Selected Topics of New Frontiers in Computational Systems Biology

Yinghao Wu
Albert Einstein College of Medicine

Department of systems and computational biology

Outline

- **Modeling and Simulating Larger Molecular Systems with Longer Time Scales**
  - **Coarse-grained simulation**
  - **Multi-scale modeling**
- **Extend the Applications of Structural Biology to Newly Emerging Fields**
  - **Fold proteins by information from evolution**
  - **Simulating crowded cellular environment**
  - **Combining structural biology with systems biology**
- **Summary and Perspective**

Outline

- Modeling and Simulating Larger Molecular Systems with Longer Time Scales
  - Coarse-grained simulation
  - Multi-scale modeling
- Extend the Applications of Structural Biology to Newly Emerging Fields
  - Fold proteins by information from genetics
  - Simulating crowded cellular environment
  - Combining structural biology with systems biology
- Summary and Perspective

Background: Why Coarse-graining (CG) ?

All atom:
time step of 1-2 fs,
time frame sampled ~100 ns

All atom:
each particle represent one atom
Maximally can handle n-atom system

- Reduction in degrees of freedom
- Fast frequency movements removed
- Smoother potential surface
- Longer time steps can be taken
- Microsecond simulations possible

Coarse-grained:
time step of 20-50 fs,
time frame sampled ~1 μs

Coarse-grained:
each particle represent m atom
Maximally can handle mn-atom system

Background: How Coarse-graining (CG) ?

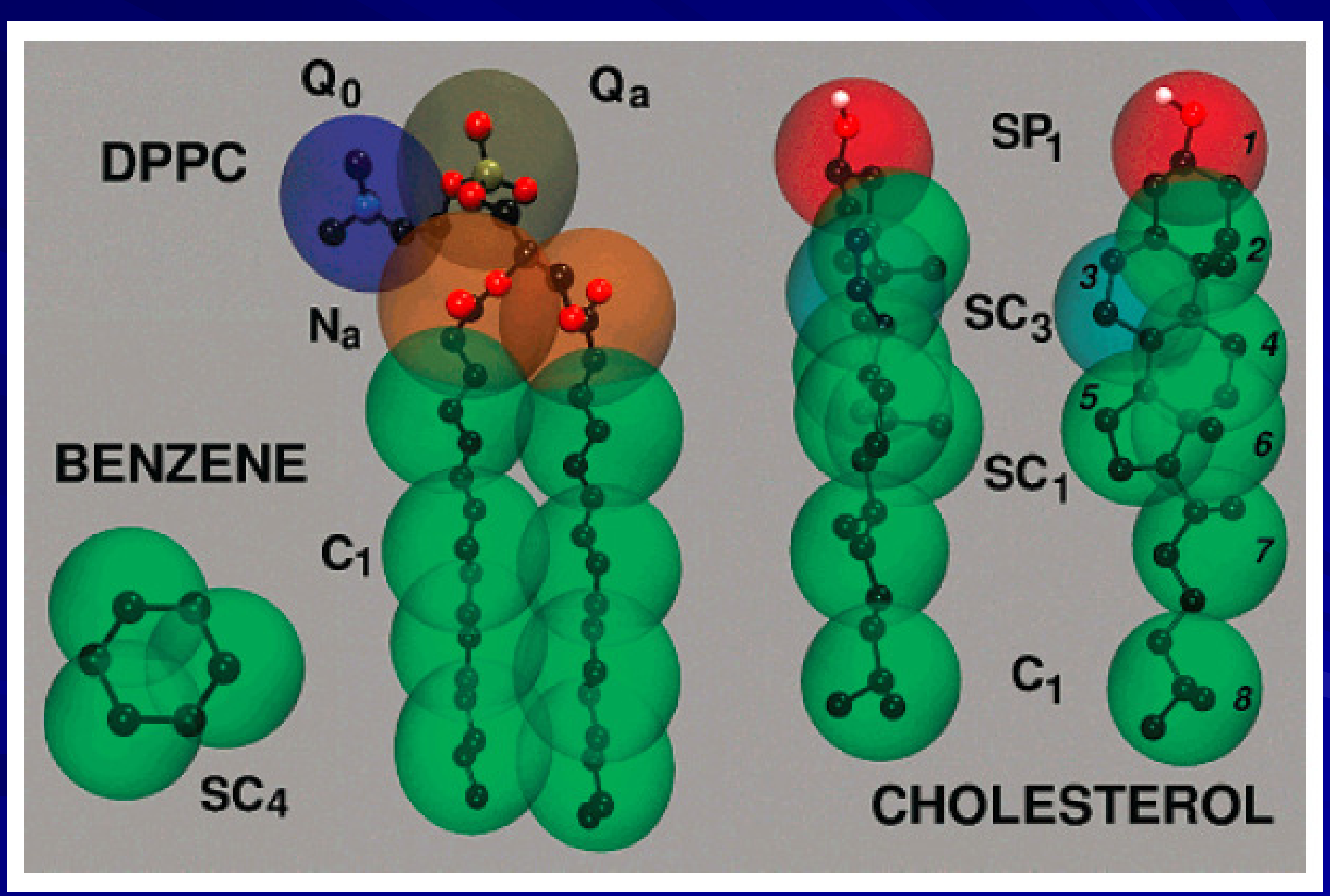
- **How to give a simplified representation for biomolecule with reasonable accuracy (CG model).**

- **How to get the correct interactions between different simplified sites within the models (Force Mapping).**

- **MATINI is a software to do CG simulations.**

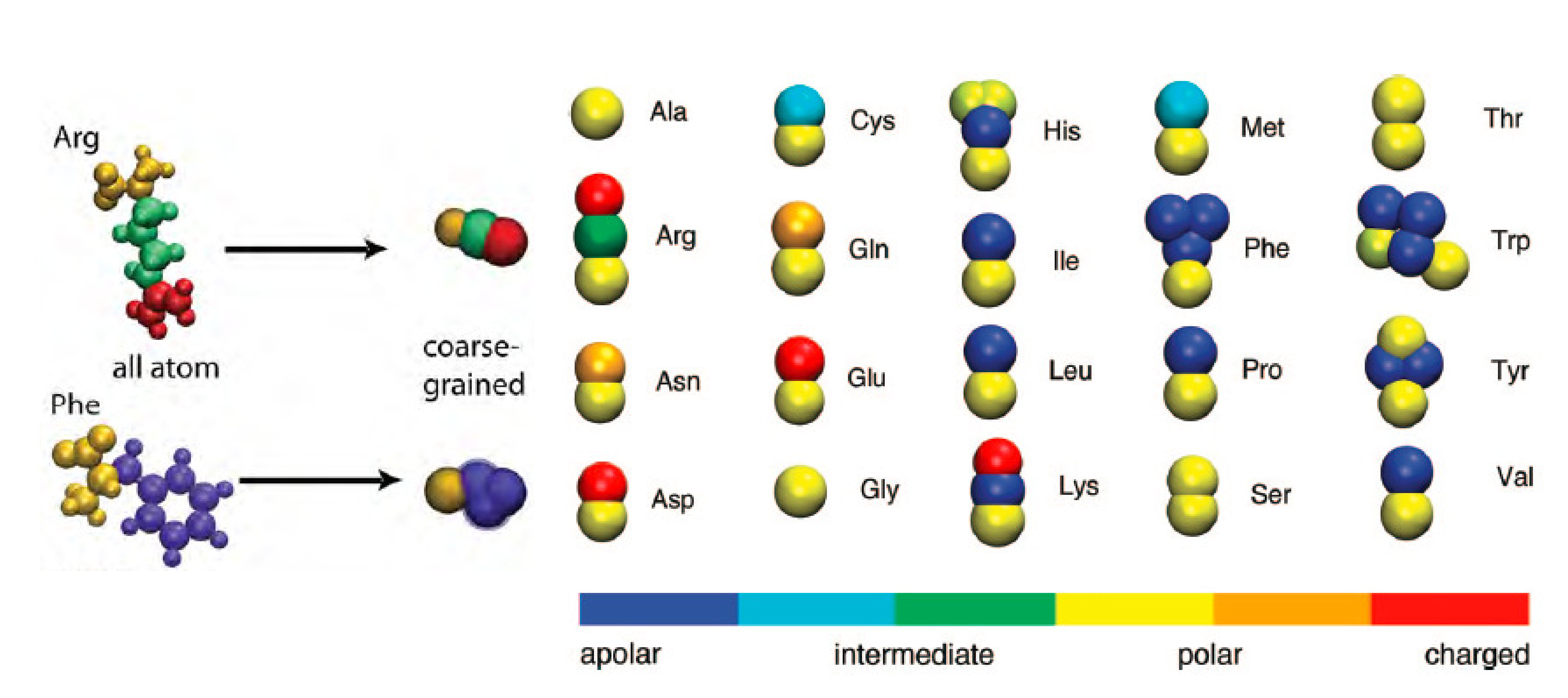
CG Simulation: Simplified Representation

Coarse-grained Model of Phosphor-lipid



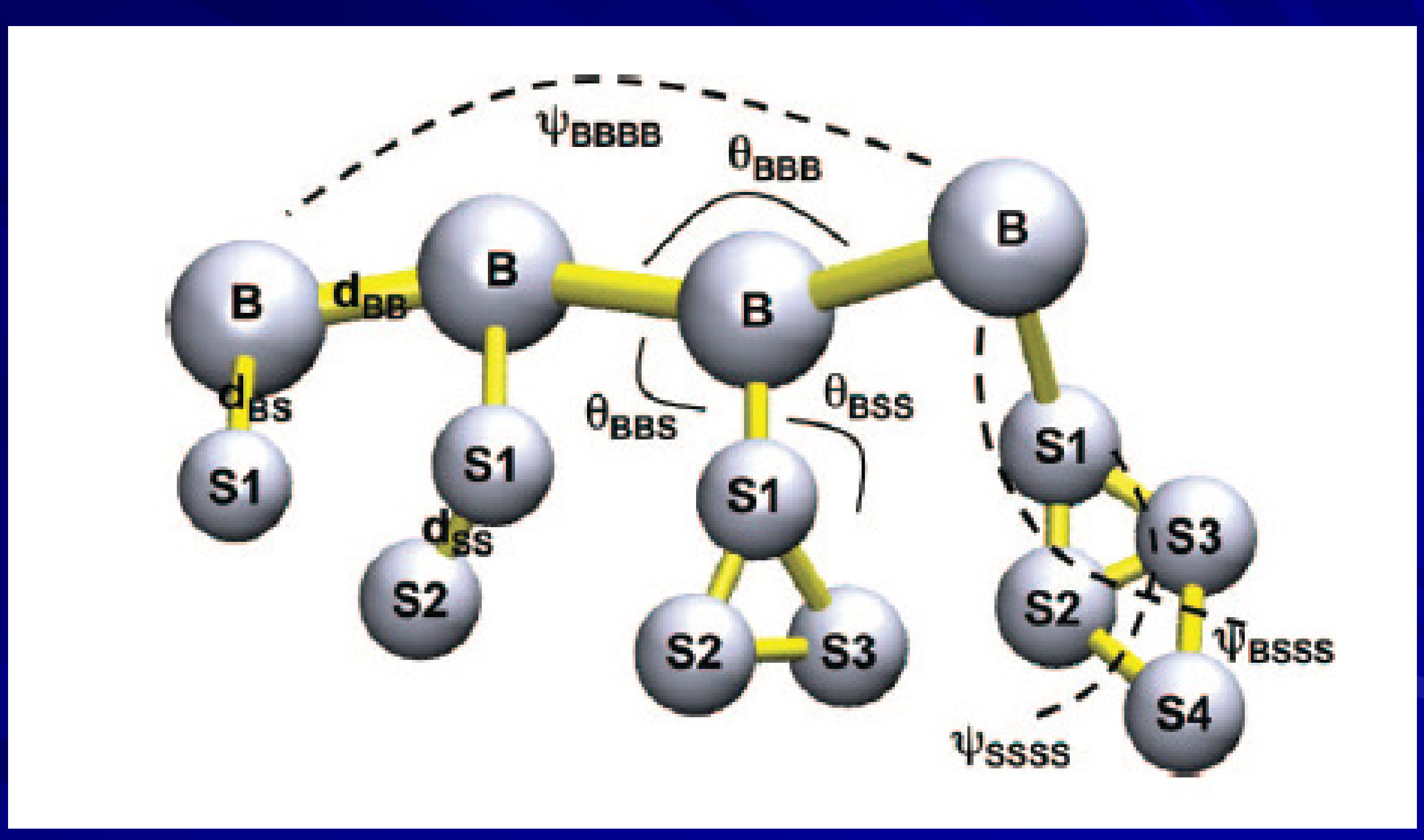
CG Simulation: Simplified Representation

Coarse-grained Model of Protein Side-chain: Mapping Sites



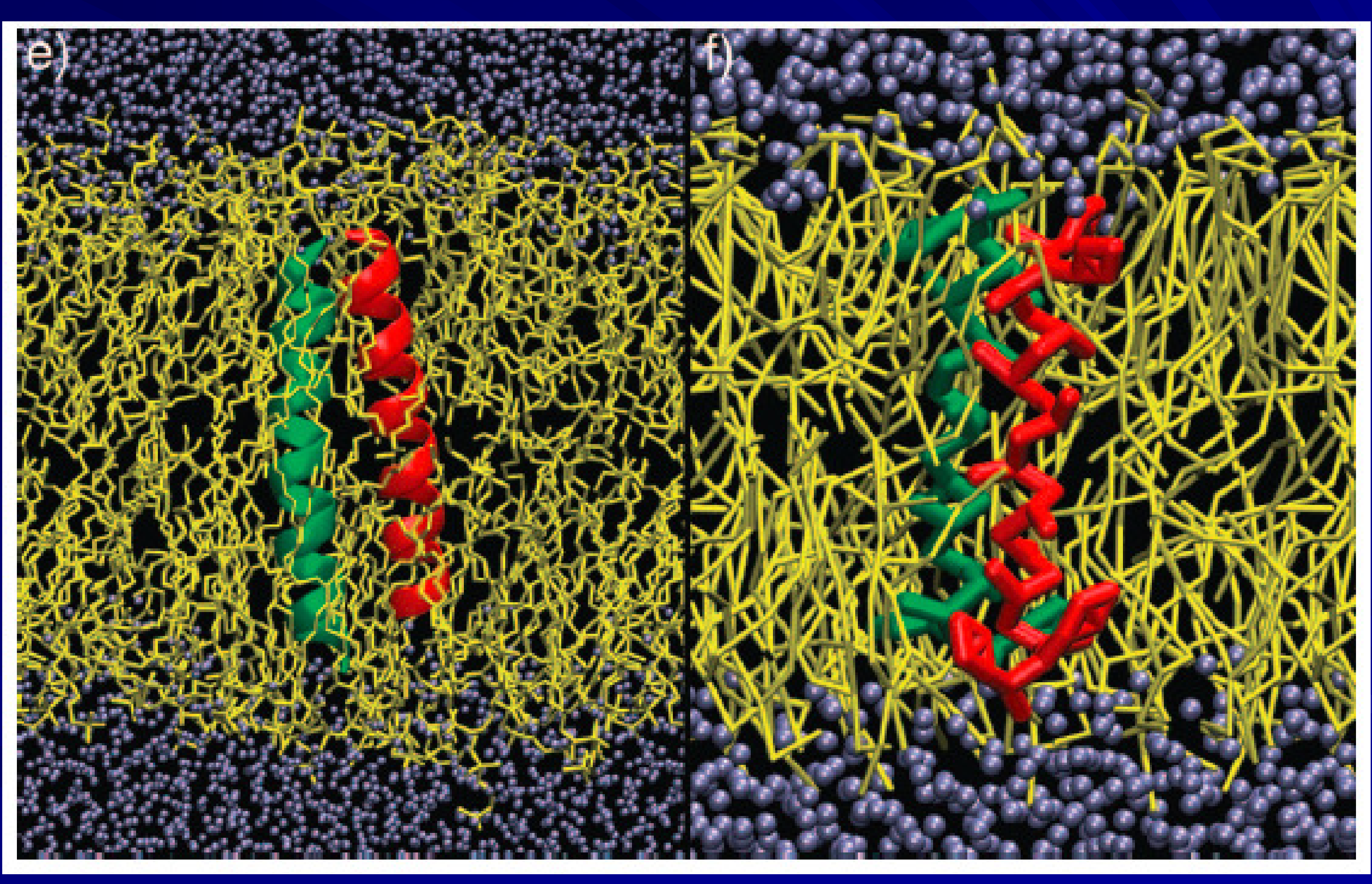
CG Simulation: Simplified Representation

Coarse-grained Model of Protein Main-chain: Degrees of Freedom



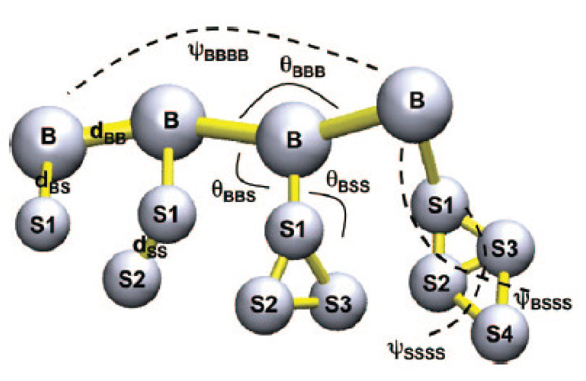
CG Simulation: Simplified Representation

Comparison between All Atom Model and Coarse-grained Model



CG Simulation: Force Field Parameterization

Bonded Interactions



$$V_{\mathrm{b}} = \frac{1}{2} K_{\mathrm{b}} (d_{ij} - d_{\mathrm{b}})^2$$

$$V_{\mathrm{a}} = \frac{1}{2} K_{\mathrm{a}} [\cos(\varphi_{ijk}) - \cos(\varphi_{\mathrm{a}})]^2$$

$$V_{\mathrm{d}} = K_{\mathrm{d}} [1 + \cos(n\psi_{ijkl} - \psi_{\mathrm{d}})]$$

$$V_{\mathrm{id}} = K_{\mathrm{id}} (\psi_{ijkl} - \psi_{\mathrm{id}})^2$$

- **Bonded interactions are described by the following set of potential energy functions acting between bonded sites i, j, k, and l with equilibrium distance, angle, and dihedral.**

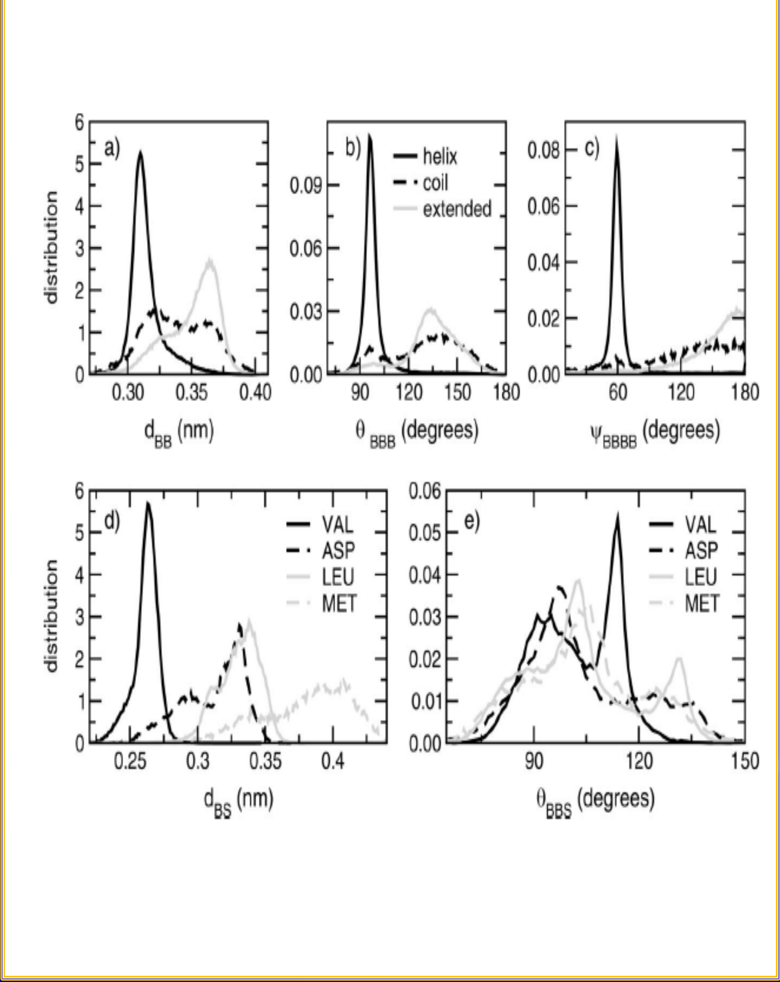
- **The force constants K are generally weak, inducing flexibility of the molecule at the coarse-grained level mimicking the collective motions at the fine-grained level. The bonded potential is used for chemically bonded sites and the angle potential to represent chain stiffness.**

- **The improper dihedral angle potential is used to prevent out-of-plane distortions of planar groups. Proper dihedrals are used to impose secondary structure of the peptide backbone.c**

CG Simulation: Force Field Parameterization

Calculating bonded force constants: 1) deriving distributions from PDB

- The bond lengths, bond angles, dihedral angles, and their respective force constants, collectively referred to as the bonded parameters, were obtained from distributions derived from the PDB.
- Chose a representative subset of approximately 2000 proteins from the PDB as the basis set for our parametrization.
- The secondary structure of every residue of these proteins was determined using the program DSSP.
- Using the center of mass of the atoms representing each coarse-grained bead, calculated the distributions of the bond lengths, bond angles, and dihedral angles.



CG Simulation: Force Field Parameterization

Calculating bonded force constant: 2) deriving force constants against distributions of PDB

- After the distributions were obtained, simulations were performed on short test peptides, with different sequences and secondary structure characteristics.
- All of the bonded parameters were optimized by matching the PDB distributions of the bonds angles and dihedrals with the distributions obtained from the simulations, using an iterative procedure.

Table 3. Backbone Bonded Parameters

backbone	d_{BB} (nm)	K_{BB} (kJ nm^{-2} mol^{-1})	θ_{BBB} (deg)	K_{BBB} (kJ mol^{-1})	ψ_{BBBB} (deg)	K_{BBBB} (kJ mol^{-1})
helix	0.35	1250	96[a]	700	60	400
coil	0.35	200	127	25		
extended	0.35	1250	134	25	180	10
turn	0.35	500	100	25		
bend	0.35	400	130	25		

[a] $\theta_{BBB} = 98°$ when Proline is in the helix; $K_{BB} = 100$ kJ mol^{-1}.

Table 5. Equilibrium Angles, Improper Dihedral Angles and Force Constants for Side Chains

side chain	θ (deg)	K (kJ mol^{-1})
θ_{BBS} (all)	100	25
θ_{BSS} (Lys, Arg)	180	25
θ_{BSS} (His, Tyr, Phe)	150	50
θ_{BSS} (Trp)	90, 210	50, 50

side chain	ψ (deg)	K (kJ rad^{-2} mol^{-1})
ψ_{BSSS} (His, Tyr, Phe)	0	50
ψ_{BSSS} (Trp)	0, 0	50, 200

CG Simulation: Force Field Parameterization

Non-bonded Interactions: Lennard-Jones Potential

$$V_{\text{Lennard-Jones}}(r_{ij}) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$$

- The strength of the interaction, determined by the value of the well depth εij depends on the interacting particle types.

- The value of ε ranges from εij) 5.6 kJ/mol for interactions between strongly polar groups to εij) 2.0 kJ/mol for interactions between polar and apolar groups mimicking the hydrophobic effect.

- The effective size of the particles is governed by the LJ parameter: σ) 0.47 nm for all normal particle types.

CG Simulation: Force Field Parameterization

Non-bonded Interactions: Interaction Matrix

TABLE 1: Interaction Matrix[a]

		Q				P					N				C				
	sub	da	d	a	0	5	4	3	2	1	da	d	a	0	5	4	3	2	1
Q	da	O	O	O	II	O	O	O	I	I	I	I	I	IV	V	VI	VII	IX	IX
	d	O	I	O	II	O	O	O	I	I	I	III	I	IV	V	VI	VII	IX	IX
	a	O	O	I	II	O	O	O	I	I	I	I	III	IV	V	VI	VII	IX	IX
	0	II	II	II	IV	I	O	I	II	III	III	III	III	IV	V	VI	VII	IX	IX
P	5	O	O	O	I	O	O	O	O	O	I	I	I	IV	V	VI	VI	VII	VIII
	4	O	O	O	O	O	I	I	II	II	III	III	III	IV	V	VI	VI	VII	VIII
	3	O	O	O	I	O	I	I	II	II	II	II	II	IV	IV	V	V	VI	VII
	2	I	I	I	II	O	II	II	II	II	II	II	II	III	IV	IV	V	VI	VII
	1	I	I	I	III	O	II	II	II	II	II	II	II	III	IV	IV	IV	V	VI
N	da	I	I	I	III	I	III	II	II	II	II	II	II	IV	IV	V	VI	VI	VI
	d	I	III	I	III	I	III	II	II	II	II	III	II	IV	IV	V	VI	VI	VI
	a	I	I	III	III	I	III	II	II	II	II	II	III	IV	IV	V	VI	VI	VI
	0	IV	IV	IV	IV	IV	IV	IV	III	III	IV	IV	IV	IV	IV	IV	IV	V	VI
C	5	V	V	V	V	V	V	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	V	V
	4	VI	VI	VI	VI	VI	VI	V	IV	IV	V	V	V	IV	IV	IV	IV	V	V
	3	VII	VII	VII	VII	VI	VI	V	V	IV	VI	VI	VI	IV	IV	IV	IV	IV	IV
	2	IX	IX	IX	IX	VII	VII	VI	VI	V	VI	VI	VI	V	V	V	IV	IV	IV
	1	IX	IX	IX	IX	VIII	VIII	VII	VII	VI	VI	VI	VI	VI	V	V	IV	IV	IV

[a] Level of interaction indicates the well depth in the LJ potential: O, $\epsilon = 5.6$ kJ/mol; I, $\epsilon = 5.0$ kJ/mol; II, $\epsilon = 4.5$ kJ/mol; III, $\epsilon = 4.0$ kJ/mol; IV, $\epsilon = 3.5$ kJ/mol; V, $\epsilon = 3.1$ kJ/mol; VI, $\epsilon = 2.7$ kJ/mol; VII, $\epsilon = 2.3$ kJ/mol; VIII, $\epsilon = 2.0$ kJ/mol; IX, $\epsilon = 2.0$ kJ/mol. The LJ parameter $\sigma = 0.47$ nm for all interacion levels except level IX for which $\sigma = 0.62$ nm. Four different CG sites are considered: charged (Q), polar (P), nonpolar (N), and apolar (C). Subscripts are used to further distinguish groups with different chemical nature: 0, no hydrogen-bonding capabilities are present; d, groups acting as hydrogen bond donor; a, groups acting as hydrogen bond acceptor; da, groups with both donor and acceptor options; 1−5, indicating increasing polar affinity.

