Overview on structure determination

François Serra, David Castillo & Marc A. Marti-Renom
Structural Genomics Group (CNAG-CRG)

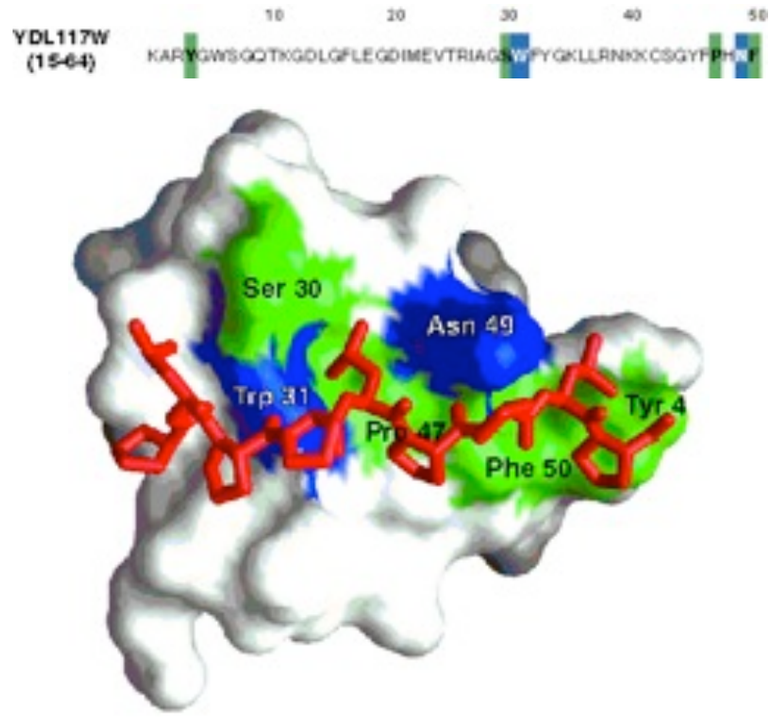
The importance of the 3D structure

The biochemical function of a molecule is defined by its interactions

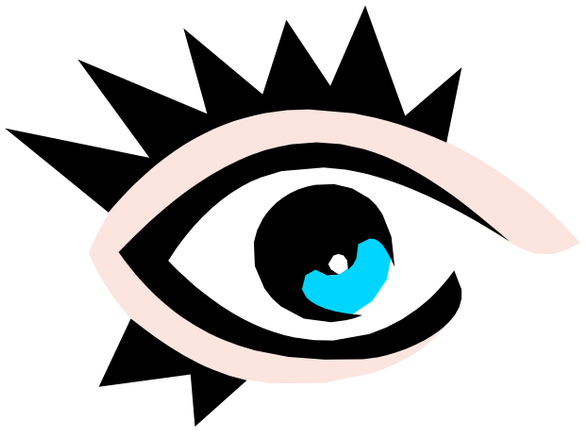
The biological function is in large part a consequence of these interactions

The 3D structure is more informative than sequence alone

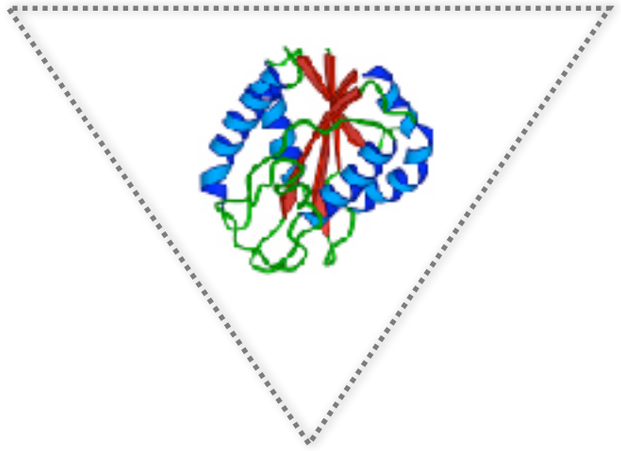
Evolution tends to conserve function and function depends more directly on structure than on sequence

