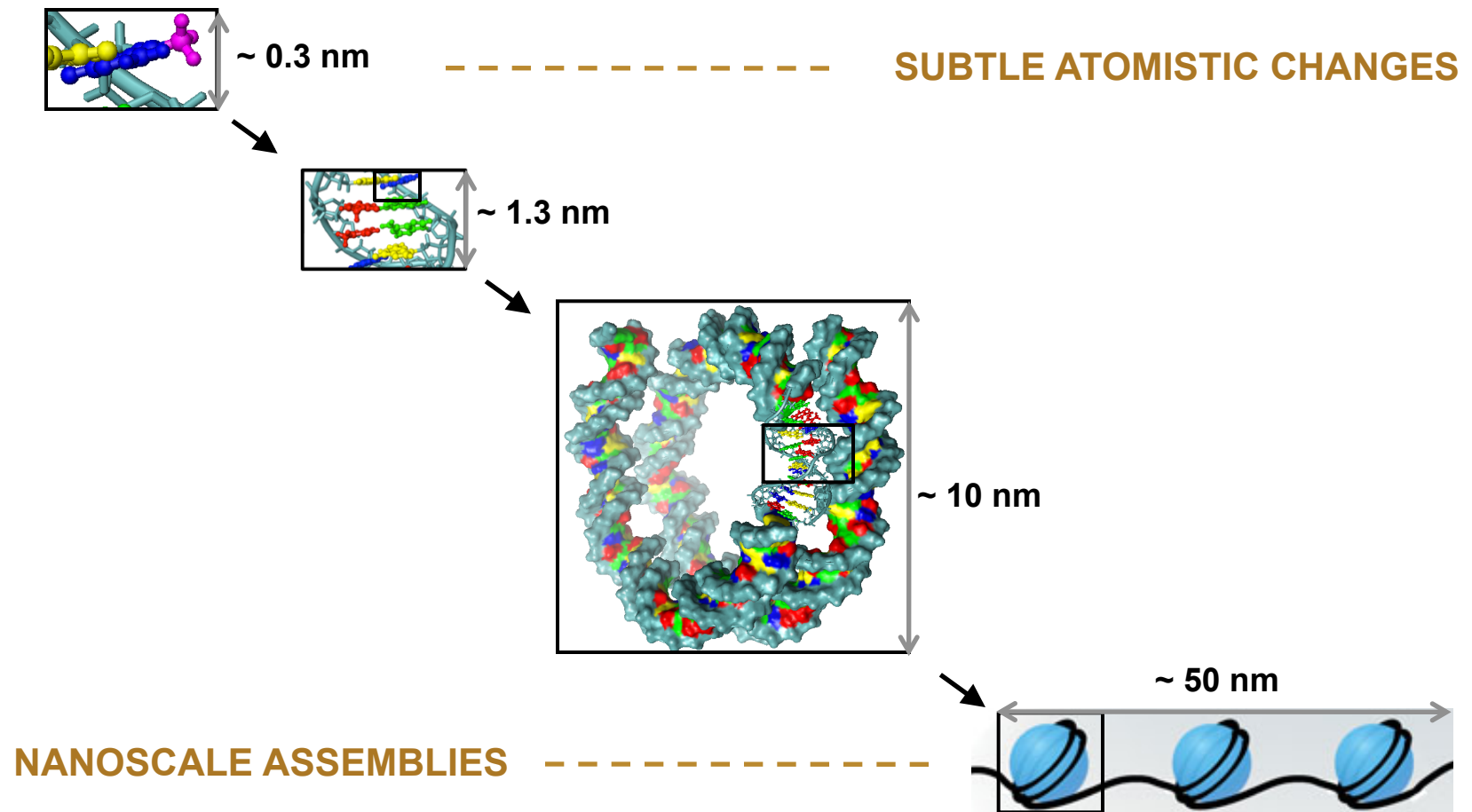# MODELLING BY HIERARCHICAL NATURAL MOVES

Peter Minary
Department of Computer Science
Oxford University

Molecular Dynamics Day
Department of Statistics, Oxford University
13th Jun, 2016

MODEL ATOMIC EFFECTS ON THE NANOSCALE

~ 0.3 nm – – – – – – – – – – – – – SUBTLE ATOMISTIC CHANGES

~ 1.3 nm

~ 10 nm

~ 50 nm

NANOSCALE ASSEMBLIES – – – – – – – – – – – –

NUMBER OF DEGREES OF FREEDOM

– **The Integral of Interest:** **A general system**

$$Q : (\Omega, E) \to \mathbb{R}$$
$$E : \Omega \to \mathbb{R}$$

$$Q = \int_{\Omega} \dots \int \exp(-\beta E(\mathbf{x})) \, d\mathbf{x}$$

– **A Case Study:** **A Helix Pair with 1000 atoms each**
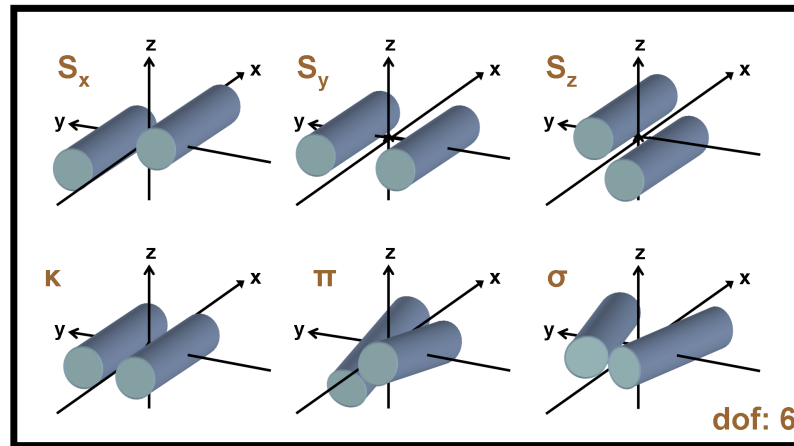
- **Cartesian DoFs**

  $$\mathbf{x} \in \Omega \subset \mathbb{R}^n$$

  $$n = 6000$$

- **Essential Moves**

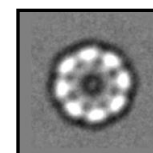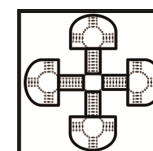 $$\mathbf{x} \in \Omega \subset \mathbb{R}^m$$
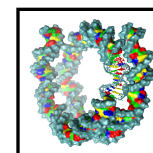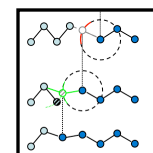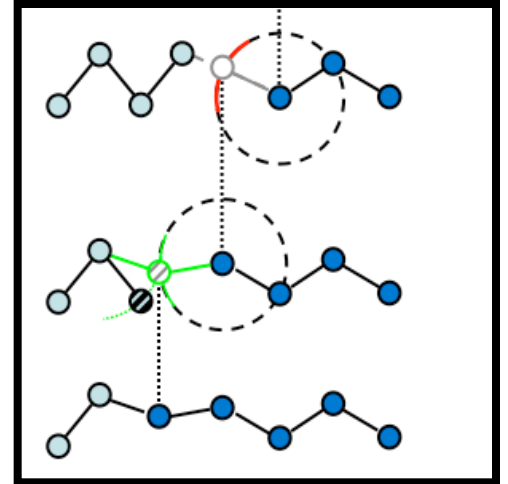
  $$m = 6$$

SOLUTION: NMMC (Natural Move Monte Carlo)

THE PROBLEM

THE OUTLINE

- Natural Move Monte Carlo, Basic Concepts & the Algorithm

- *In Silico* Epigenetics, Nucleosome Positioning / Methylation

- Hierarchical Modeling of RNA Junctions & Nanotechnology

- Multi Scale Cryo-EM Refinement Against Single 2D Images
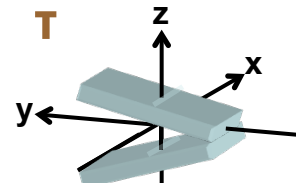
NATURAL MOVE MONTE CARLO

NATURAL DEGREES OF FREEDOM FOR NUCLEIC ACIDS

D_x Shift
D_y Slide
D_z Rise

τ Tilt
ρ Roll
ω Twist

S_x Shear
S_y Stretch
S_z Stagger

κ Buckle
π Propeller
σ Opening

dof: 10
(4+12x½)

Moves break the chain!