- **Deriving the parameters in the matrix by the same procedure that was used to derive the parameters for the bonded interactions**

CG Simulation: Force Field Parameterization

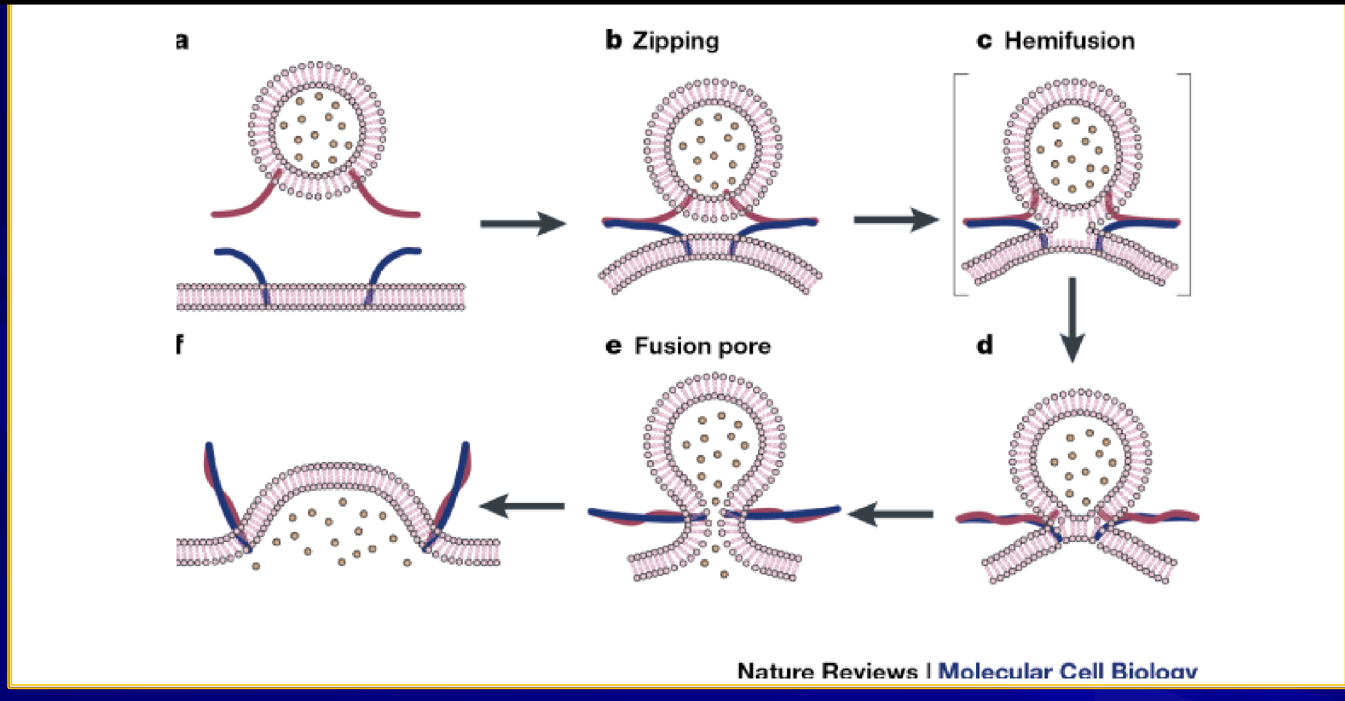
Non-bonded Interactions: Coulombic Potential

$$V_{el} = \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_{rel} r_{ij}}$$

- **charged groups (type Q) bearing a charge q interact via a Coulombic energy function with a relative dielectric constant εrel) 15 for explicit screening.**

CG Simulation: Application

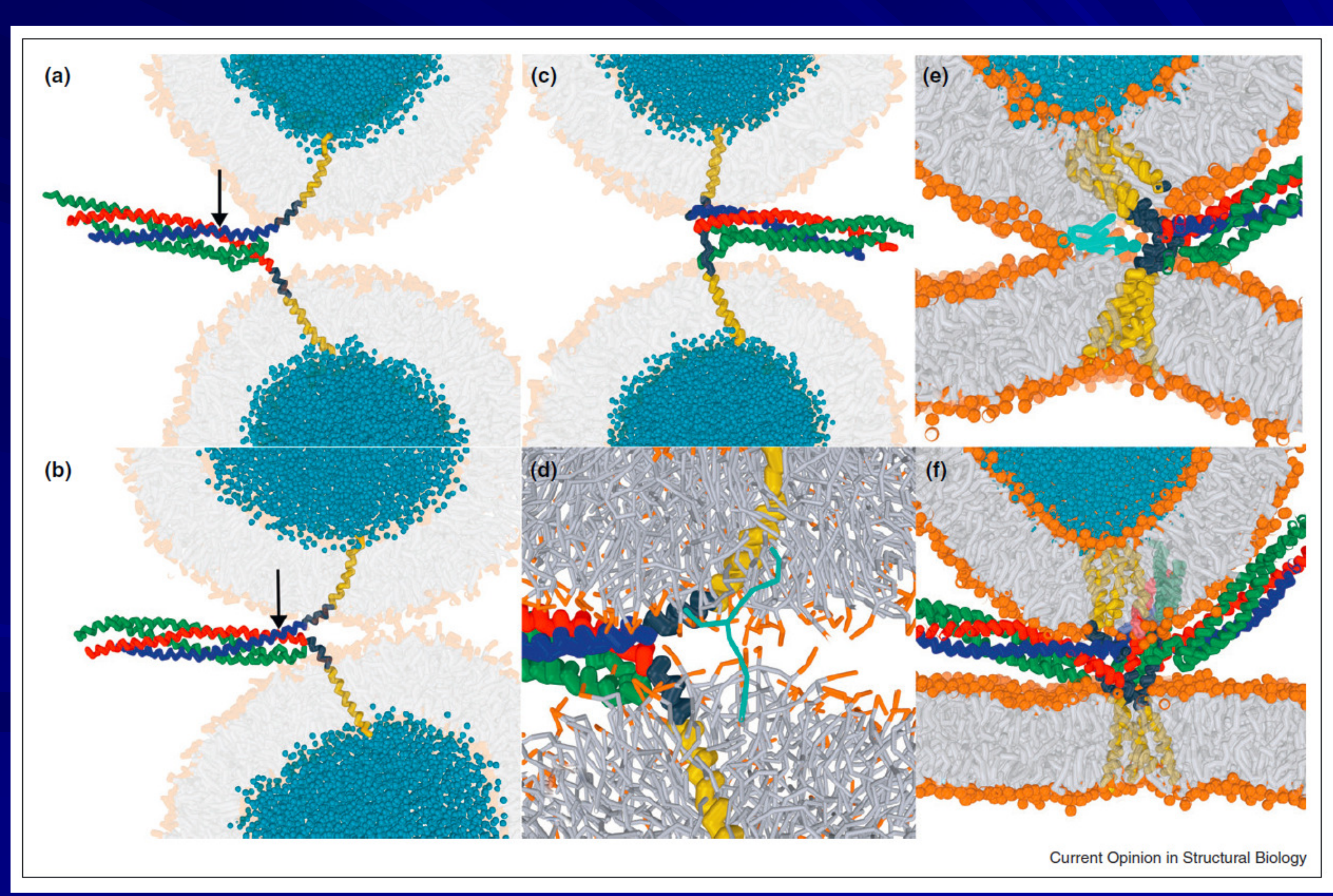
SNARE-mediated membrane fusion



a) The two membranes are in the vicinity of each other but the SNAREs are not yet in contact.
b) SNARE complexes start zipping from the amino-terminal end, which draws the two membranes further towards each other.
c) Zipping proceeds, causing increased curvature and lateral tension of the membranes, exposing the bilayer interior.
d) The highly unfavorable void space at the membrane junction in (c) causes the establishment of contacts between the distal membrane leaflets.
e) The lateral tension in the trans-bilayer contact area induces membrane breakdown, yielding a fusion pore.
f) The fusion pore expands and the membrane relaxes.

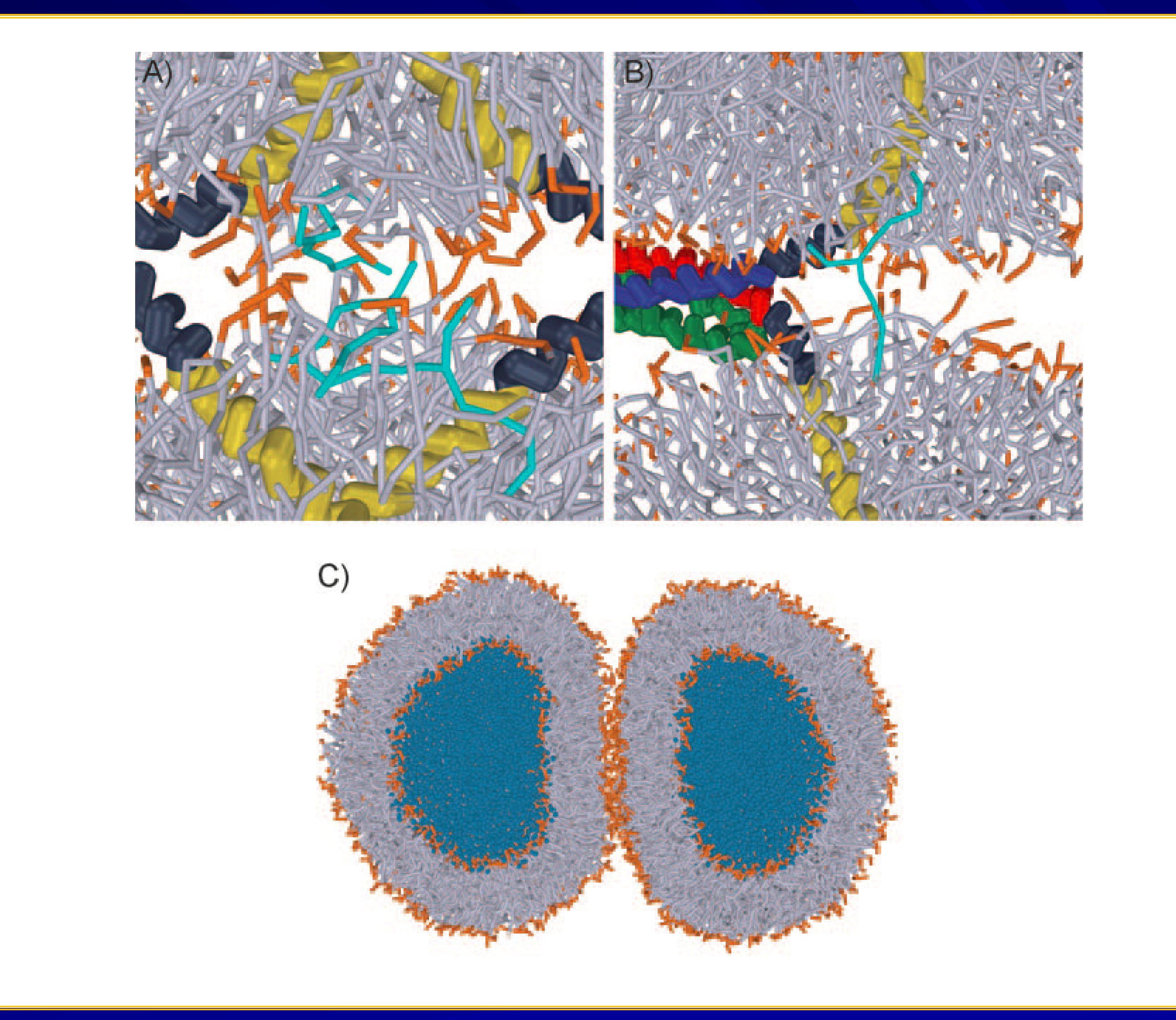
CG Simulation: Application

SNARE-mediated membrane fusion: initial set-up



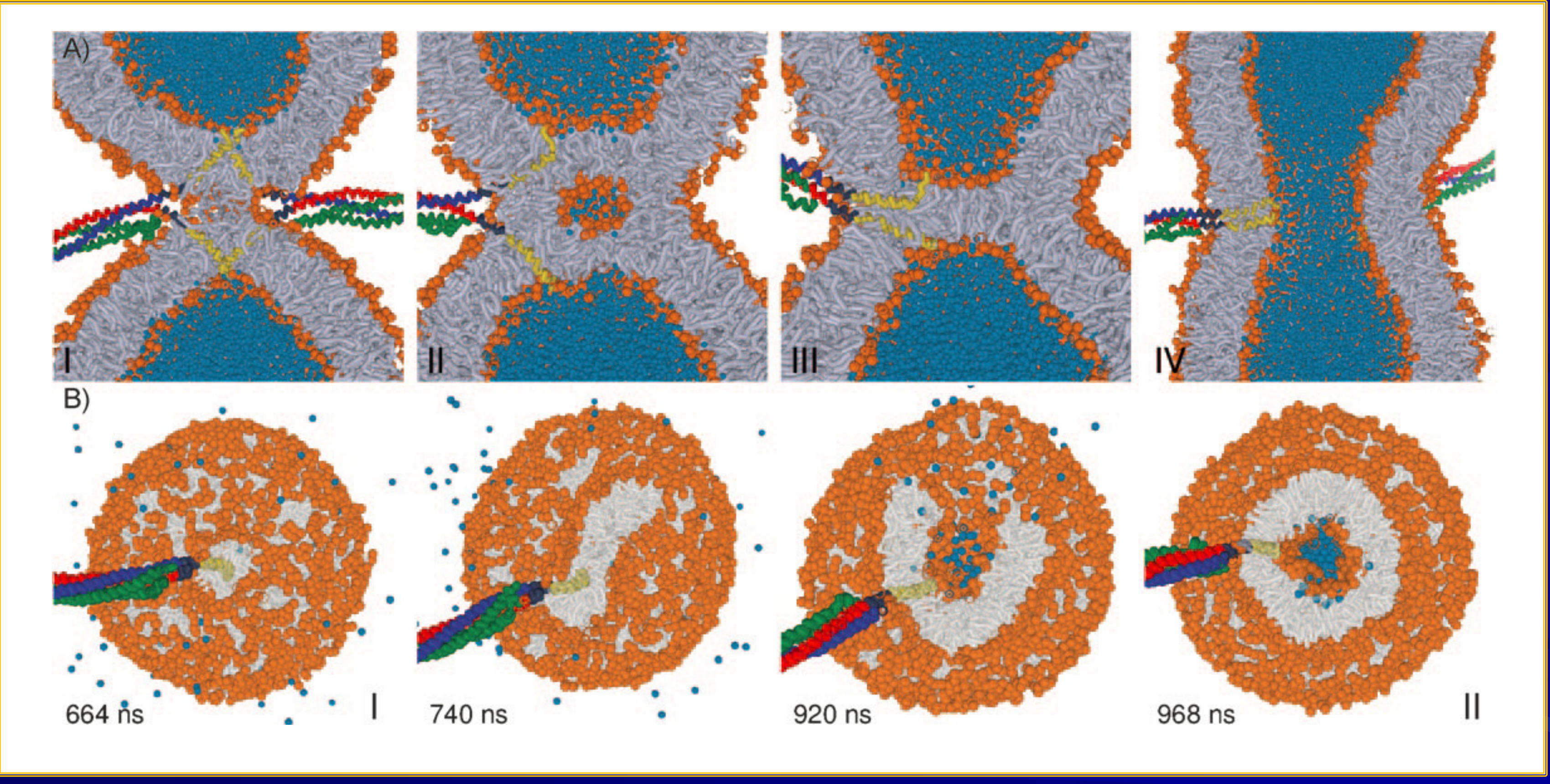
CG Simulation: Application

SNARE-mediated membrane fusion: stalk formation during simulation



CG Simulation: Application

SNARE-mediated membrane fusion: four stages of fusion



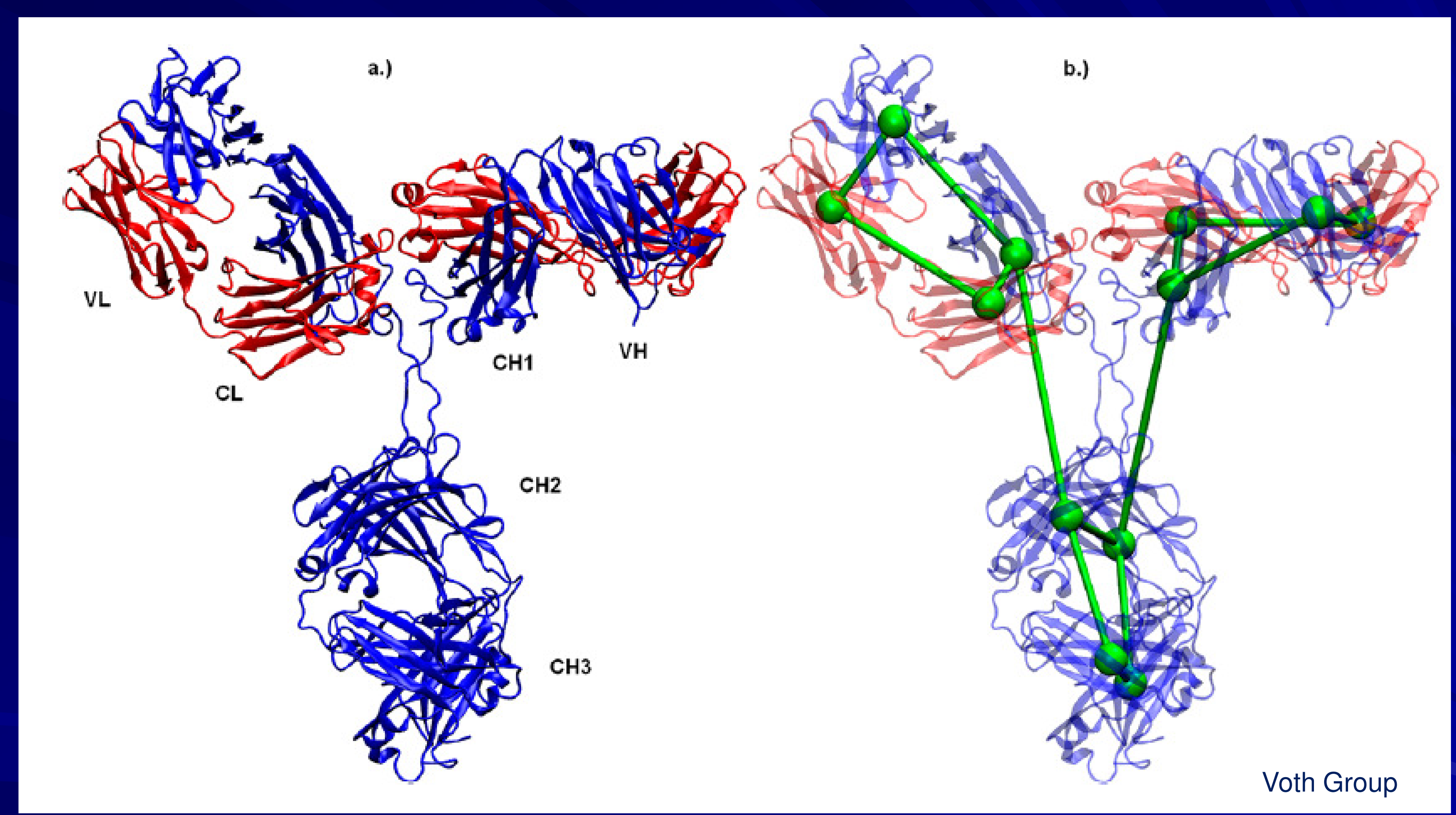
A) I: Stalk, II : inverted micelle intermediate (IMI; single SNARE complex), III: Hemifusion-diaphragm (single SNARE complex), IV: fusion. B) Cross-sections of the fusion plane showing the transition from stage I (stalk) to stage II (inverted micelle) in more detail. The stalk (664 ns), shows a worm-like expansion (740 ns) and eventually, after bending (920 ns), forms a closed ring that encapsulates exterior solvent forming an inverted micelle (968 ns).

CG Simulation: Application

SNARE-mediated membrane fusion: conclusions

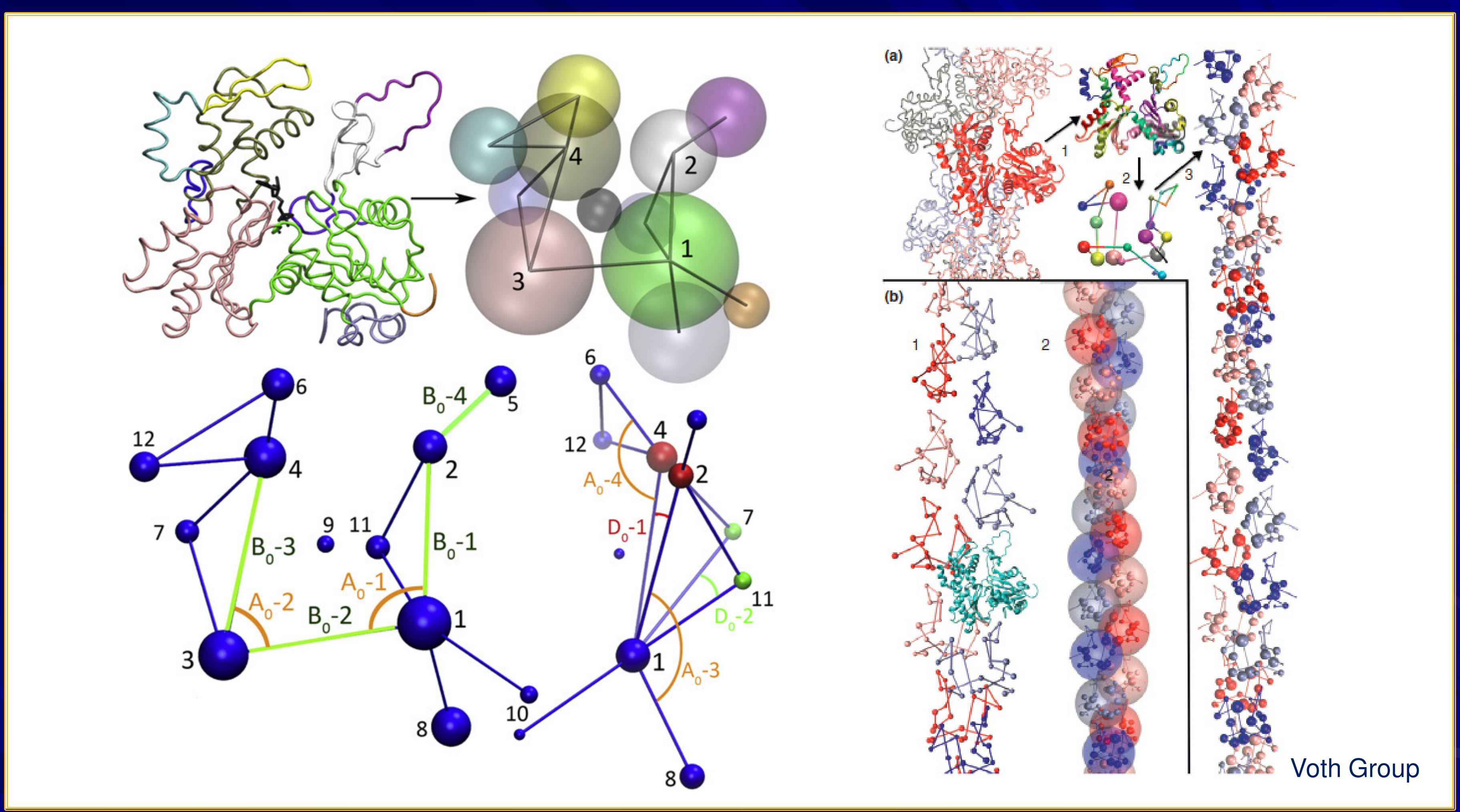
- SNARE complex is driven towards the perimeter of the fusion region
- SNARE complexes induce stalk formation and actively opens fusion pores
- Intrinsic properties of trans-membrane regions facilitate stalk formation
- SNARE complexes release their energy throughout the whole fusion process up to the expansion of the fusion pore

