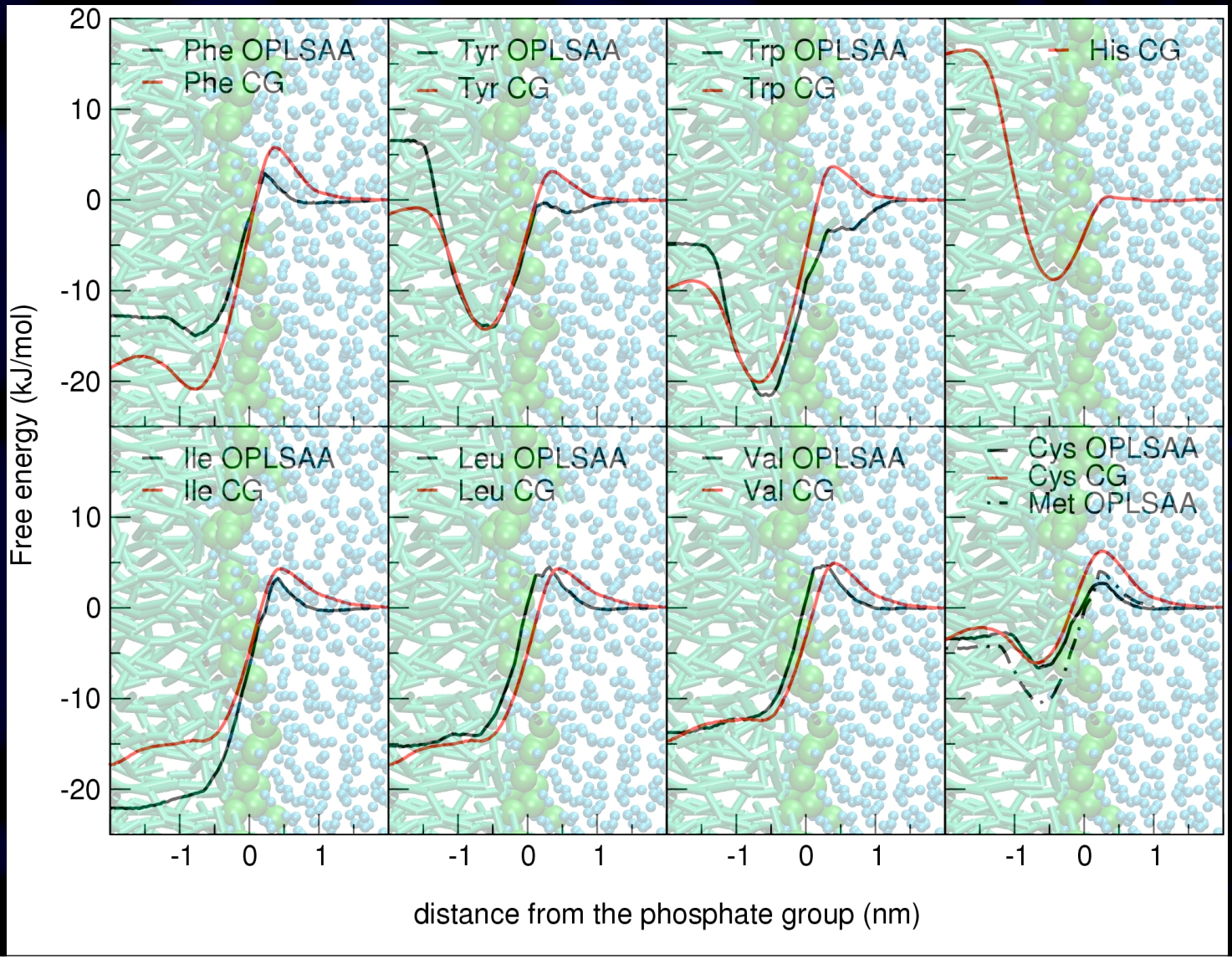
The backbone bead type depends on the secondary structure!!!

