

Structural Bioinformatics

Davide Baù

Staff Scientist

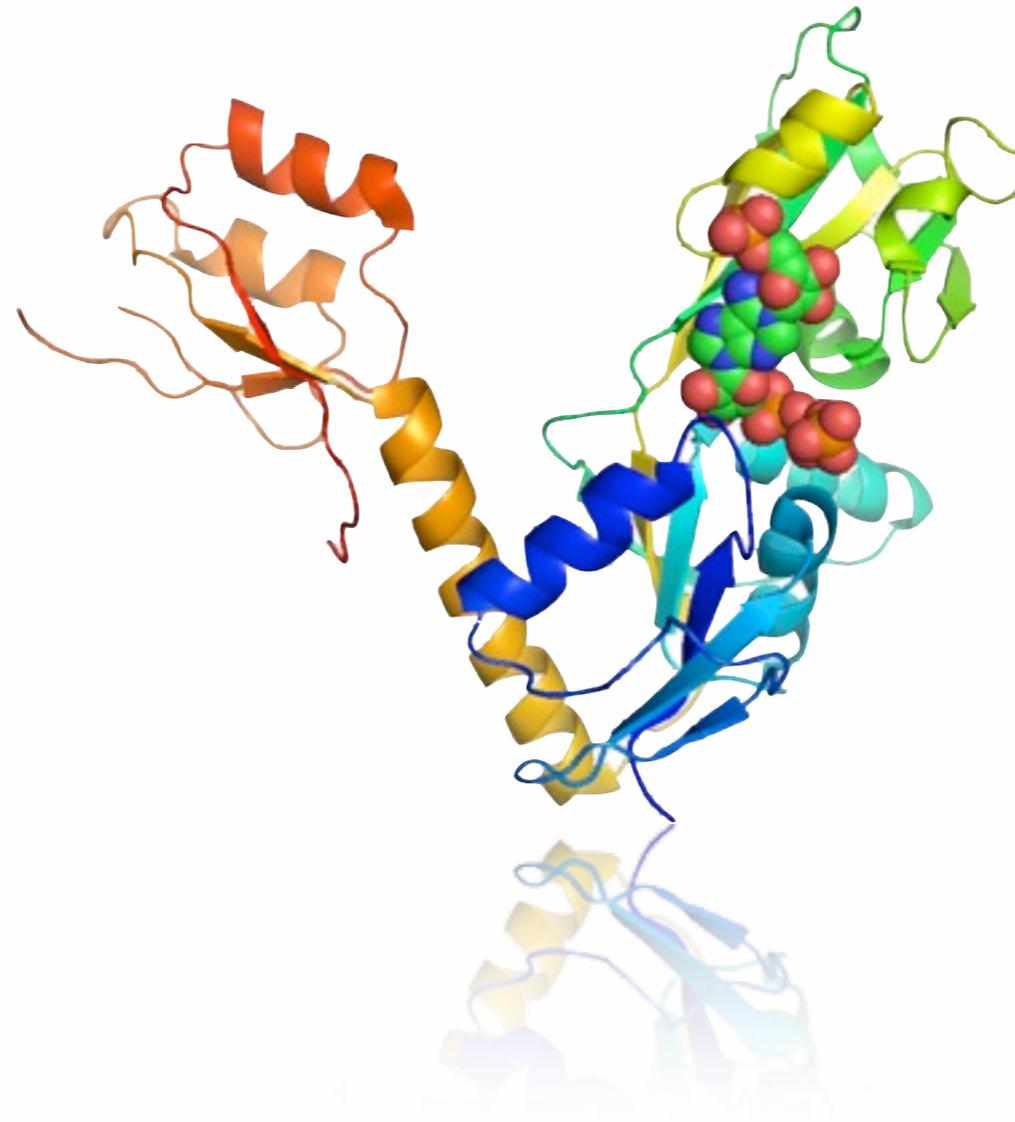
Genome Biology Group (CNAG)

Structural Genomics Group (CRG)

