Hybrid [integrative] **methods for structure determination**

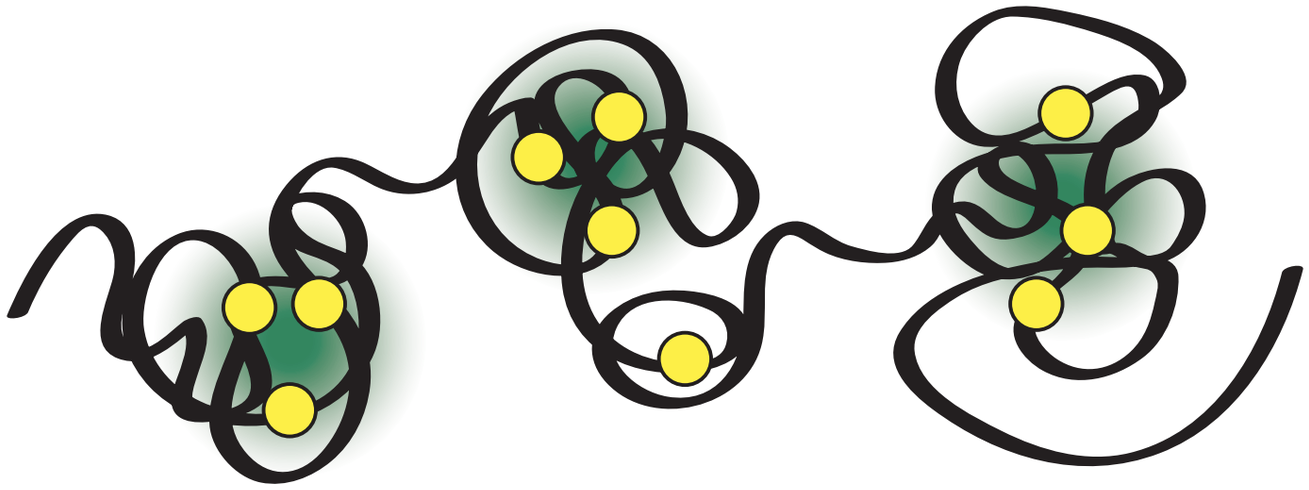
Marc A. Marti-Renom
Genome Biology Group (CNAG)
Structural Genomics Group (CRG)

*iCrea
INSTITUCIÓ CATALANA DE RECERCA I ESTUDIS AVANÇATS

cnag CRG Centre for Genomic Regulation

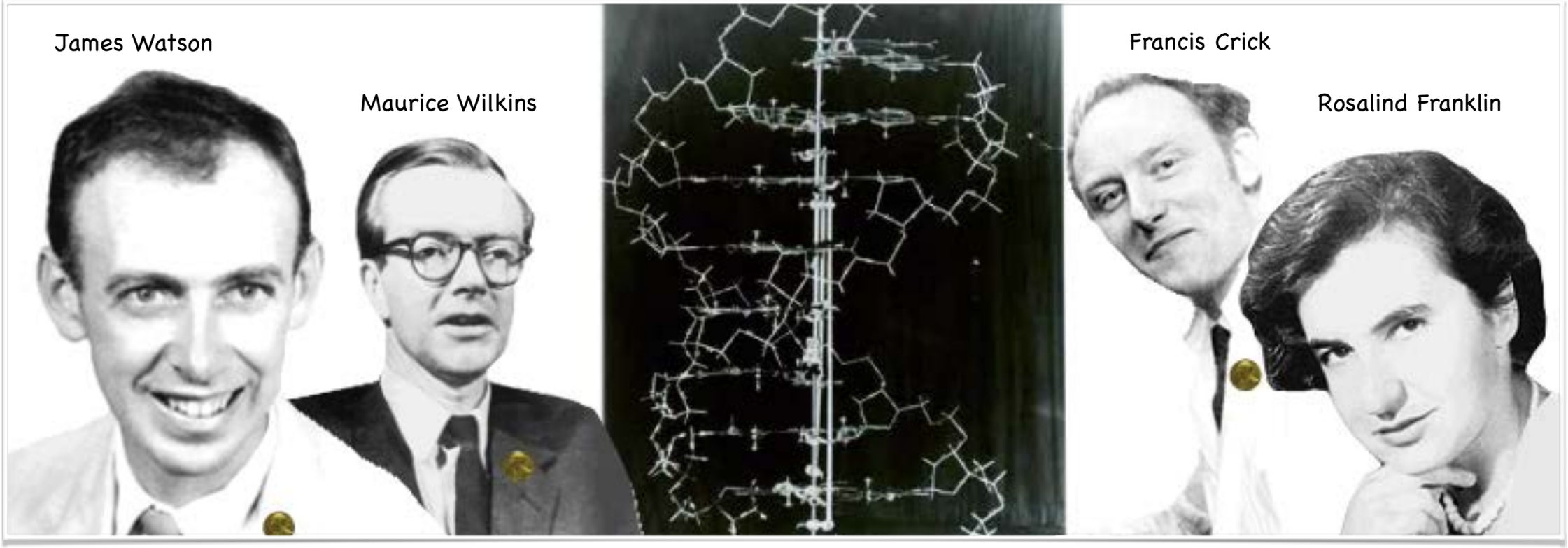
IMP

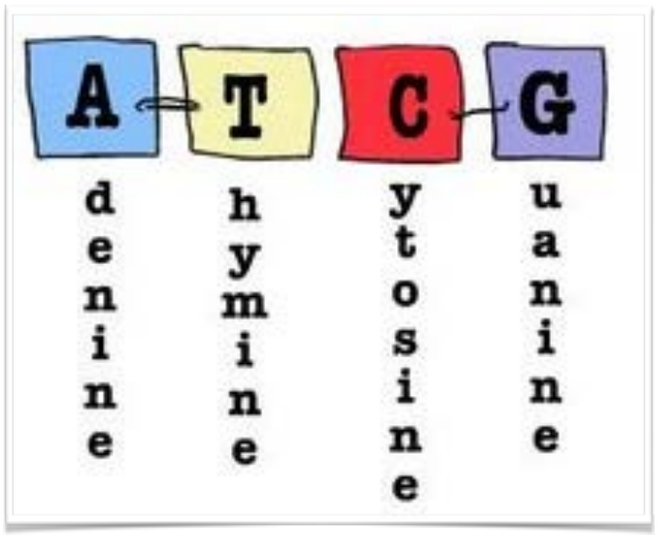
TADbit

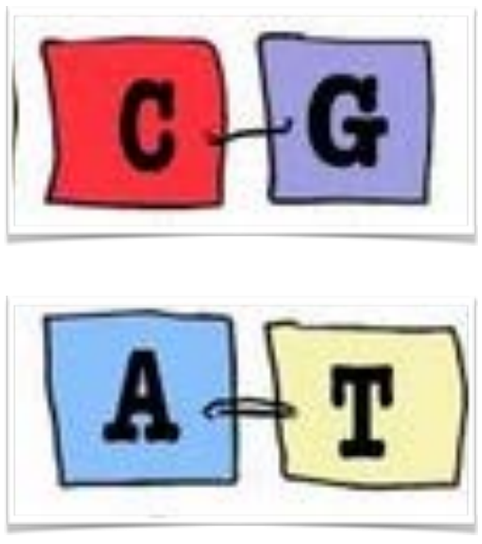


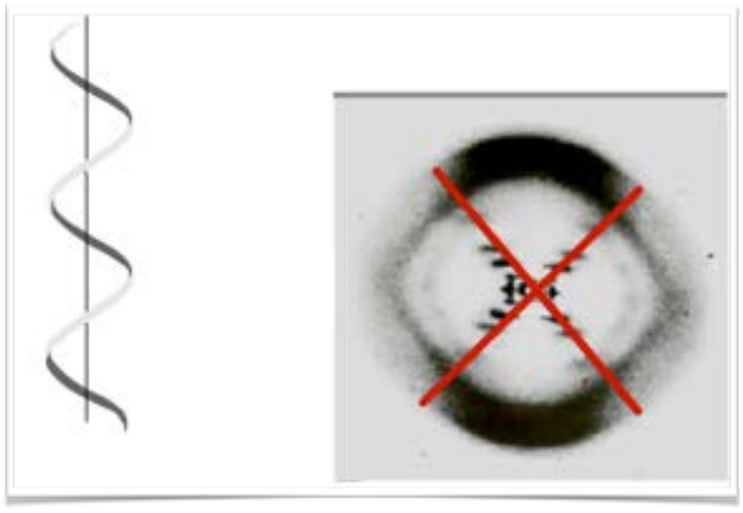
Since 1956

Integrative Modeling

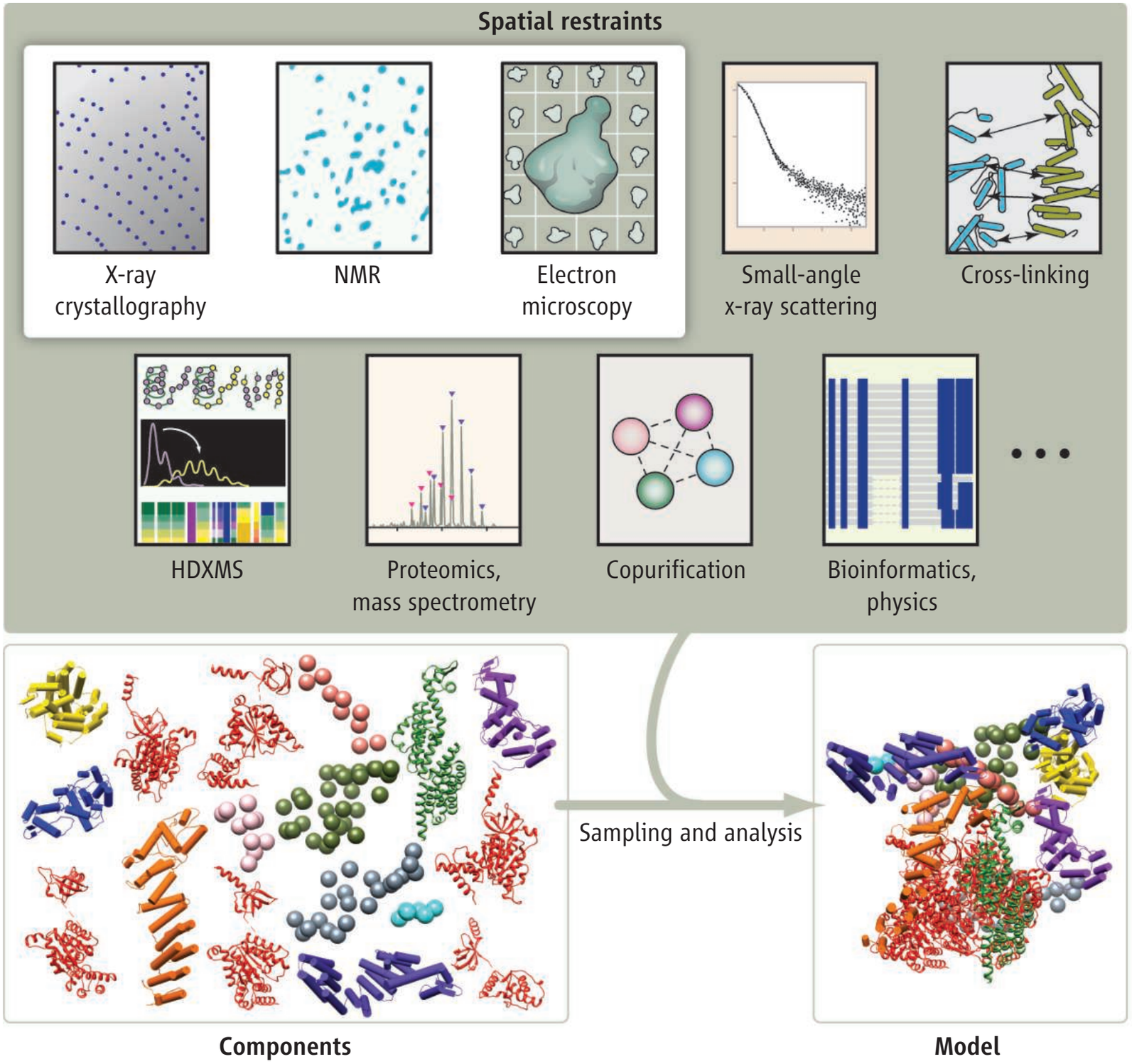






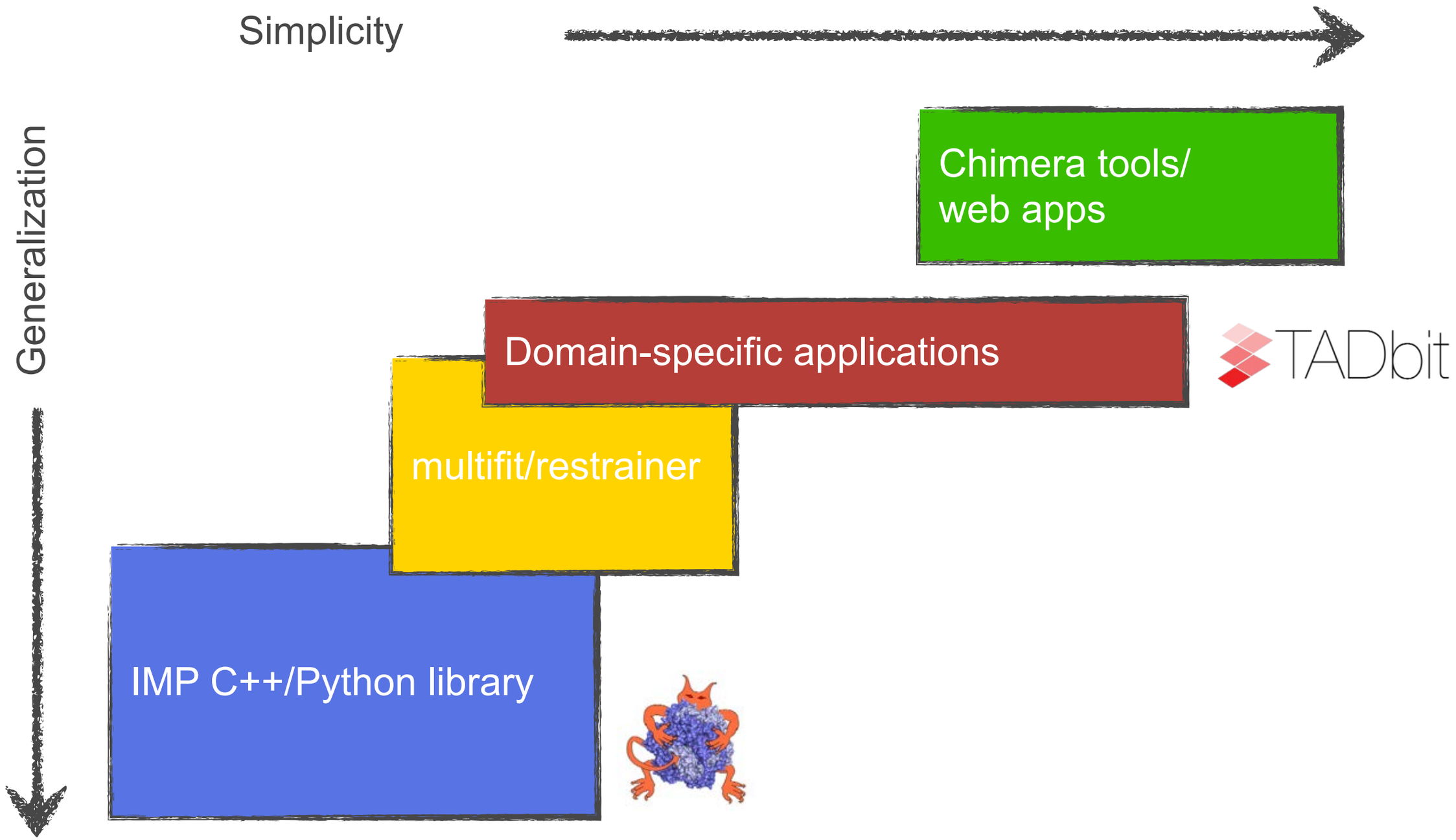


Build geometric models



The Integrative Modeling Platform (IMP)

http://www.integrativemodeling.org



Russel et al, PLoS Biology, 2012

The Integrative Modeling Platform (IMP)

http://www.integrativemodeling.org

Stage 1: Gathering information.

Stage 2: Choosing how to represent and evaluate models.

Stage 3: Finding models that score well.

Stage 4: Analyzing resulting models and information.