Beyond MARTINI



Domain-based Coarse-graining

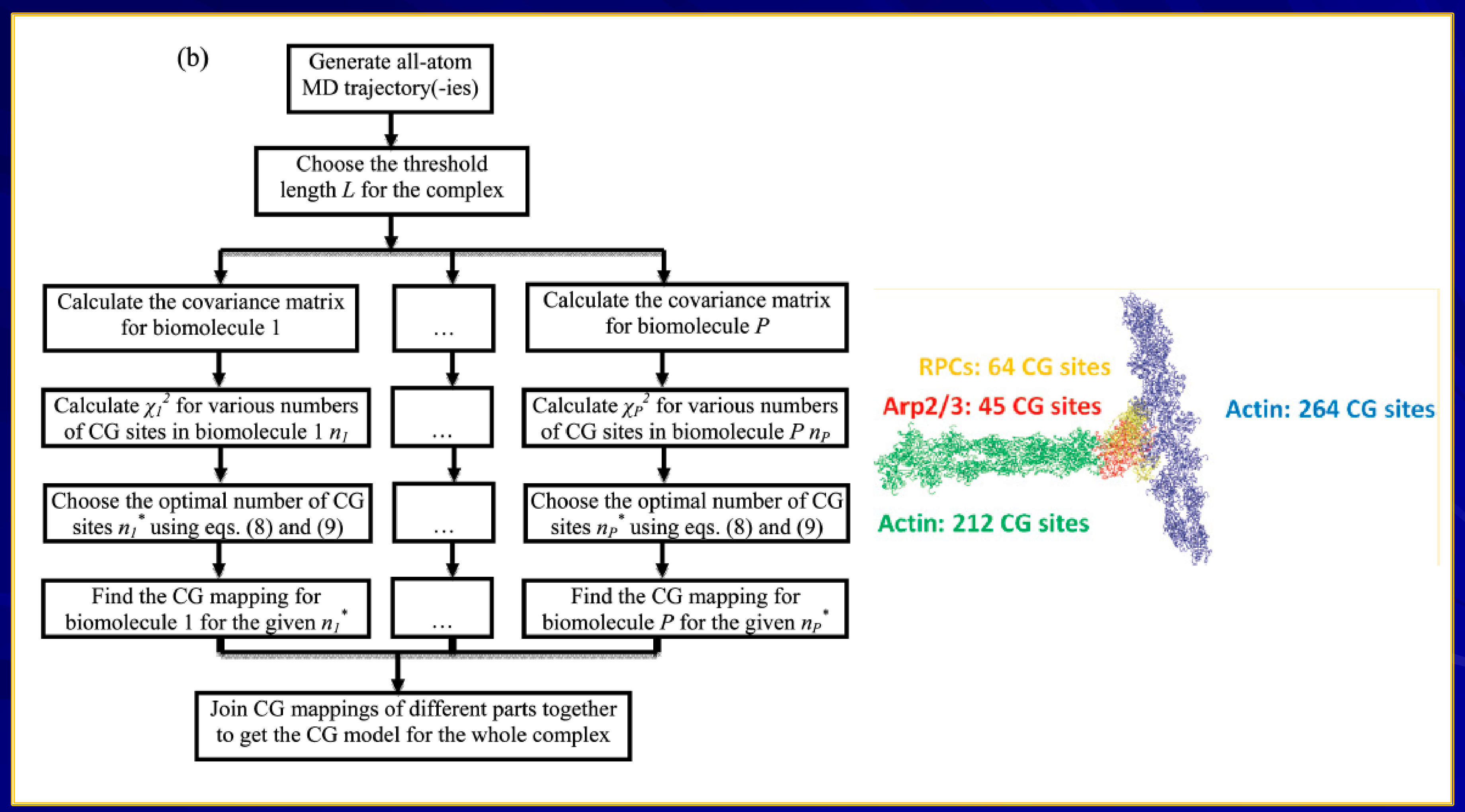
Beyond MARTINI



Coarse-graining the molecule into functional units

Beyond MARTINI

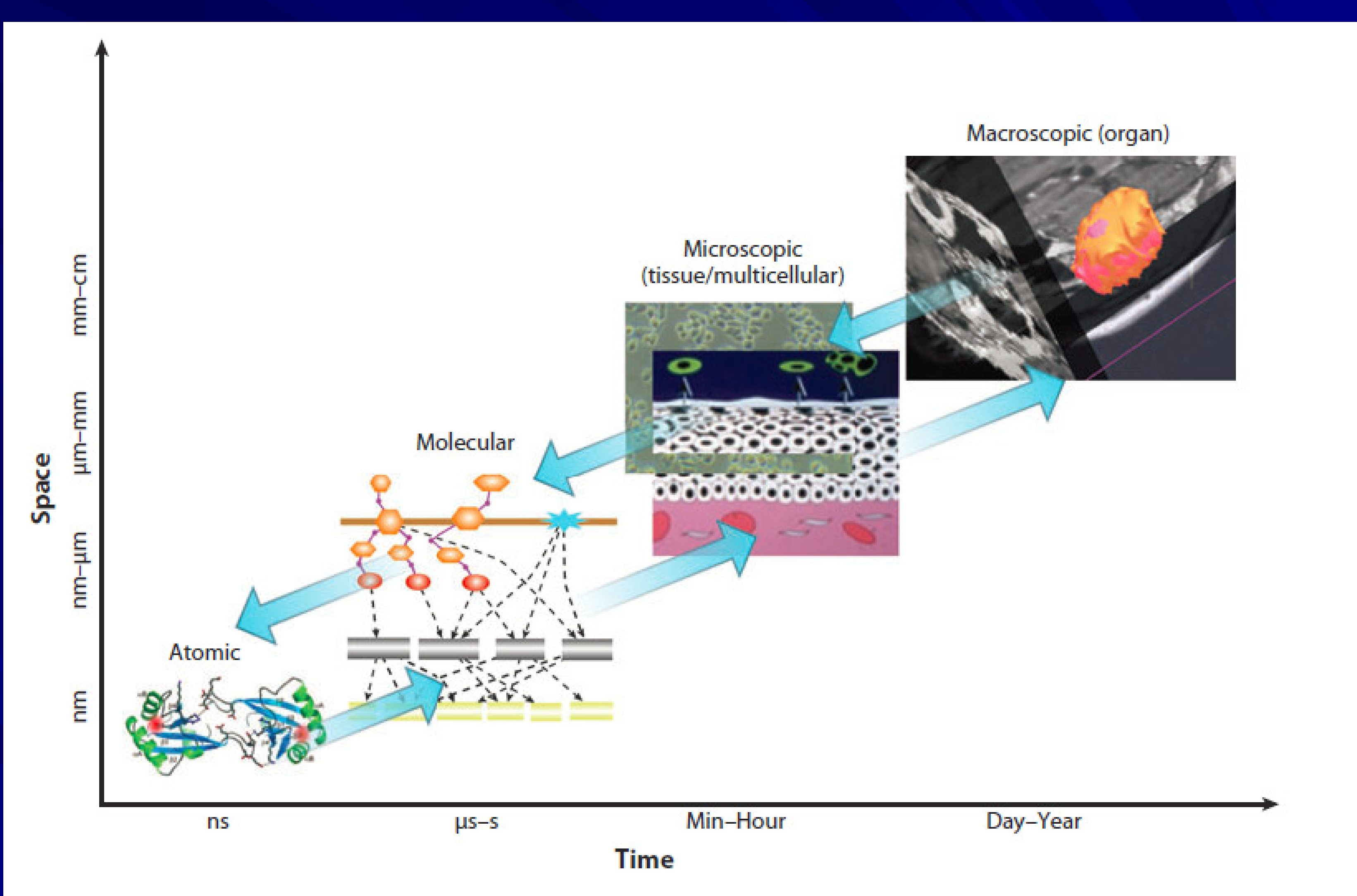
Optimizing the number of coarse-grained sites in different components of large bio-molecular complexes

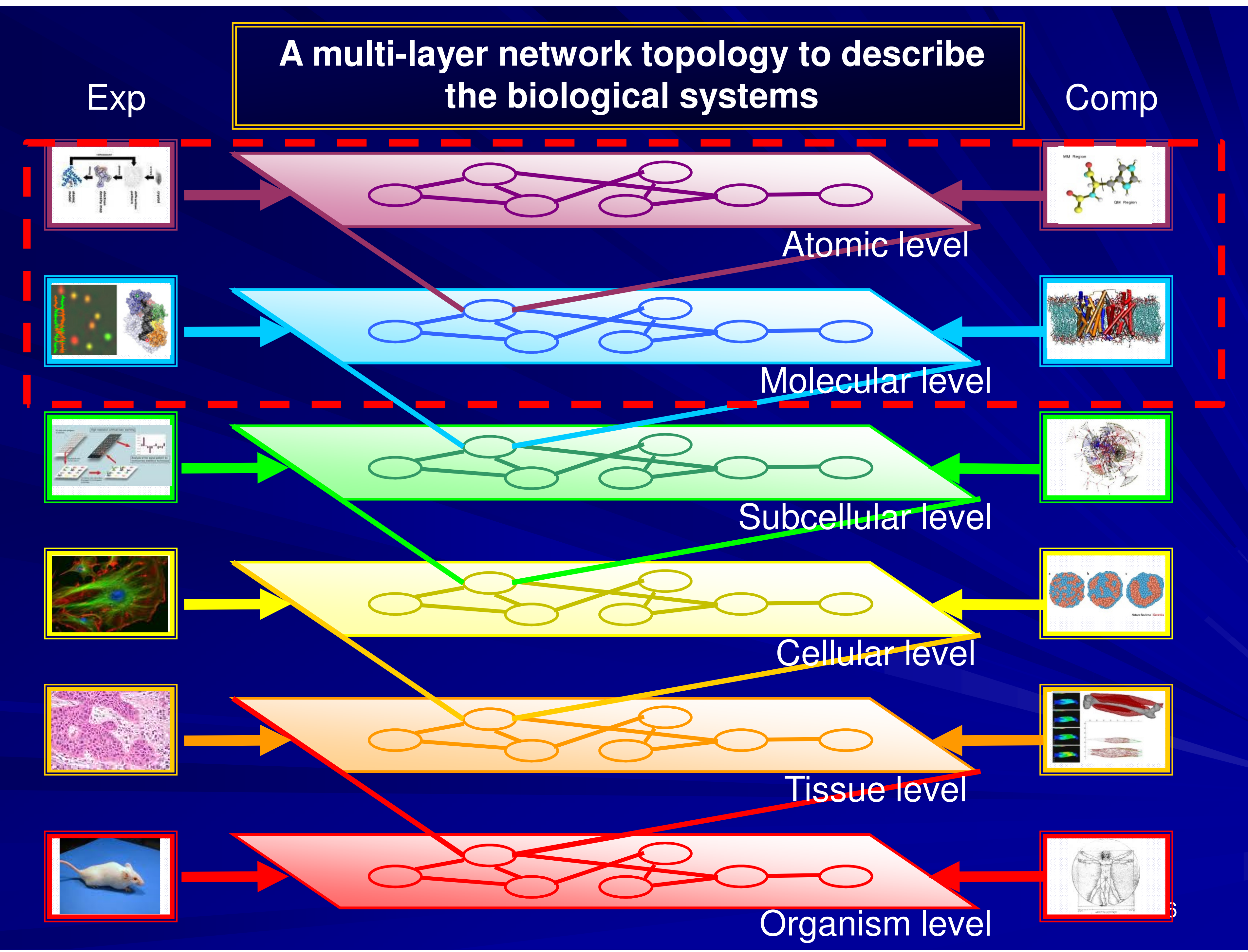


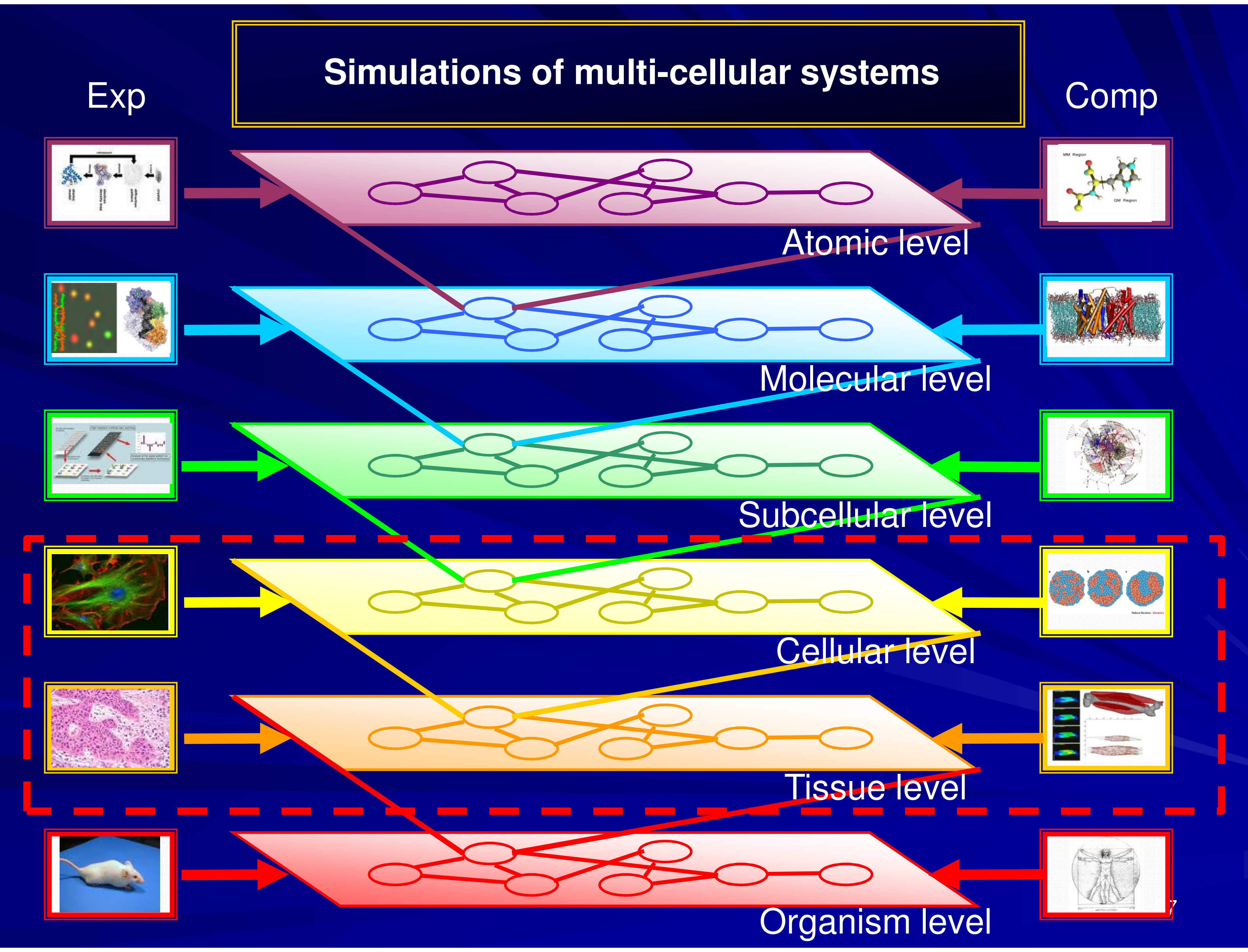
Outline

- Modeling and Simulating Larger Molecular Systems with Longer Time Scales
  - Coarse-grained simulation
  - Multi-scale modeling
- Extend the Applications of Structural Biology to Newly Emerging Fields
  - Fold proteins by information from genetics
  - Simulating crowded cellular environment
  - Combining structural biology with systems biology
- Summary and Perspective

A multi-scale nature of biological systems

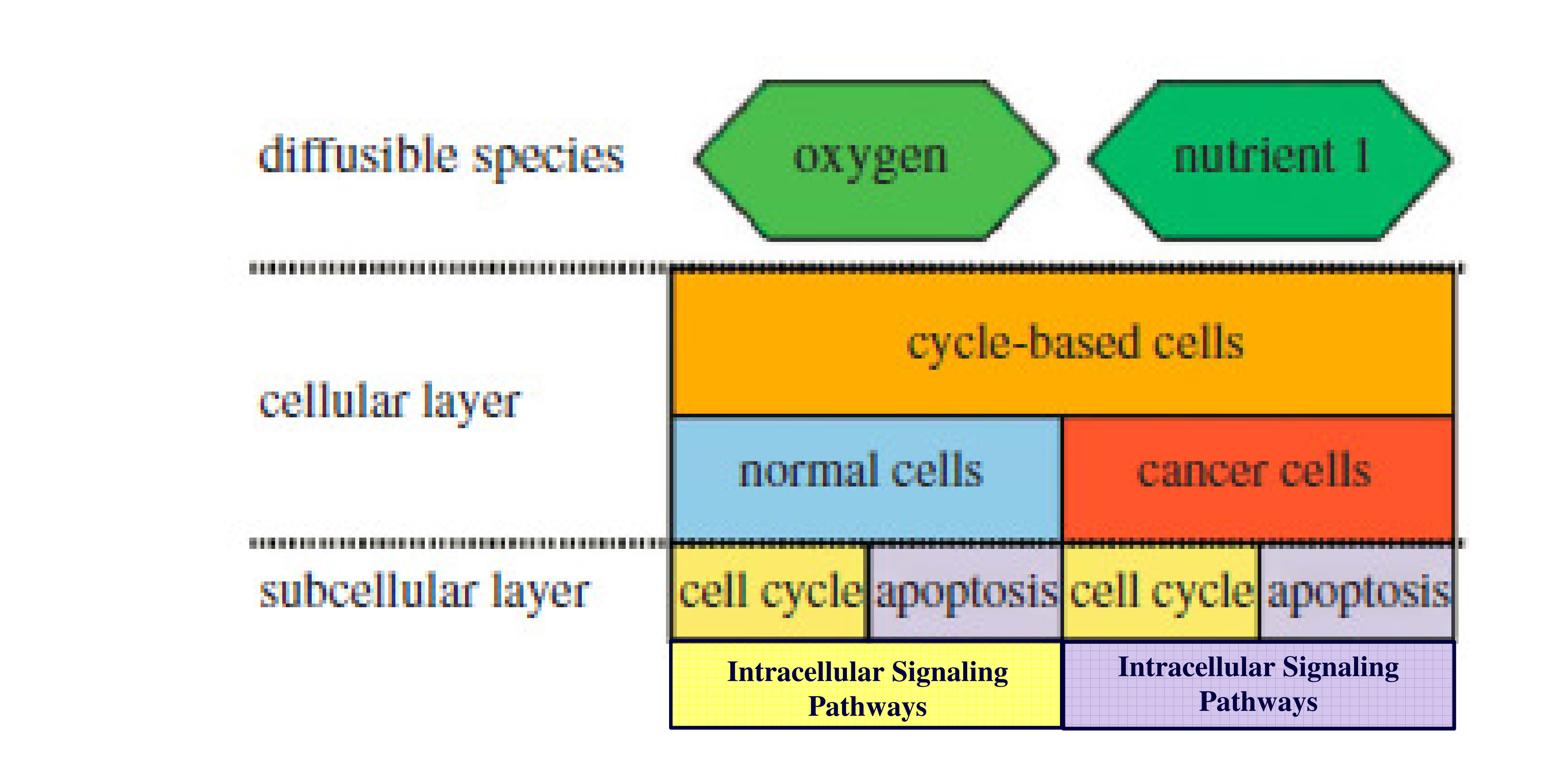



A multi-layer network topology to describe the biological systems
Exp
Comp
Atomic level
Molecular level
Subcellular level
Cellular level
Tissue level
Organism level


Simulations of multi-cellular systems
Exp
Comp
Atomic level
Molecular level
Subcellular level
Cellular level
Tissue level
Organism level

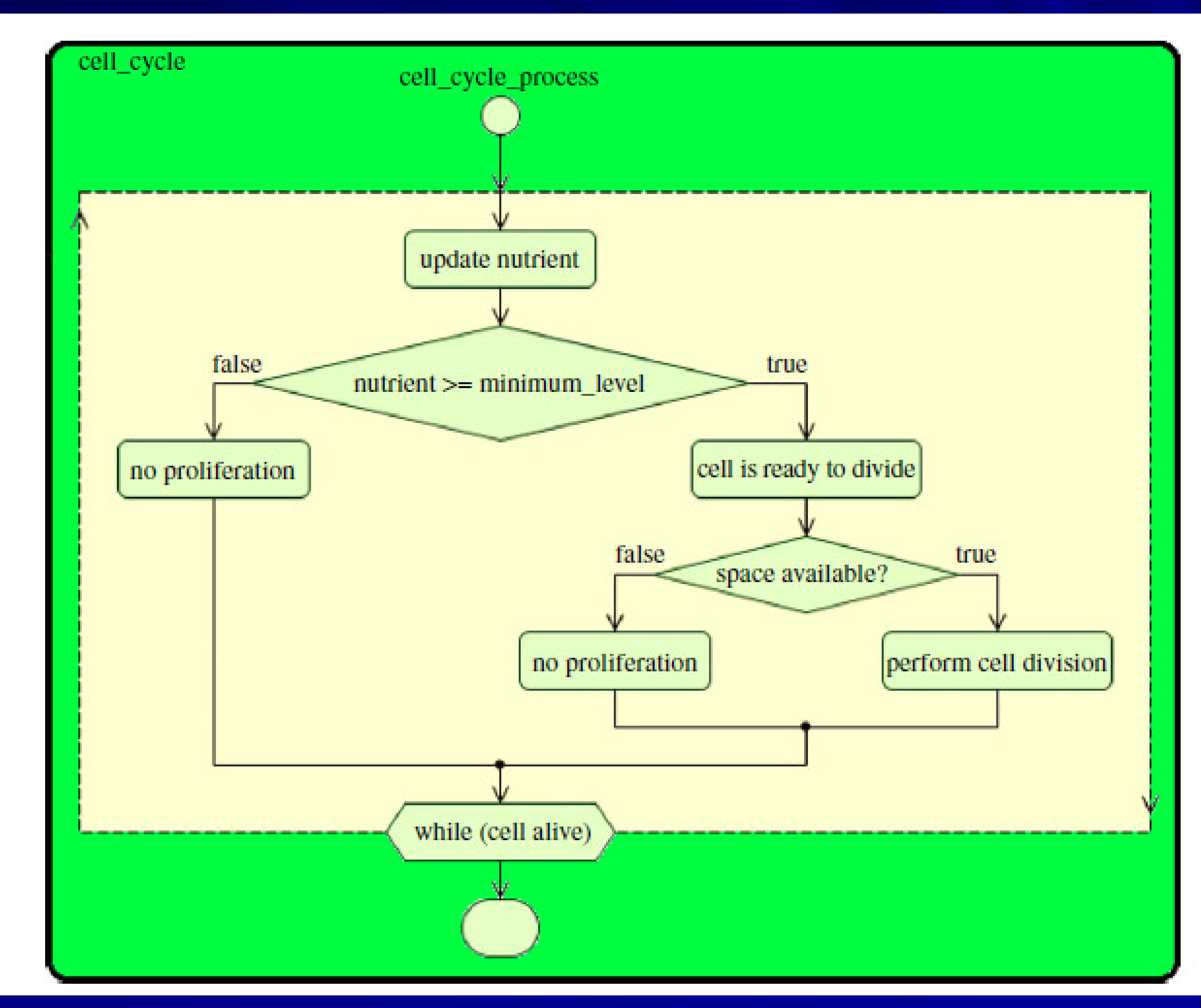
Multi-scale Modeling: Cellular Simulation

The multi-scale model of cell population growth



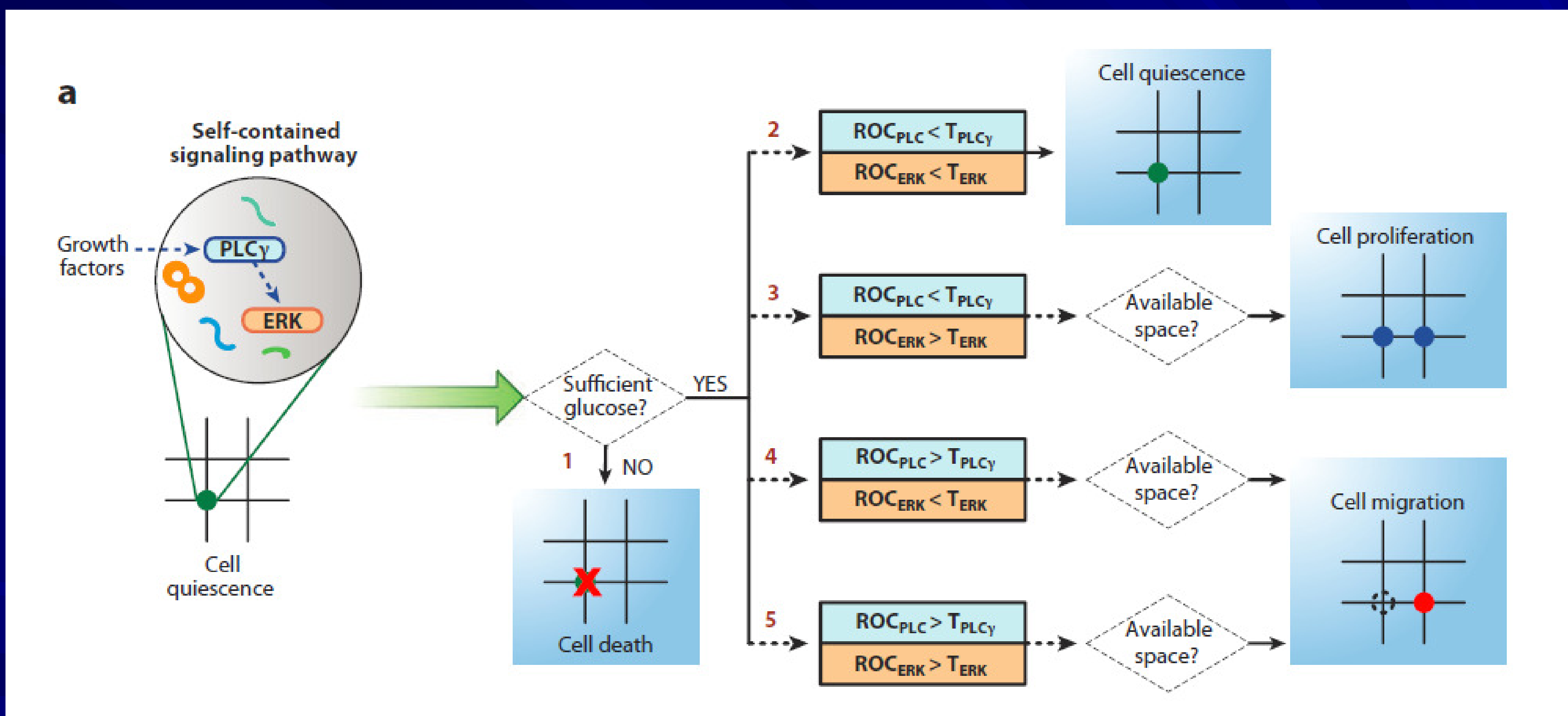
Multi-scale Modeling: Cellular Simulation

Moving Sets



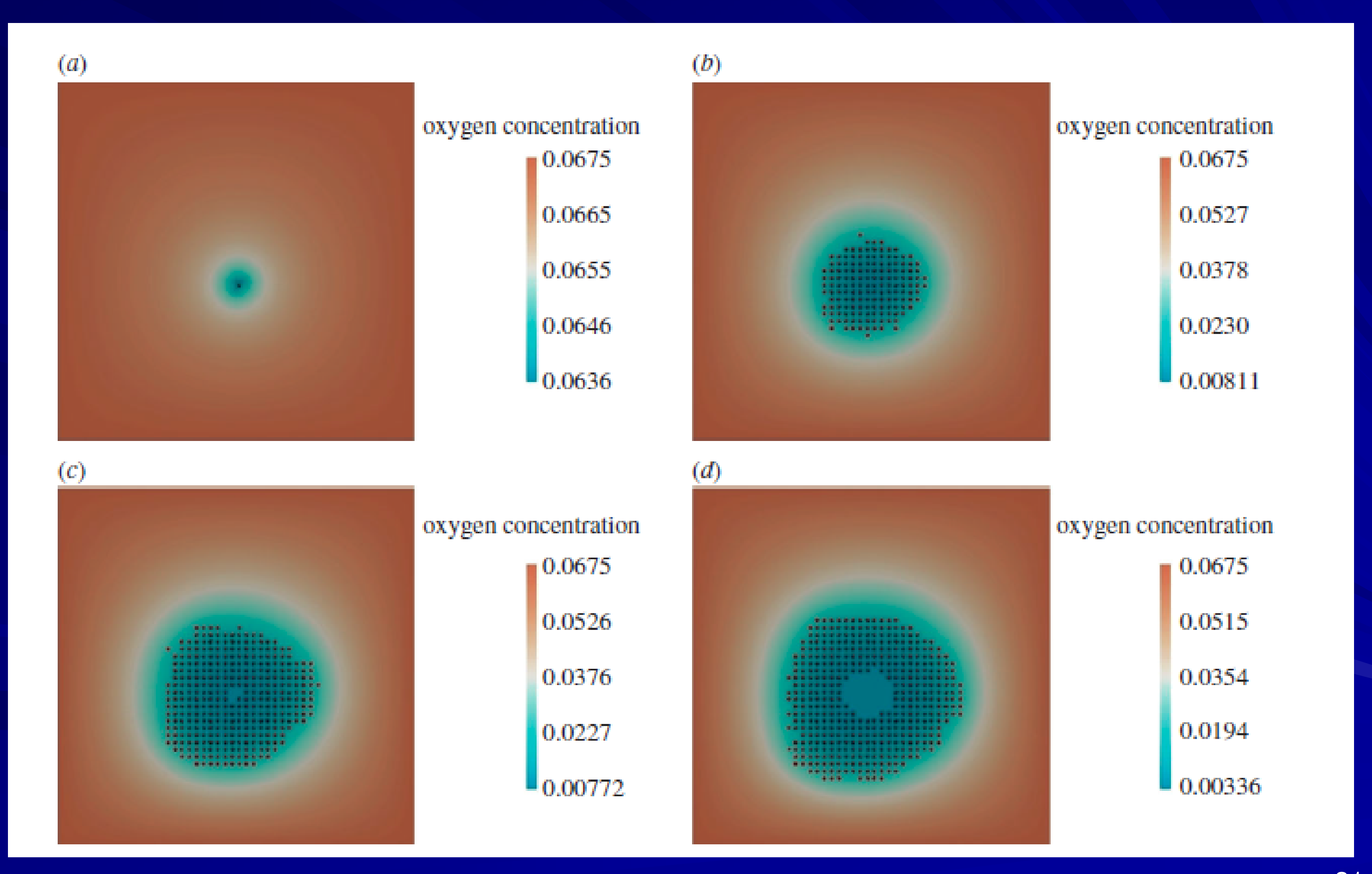
Multi-scale Modeling: Cellular Simulation