dbau@pcb.ub.cat

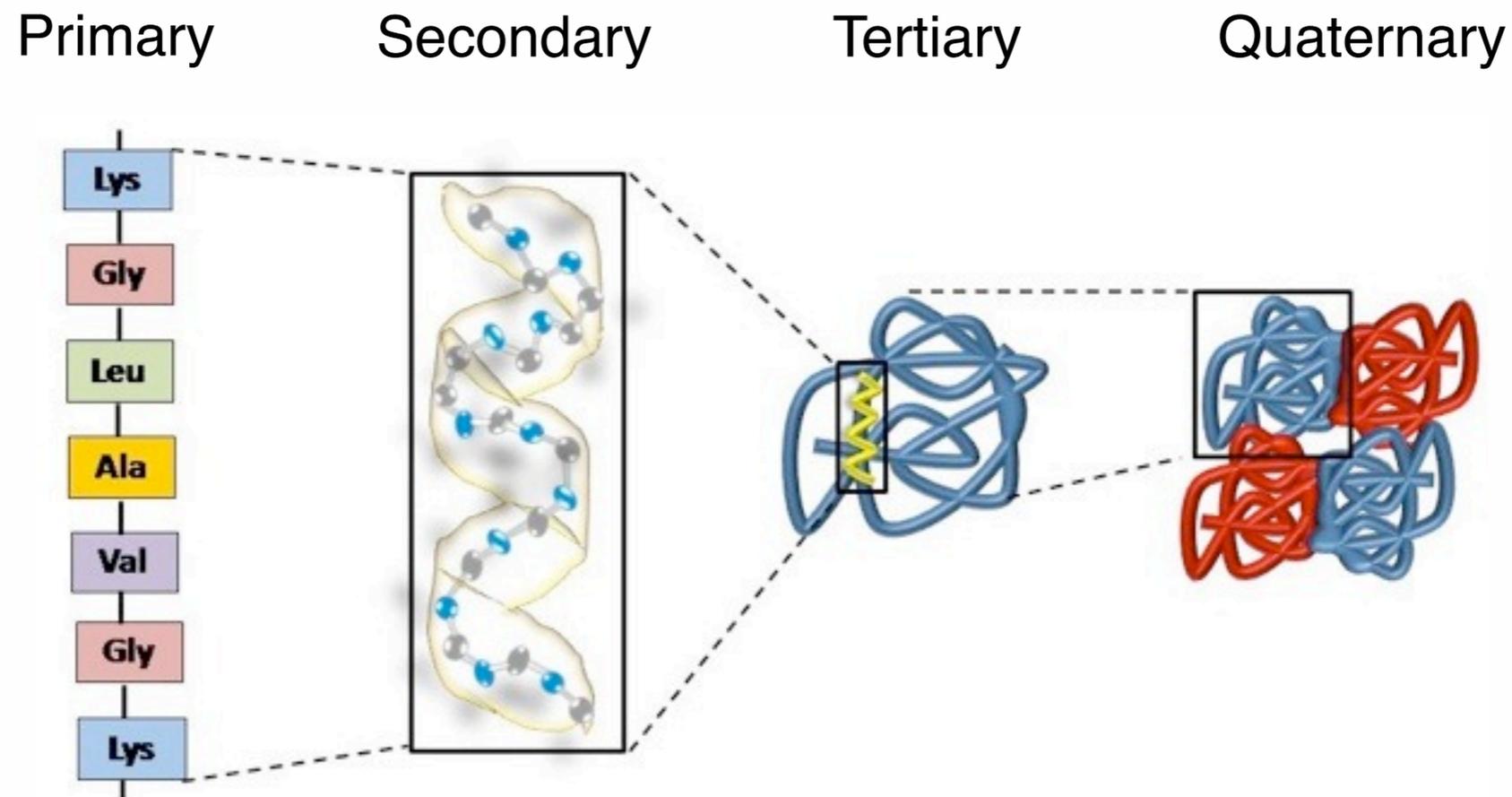


Proteins



Summary

Protein structural levels



Take home message

Biochemical function

Activity depends on the 3D structure