IMP

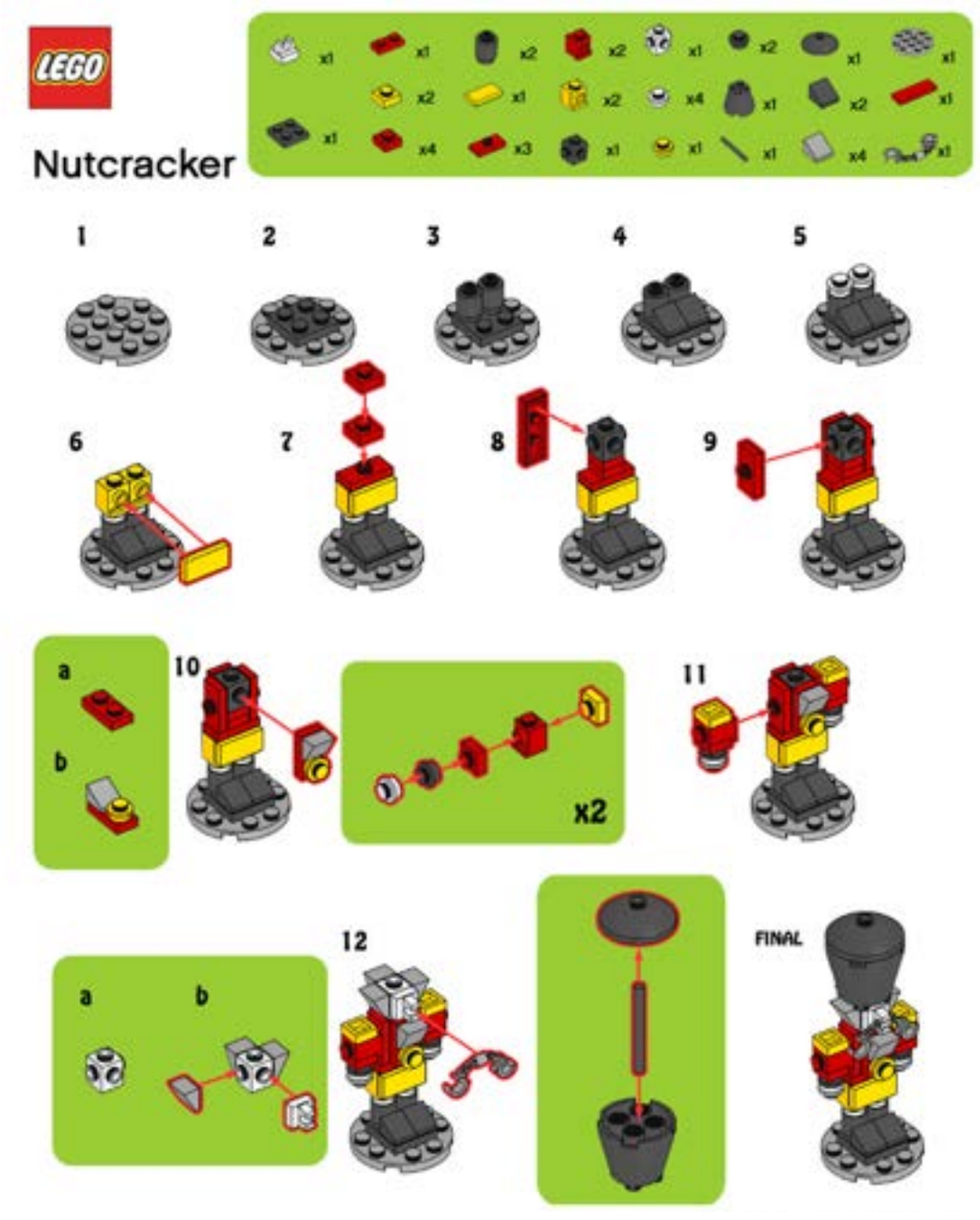
Russel et al, PLoS Biology, 2012

UCSF
University of California
San Francisco

Representation

- Atomic
- Rigid bodies
- Coarse grained
- Multi-scale
- Symmetry/periodicity
- Multi-state systems

Scoring



- Proteomics
- Density maps
- EM images
- FRET
- Chemical cross linking
- Homology-derived restraints
- SAXS
- Native mass spec
- Statistical potentials
- Molecular mechanics forcefields
- Bayesian scoring functions
- Library of functional forms (ambiguity, …)

UCSF
University of California
San Francisco

Sampling

- Monte-Carlo
- Conjugate Gradients
- Quasi-Newton
- Simplex
- Divide and conquer sampler

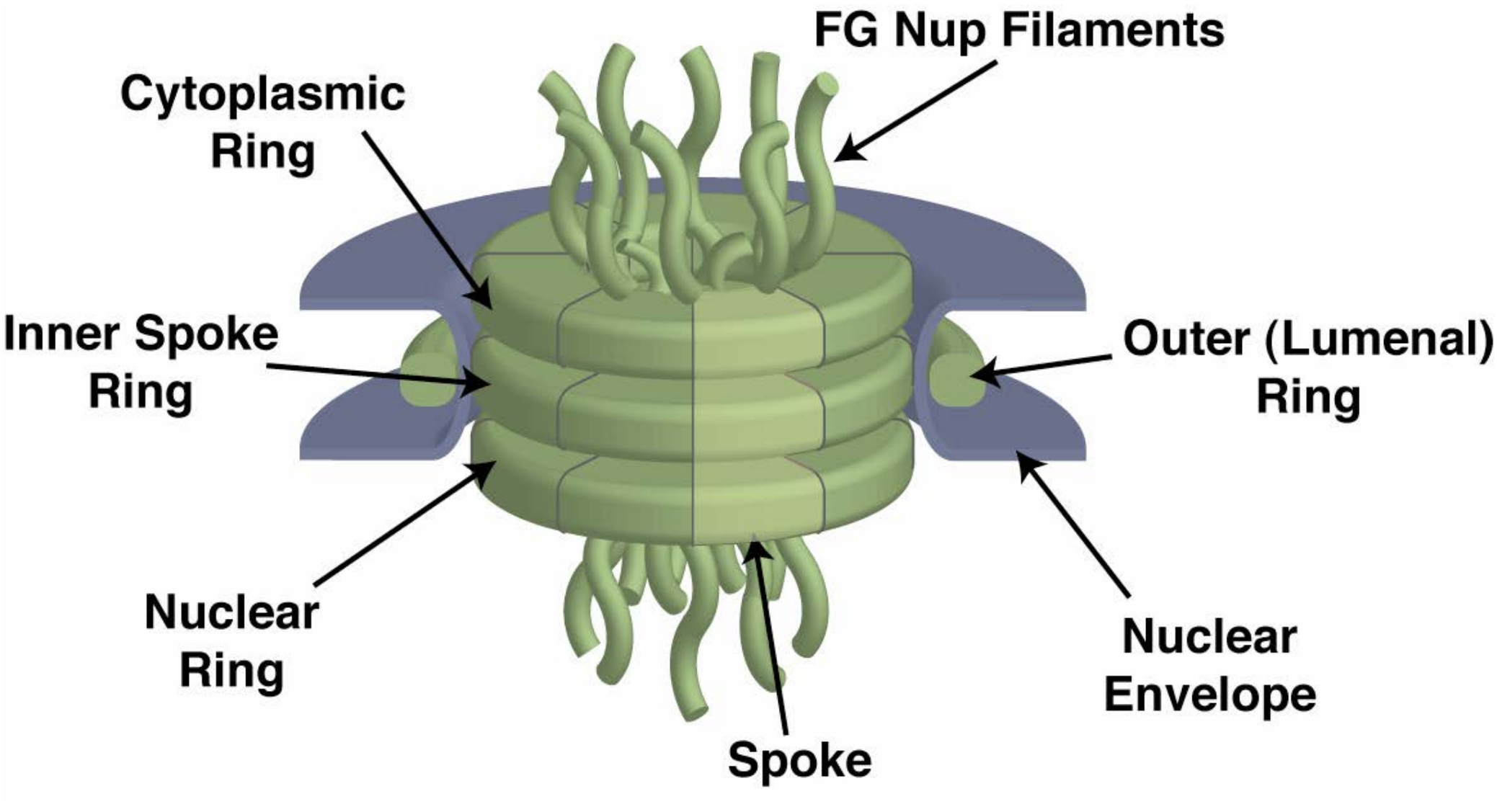
UCSF
University of California
San Francisco

Analysis

- Clustering
- Output
  - Chimera
  - Pymol
  - PDBs
  - Density maps

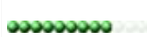
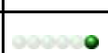
The NPC

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



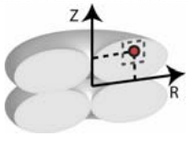
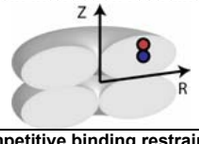
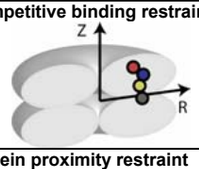
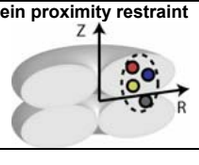
Representation

436 proteins!

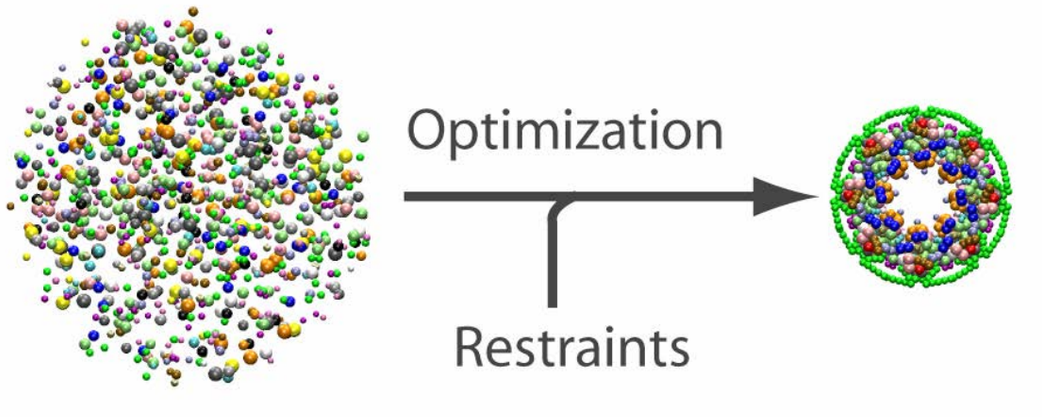
τ	N_τ^1	N_τ^2	κ	$\{B_j^\kappa\}$	n_κ	r	τ	N_τ^1	N_τ^2	κ	$\{B_j^\kappa\}$	n_κ	r
Nup192	1	1	1,2,5		2	3.0	**Nup1**	0	1	1,5		9	1.5
			3	-	1	-				2		2	1.5
Nup188	1	1	1,2,5		2	3.0				3	-	1	-
			3	-	1	-				4		7	1.5
Nup170	1	1	1,2,5		2	2.9	**Nsp1**	2	2	1,5		12	1.3
			3	-	1	-				2		3	1.3
Nup157	1	1	1,2,5		3	2.5				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

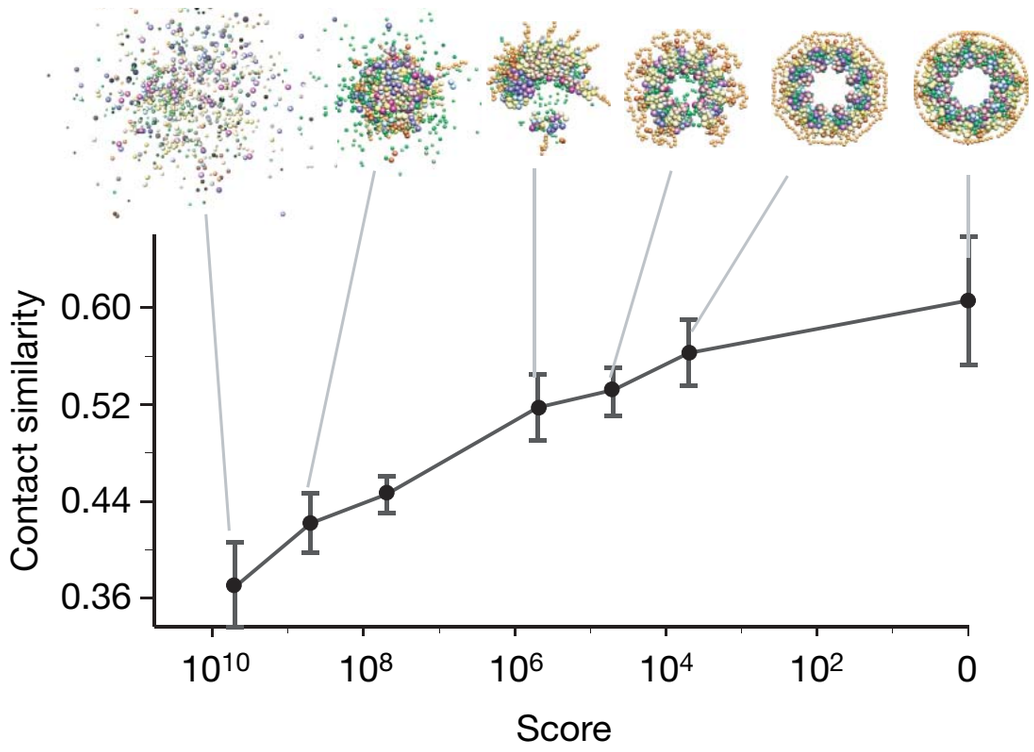
UCSF
University of California
San Francisco

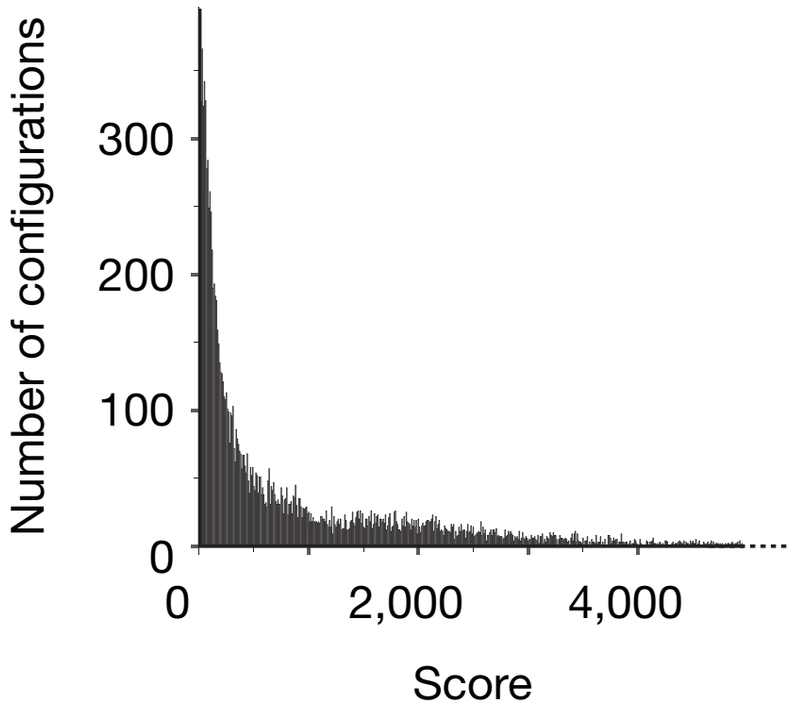
Scoring

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: $B^{ms} = \{B_j^{\kappa=1}(\theta,s,\tau,i)\}$
	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: $B^{ms} = \{B_j^{\kappa=6}(\theta,s,\tau,i) \mid \tau \in (\text{Ndc1,Pom152,Pom34})\}$
	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: $B^{ms} = \{B_j^{\kappa=8}(\theta,s,\tau,i) \mid \tau \in (\text{Ndc1,Pom152,Pom34})\}$
			-	-	**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: $B^{ms} = \{B_j^{\kappa=7}(\theta,s,\tau,i)\tau \in (\text{Pom152})\}$
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}: $B^{ms} = \{B_j^{\kappa=1}(\theta,s,\tau,i) \mid \tau \in C_{51}\}$
	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: $B = \{B_j^{\kappa}(\theta,s,\tau,i) \mid \theta \in (1,2,3), \kappa \in (2,4,9)\}$

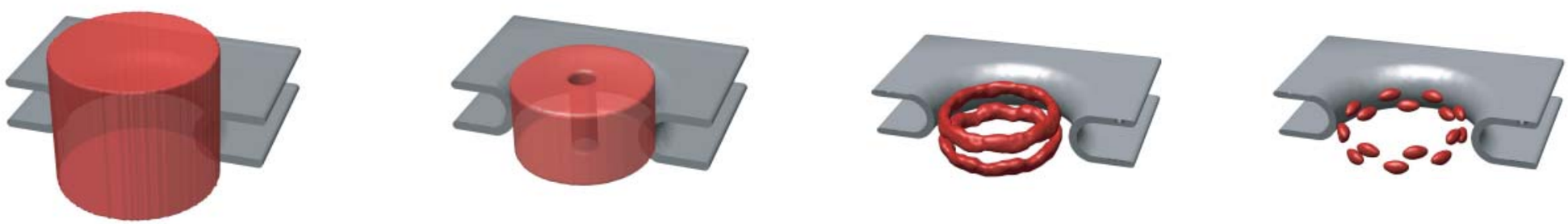
Optimization







Integrating data



Nuclear envelope pore volume

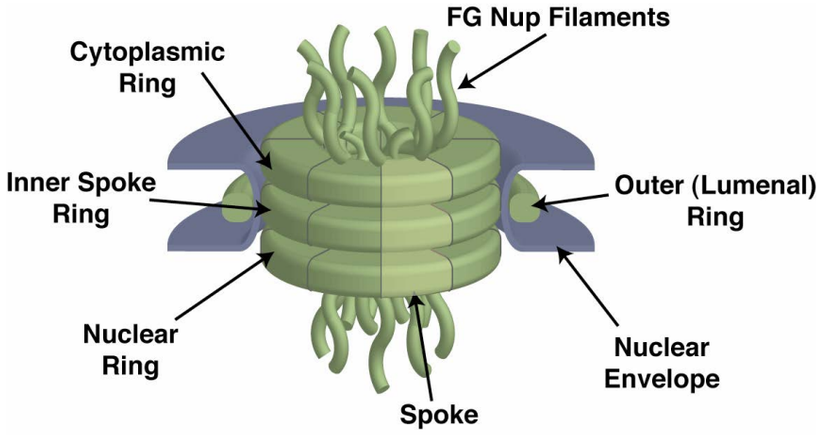
+ Immuno-EM

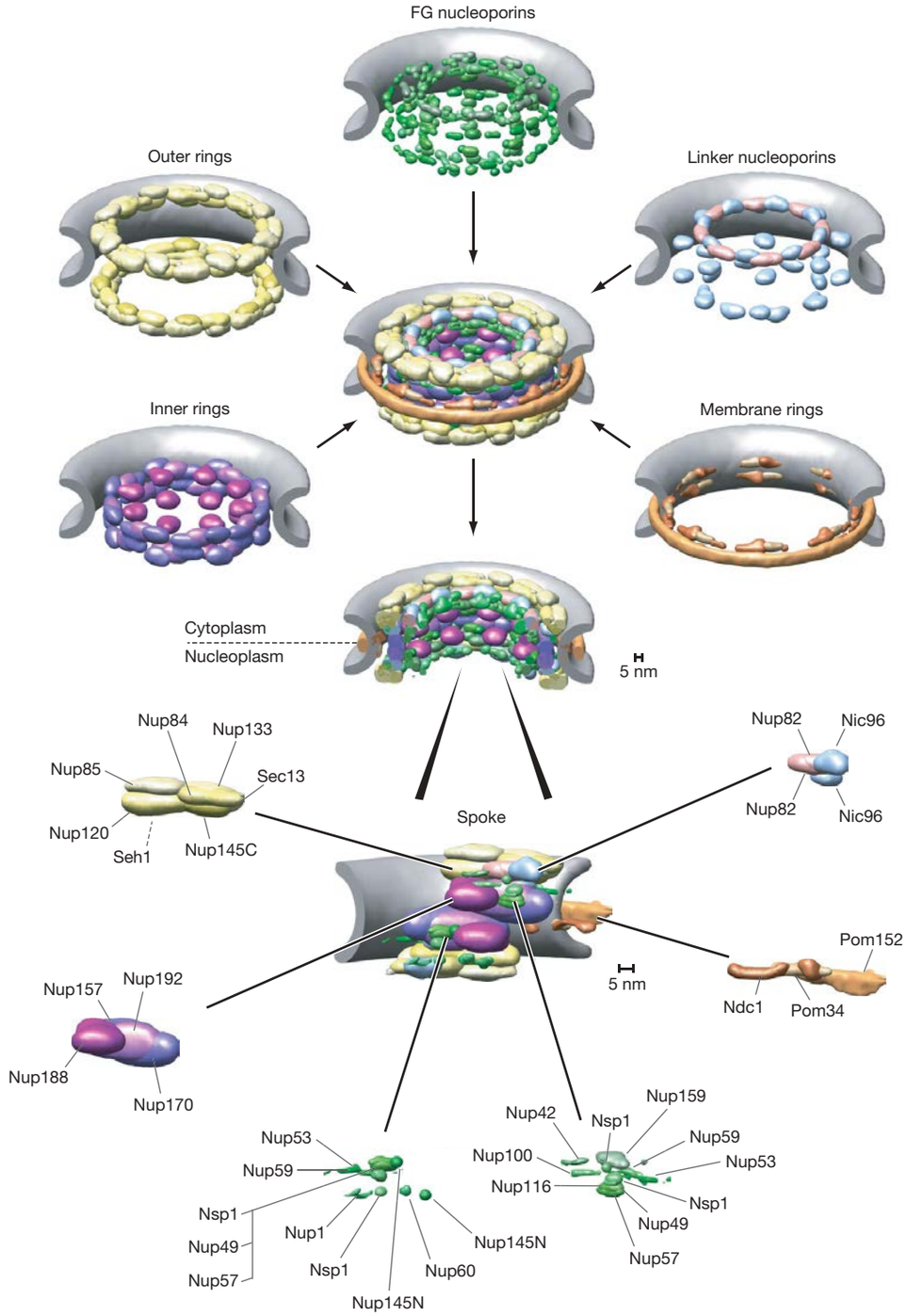
+ Ultracentrifugation
Nucleoporin stoichiometry
NPC symmetry