7
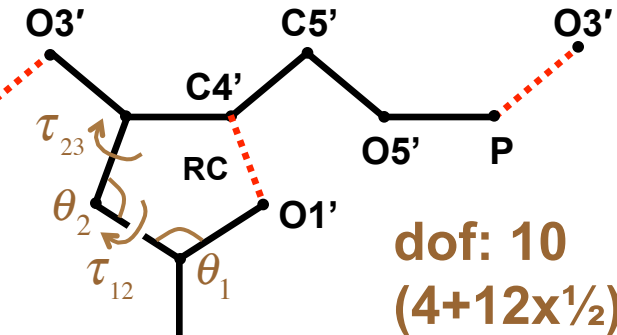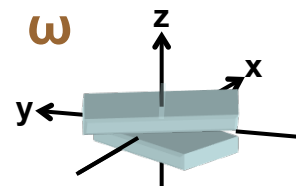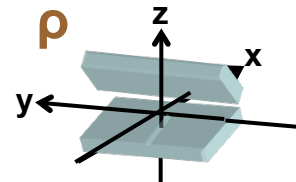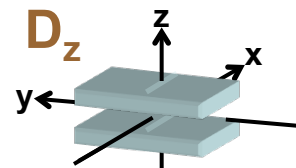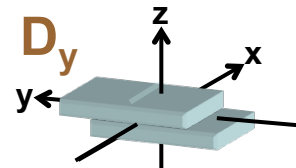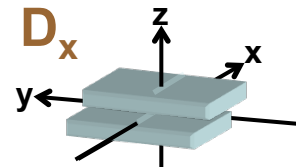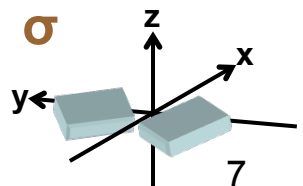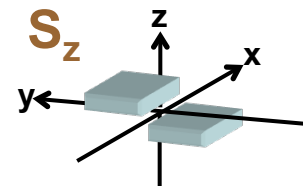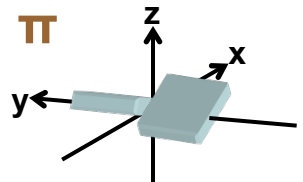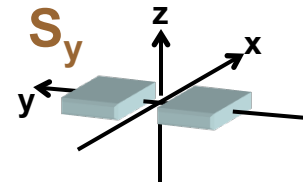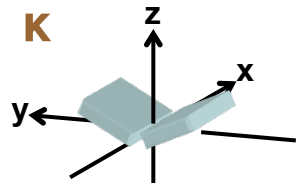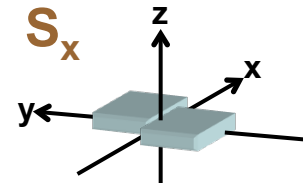
NATURAL DEGREES OF FREEDOM FOR PROTEINS

S_x Shear
S_y Stretch
S_z Stagger

κ   Buckle
π   Propeller
σ   Opening

S_x
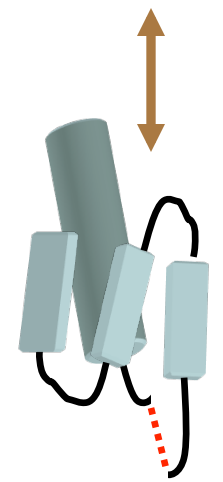
Moves break the chain!
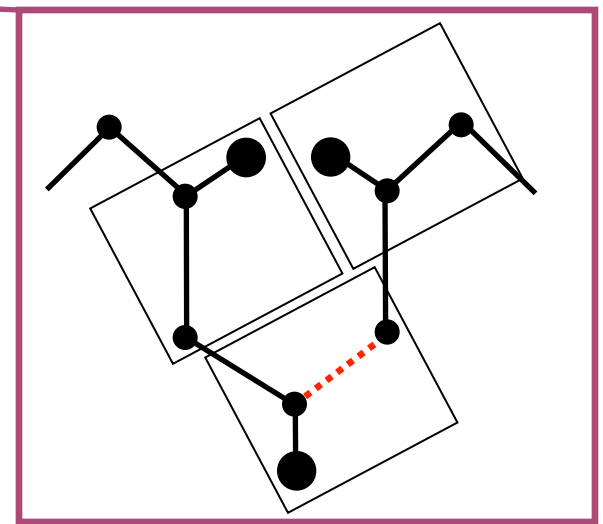
RECURSIVE STOCHASTIC CLOSURE

RSC = DFC[SPC[SPC[SPC[…]]]]
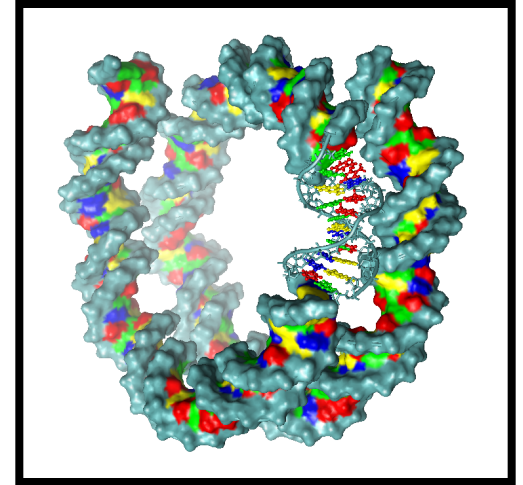
Molten zone

SPC$_2$

SPC$_3$

DFC

RSC

Molten zone

APPLICATIONS

NUCLEOSOME POSITIONING

THE COMPUTATIONAL PIPELINE

Threading

C A A G
G T T C

S_n GCTT..TCCA GTC TT..ACAC

Q TATCCAGTTCTTACACAGTCTACTAGTTCATAAGTCGCTATC

Template **Sequence** **Structure**

$E(i)$ $O(i)$ $i = n$ $S(i)$

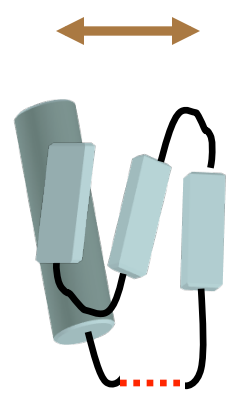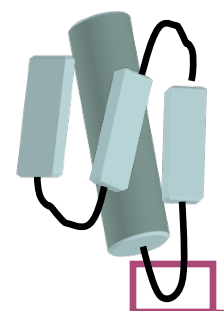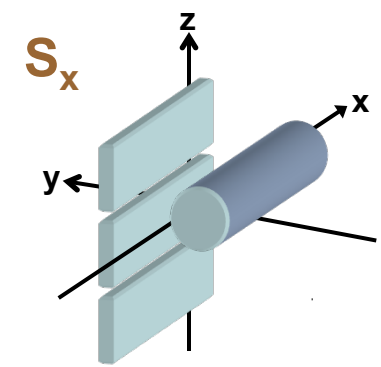# NUCLEOSOME FORMATION ENERGY

S_j

$E = E (\ \bigcirc\ ) - E (\ \rule{1cm}{0.3cm}\ )$

T_k

Q TATCCAGTTCTTACACATCTACTAGTTCATAAGTCGCTATCTATCCAGTTCTTACACAGTCTACTAGTTCATAAGTCGCTATCTATCCAGTTCTTACACAGTCTACTAGTTCATAAGTCGCTATC