Evolution conserve

Structure is more conserved than sequence

Protein types

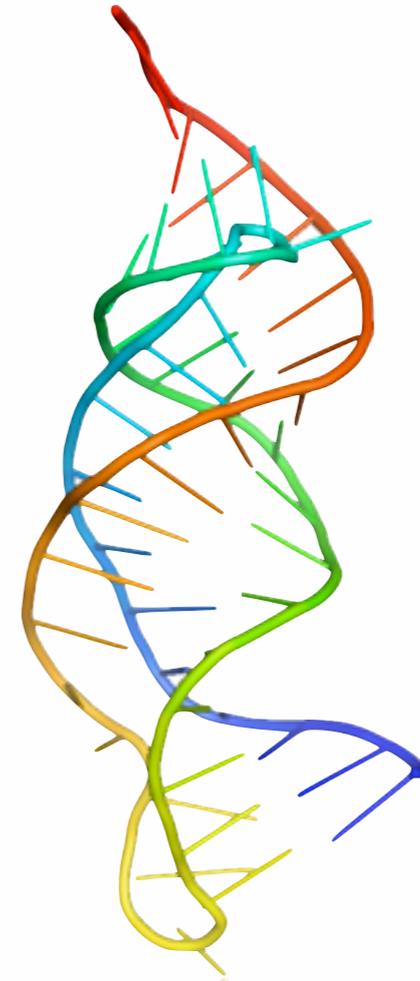
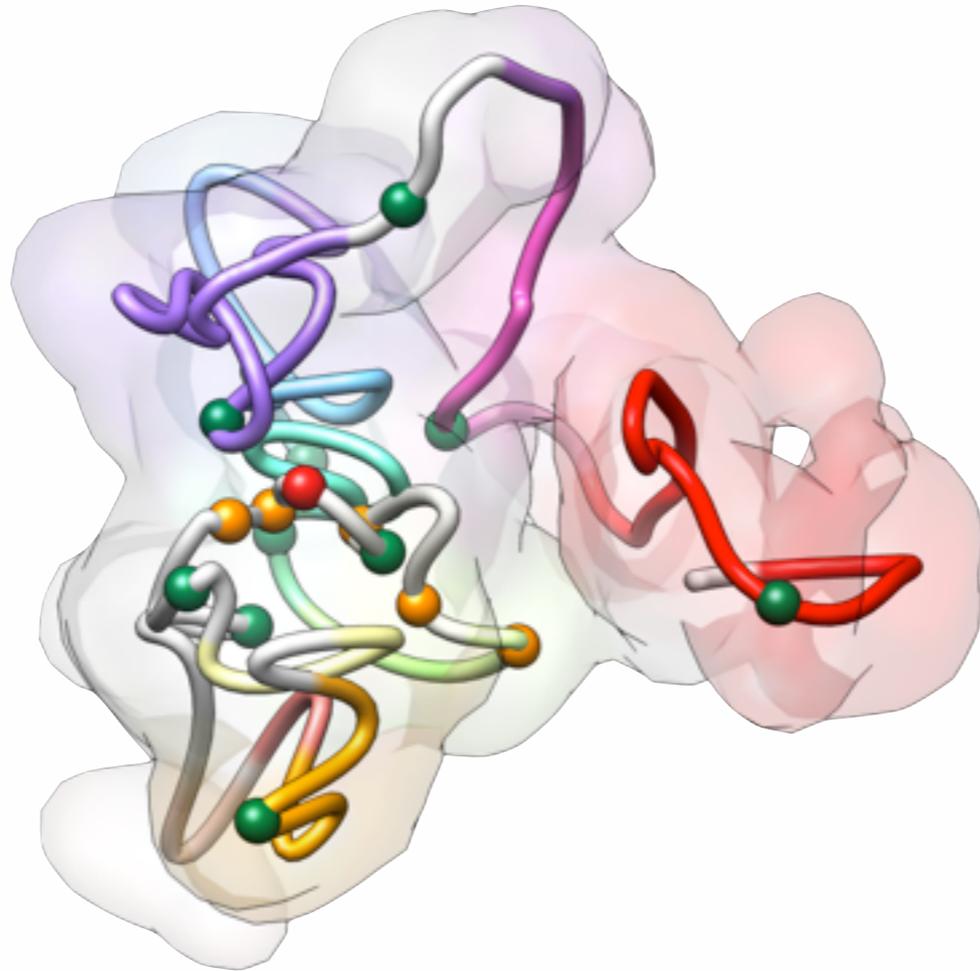
Fibrous

