Course outline

Theory **Practice**

Day 1

- **Introduction to structure determination**
- **Chromatin structure and Hi-C data**
- **Introduction to linux and python (FACULTATIVE)**
- **The Integrative Modeling Platform and Chimera**

Day 2

- **The Integrative Modeling Platform applied to chromatin**
- **TADbit introduction and installation**
- **Topologically Associated Domains detection and analysis**

Day 3

- **The TADbit documentation: examples and code snippets**
- **3D modeling of real Hi-C data**
- **Analysis of the results**

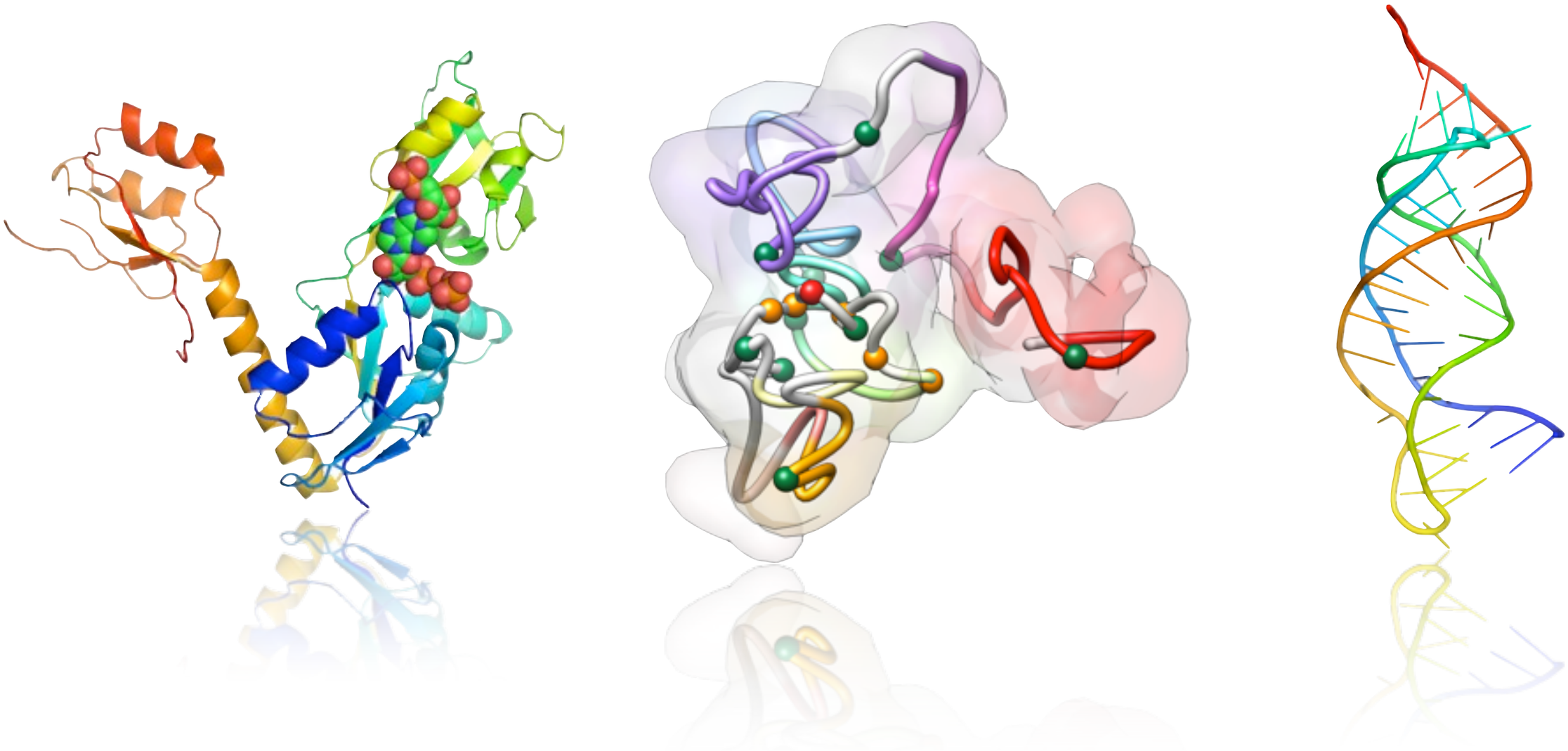
3D structure determination

Davide Baù & François Serra

Genome Biology Group (CNAG)
Structural Genomics Group (CRG)

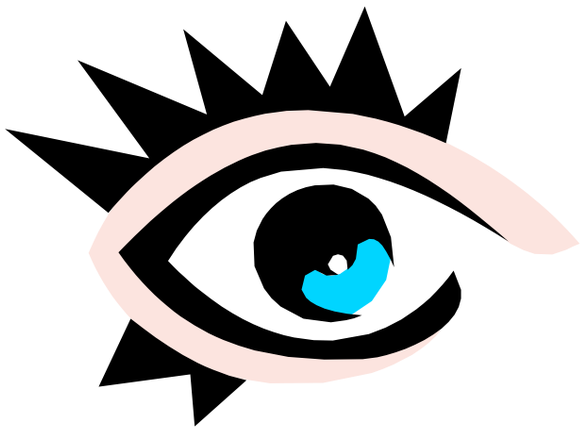
Structural Genomics Group

http://www.marciuslab.org

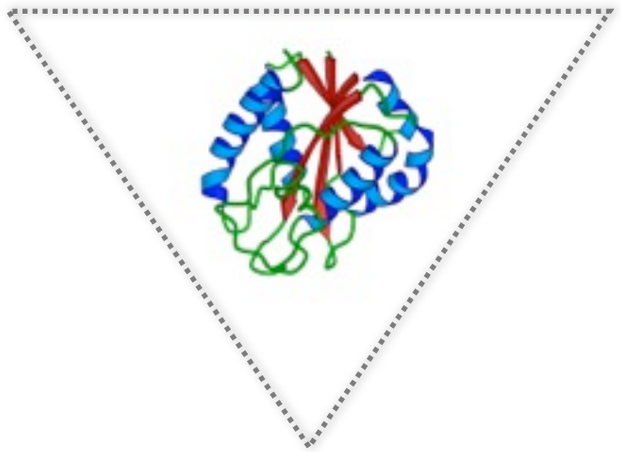


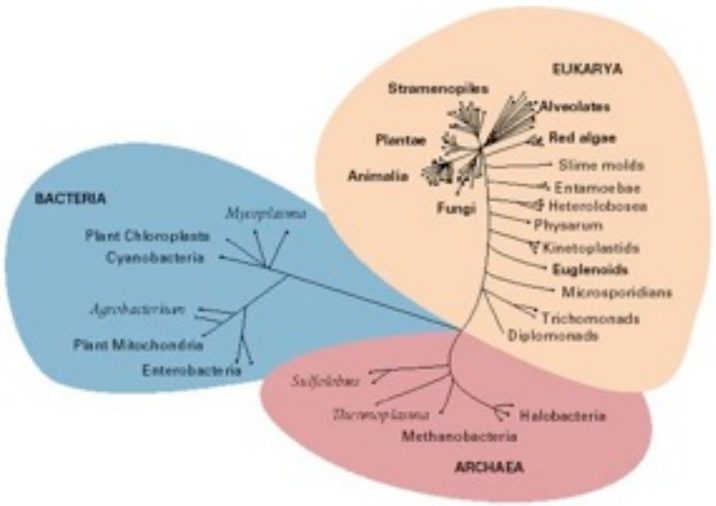
cnag CRG

Data groups

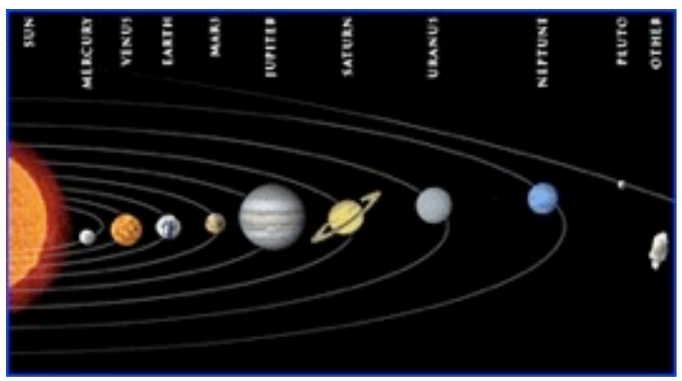


Experimental observations





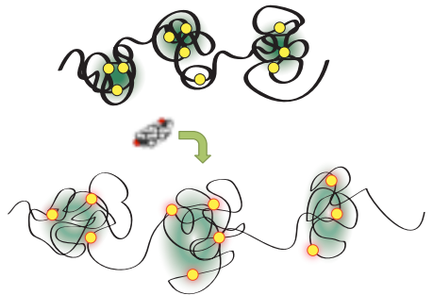
Statistical rules



Laws of physics

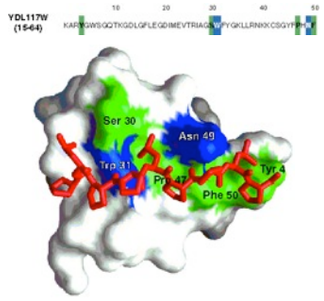
The importance of the 3D structure

The biochemical function of a molecule is defined by its interactions



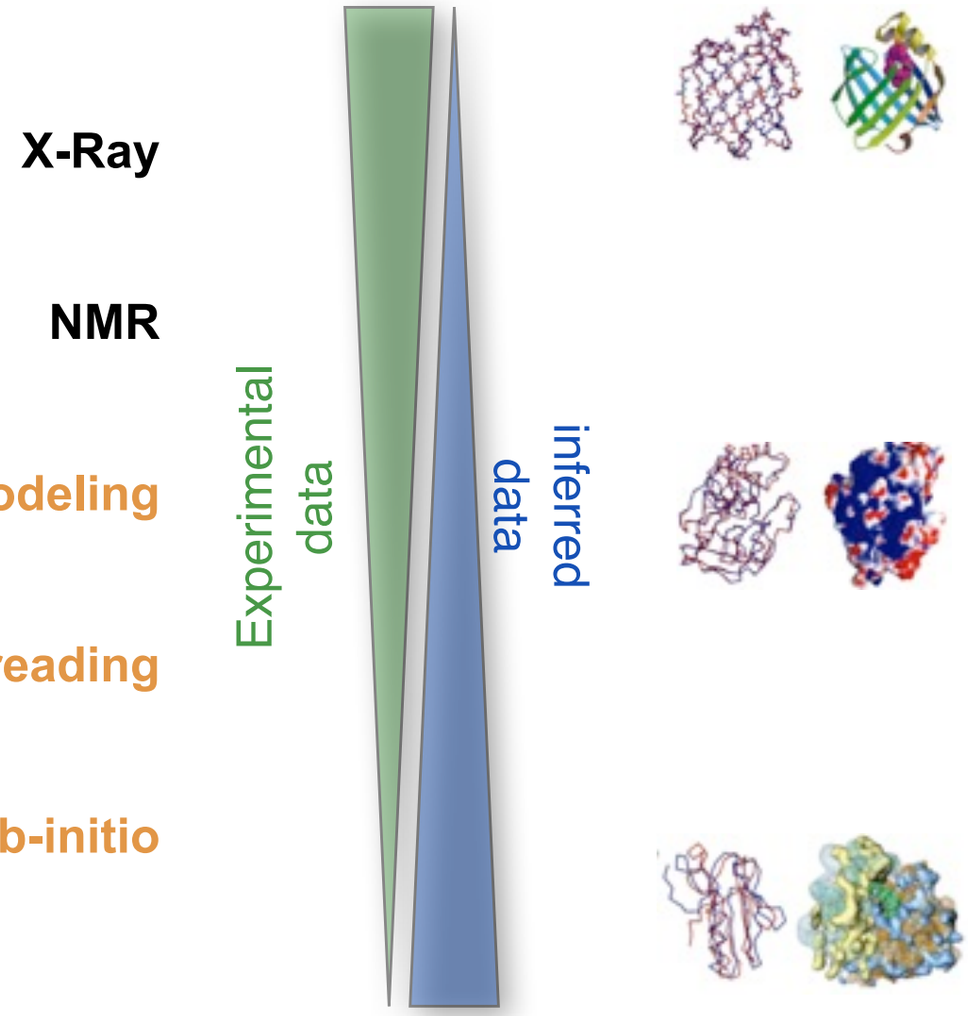
The biological function is in large part a consequence of these interations

The 3D structure is more informative than sequence alone

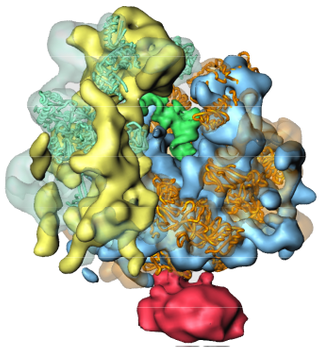


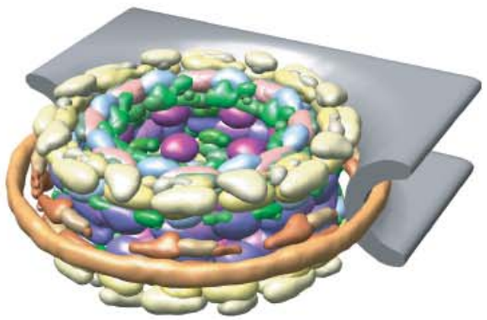
Evolution tends to conserve function and function depends more directly on structure than on sequence

