RICE | CTBP
Center for Theoretical Biological Physics

Modeling Genome Organization from an "Agnostic" Point of View

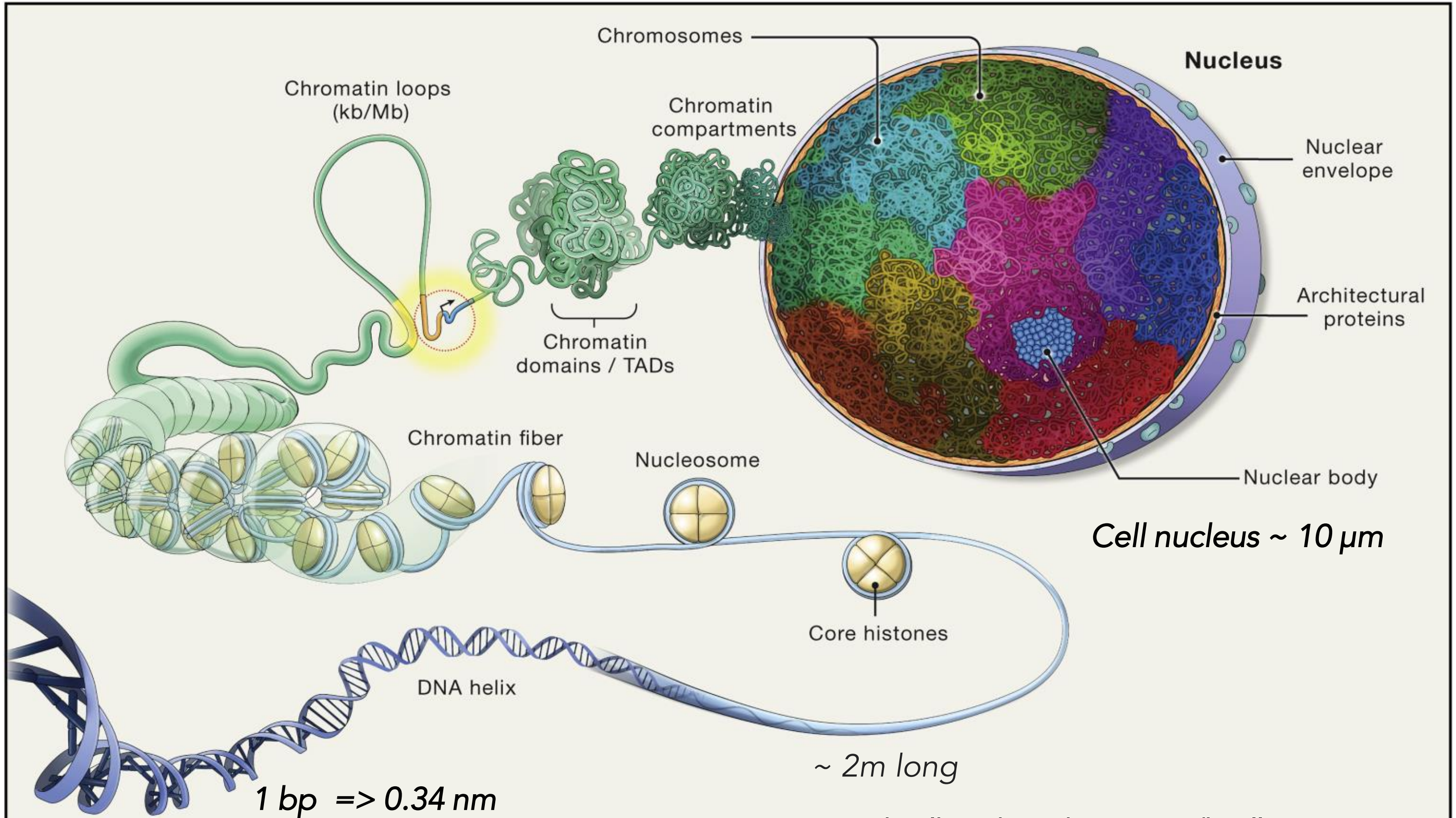
Antonio Bento de Oliveira Junior

Symposium on Current Topics in Molecular Biophysics

THE Welch FOUNDATION

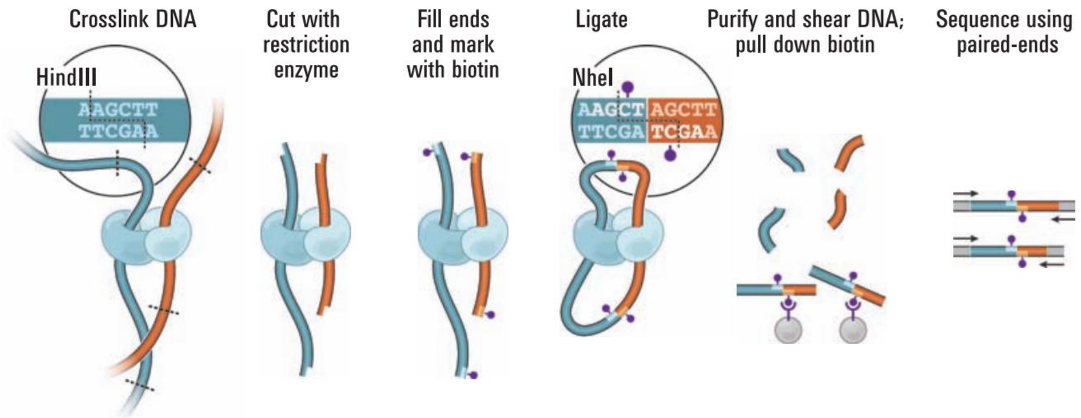
antonio.oliveira@rice.edu

Genome Organization



Misteli, T. (2020). 10.1016/j.cell.2020.09.014

Hi-C – Chromosome "contact map"



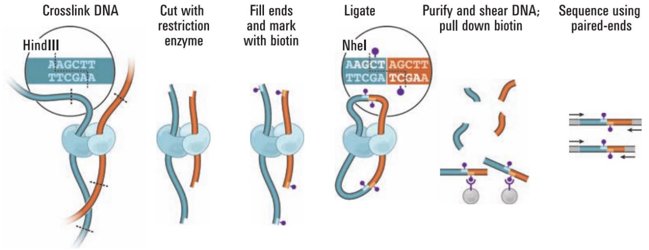
Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome

Erez Lieberman-Aiden,[1,2,3,4]* Nynke L. van Berkum,[5]* Louise Williams,[1] Maxim Imakaev,[2] Tobias Ragoczy,[6,7] Agnes Telling,[6,7] Ido Amit,[1] Bryan R. Lajoie,[5] Peter J. Sabo,[8] Michael O. Dorschner,[8] Richard Sandstrom,[8] Bradley Bernstein,[1,9] M. A. Bender,[10] Mark Groudine,[6,7] Andreas Gnirke,[1] John Stamatoyannopoulos,[8] Leonid A. Mirny,[2,11] Eric S. Lander,[1,12,13]† Job Dekker[5]†

Hi-C – Chromosome "contact map"

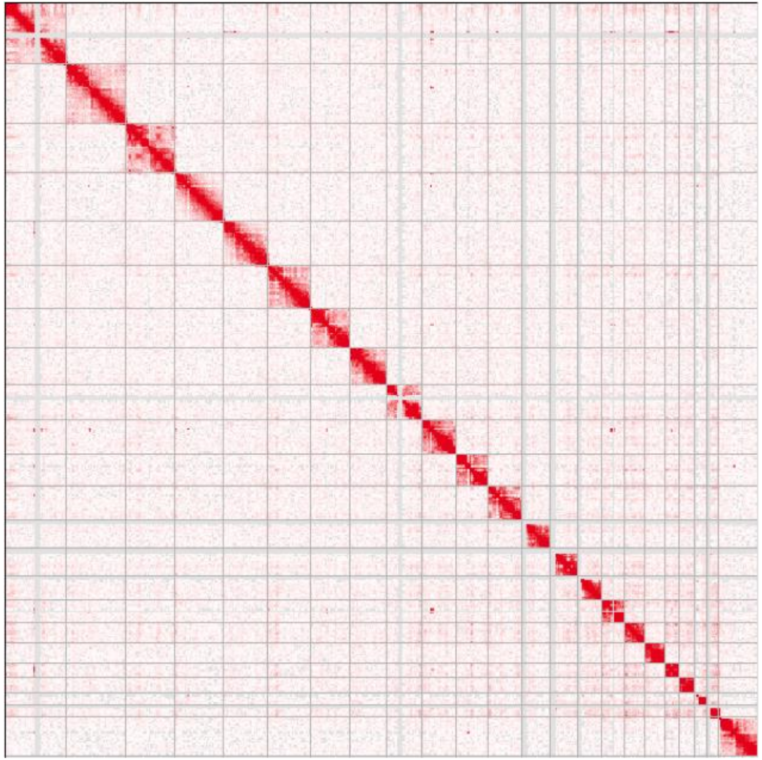
Hi-C

GENERATES GENOME-WIDE CONTACT MAPS



Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome

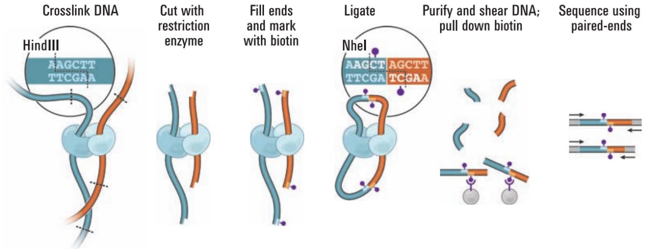
Erez Lieberman-Aiden,[1,2,3,4]* Nynke L. van Berkum,[5]* Louise Williams,[1] Maxim Imakaev,[2] Tobias Ragoczy,[6,7] Agnes Telling,[6,7] Ido Amit,[1] Bryan R. Lajoie,[5] Peter J. Sabo,[8] Michael O. Dorschner,[8] Richard Sandstrom,[8] Bradley Bernstein,[1,9] M. A. Bender,[10] Mark Groudine,[6,7] Andreas Gnirke,[1] John Stamatoyannopoulos,[8] Leonid A. Mirny,[2,11] Eric S. Lander,[1,12,13]† Job Dekker[5]†

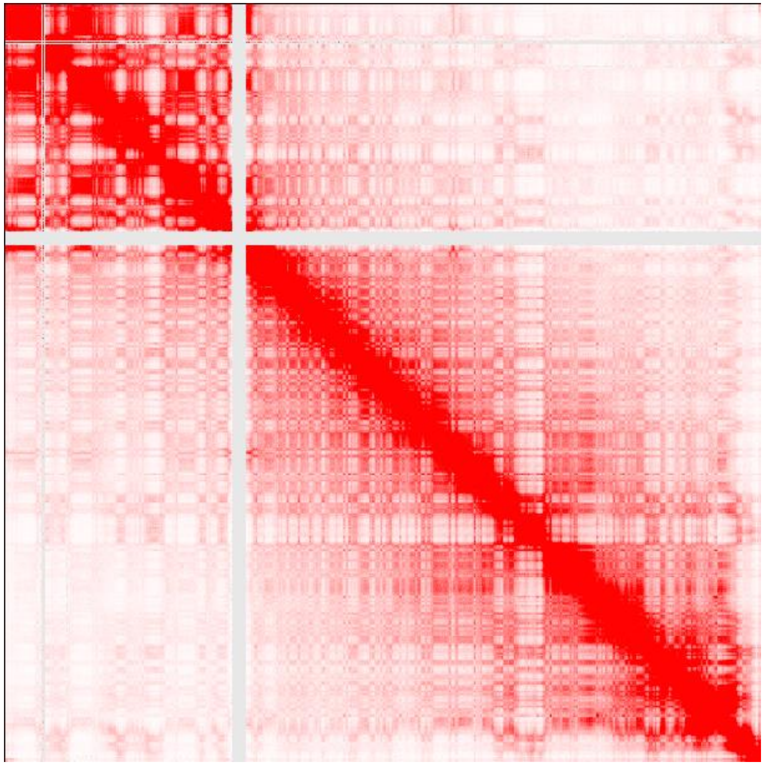


Hi-C – Chromosome "contact map"

Hi-C

GENERATES GENOME-WIDE CONTACT MAPS





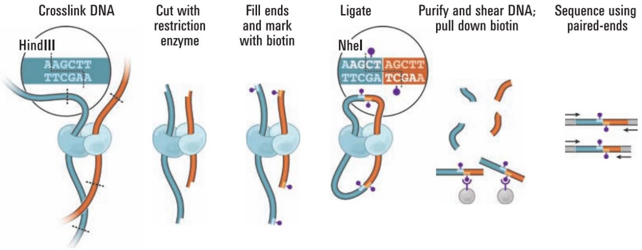
Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome

Erez Lieberman-Aiden,[1,2,3,4]* Nynke L. van Berkum,[5]* Louise Williams,[1] Maxim Imakaev,[2] Tobias Ragoczy,[6,7] Agnes Telling,[6,7] Ido Amit,[1] Bryan R. Lajoie,[5] Peter J. Sabo,[8] Michael O. Dorschner,[8] Richard Sandstrom,[8] Bradley Bernstein,[1,9] M. A. Bender,[10] Mark Groudine,[6,7] Andreas Gnirke,[1] John Stamatoyannopoulos,[8] Leonid A. Mirny,[2,11] Eric S. Lander,[1,12,13]† Job Dekker[5]†

Hi-C – Chromosome "contact map"

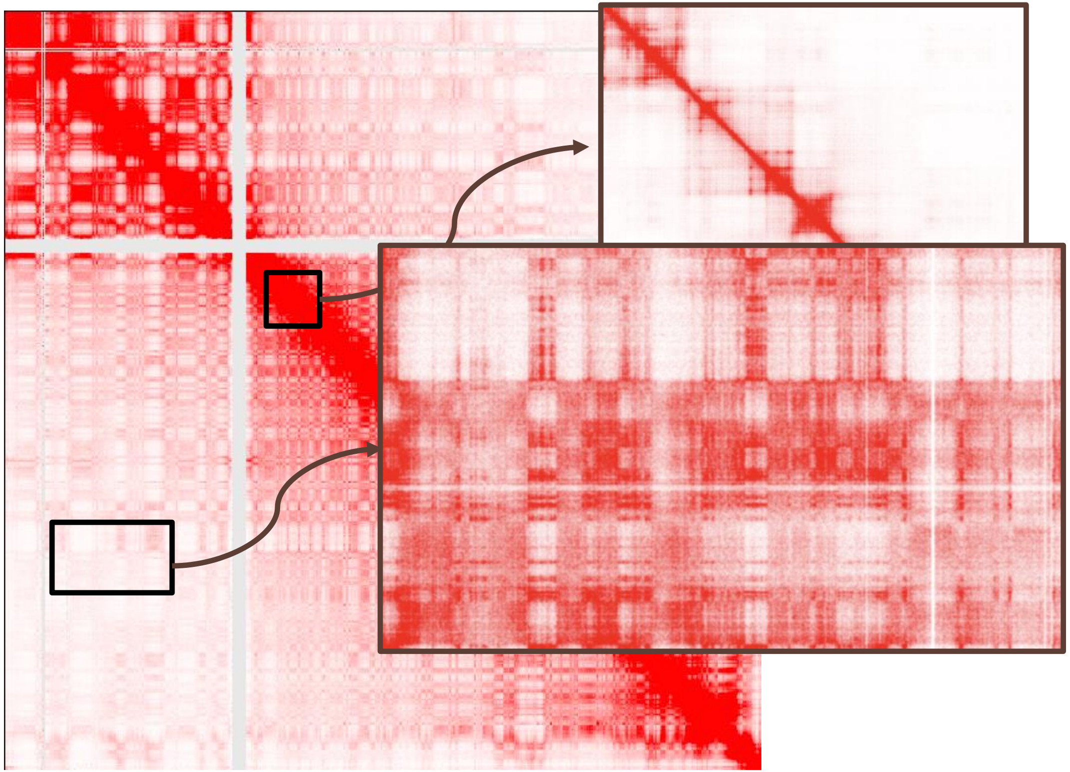
Hi-C

GENERATES GENOME-WIDE CONTACT MAPS



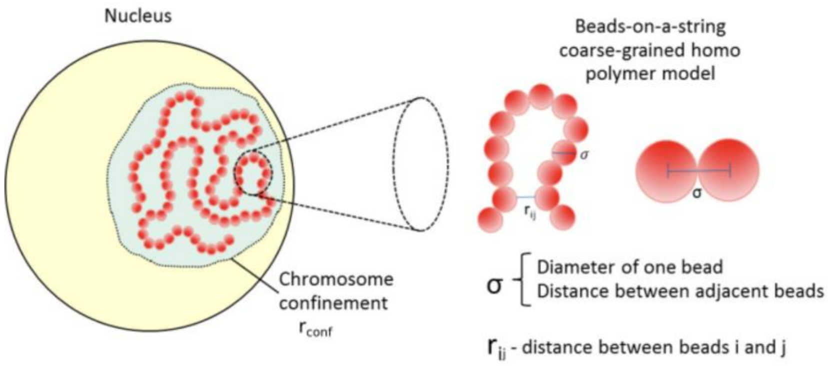
Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome

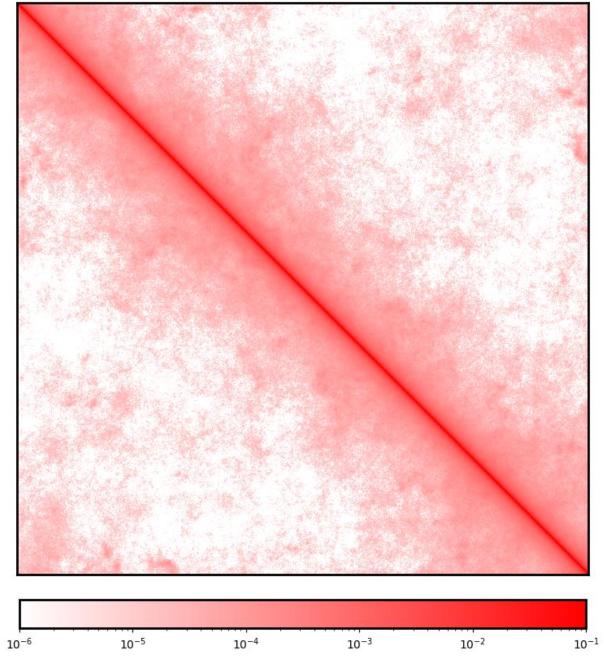
Erez Lieberman-Aiden,[1,2,3,4*] Nynke L. van Berkum,[5*] Louise Williams,[1] Maxim Imakaev,[2] Tobias Ragoczy,[6,7] Agnes Telling,[6,7] Ido Amit,[1] Bryan R. Lajoie,[5] Peter J. Sabo,[8] Michael O. Dorschner,[8] Richard Sandstrom,[8] Bradley Bernstein,[1,9] M. A. Bender,[10] Mark Groudine,[6,7] Andreas Gnirke,[1] John Stamatoyannopoulos,[8] Leonid A. Mirny,[2,11] Eric S. Lander,[1,12,13†] Job Dekker[5†]



Chromosome modeling – Top-down approach

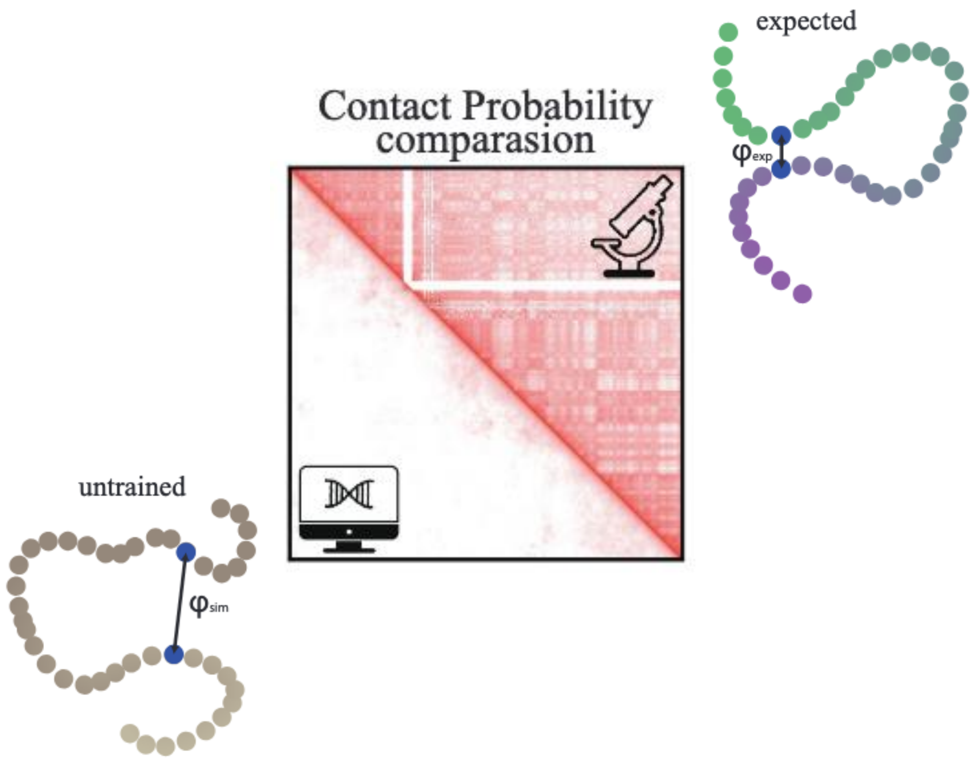
Homopolymer model





$$U_{HP}(\vec{r}) = \sum_{i\in\{\text{Loci}\}} U_{\text{FENE}}(\vec{r}_{i,i+1}) + \sum_{i\in\{\text{Loci}\}} U_{hc}(\vec{r}_{i,i+1}) + \sum_{i\in\{\text{Angles}\}} U_{\text{Angle}}(\theta_i) + \sum_{\substack{i,j\in\{\text{Loci}\}\\ j>i+2}} U_{sc}(\vec{r}_{i,j}) + \sum_{i\in\{\text{Loci}\}} U_c(\vec{r}_i),$$

Chromosome modeling – Top-down approach



$$U_{HP}(\vec{r}) = \sum_{i\in\{\text{Loci}\}} U_{\text{FENE}}(\vec{r}_{i,i+1}) + \sum_{i\in\{\text{Loci}\}} U_{hc}(\vec{r}_{i,i+1}) + \sum_{i\in\{\text{Angles}\}} U_{\text{Angle}}(\theta_i) + \sum_{\substack{i,j\in\{\text{Loci}\}\\ j>i+2}} U_{sc}(\vec{r}_{i,j}) + \sum_{i\in\{\text{Loci}\}} U_c(\vec{r}_i),$$

Maximum Entropy Approach

$$U(\vec{r}) = U_{HP}(\vec{r}) + \sum_{ij} \lambda_{ij} f(r_{ij})$$

Observables

Chromosome modeling – Top-down approach

Maximum Entropy Approach

$$U(\vec{r}) = U_{HP}(\vec{r}) + \sum_{ij} \lambda_{ij} f(r_{ij})$$

Observables

Physics Assumptions
MiChroM Potential

$$U_{MiChroM}(\vec{r}) = U_{HP}(\vec{r}) + \sum_{\substack{k \geq l \\ k,l \in \text{Types}}} \alpha_{kl} \sum_{\substack{i \in \{\text{Loci of Type } k\} \\ j \in \{\text{Loci of Type } l\}}} f(r_{ij})$$

$$+ \chi \sum_{(i,j) \in \{\text{Loop Sites}\}} f(r_{ij}) + \sum_{d=3}^{d_{cutoff}} \gamma(d) \sum_i f(r_{i,i+d}),$$

Di Pierro, M., Zhang, B., Aiden, E. L., Wolynes, P. G., & Onuchic, J. N. (2016). Proc. Natl. Acad. Sci

Agnostic Potential

$$U(\vec{r}) = U_{HP}(\vec{r}) + \sum_{ij} \lambda_{ij} f(r_{ij})$$

Zhang, B., & Wolynes, P. G. (2015). Proc. Natl. Acad. Sci.