Membrane

Globular

Nucleic acids

DNA and RNA



Take home message

DNA and RNA

Polymers of nucleotide units

Nucleotides

Nucleobase (G,C,A,T - U)
+ sugar +phosphate

DNA

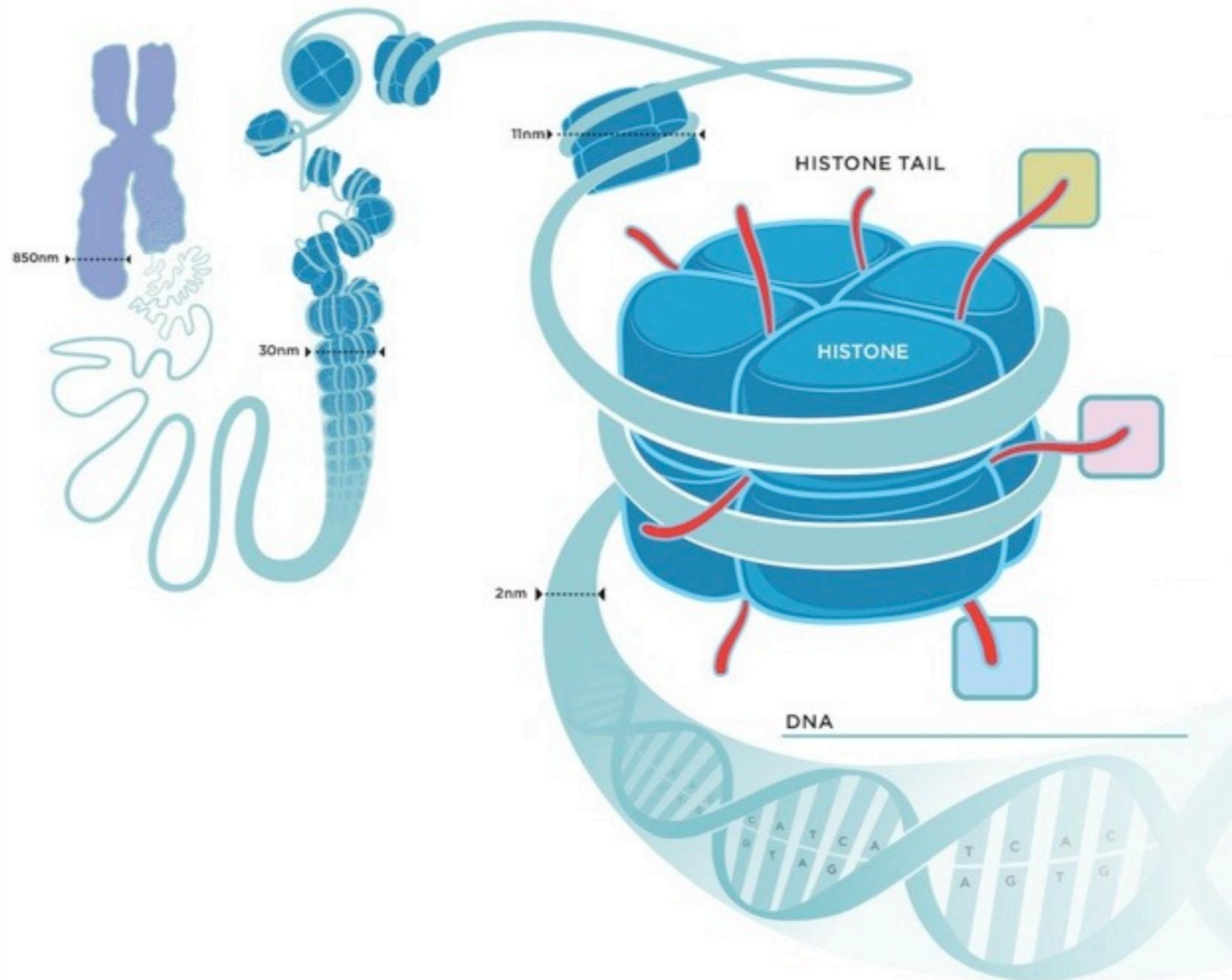
Store the genetic information

RNA

Implicated in various
biological processes

The nuclear organization of DNA

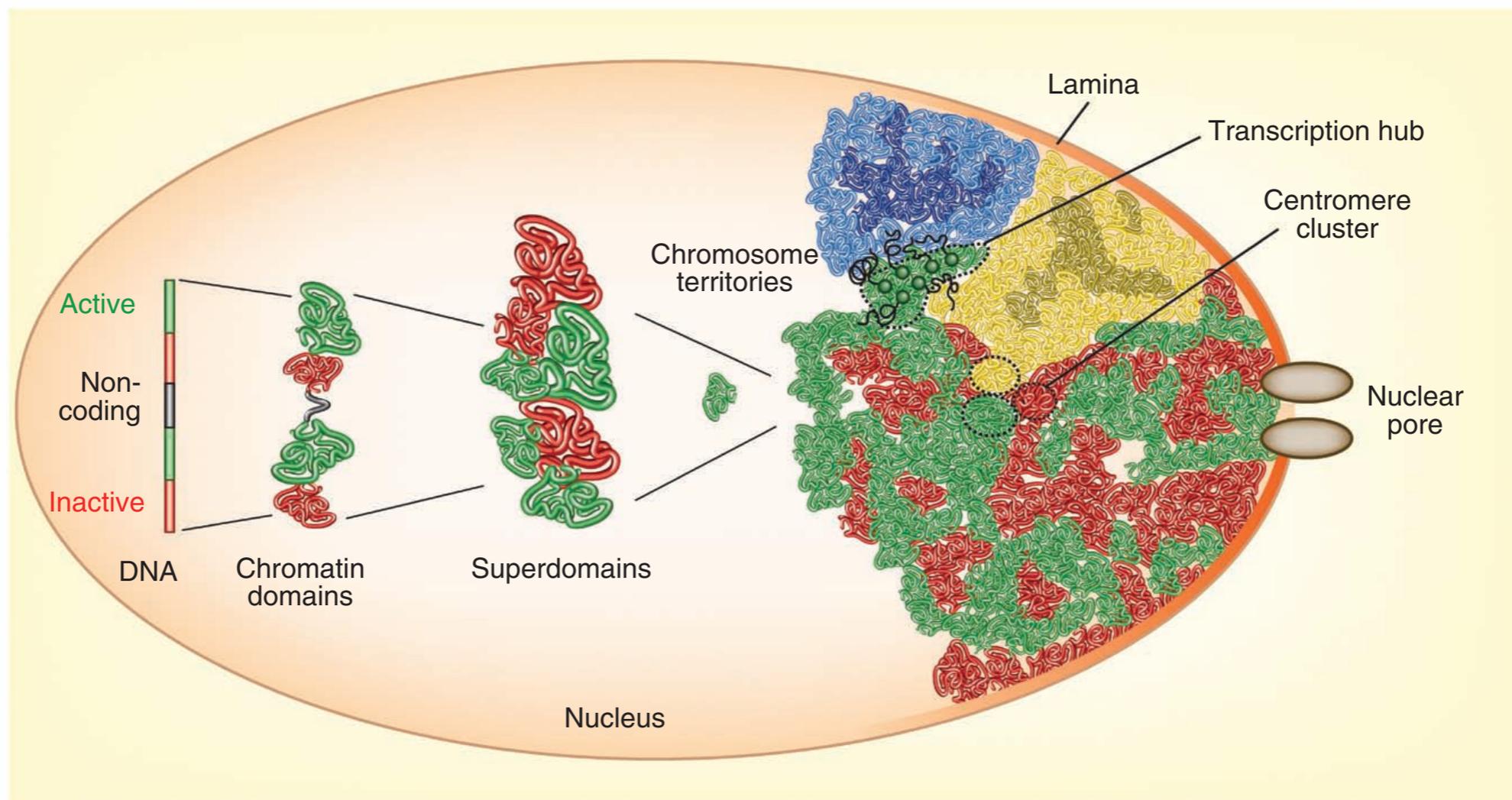
Chromosome Chromatin fibre Nucleosome



Adapted from Richard E. Ballermann, 2012

Complex genome organization

Cavalli, G. & Misteli, Nat Struct Mol Biol 20, 290–299 (2013)