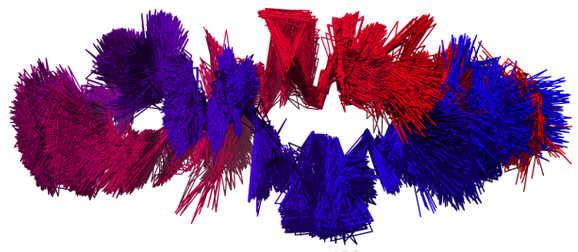
Structure prediction vs determination

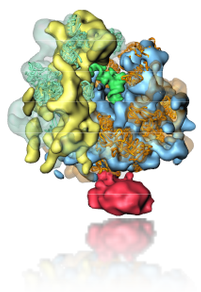


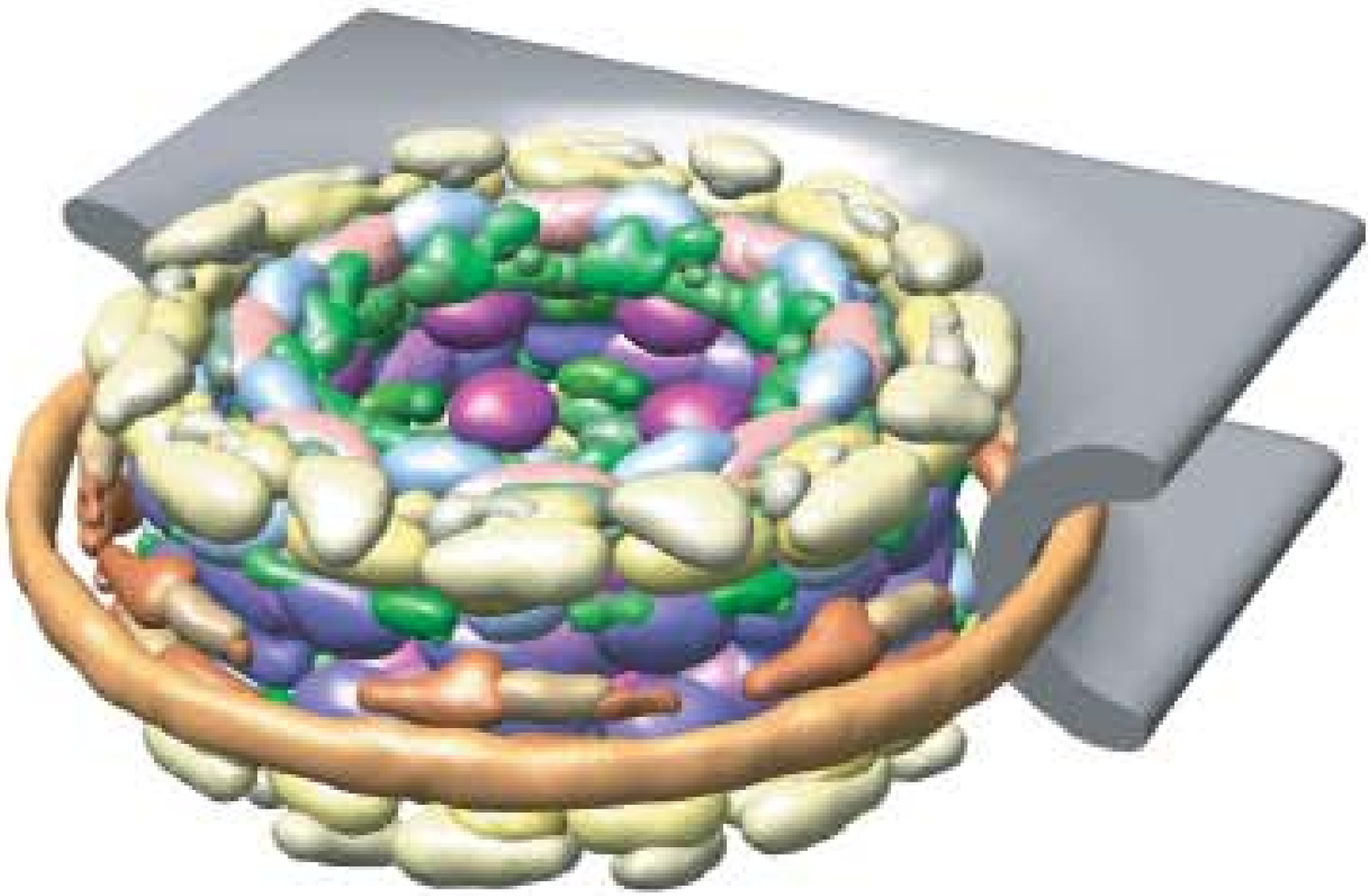
Data integration





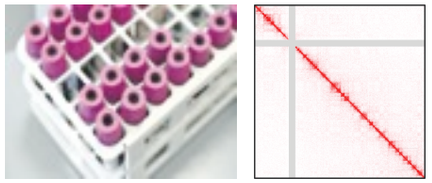






The four stages of integrative modeling

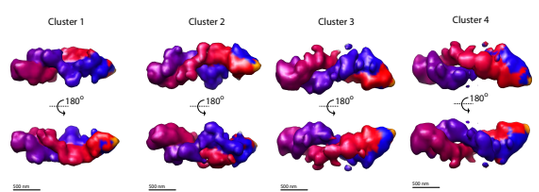
Stage 1: Gathering experimental and statistical Information



Stage 2: Choosing How To Represent And Evaluate Models

Stage 3: Finding Models That Score Well

Stage 4: Analyzing Resulting Models and Information



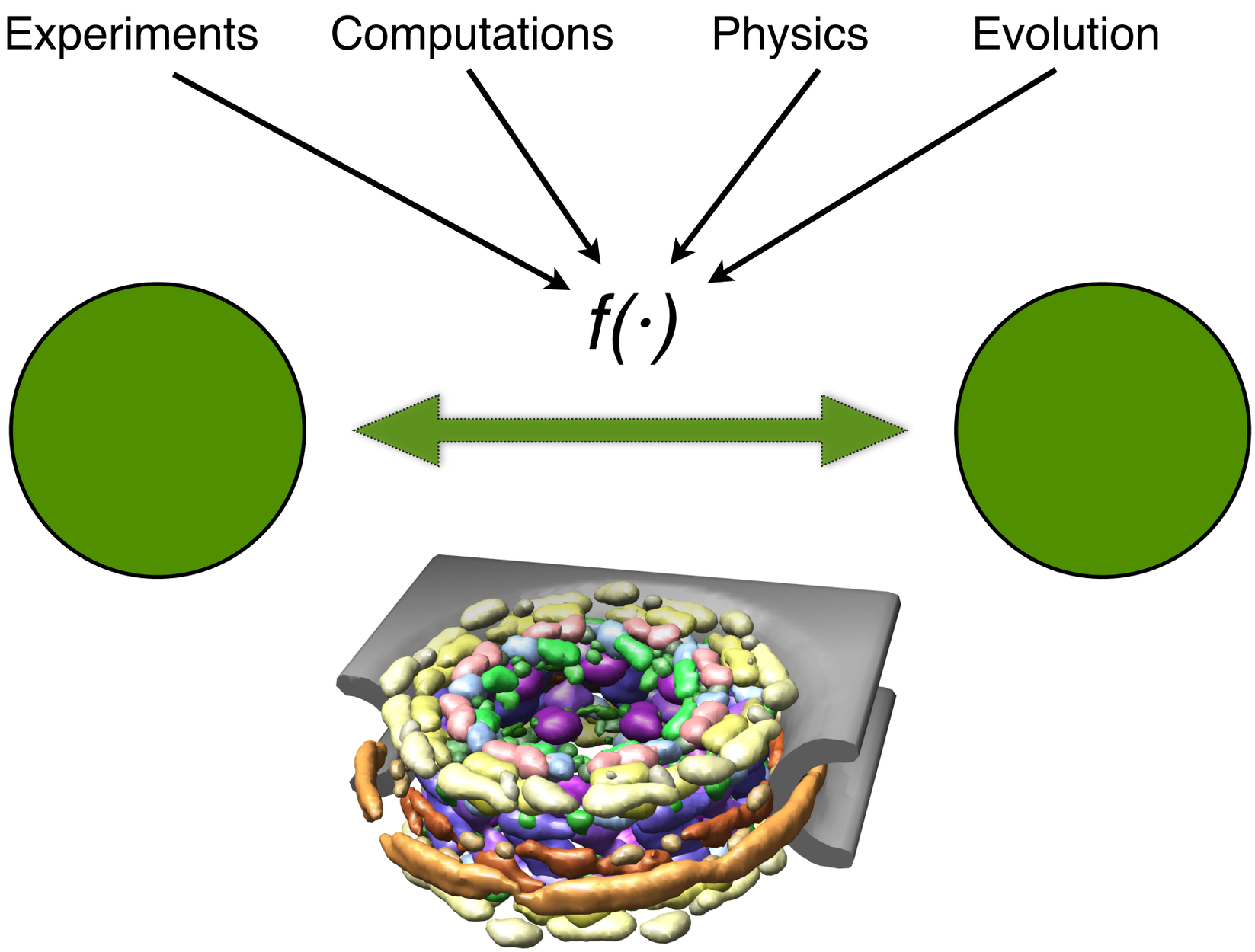
Advantages of integrative modeling

- **It facilitates the use of new information**
- **It maximizes accuracy, precision and completeness of the models**
- **It facilitates assessing the input information and output models**
- **It helps in understanding and assessing experimental accuracy**

Russel, D., Lasker, K., Webb, B., Velázquez-Muriel, J., Tjioe, E., Schneidman-Duhovny, D., Peterson, B., et al. (2012). *PLoS Biology*, *10*(1), e1001244