T_j

$O_i / O_k \sim \exp(-\beta\ \Delta E_{jk})$

S_k

12
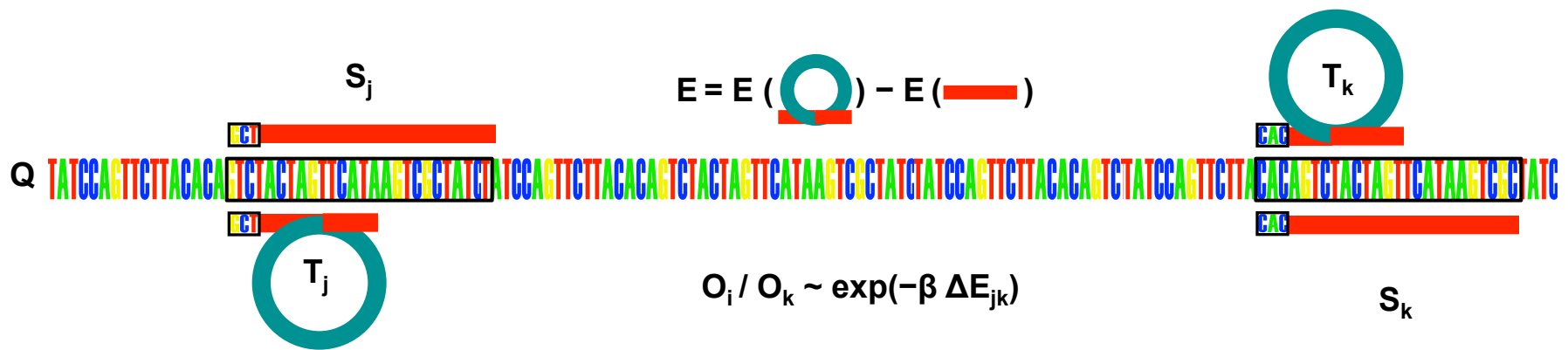
IN VITRO EXPERIMENT

Pearson

CC = −0.62

$E_n − E_l$: **Nucleosome Formation Energy**

InVitro : Occupancy From MNase Study
Kaplan et al., *Nature* **458** 362-366 (2009).
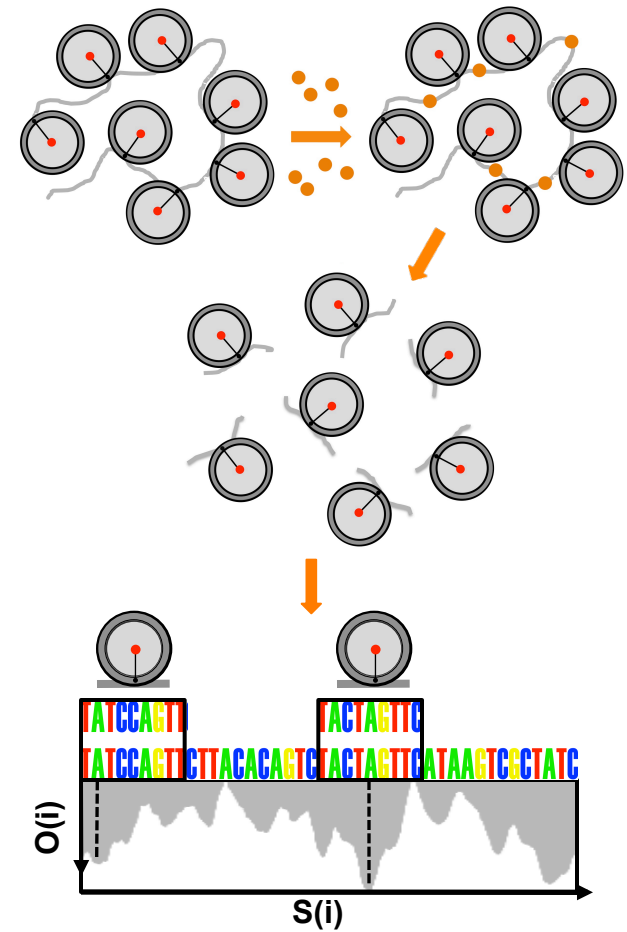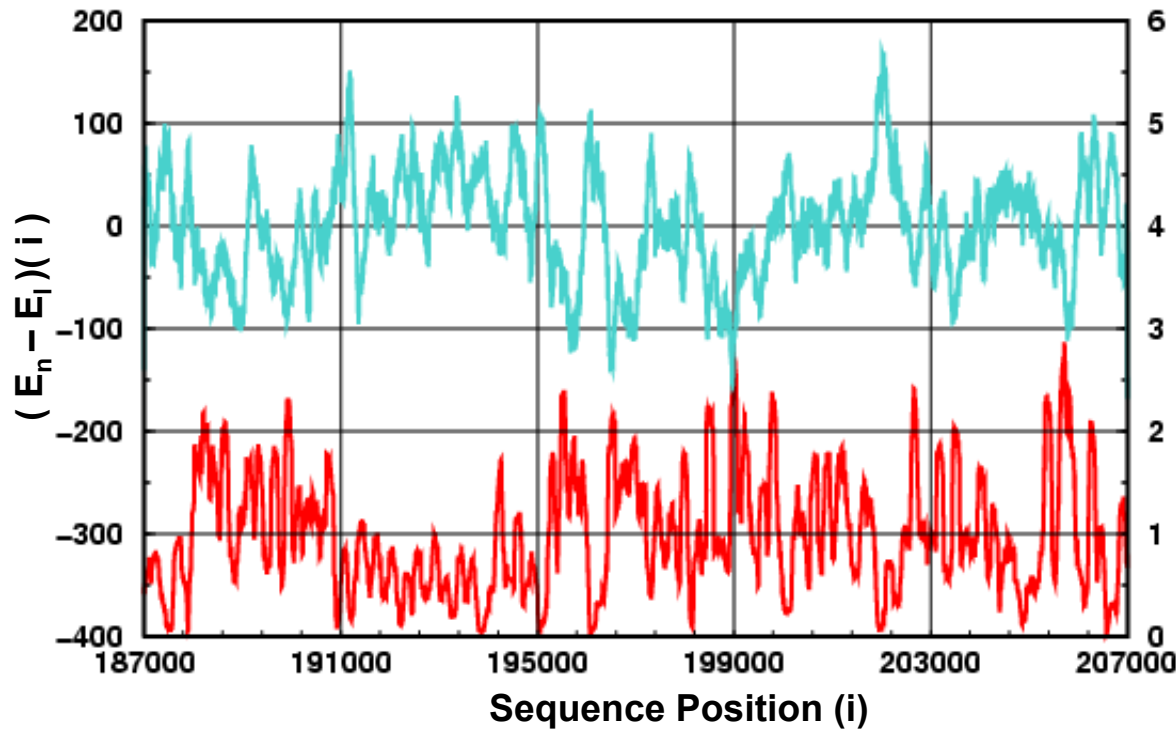
NUCLEOSOME POSITIONING SEQUENCES

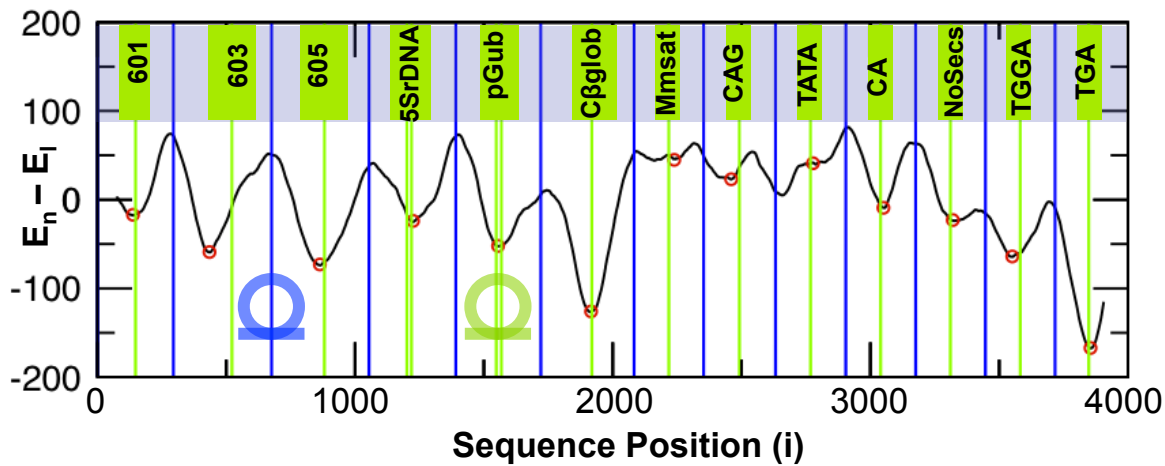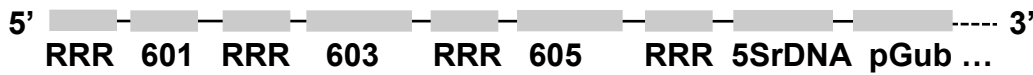| Name	Length	Dyad	Name	Length	Dyad
601	147	74	CAG	132	0
603	234	154	TATA	126	0
605	231	132	CA	124	0
5S_rDNA	190	74	NoSecs	124	0
pGub	183	84	TGGA	123	0
Cβglobin	215	125	TGA	115	0
Mmsat	123	0	RRR	147	74

van der Heijden et al., *PNAS* **109** E2514-E2522 (2012).