Chromosome modeling – Top-down approach

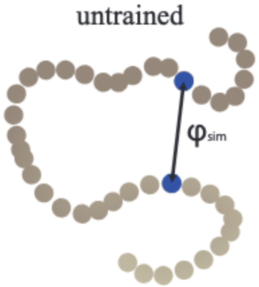
Maximum Entropy Approach

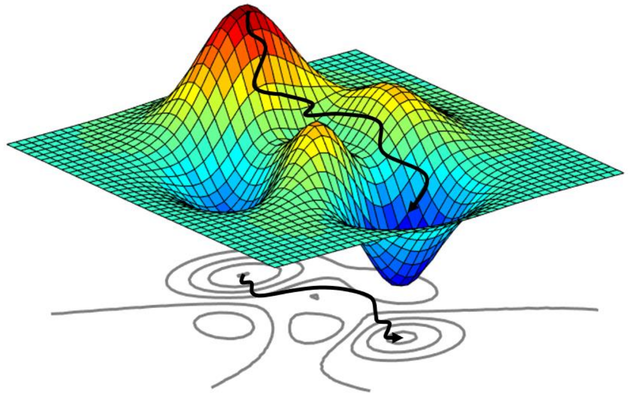
$$U(\vec{r}) = U_{HP}(\vec{r}) + \sum_{ij} \lambda_{ij} f(r_{ij})$$

Observables

$$f(r_{i,j}) = \frac{1}{2}\left(1 + \tanh\left[\mu(r_c - r_{i,j})\right]\right)$$

Parameters λ_{ij} optimization



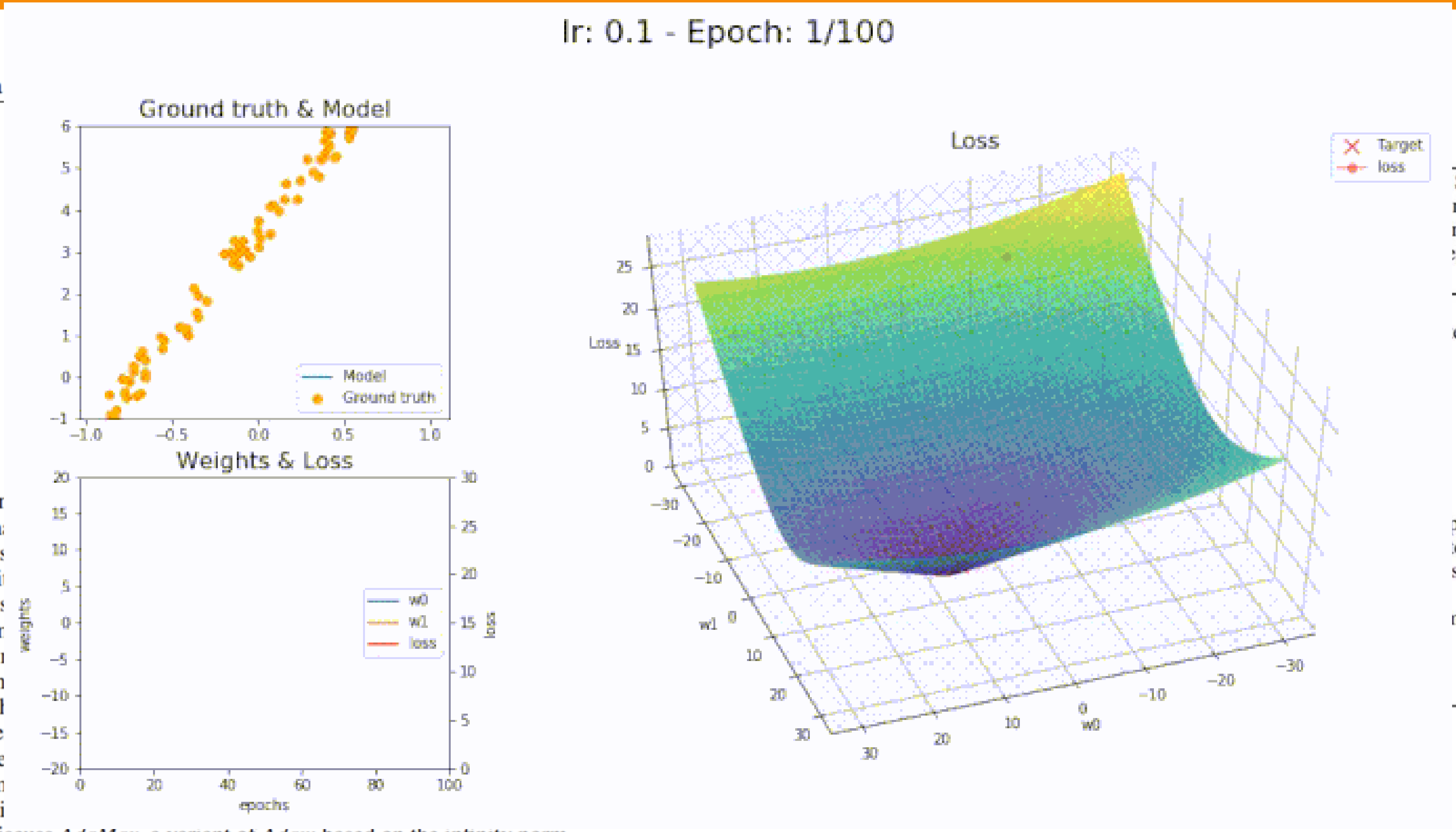


Zhang, B., & Wolynes, P. G. (2015). Proc. Natl. Acad. Sci.

First-Order Direct Inversion

Published as a

ADAM:



See section 2 for details,
ndicates the elementwise
roblems are $\alpha = 0.001$,
nt-wise. With β_1^t and β_2^t

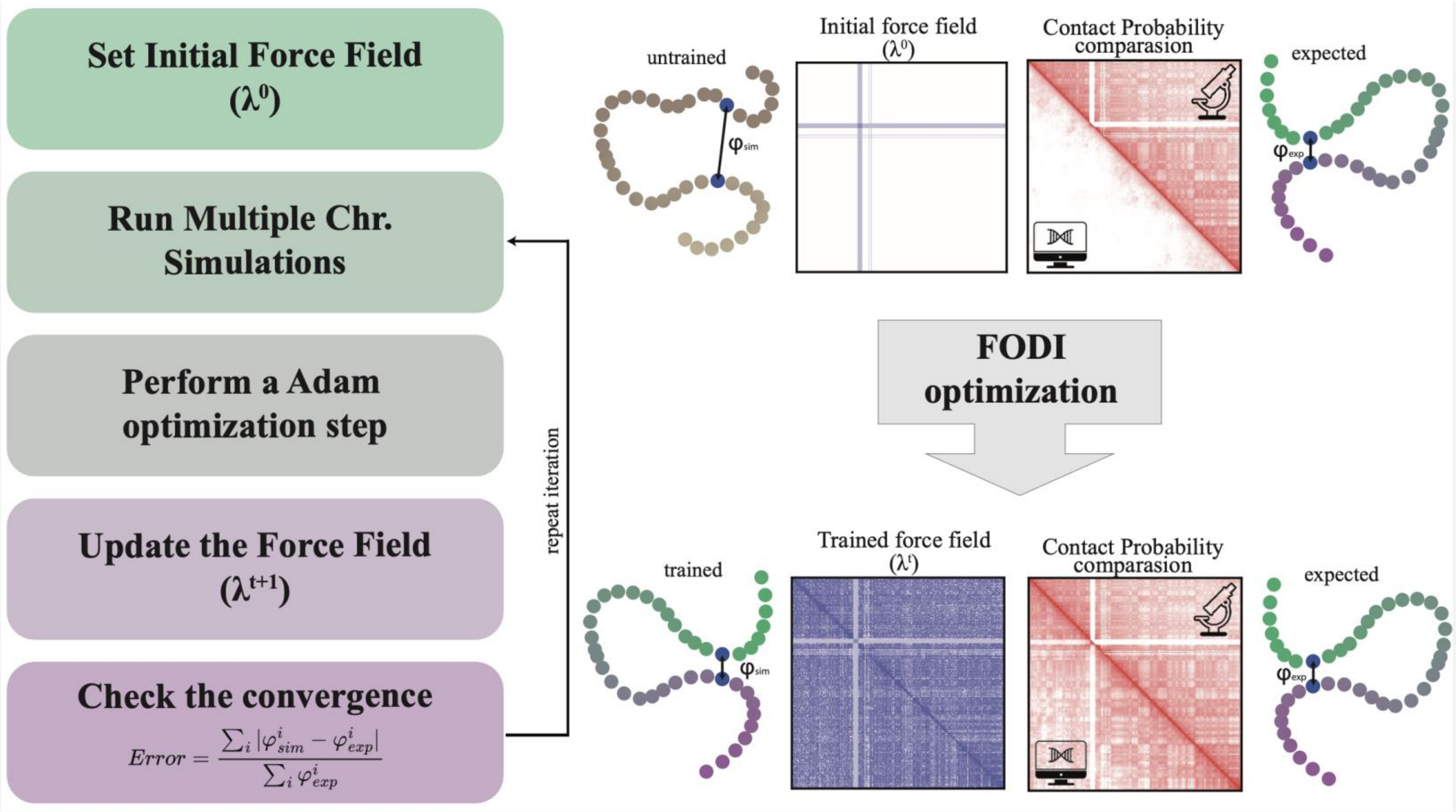
es

p t)
e)
stimate)

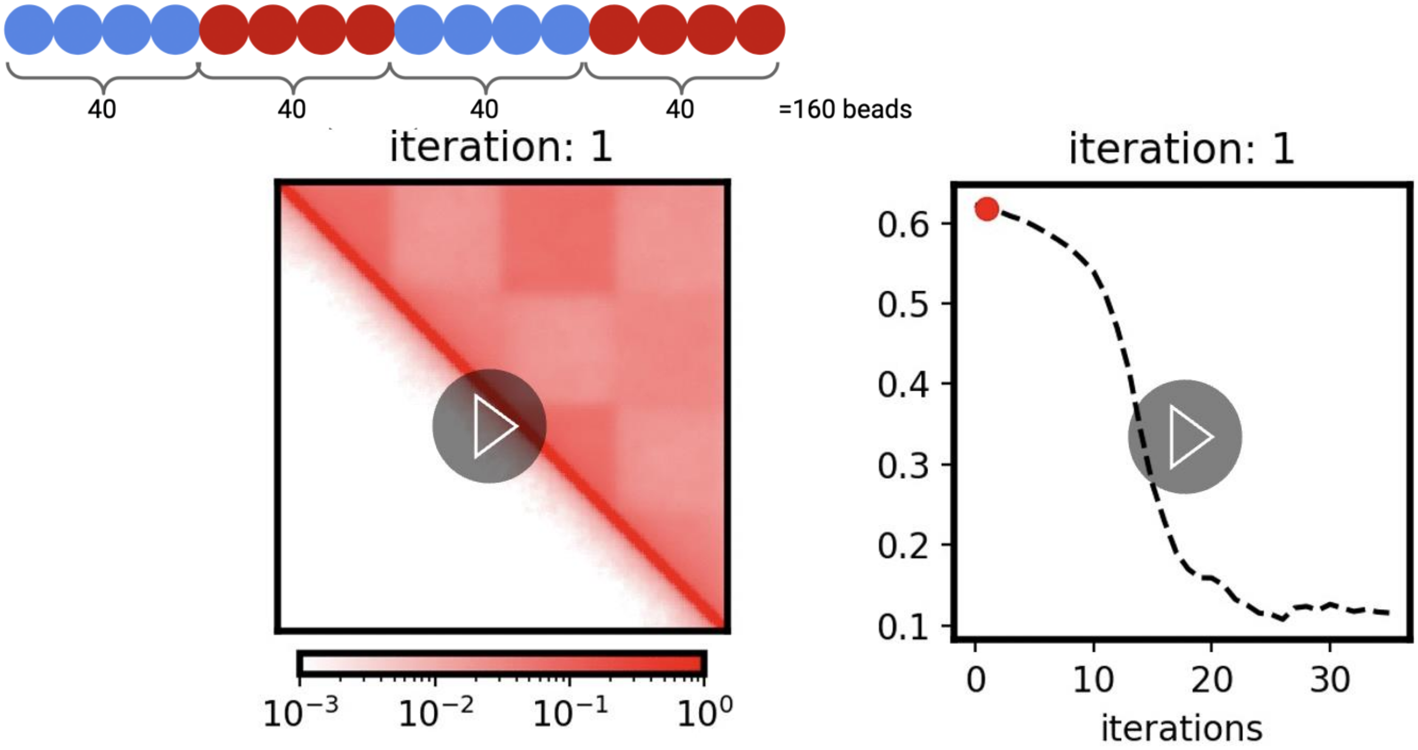
nate)

We i
stoch
ments
has li
and i
The r
very
tation
on wh
verge
gence
optim
practi
we discuss *AdaMax*, a variant of *Adam* based on the infinity norm.