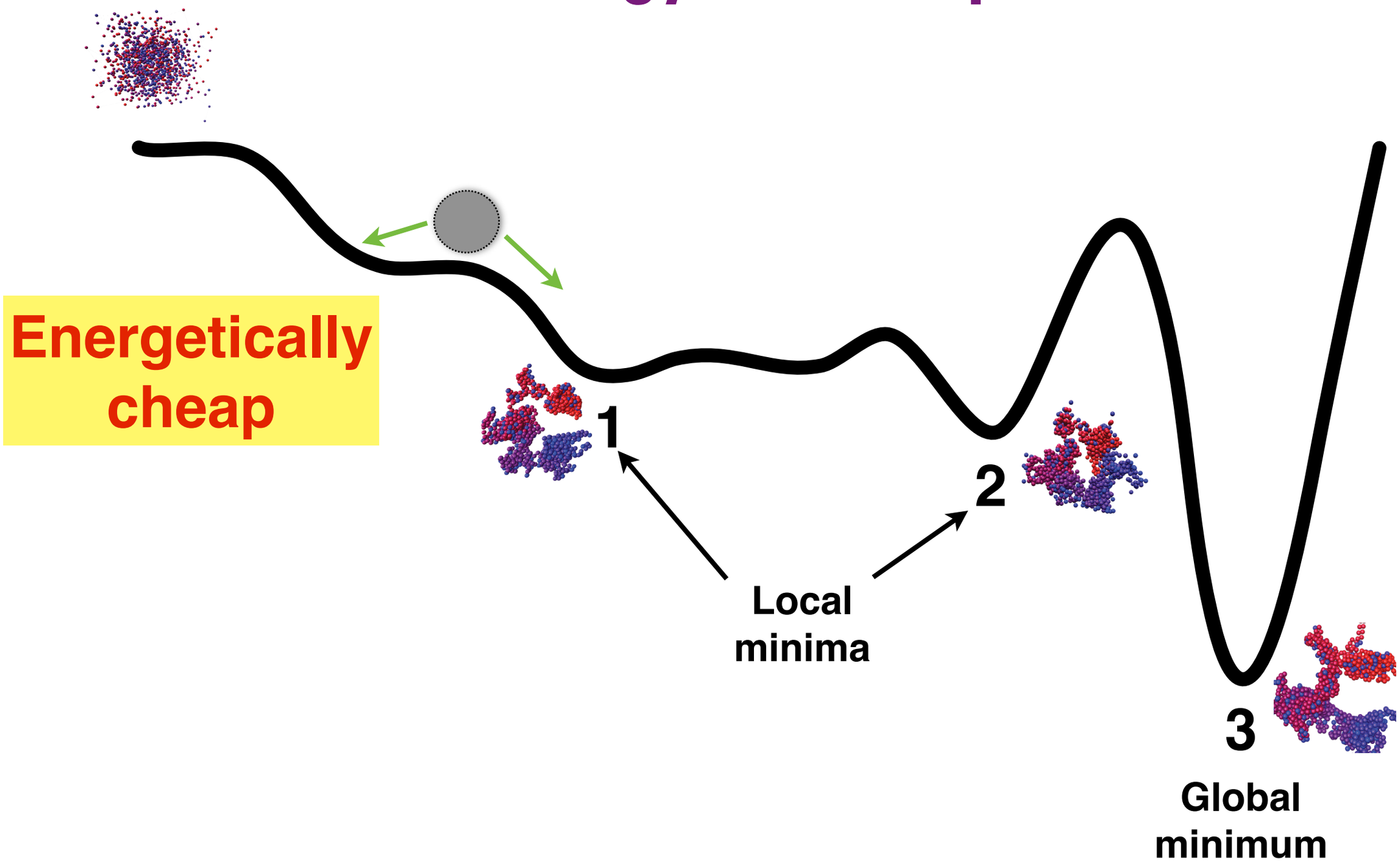
Integrative Modeling Platform

http://www.integrativemodeling.org

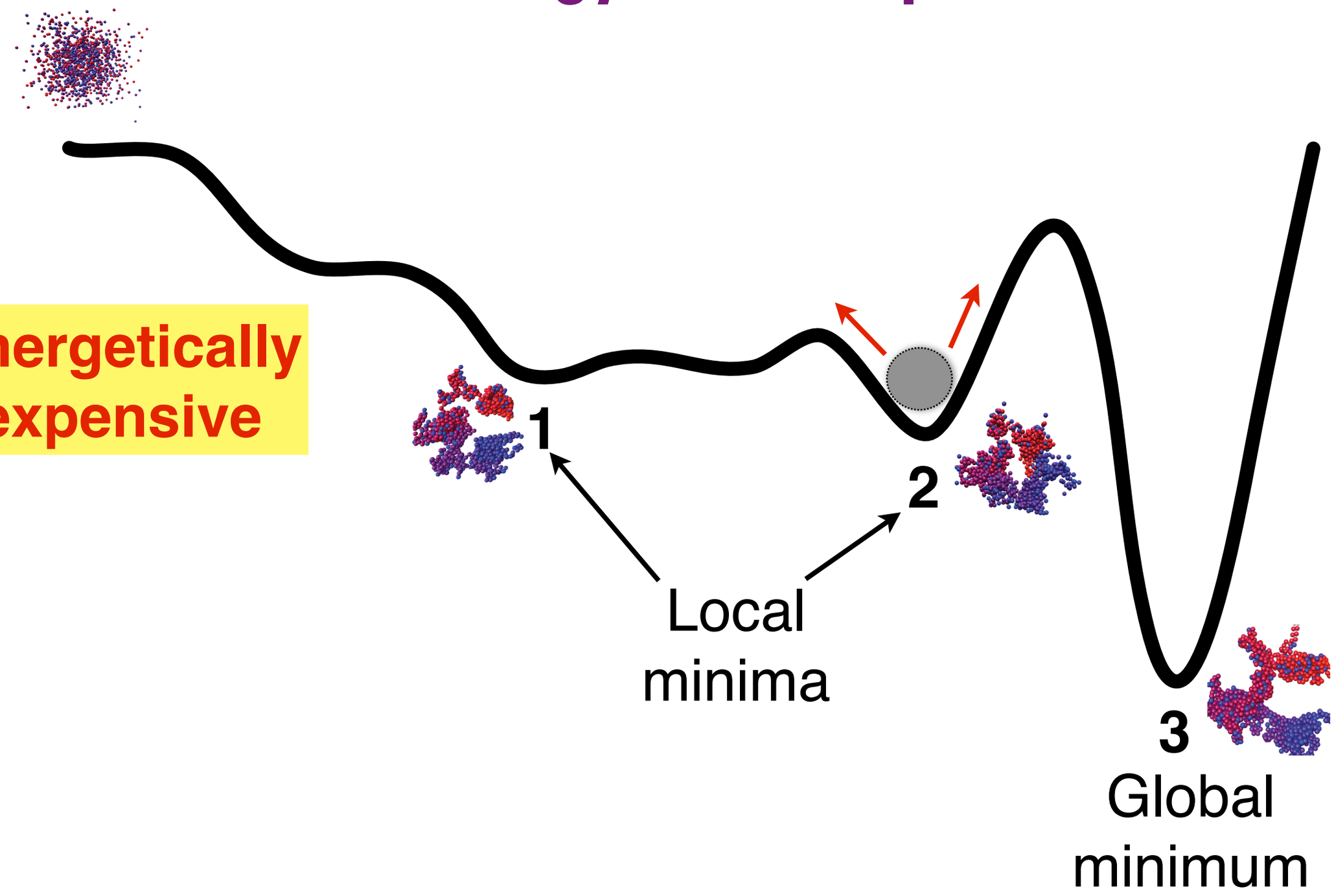


From: Russel, D. et al. PLOS Biology 10, e1001244 (2012).