CONSTRUCT THE QUERY SEQUENCE BY CONCATENATING N.P. ELEMENTS
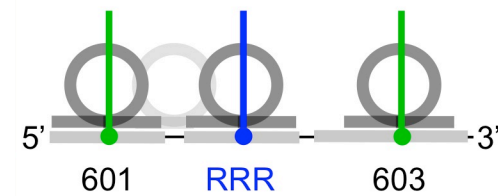
5' 601 RRR 603 -3'

QUERY SEQUENCE

5' RRR 601 RRR 603 RRR 605 RRR 5SrDNA pGub ... 3'

	601	603	605	5Sr	pGu	Cβg
L	147	234	231	190	183	215
D	74	154	132	92	104	125
dD	-10	-87	-18	6	7	-3

Mms	CAG	TAT	CA	NoS	TGG	TGA
123	132	126	124	124	123	115
N/A	N/A	N/A	N/A	N/A	N/A	N/A
21	-32	9	11	9	-31	7

14

EPIGENETIC EFFECTS

**Small Atomic Changes
(5Me-C, 5hMe-C, HXXxx, …)**

↓

**Large Meso-Scale Assemblies
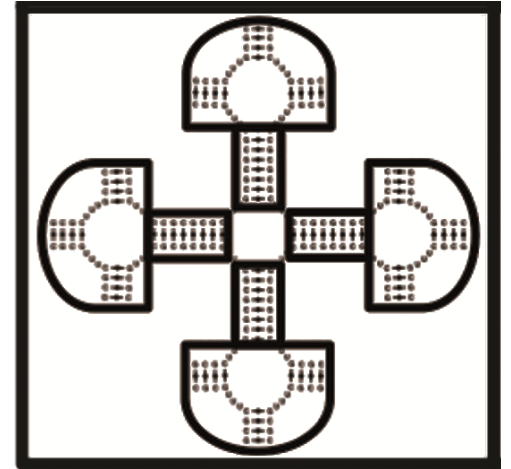(Nucleosome, Hetero/EuChromatin)**

NUCLEOSOME POSITIONING SEQUENCES-II
DNA METHYLATION

Name	Length	Dyad	Name	Length	Dyad
601	147	74	CAG	132	0
603	234	154	TATA	126	0
605	231	132	CA	124	0
5S_rDNA	190	74	NoSecs	124	0
pGub	183	84	TGGA	123	0
Cβglobin	215	125	TGA	115	0
Mmsat	123	0	RRR	147	74

	601	603	605	5Sr	pGu	Cβg
L	147	234	231	190	183	215
D	74	154	132	92	104	125
dD	-10	-87	-18	6	7	-3
dD_M	4	-43	-108	32	-21	-79
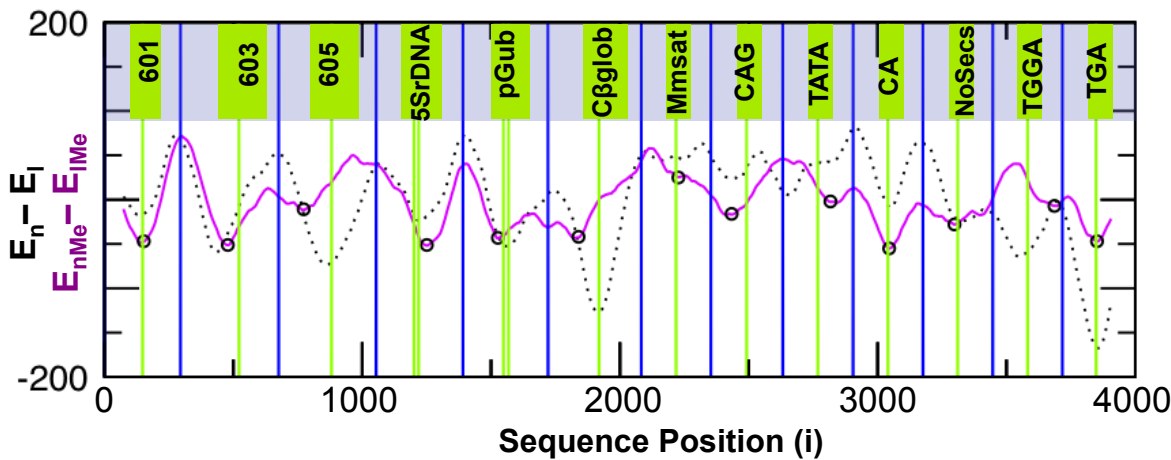
Mms	CAG	TAT	CA	NoS	TGG	TGA
123	132	126	124	124	123	115
N/A	N/A	N/A	N/A	N/A	N/A	N/A
21	-32	9	11	9	-31	7
9	-58	49	5	-13	105	4

Plot axes: $E_n - E_i$ and $E_{nMe} - E_{IMe}$ (vertical, ranging -200 to 200) vs Sequence Position (i) (horizontal, 0 to 4000). Labels along top: 601, 603, 605, 5SrDNA, pGub, Cβglob, Mmsat, CAG, TATA, CA, NoSecs, TGGA, TGA.
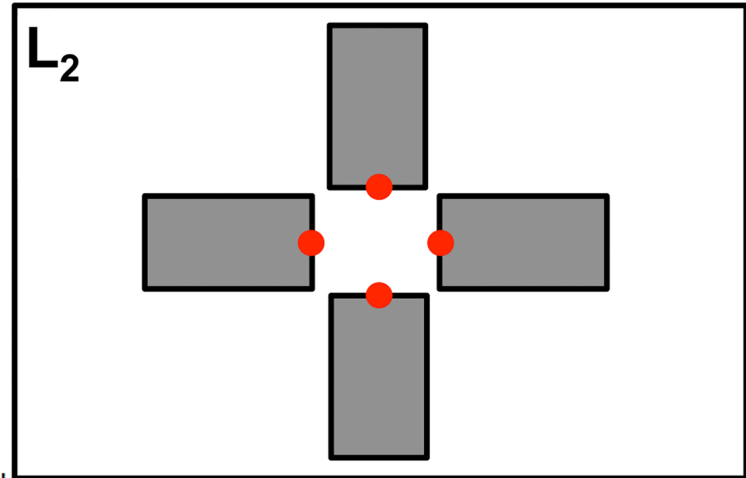
16