+ Overlay assays
Affinity purifications

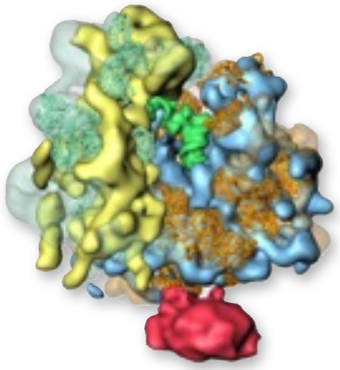
University of California
San Francisco

The STRUCTURE of NPC

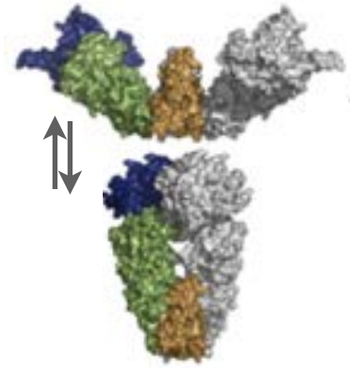




IMP-based efforts

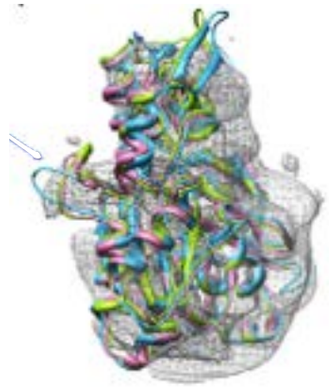


Ribosomes,
Sali, Frank; Sali, Akey

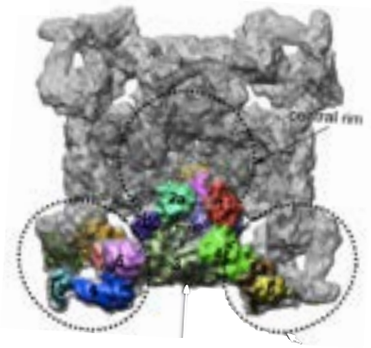


Hsp90 landscape
Sali, Agard

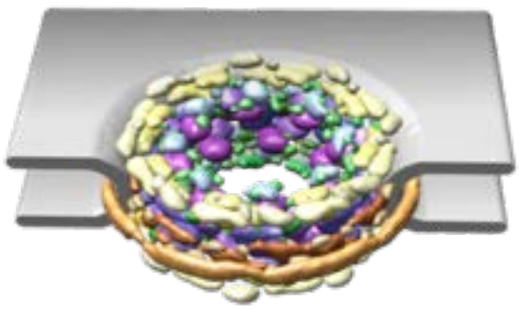
TRiC/CCC
Sali, Frydman, Chiu



Actin
Sali, Chiu

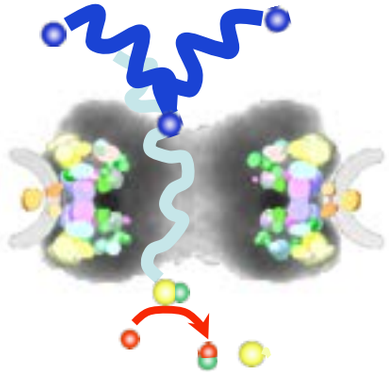


RyR channel
Sali, Serysheva, Chiu

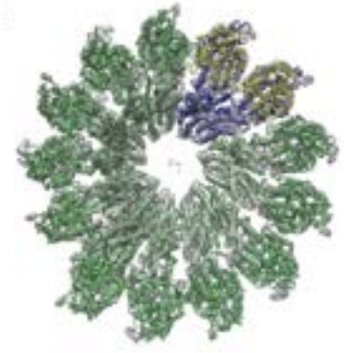


Nuclear Pore Complex,
Sali, Rout, Chait

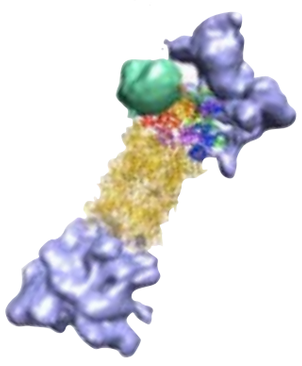
Nup84 complex,
Sali, Rout, Chait



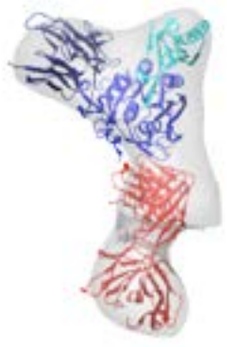
Nuclear Pore Complex transport,
Sali, Rout, Chait, Chook, Liphardt



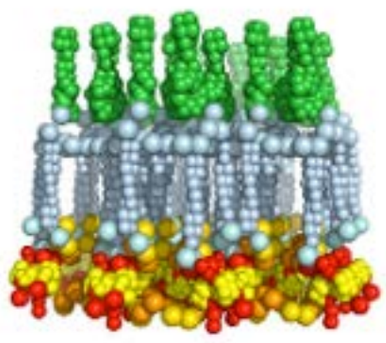
Microtubule nucleation
Sali, Agard



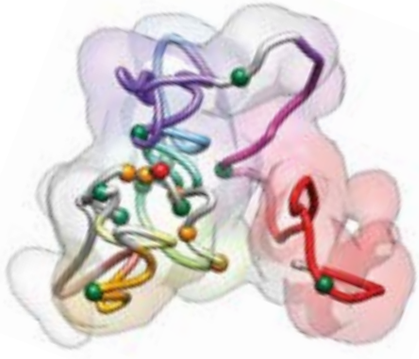
26 Proteasome
Sali, Baumeister



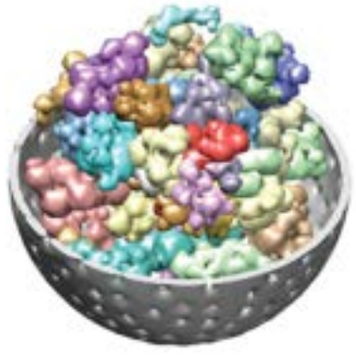
PCS9K-Fab complex
Sali, Cheng, Agard, Pons



Spindle Pole Body
Sali, Davis, Muller

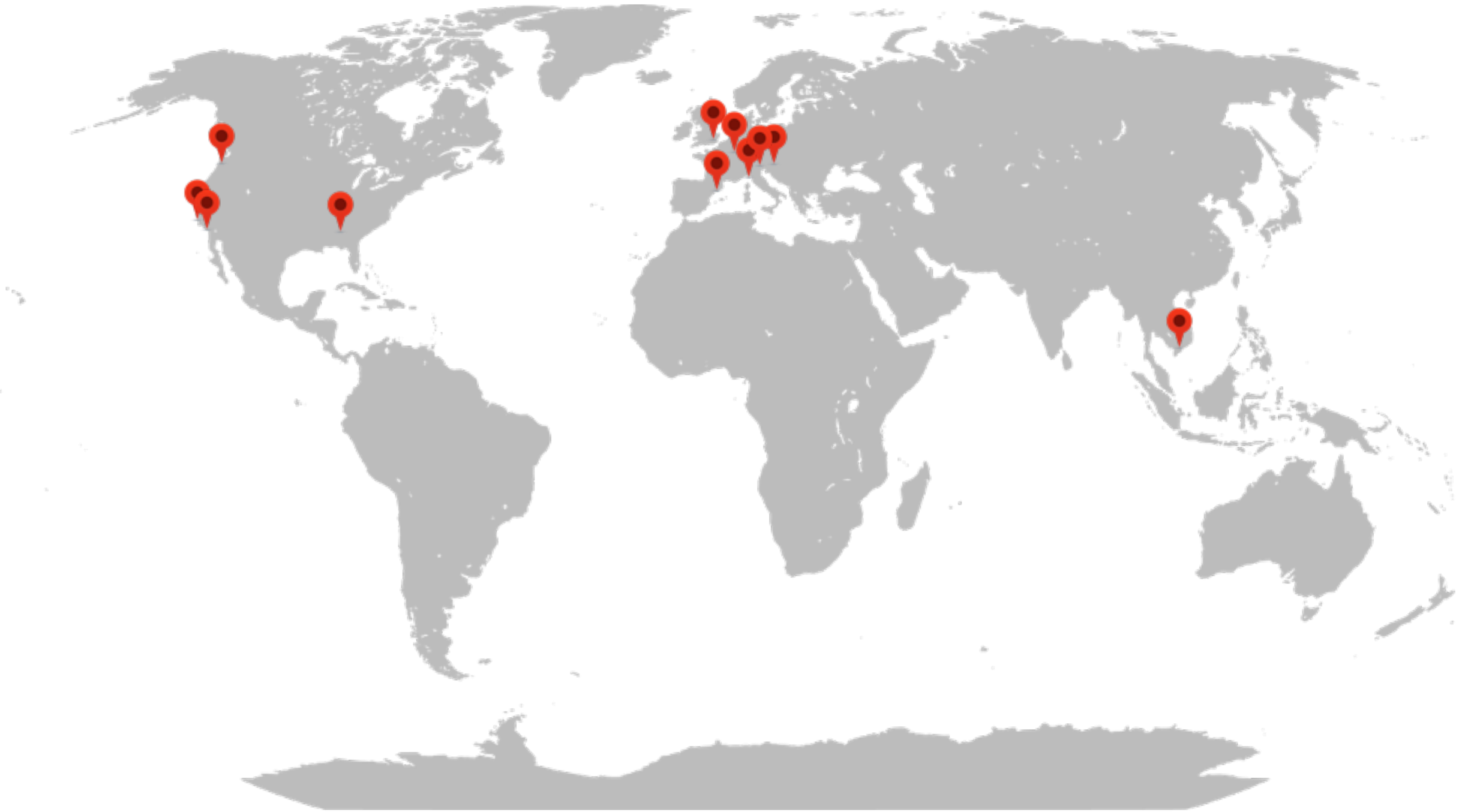


Chromatin globin domain
Marti-Renom

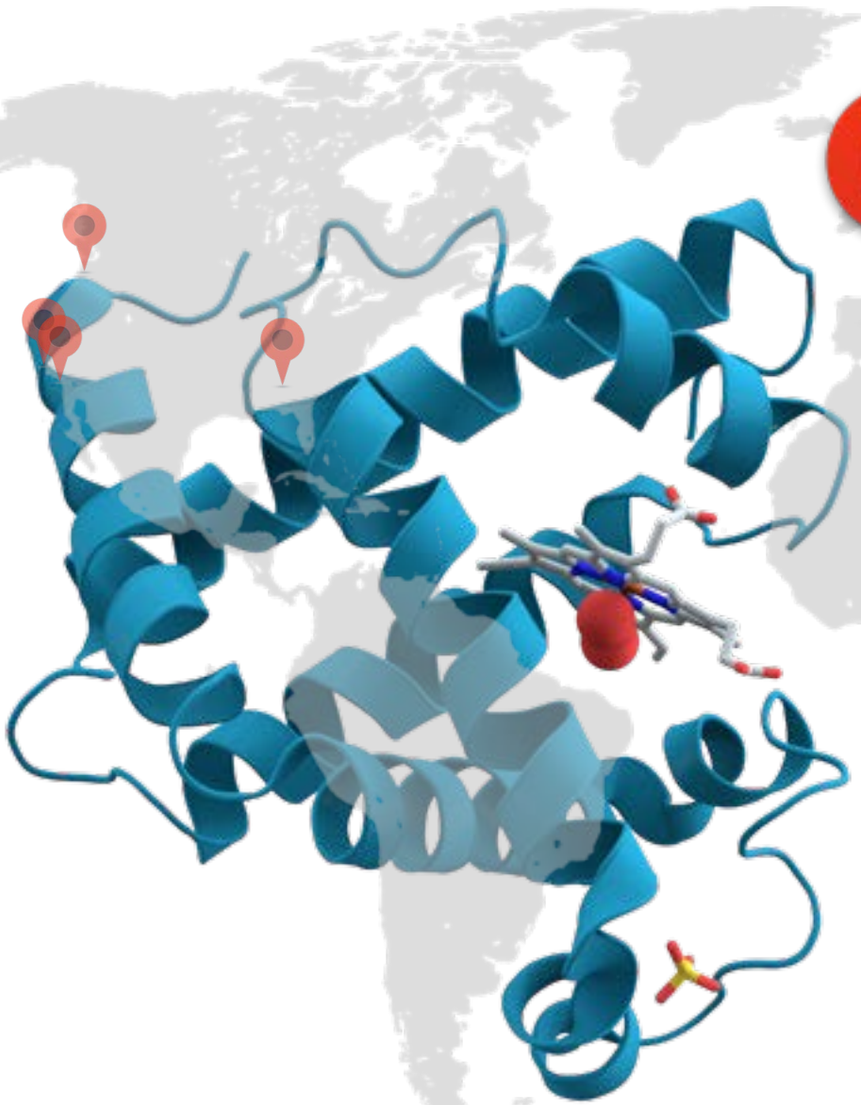


Lymphoblastoid cell genome
Alber, Chen

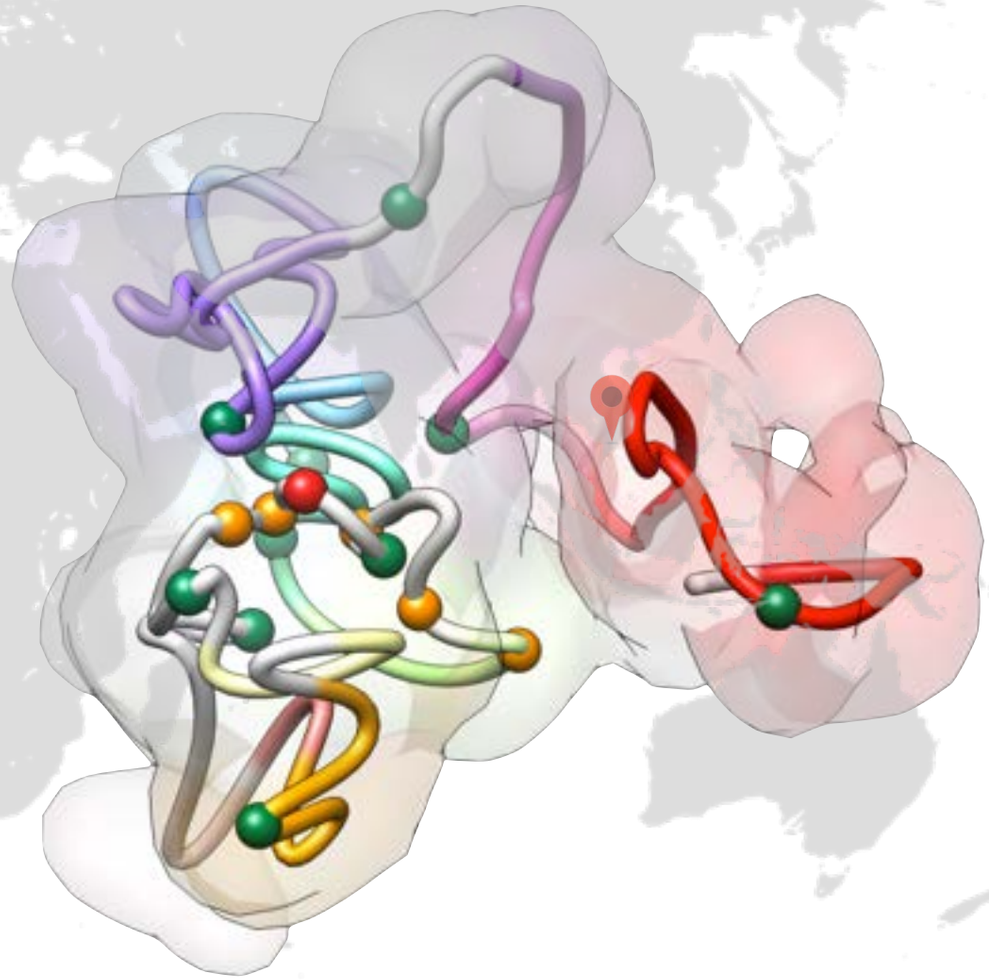
Who Is developing with IMP?