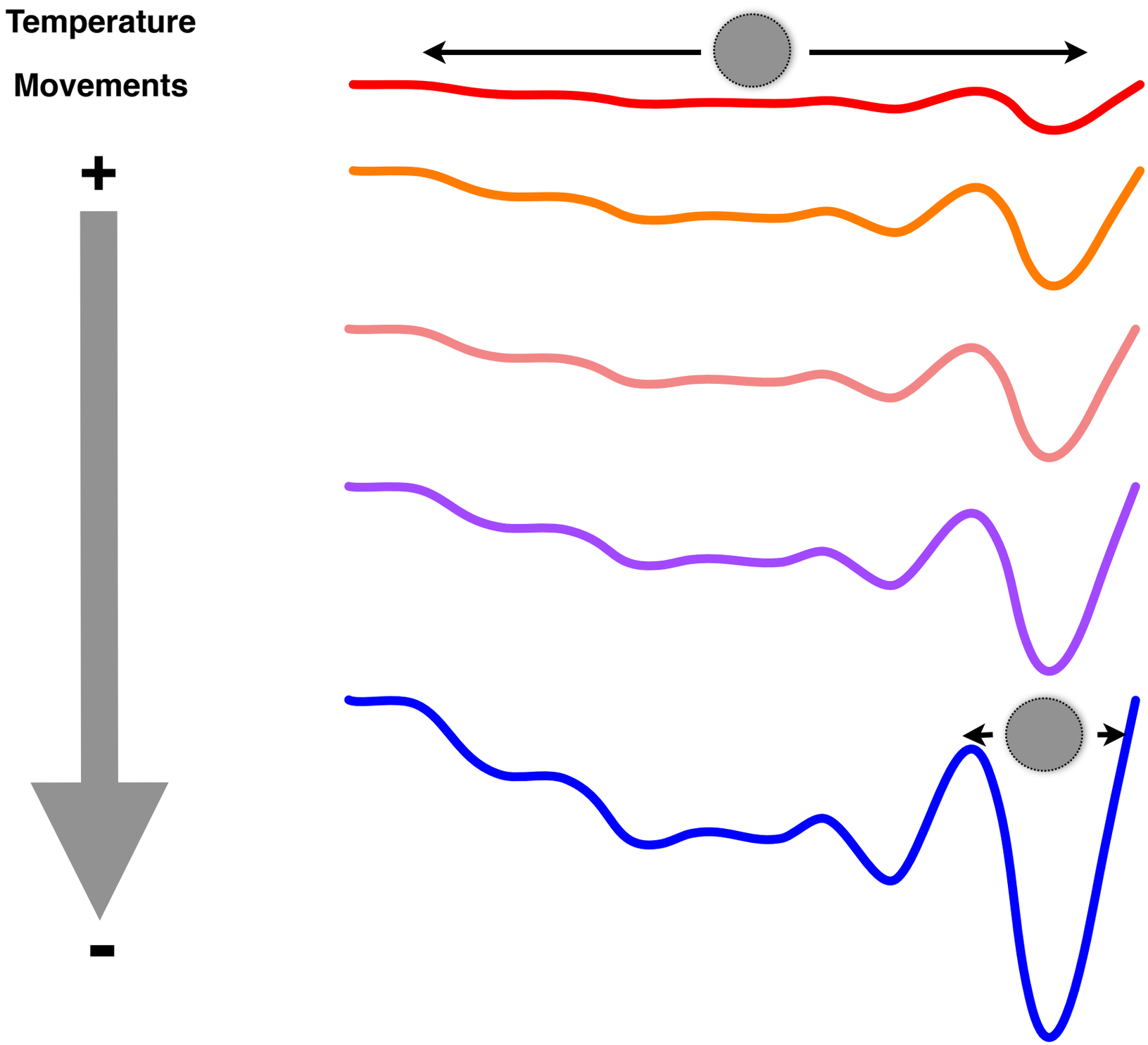
Energy landscape



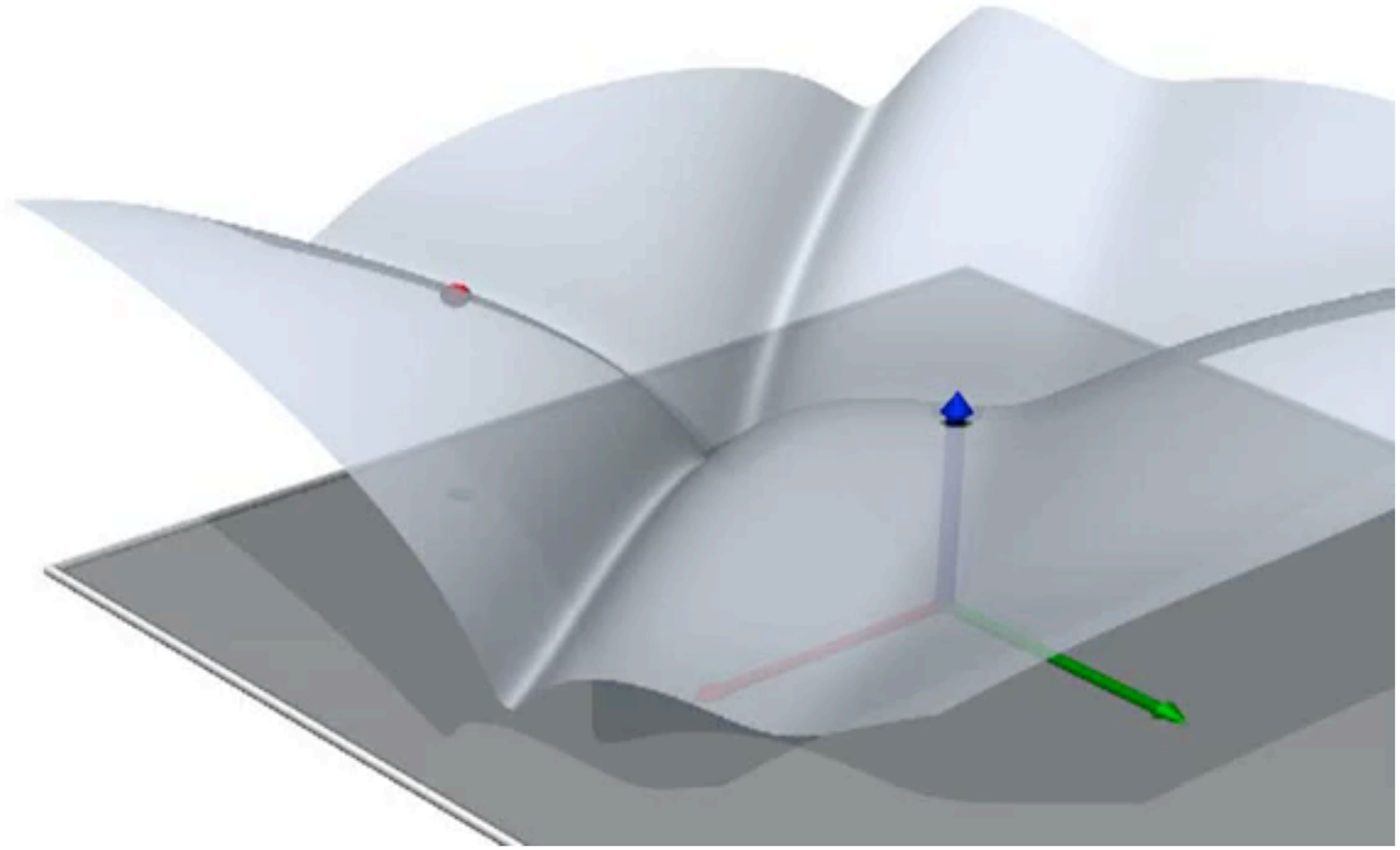
Energy landscape



The simulating annealing procedure

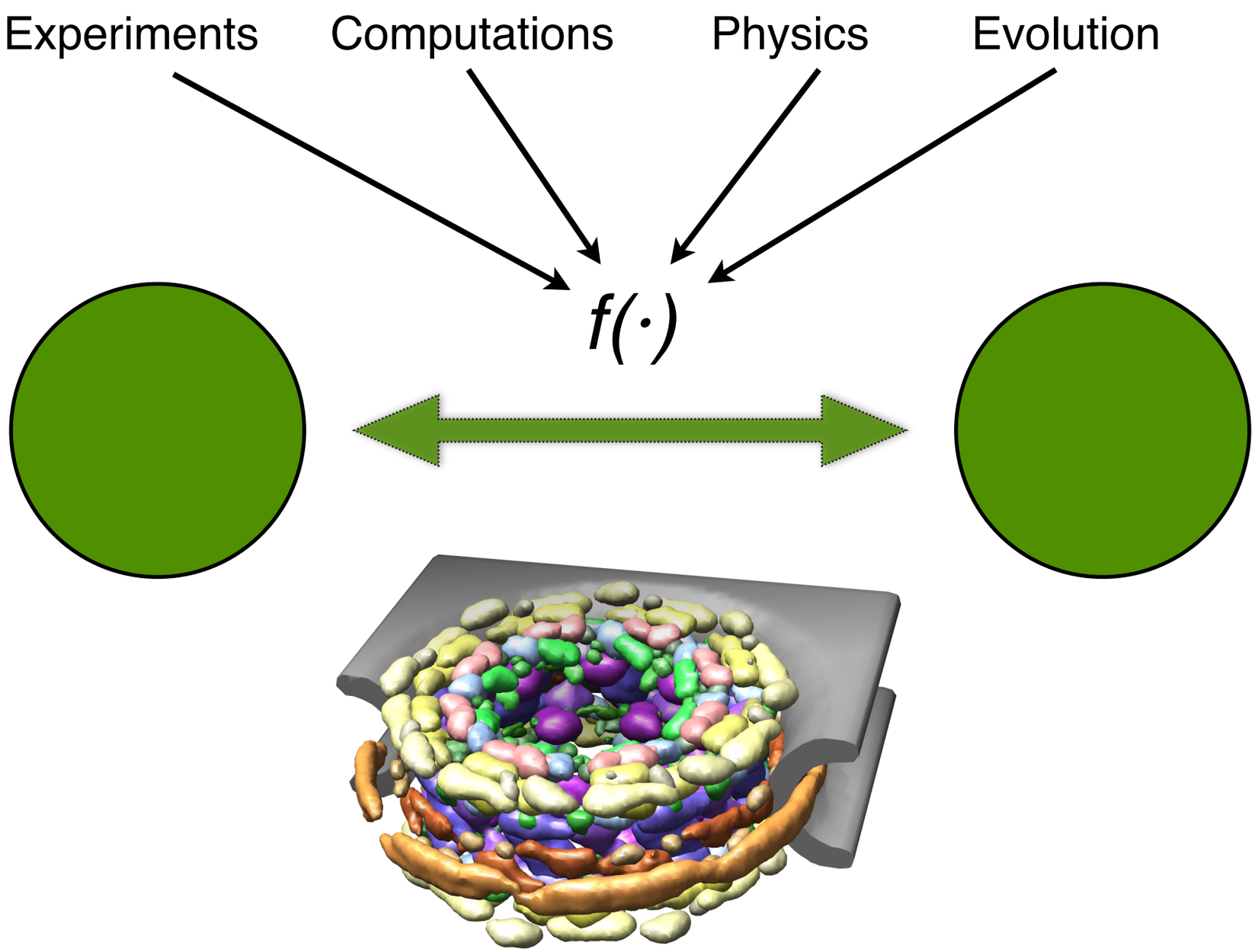


En example of nergy optimization



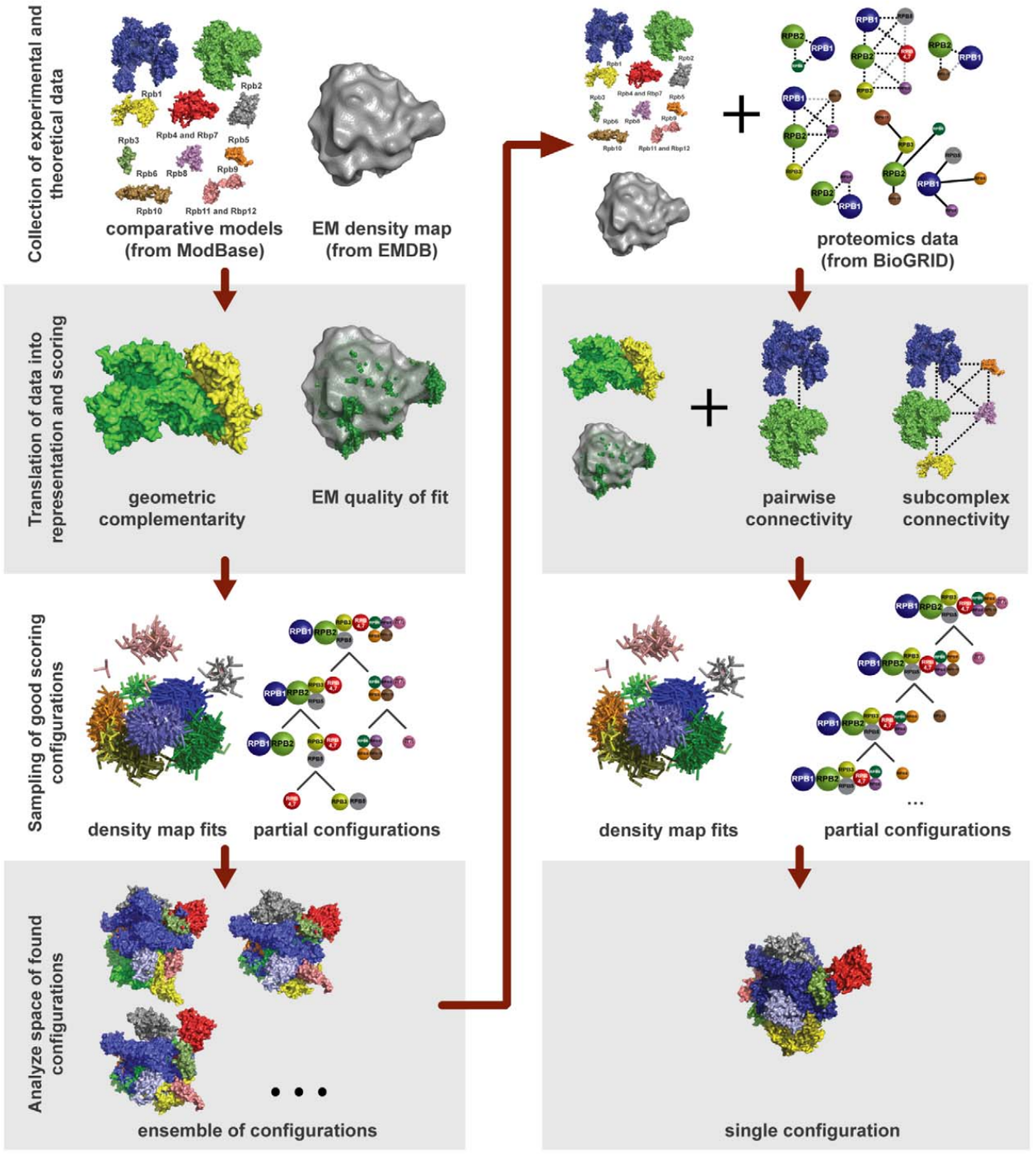
Integrative Modeling Platform

http://www.integrativemodeling.org



From: Russel, D. et al. PLOS Biology 10, e1001244 (2012).

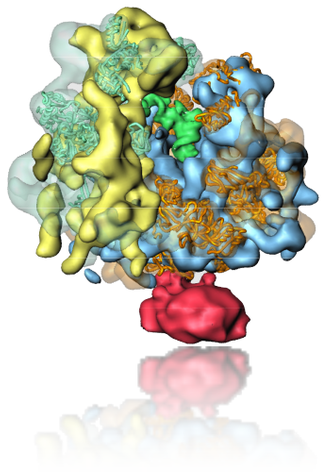
"Toy" example...

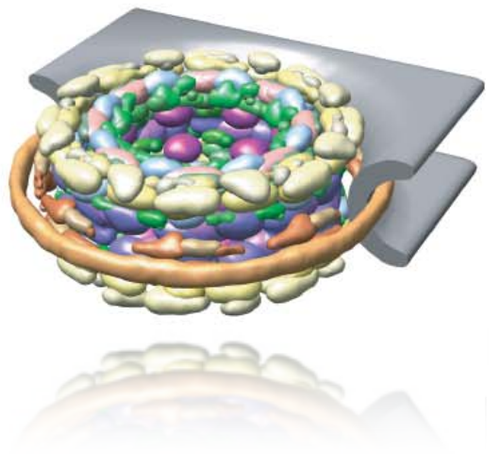


Russel, D., Lasker, K., Webb, B., Velázquez-Muriel, J., Tjioe, E., Schneidman-Duhovny, D., Peterson, B., et al. (2012). *PLoS Biology*, *10*(1), e1001244