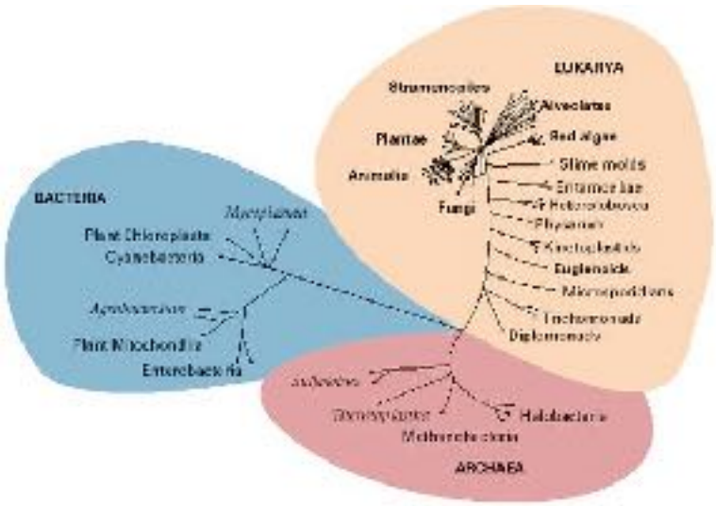


Data groups



Experimental observations





Statistical rules

Laws of physics

Structure prediction vs determination

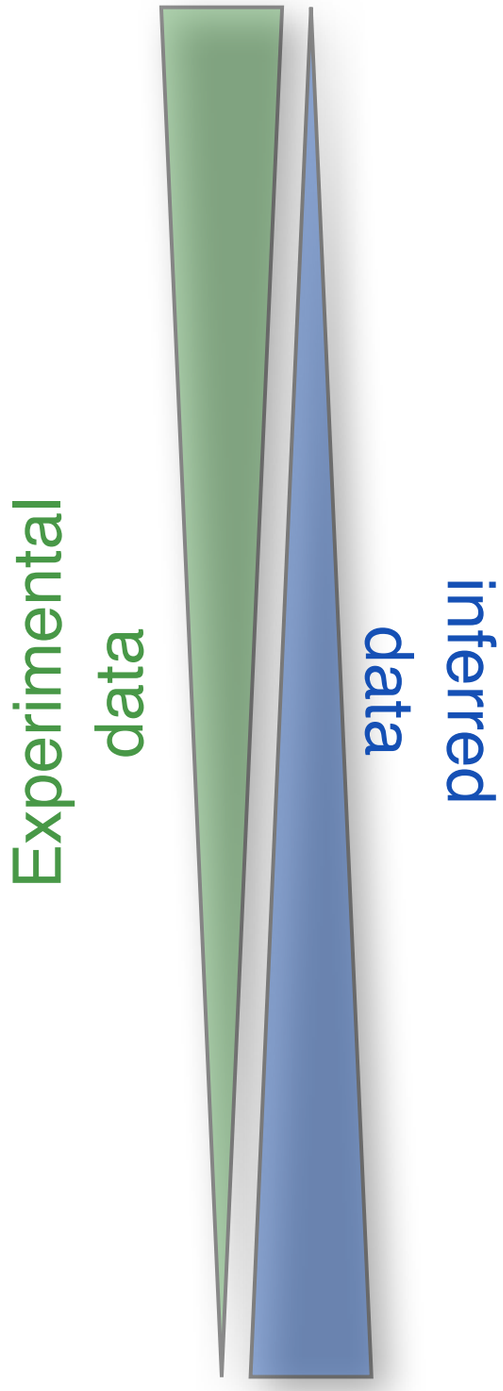
X-Ray

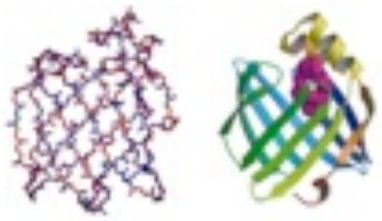
NMR

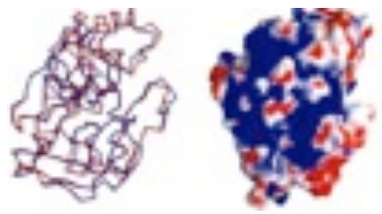
Comparative Modeling

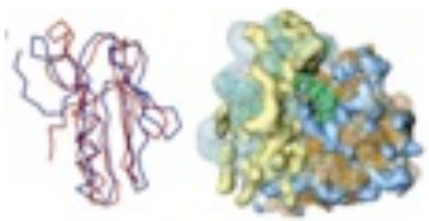
Threading

Ab-initio

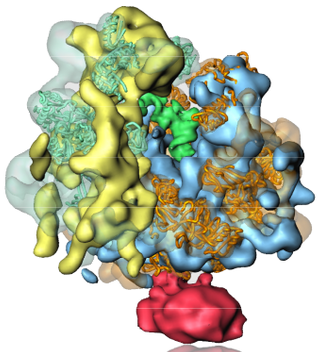


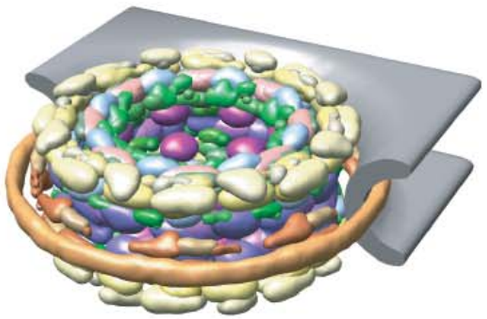


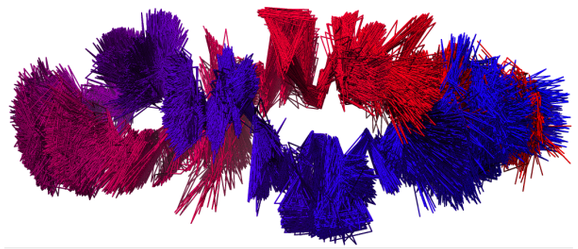


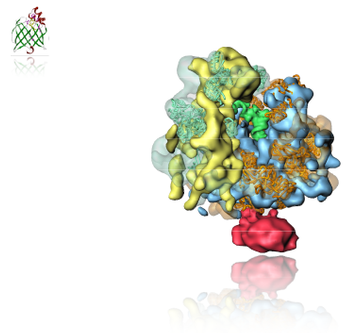


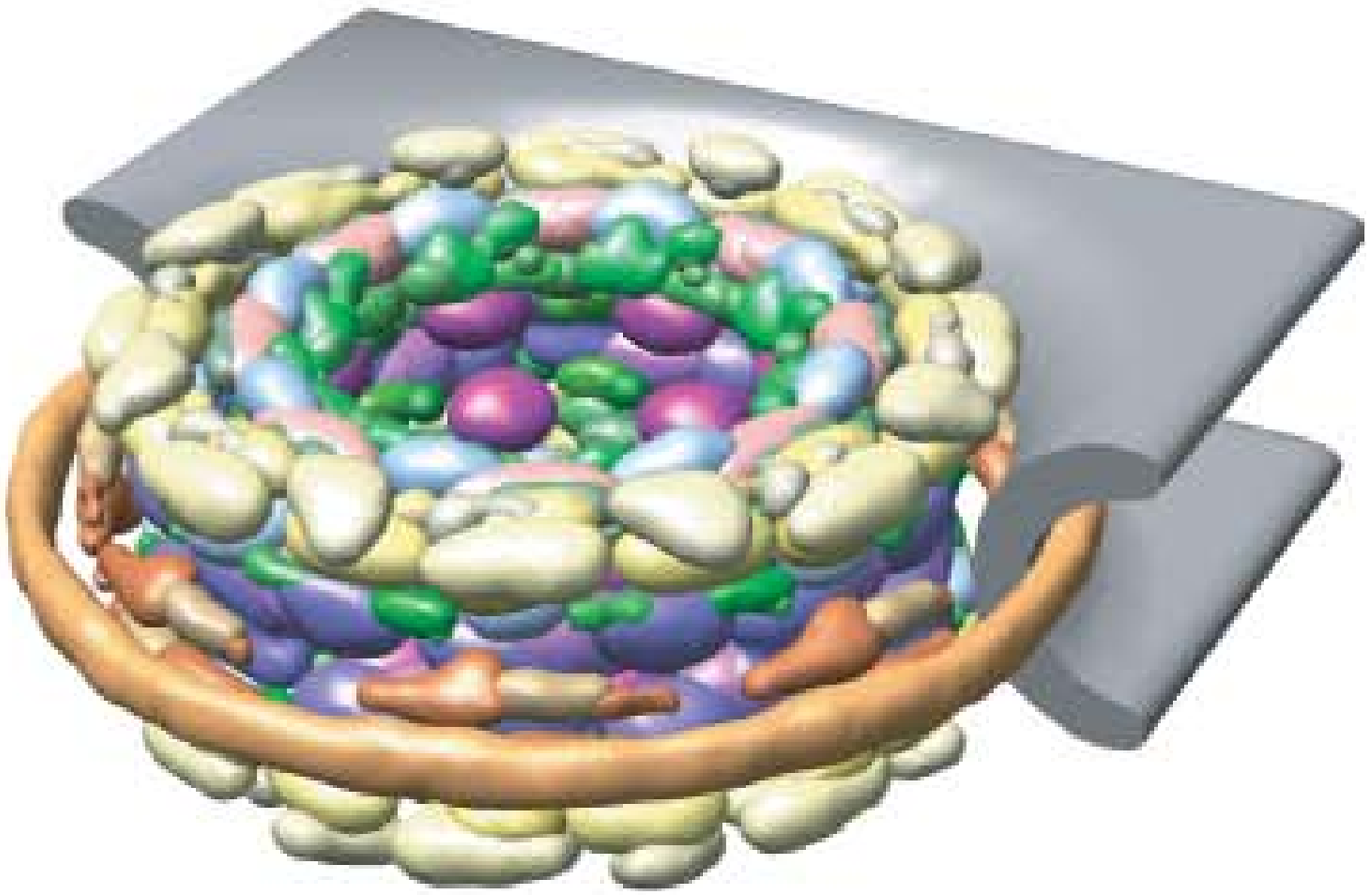
Data integration

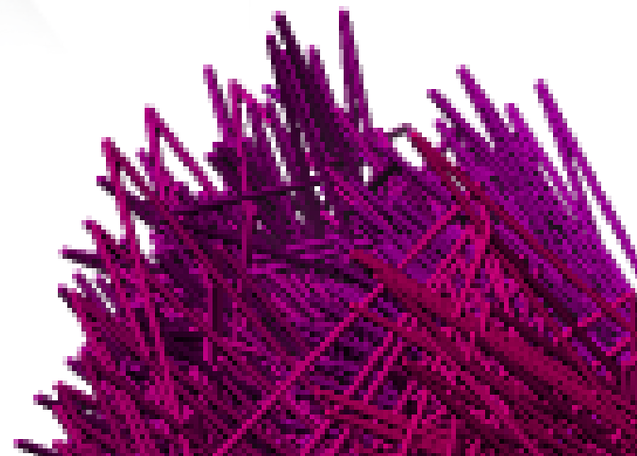












Advantages of integrative modeling

- **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**

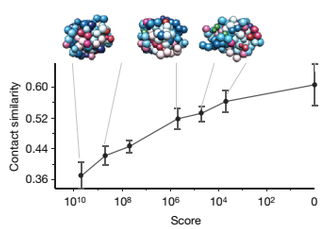
Russel, D., Lasker, K., Webb, B., Velázquez-Muriel, J., Tjioe, E., Schneidman-Duhovny, D., Peterson, B., et al. (2012). *PLoS Biology*, *10*(1), e1001244