Table 2. Backbone Particle Type in Different Kinds of Secondary Structure[a]

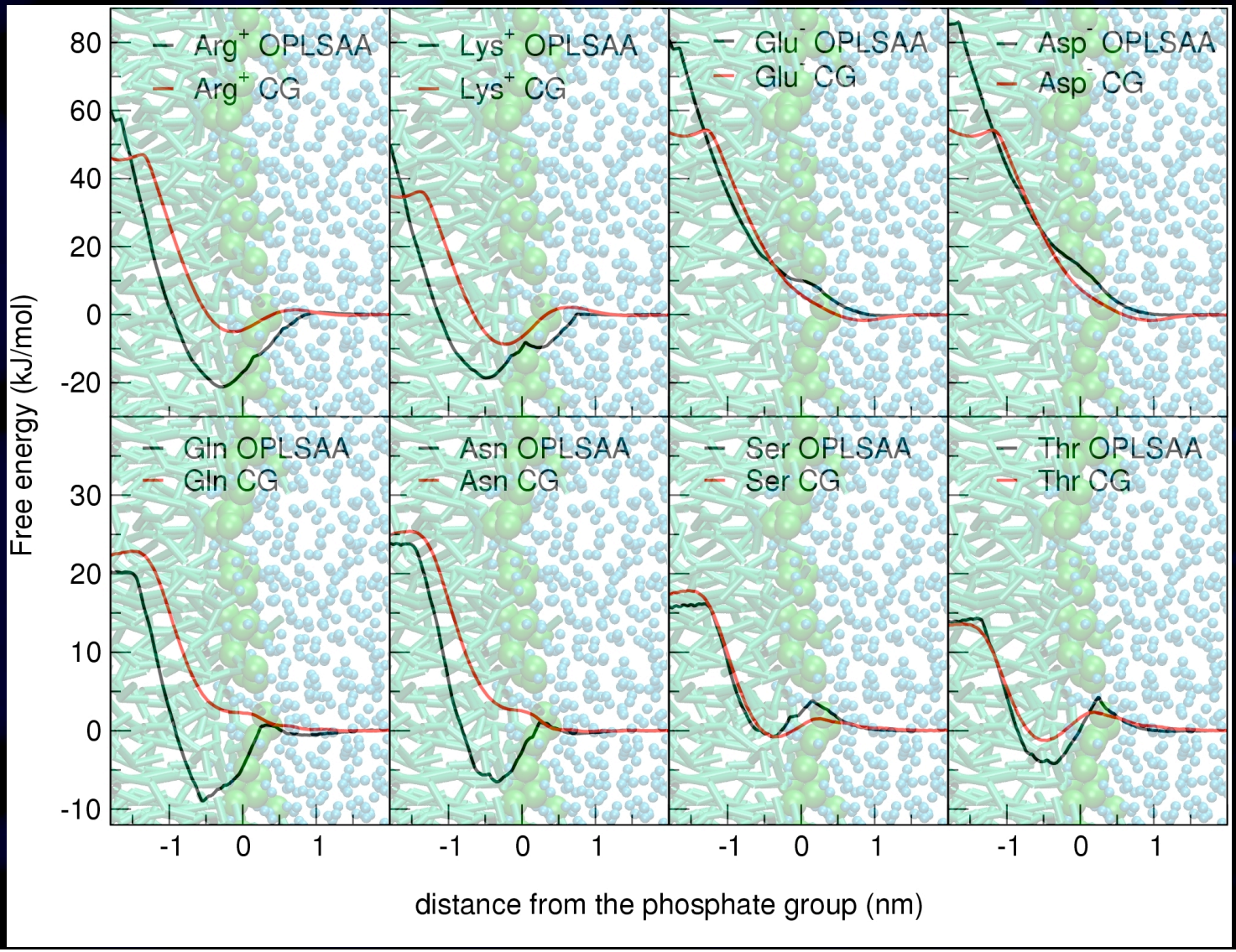
backbone	coil bend free	helix	helix (N-terminus/C-terminus)	β-strand turn
backbone	P5	N0	Nd/Na	Nda
Gly	P5	N0	Nd/Na	Nda
Ala	P4	C5	N0	N0
Pro	Na	C5	N0/Na	N0

[a] Both glycine and alanine have no side chain.