Lattice-based Simulations



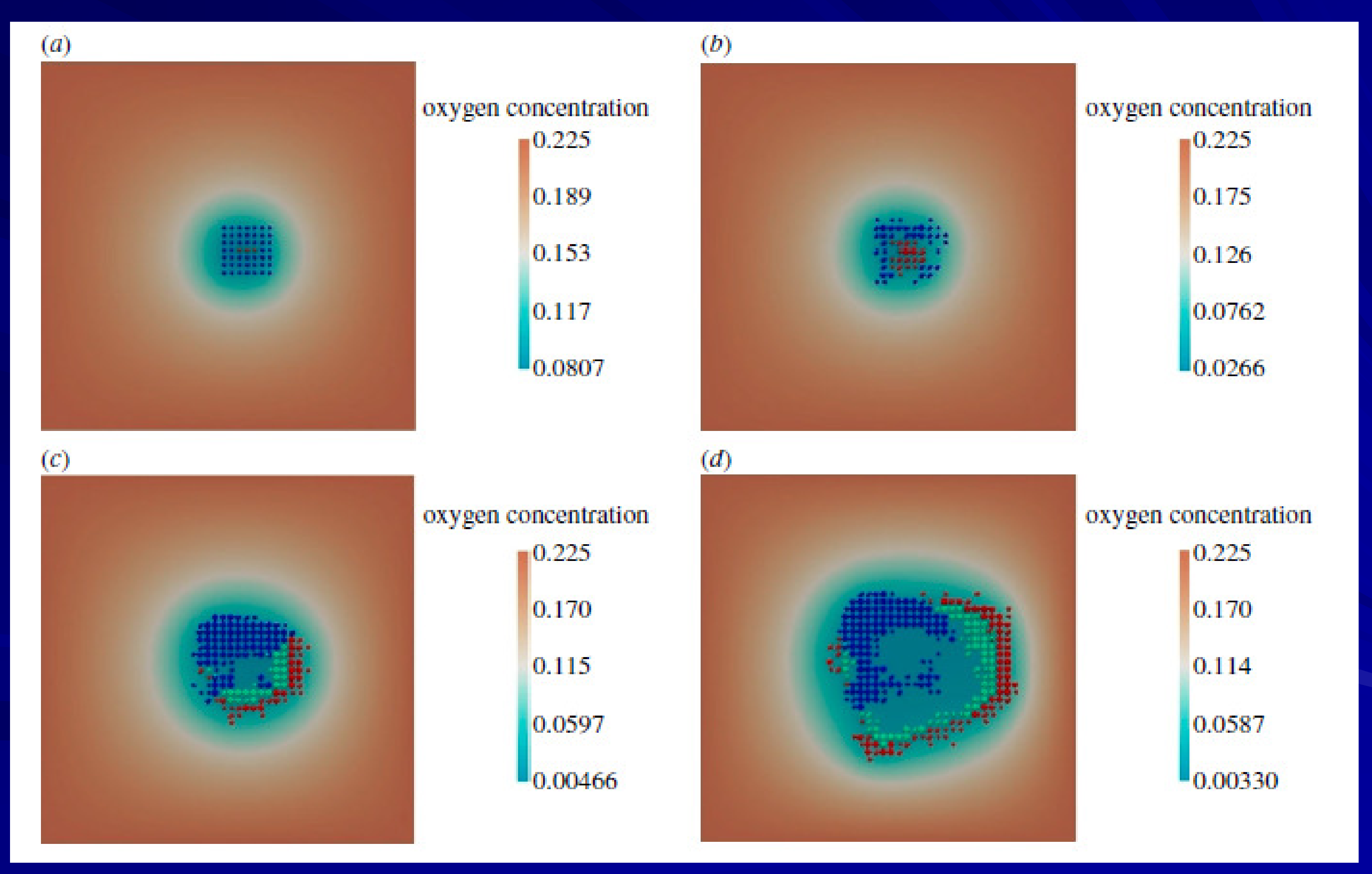
Multi-scale Modeling: Cellular Simulation

Simulation results of normal cells



Multi-scale Modeling: Cellular Simulation

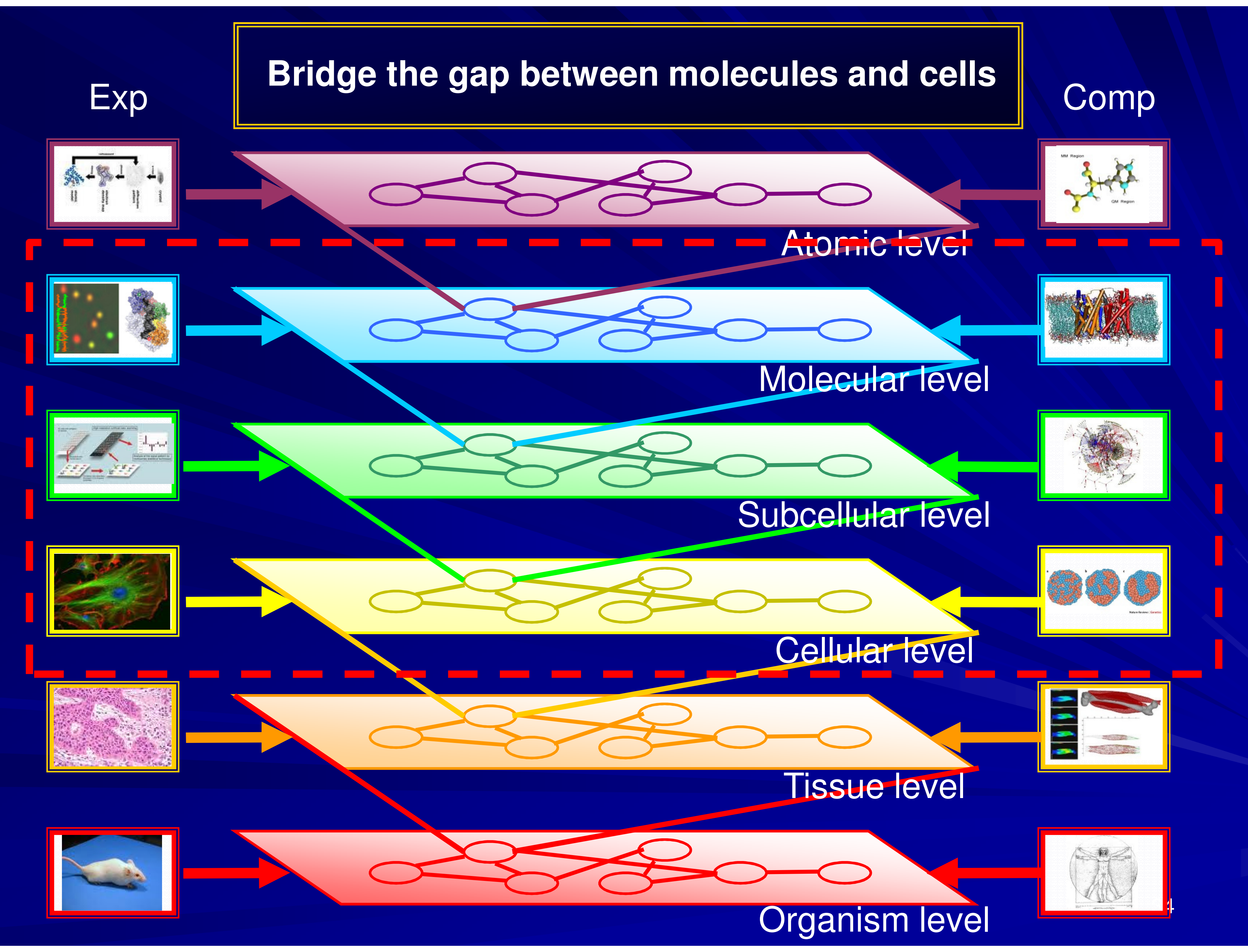
Simulation results of tumor cells



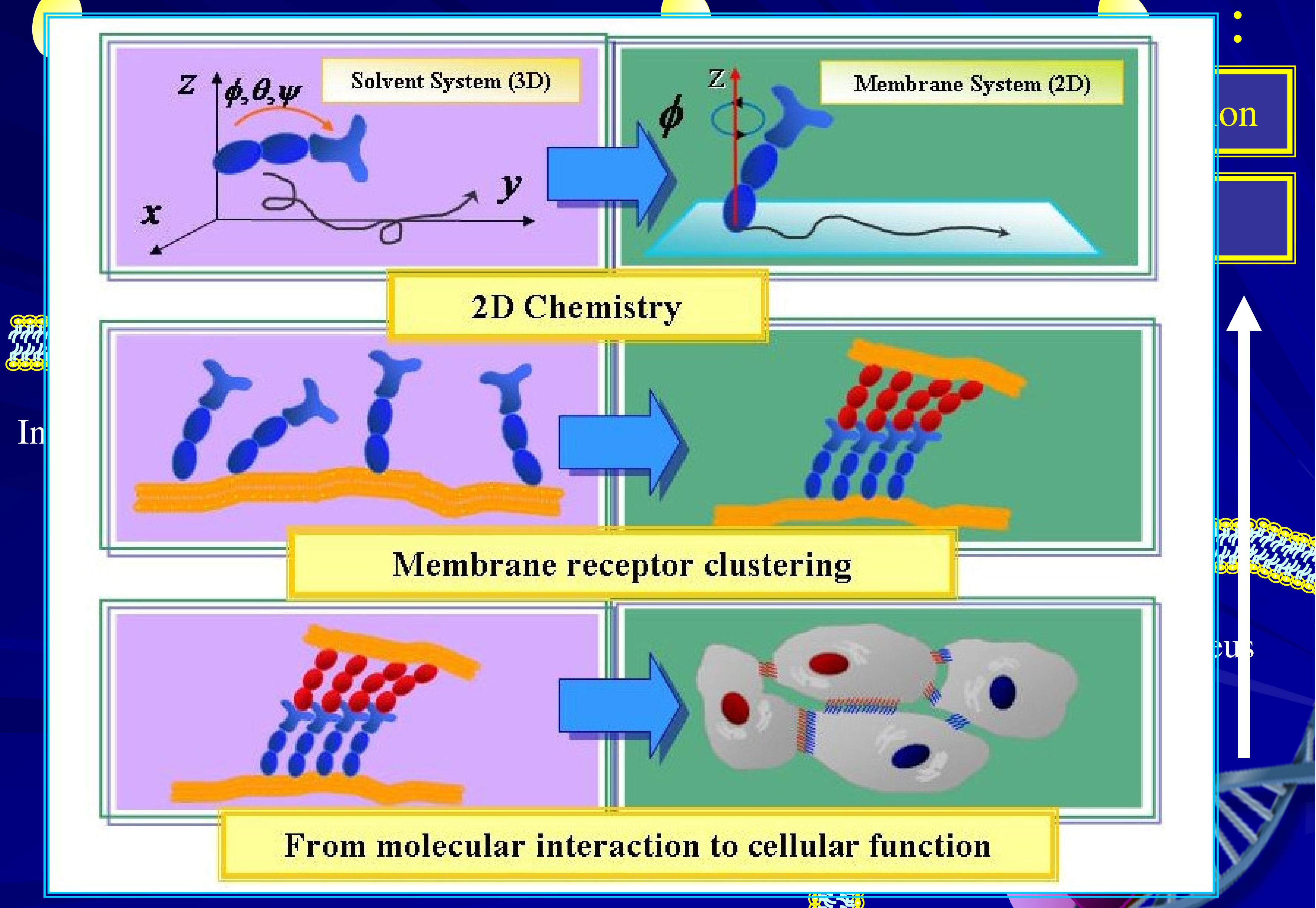
Multi-scale Modeling: Cellular Simulation

Conclusions

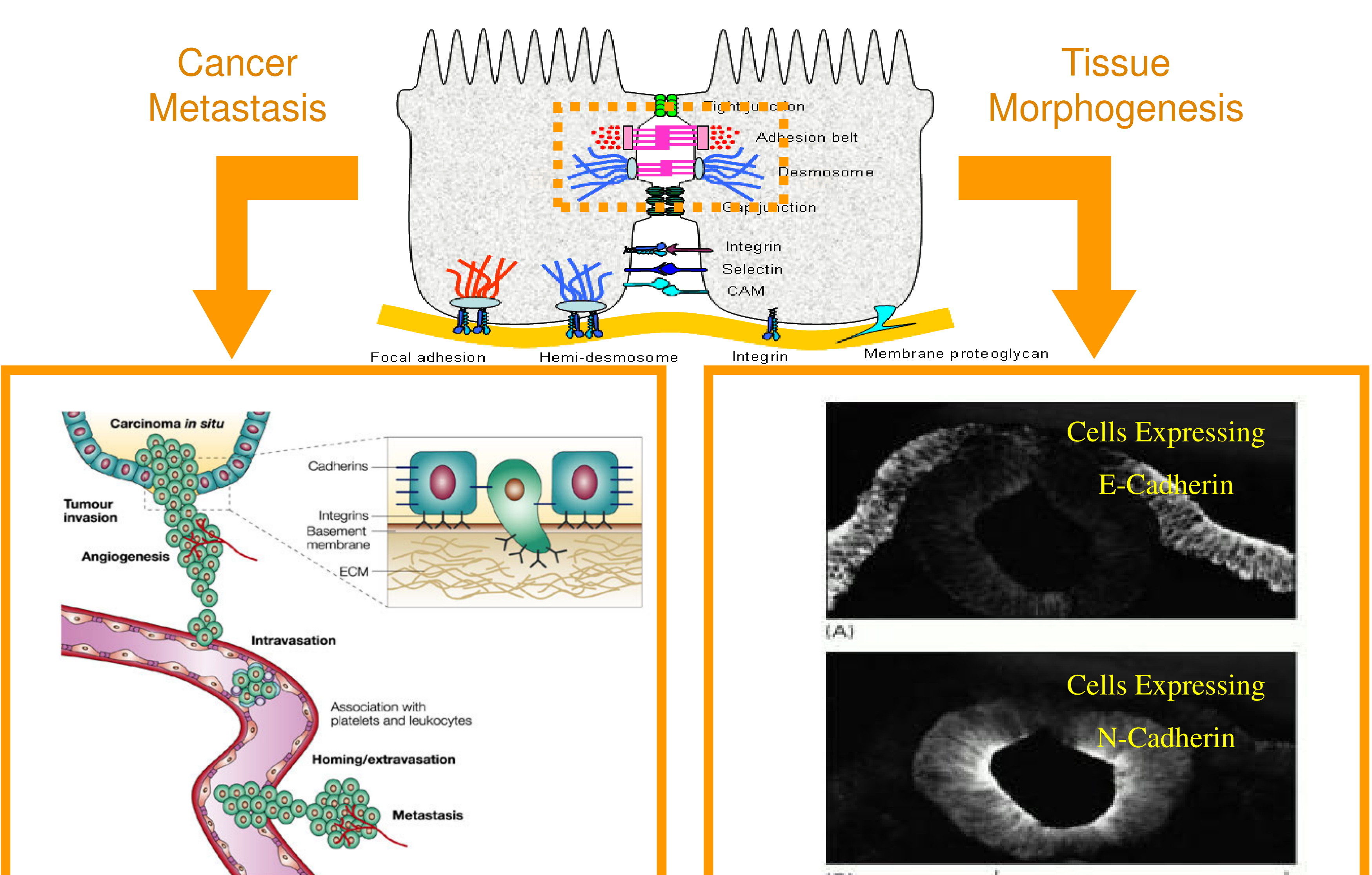
- The multi-scale simulations show the dynamics of cell populations and their responses to nutrient shortage.
- Tumour cells take over the space originally occupied by normal cells and keep growing in the hypoxic areas, owing to their higher rates of oxygen consumption.


Bridge the gap between molecules and cells
Exp
Comp
Atomic level
Molecular level
Subcellular level
Cellular level
Tissue level
Organism level

Background



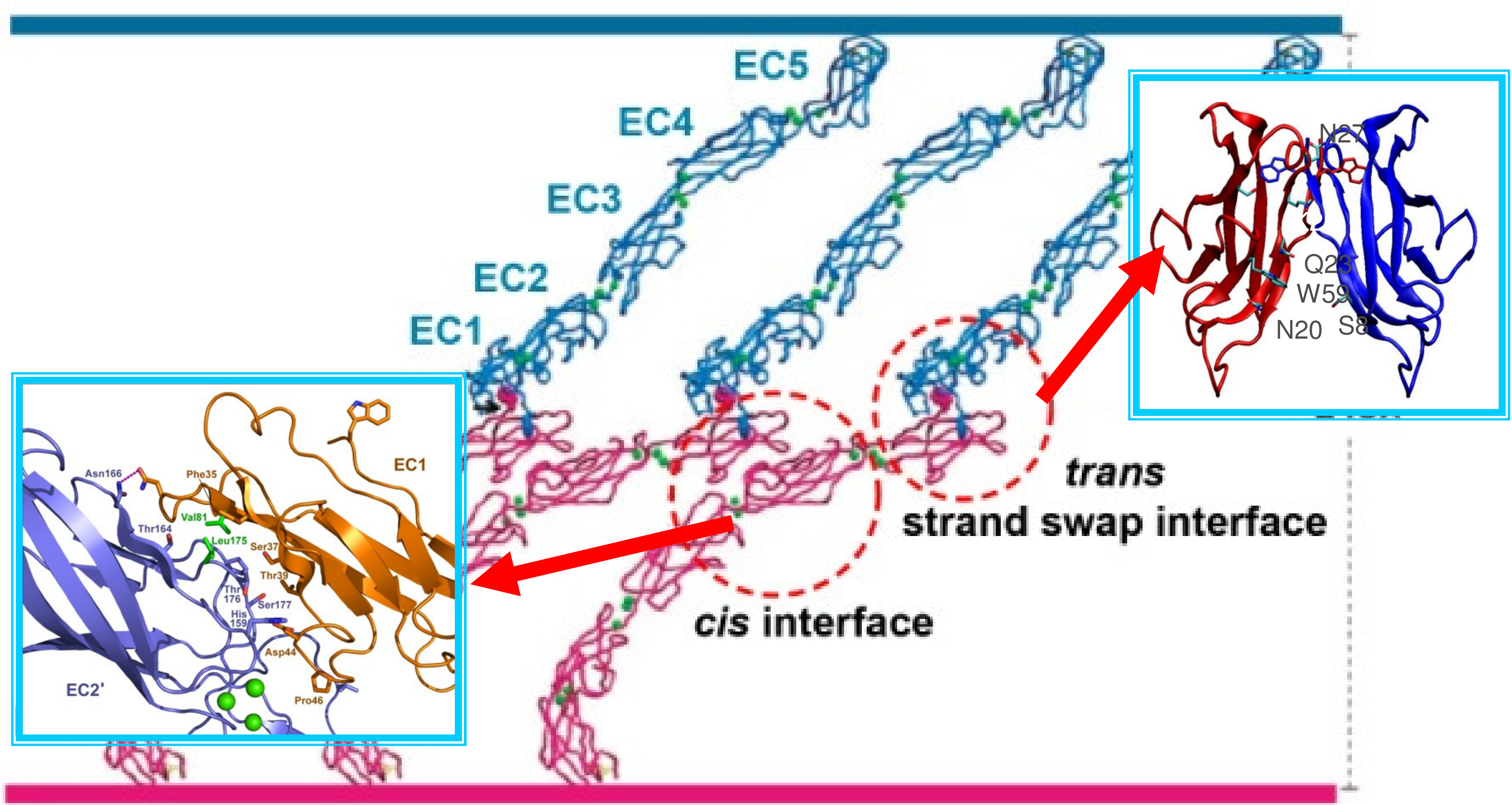
Cadherin-mediated Cell Adhesion



Nature Reviews Molecular Cell Biology, October, (2004).

K. Hatta e.t. (1986) Nature.

Molecular Structure of Type-I Classical Cadherin

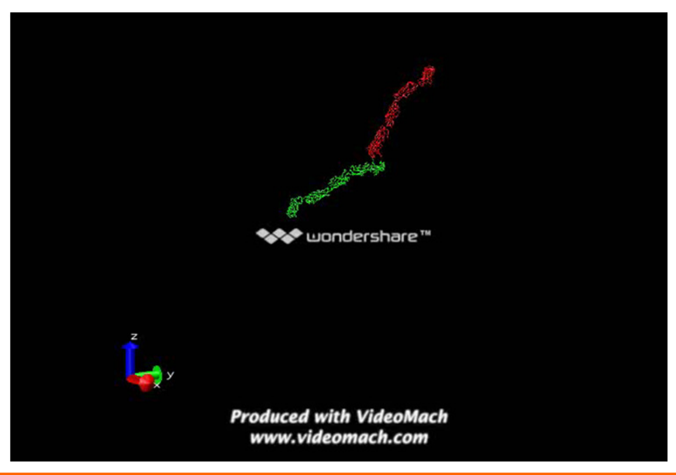


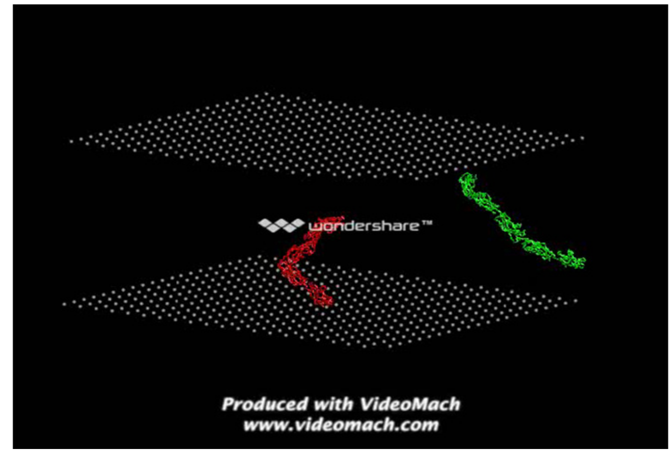
AUC can only measure 3D binding affinity

Cis binding too weak to be measured

Jin X, Harrison O, et. al., *Structure* (2011)

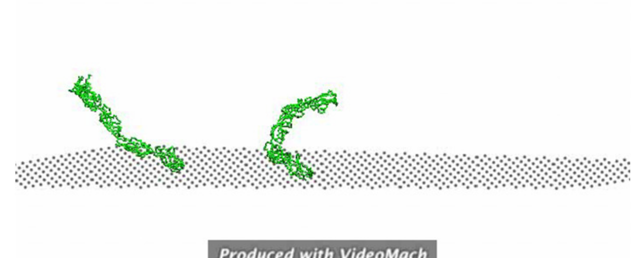
From 3D to 2D Binding Affinity: Theory and Simulations

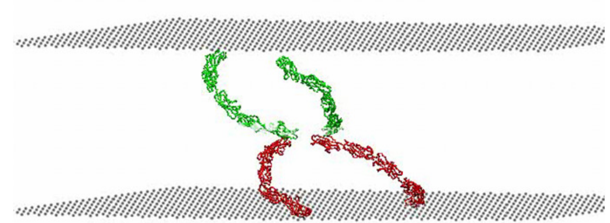




$$\Delta G^{(2D)}(trans) = \Delta G^{(3D)}(trans) - RT\left[\ln\left(\frac{C_0 h_M^2}{\rho_0 h_T}\right) + \ln\left(\frac{(\Delta\psi_M(1-\cos\Delta\theta_M))^2}{4\pi\Delta\psi_T(1-\cos\Delta\theta_T)}\right)\right]$$