"Real" examples

PROTEINS





COMPLEXES

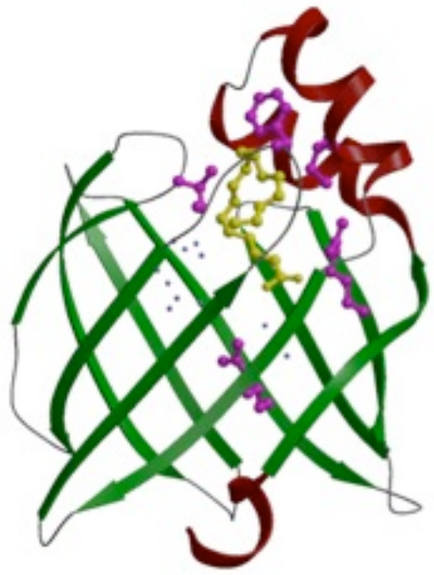
GENOMES

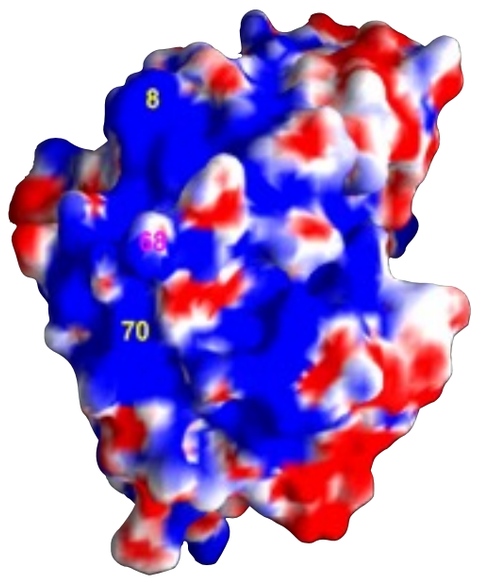
Proteins

Single data type

Amino Acids

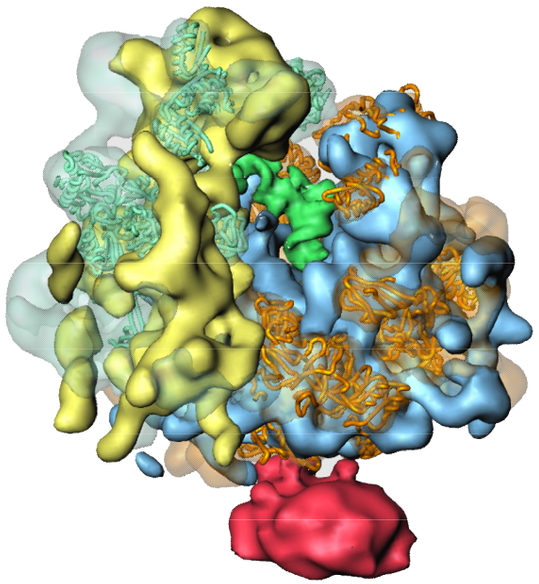
X-Ray;
NMR;
Modeling

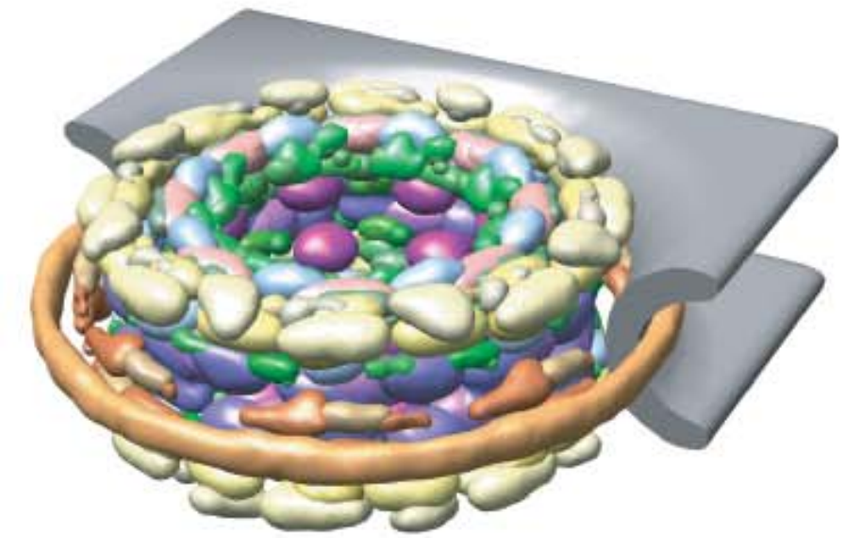




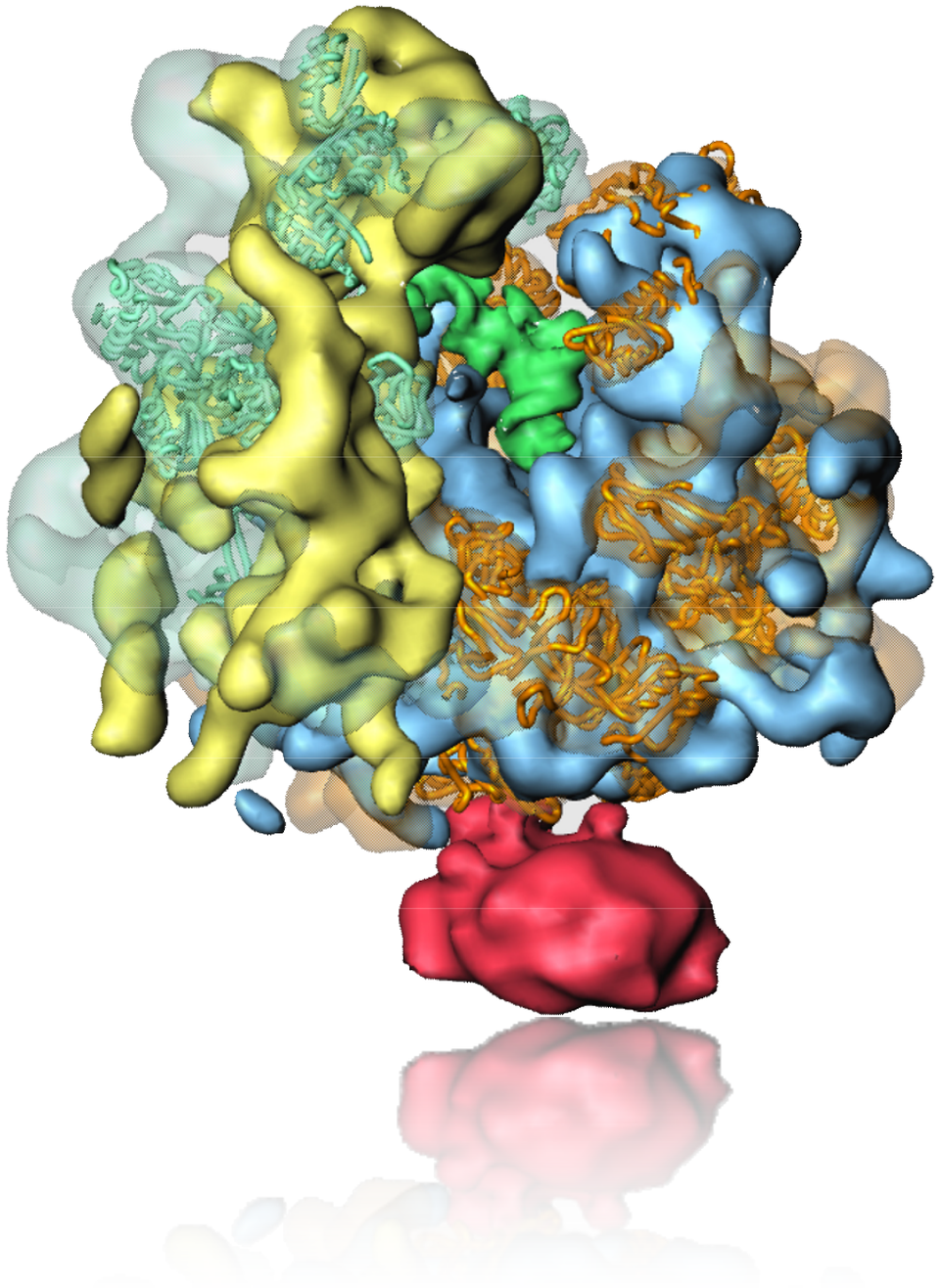
Complexes

Multiple data types





S. cerevisiae ribosome



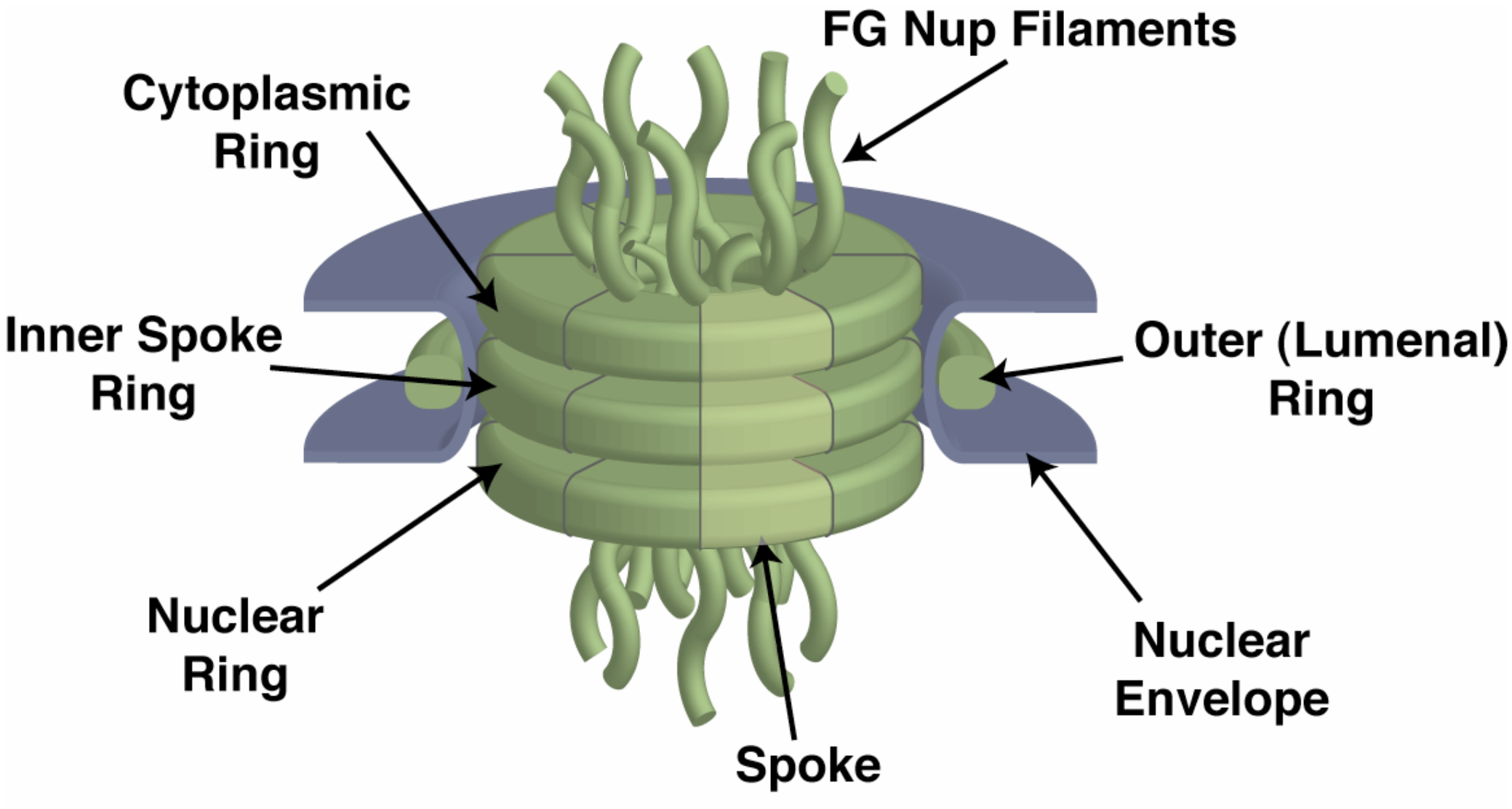
Fitting of comparative models into 15Å cryo-electron density map.

43 proteins could be modeled on 20-56% seq.id. to a known structure.

The modeled fraction of the proteins ranges from 34-99%.

C. Spahn, R. Beckmann, N. Eswar, P. Penczek, A. Sali, G. Blobel, J. Frank. Cell 107, 361-372, 2001.

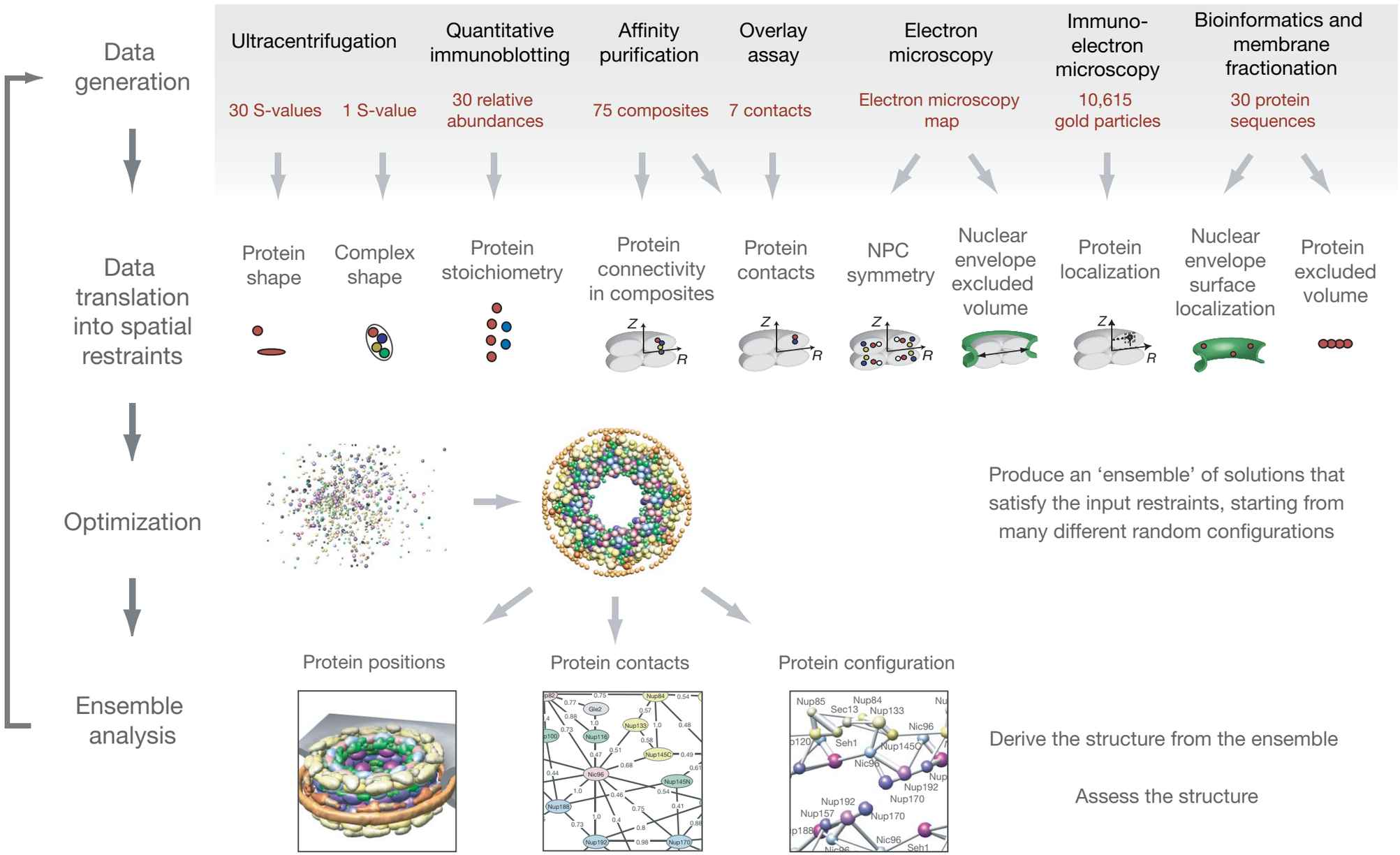
The nuclear pore complex



Alber, F., Dokudovskaya, S., Veenhoff, L. M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., et al. (2007). *Nature*, *450*(7170), 695–701

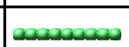
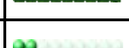
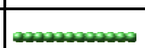
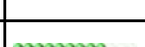
Integrative Modeling of the NPC

F. Alber et al. Natute (2007) Vol 450



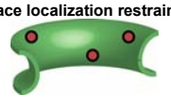
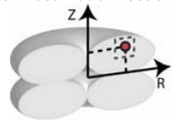
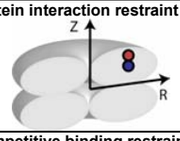
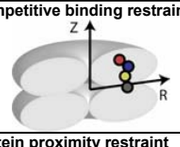
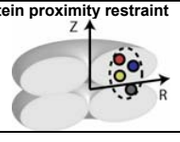
Representation

436 proteins!