HIERARCHICAL
NATURAL MOVE MONTE CARLO

Sim, A. S. L., Levitt, M., **Minary**, **P**. Modeling and Design by Hierarchical Natural Moves. *PNAS* **109**: 2890-2895 (2012).
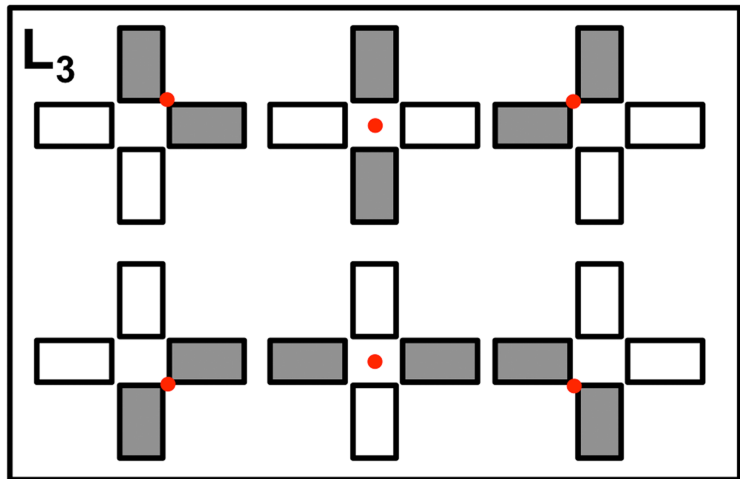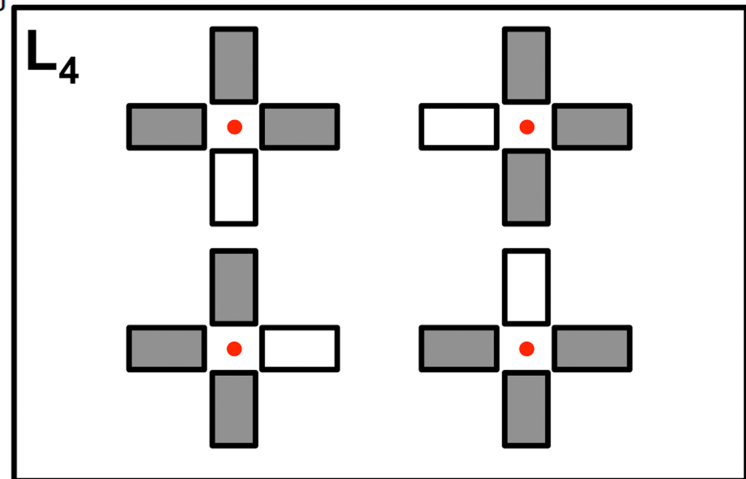
RNA 4-WAY JUNCTION

18
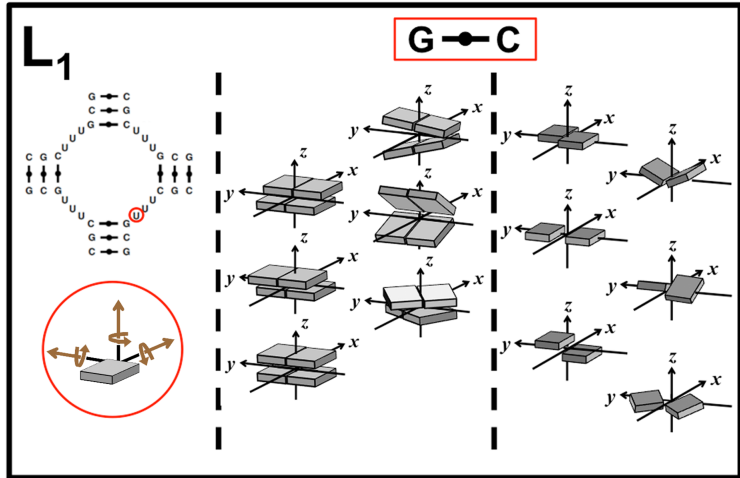
TESTING THE APPROACH / SOFTWARE

NM-MOSAICS   **FA-MCSym**   **FA-Rosetta**   **HNM-MOSAICS**

L_1

$L_1 - L_4$

Distance (Å)   Distance (Å)   Distance (Å)   Distance (Å)
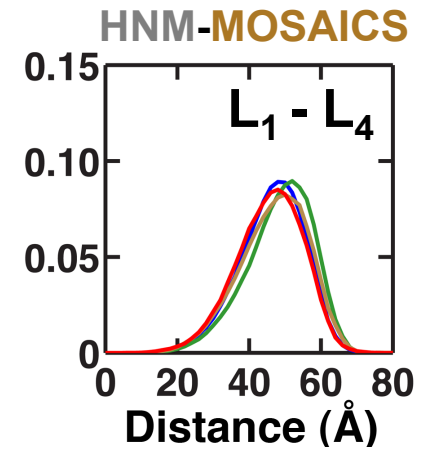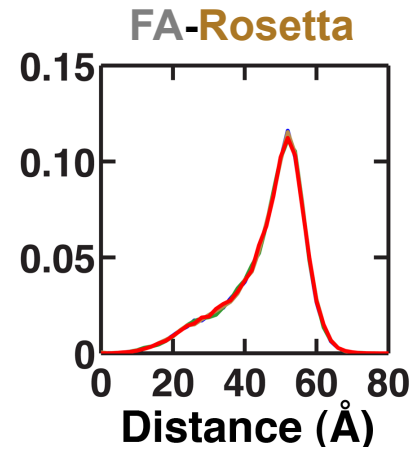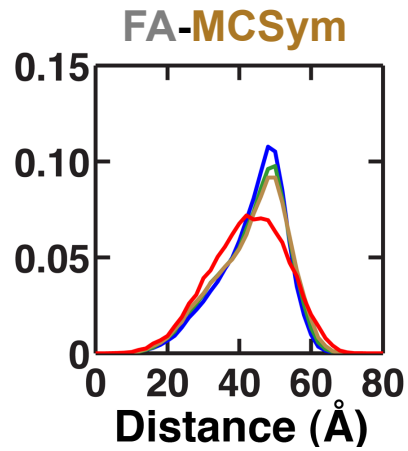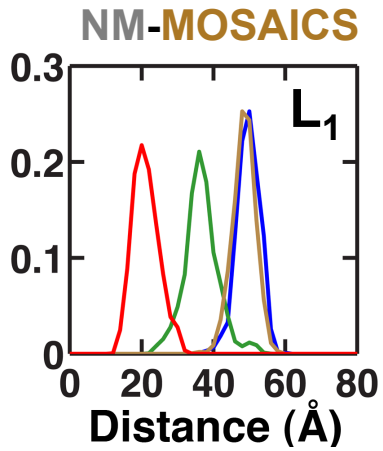
Distance Distributions

19
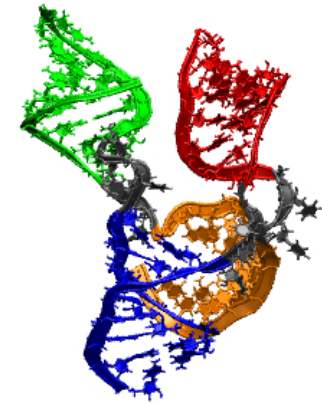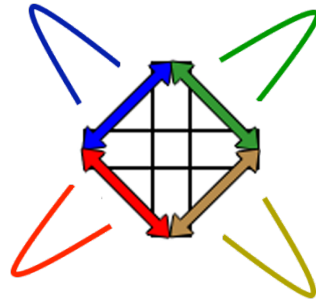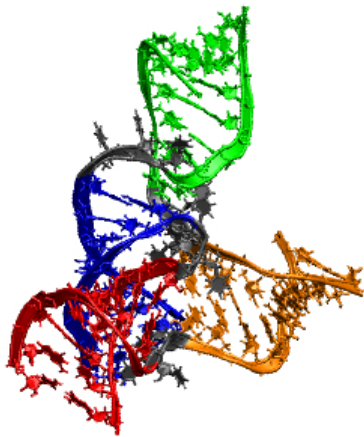
FRACTAL-LIKE RNA STRUCTURE