From proteins to genomes



Whale sperm myoglobin structure (1960)

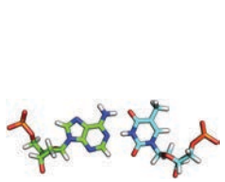


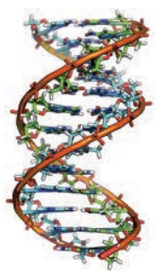
alpha-globin genomic domain structure (2011)

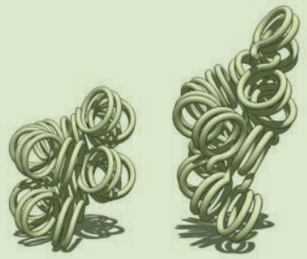
Resolution Gap

Marti-Renom, M. A. & Mirny, L. A. PLoS Comput Biol 7, e1002125 (2011)

Knowledge

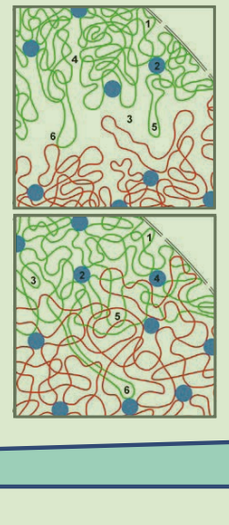


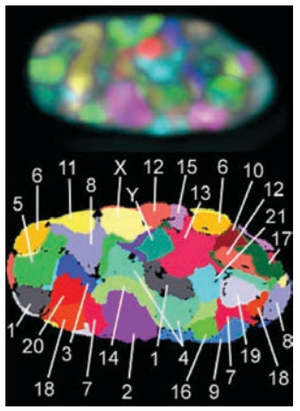




IDM

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

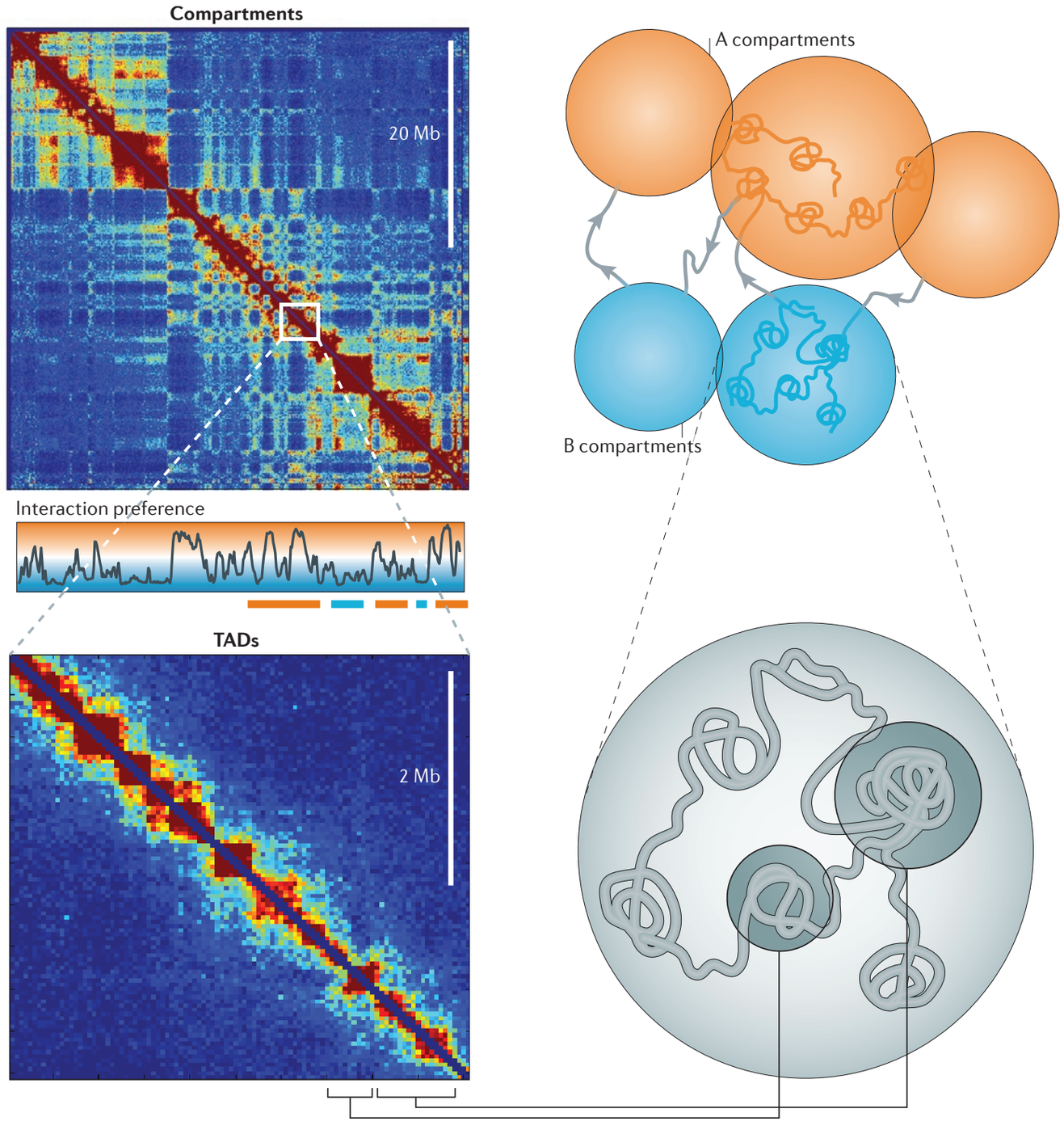
10^{-10} 10^{-8} 10^{-6} 10^{-4} 10^{-2} 10^0 10^2 10^3 s

Resolution

10^{-3} 10^{-2} 10^{-1} μ

"Bridging" the Resolution Gap

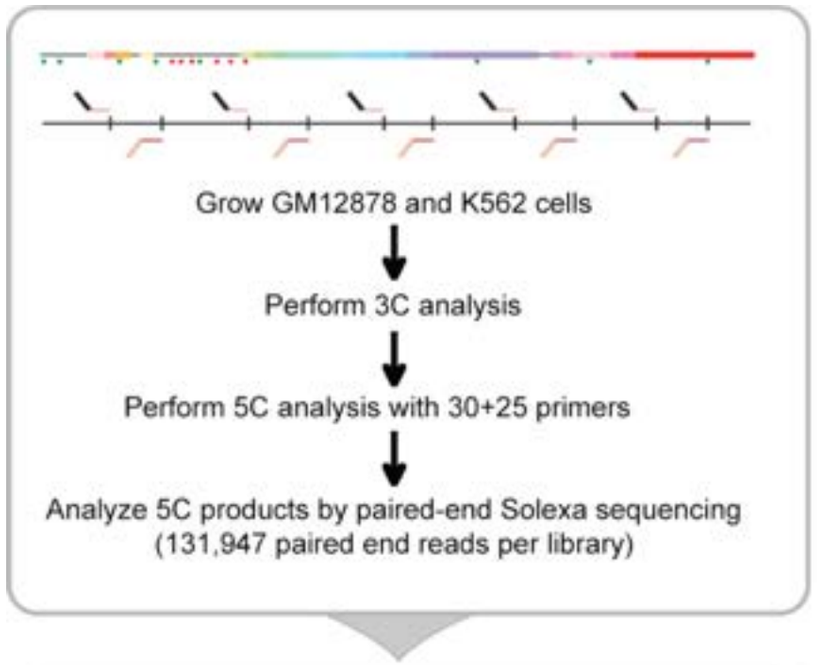
Dekker, J., Marti-Renom, M. A., & Mirny, L. A. (2013). Nature Reviews Genetics, 14(6), 390–403.

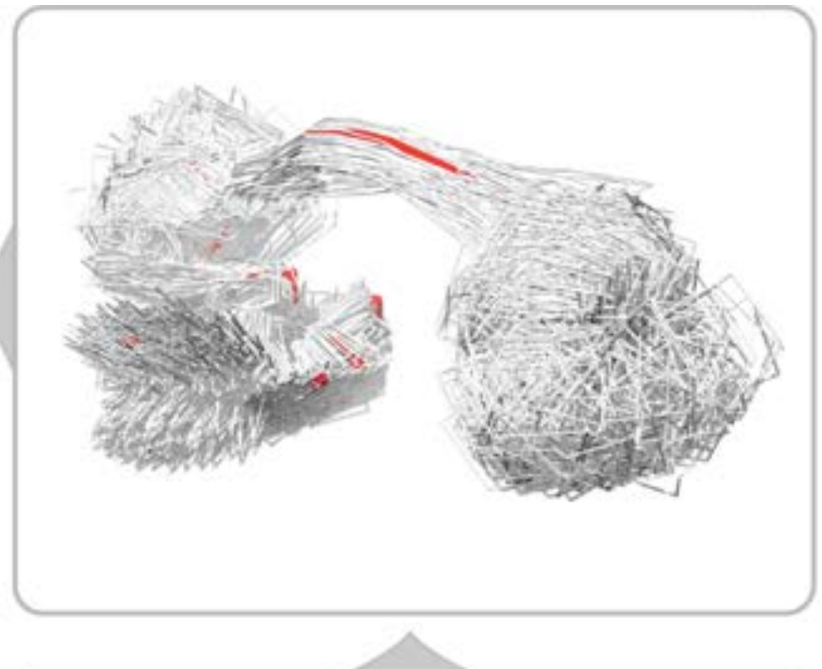


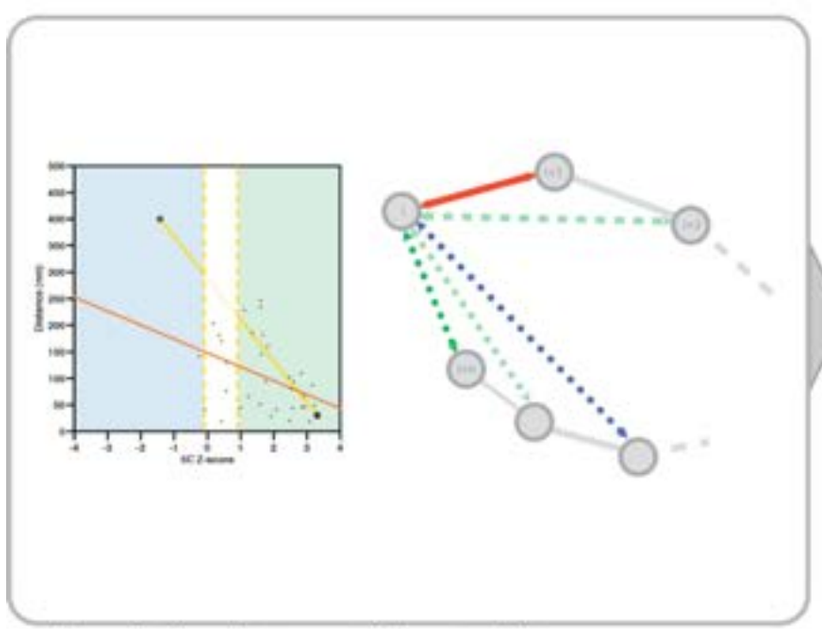
Hybrid Method

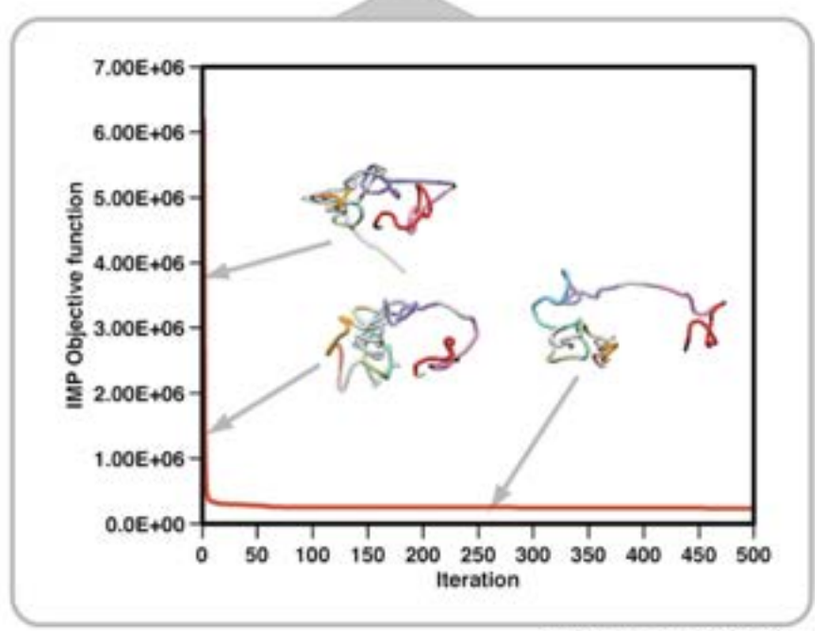
Baù, D. & Marti-Renom, M. A. Methods 58, 300–306 (2012).

Experiments







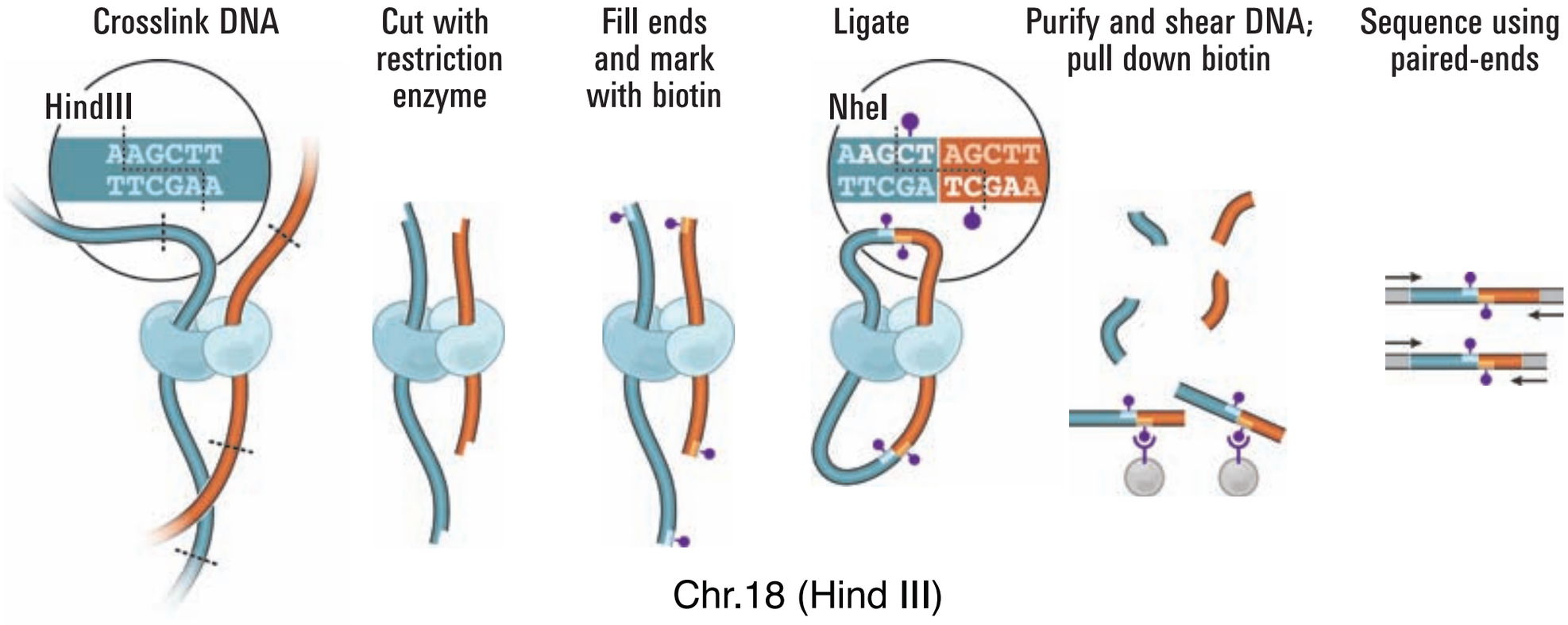


Computation

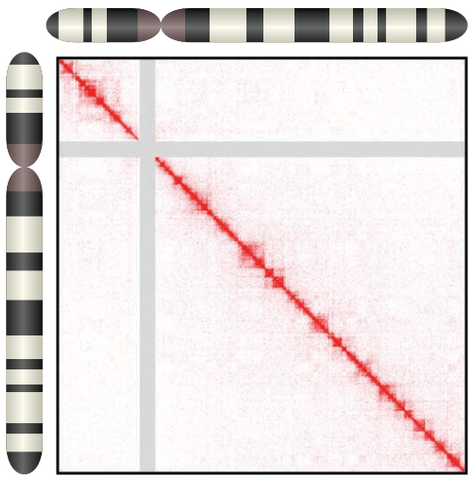
Hi-C technology

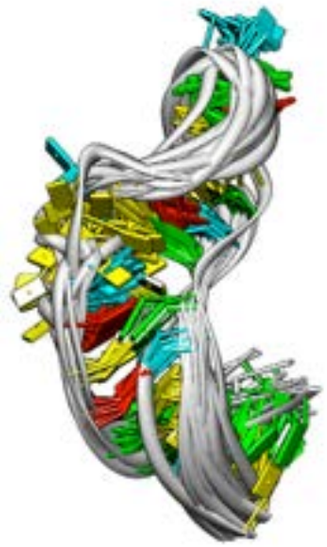
Lieberman-Aiden, E. et al. Science 326, 289–293 (2009).