τ	N_τ^1	N_τ^2	κ	$\{B_j^\kappa\}$	n_κ	r	τ	N_τ^1	N_τ^2	κ	$\{B_j^\kappa\}$	n_κ	r
Nup192	1	1	1,2,5		2	3.0	**Nup1**	0	1	1,5		9	1.5
			3	-	1	-				2		2	1.5
Nup188	1	1	1,2,5		2	3.0				3	-	1	-
			3	-	1	-				4		7	1.5
Nup170	1	1	1,2,5		2	2.9	**Nsp1**	2	2	1,5		12	1.3
			3	-	1	-				2		3	1.3
Nup157	1	1	1,2,5		3	2.5				3	-	1	-
			3	-	1	-				4		9	1.3
Nup133	1	1	1,2,5		2	2.7	**Gle1**	1	0	1,2,5		2	2.1
			3	-	1	-				3	-	1	-
Nup120	1	1	1,2,5		2	2.6	**Nup60**	0	1	1,5		4	1.6
			3	-	1	-				2,3		1	1.6
Nup85	1	1	1,2,5		3	2.0				4		3	1.6
			3	-	1	-	**Nup59**	1	1	1,5		4	1.6
Nup84	1	1	1,2,5		3	2.0				2		2	1.6
			3	-	1	-				3	-	1	-
Nup145C	1	1	1,2,5		2	2.3				4		2	1.6
			3	-	1	-	**Nup57**	1	1	1,5		3	1.8
Seh1	1	1	1,2,3,5		1	2.2				2,3		1	1.8
Sec13	1	1	1,2,3,5		1	2.1				4		2	1.8
Gle2	1	1	1,2,3,5		1	2.3	**Nup53**	1	1	1,5		3	1.7
Nic96	2	2	1,2,5		2	2.4				2,3		1	1.7
			3	-	1	-				4		2	1.7
Nup82	1	1	1,2,5		2	2.3	**Nup145N**	0	2	1,5		6	1.5
			3	-	1	-				2,3		1	1.5

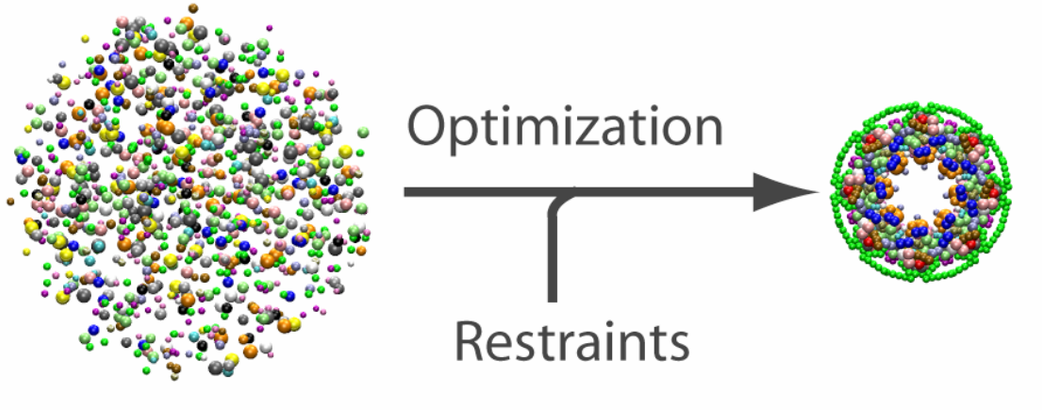
Alber, F., Dokudovskaya, S., Veenhoff, L. M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., et al. (2007). *Nature*, *450*(7170), 695–701

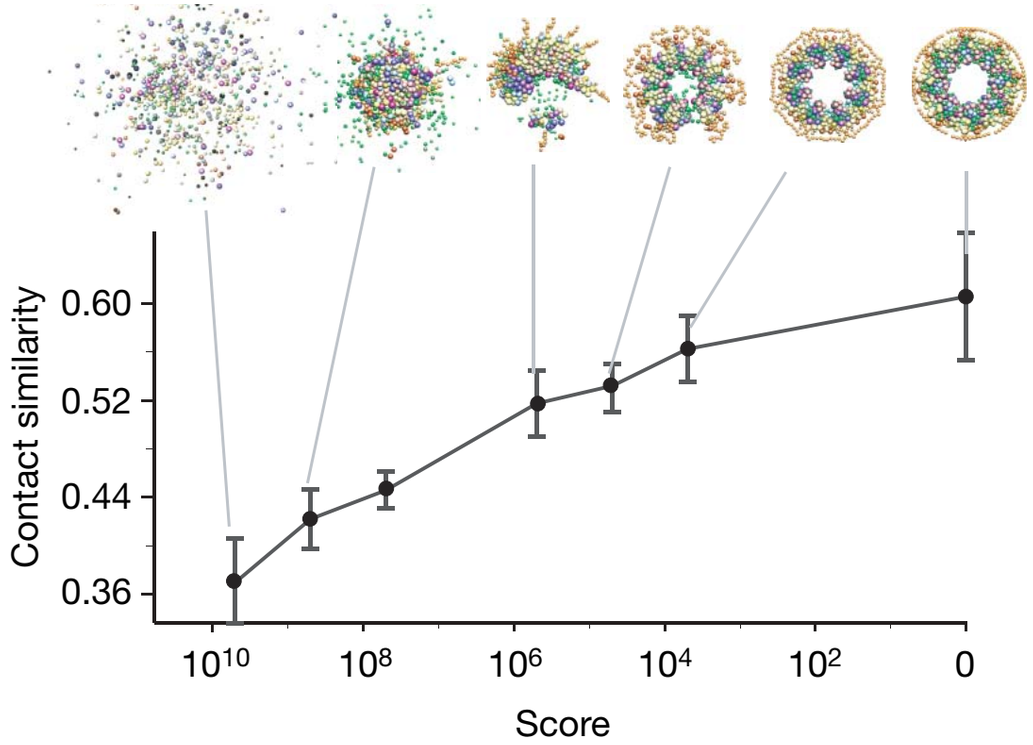
Scoring

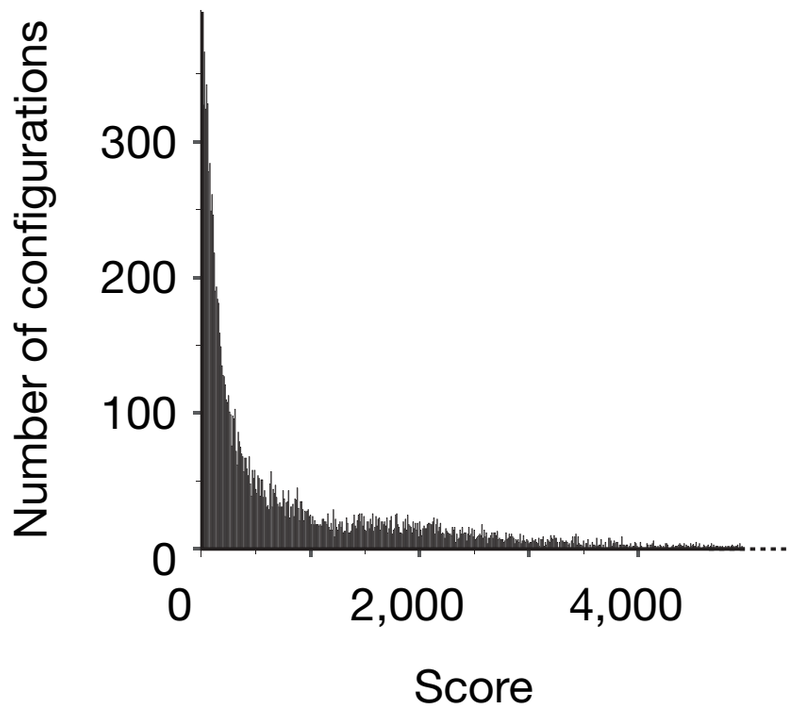
Data generation		Data interpretation				
Method	**Experiments**	**Restraint**	R_C	R_O	R_A	**Functional form of activated feature restraint**
Bioinformatics and Membrane fractionation	**30 nup sequences**	**Protein excluded volume restraint** 	-	-	**1,864 *1,863/2**	**Protein-protein**: Violated for $f < f_o$. f is the distance between two beads, f_o is the sum of the bead radii, and σ is 0.01 nm. Applied to all pairs of particles in representation κ=1: $B^{res} = \{B_j^{\kappa=1}(\theta,s,\tau,i)\}$
	30 nup sequences	**Surface localization restraint** 	-	-	**48**	**Membrane-surface location:** Violated if $f \neq f_o$. f is the distance between a protein particle and the closest point on the NE surface (half-torus), f_o = 0 nm, and σ is 0.2 nm. Applied to particles: $B^{res} = \{B_j^{\kappa=6}(\theta,s,\tau,i) \mid \tau \in (\text{Ndc1,Pom152,Pom34})\}$
	30 Nup sequences and immuno-EM (see below)		-	-	**64**	**Pore-side volume location:** Violated if $f < f_o$. f is the distance between a protein particle and the closest point on the NE surface (half-torus), f_o = 0 nm, and σ is 0.2 nm. Applied to particles: $B^{res} = \{B_j^{\kappa=8}(\theta,s,\tau,i) \mid \tau \in (\text{Ndc1,Pom152,Pom34})\}$
			-	-	**80**	**Perinuclear volume location:** Violated if $f > f_o$,, f is the distance between a protein particle and the closest point on the NE surface (half-torus), f_o = 0 nm, and σ is 0.2 nm. Applied to particles: $B^{res} = \{B_j^{\kappa=7}(\theta,s,\tau,i)\tau \in (\text{Pom152})\}$
Hydrodynamics experiments	**1 S-value**	**Complex shape restraint** 	**1**	**164**	**1**	**Complex diameter** Violated if $f < f_o$. f is the distance between two protein particles representing the largest diameter of the largest complex, f_o is the complex maximal diameter $D=19.2\text{-}R$, where R is the sum of both particle radii, and σ is 0.01 nm. Applied to particles of proteins in composite C_{45}: $B^{res} = \{B_j^{\kappa=1}(\theta,s,\tau,i) \mid \tau \in C_{51}\}$
	30 S-values	**Protein chain restraint** 	-	-	**1,680**	**Protein chain** Violated if $f \neq f_o$. f is the distance between two consecutive particles in a protein, f_o is the sum of the particle radii, and σ is 0.01 nm. Applied to particles: $B = \{B_j^{\kappa}(\theta,s,\tau,i) \mid \kappa = 1\}$
Immuno-Electron microscopy	**10,940 gold particles**	**Protein localization restraint** 	-	-	**456**	**Z-axial position** Violated for $f < f_o$. f is the absolute Cartesian Z-coordinate of a protein particle, f_o is the lower bound defined for protein type τ, and σ is 0.1 nm. Applied to particles: $B = \{B_j^{\kappa}(\theta,s,\tau,i) \mid \kappa = 1, j = 1\}$
					456	Violated for $f > f_o$. f is the absolute Cartesian Z-coordinate of a protein particle, f_o is the upper bound defined for protein type τ, and σ is 0.1 nm. Applied to particles: $B = \{B_j^{\kappa}(\theta,s,\tau,i) \mid \kappa = 1, j = 1\}$
			-	-	**456**	**Radial position** Violated for $f < f_o$. f is the radial distance between a protein particle and the Z-axis in a plane parallel to the X and Y axes, f_o is its lower bound defined for protein type τ, and σ is 0.1 nm. Applied to particles: $B = \{B_j^{\kappa}(\theta,s,\tau,i) \mid \kappa = 1, j = 1\}$
					456	Violated for $f > f_o$. f is the radial distance between a protein particle and the Z-axis in a plane parallel to the X and Y axes, f_o is its upper bound defined for protein type τ, and σ is 0.1 nm. Applied to particles: $B = \{B_j^{\kappa}(\theta,s,\tau,i) \mid \kappa = 1, j = 1\}$
Overlay assays	**13 contacts**	**Protein interaction restraint** 	**20**	**112**	**20**	**Protein contact** Violated for $f > f_o$. f is the distance between two protein particles, f_o is the sum of the particle radii multiplied by a tolerance factor of 1.3, and σ is 0.01 nm. Applied to particle: $B = \{B_j^{\kappa}(\theta,s,\tau,i) \mid \kappa \in (2,4,9), \theta \in (1,2,3)\}$
Affinity purification	**4 complexes**	**Competitive binding restraint** 	**1**	**132**	**4**	**Protein contact** Violated for $f > f_o$. f is the distance between two protein particles, f_o is the sum of the particle radii multiplied by a tolerance factor of 1.3, and σ is 0.01 nm. Applied to : $B = \{B_j^{\kappa}(\theta,s,\tau,i) \mid \theta \in (1,2,3), \kappa \in (2,4,6), \tau = (Nup82, Nic96, Nup49, Nup57)\}$
	64 complexes	**Protein proximity restraint** 	**692**	**25,348**	**692**	**Protein proximity** Violated for $f > f_o$. f is the distance between two protein particles, f_o is the maximal diameter of a composite complex, and σ is 0.01 nm. Applied to particles: $B = \{B_j^{\kappa}(\theta,s,\tau,i) \mid \theta \in (1,2,3), \kappa \in (2,4,9)\}$

Alber, F., Dokudovskaya, S., Veenhoff, L. M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., et al. (2007). *Nature*, *450*(7170), 695–701