The four stages of integrative modeling

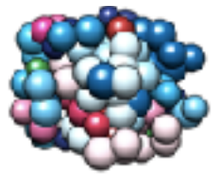
Data collection

Representation and scoring

Optimization

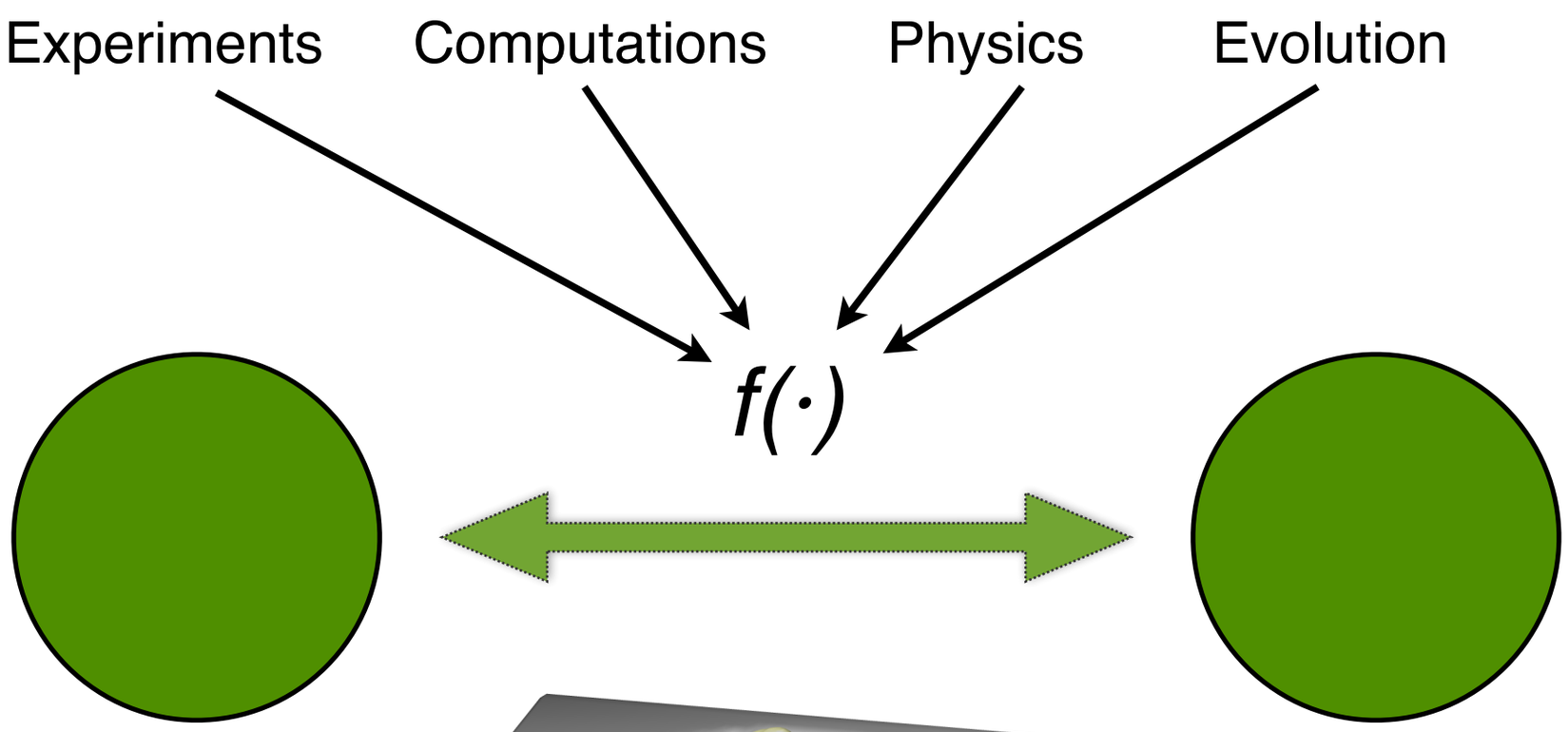


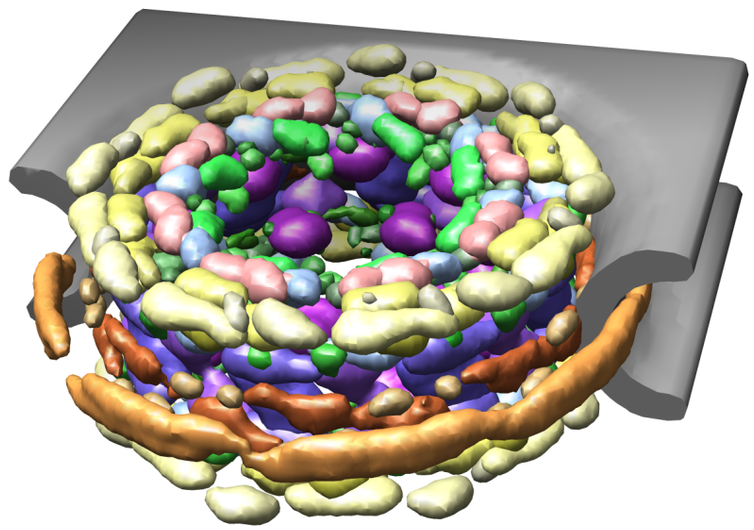
Analysis of the results



Integrative Modeling Platform

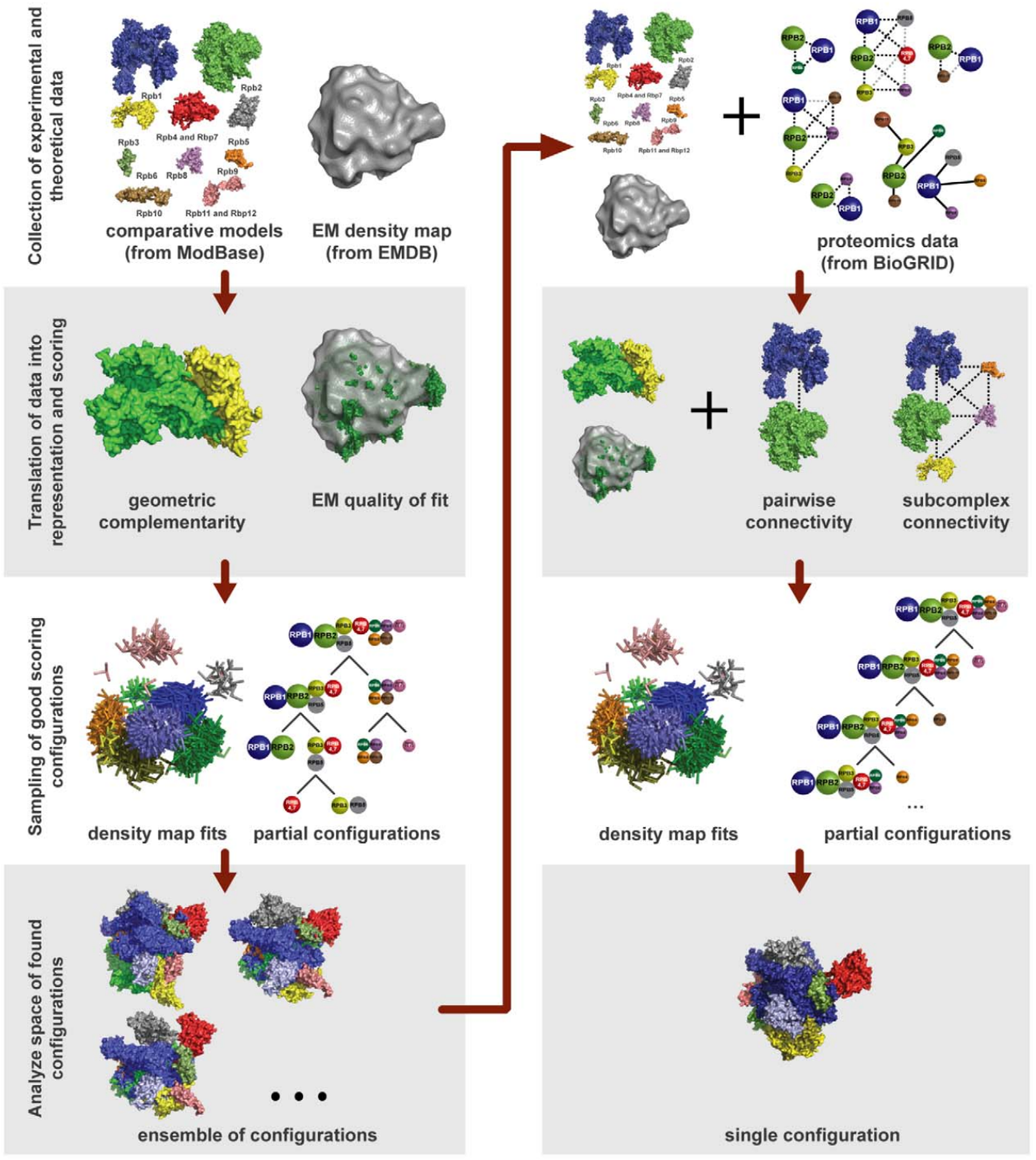
http://www.integrativemodeling.org





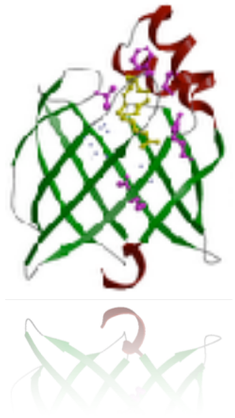
From: Russel, D. et al. PLOS Biology 10, e1001244 (2012).

“Toy” example...

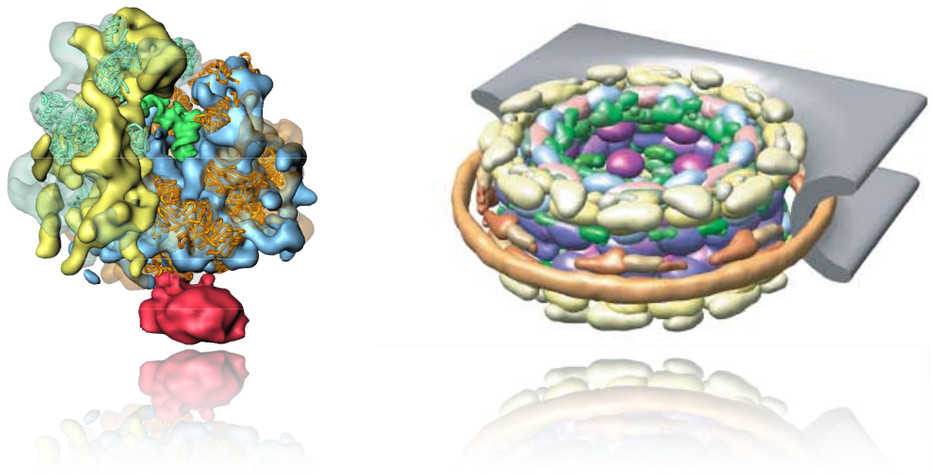


Russel, D., Lasker, K., Webb, B., Velázquez-Muriel, J., Tjioe, E., Schneidman-Duhovny, D., Peterson, B., et al. (2012). *PLoS Biology*, *10*(1), e1001244