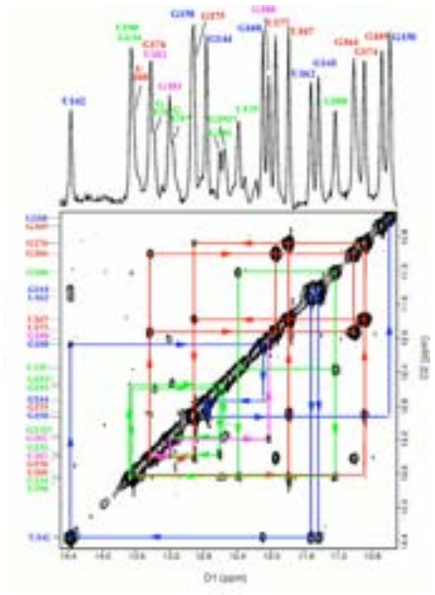
http://3dg.umassmed.edu



Chr.18 (Hind III)

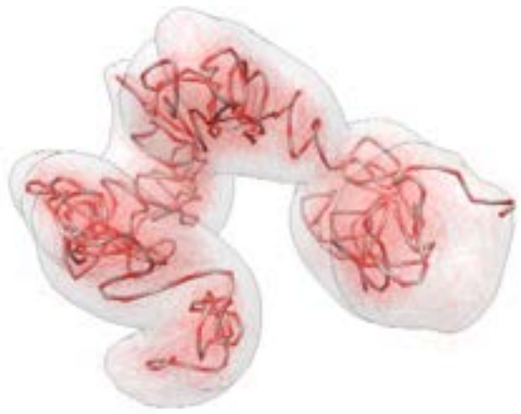


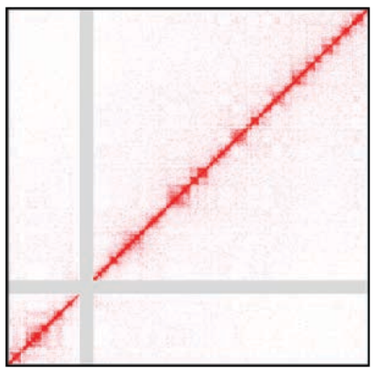




Biomolecular structure determination

2D-NOESY data



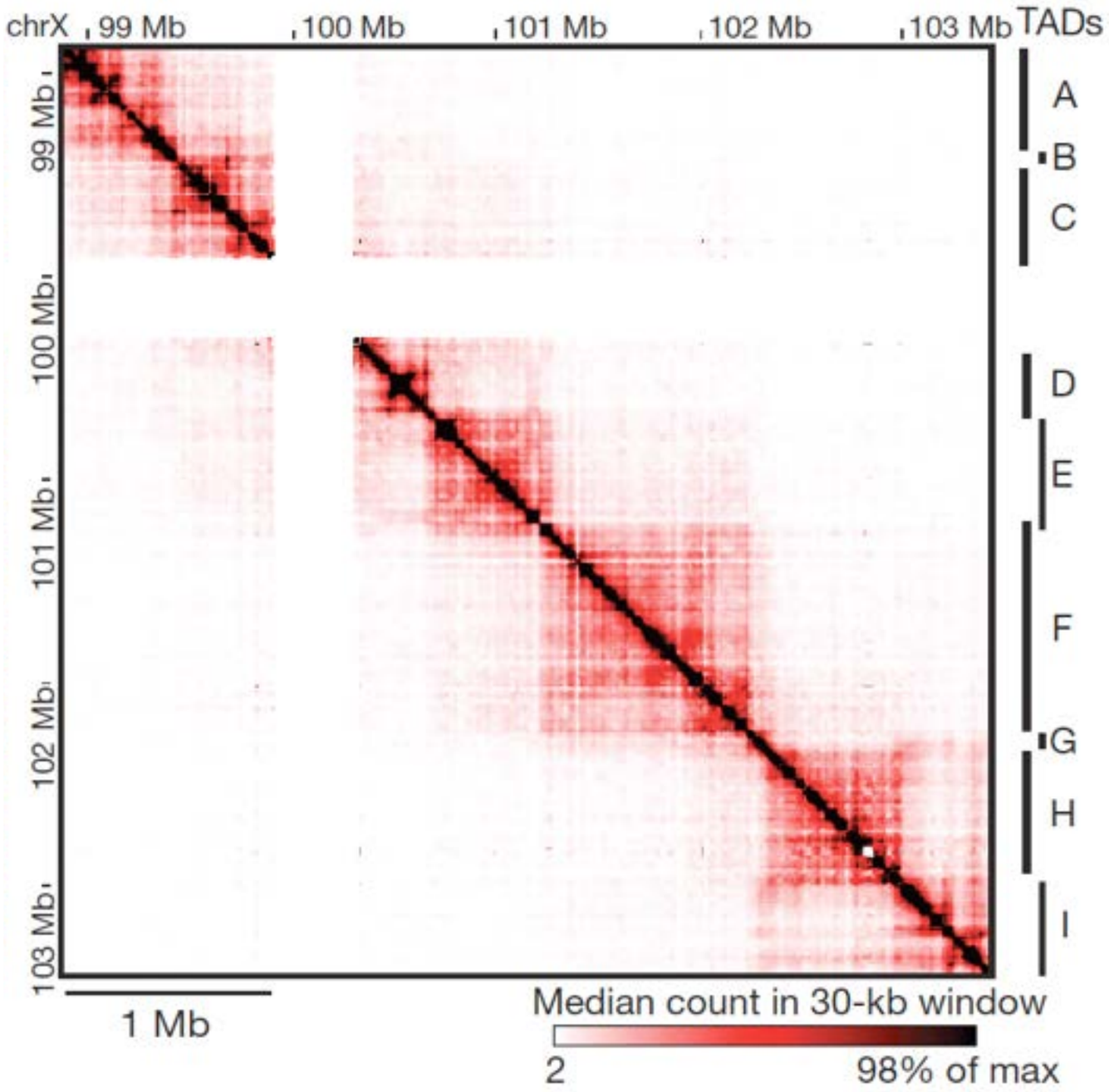


Chromosome structure determination

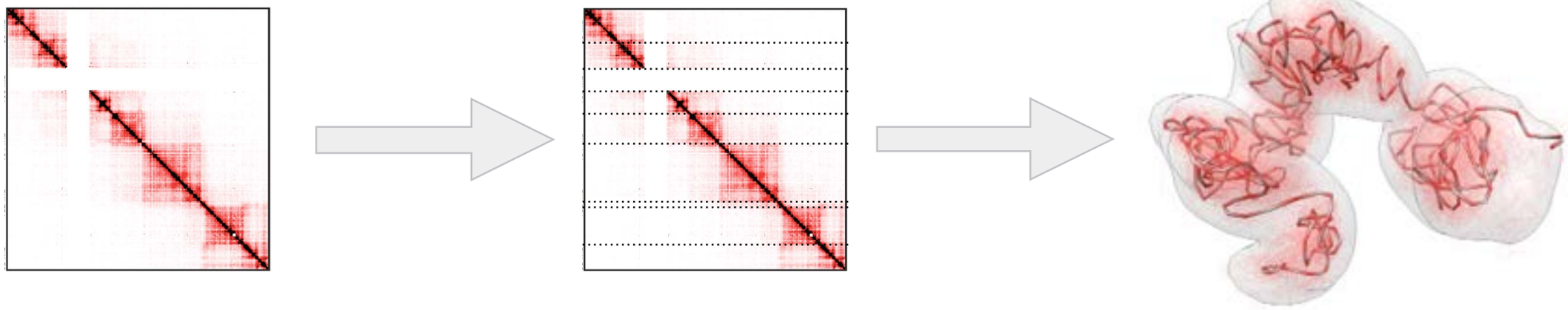
3C-based data

3C-like data

Nora, E. P., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature

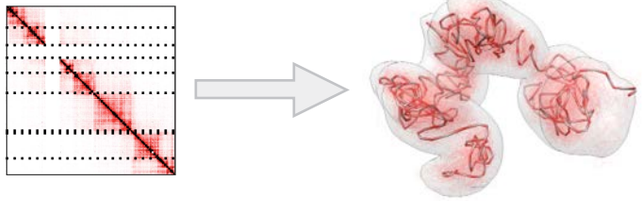


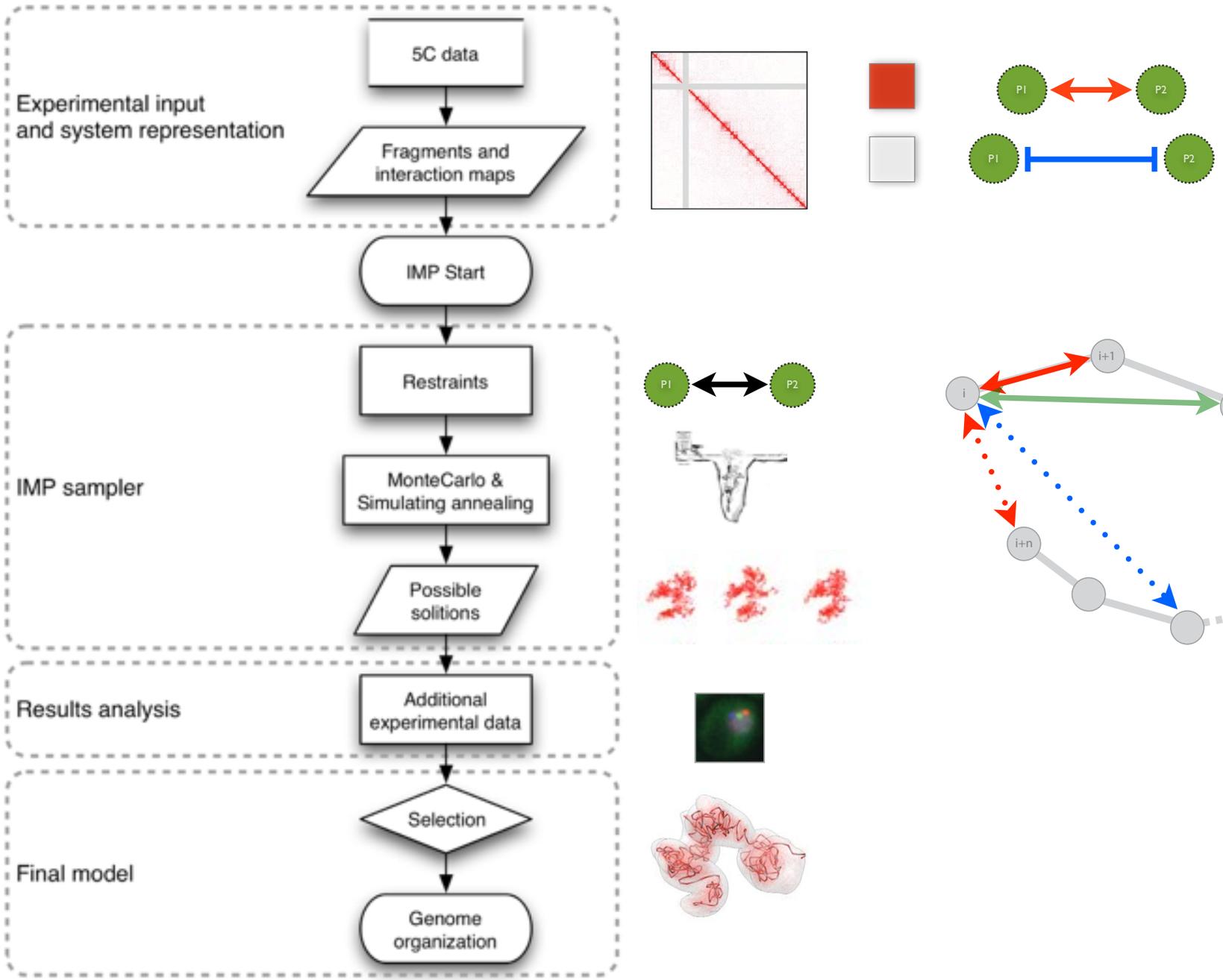
TADbit



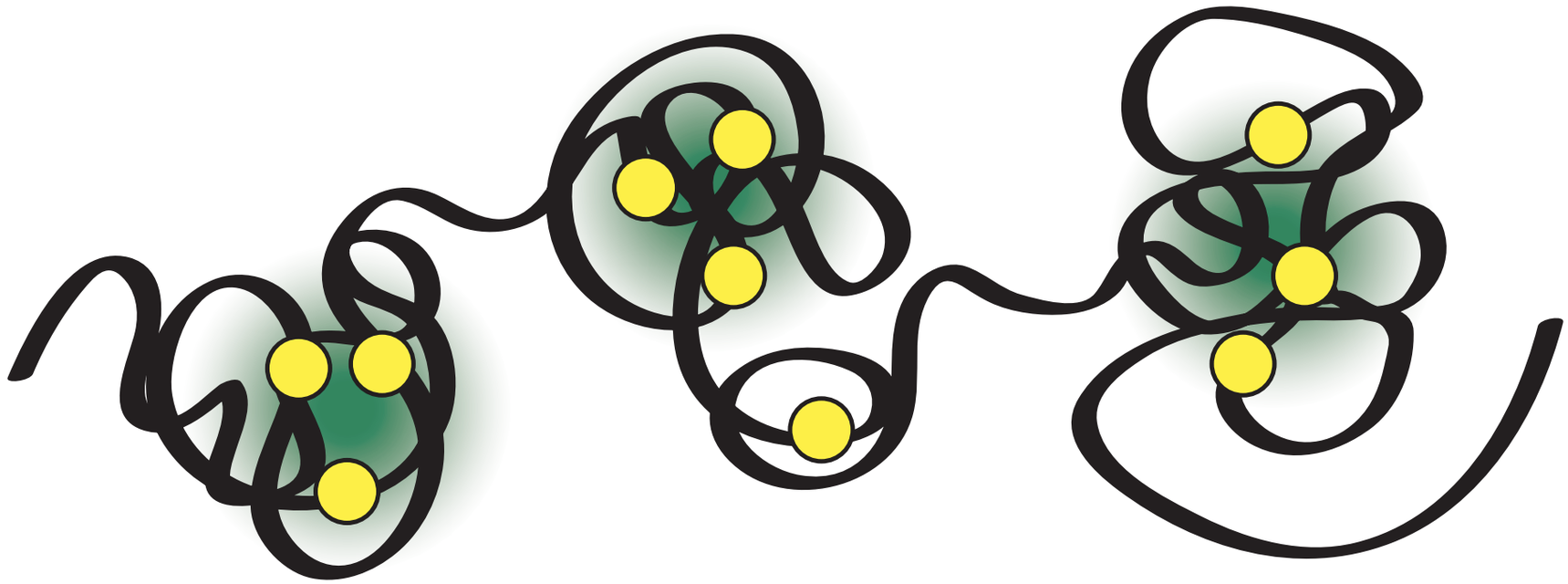
http://www.3Dgenomes.org

TADbit




Experimental input
and system representation
5C data
Fragments and
interaction maps
IMP Start
IMP sampler
Restraints
MonteCarlo &
Simulating annealing
Possible
solitions
Results analysis
Additional
experimental data
Final model
Selection
Genome
organization
P1
P2
P1
P2
P1
P2
i
i+1
i+2
i+n

On TADs and hormones



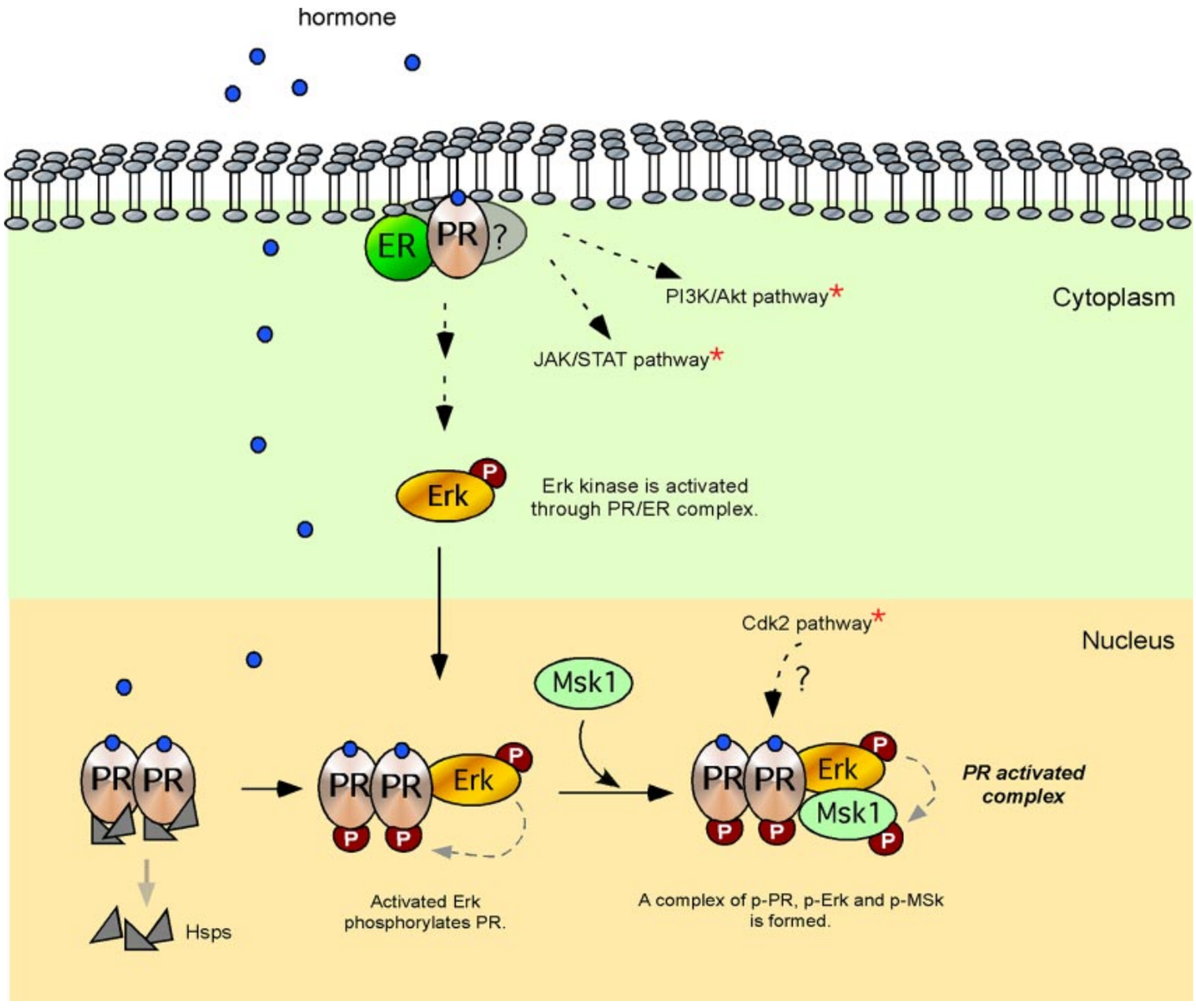
François le Dily

Davide Baù

François Serra

cnag CRG Centre for Genomic Regulation

Progesterone-regulated transcription in breast cancer



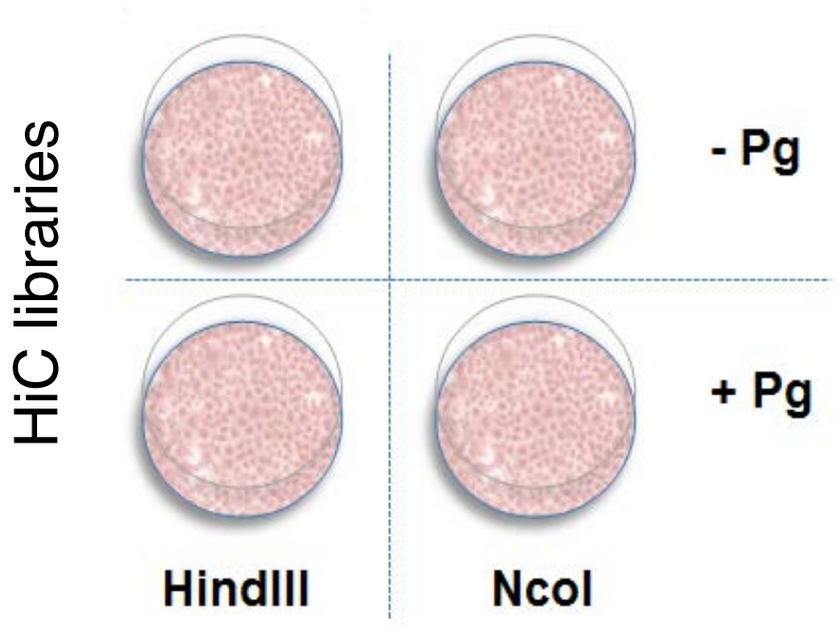
> 2,000 genes **Up**-regulated

> 2,000 genes **Down**-regulated

Regulation in 3D?