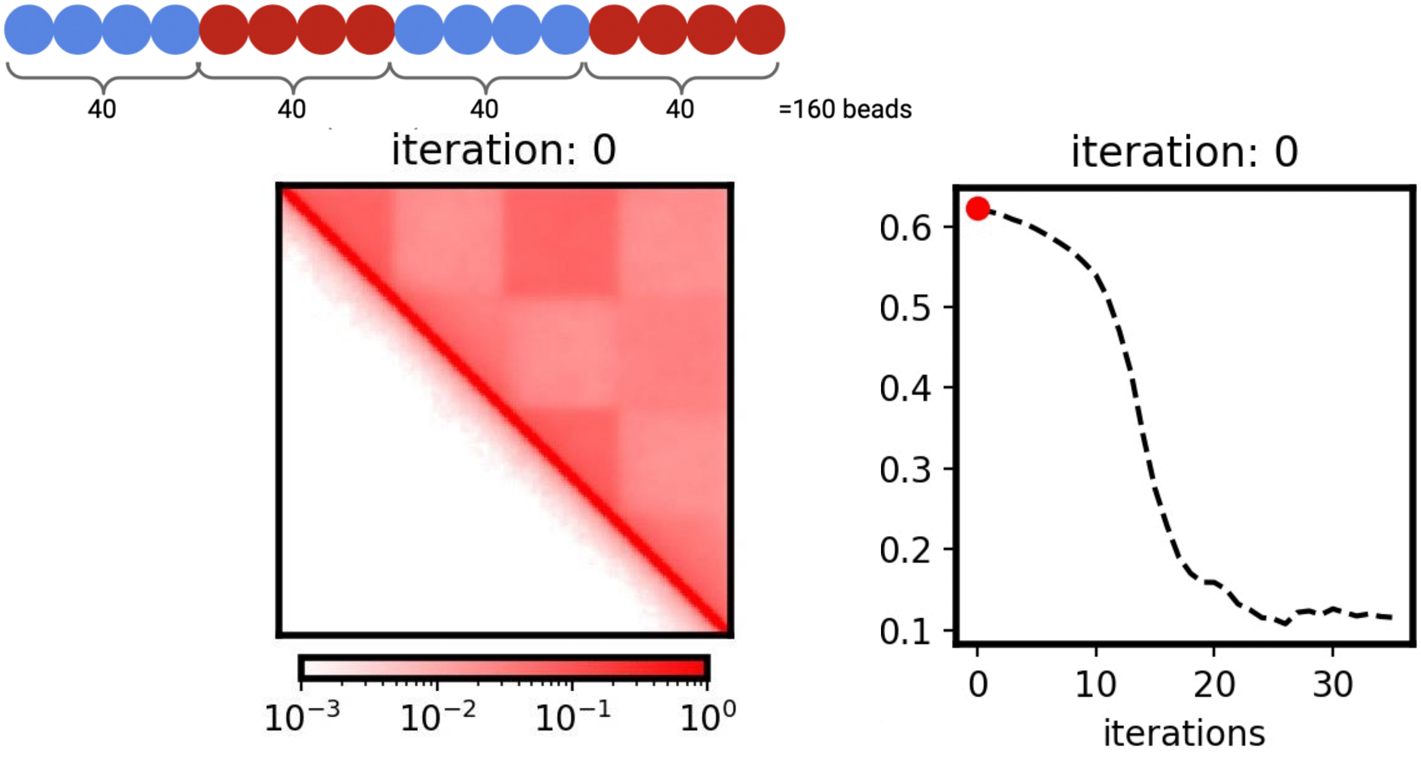
First-Order Direct Inversion - Protocol



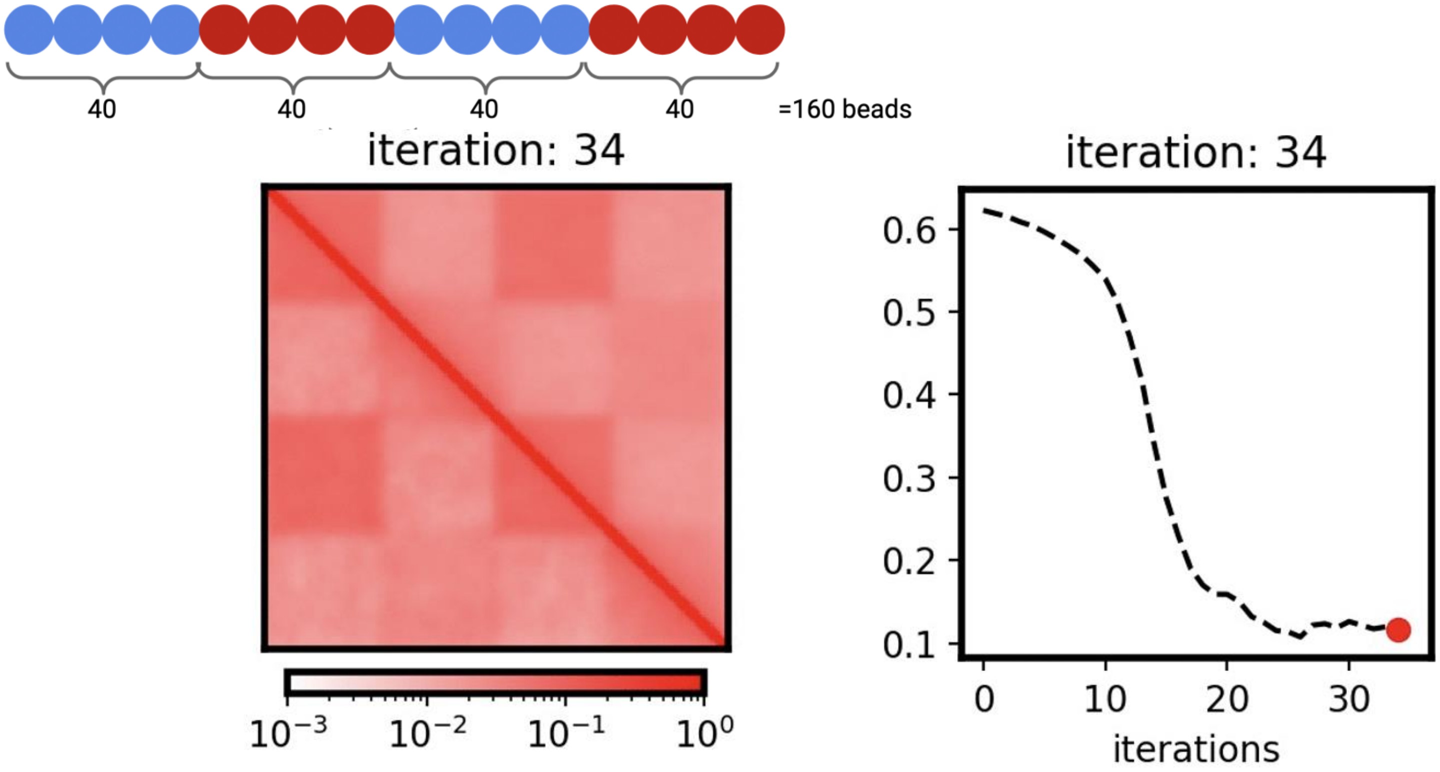
First-Order Direct Inversion - Toy model



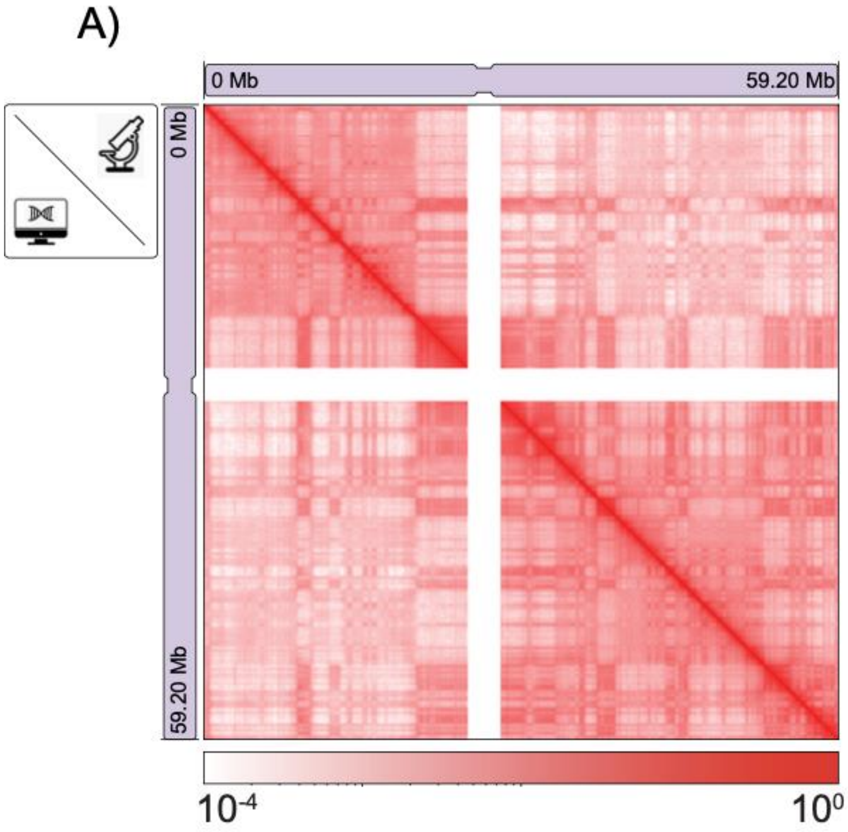
First-Order Direct Inversion - Toy model



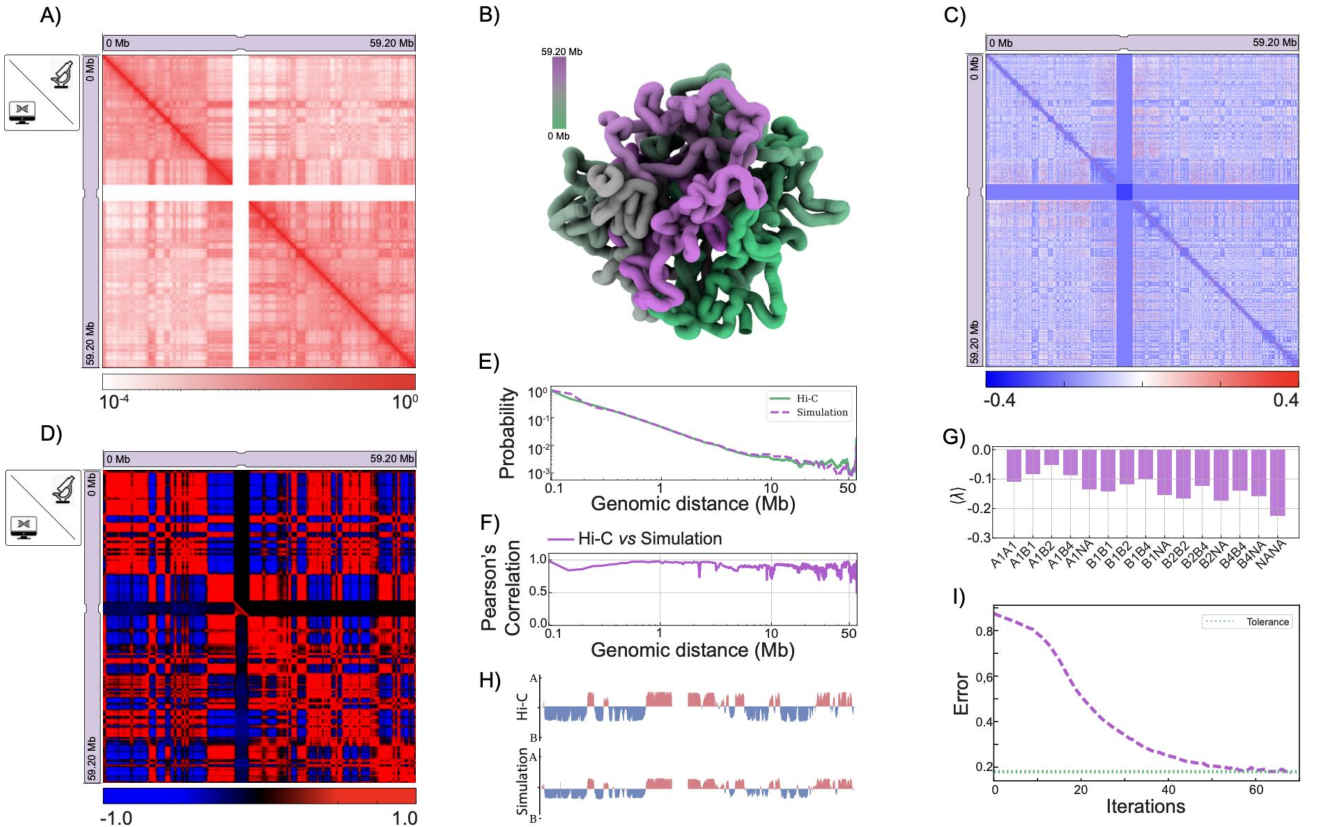
First-Order Direct Inversion - Toy model