http://www.cs.ox.ac.uk/mosaics/

$L_1 - L_7$

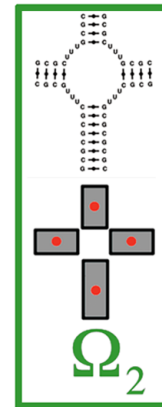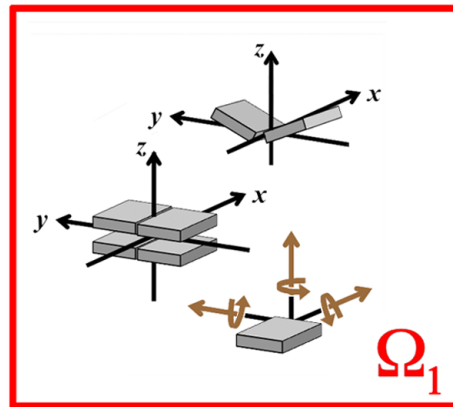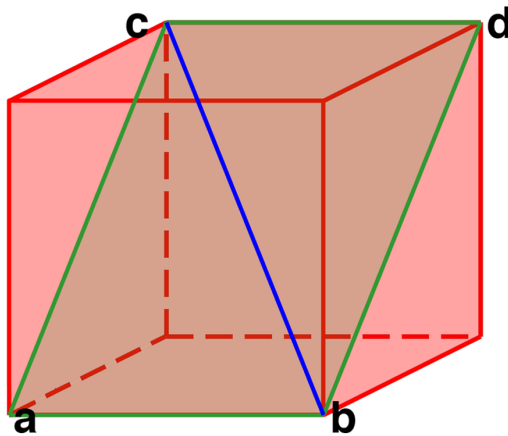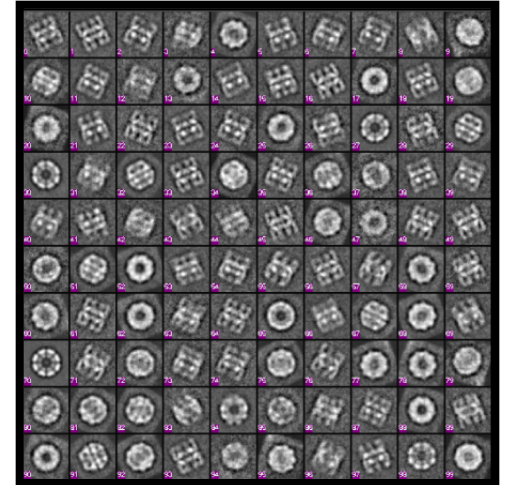$$\varepsilon(i) = \sum_{i<j} \delta \left[\Lambda_i, \Lambda_j \right]$$

20
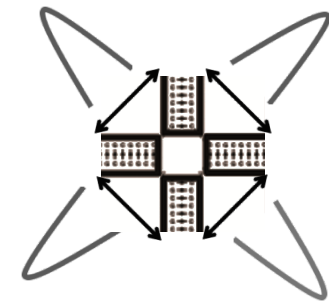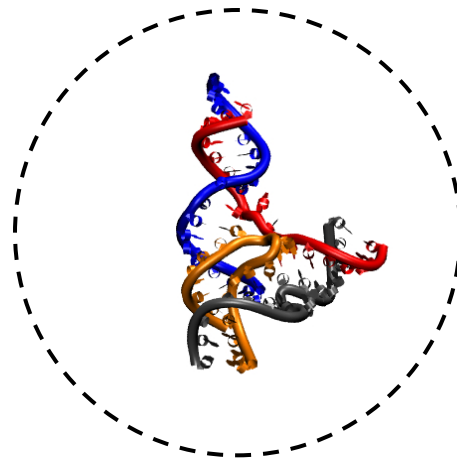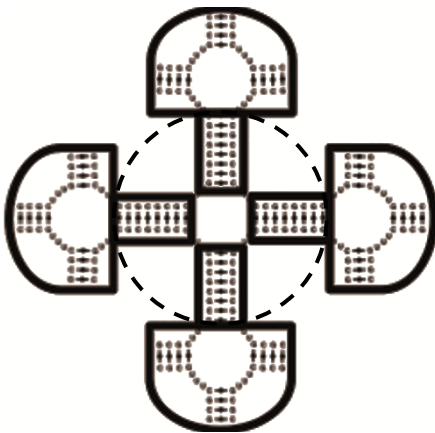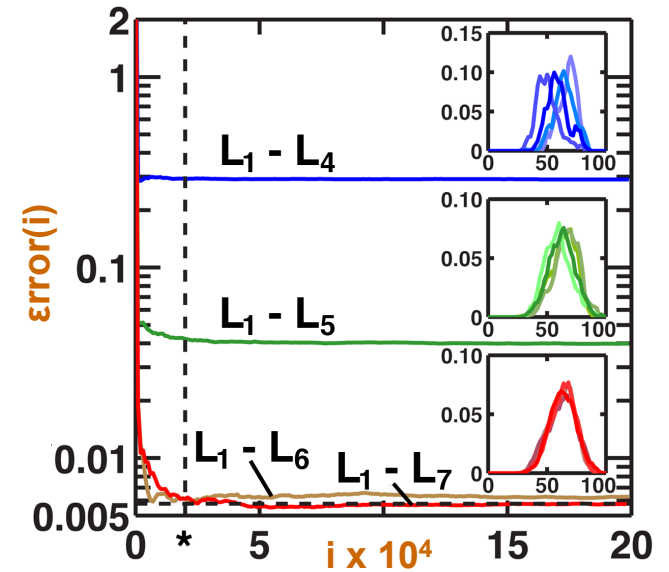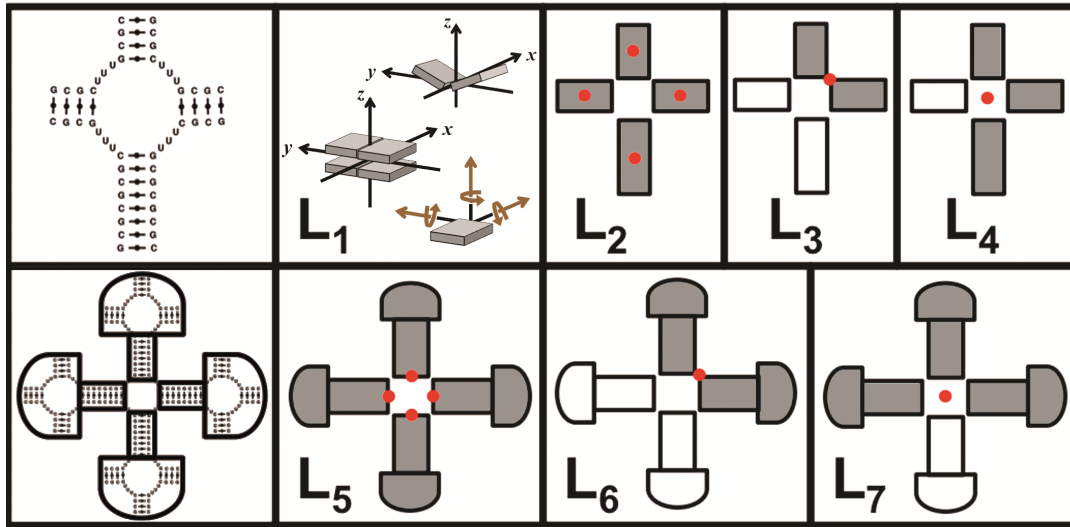
HIERARCHY OF (SUB)SPACES
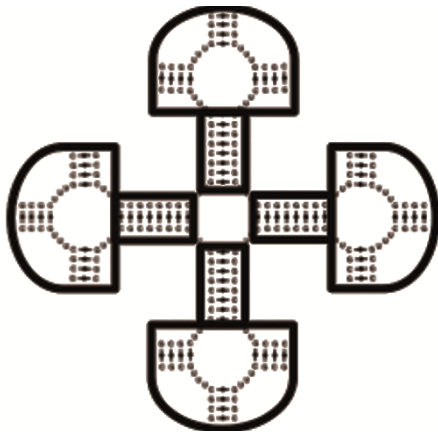
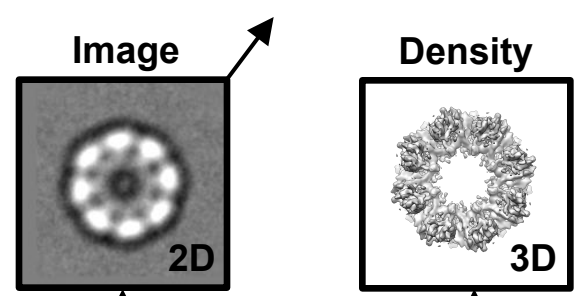$$\Omega_1 \supset \Omega_2 \supset \Omega_3$$
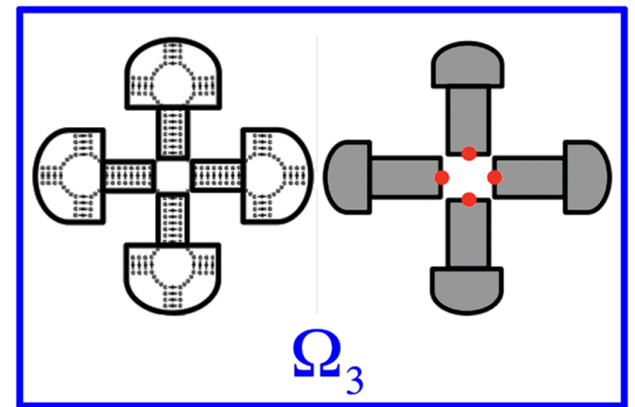
CRYO-EM REFINEMENT