Marina Corral

Take home message

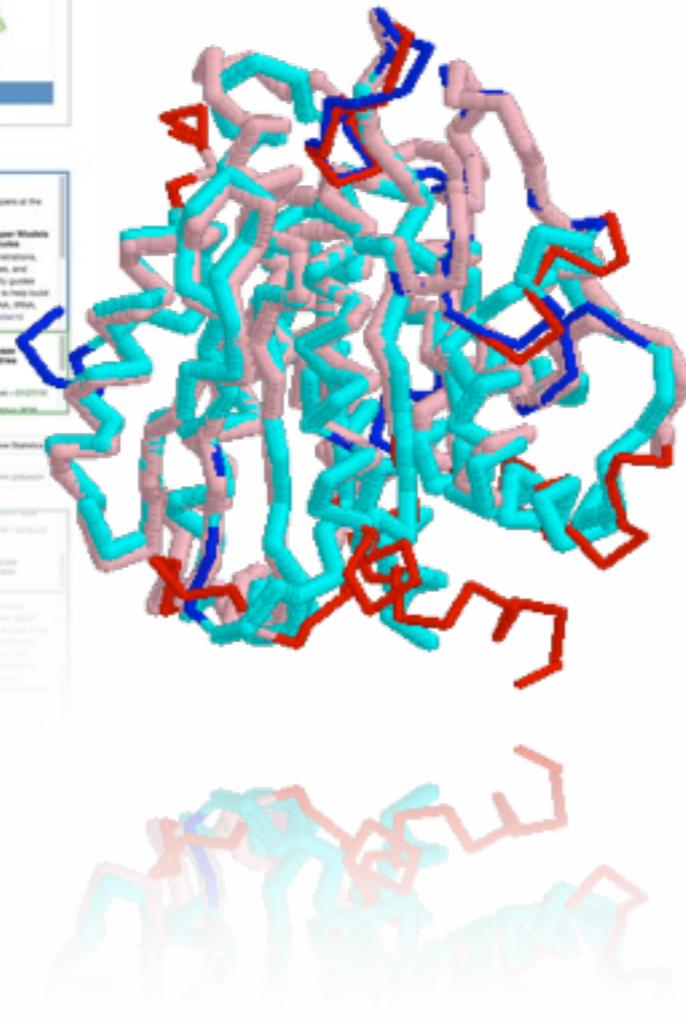
**Chromatin = DNA + (histone) proteins
+ other biomolecules**