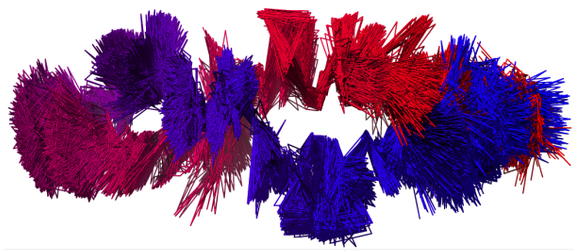
“Real” examples



PROTEINS



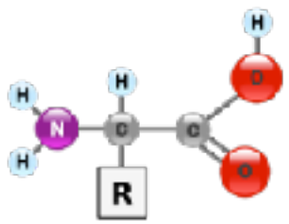
COMPLEXES



GENOMES

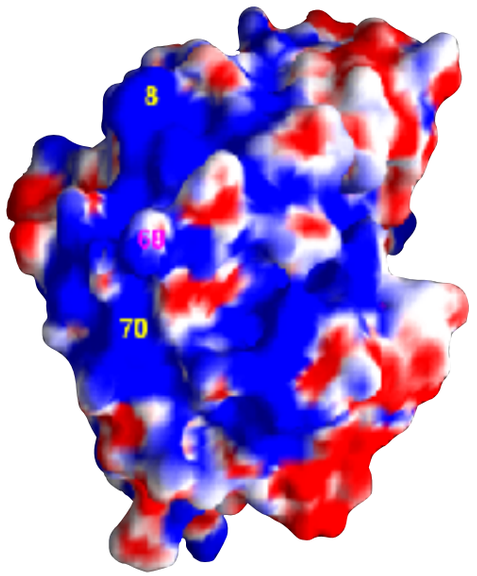
Proteins

Single *data type*



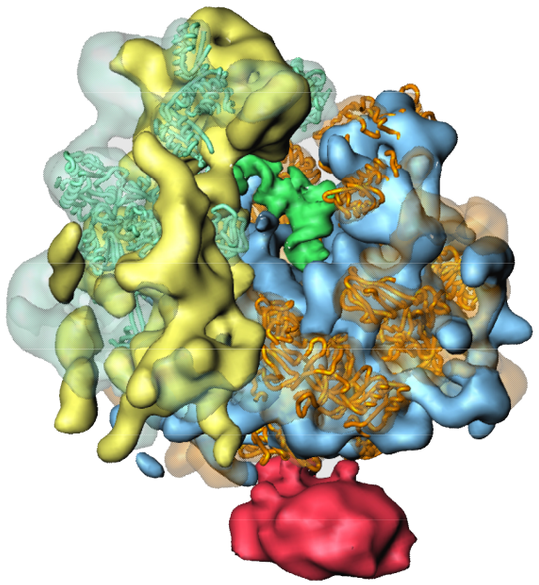
Amino Acids

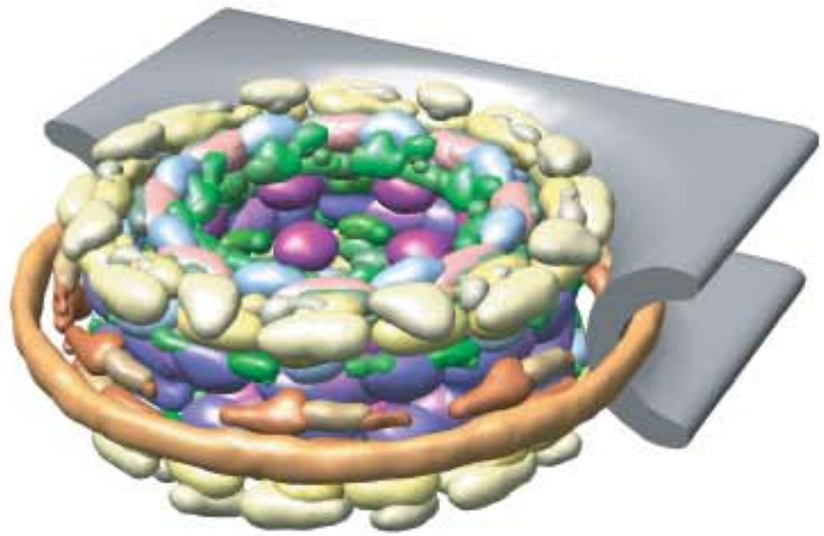
X-Ray;
NMR;
Modeling



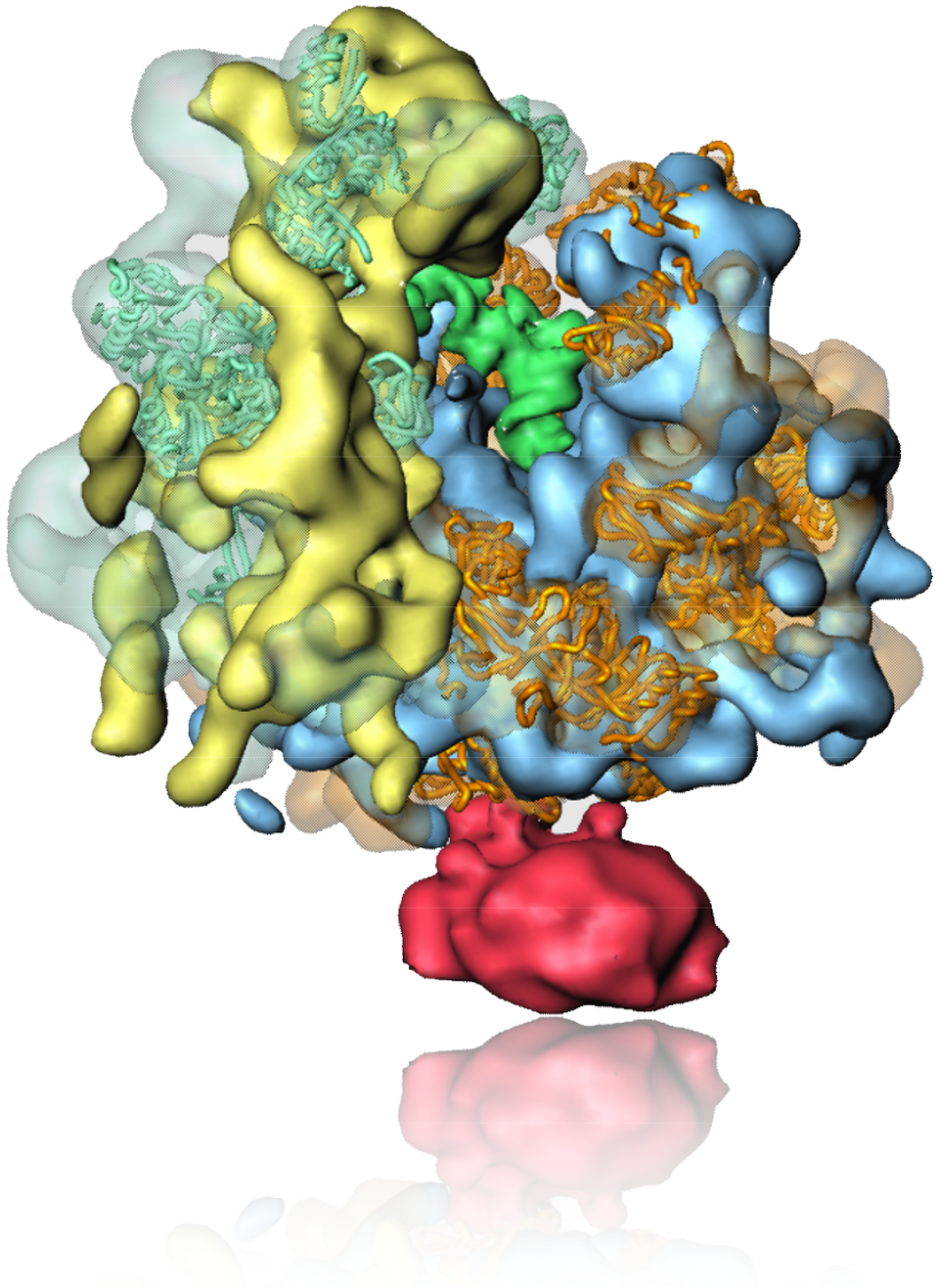
Complexes

Multiple data types





S. cerevisiae ribosome



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

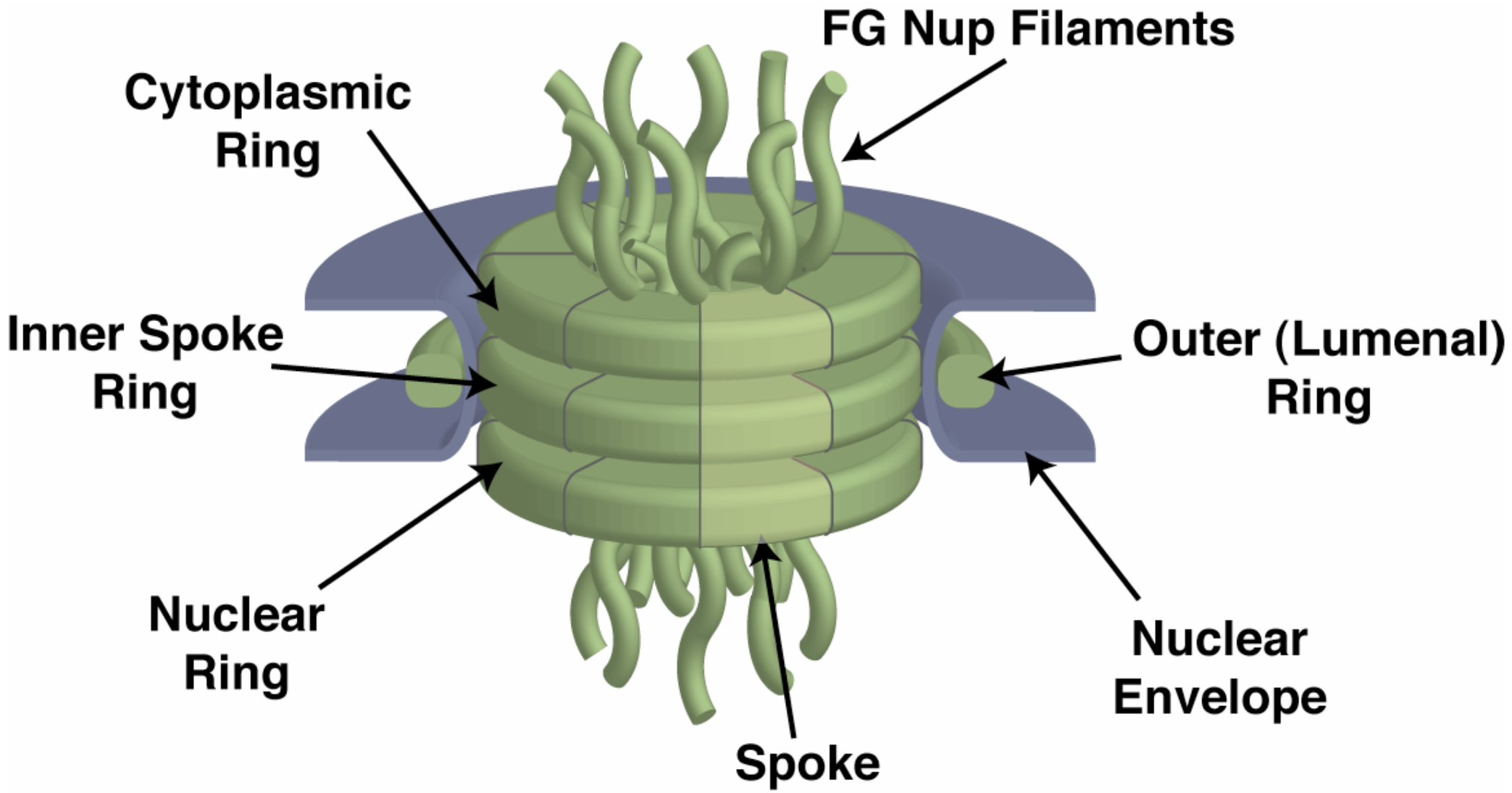
43 proteins could be modeled on 20-56% seq.id. to a known structure.