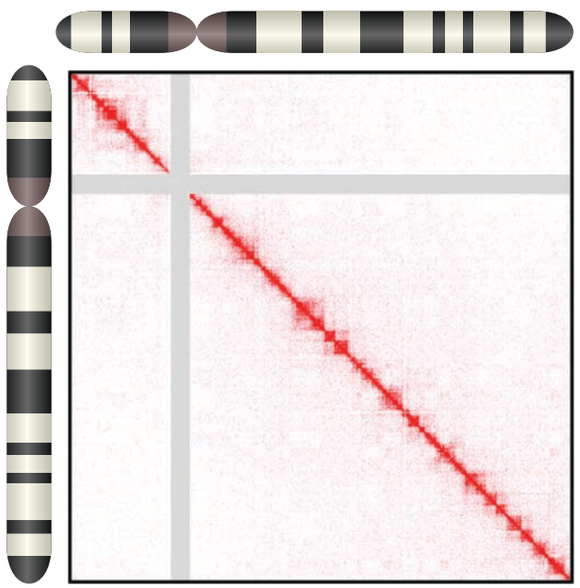
Vicent *et al* 2011, Wright *et al* 2012, Ballare *et al* 2012

Experimental design

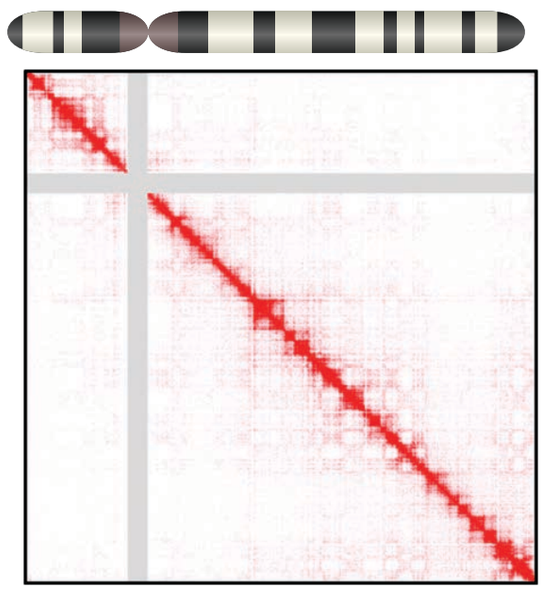


ChIP-Seq
RNA-Seq
Hi-C

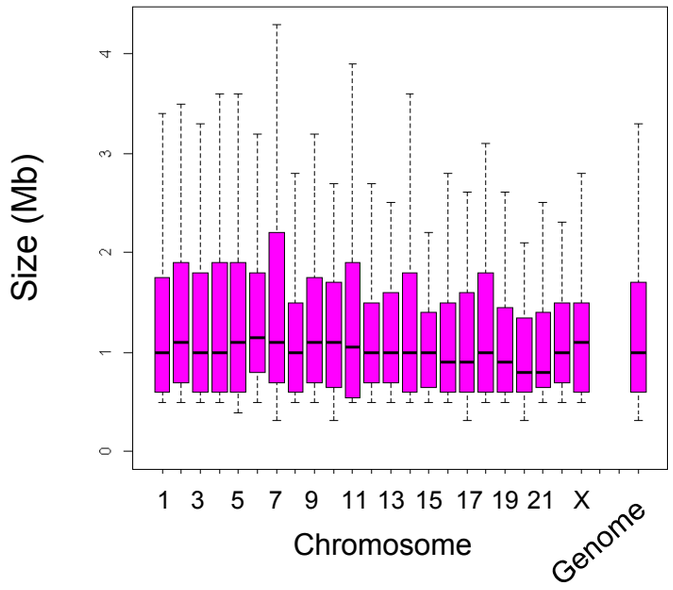
Chr.18 (Hind III)



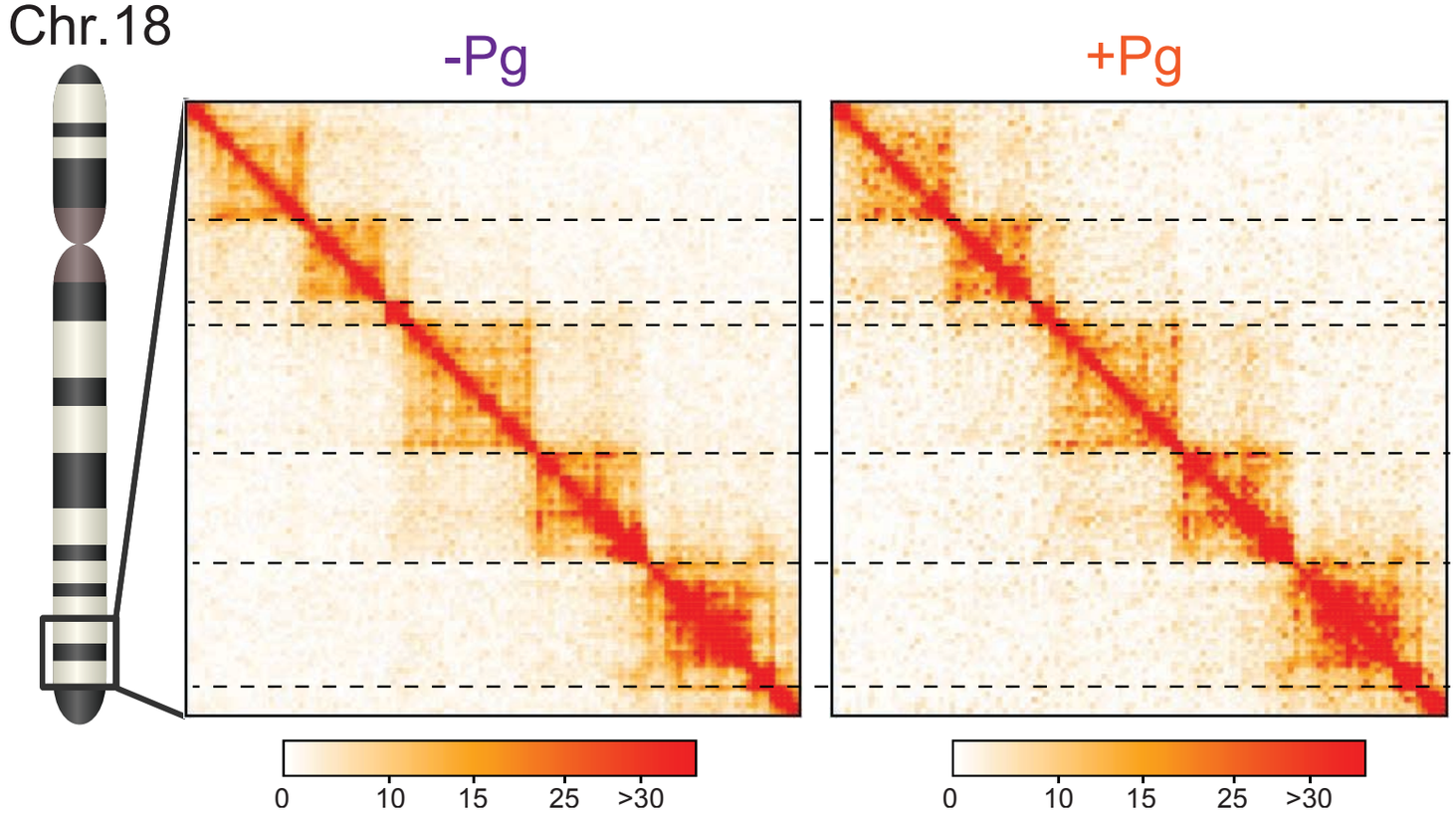
Chr.18 (NcoI)

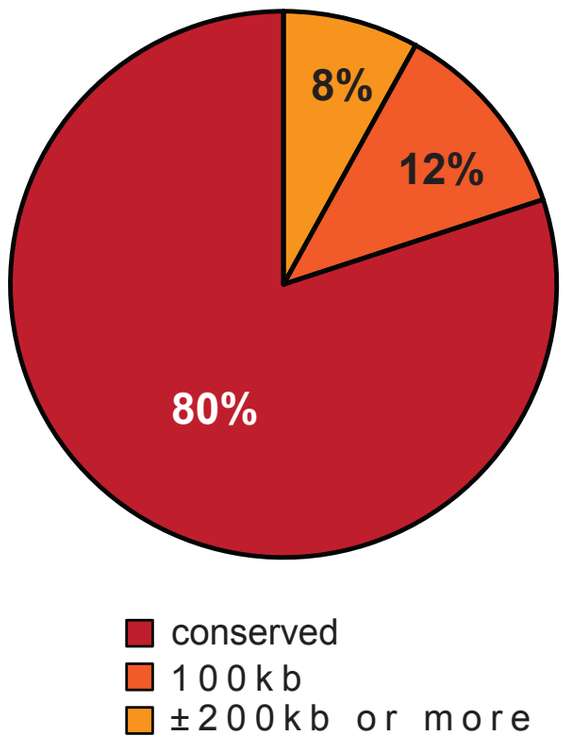


Are there TADs? how robust?