Optimization

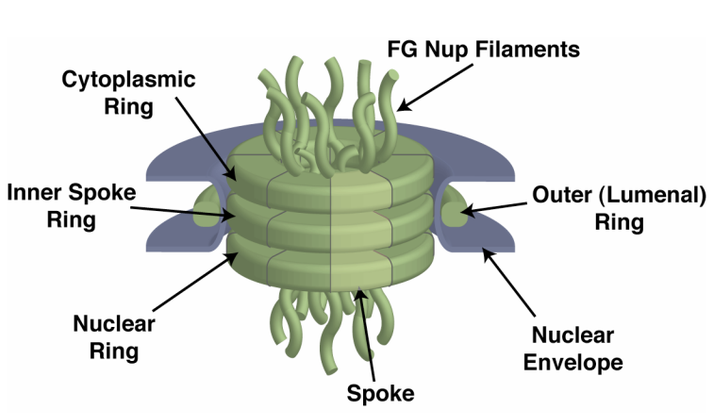


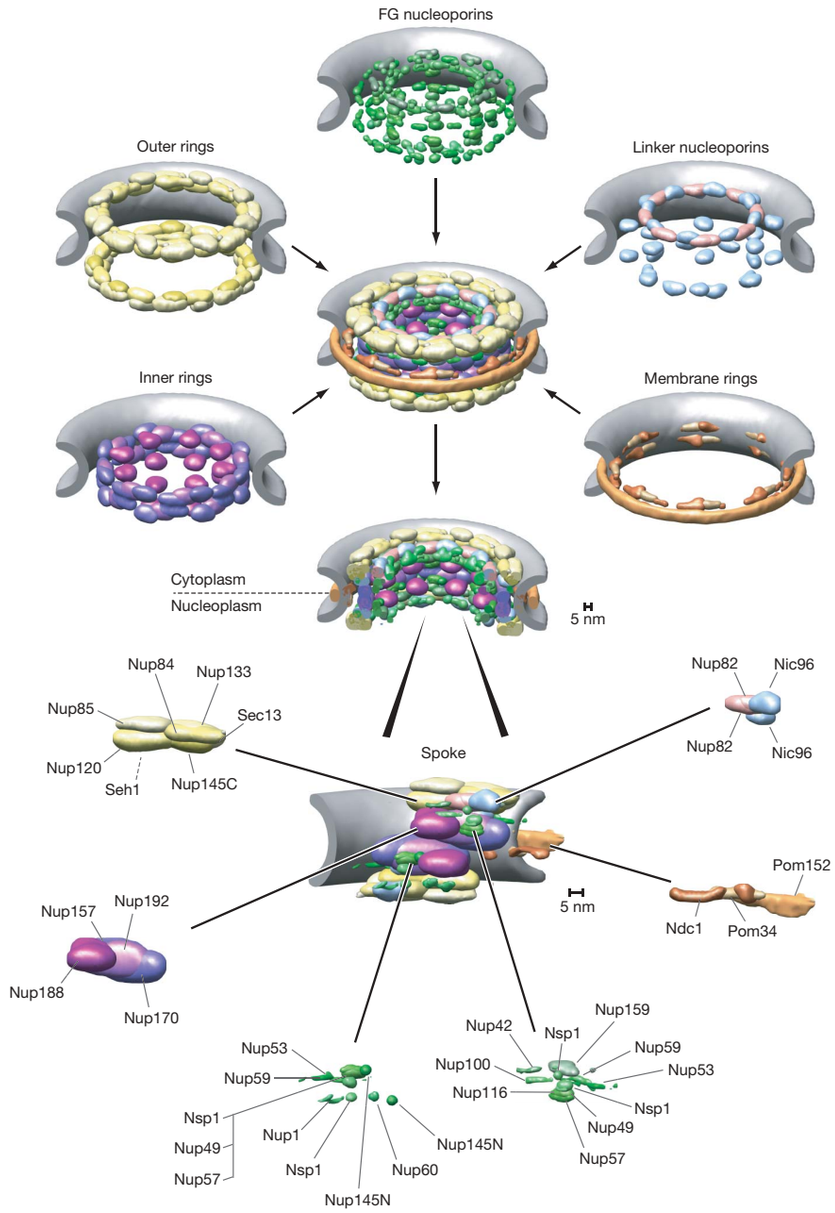




Alber, F., Dokudovskaya, S., Veenhoff, L. M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., et al. (2007). *Nature*, *450*(7170), 695–701

The structure of the nuclear pore complex

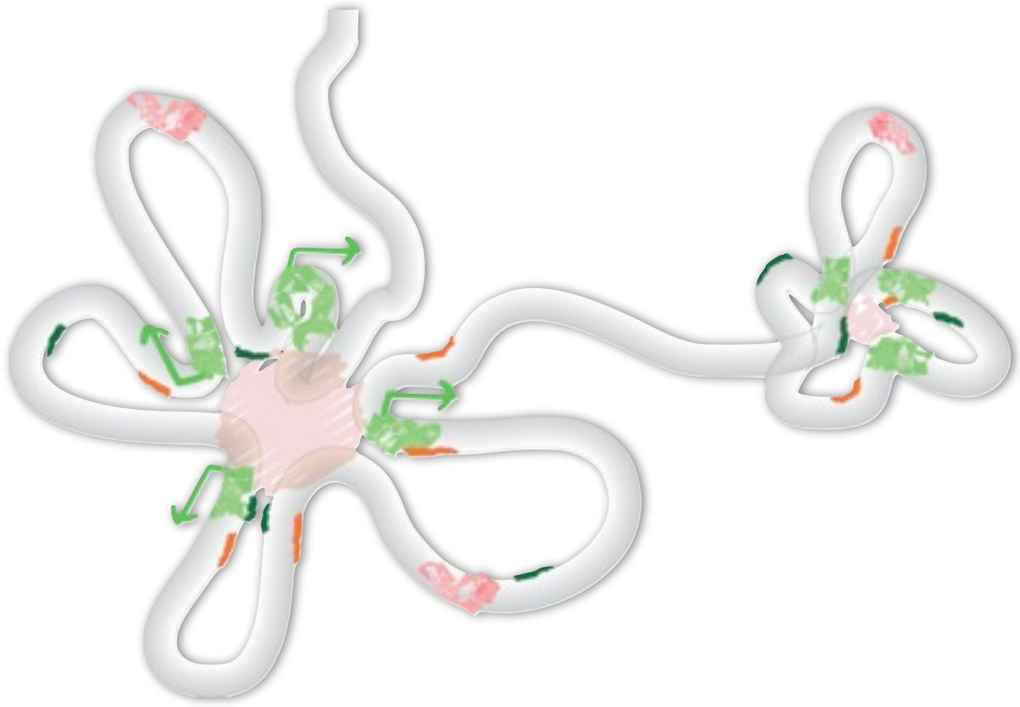


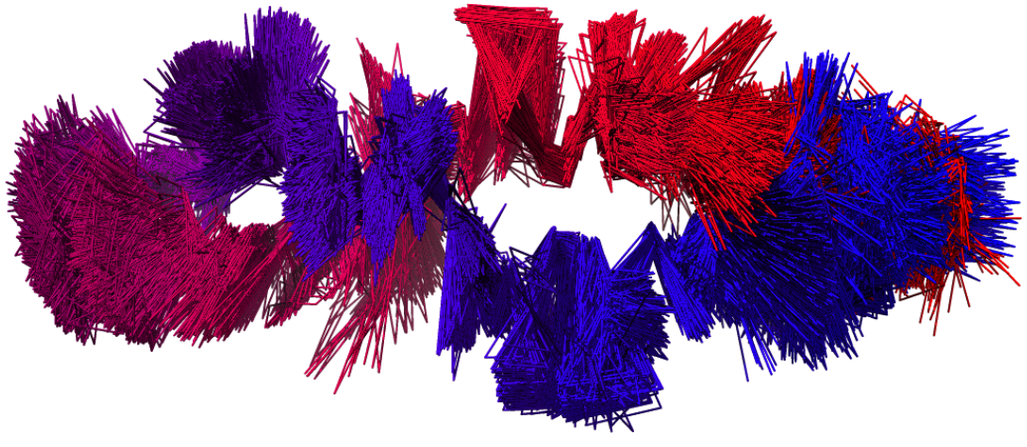


Alber, F., Dokudovskaya, S., Veenhoff, L. M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., et al. (2007). *Nature*, *450*(7170), 695–701

Genomes

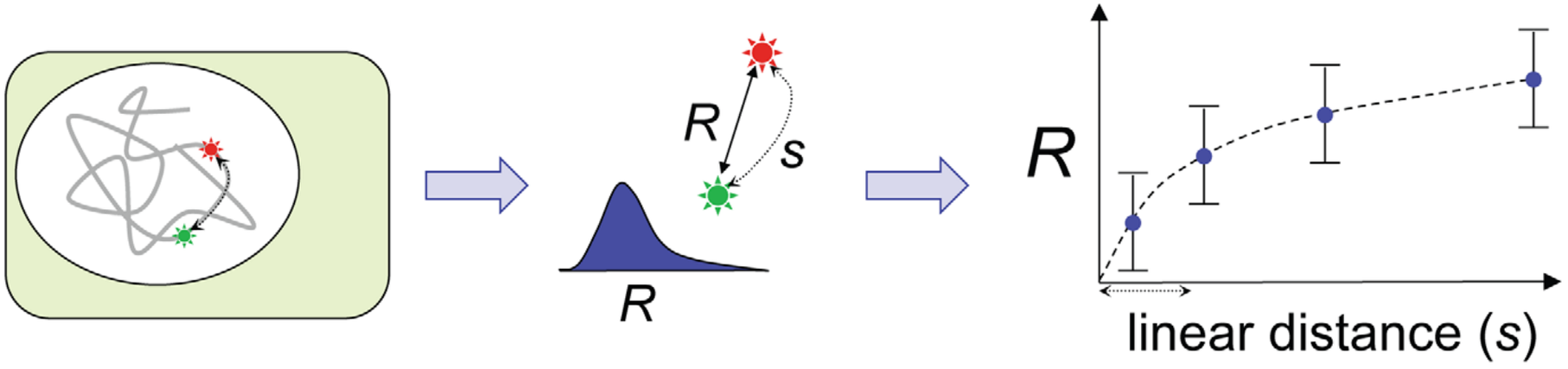
Limited data types



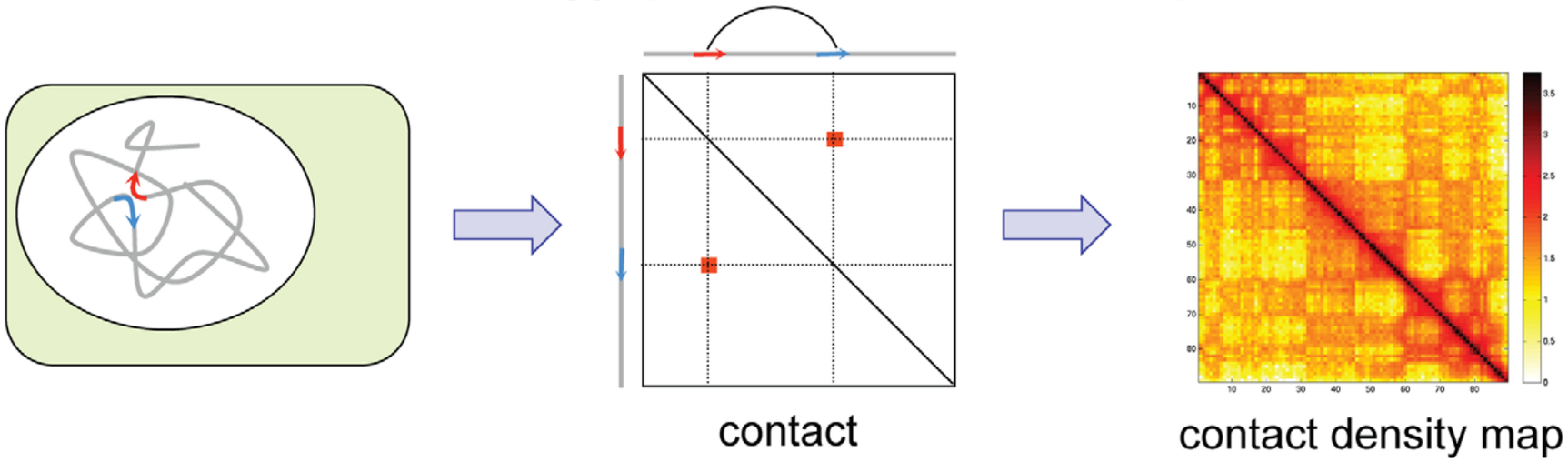


Main approaches

Light microscopy (FISH)

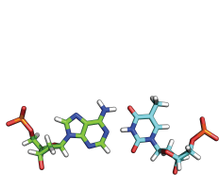


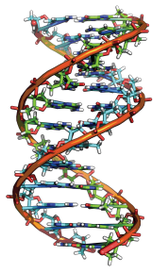
Cell/molecular biology (3C-based methods)