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.

cnag CRG

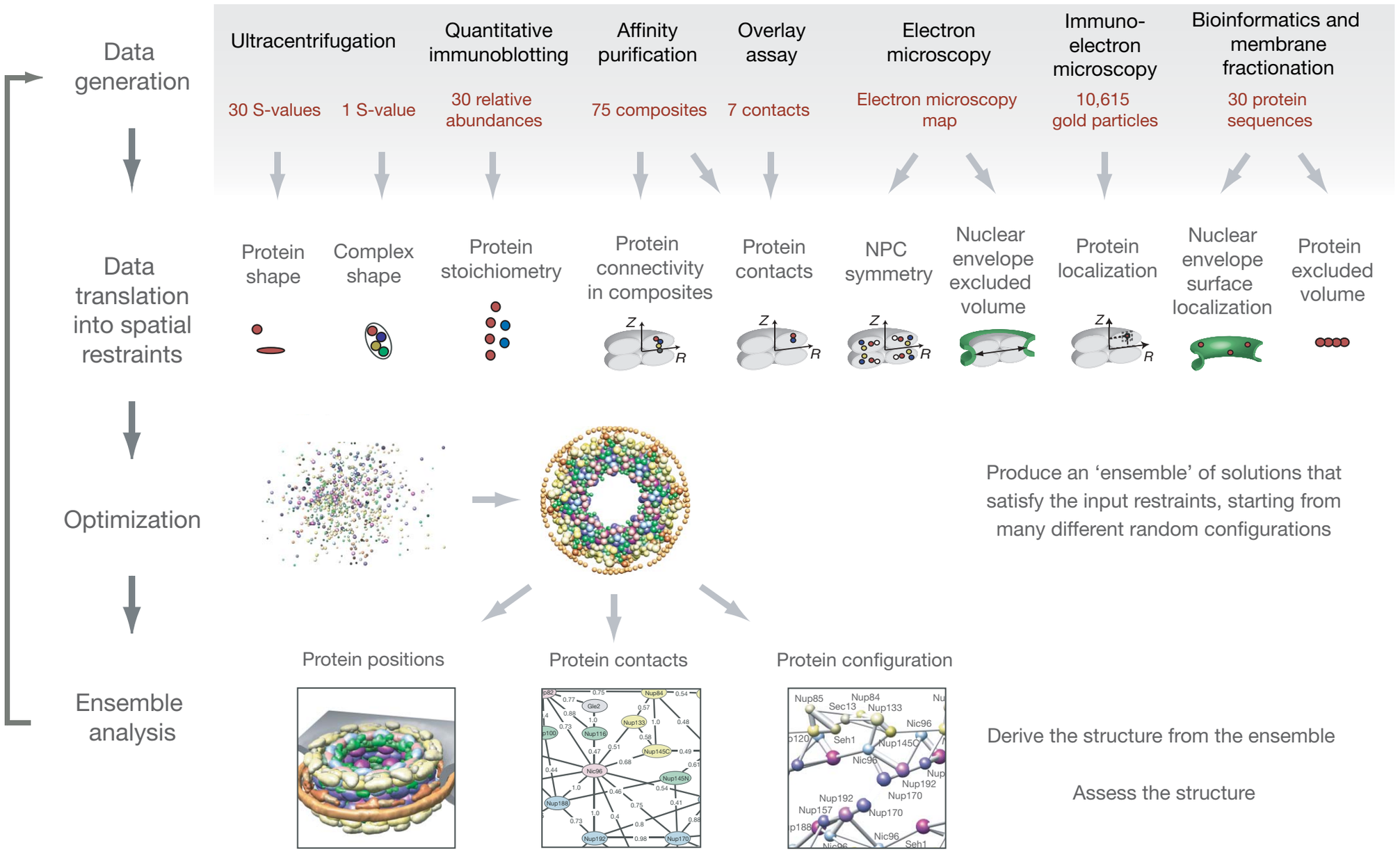
The nuclear pore complex

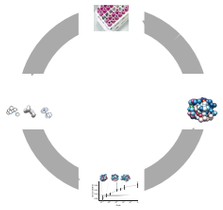


Alber, F., Dokudovskaya, S., Veenhoff, L. M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., et al. (2007). *Nature*, *450*(7170), 695–701

Integrative Modeling of the NPC

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





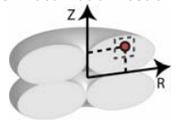
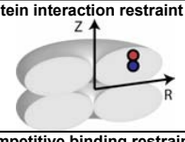
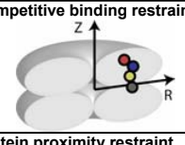
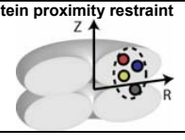
Representation

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

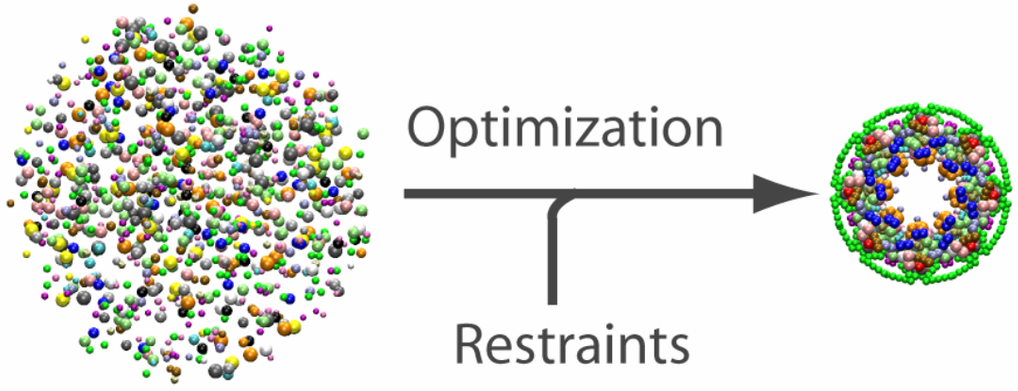
Alber, F., Dokudovskaya, S., Veenhoff, L. M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., et al. (2007). *Nature*, *450*(7170), 695–701

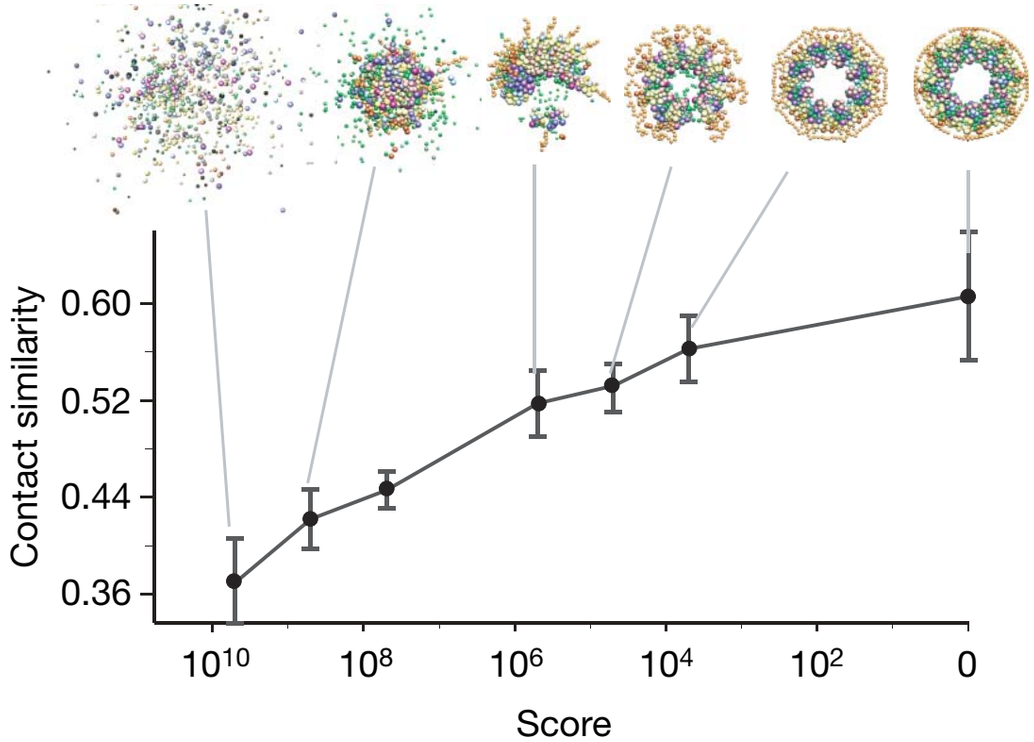
Scoring

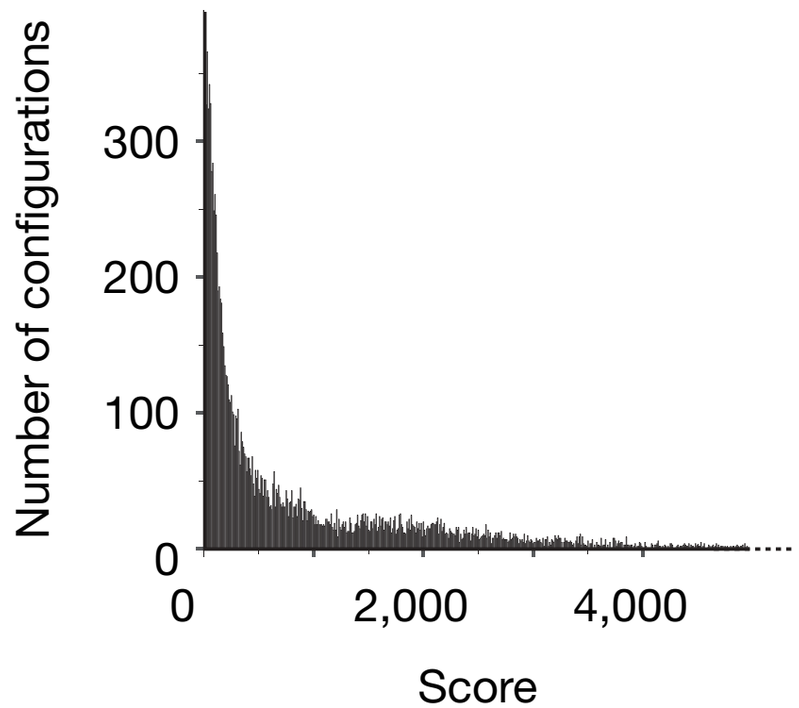
Data generation		Data interpretation				
Method	**Experiments**	**Restraint**	R_C	R_O	R_A	**Functional form of activated feature restraint**
Bioinformatics and Membrane fractionation	**30 nup sequences**	**Protein excluded volume restraint**	-	-	**1,864 *1,863/2**	**Protein-protein**: Violated for $f < f_o$. 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 $D=19.2-R$, 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