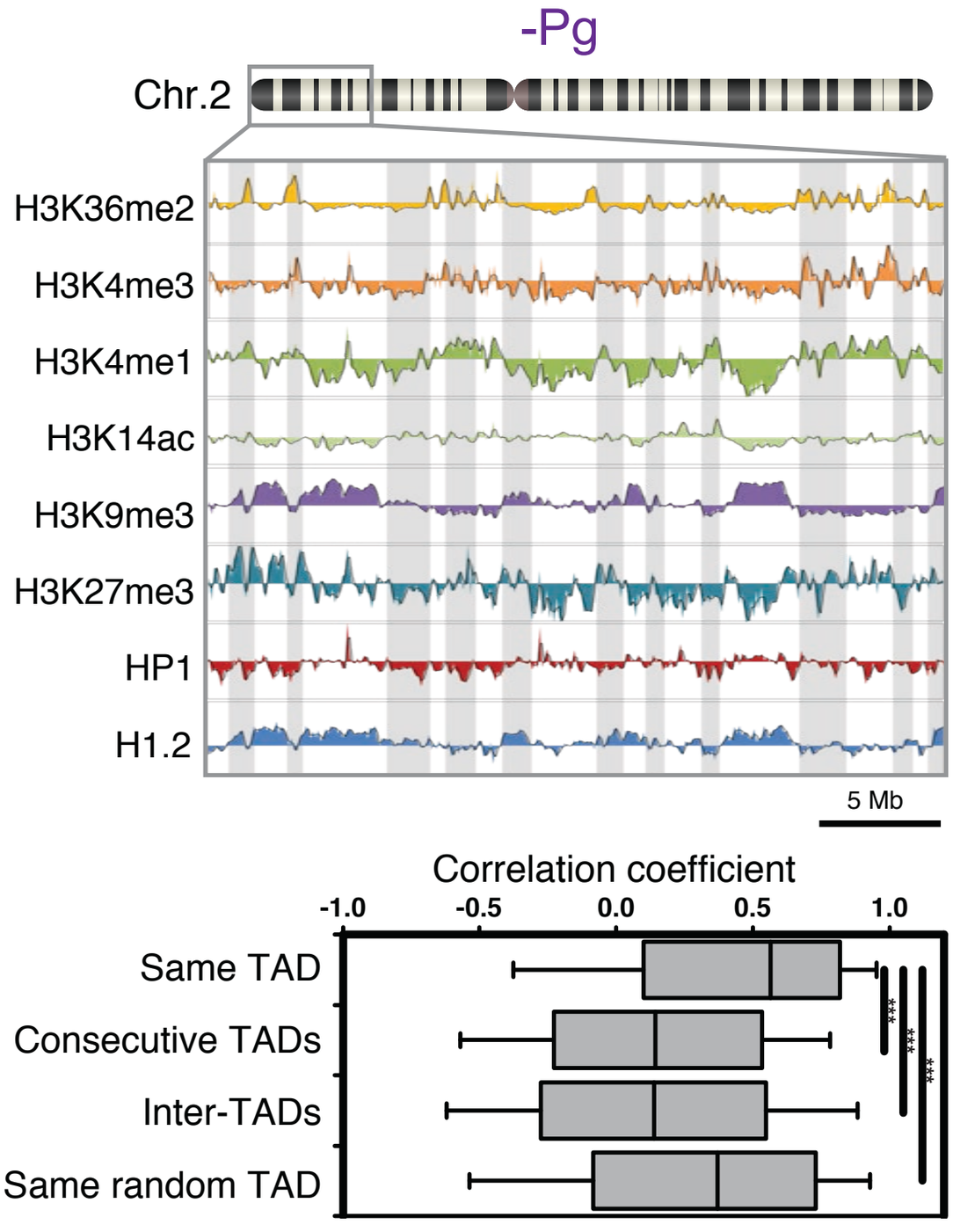


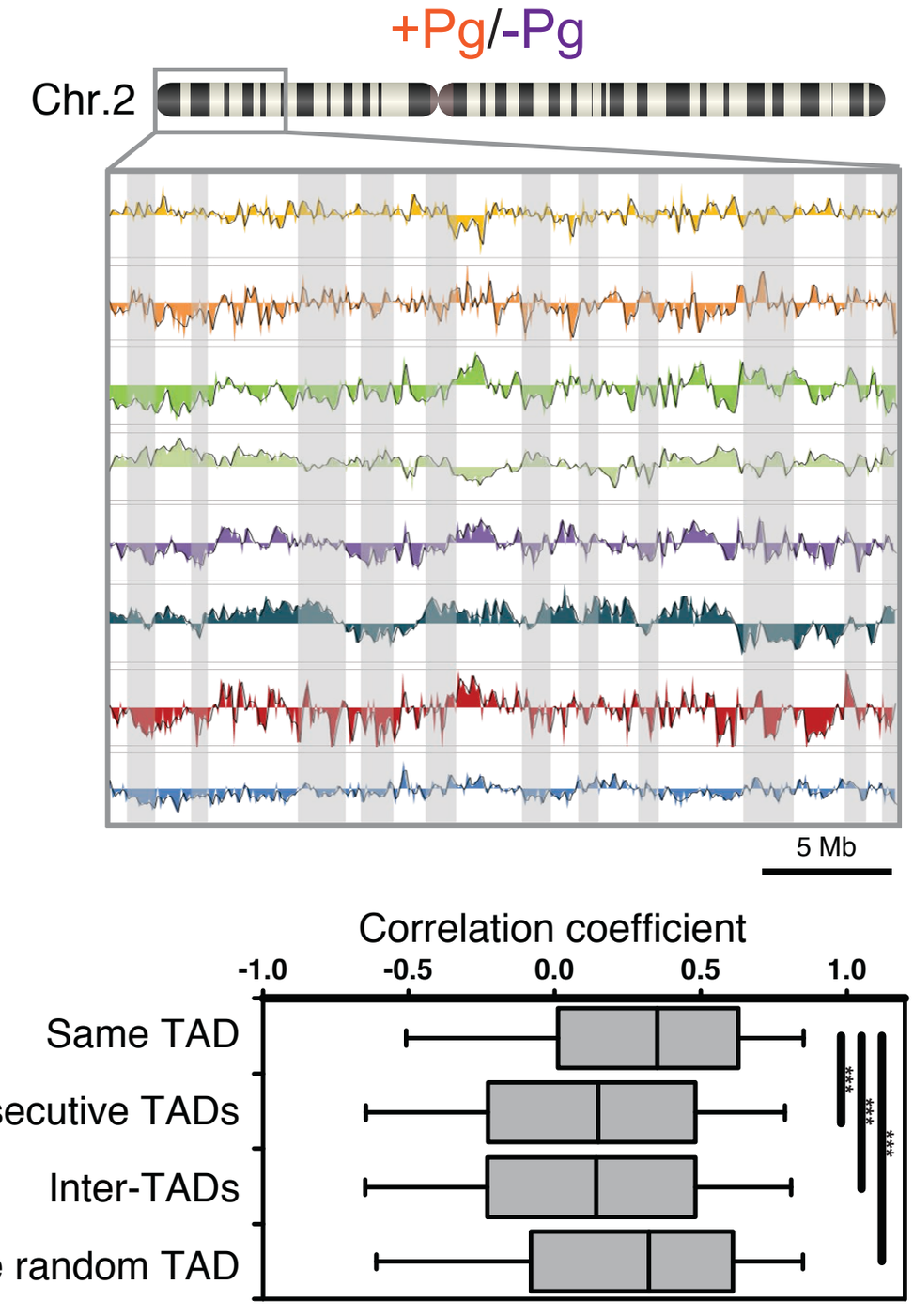
>2,000 detected TADs



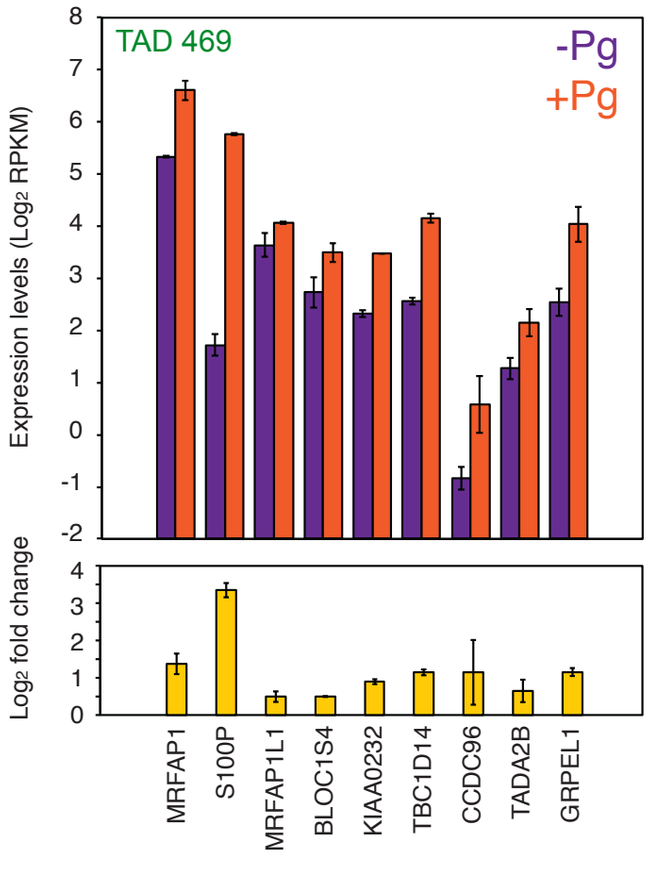


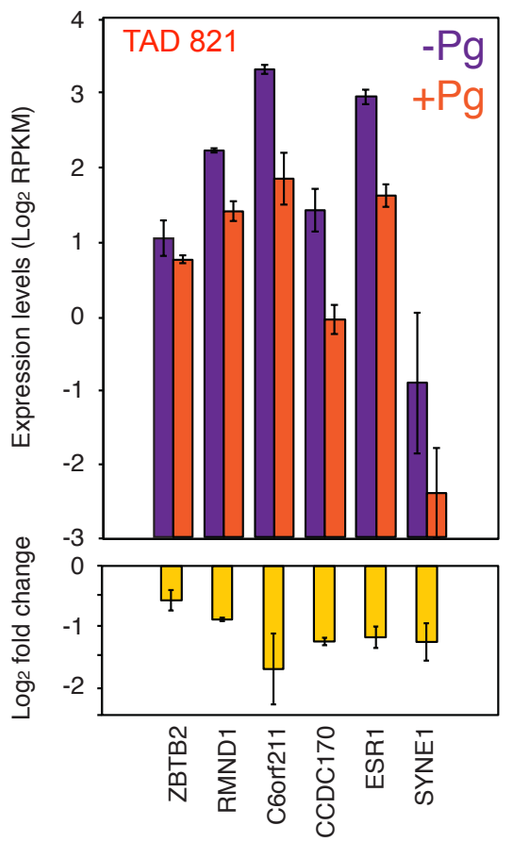
Are TADs homogeneous?

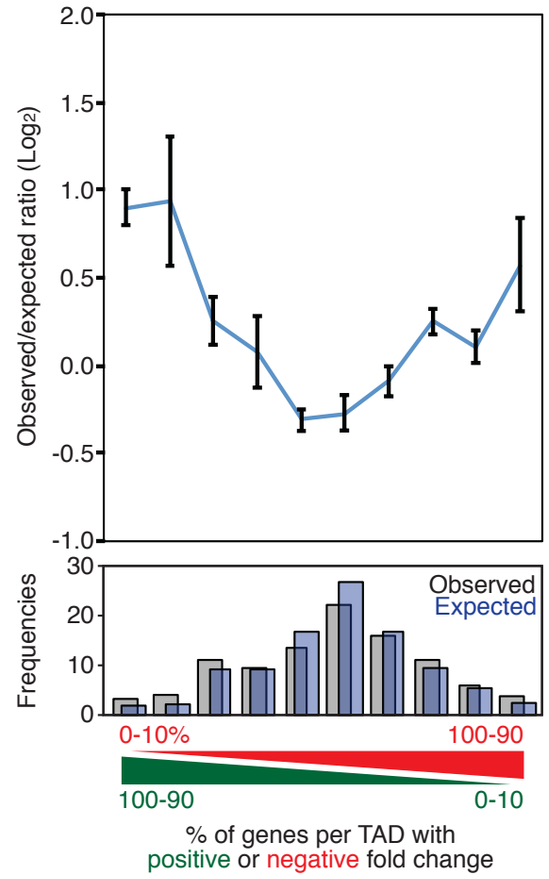




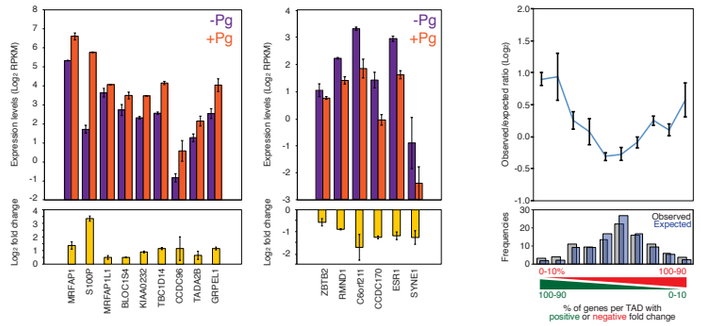
Do TADs respond differently to Pg treatment?



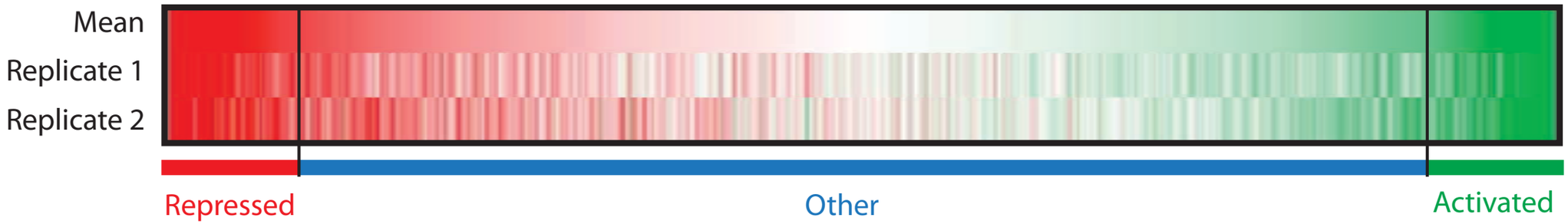




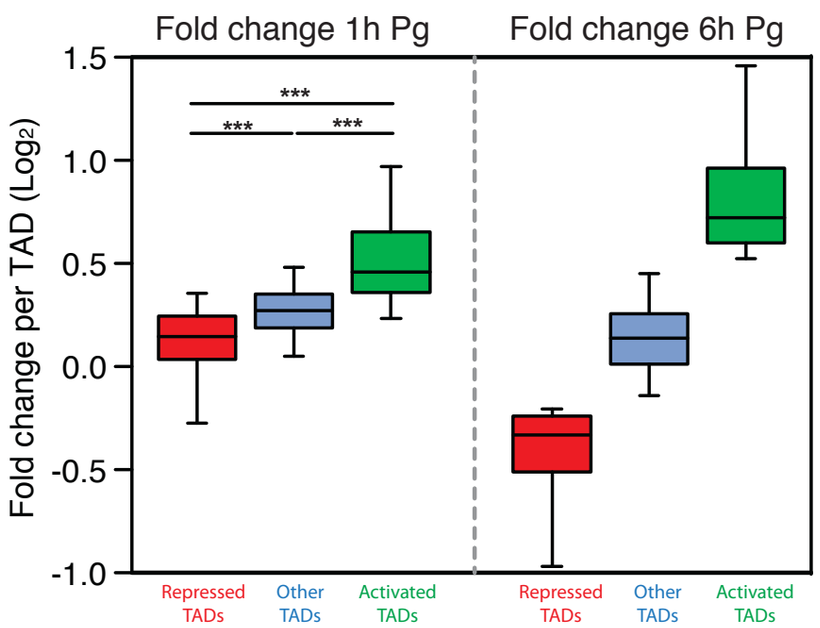
Do TADs respond differently to Pg treatment?

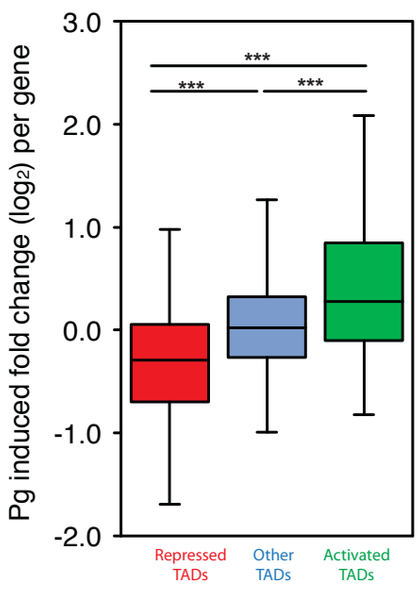


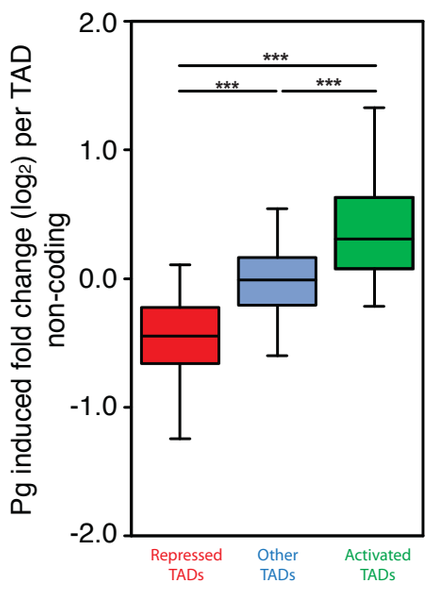
Pg induced fold change per TAD (6h)

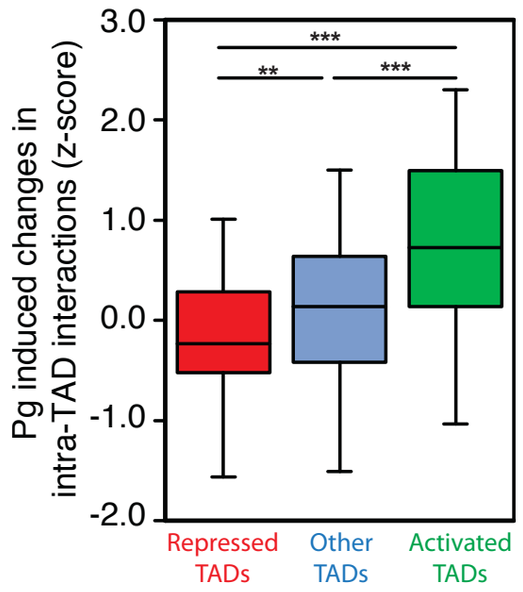


Repressed TADs | Other TADs | Activated TADs



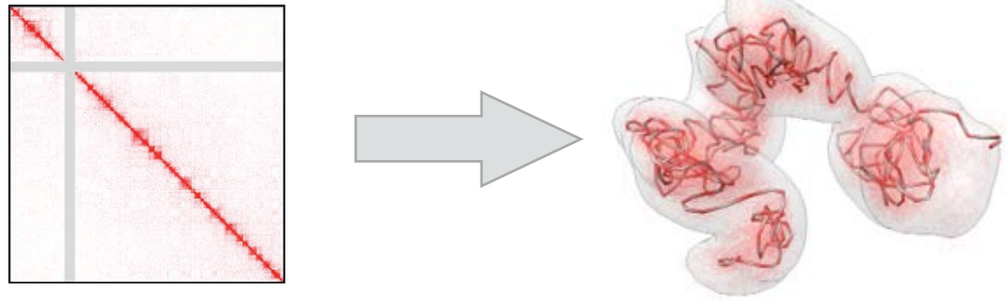




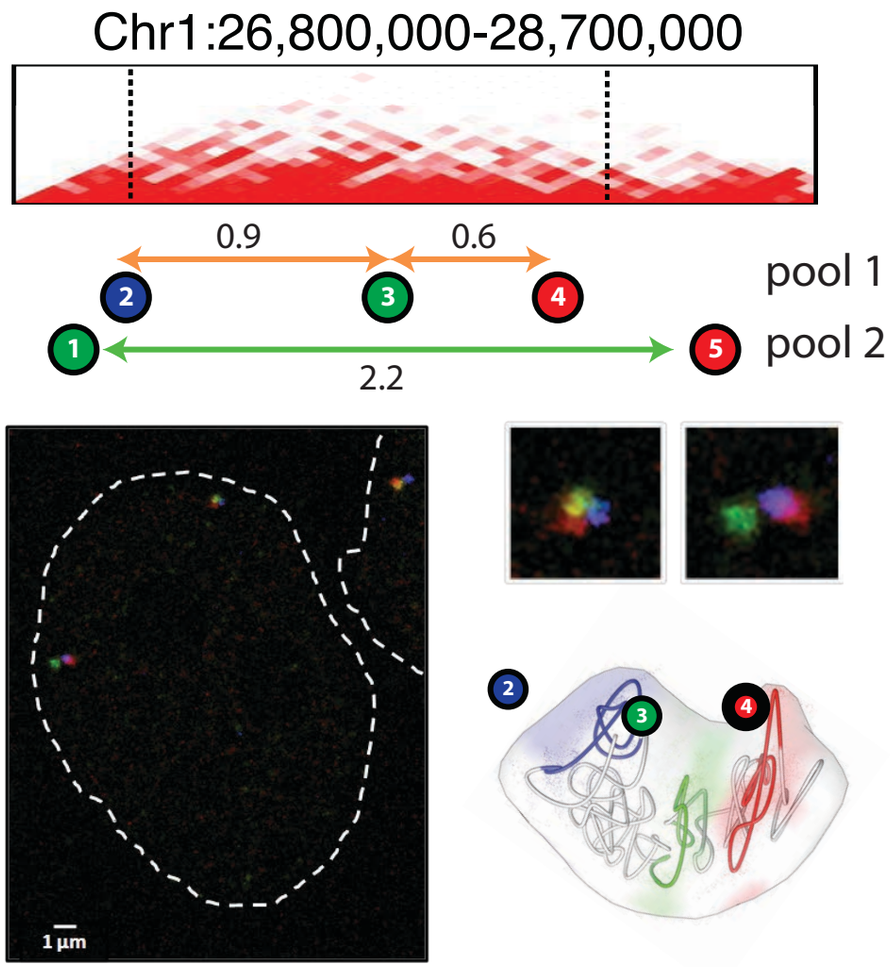


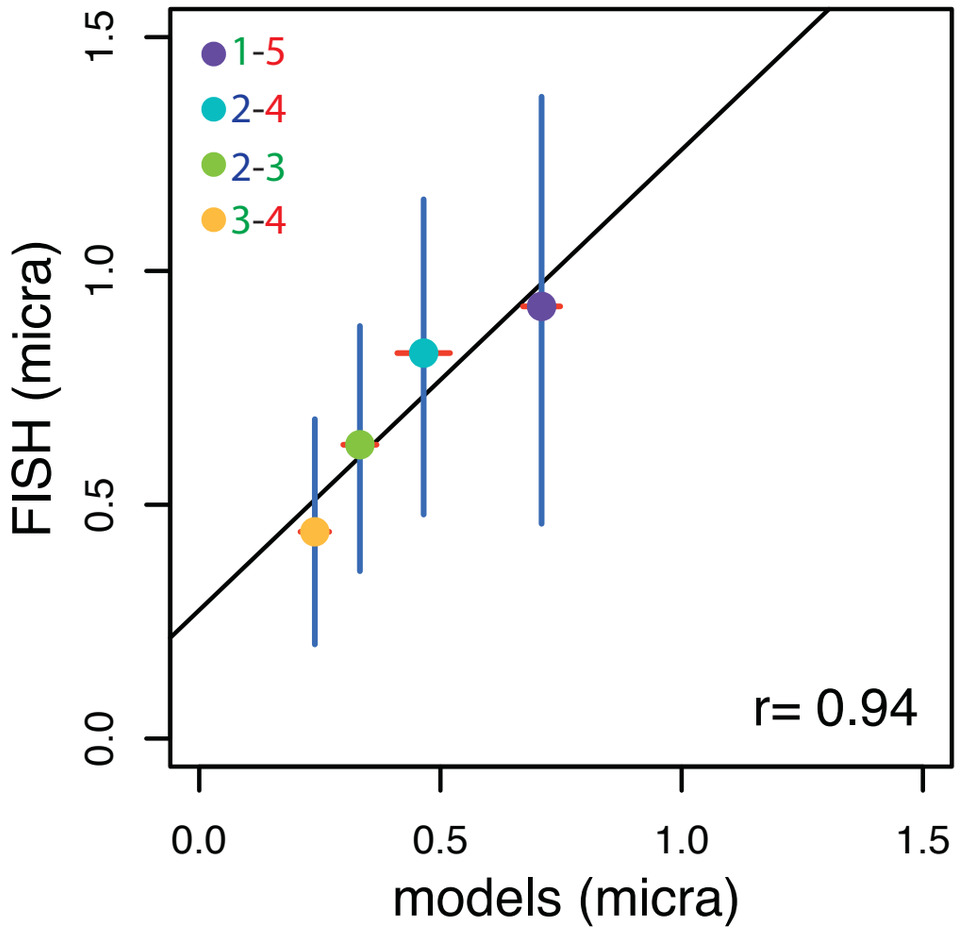
cnag CRG Centre for Genomic Regulation

Modeling 3D TADs



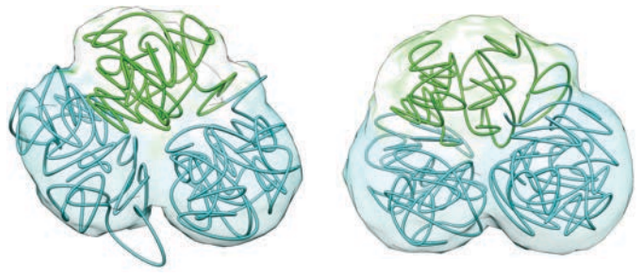
61 genomic regions containing 209 TADs covering 267Mb