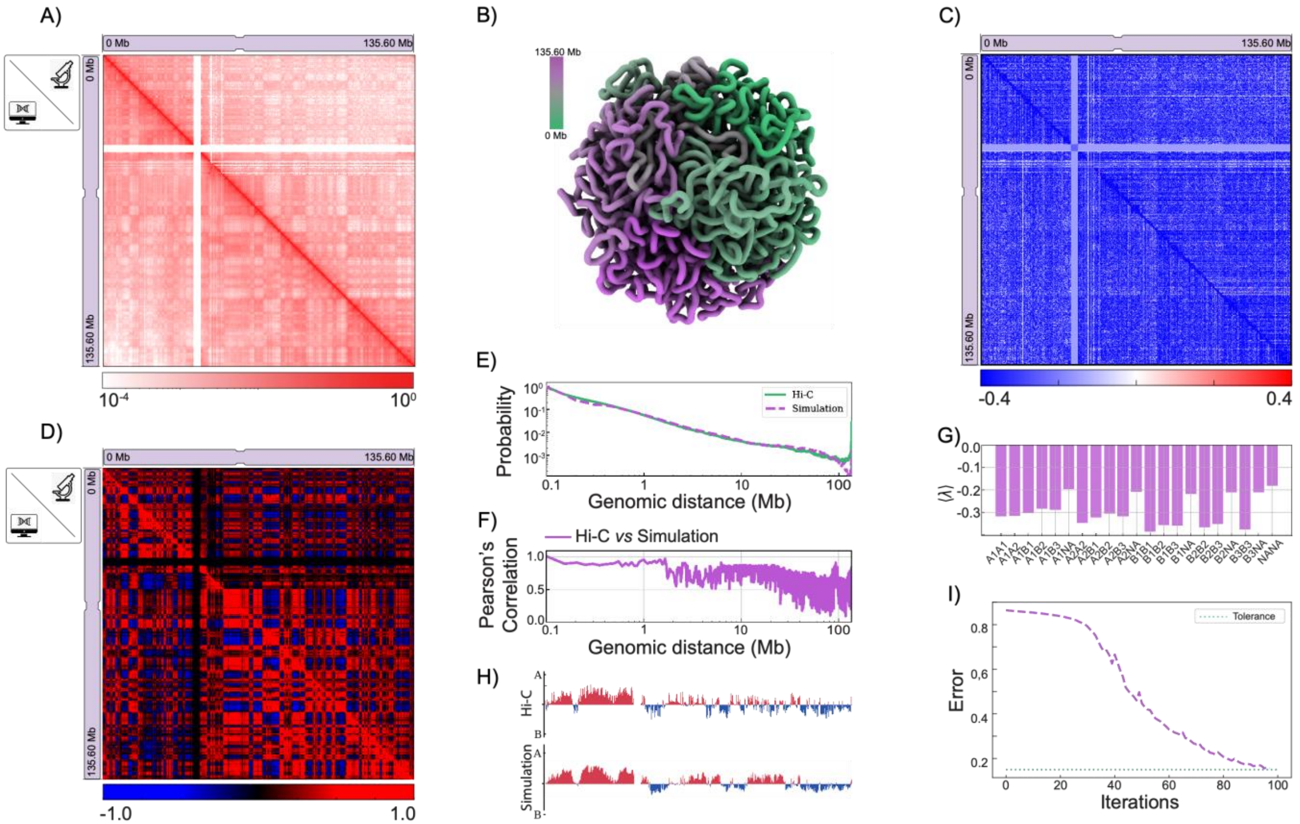
First-Order Direct Inversion – chr19



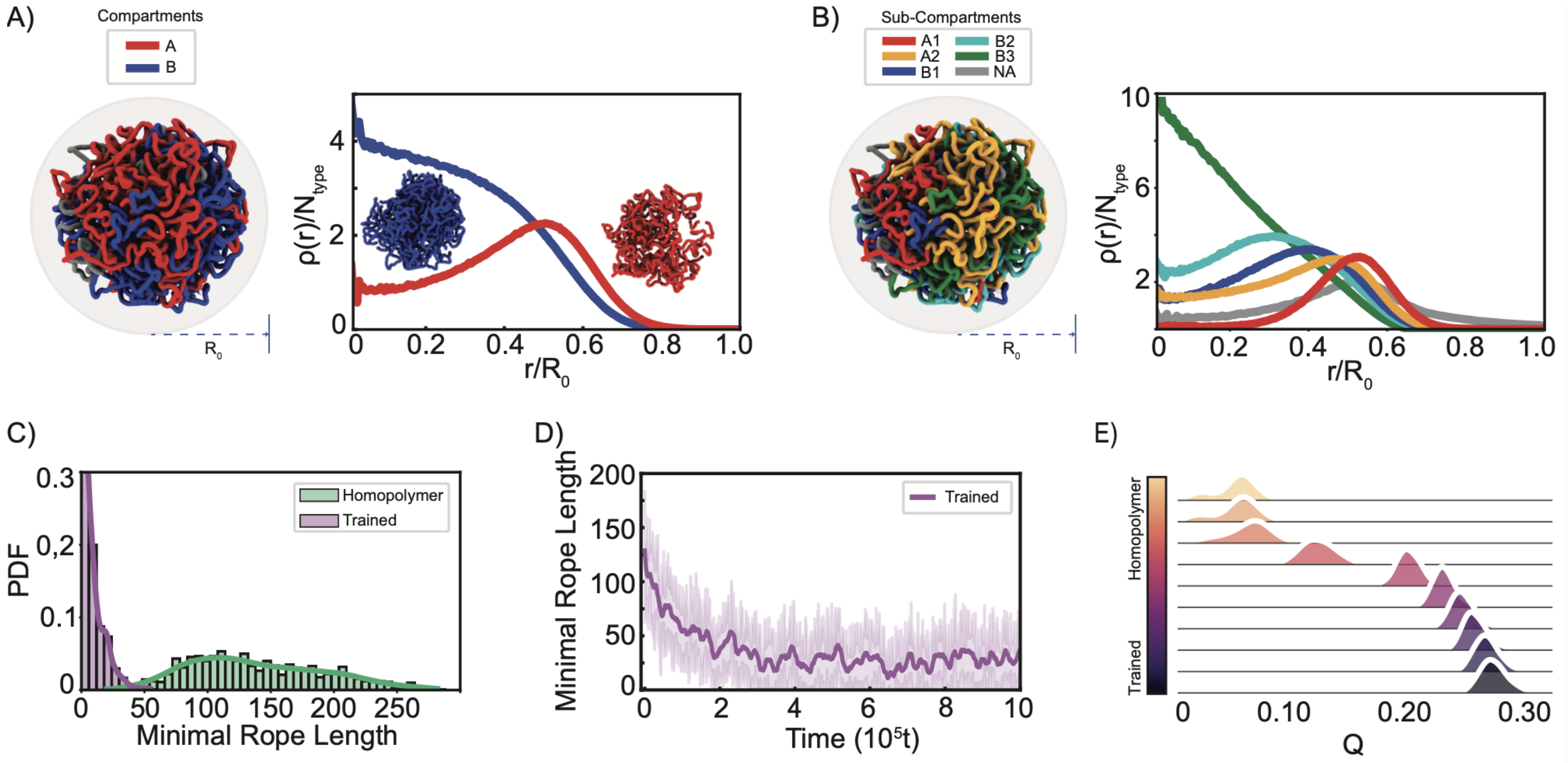
First-Order Direct Inversion – chr19



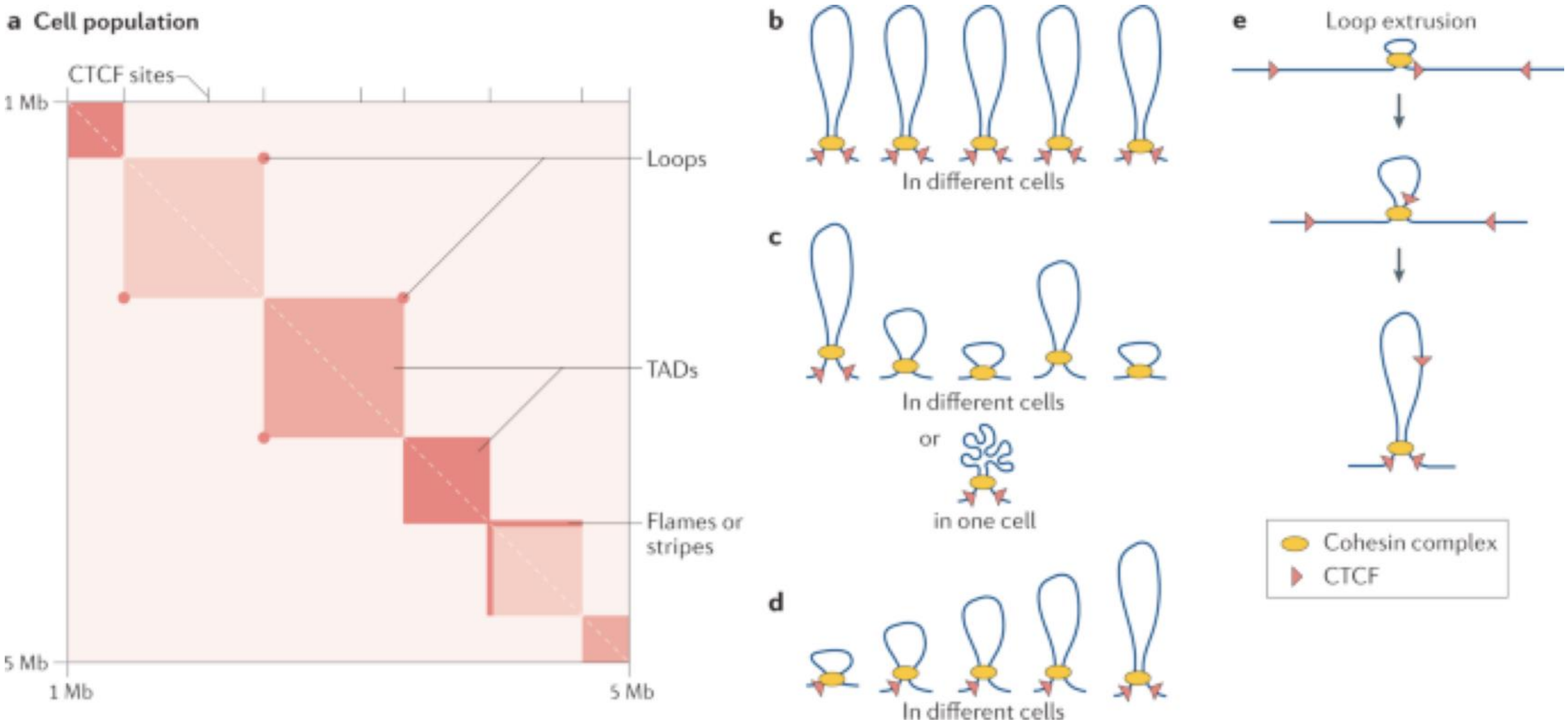
First-Order Direct Inversion – chr10



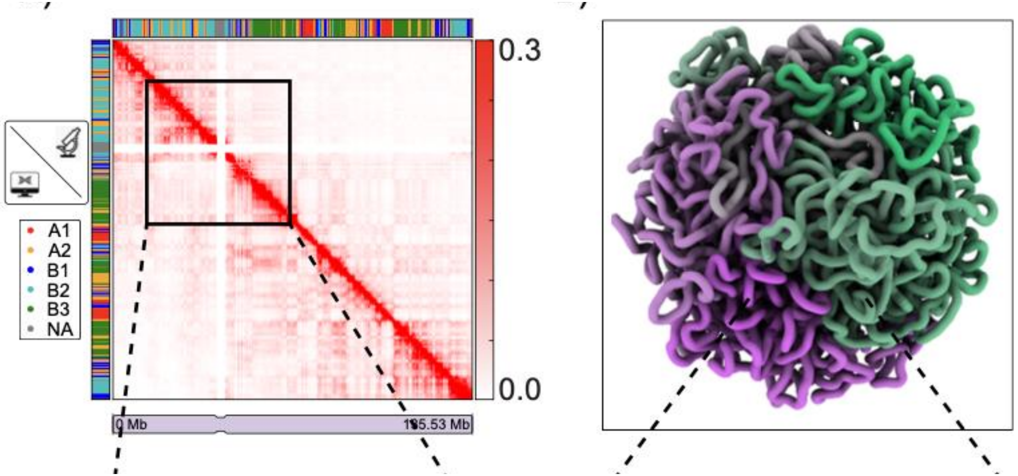
First-Order Direct Inversion – chr10



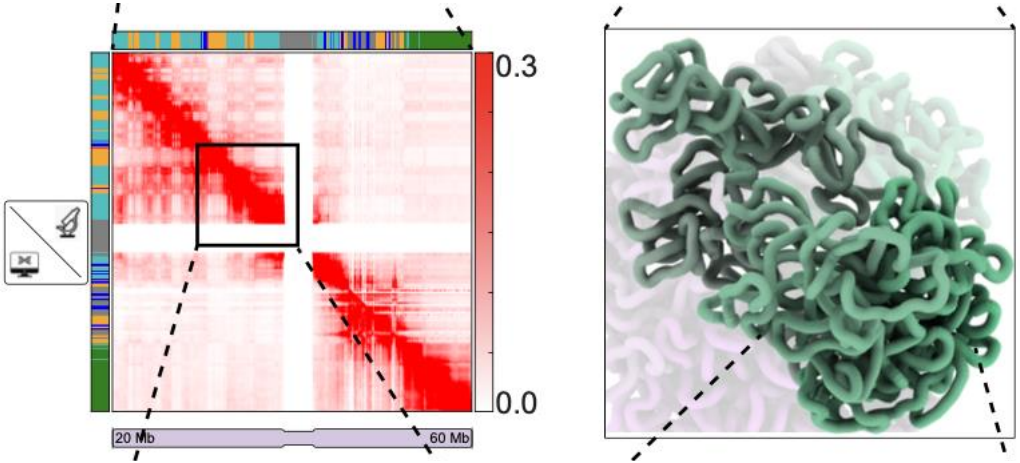
First-Order Direct Inversion



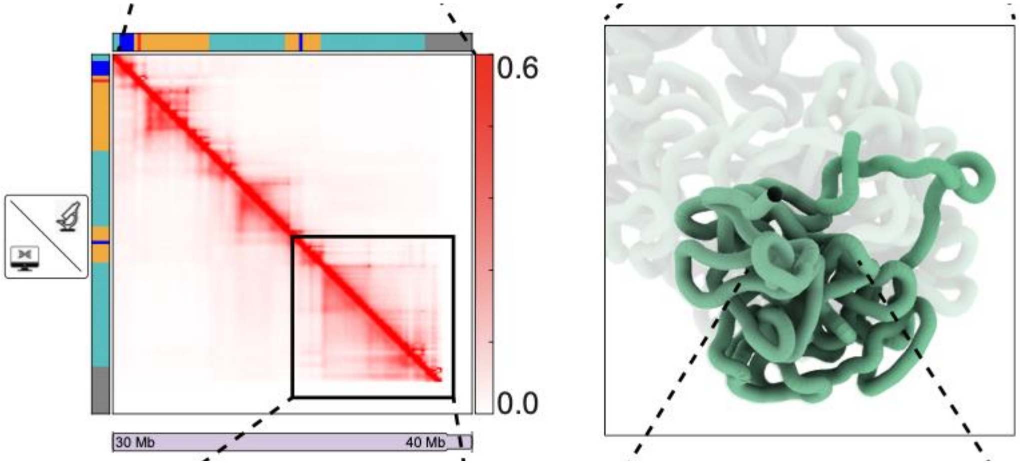
First-Order Direct Inversion



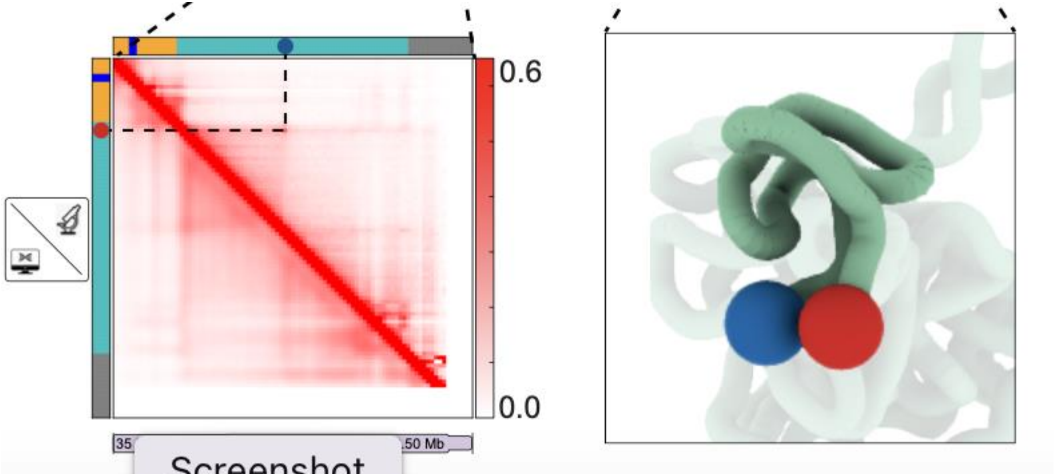
First-Order Direct Inversion



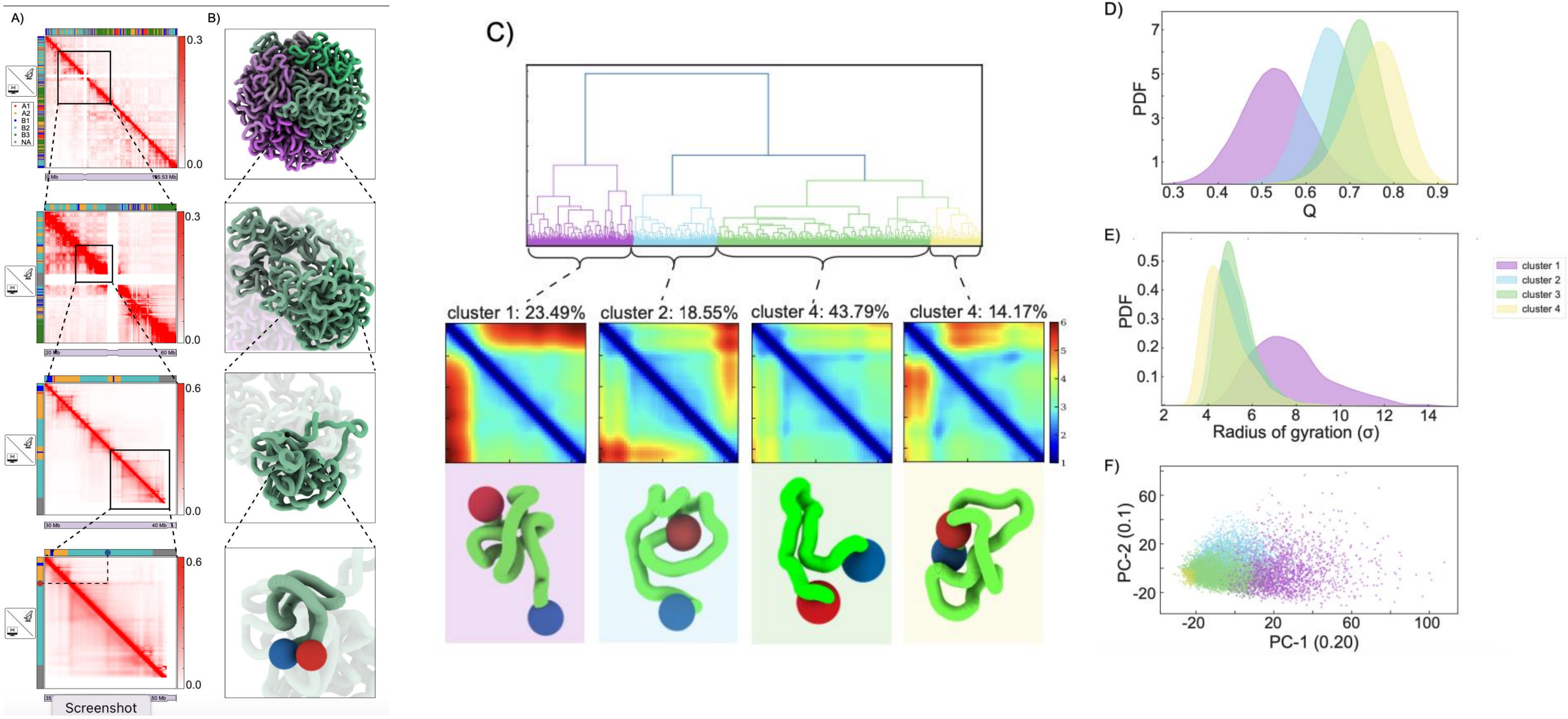
First-Order Direct Inversion



First-Order Direct Inversion