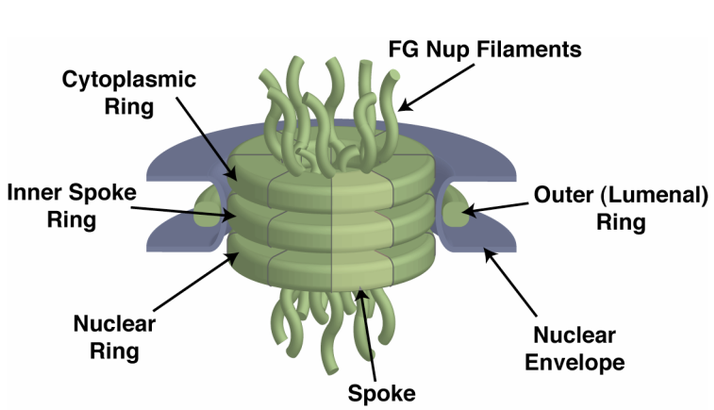


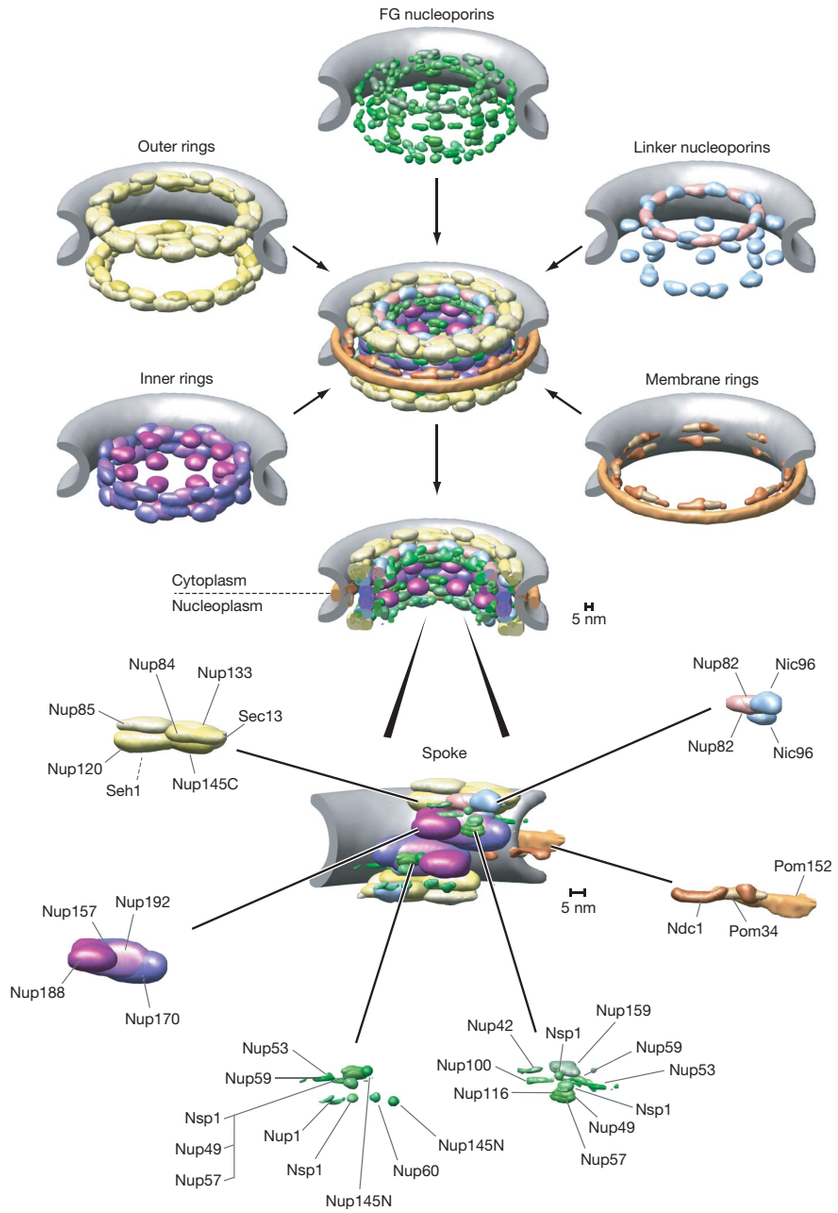




Alber, F., Dokudovskaya, S., Veenhoff, L. M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., et al. (2007). *Nature*, *450*(7170), 695–701

The structure of the nuclear pore complex

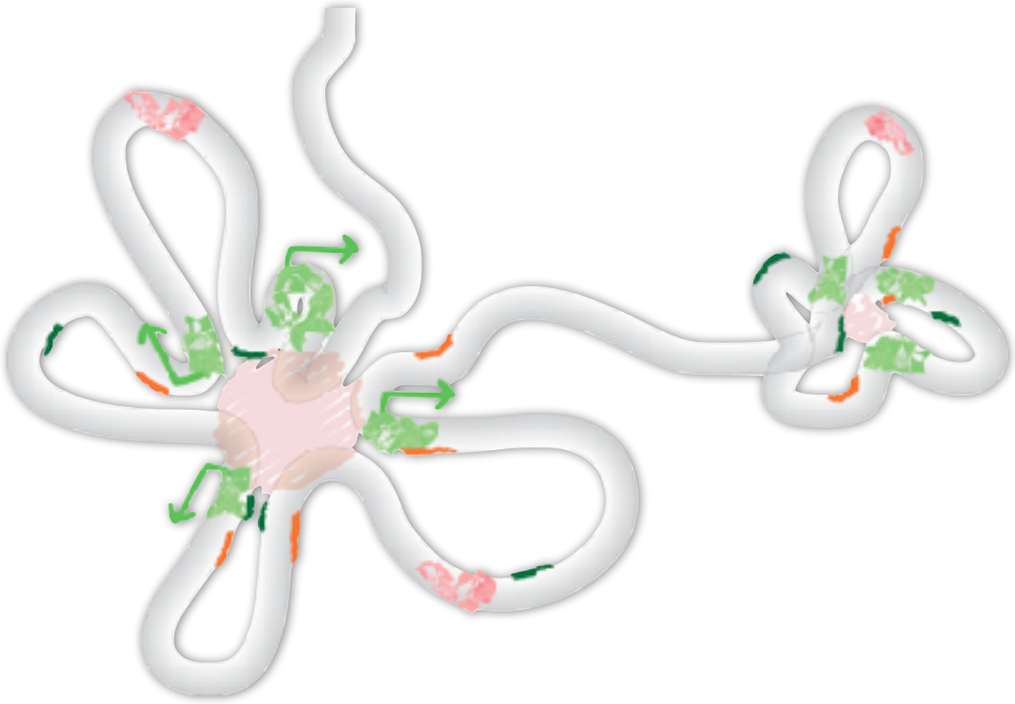


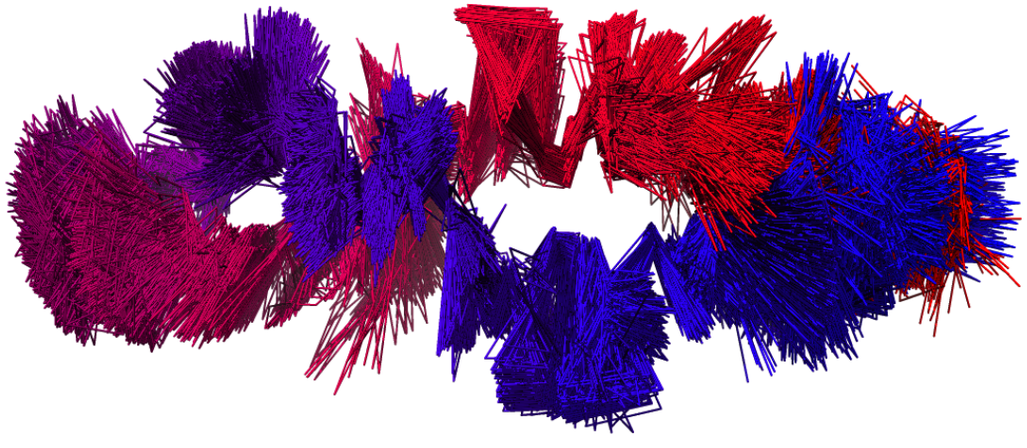


Alber, F., Dokudovskaya, S., Veenhoff, L. M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., et al. (2007). *Nature*, *450*(7170), 695–701