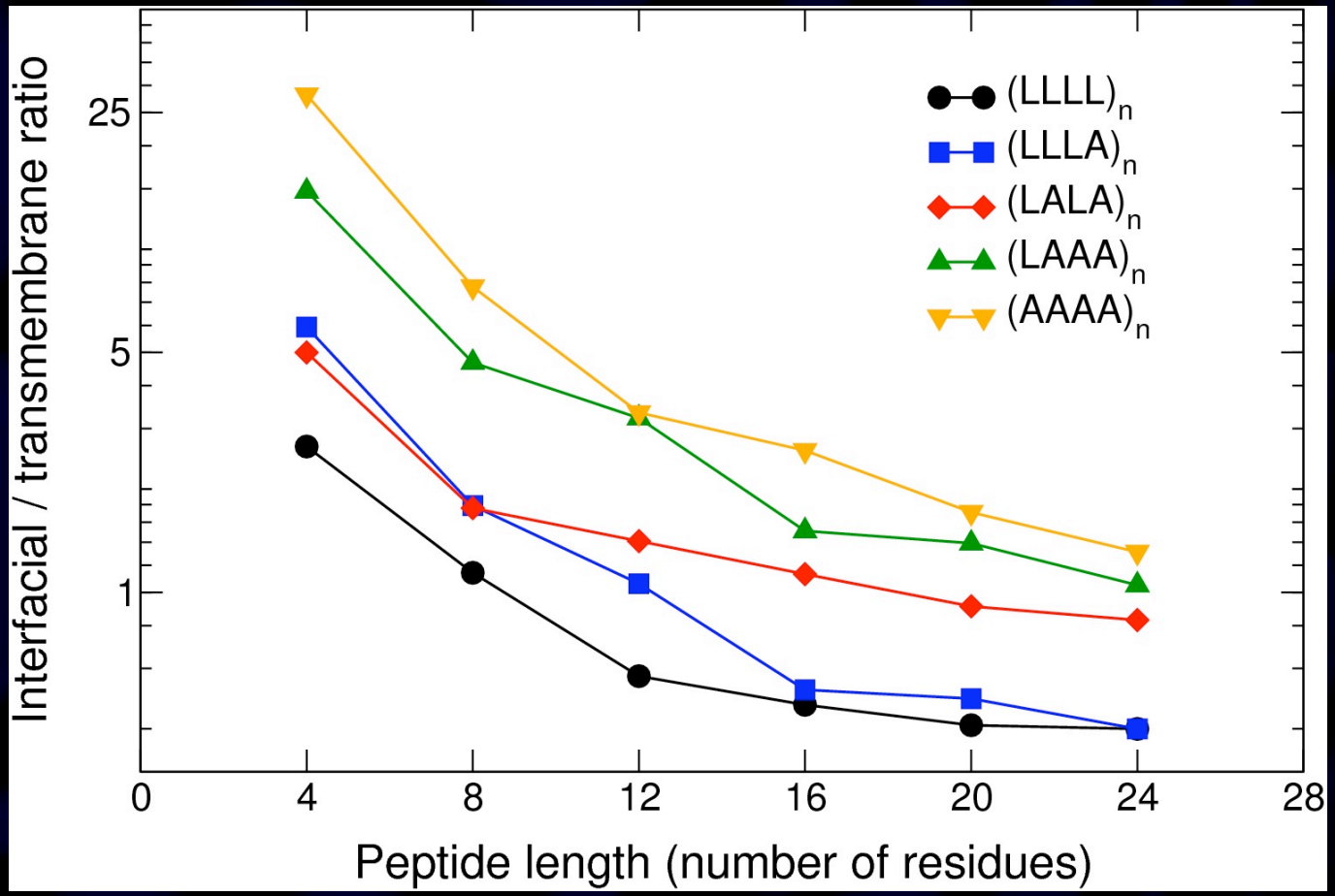
Validation: partitioning in bilayers



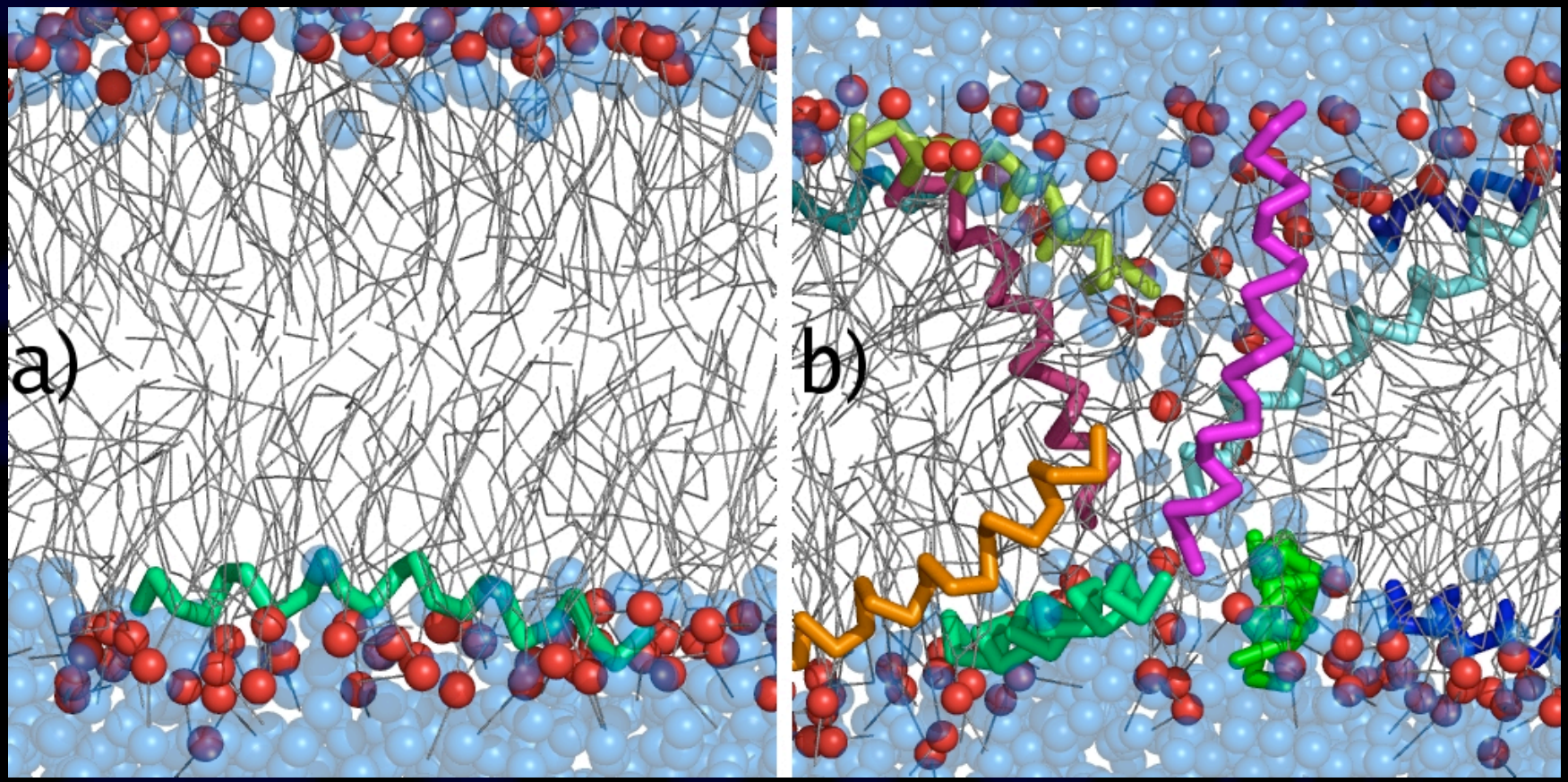
Validation: partitioning in bilayers



Validation: partitioning of hydrophobic peptides in bilayers



Validation: antimicrobial peptides can form toroidal pores



Magainin H2 in a DPPC bilayer, at low concentration (a) and high concentration.