First-Order Direct Inversion



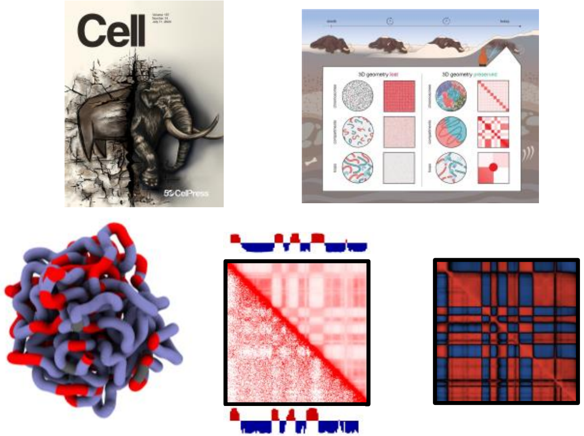
First-Order Direct Inversion – Projects

OpenMiChroM

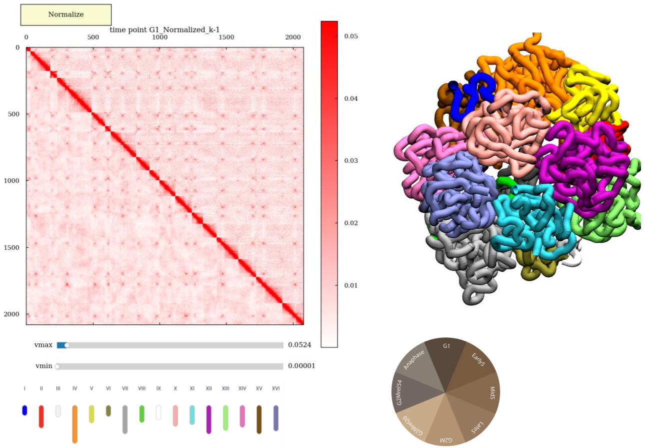
Chromatin Dynamics Software

Modeling the Wooly Mammoth Genome

Sandoval-Velasco, Dudchenko, O... Aiden, E. L. (2024). Cell, 187(14), 3541–3562.e51.

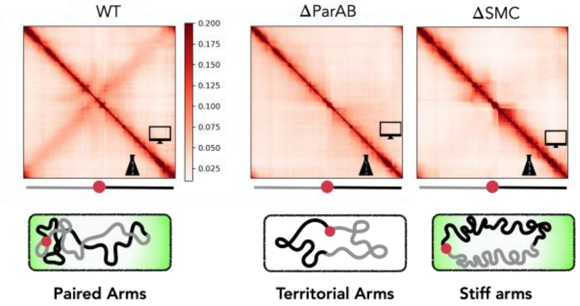


Yeast Genome cell cycle



Bacterial Chromosome Replication

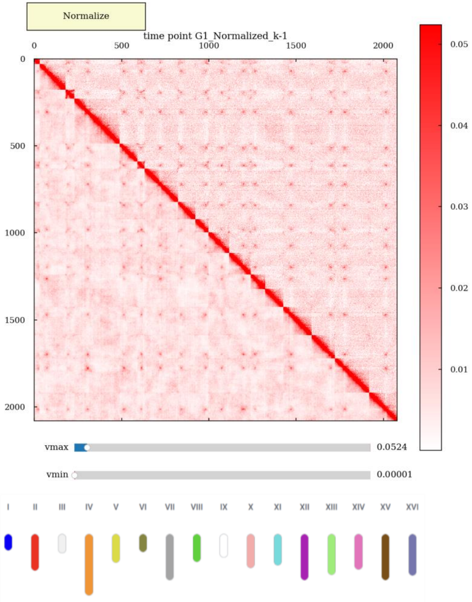
Brahmachari, S., Oliveira, A. B., Onuchic, J. N. (2024). Compaction-mediated segregation of partly replicated bacterial chromosome. bioRxiv, 2024.07.27.604869.