**The genome is well organized and
hierarchically packaged**

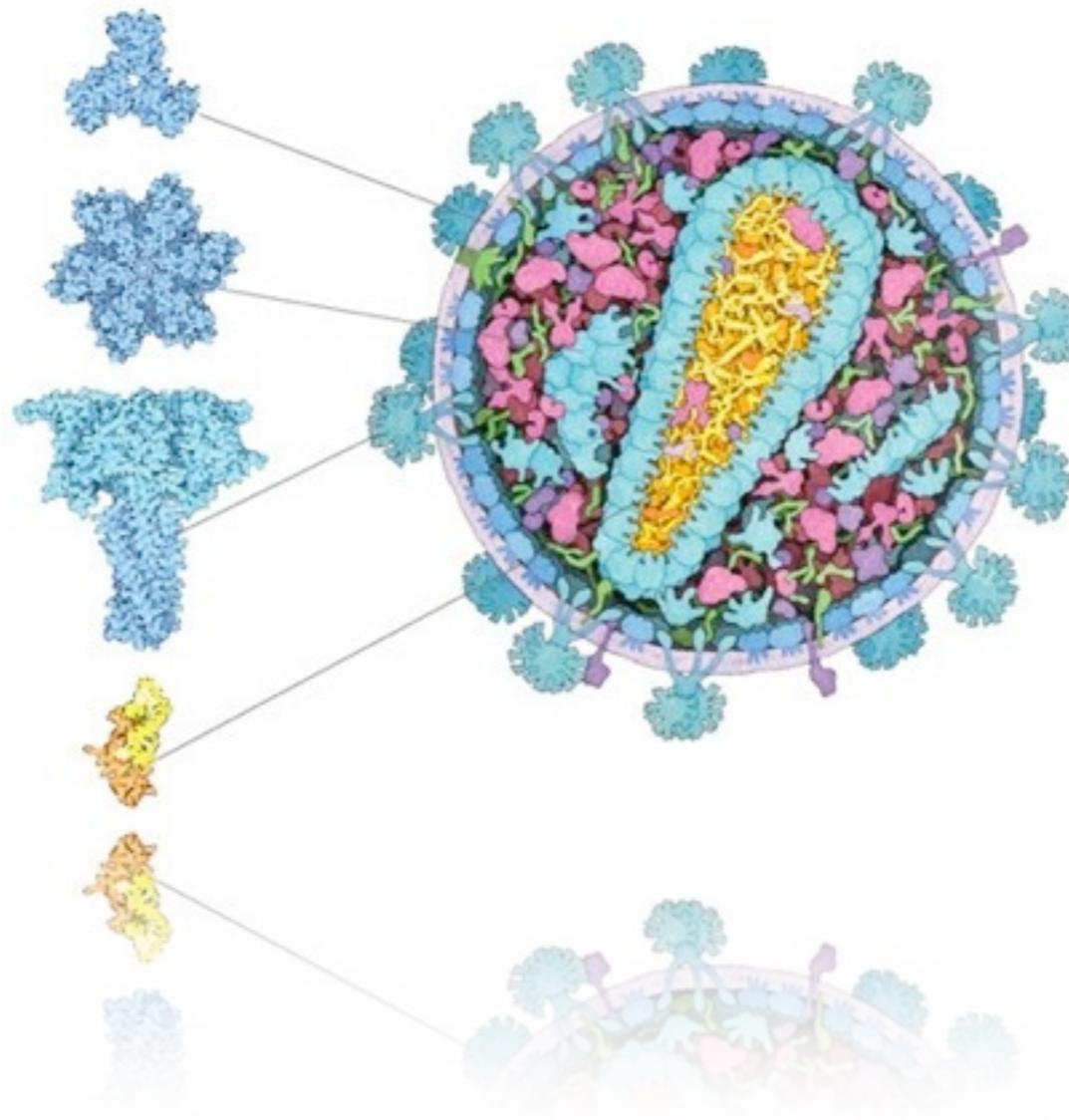
**Histone modifications affect
chromatin structure and activity**

**3C-like data measure the frequency of
interaction between distant loci**

Databases, alignments and structure classification



Known structures



The Protein Data Bank

<http://www.pdb.org>

The screenshot displays the RCSB PDB website homepage. At the top, a navigation bar includes links for Deposit, Search, Visualize, Analyze, Download, Learn, and More, along with a MyPDB Login button. Below this is the PDB logo and a search bar with the text "Search by PDB ID, author, macromolecule, sequence, or ligands". A sidebar on the left contains a "Welcome" section and a list of navigation options: Deposit, Search, Visualize, Analyze, Download, and Learn. The main content area is divided into several sections: "A Structural View of Biology" with a descriptive paragraph and a "Structure and Health Focus: Ebola Virus Proteins" section featuring a "Video Tour" and a "Molecule of the Month Article"; "March Molecule of the Month" featuring a 3D model of Phototropin; "Latest Entries" showing a 3D model of the crystal structure of D-threonine aldolase (4V15); "New Features" listing updates from December and November 2014, such as "Improved Representation of Large Structures" and "Redesigned Home Page"; and "News" with articles like "Join the RCSB PDB Team" and "Build 3D Paper Models of Biomolecules". At the bottom, a statistics bar provides key metrics: 35471 Distinct Protein Sequences, 28030 Structures of Human Sequences, and 7595 Nucleic Acid Containing Structures.

RCSB PDB Deposit - Search - Visualize - Analyze - Download - Learn - More - MyPDB Login

RCSB PDB PROTEIN DATA BANK An Information Portal to 107620 Biological Macromolecular Structures

Search by PDB ID, author, macromolecule, sequence, or ligands Go

Advanced Search | Browse by Annotations

PDB-101 PDB EMDatabank Structural Biology KnowledgeBase

Facebook Twitter YouTube LinkedIn

Welcome

- Deposit
- Search
- Visualize
- Analyze
- Download
- Learn

A Structural View of Biology

This resource is powered by the Protein Data Bank archive—information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

As a member of the wwPDB, the RCSB PDB curates and annotates PDB data. The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

Structure and Health Focus: Ebola Virus Proteins

Video Tour Molecule of the Month Article

March Molecule of the Month

Phototropin

Latest Entries

As of Tuesday, Mar 24

4V15 PDB Entry

Crystal structure of D-threonine aldolase from *Alcaligenes xylosoxidans*

View in 3D

New Features

December 2014 Release

- Improved Representation of Large Structures**
Updates to PDB Archive, Visualization, New Queries for Ribosomes and Viruses

November 2014 Release

- Redesigned Home Page**
Simple. Clean. Usable. Tools & Functions More Visible.
- Reference Management**
Quickly save references to Mendeley & Endnote.