Zhang, J., **Minary**, **P**., Levitt, M. Multi-Scale Natural Moves Refine Macromolecules Using Single Particle Electron Microscopy Projection Images. *PNAS* **109**, 9845-9850 (2012).
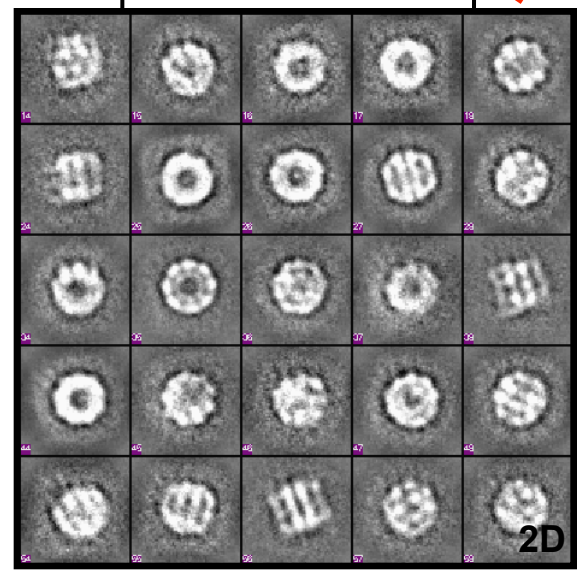
REFINEMENT AGAINST 2D IMAGES

Optimize Against

Image   **Density**

2D   3D

Reconstruction

2D

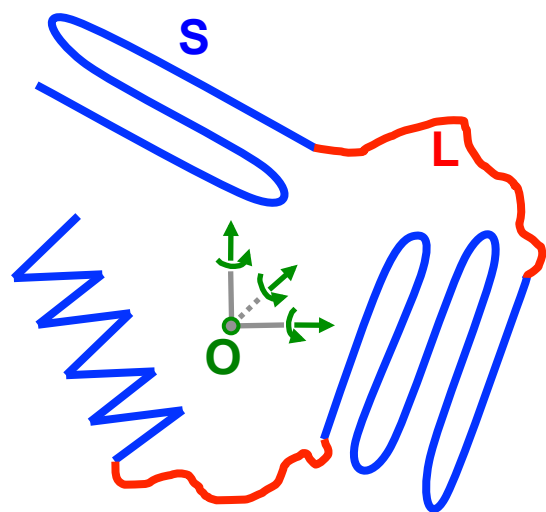**NOVEL 2D FITTING ENABLED BY NATURAL DOF**

TRADITIONAL 3D FITTING DONE BY MD, NM

ORIENTATIONAL NATURAL MOVE MC

S: Segment

L: Loop

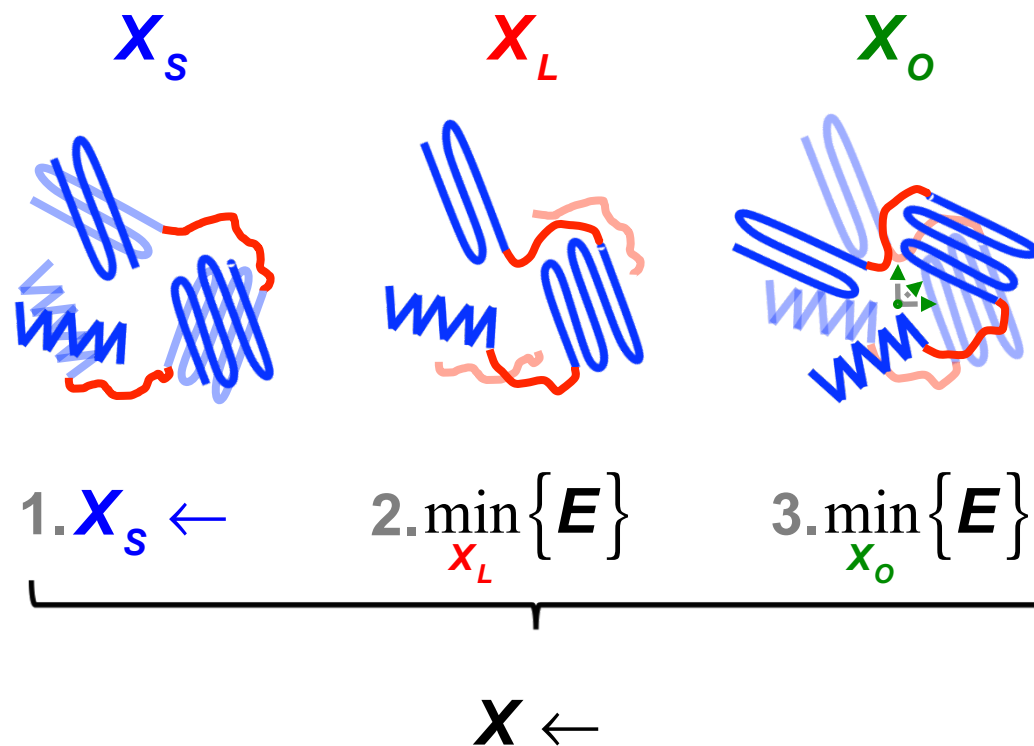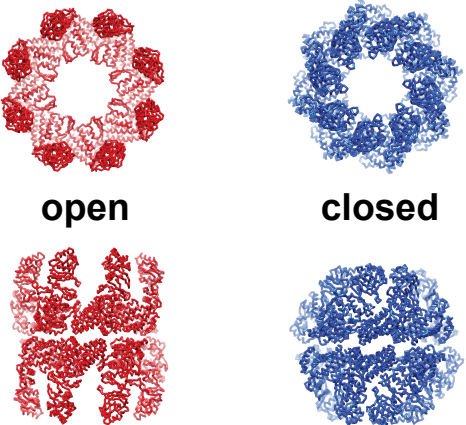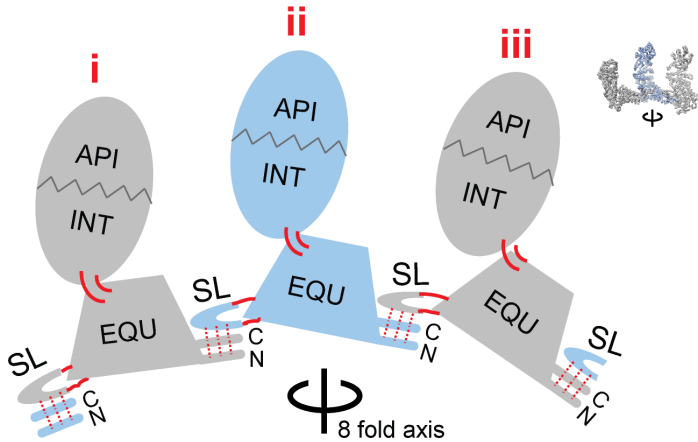O: Orientation