First-Order Direct Inversion – Projects

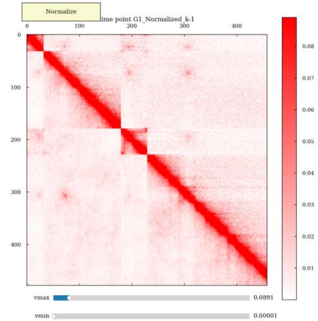
OpenMiChroM

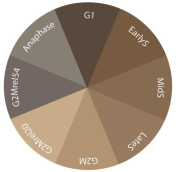
Chromatin Dynamics Software

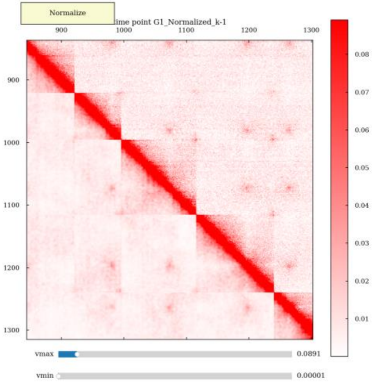


Yeast Genome

cell cycle

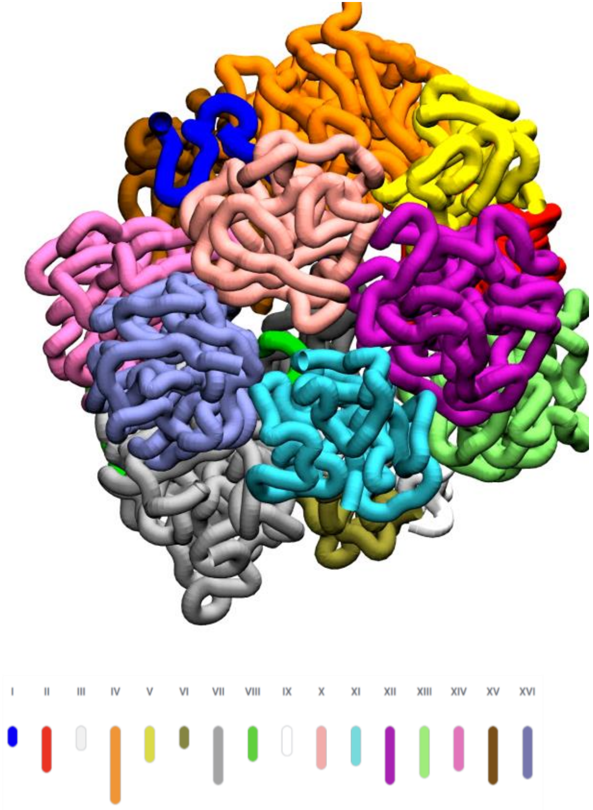






Vittore Scolari

Yeast Genome cell cycle - UNBIASED

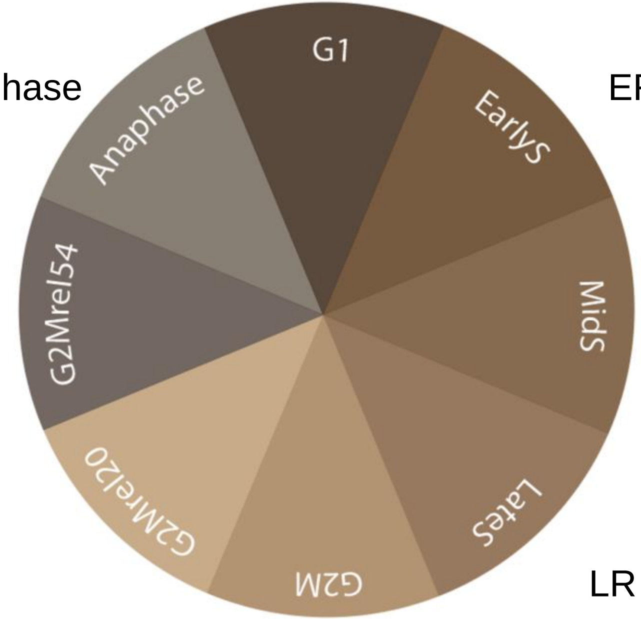


G1

Anaphase

G0

ER

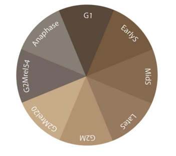


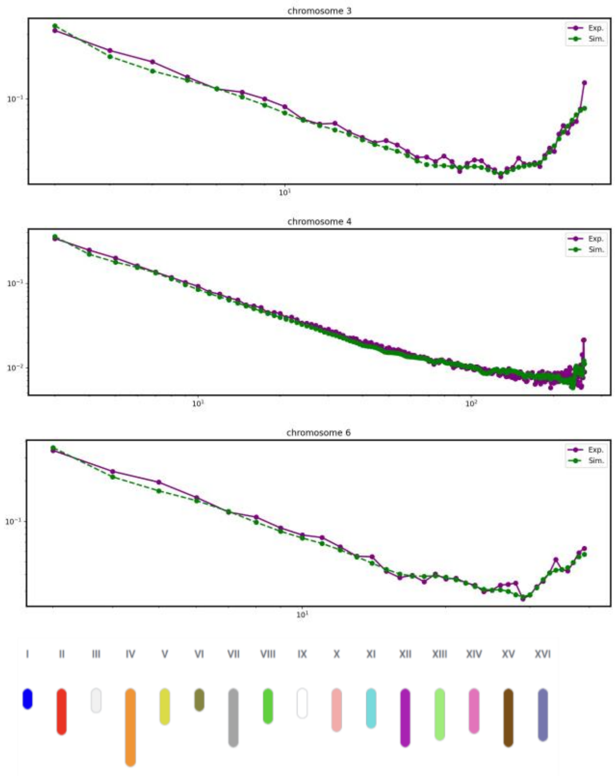
MR

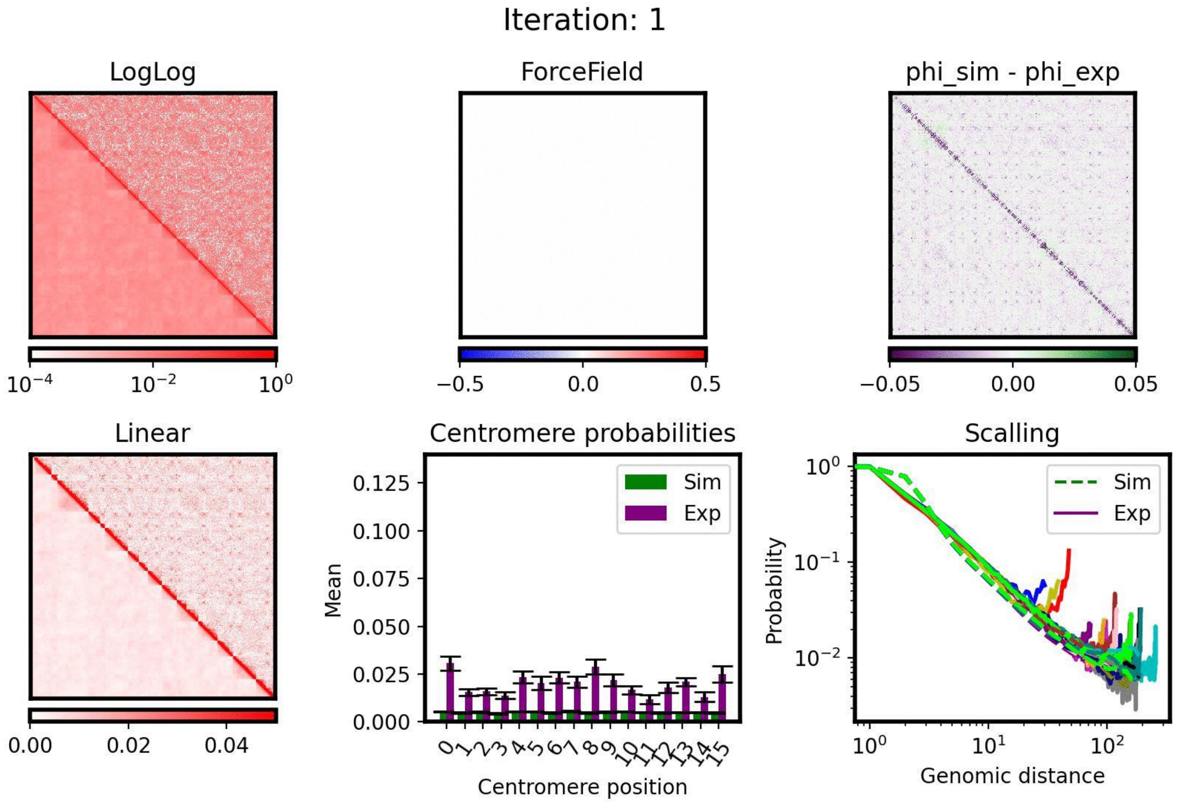
LR

G2M

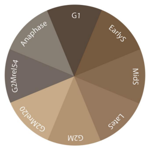
Yeast Genome cell cycle - G1

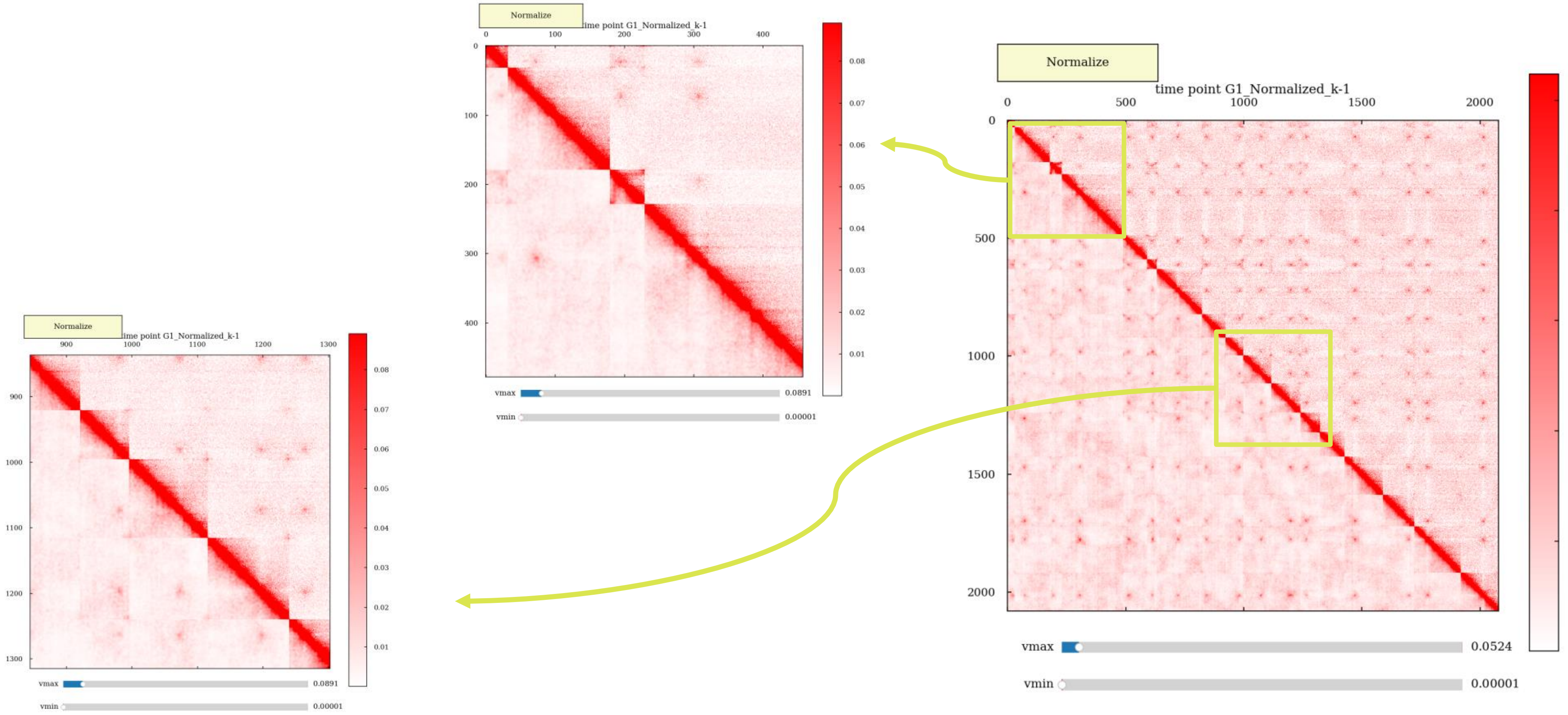




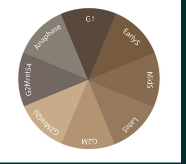


Yeast Genome cell cycle - G1

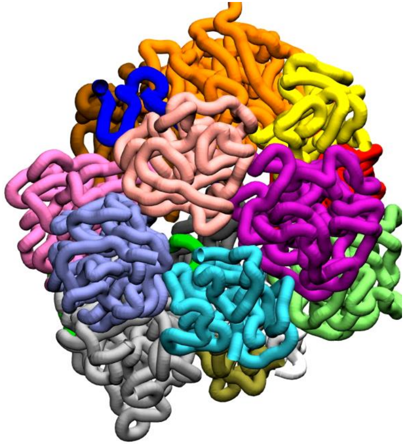


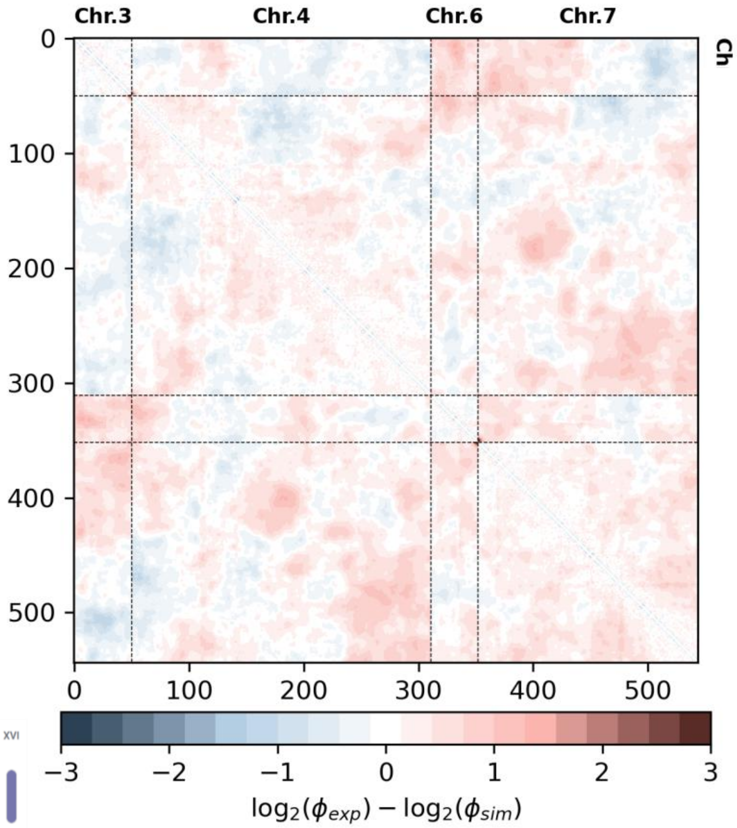


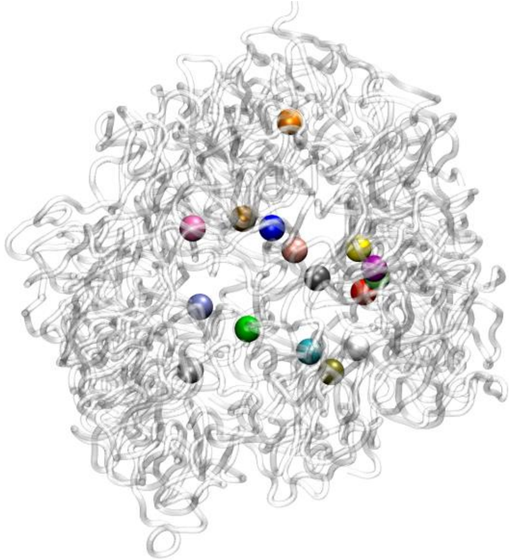
Yeast Genome cell cycle - G1



Serpentine



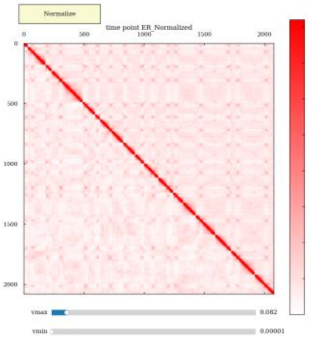




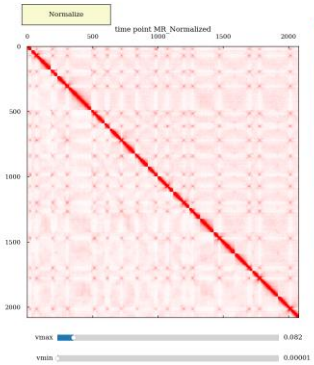
Baudry, L., Millot, G. A., Thierry, A., Koszul, R., & Scolari, V. F. (2020). Serpentine: a flexible 2D binning method for differential Hi-C analysis. Bioinformatics, 36(12), 3645–3651. doi: 10.1093/bioinformatics/btaa249

Yeast Genome cell cycle

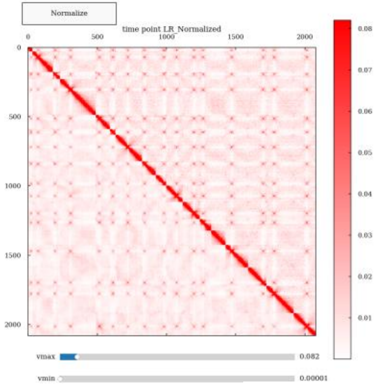
ER



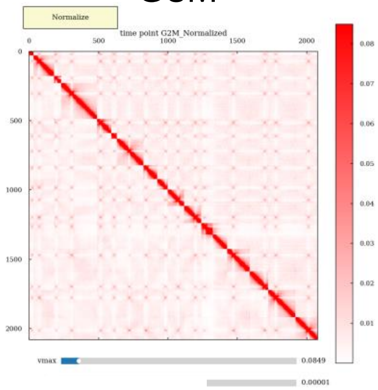
MR



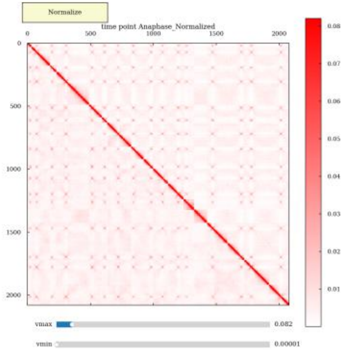
LR

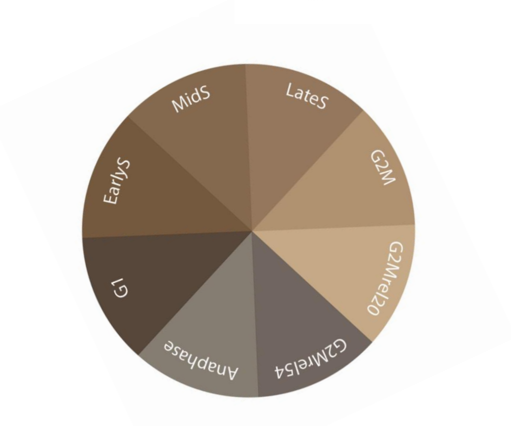


G3M

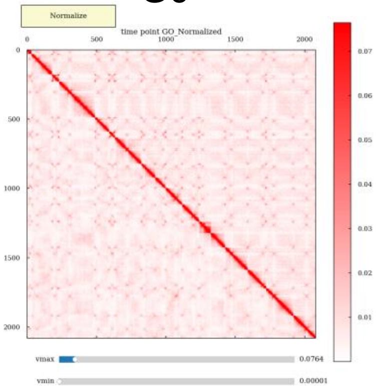


Anaphase





G0



Collaborators

José N. Onuchic (Rice Univ. - CTBP)

Peter G. Wolynes (Rice Univ. - CTBP)

Vinicius G. Contessoto (Rice Univ. - CTBP)

Sumitabha Brahmachari (Rice Univ. – CTBP)

Esteban Rojas (Rice Univ. - CTBP)

Matheus Mello (Rice Univ. - CTBP)

Ryan R. Cheng (Univ. Kentucky)

Erez Aiden (Baylor)

Olga Dudchenko (Baylor)

Ragini Mahajan (Baylor)

Michele Di Pierro (Northeastern Univ.)

Vittore Scolari (Institut Curie)

Paul C. Whitford (Northeastern Univ.)

Cynthia P. Estrada (Karolinska Univ.)

Ronaldo Oliveira (UFTM)

Vitor Leite (Unesp – SJRP)

Pedro Pascutti (UFRJ)

Leandro Martinez (UNICAMP)