News

Publications -

- Join the RCSB PDB Team**
We have open positions for Web Developers at the UC San Diego site. - 03/24/15
- Build 3D Paper Models of Biomolecules**
Video demonstrations, PDF templates, and related activity guides are available to help build models of DNA, tRNA, HIV Capsid, and more at PDB-101. - 03/24/15
- Changes to the Release Process for PDB Entries**
- 02/12/15
- Time-stamped Copies of the PDB Archive - 01/27/15
- Integration of Large Structures with the Main PDB

PDB at a Glance | 35471 Distinct Protein Sequences | 28030 Structures of Human Sequences | 7595 Nucleic Acid Containing Structures | More Statistics

PDB40 Symposium
More than 77000 structures in the archive

Deposition of chemical shift data mandatory

Deposition of experimental data mandatory
50000th structure is released

First release of remediated data

 **BMRB** joins wwPDB

Last PDB archive distribution by postal mail (8 DVDs)

wwPDB established by RCSB PDB, PDBe, PDBj

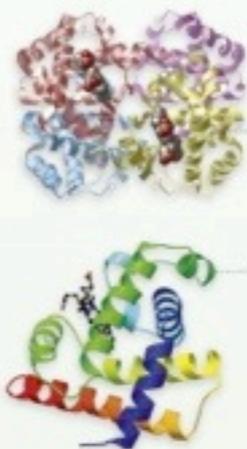
    **EMDB** established at MSD-EBI

Osaka University opens a PDB data deposition center

10000th structure is released

PDB moves to **RCSB PDB**

MSD at the EBI becomes a deposition center for PDB data



Early structures include
carboxypeptidase
chymotrypsin
cytochrome b5
hemoglobin
lactate dehydrogenase
myoglobin
rubredoxin
subtilisin
trypsin inhibitor

PDB established

2011

2010

2009

2008

2007

2006

2005

2003

2002

2000

1999

1998

1997

1991

1989

1988

1982

1981

1979

1973

1971

Nobel prize awarded to V. Ramakrishnan, T.A. Steitz, A.E. Yonath for *Studies of the structure and function of the ribosome*

Nobel prize awarded to R.D. Kornberg for *Studies of the molecular basis of eukaryotic transcription*

Nobel prize awarded to R. MacKinnon for *Potassium channels*

Nobel prize awarded to K. Wüthrich for *Development of NMR spectroscopy for determining the 3D structure of biological macromolecules in solution*

First **ribosome structures** determined (Ban *et al.*, 2000; Carter *et al.*, 2000; Schluenzen *et al.*, 2000)

Nobel prize awarded to P.D. Boyer, J.E. Walker, J.C. Skou for *Elucidation of the enzymatic mechanism underlying the synthesis of adenosine triphosphate (ATP) and discovery of an ion-transporting enzyme*

First **EM entry** released in PDB: bacteriorhodopsin (Henderson *et al.*, 1990)

First **NMR entry** released in PDB: protein BDS-I (Driscoll *et al.*, 1989)

Nobel prize awarded to J. Deisenhofer, R. Huber, H. Michel for *The determination of the 3D structure of a photosynthetic reaction centre*

Nobel prize awarded to A. Klug for *Development of crystallographic electron microscopy and discovery of the structure of biologically important nucleic acid-protein complexes*

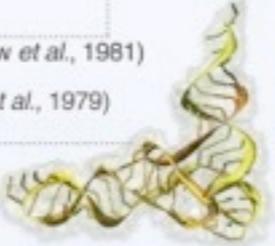
First **B-DNA** structure determined (Drew *et al.*, 1981)

First **DNA (Z-DNA)** determined (Wang *et al.*, 1979)

First **tRNA** structures determined (Kim *et al.*, 1973; Robertus *et al.*, 1974)