$$\Delta G(3D) = 8.9RT$$
$$\Delta G(2D) = 5.7RT$$

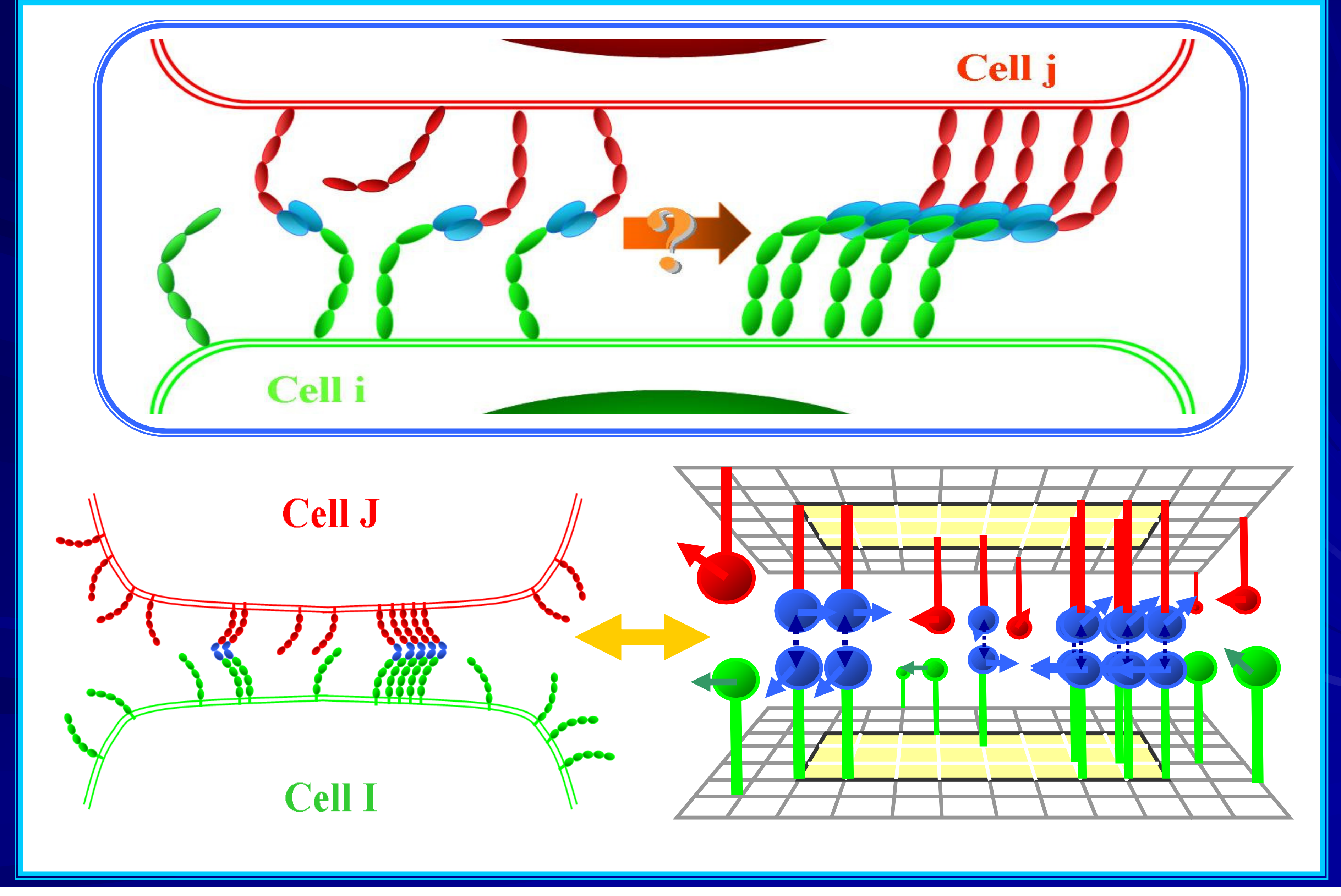




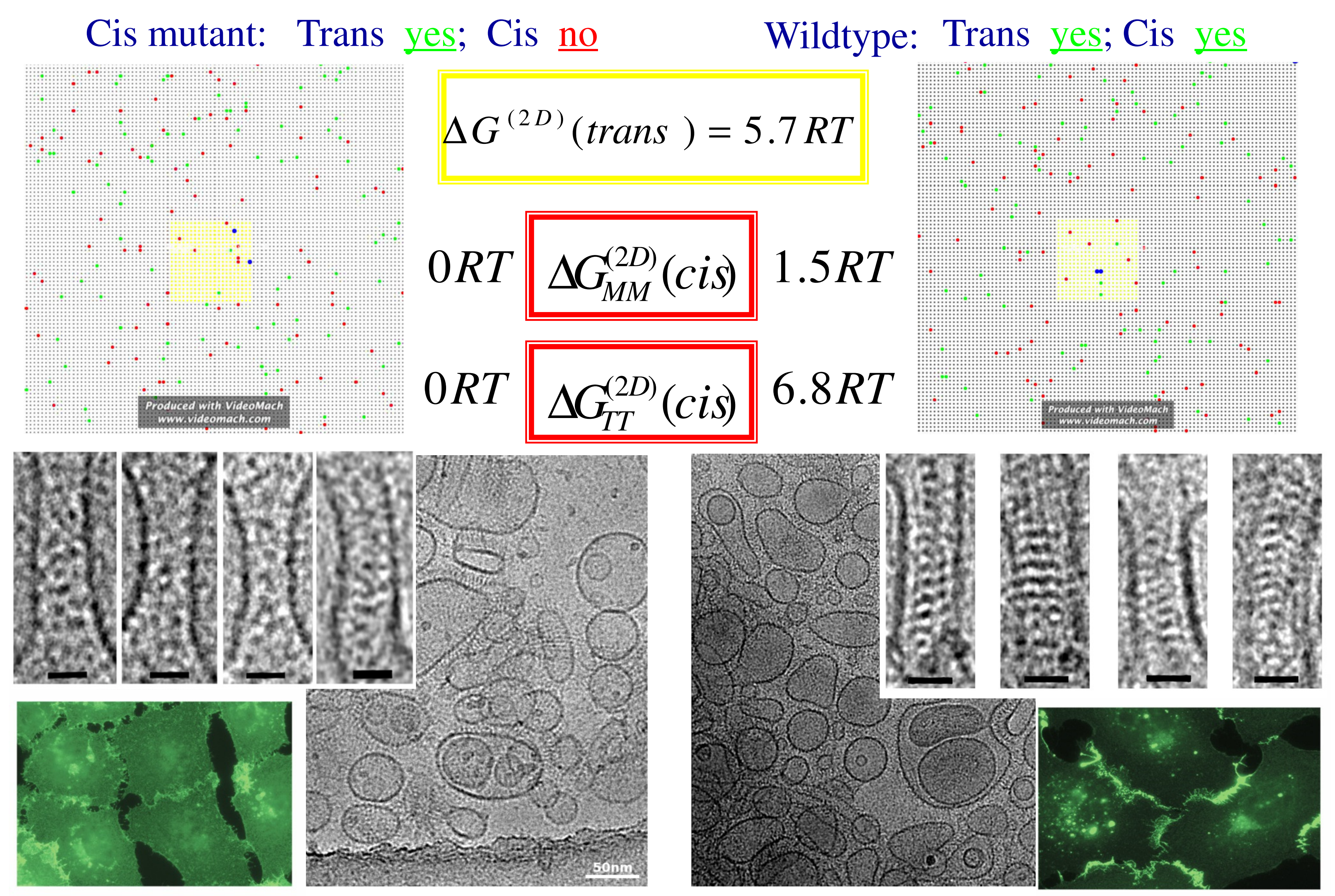
$$\Delta G_{TT}^{(2D)}(cis) - \Delta G_{MM}^{(2D)}(cis) = 2RT\ln\left(\frac{\Delta\psi_M(1-\cos\Delta\theta_M)h_M}{\Delta\psi_T(1-\cos\Delta\theta_T)h_T}\right)$$

$$\Delta G_{MM}^{(2D)}(cis) = 1.5RT$$
$$\Delta G_{TT}^{(2D)}(cis) = 6.8RT$$

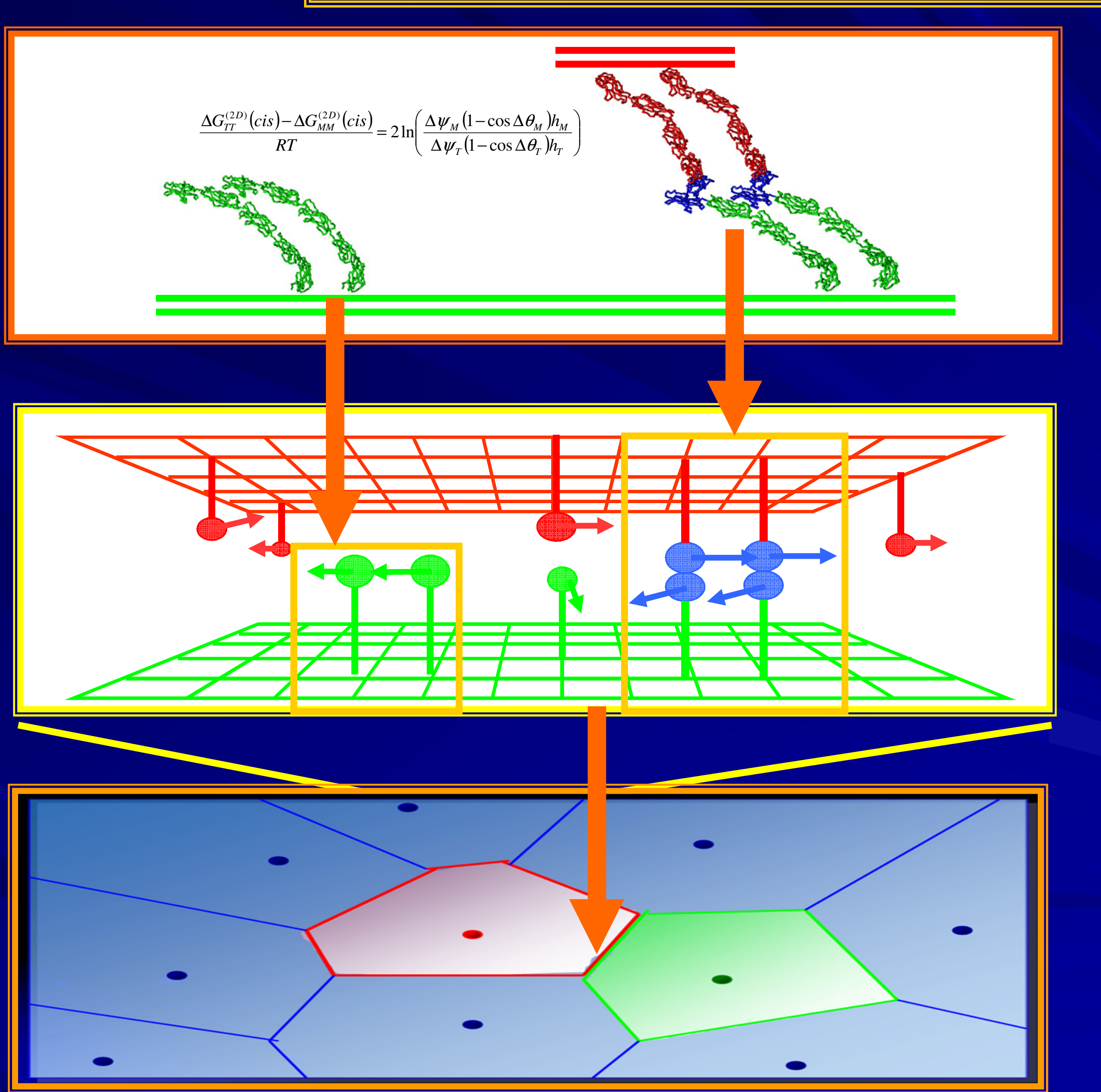
Cadherin clustering: Anisotropic Lattice MC Simulation



Cadherin clustering: Simulation versus Experiements



Beyond the molecular level



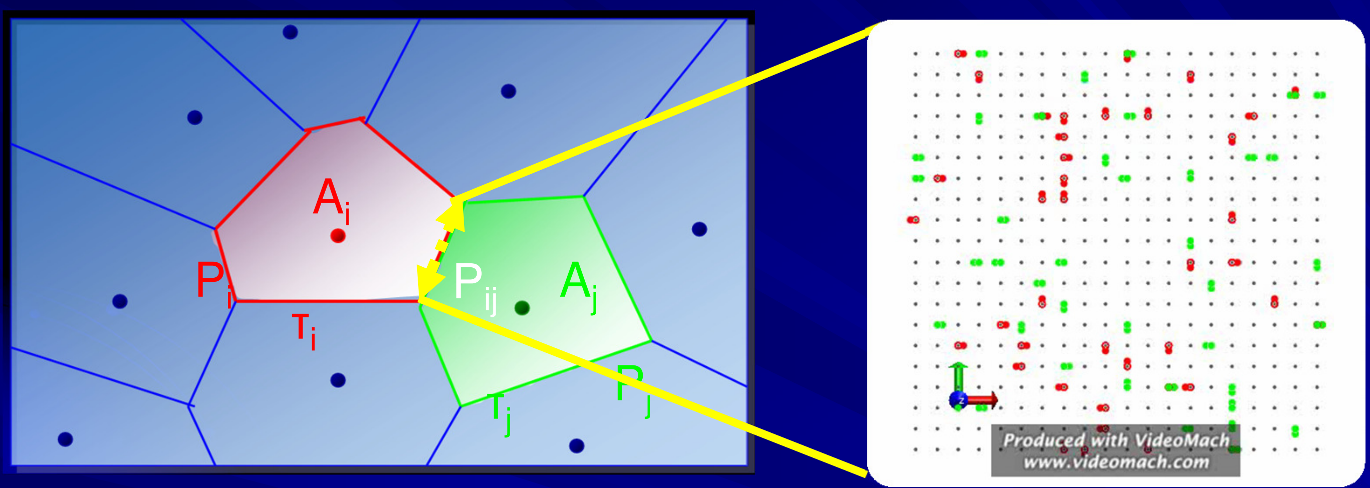
Molecular Level

Subcellular Level

Cellular Level

Individual-based Cellular Simulation

- Voronoi tessellation as the representation of cell geometry



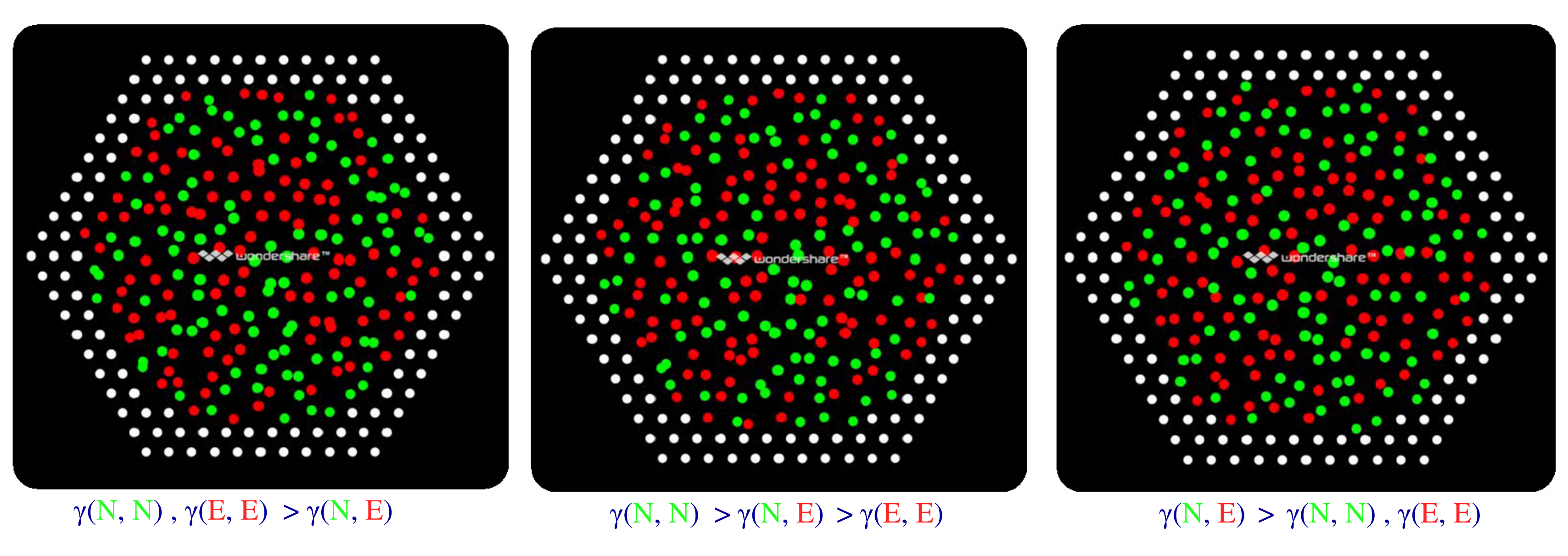
Junction simulation

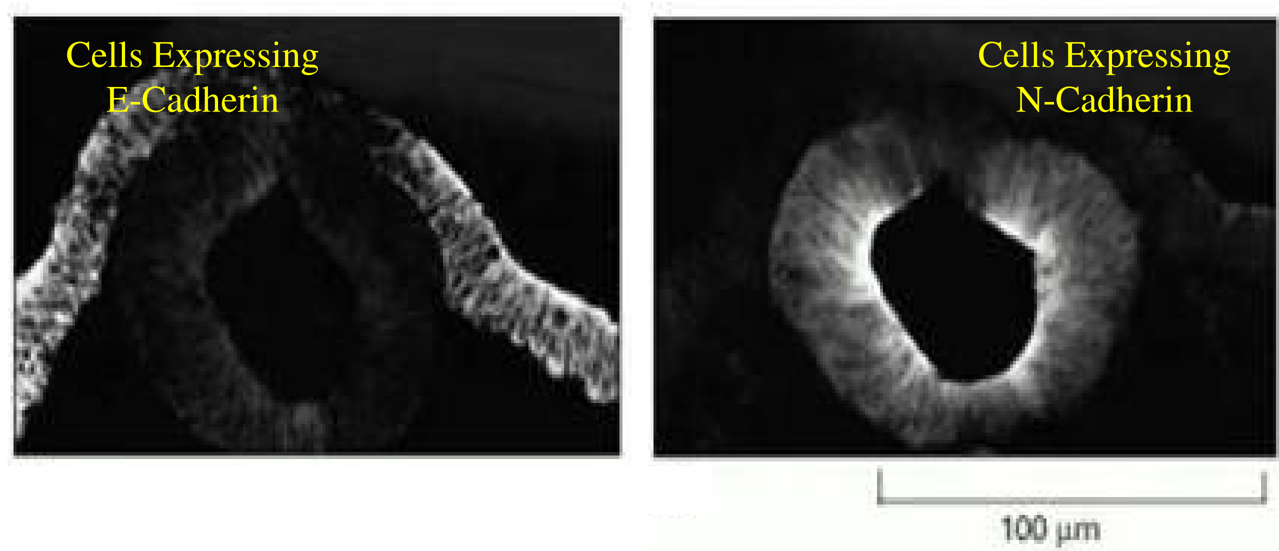
- Hamiltonian function describing cell-cell interactions

$$U_{Total} = -\sum_{ij} \gamma(\tau_i, \tau_j) P_{ij} + \lambda_A \sum_i (A_i - A_i^0)^2 + \lambda_P \sum_i (P_i - P_i^0)^2$$

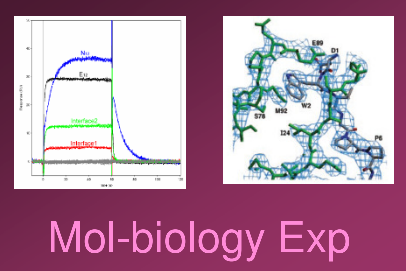
- Use simulated annealing to search the energy minimal state

From Cadherin Binding Specificity to Cell Aggregation



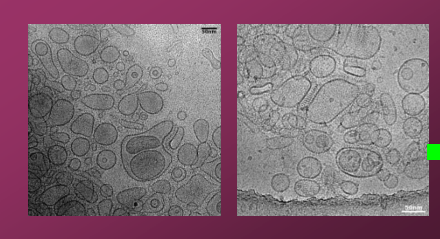


Summary

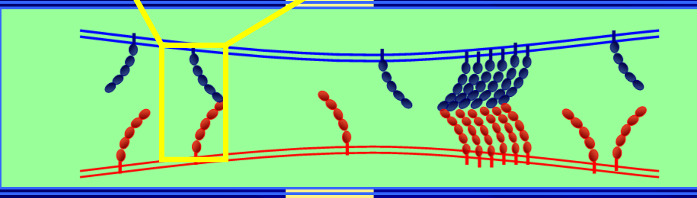


Mol-biology Exp

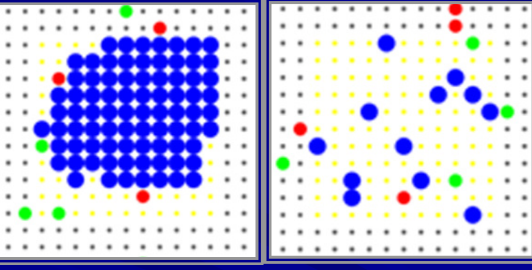
Molecular Simulation

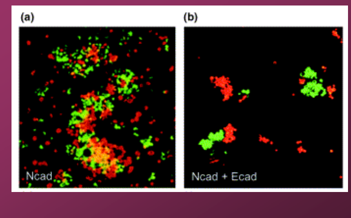


Liposome Exp

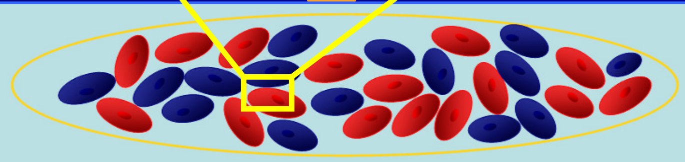


Subcellular Simulation

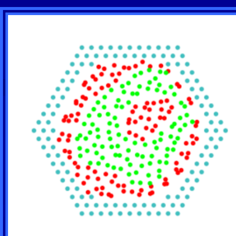


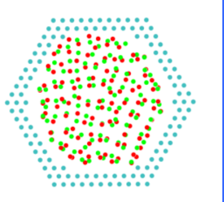


Cell-aggregation Exp



Cellular Simulation





Outline

- Modeling and Simulating Larger Molecular Systems with Longer Time Scales
  - Coarse-grained simulation
  - Multi-scale modeling
- **Extend the Applications of Structural Biology to Newly Emerging Fields**
  - **Fold proteins by information from evolution**
  - Simulating crowded cellular environment
  - Combining structural biology with systems biology
- Summary and Perspective

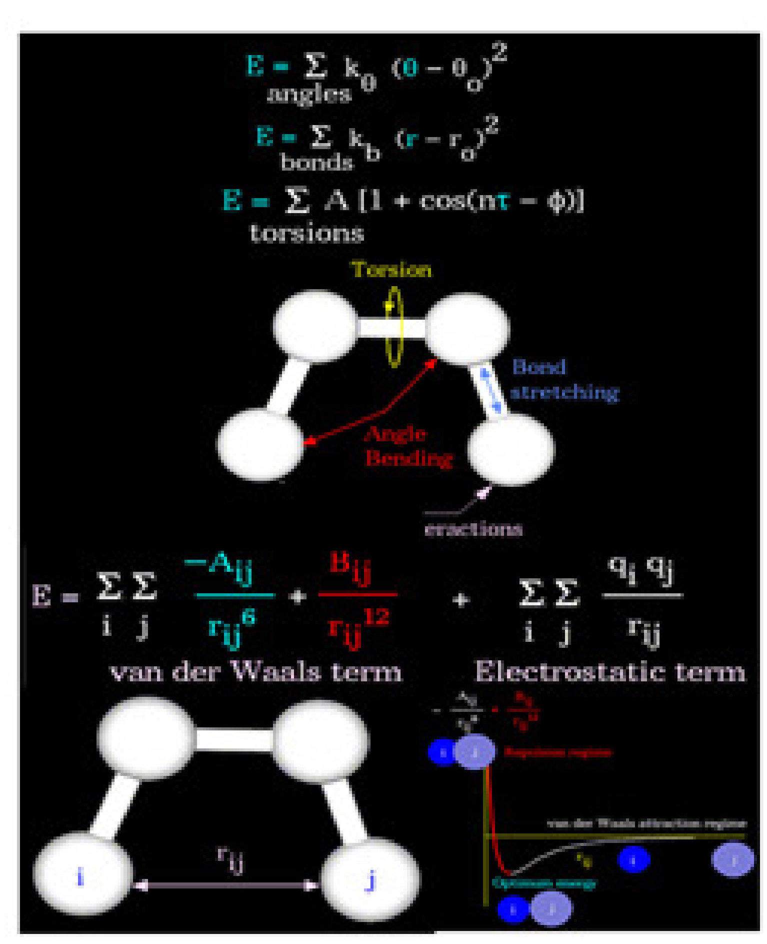
Protein Folding

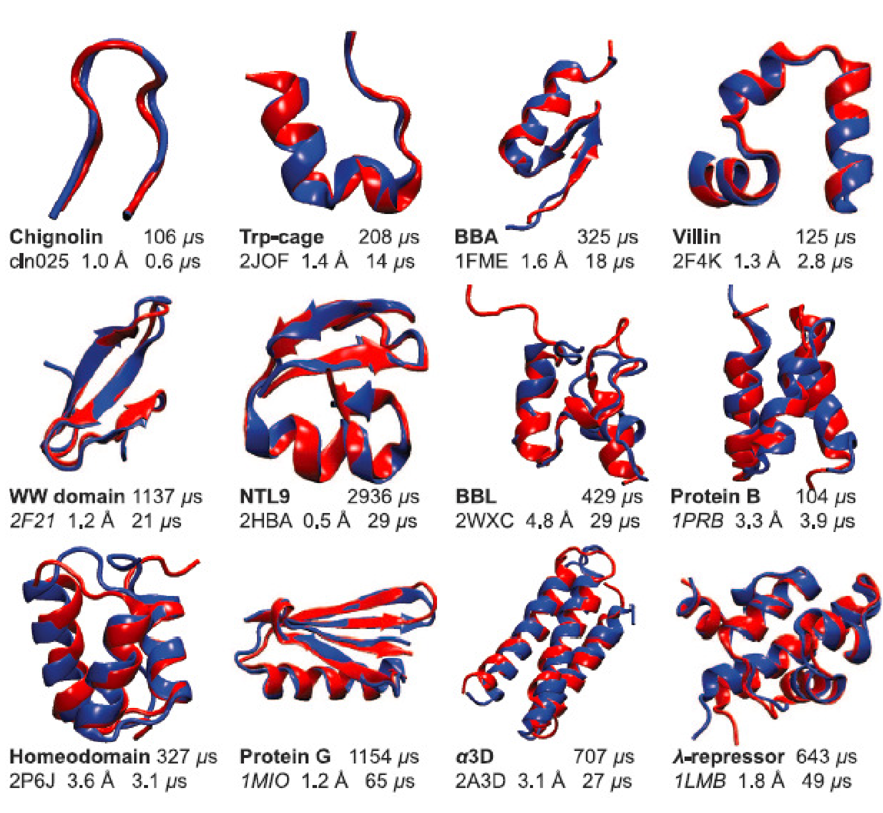
Proteins don't have a "folding problem"...
... it's we humans that do!