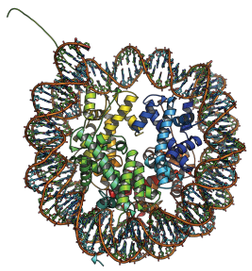


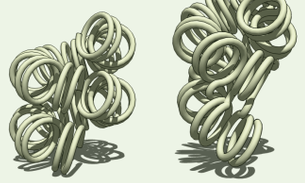
The resolution gap

Knowledge



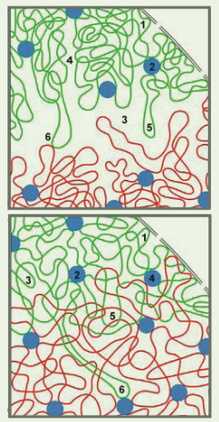


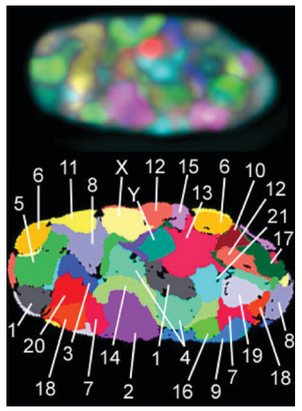




IDM

INM





DNA length

10^0 10^3 10^6 10^9 nt

Volume

10^{-9} 10^{-6} 10^{-3} 10^0 10^3 μm^3

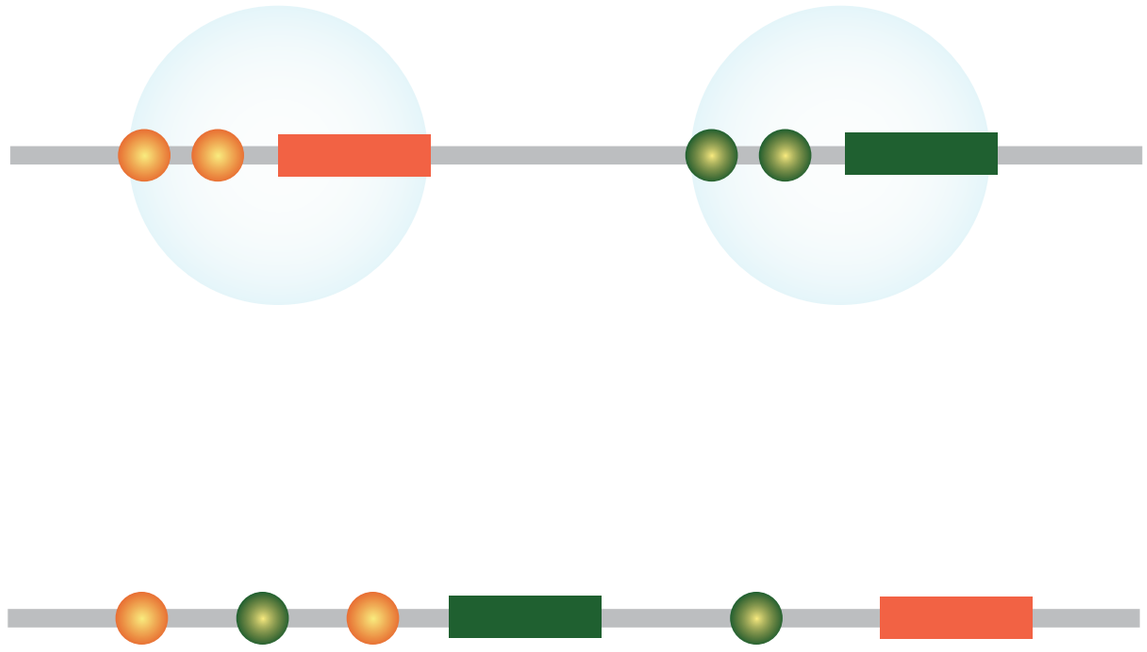
Time

10^{-10} 10^{-8} 10^{-6} 10^{-4} 10^{-2} 10^0 10^2 10^3 s

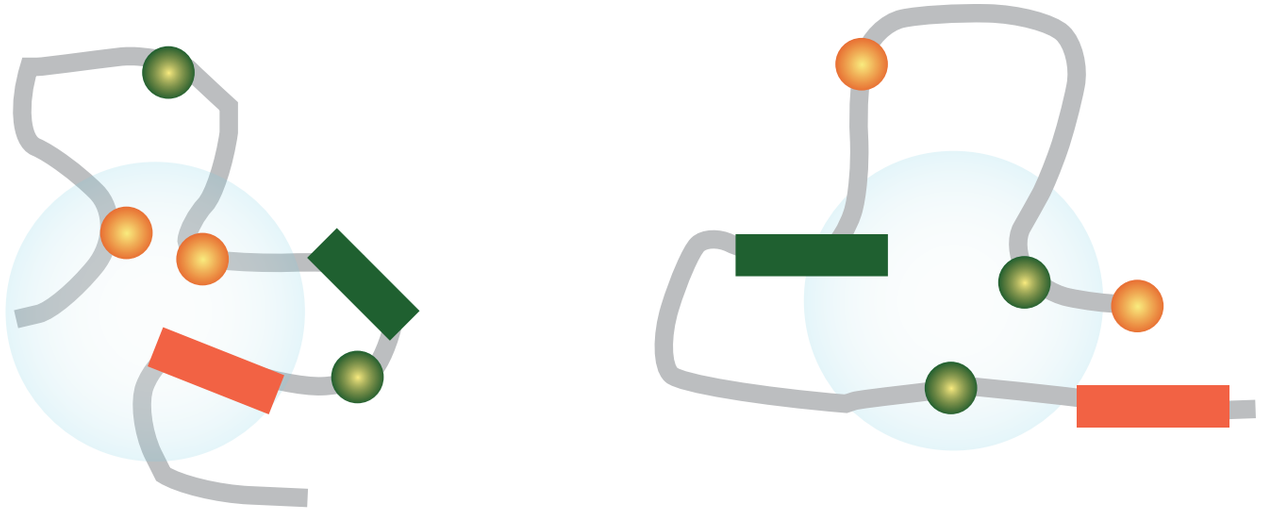
Resolution

10^{-3} 10^{-2} 10^{-1} μ

Simple genomes



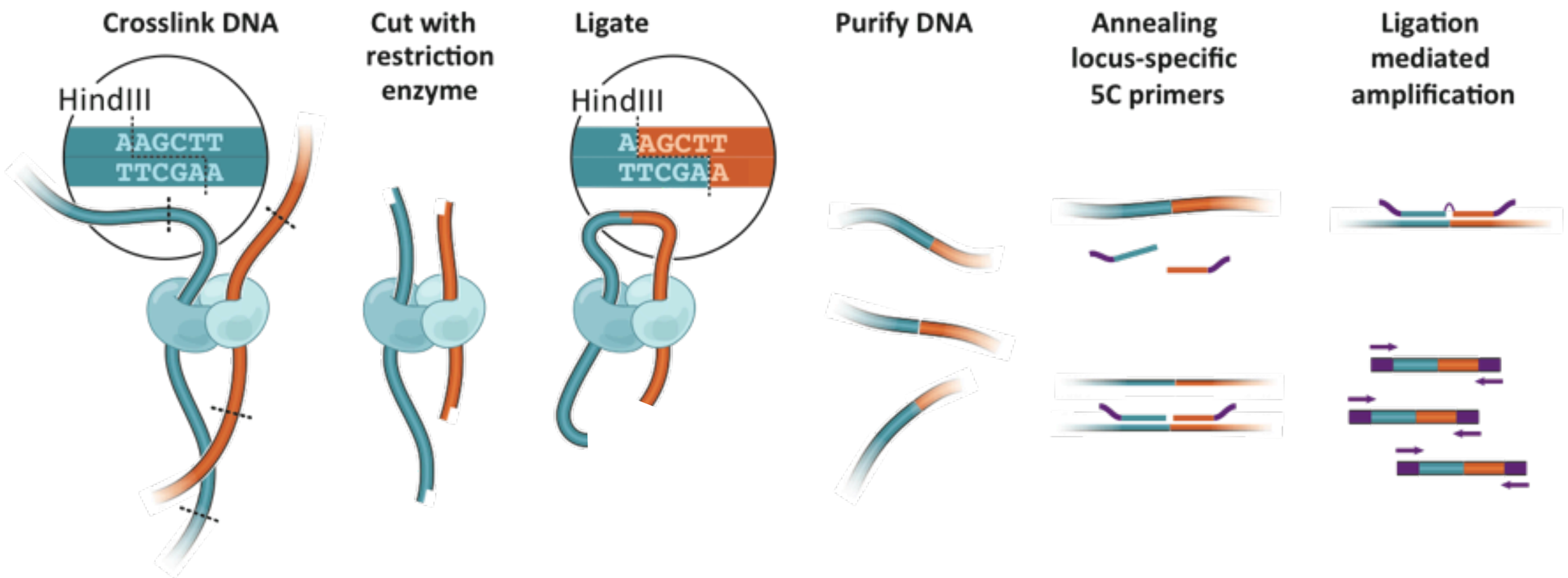
Complex genomes



5C technology

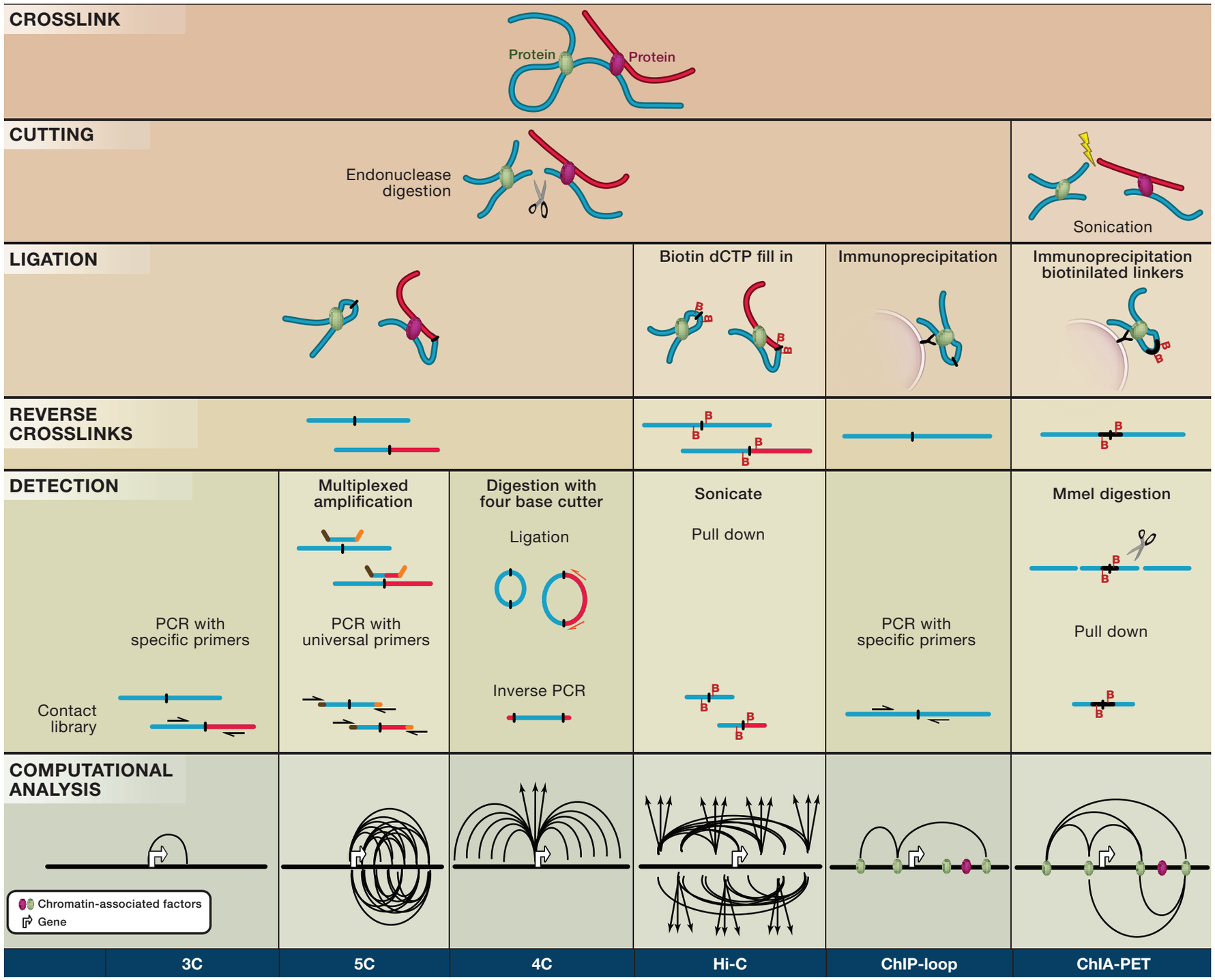
http://my5C.umassmed.edu

Job Dekker



Dostie et al. Genome Res (2006) vol. 16 (10) pp. 1299-309

3C-like technologies



Hakim and Misteli Cell (2012) vol. 148, March 2

3C-like technologies

	3C	5C	4C	Hi-C	ChIP-loop	ChIA-PET
Principle	Contacts between two defined regions[3,17]	All against all[4,18]	All contacts with a point of interest[14]	All against all[10]	Contacts between two defined regions associated with a given protein[8]	All contacts associated with a given protein[6]
Coverage	Commonly < 1Mb	Commonly < 1Mb	Genome-wide	Genome-wide	Commonly < 1Mb	Genome-wide
Detection	Locus-specific PCR	HT-sequencing	HT-sequencing	HT-sequencing	Locus-specific qPCR	HT-sequencing
Limitations	Low throughput and coverage	Limited coverage	Limited to one viewpoint		Rely on one chromatin-associated factor, disregarding other contacts	
Examples	Determine interaction between a known promoter and enhancer	Determine comprehensively higher-order chromosome structure in a defined region	All genes and genomic elements associated with a known LCR	All intra- and interchromosomal associations	Determine the role of specific transcription factors in the interaction between a known promoter and enhancer	Map chromatin interaction network of a known transcription factor
Derivatives	PCR with TaqMan probes[7] or melting curve analysis[1]		Circular chromosome conformation capture[20], open-ended chromosome conformation capture[19], inverse 3C[12], associated chromosome trap (ACT)[11], affinity enrichment of bait-ligated junctions[2]	Yeast[5,15], tethered conformation capture[9]		ChIA-PET combined 3C-ChIP-cloning (6C),[16] enhanced 4C (e4C)[13]

Take home message