Genomes

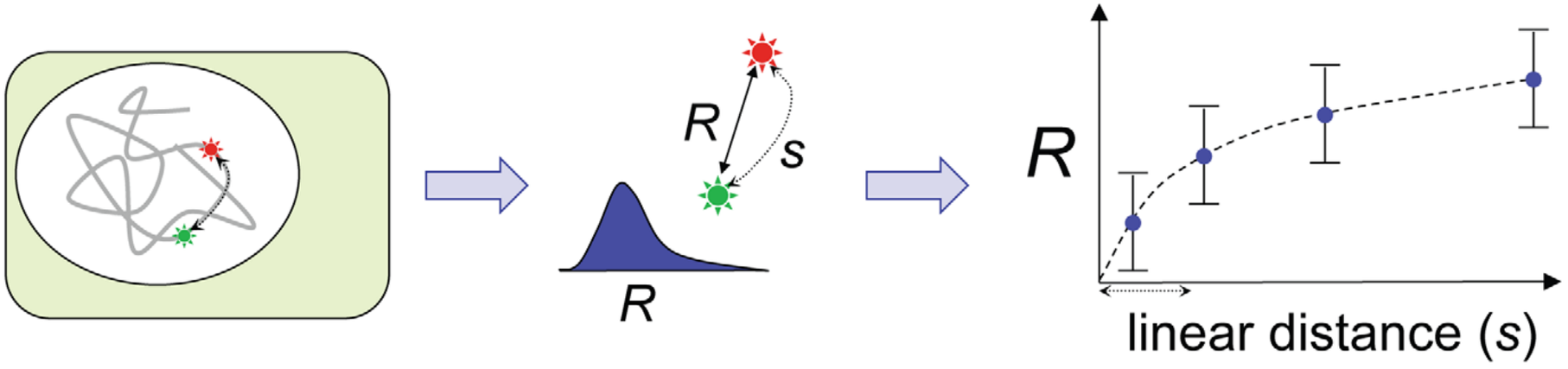
Limited *data types*



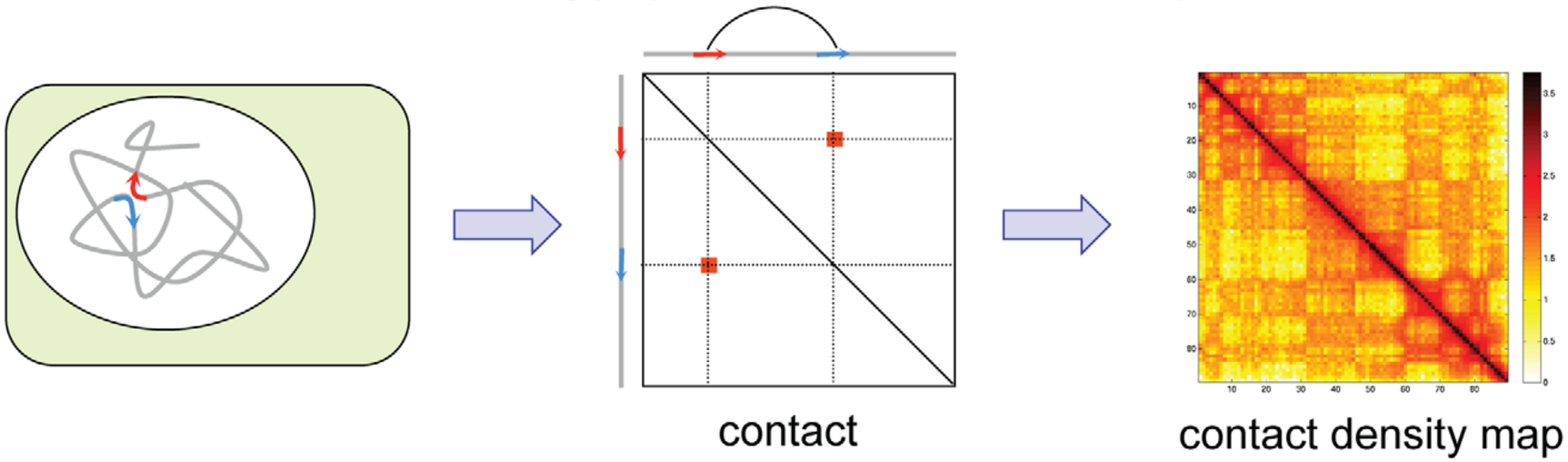


Main approaches

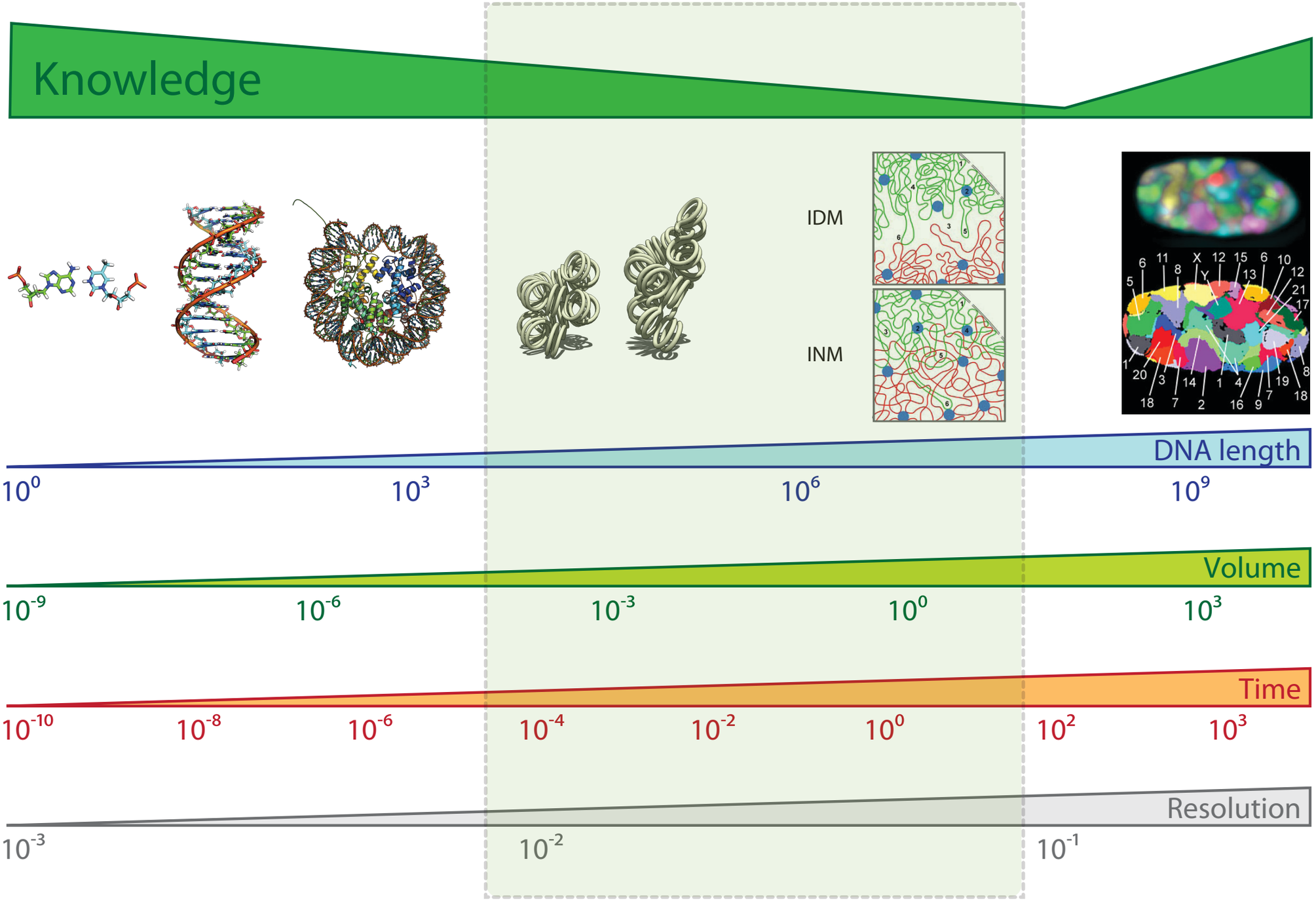
Light microscopy (FISH)



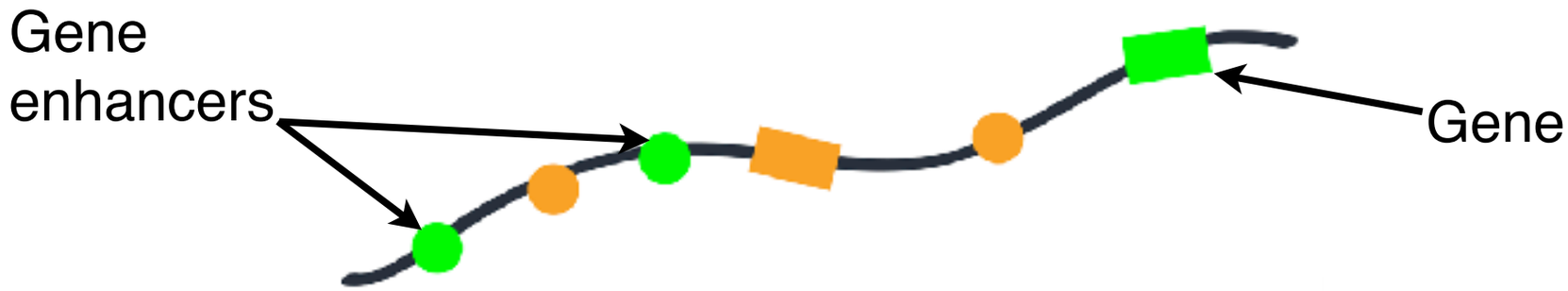
Cell/molecular biology (3C-based methods)