X_S X_L X_O

S

L

O

$X = X_S \cup X_L \cup X_O$

$1. X_S \leftarrow$ $2. \min_{X_L} \{E\}$ $3. \min_{X_O} \{E\}$
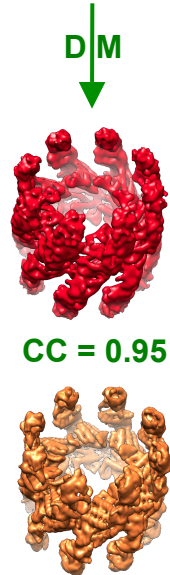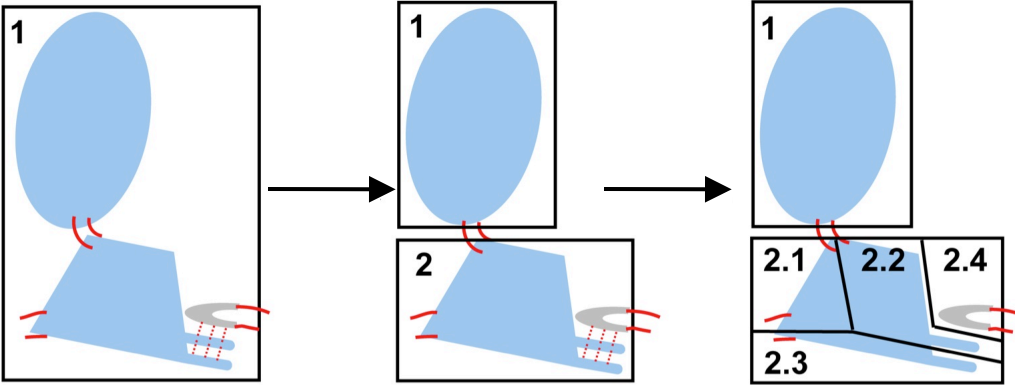
$X \leftarrow$

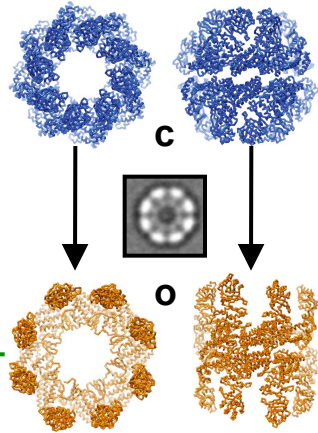$p_{acc} \leftarrow \exp\left[-\beta \; \Delta\widetilde{E}(X)\right]$

24

MM-CHAPERONIN

The Structure
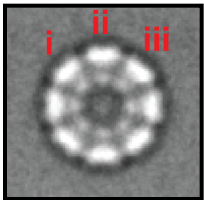
Multi Scale Natural Degrees of Freedom

open   closed

Refinement

CC = 0.95

MM-CHAPERONIN

5000 o + c

ATP/AlFx state

+

ATP-free state

100 class average

Examples
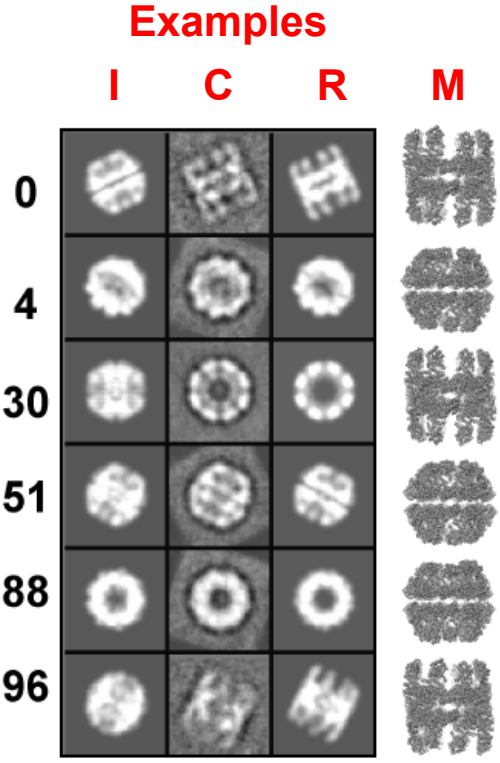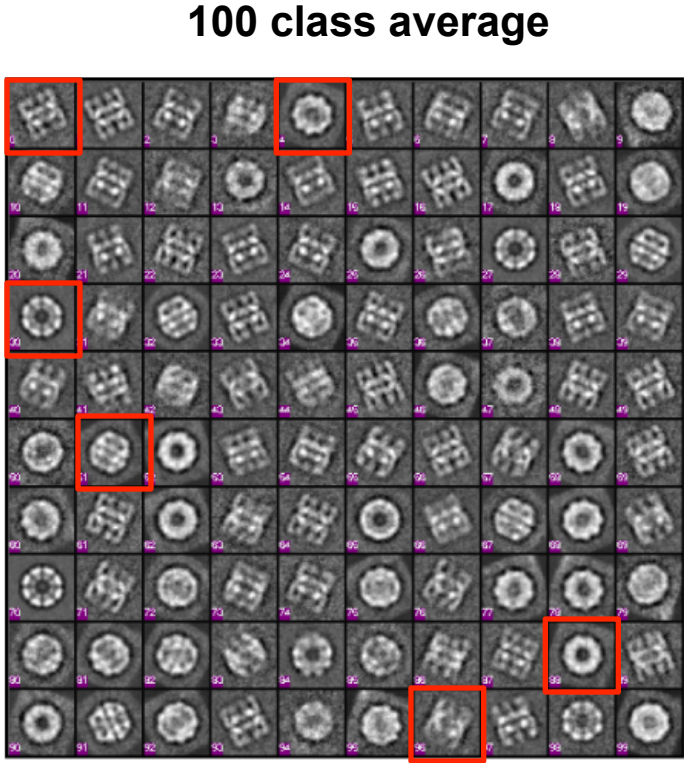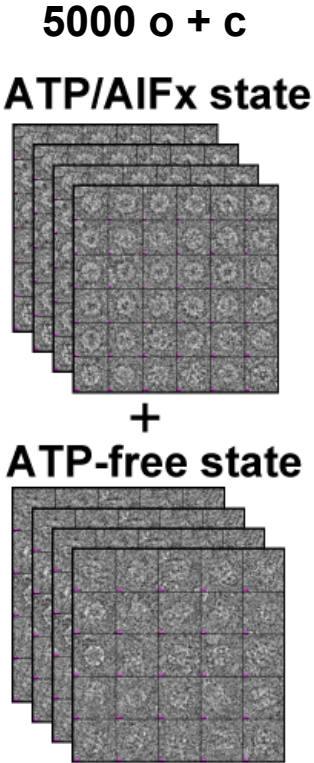
I : initial projection; C : class average; R : refined projection; M : refined model

CONCLUSION

- *In Silico* Nucleosome Positioning using Structural Data
  - Training-free methods offer the study epigenetic effects
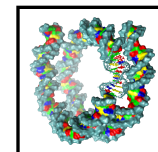
- Hierarchical Modeling of RNA Junctions & Nanotechnology
  - Explores the entire conformational space of RNA constructs
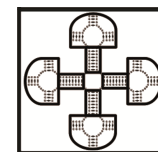
- Multi Scale Cryo-EM Refinement Against Single 2D Images
  - Eliminates data heterogeneity and preferred orientation issues

ACKNOWLEDGEMENTS

- ## Michael Levitt
  Robert W. and Vivian K. Cahill Professor in Cancer Research in the School of Medicine
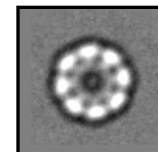 *Professor, by courtesy, of Computer Science.* **Stanford University**

- ## Adelene Sim
  Bioinformatics Institute, **A*STAR, Singapore**

- ## Junjie Zhang
  Department of Biochemistry and Biophysics, **Texas A&M University**

Methodologies for   MOSAICS
Optimization and
SAmpling
In   http://www.cs.ox.ac.uk/mosaics
Computational
Studies
                                    SOFTWARE APPLIED

MOSAICS

COMPUTATIONAL STRUCTURAL BIOLOGY