Limitations of the model

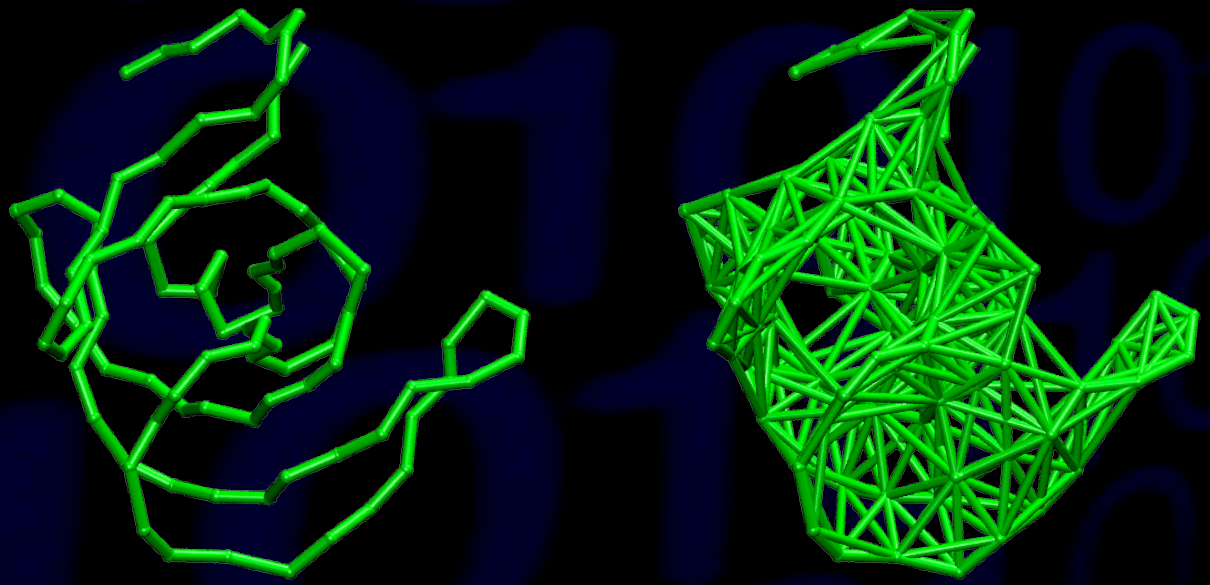
1. The model cannot predict the local structure and changes in it. So it cannot be used when protein secondary structure changes!
2. Temperature dependence of properties is not always correct (model based on free energies)
3. The model is parameterized for the fluid phase, at T around 300K; it is not expected to work well in other conditions

Limitations of the model

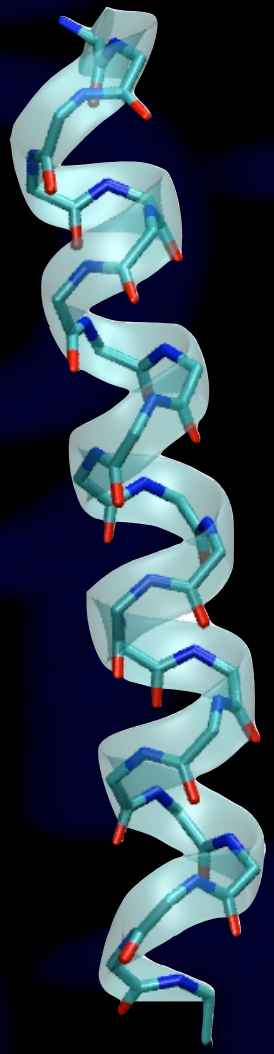
4. Partitioning of (polar and) charged species shows significant discrepancies with atomistic models (free energy and hydration)
5. Instability of beta-sheet structures when only improper dihedrals are used
6. Molecular surfaces are not very accurate (might be a problem for some protein-protein interactions)