Physical-based potential

Knowledge-based potential

Physical-based Potential





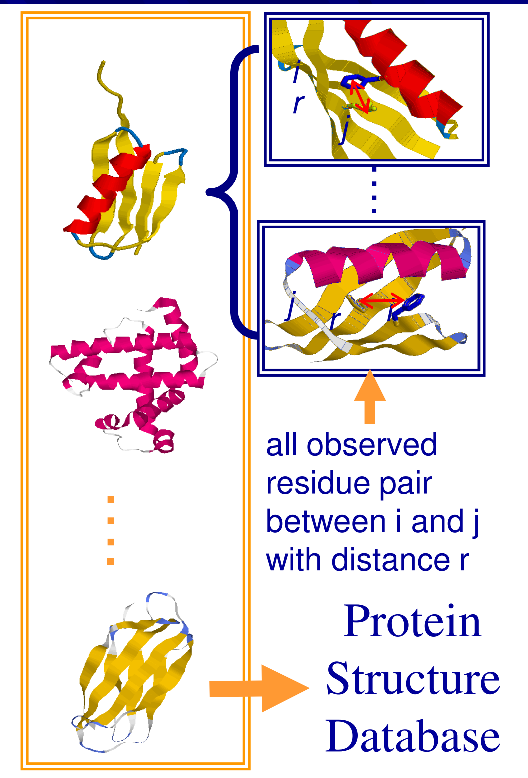
David Shaw Group

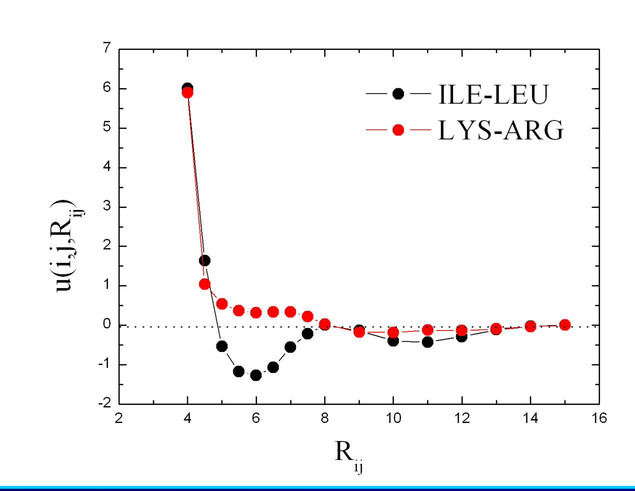
Knowledge-based Potential

The energy of a residue pair between residue type i and j with distance r

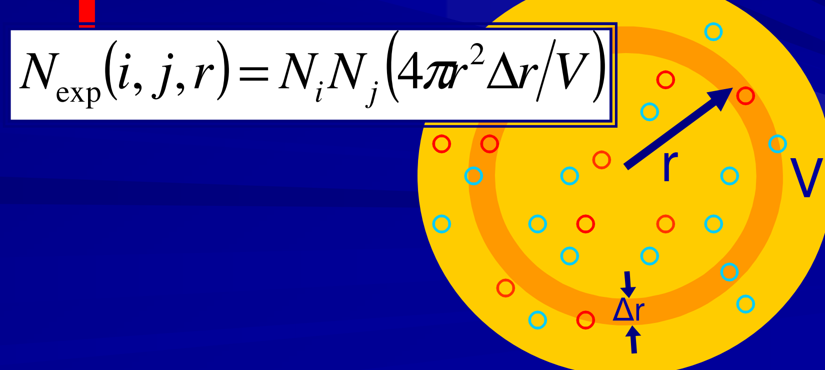
$$\bar{u}(i,j,r) = -RT \ln g_{ij}(r)$$

$$g_{ij}(r) = \frac{N_{obs}(i,j,r)}{N_{exp}(i,j,r)}$$

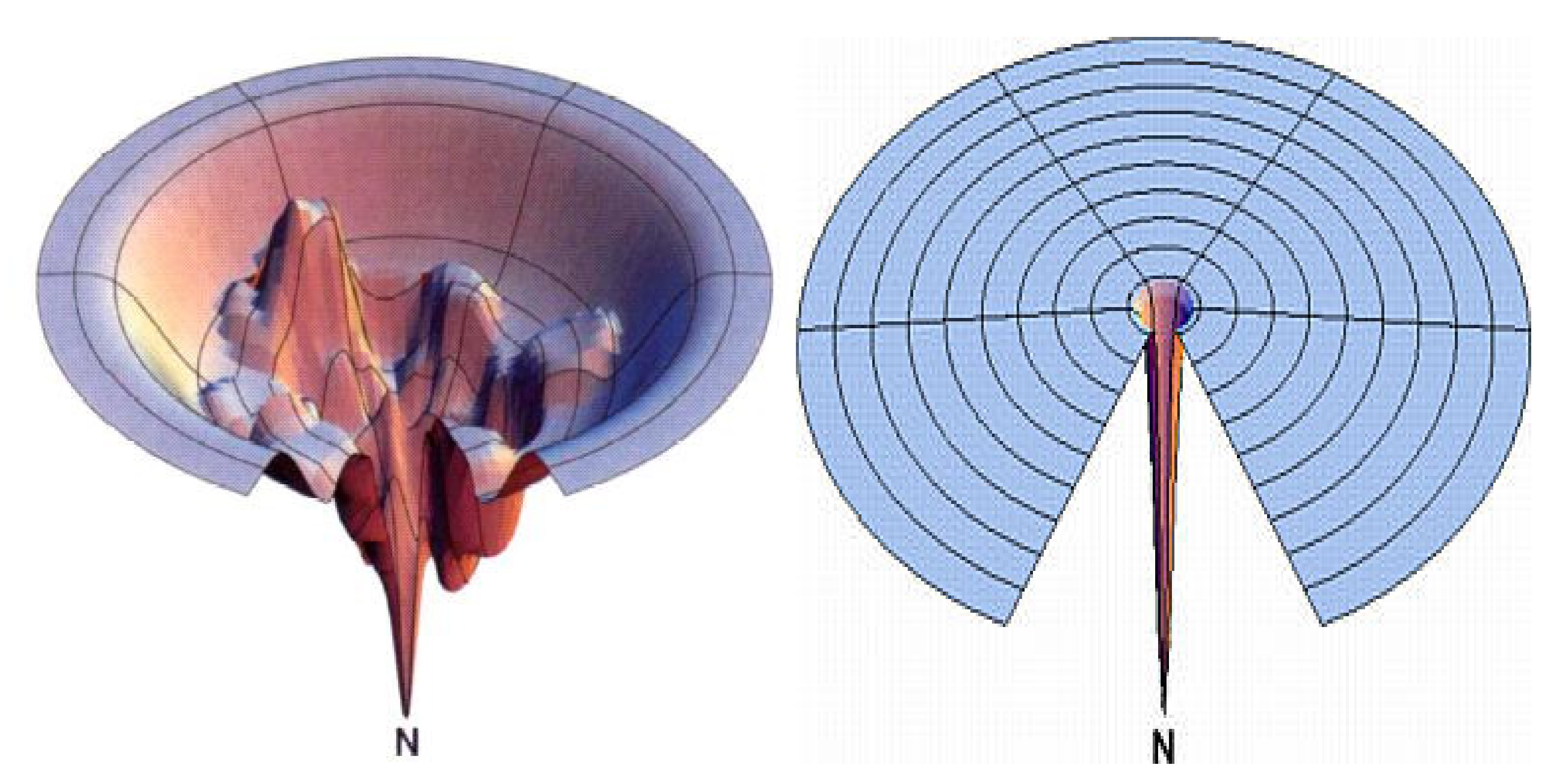




$$N_{exp}(i,j,r) = N_i N_j \left(4\pi r^2 \Delta r / V\right)$$



The problem of knowledge-based potential to study protein folding

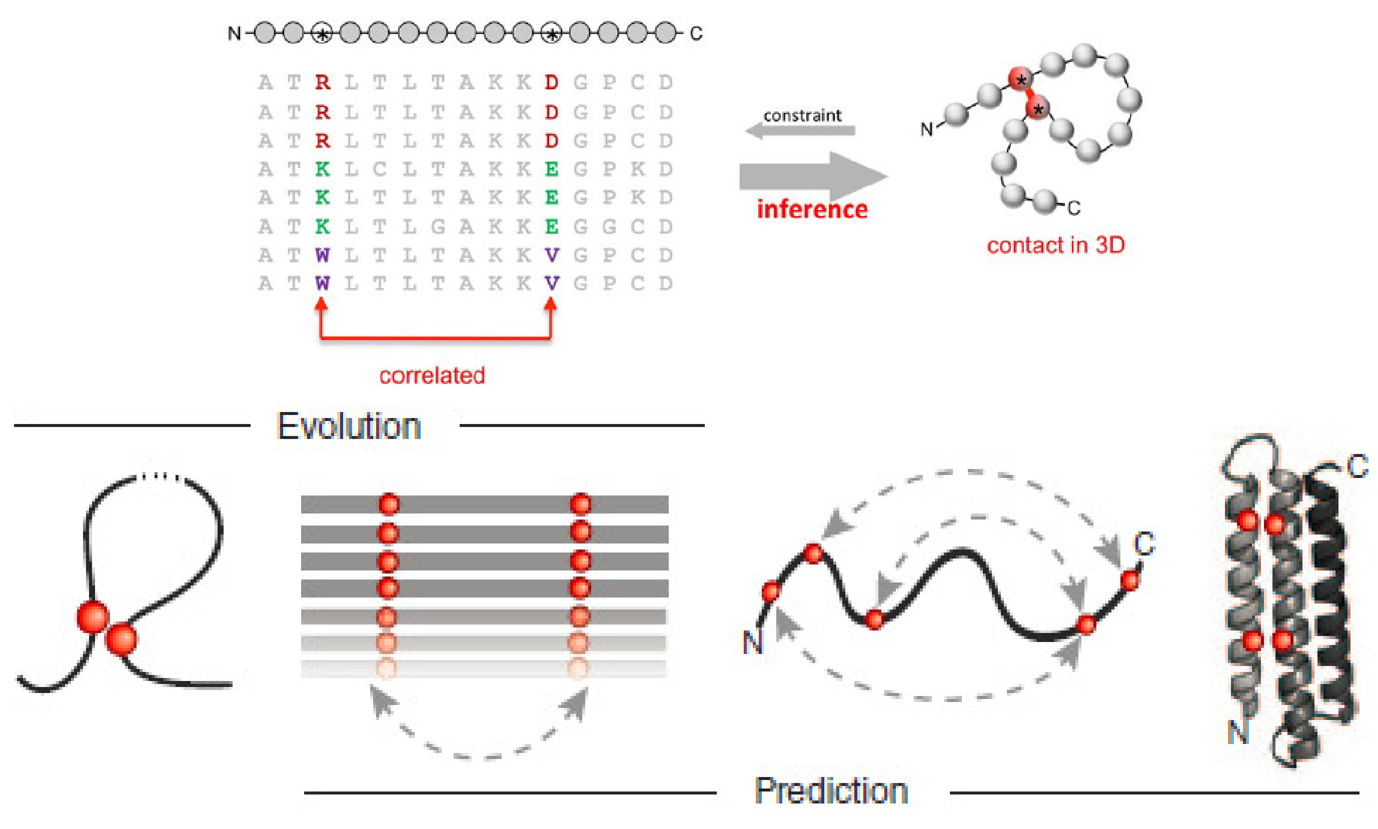


Can we get insights from other disciplines ?

Fold Proteins by Genetic Evolution

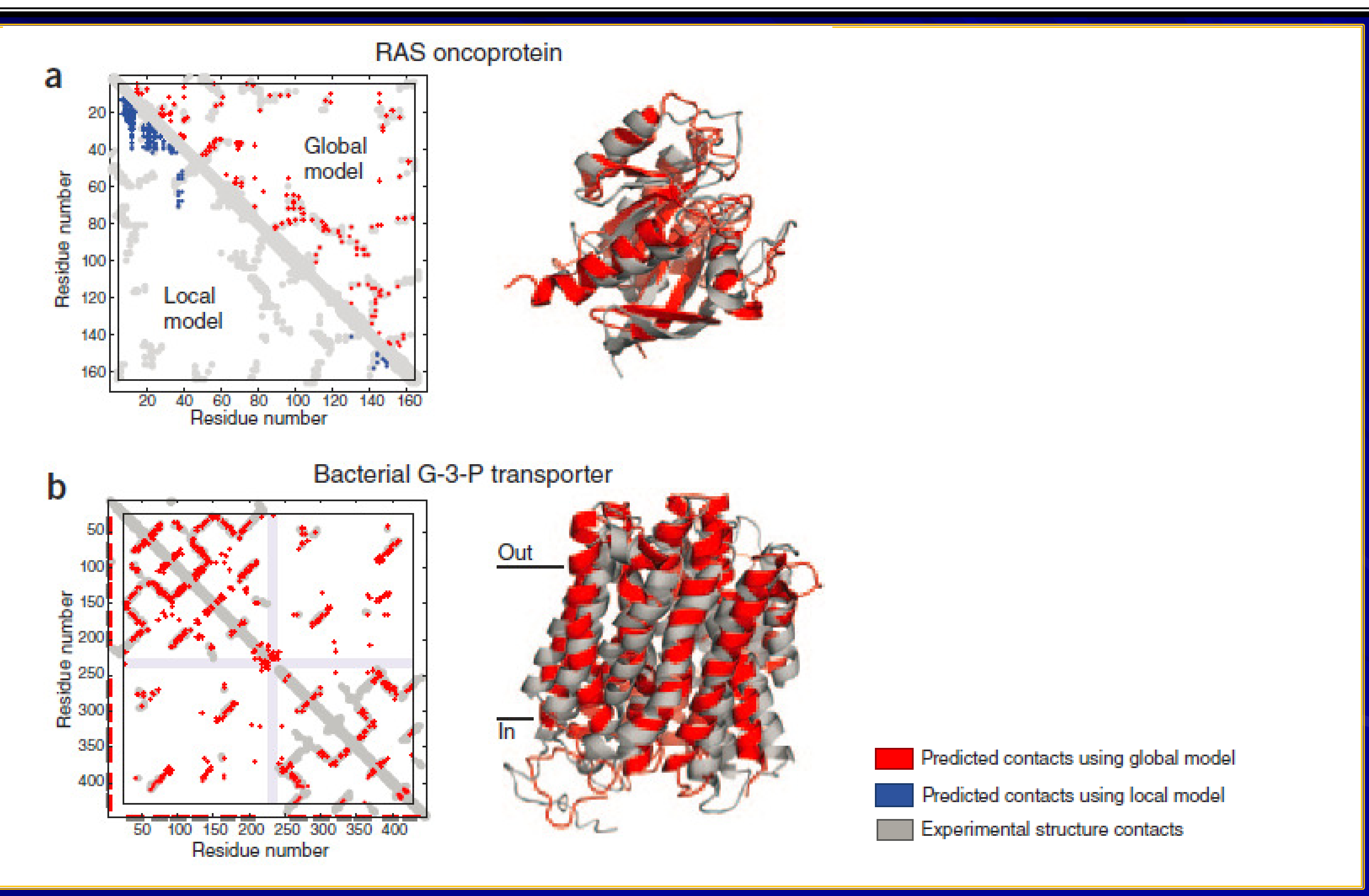
Evolution Patterns

Correlated mutations carry information about distance relationships in protein structure.



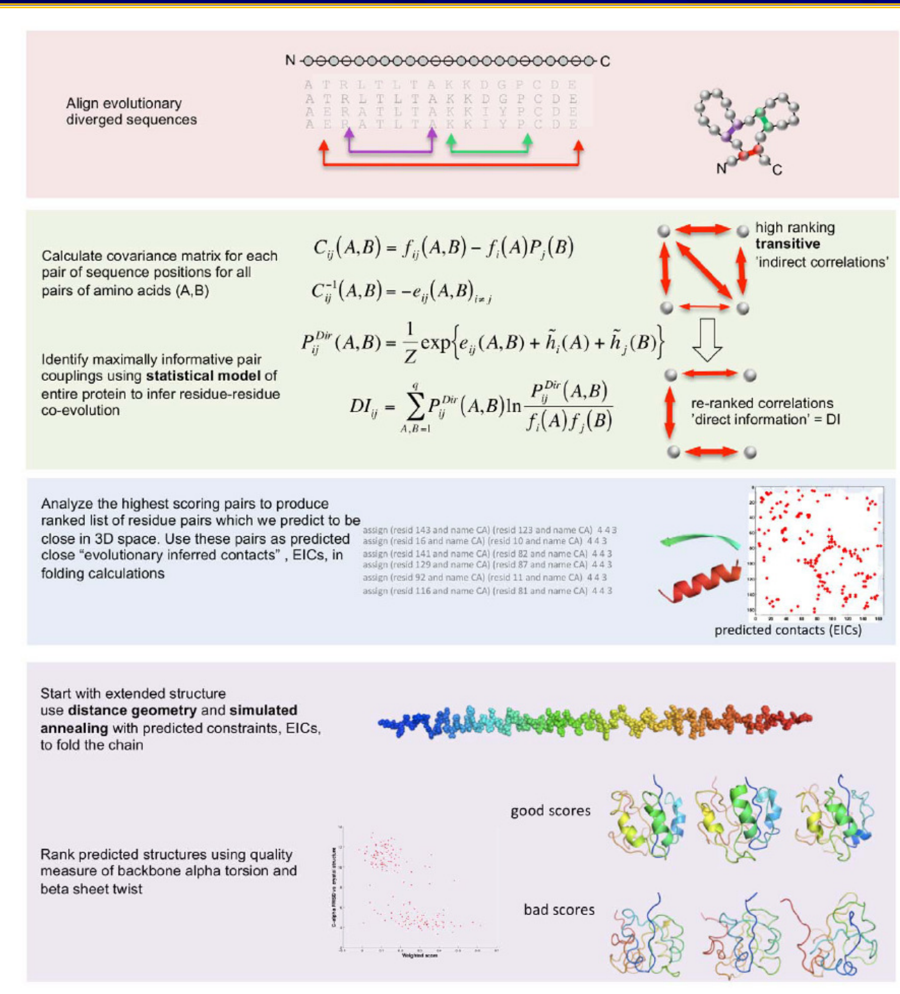
Fold Proteins by Genetic Evolution

These patterns can be used to predict residue contacts in proteins



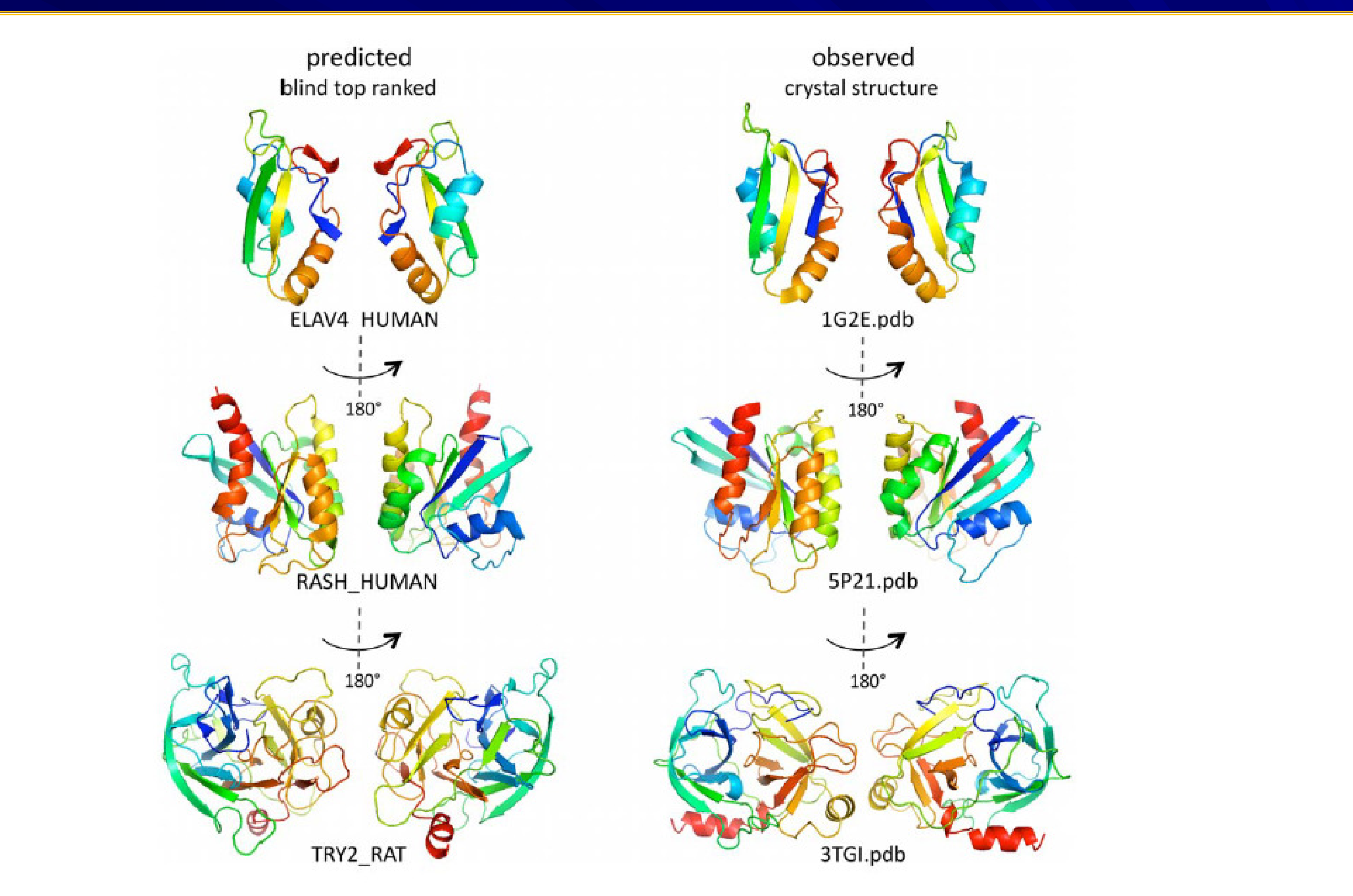
Fold Proteins by Genetic Evolution

Predicted contacts can be transferred into potential to fold proteins



Fold Proteins by Genetic Evolution

Results



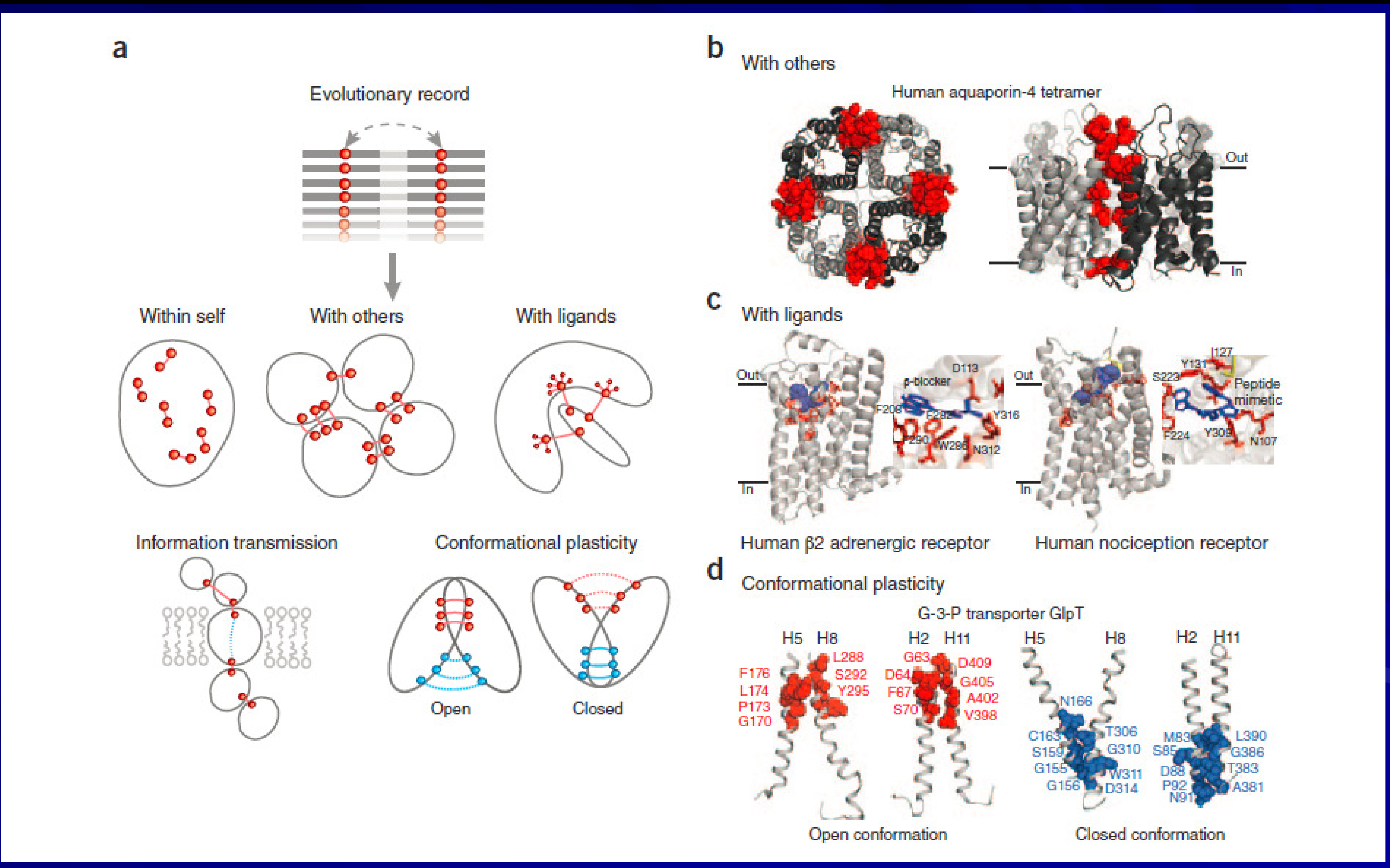
Fold Proteins by Genetic Evolution

Conclusions

- **Three-dimensional structure from coevolution patterns—why does it work?**
  - **Precise information in the evolutionary sequence record.**
  - **Growth in sequence databases from massively parallel sequencing.**
  - **Reduction of conformational search space by cooperative probability models.**

Fold Proteins by Genetic Evolution

Future Directions

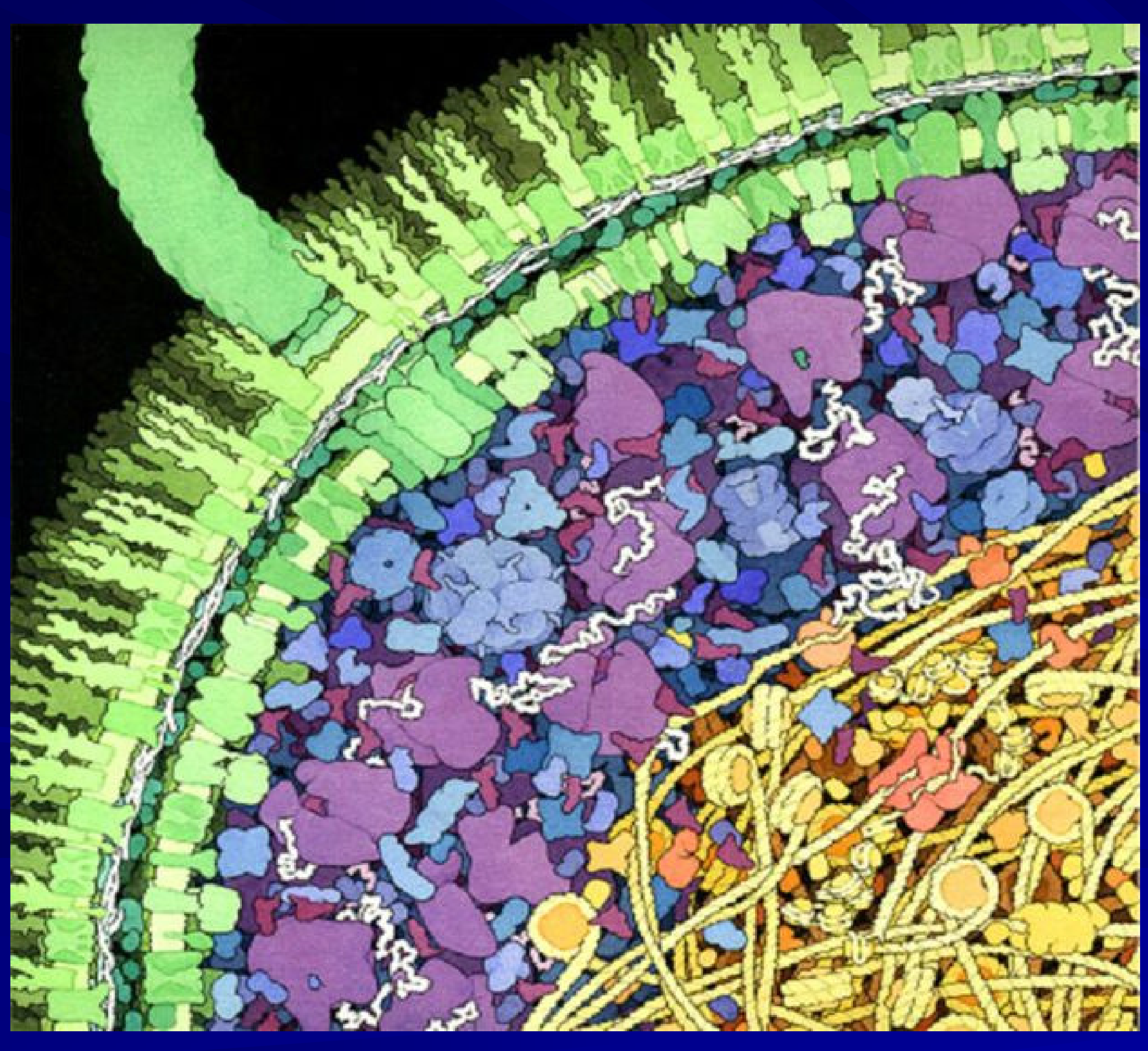


Outline

- Modeling and Simulating Larger Molecular Systems with Longer Time Scales
  - Coarse-grained simulation
  - Multi-scale modeling
- **Extend the Applications of Structural Biology to Newly Emerging Fields**
  - Fold proteins by information from genetics
  - **Simulating crowded cellular environment**
  - Combining structural biology with systems biology
- Summary and Perspective

Background

There are high concentrations of macromolecules in living cells



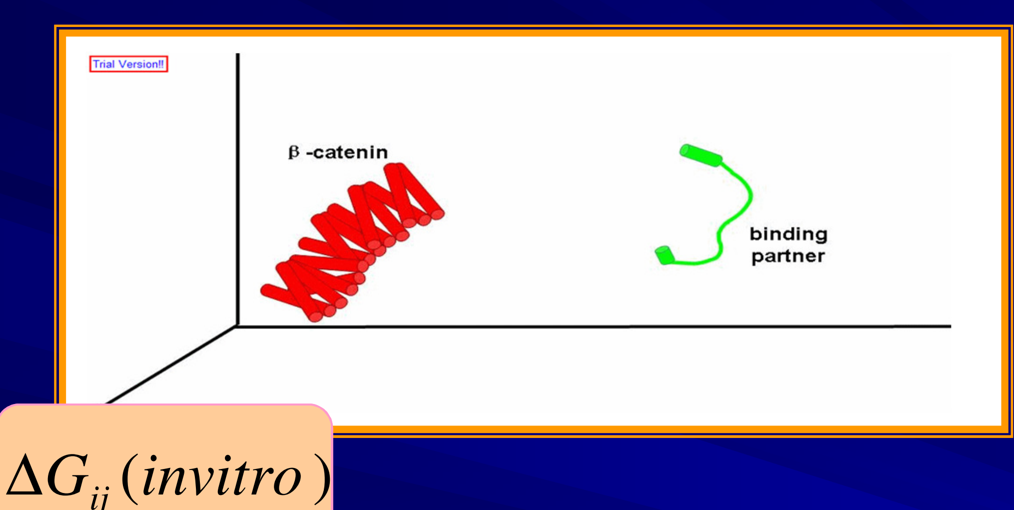
The crowding effect can make molecules in cells behave in radically different ways than in test-tube assays.