Fernando Bruno (University of Warsaw)

Osvaldo N. Oliveira Jr. (USP – Sao Carlos)

Obrigado!

Thank you!

Merci!

antonio.oliveira@rice.edu

ndb.rice.edu

open-michrom.readthedocs.io

AMD HPC High Performance Compute Fund

THE Welch FOUNDATION

NSF

RICE

MaxEnt approach

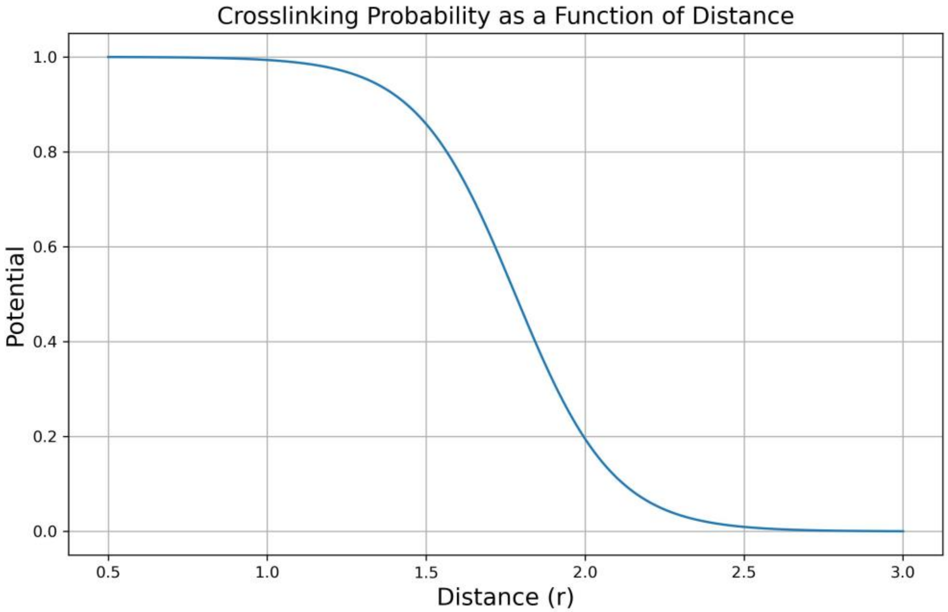
$$U_{\mathrm{ME}}(r) = U(r) + \sum_i \alpha_i f_i(r_{ab}),$$

Zhang, B., & Wolynes, P. G. PNAS (2015).

MaxEnt approach

$$U_{\mathrm{ME}}(r) = U(r) + \sum_i \alpha_i f_i(r_{ab})$$

$$f(r_{i,j}) = \frac{1}{2}\left(1 + \tanh\left[\mu(r_c - r_{i,j})\right]\right)$$



Zhang, B., & Wolynes, P. G. PNAS (2015).

MaxEnt approach

$$U_{\mathrm{ME}}(r) = U(r) + \sum_i \alpha_i f_i(r_{ab})$$

$$\frac{z(\alpha)}{z_o} = \frac{\int e^{-\beta U_{\mathrm{ME}}(r)} dr}{\int e^{-\beta U(r)} dr}$$

$$= \frac{\int e^{-\beta U(r)} e^{-\beta \sum_i \alpha_i f_i(r)} dr}{\int e^{-\beta U(r)} dr} = \left\langle e^{-\beta \sum_i \alpha_i f_i(r)} \right\rangle_U$$

$$= e^{\sum_{n=1}^{\infty} \frac{(-\beta)^n}{n!} \langle\langle (\sum_i \alpha_i f_i(r))^n \rangle\rangle}.$$

•Taking the first two term from the cumulant expansion

$$c_1 = -\beta \sum_i \alpha_i \langle f_i(r) \rangle, \quad c_2 = \frac{\beta^2}{2} \sum_i \sum_j \alpha_i \alpha_j \left[\langle f_i f_j \rangle - \langle f_i \rangle \langle f_j \rangle \right].$$

$$\Gamma(\alpha) = \ln\left(\frac{z(\alpha)}{z_o}\right) + \beta \sum_i \alpha_i f_{i,\,\mathrm{exp}}$$

$$\Gamma(\alpha) = \frac{\beta^2}{2} \alpha^T B \alpha - \beta \left[\langle f_i \rangle - f_{i,\mathrm{exp}} \right]^T \alpha,$$

$$\alpha = \frac{1}{\beta} B^{-1} \left[\langle f_i \rangle - f_{i,\mathrm{exp}} \right]^T.$$

$$B_{ij} = \langle f_i f_j \rangle - \langle f_i \rangle \langle f_j \rangle.$$

Hessian Matrix

Zhang, B., & Wolynes, P. G. PNAS (2015).

First-Order Direct Inversion

Taking the first term from the cumulant expansion

$$c_1 = -\beta \sum_i \alpha_i \langle f_i(r) \rangle,$$

$$\Gamma(\alpha) = -\beta[\langle f_i \rangle - f_{i,exp}]a_i$$

Taking the gradient Γ(α)

$$g_t = -\beta[\langle f_i \rangle - f_{i,exp}]$$