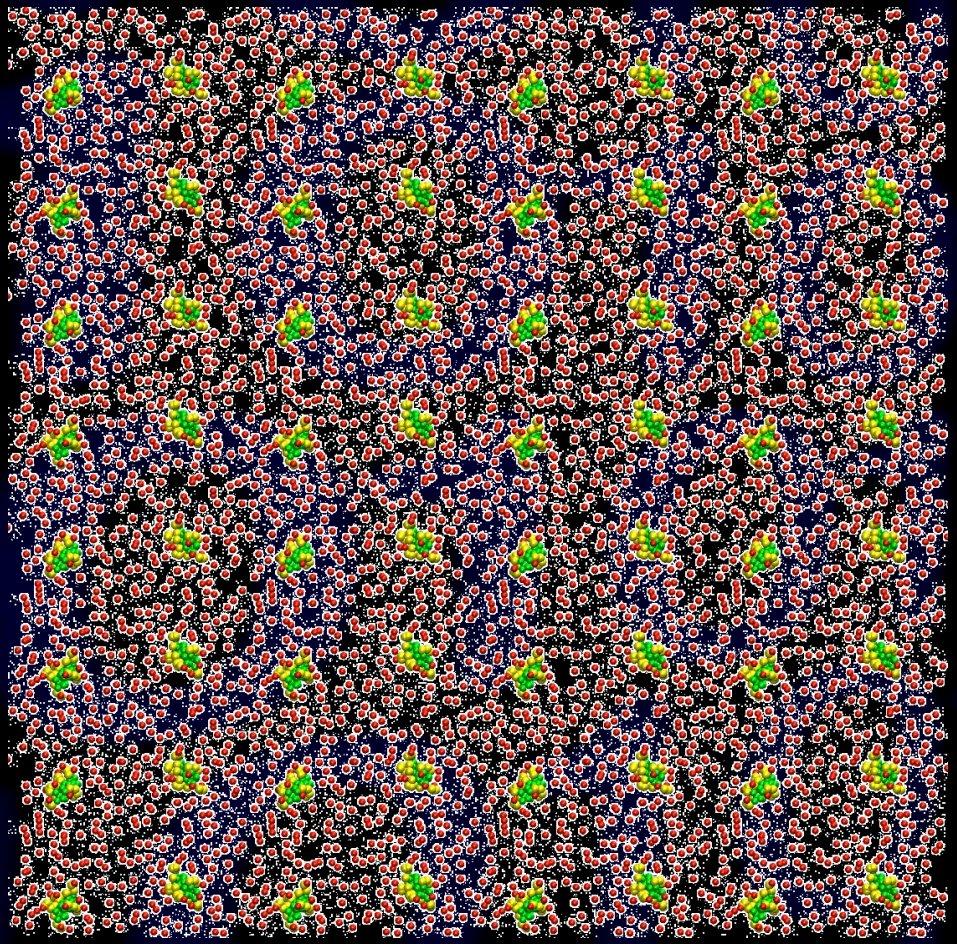
New developments

Application of elastic networks: either on top of "regular" MARTINI, or based on atomistic coordinates (developed by Xavier Periole).



Does WALP23 aggregate?





System:
64 WALP23 peptides (antiparallel orientation), 4608 DOPC, 114008 WATER

Simulations:
6 MD runs
16 μs each

Acknowledgements

Dr. S. Kandasamy (U. Michigan)
Dr. X. Periole (U.Groningen)
Prof. R.G. Larson (U. Michigan)
Prof. D.P.Tieleman (U. Calgary)
Prof. S.J. Marrink (U. Groningen)

AHFMR
ALBERTA HERITAGE FOUNDATION FOR MEDICAL RESEARCH

ACADEMY OF FINLAND

Instituts thématiques
Inserm
Institut national de la santé et de la recherche médicale