Consequently, measurements of the properties of enzymes or processes in metabolism that are made in the laboratory (*in vitro*) in dilute solutions may be different by many orders of magnitude from the true values seen in living cells (*in vivo*).

The study of biochemical processes under realistically crowded conditions is very important, since these conditions are a ubiquitous property of all cells and crowding may be essential for the efficient operation of metabolism.

Background

Crowding effects alter molecular properties



$\Delta G_{ij}(invitro)$

Trial Version!!

β -catenin

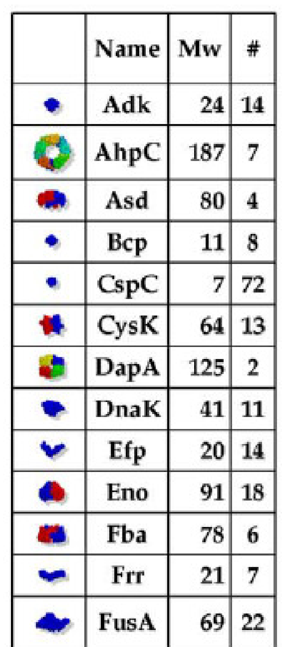
binding partner

$\Delta G_{ij}(invivo)$

- The size of the crowding effect depends on both the molecular mass and shape of the molecule involved.
- This excluded volume effect alters the rates and equilibrium constants of molecular reactions.
- This effect alters dissociation constants of macromolecules, such as when multiple proteins come together to form protein complexes, or when DNA-binding proteins bind to their targets in the genome, or enzyme reactions involving small molecules in solvent.

Simulate Crowded Cellular Environment

Model Setup

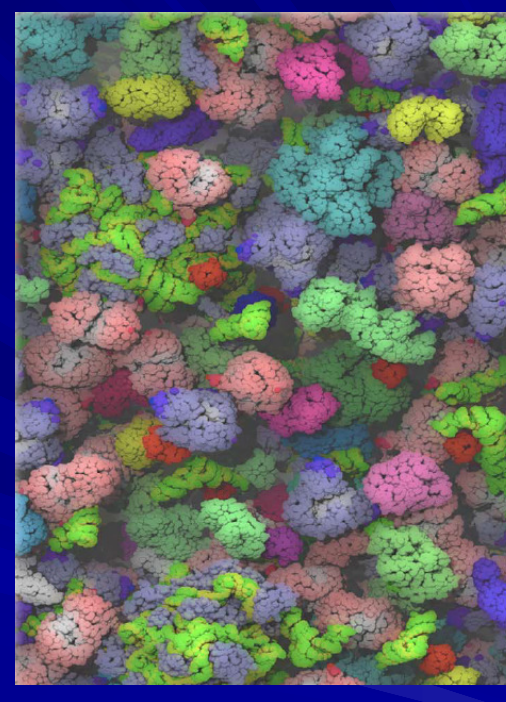


Name	Mw	#
Adk	24	14
AhpC	187	7
Asd	80	4
Bcp	11	8
CspC	7	72
CysK	64	13
DapA	125	2
DnaK	41	11
Efp	20	14
Eno	91	18
Fba	78	6
Frr	21	7
FusA	69	22

Name	Mw	#
GapA	142	10
GlnA	621	1
GltD	94	3
GlyA	91	15
GpmA	55	4
Hns	5	7
Hup	15	12
IcdA	92	43
IlvC	54	18
Mdh	65	13
MetE	84	213
Mop	845	2

Name	Mw	#
PanB	140	2
Pgk	41	26
Pnp	190	3
Ppa	116	9
PpiB	18	7
PurA	94	4
PurC	42	7
Pyr	308	3
RpiA	46	3
Rpo	260	4
SerC	79	11
SodA	46	13
SodB	42	9

Name	Mw	#
Suc	142	4
Tig	48	9
TpiA	54	5
Tsf	61	12
TufA	84	181
Upp	45	11
UspA	31	7
50S	1,355	10
30S	788	10
tRNA-C	24	37
tRNA-Q	24	37
tRNA-F	25	37
GFP	26	8



Skolnick Group & Elcock Group

Stochastic Dynamics: Equation of Motion

$$\vec{F}_{ij} = -\left(\partial V_{ij} / \partial d_{ij}\right) \times \left(\frac{d(d_{ij})}{dx}, \frac{d(d_{ij})}{dy}, \frac{d(d_{ij})}{dz} \right)$$

Forces

Langevin Equation

$$\hat{\Gamma}^{f}_{i-E} \times \vec{v}_i + \sum_{j \in (d_{ij} < (R_i + R_j))} \hat{\Gamma}^{f}_{ij} \times \left(\vec{v}_i - \vec{v}_j \right) = \sum_{j \in (d_{ij} < (R_i + R_j))} \vec{F}_{ij} + f_i(t)$$

Molecule-environment friction *Inter-molecule friction* *Random forces from environment*

$$\hat{\Gamma}^{f}_{i-E} = \eta \hat{I}$$

$$\hat{\Gamma}^{f}_{ij} = \eta^{ij}_{\parallel} \vec{n}_{ij} \vec{n}_{ij} + \eta^{ij}_{\perp} (\hat{I} - \vec{n}_{ij} \vec{n}_{ij})$$

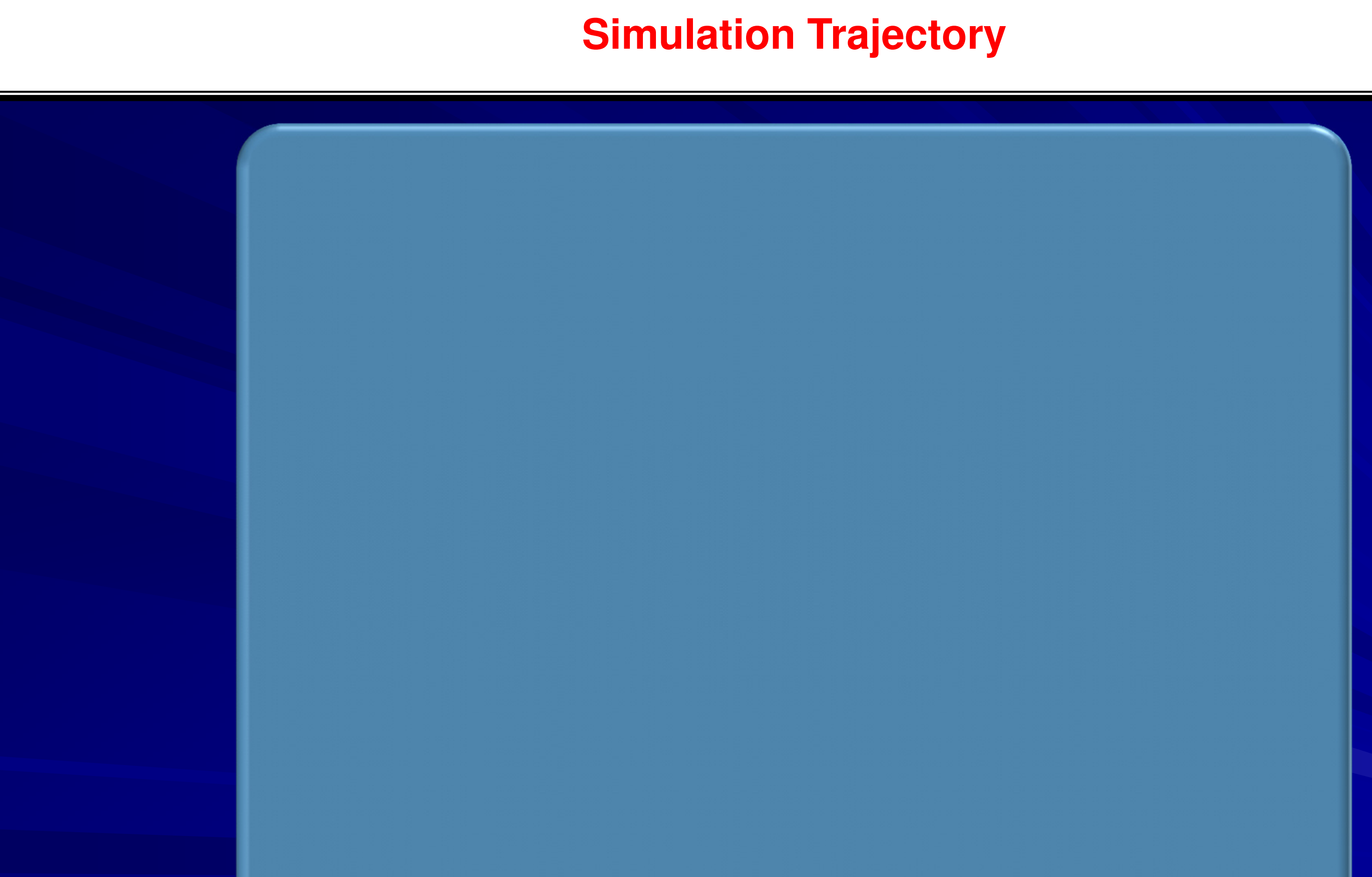
$$\vec{n}_{ij} = \left(\vec{x}_j - \vec{x}_i \right) / \left| \vec{x}_j - \vec{x}_i \right|$$

$$\left\langle f(t) f(t') \right\rangle \approx 2 \hat{\Gamma} \delta(t - t'), \quad \hat{\Gamma} \approx 2 \eta^2 D \hat{I}$$

$$\left\langle f(t) \right\rangle = 0$$

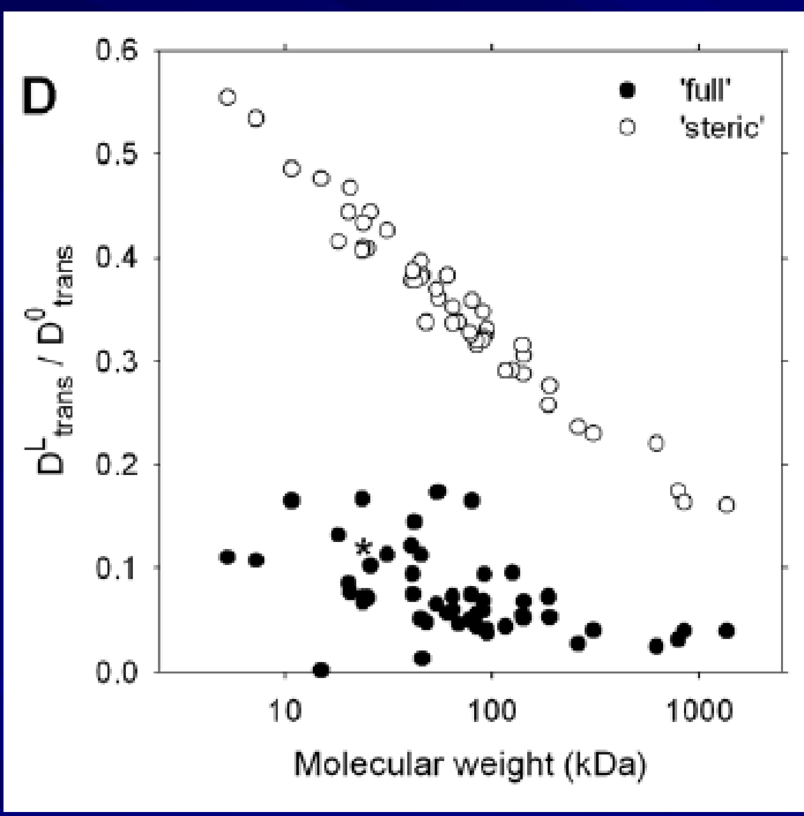
Simulate Crowded Cellular Environment

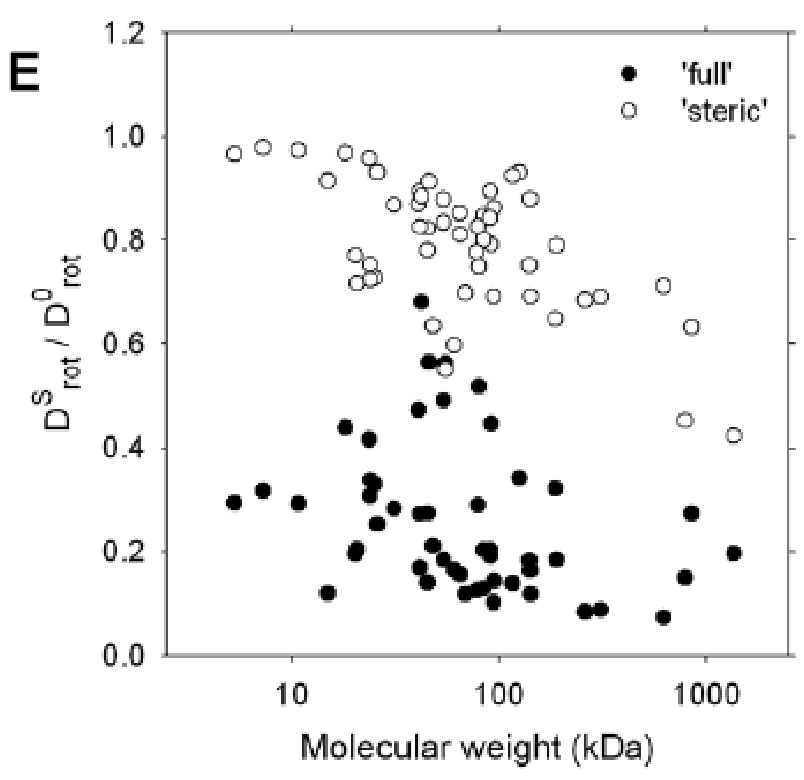
Simulation Trajectory



Simulate Crowded Cellular Environment

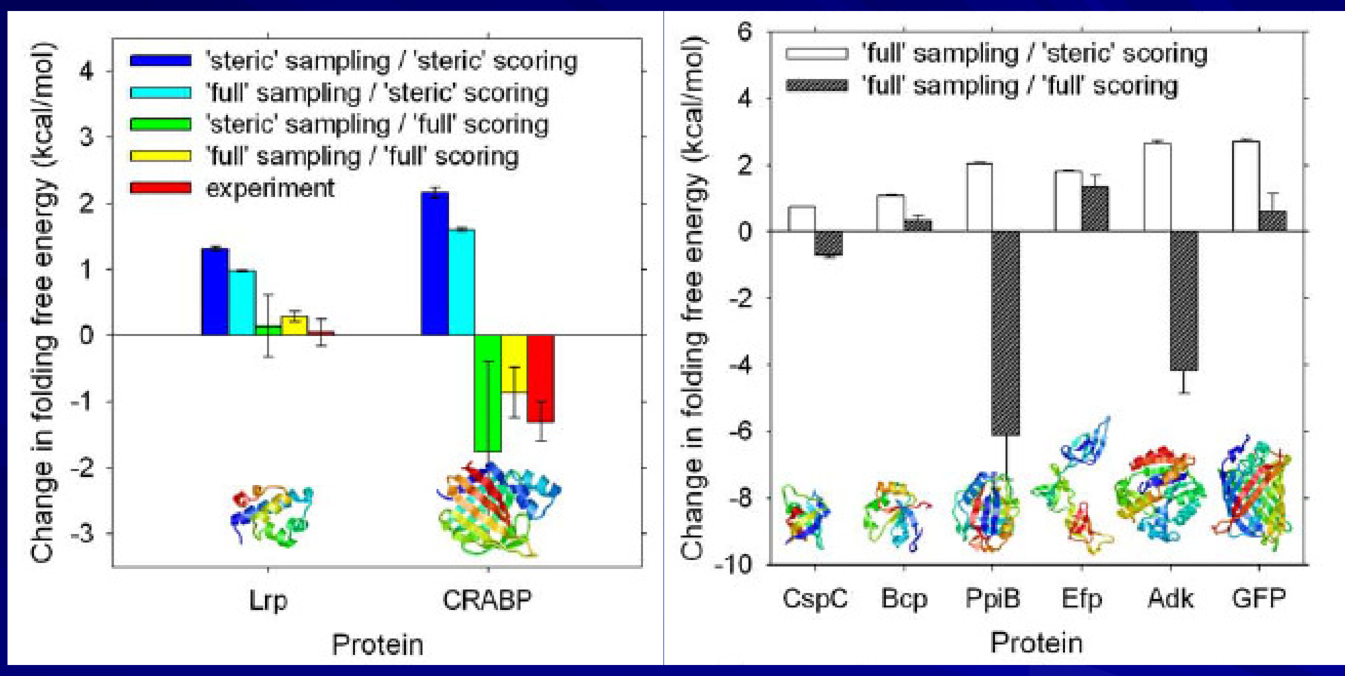
Simulation Results: Diffusion





Simulate Crowded Cellular Environment

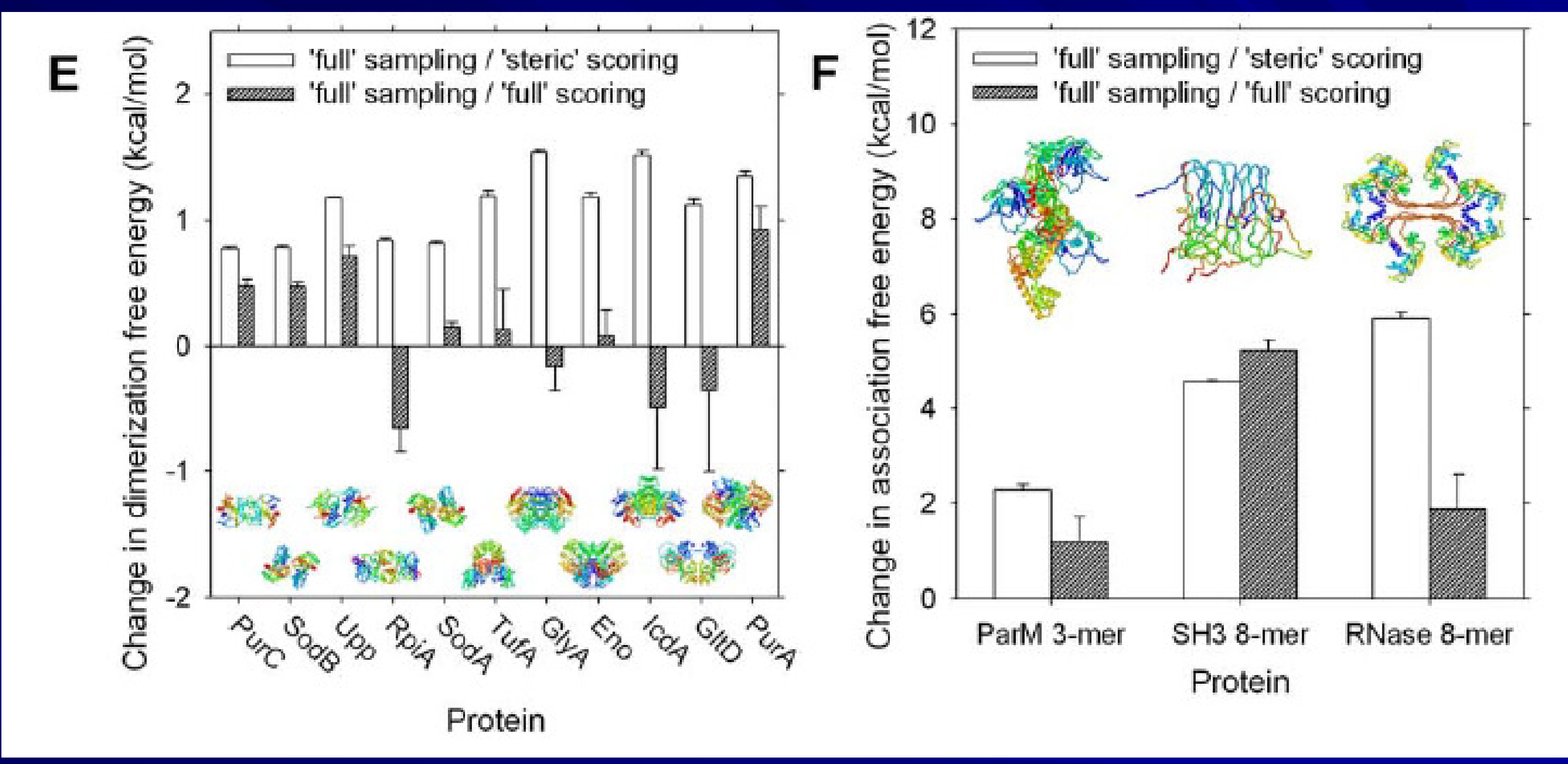
Simulation Results: Stability



- The excluded-volume (crowding) effect experienced by proteins significantly stabilizes their folded states relative to their unfolded states.
- The effect is counterbalanced by the more favorable energetic interactions engaged in by the unfolded state conformations.
- These results suggest that differences between the in vitro and in vivo thermodynamic stabilities will vary significantly with the identity of the protein.

Simulate Crowded Cellular Environment

Simulation Results: Binding



- **Excluded-volume crowding effect, which alone stabilizes dimers relative to separated monomers is largely cancelled by the more favorable energetic interactions that the monomers form with the cytoplasm constituents.**

Simulate Crowded Cellular Environment

Summary and Conclusions

- Developing computational models of intracellular environments is the route to understanding differences between biomolecular behavior observed in vitro and in vivo.
- The simulation model successfully describes the relative thermodynamic stabilities of proteins measured in E. coli, and shows that effects additional to the "crowding" effect must be included in attempts to understand macromolecular behavior in vivo.
- the simulation approach offers a potentially important complement to experimental techniques and provides a vivid illustration of molecular behavior inside a biological cell

Simulate Crowded Cellular Environment

Future Directions

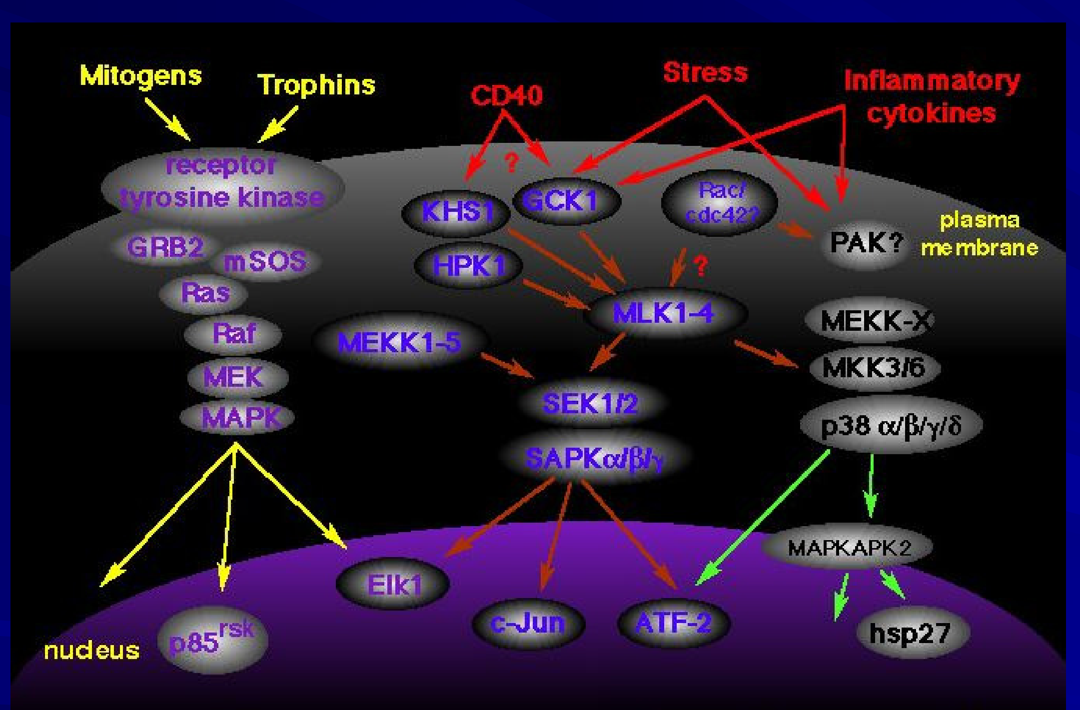
- The internal conformational variations of macromolecules or transitions between different states in molecular motors.
- Effects from cellular compartmentation and cytoskeleton.
- Real biological applications.

Outline

- Modeling and Simulating Larger Molecular Systems with Longer Time Scales
  - Coarse-grained simulation
  - Multi-scale modeling
- **Extend the Applications of Structural Biology to Newly Emerging Fields**
  - Fold proteins by information from genetics
  - Simulating crowded cellular environment
  - **Combining structural biology with systems biology**
- Summary and Perspective

Predicting protein-protein interactions

Background: Protein-protein interactions (PPIs)



PPIs are involved in many biological processes:

- Signal transduction
- Protein complexes or molecular machinery
- Protein carrier
- Protein modifications (phosphorylation)
- …

PPIs prediction: predict if two proteins interact with each other or not, given their sequences.