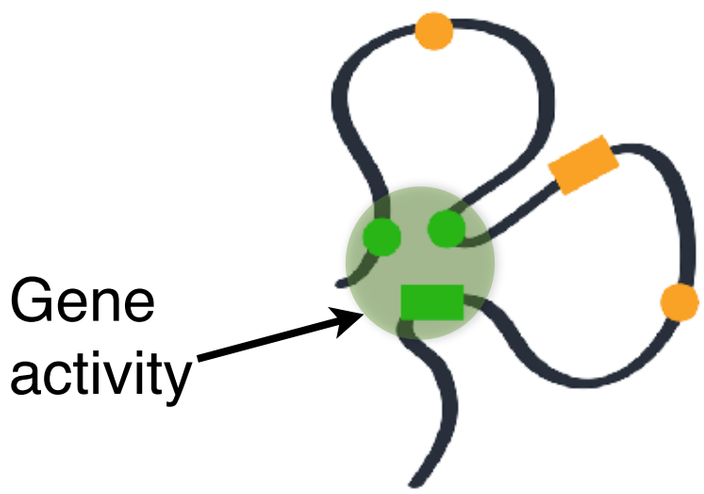


The resolution gap



The genome is not linearly organized

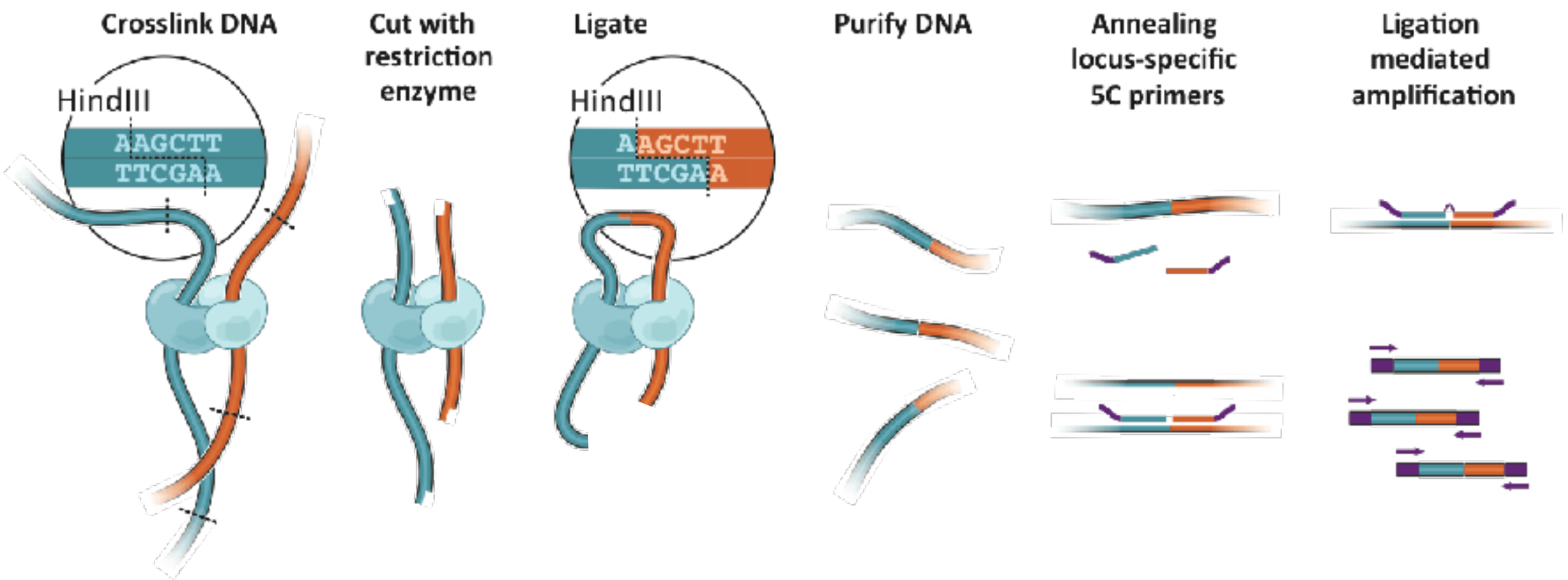




5C technology

http://my5C.umassmed.edu

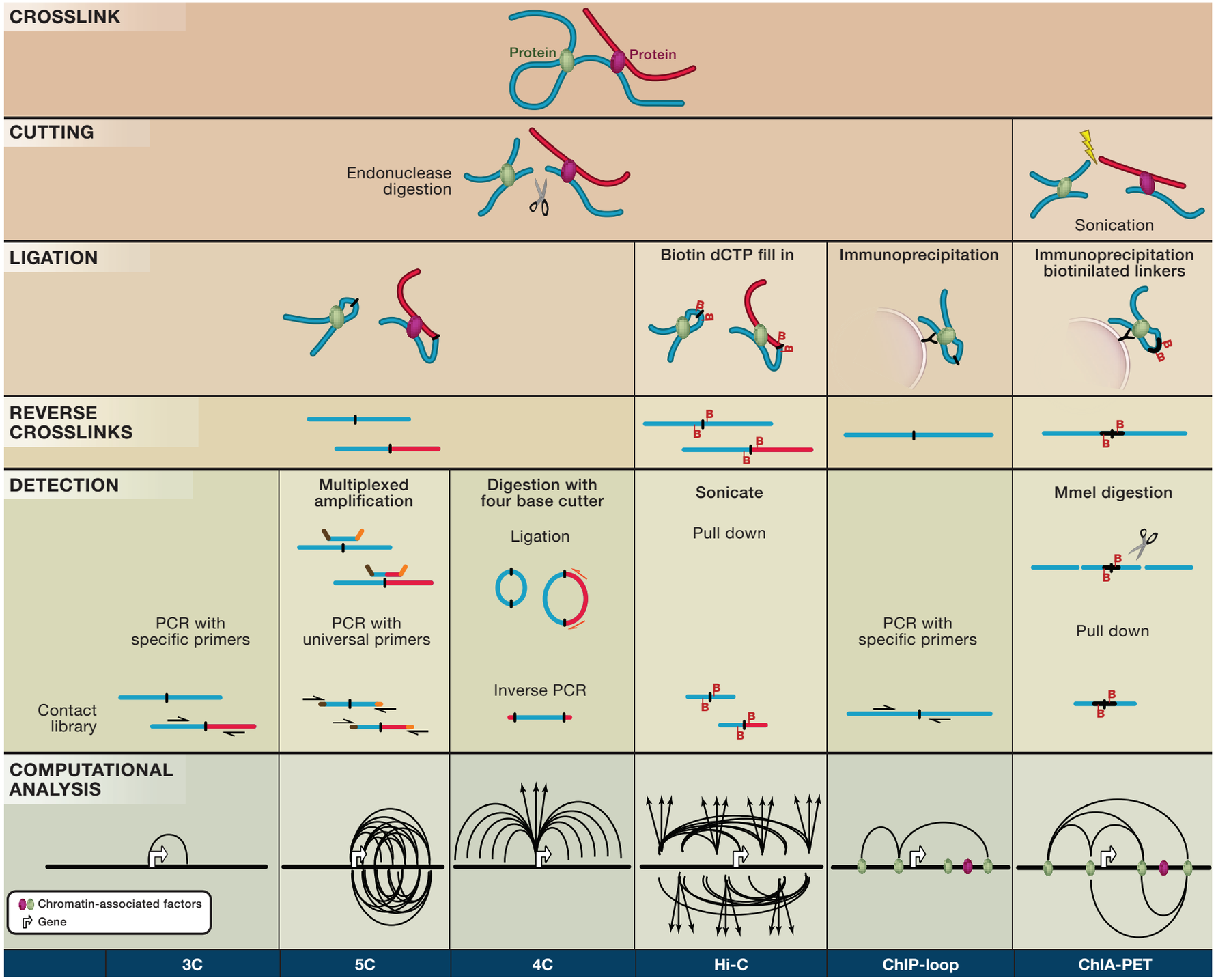
Job Dekker

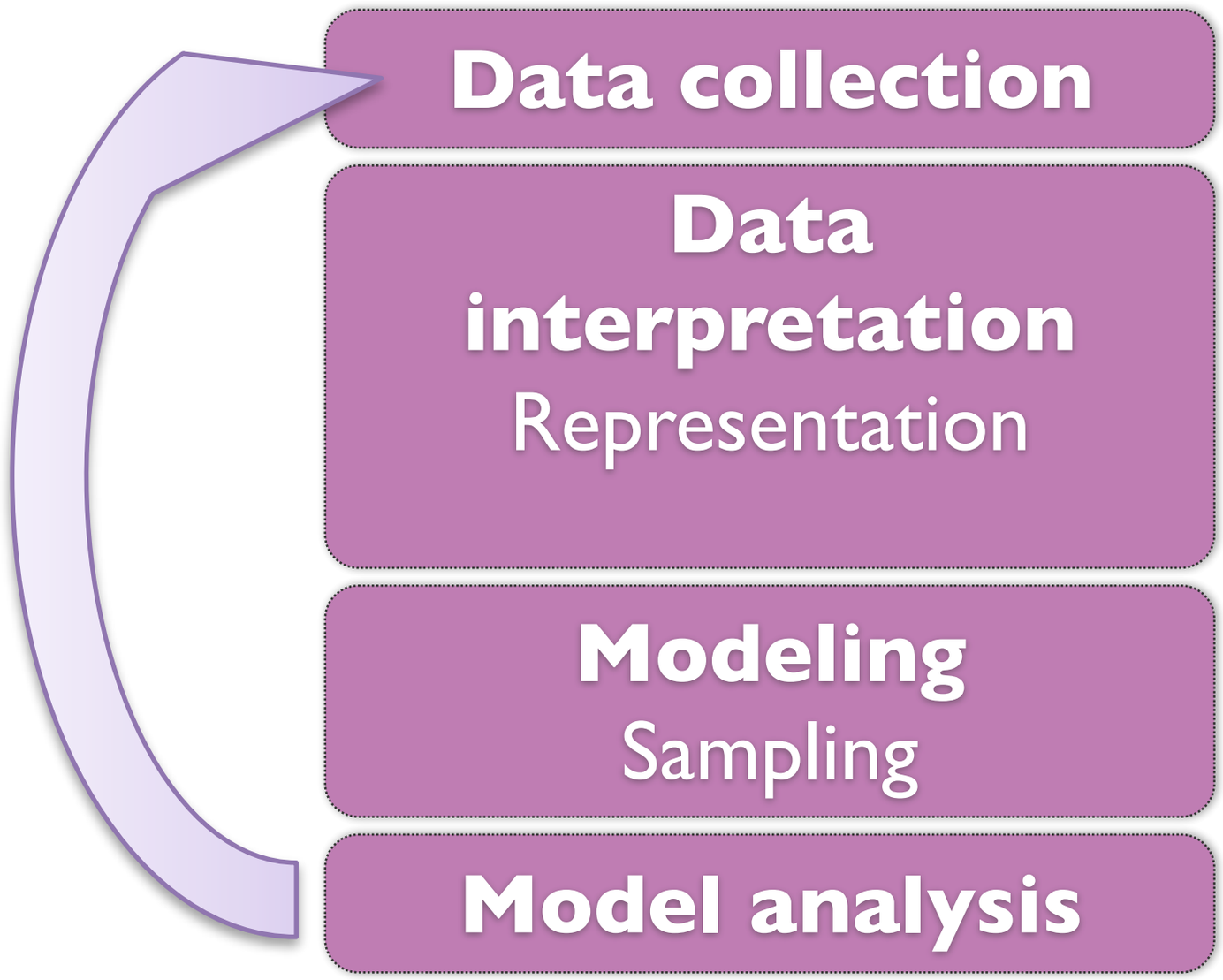


Dostie et al. Genome Res (2006) vol. 16 (10) pp. 1299-309

cnag CRG

3C-like technologies



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

Take home message