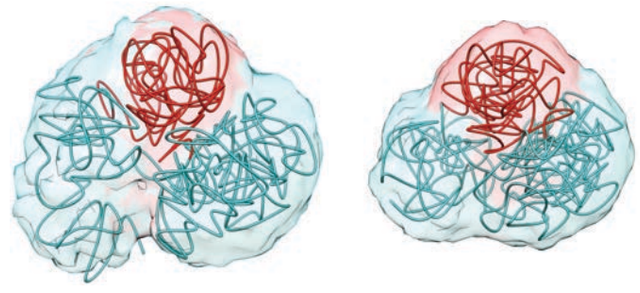


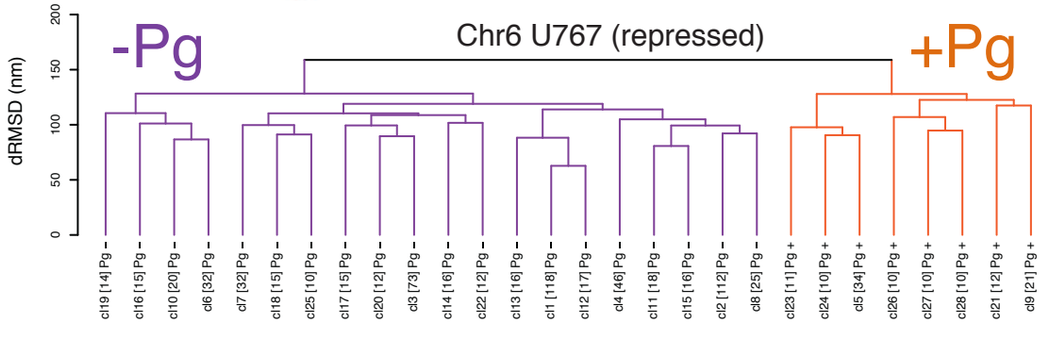
How TADs respond structurally to Pg?

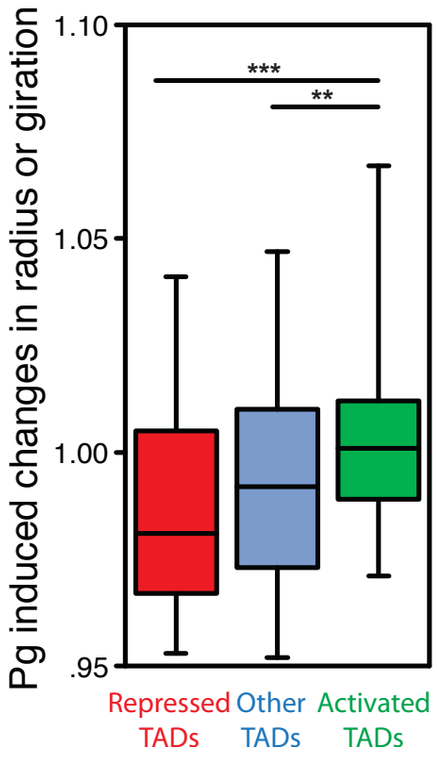
Chr2:9,600,000-13,200,000

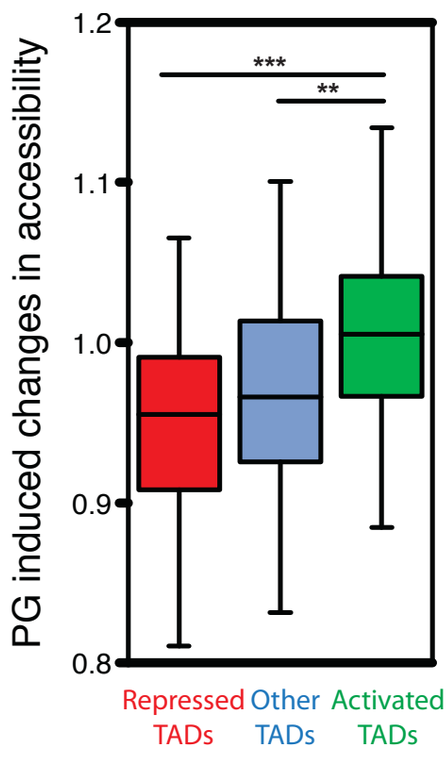


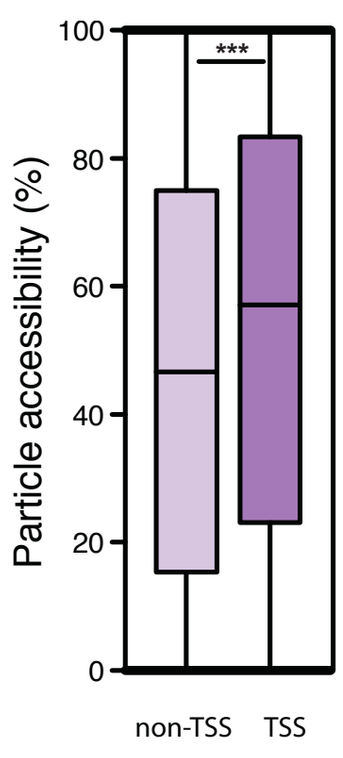
Chr6:71,800,000-76,500,000







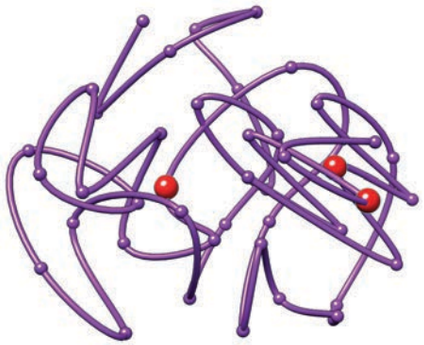


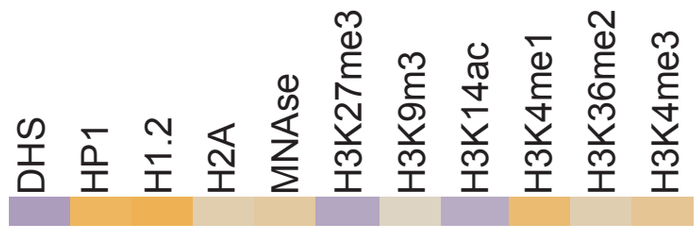


Model for TAD regulation

Repressed TAD

chr1 U41

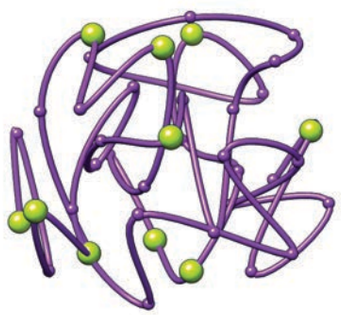


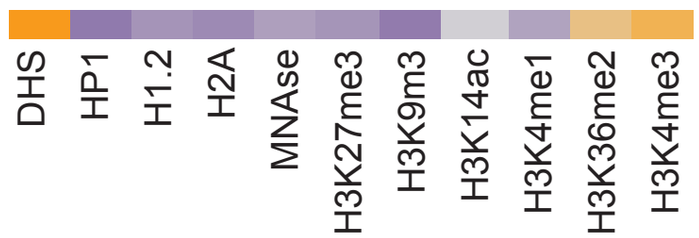


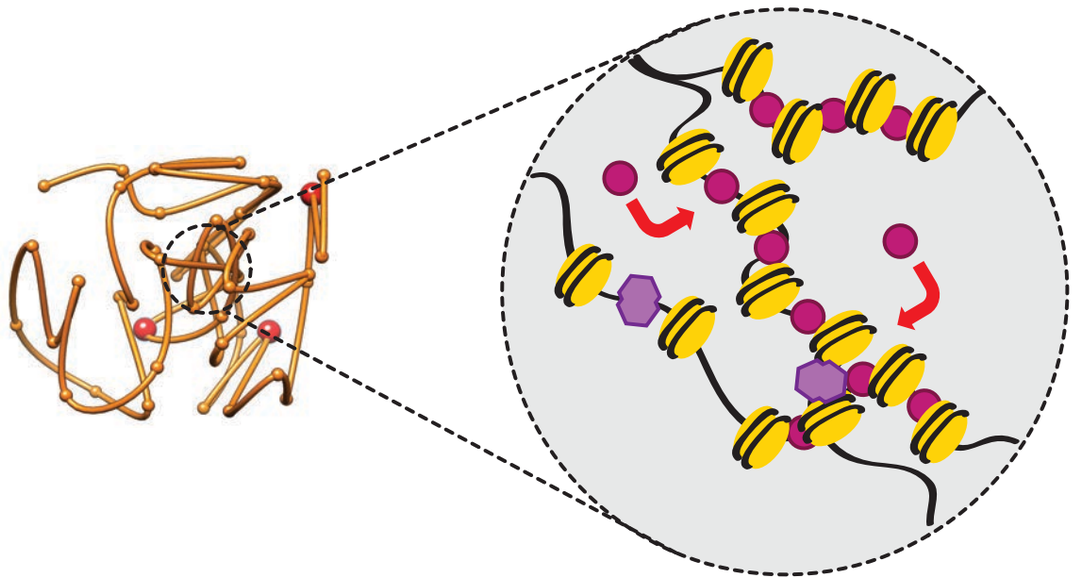
Structural transition **+Pg**

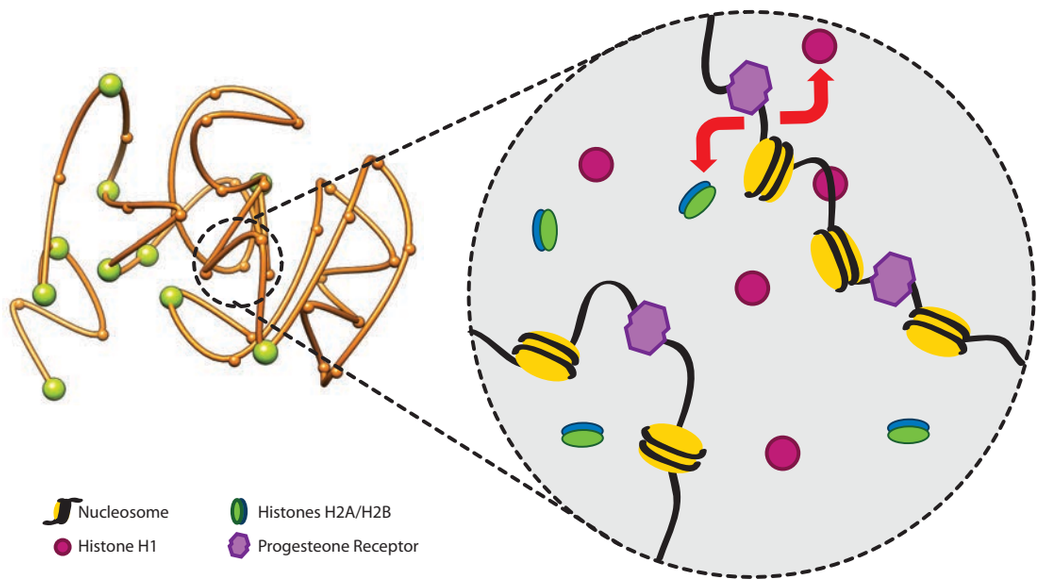
Activated TAD

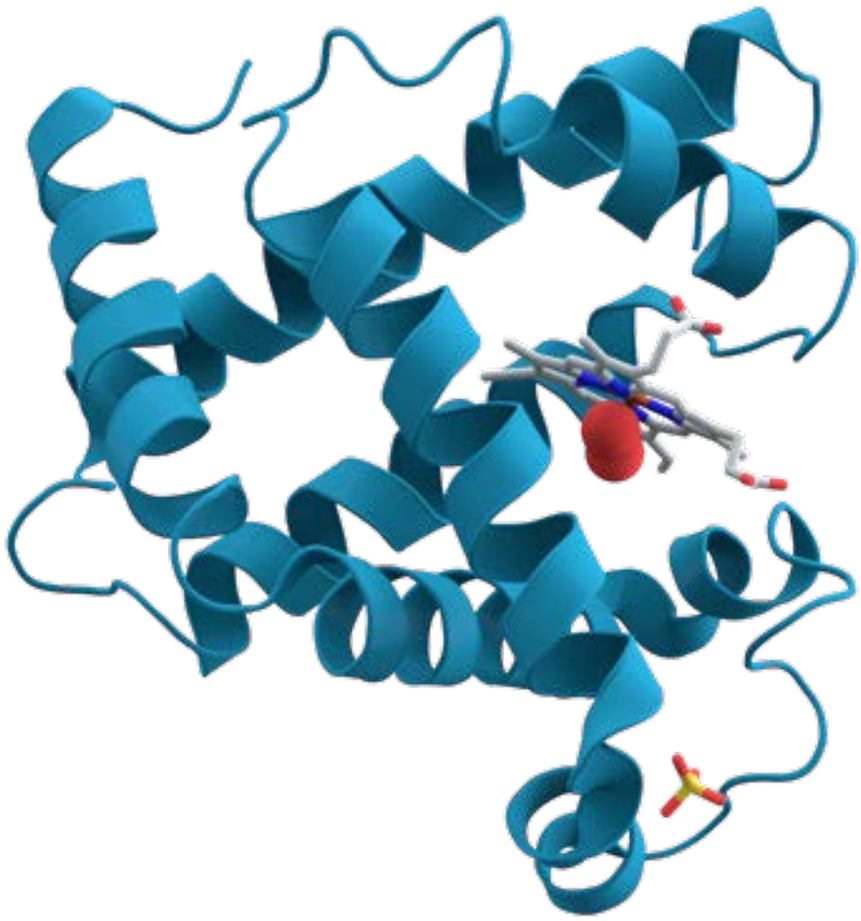
chr2 U207

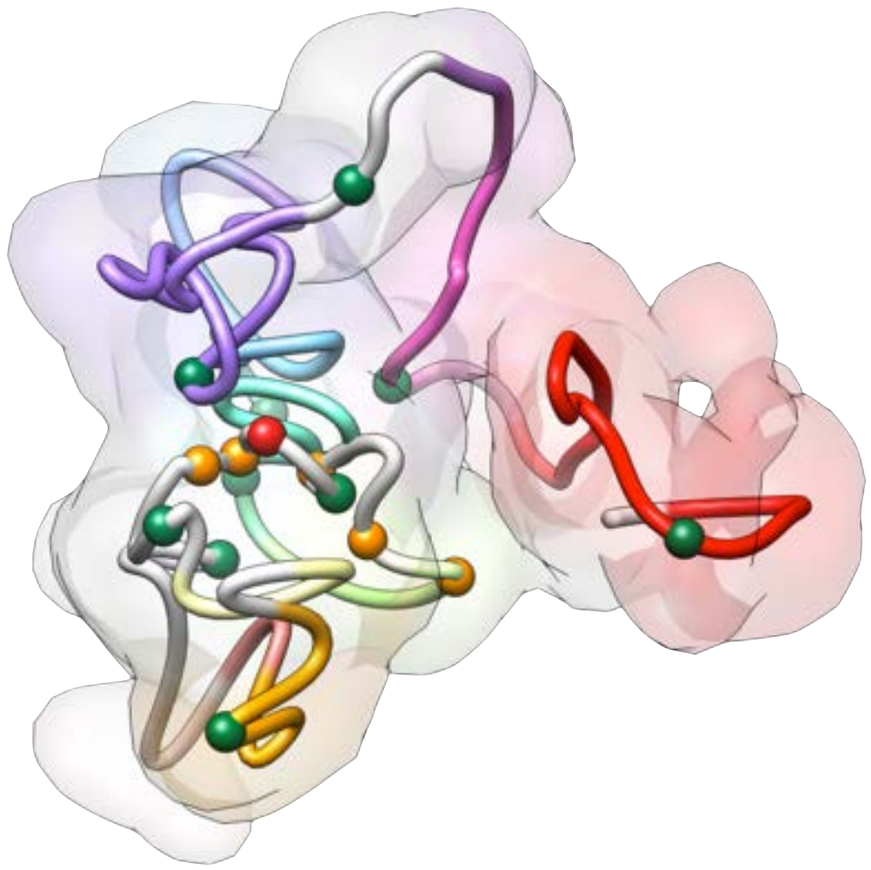












STRUCTURE ⟷ FUNCTION

Open positions
http://marciuslab.org

Acknowledgments

François Serra
Davide Baù
François le Dily

David Dufour
Mike Goodstadt
Gireesh Bogu
Francisco Martínez-Jiménez

erc

Miguel Beato, Thomas Graf and Guillaume Filion

http://marciuslab.org
http://3DGenomes.org
http://cnag.cat · http://crg.cat