Predicting protein-protein interactions

Background: experimental methods

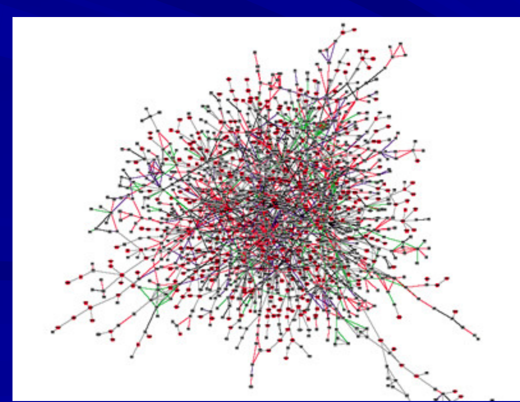
- The first comprise and ‘atomic observation’ in which the protein interaction is detected using, for example, X-ray crystallography. These experiments can yield specific information on the atoms or residues involved in the interaction.
- The second is a ‘direct interaction observation’ where protein interaction between two partners can be detected as in a two-hybrid experiment.
- At a third level of observation, multi-protein complexes can be detected using methods such as immuno-precipitation or mass-specific analysis. This type of experiment does not unveil the chemical detail of the interactions or even reveal which proteins are in direct contact but gives information as to which proteins are found in a complex at a given time.
- The fourth category comprises measurements at the cellular level, where an ‘activity bioassay’ is used to observe an interaction; for example, proliferation assays of cells by a receptor-ligand interaction.

Predicting protein-protein interactions

Background: PPIs databases

- Recent advances in proteomics technologies such as two-hybrid, phage display and mass spectrometry have enabled us to create a detailed map of biomolecular interaction networks. Initial mapping efforts have already produced a wealth of data. As the size of the interaction set increases, databases and computational methods will be required to store, visualize and analyze the information in order to effectively aid in knowledge discovery.

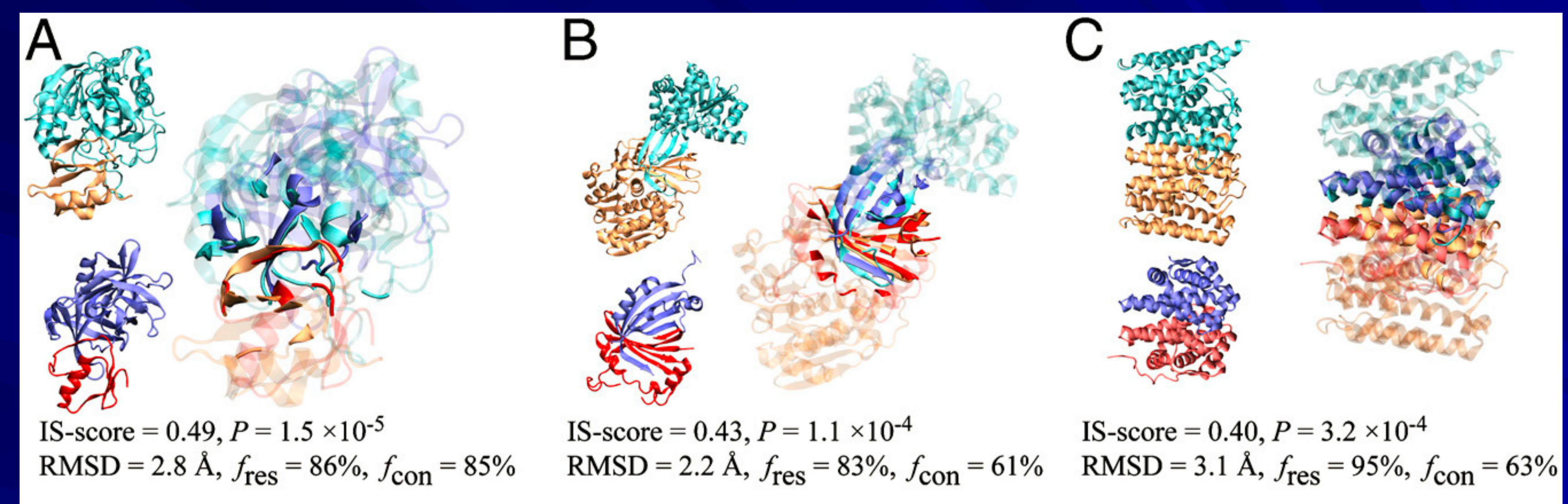
- For the protein-protein interactions, there are many websites can be reached, here I just show several.
  - BIND (Interaction Network Database)
  - DIP (Database of Interacting Proteins)
  - Protein-Protein Interaction Server
  - Protein-Protein Interface



Problem: too much information, many false positives and hard to prove.

Predicting protein-protein interactions

Idea: what computational structural biologists can do?

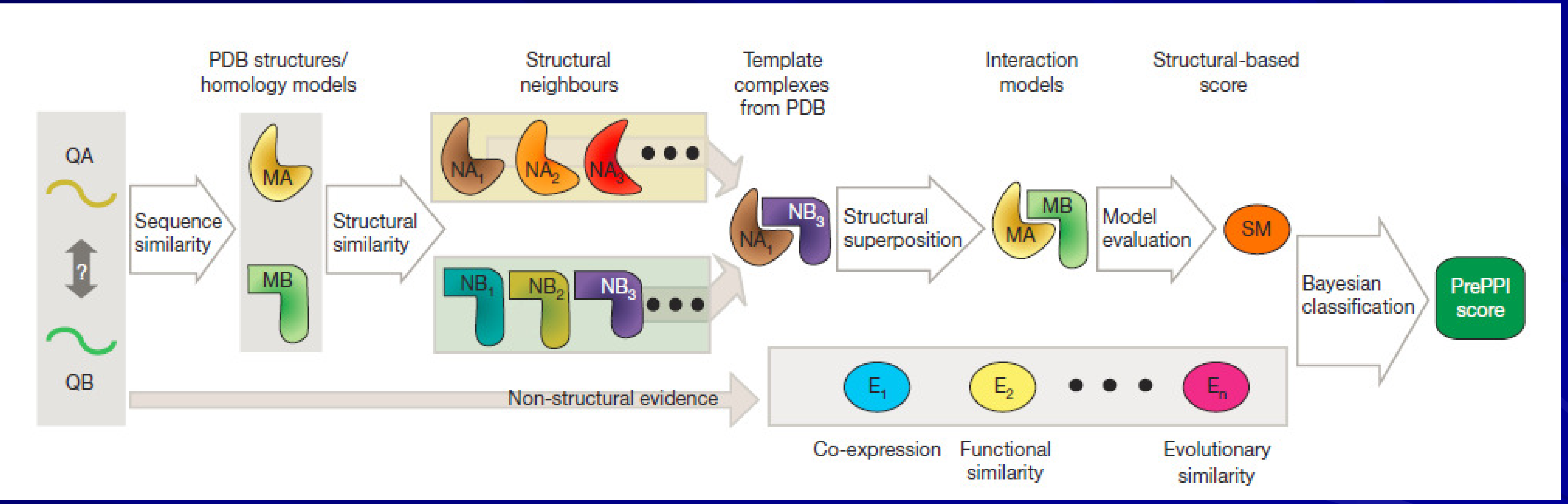


Although protein complexes only share remote structural similarity, they sometimes tend to bind with each other with conservative interfaces.

Can we use this structural information to help PPI prediction?

Predicting protein-protein interactions

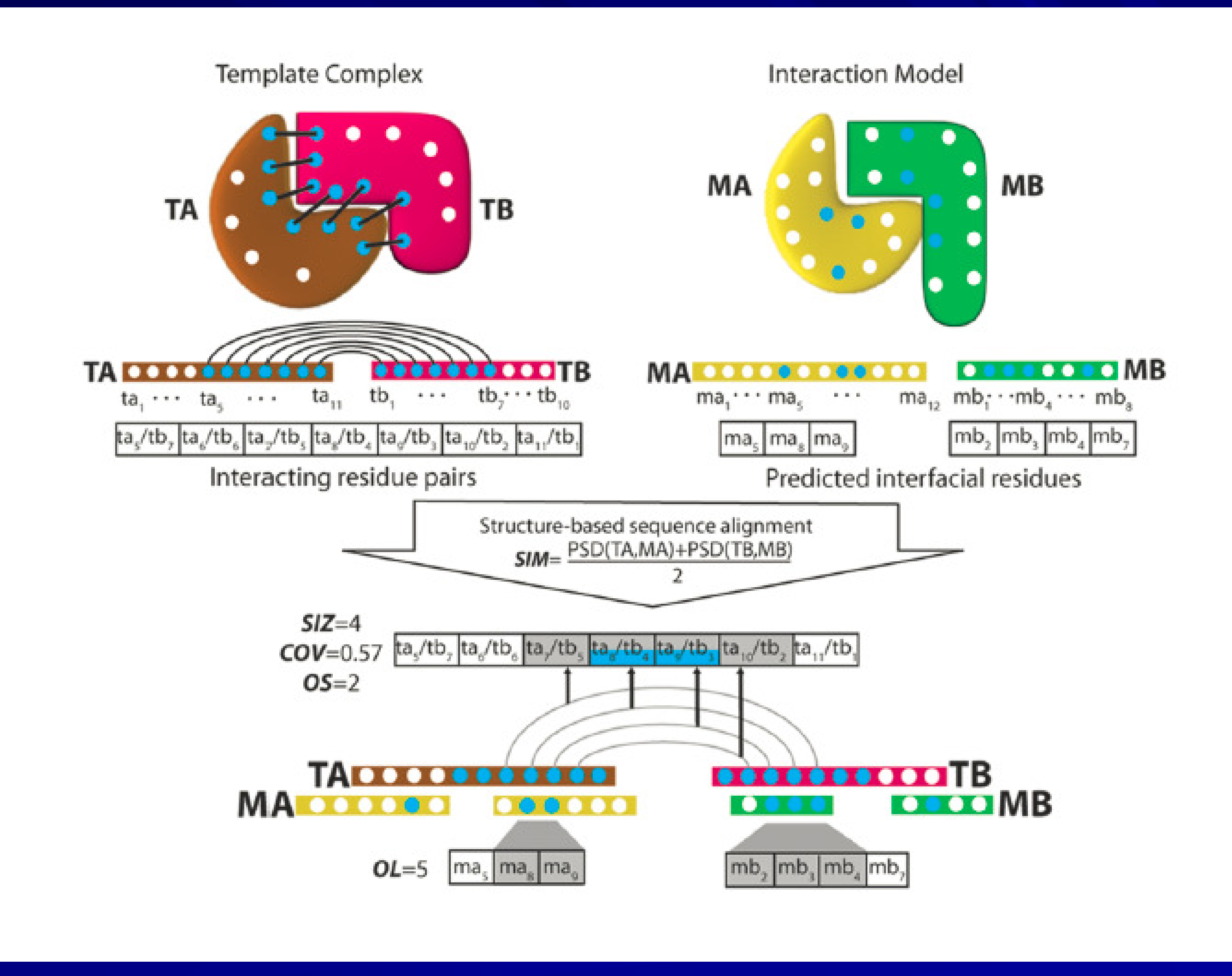
Method: prediction procedure



Honig Lab

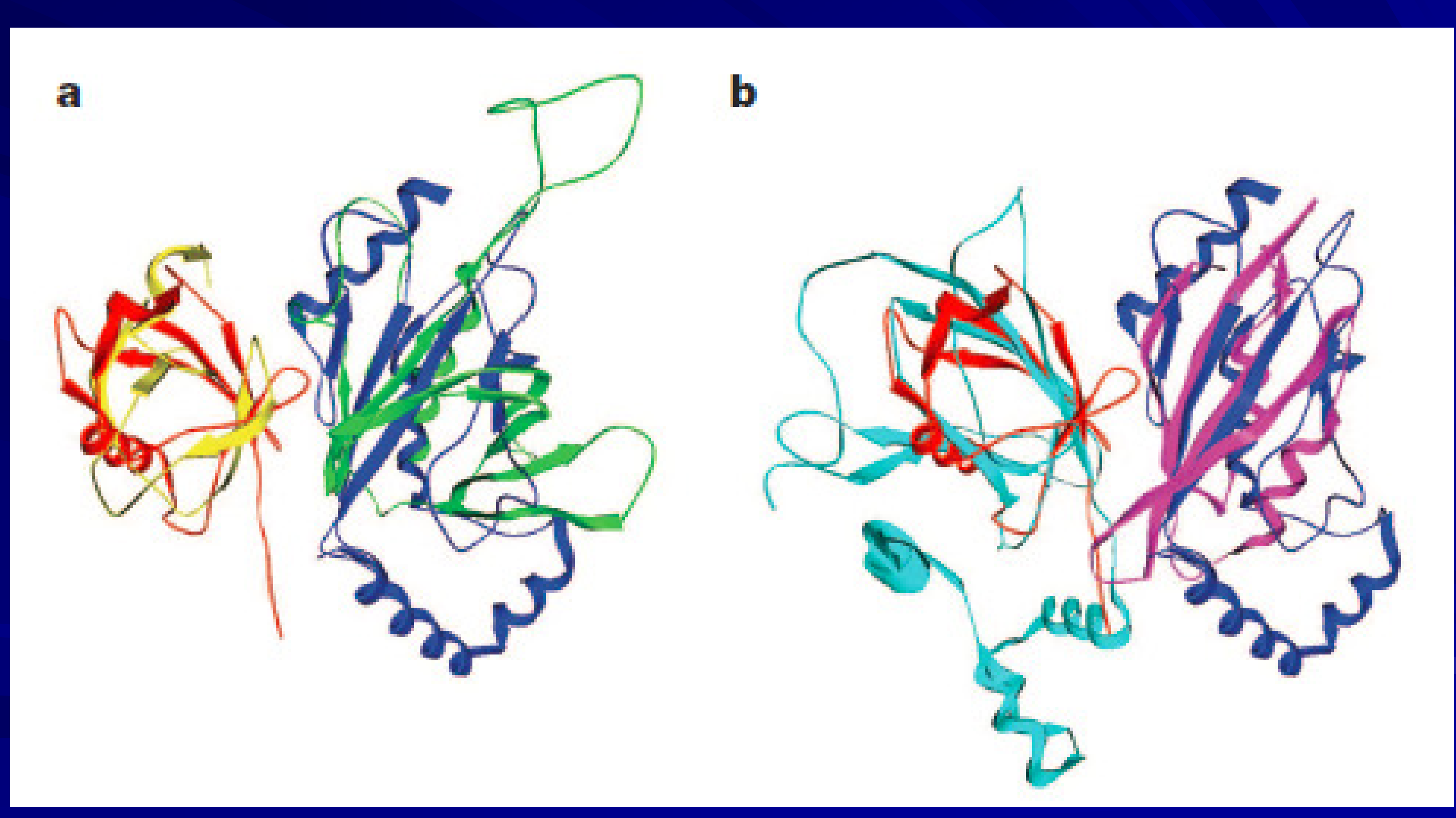
Predicting protein-protein interactions

Method: model evaluation



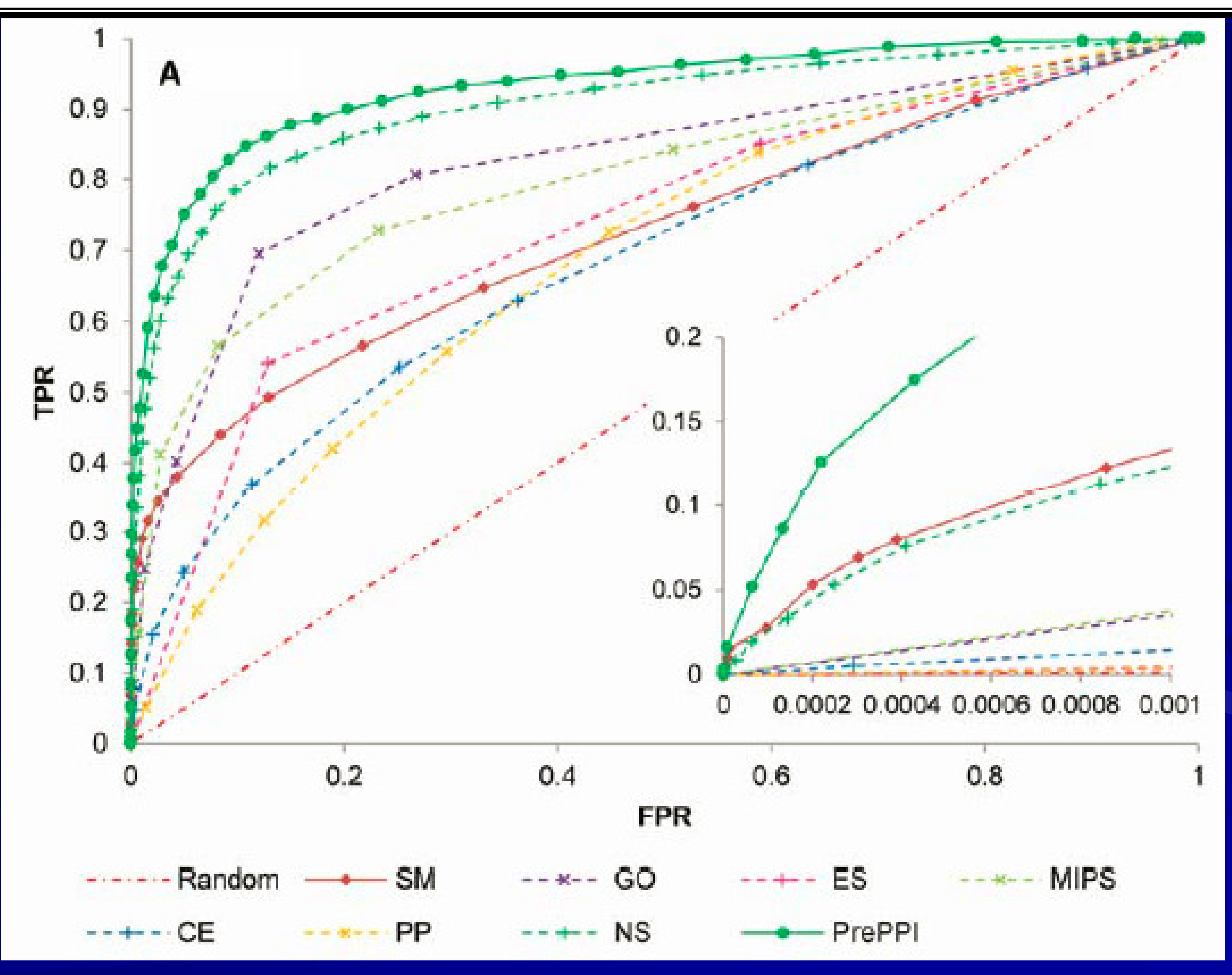
Predicting protein-protein interactions

Prediction results



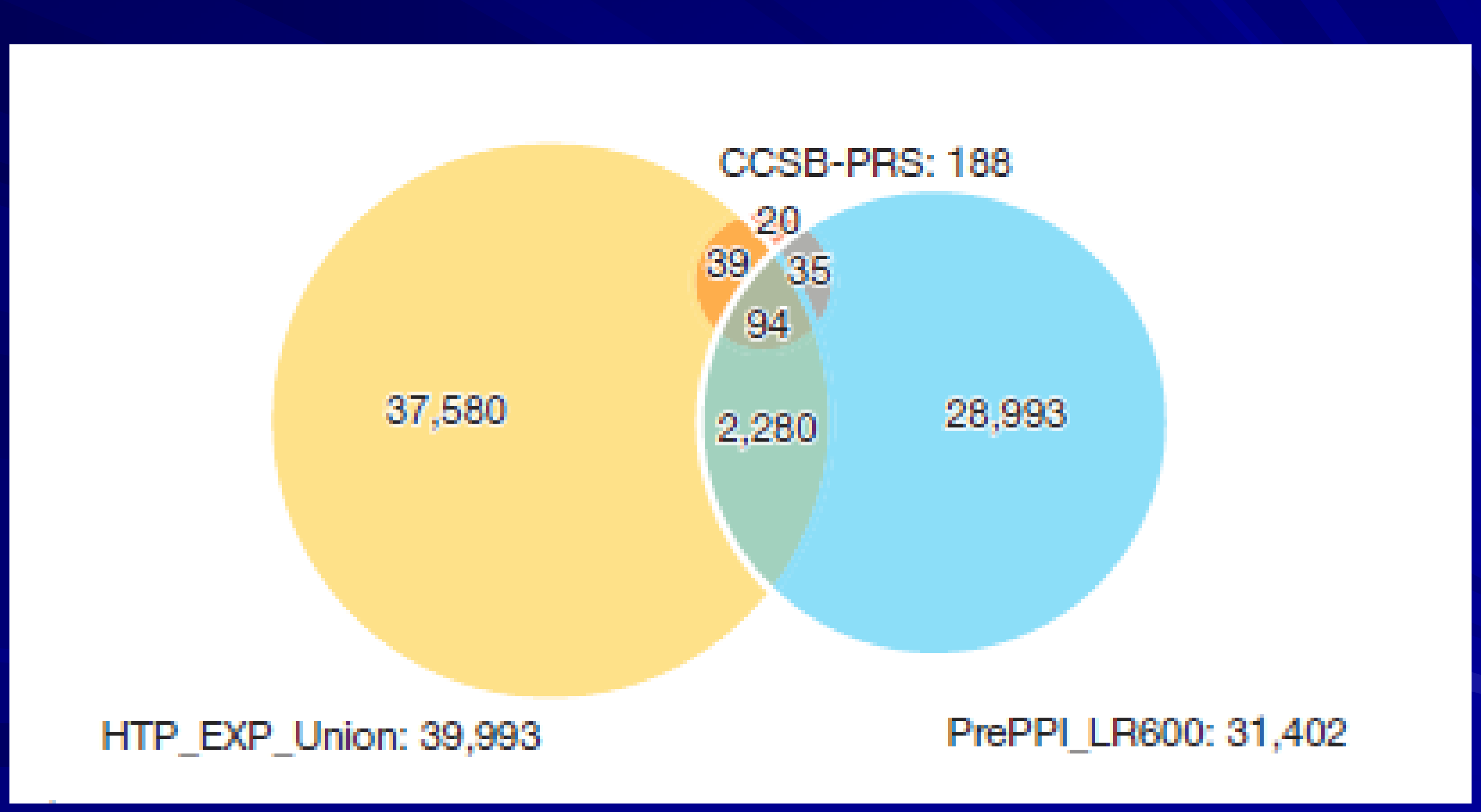
Predicting protein-protein interactions

Prediction results



Predicting protein-protein interactions

Prediction results



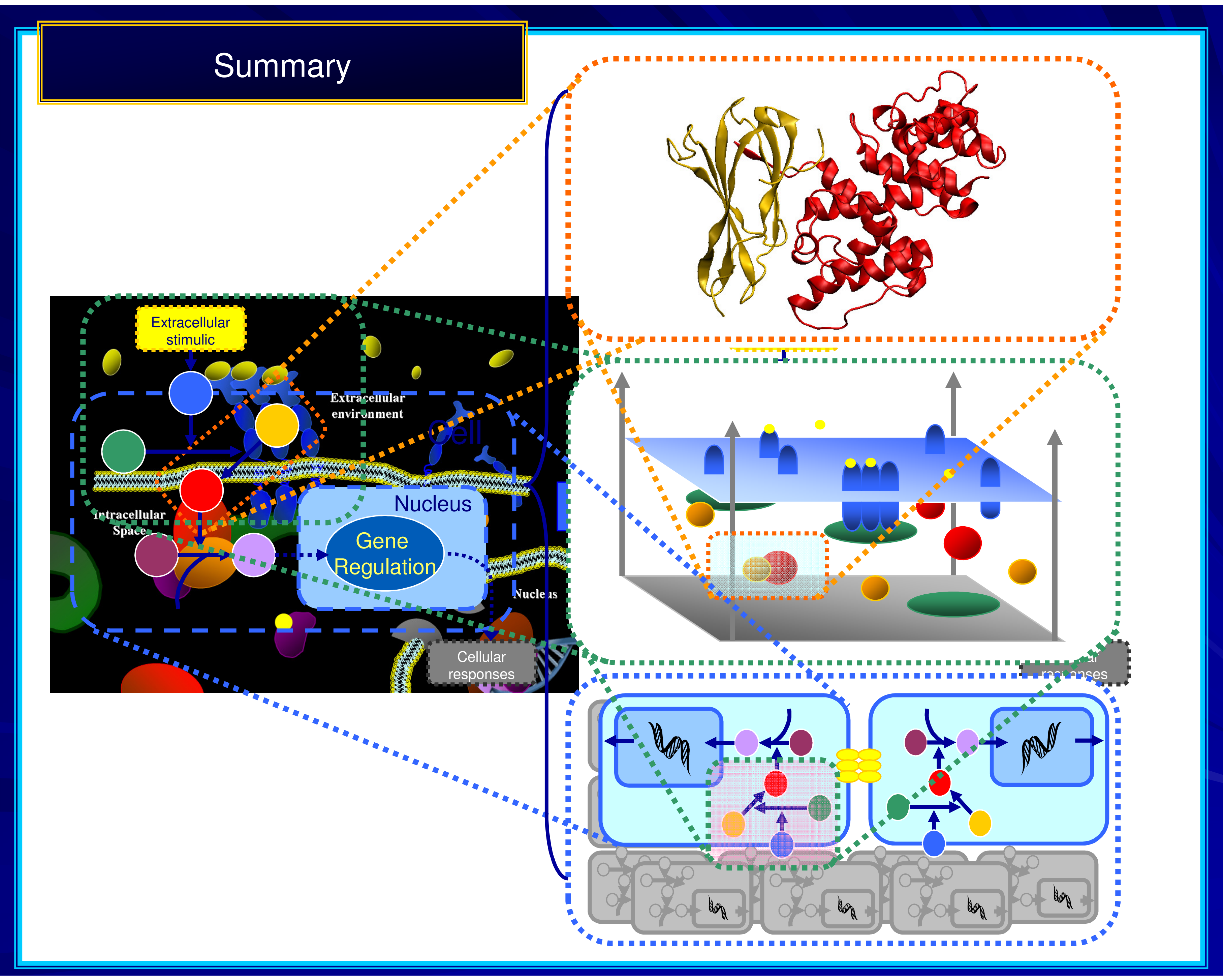
Predicting protein-protein interactions

Summary and Conclusions

- It has been demonstrated that even distantly related proteins often use regions of their surface with similar arrangements of secondary structure elements to bind to other proteins, considerably expanding the number of putative PPIs that can be identified.
- The study suggests the ability to add a structural 'face' for a large number of PPIs, and that structural biology can have an important role in molecular systems biology.
- Different methods with diverse backgrounds can be integrated to effectively remove false positives in PPI interactome.

Outline

- Modeling and Simulating Larger Molecular Systems with Longer Time Scales
  - Coarse-grained simulation
  - Multi-scale modeling
- Extend the Applications of Structural Biology to Newly Emerging Fields
  - Fold proteins by information from genetics
  - Simulating crowded cellular environment
  - Combining structural biology with systems biology
- **Summary and Perspective**

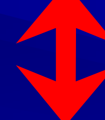

Summary
Extracellular stimulic
Extracellular environment
Cell
Intracellular Space
Nucleus
Gene Regulation
Cellular responses

Summary

Systems
Biology

Genetic
Evolution

Structural
Biology



Cell
Biology