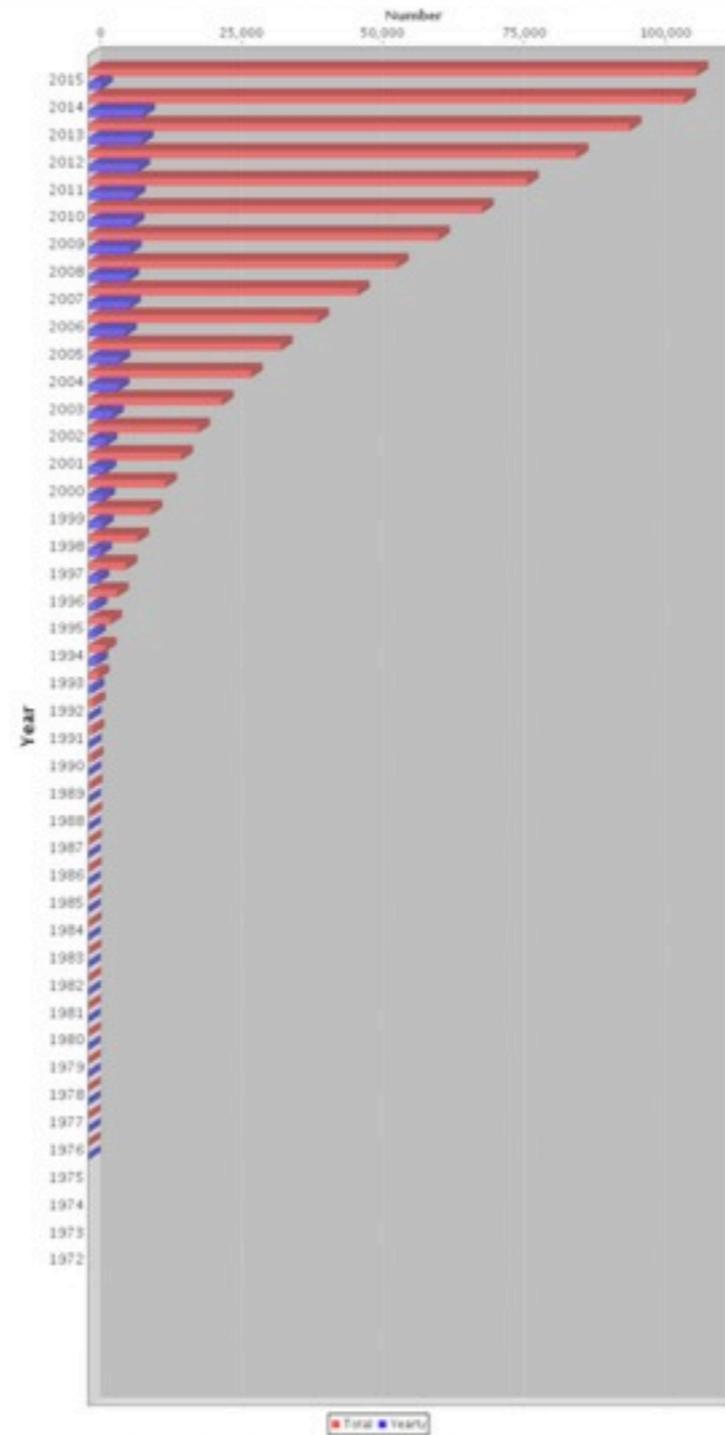






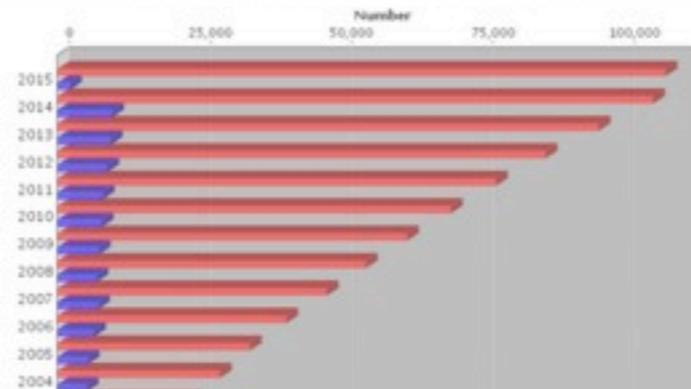
Yearly growth of total structures

<http://www.pdb.org>



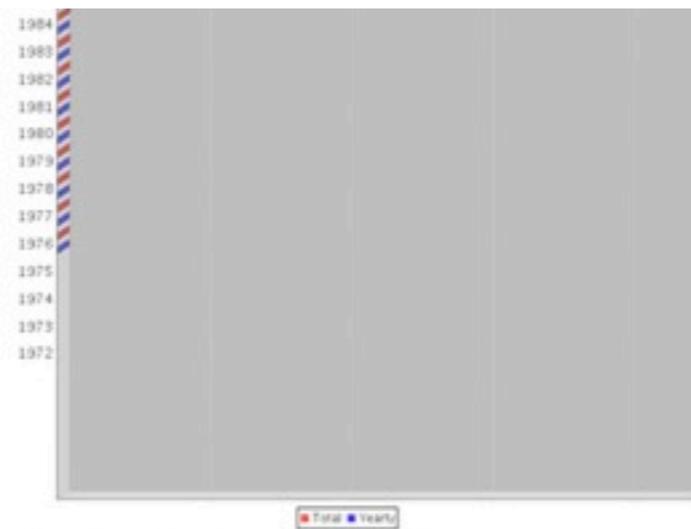
Yearly growth of total structures

<http://www.pdb.org>



PDB Current Holdings Breakdown

Exp.Method	Proteins	Nucleic Acids	Protein/NA Complexes	Other	Total
X-RAY	89684	1612	4426	4	95726
NMR	9534	1112	224	8	10878
ELECTRON MICROSCOPY	549	29	177	0	755
HYBRID	68	3	2	1	74
other	164	4	6	13	187
Total	99999	2760	4835	26	107620



PDB format

<http://www.pdb.org>

```
HEADER      EXTRACELLULAR MATRIX                22-JAN-98   1A3I
TITLE       X-RAY CRYSTALLOGRAPHIC DETERMINATION OF A COLLAGEN-LIKE
TITLE       2 PEPTIDE WITH THE REPEATING SEQUENCE (PRO-PRO-GLY)
...
EXPDTA      X-RAY DIFFRACTION
AUTHOR      R.Z.KRAMER,L.VITAGLIANO,J.BELLA,R.BERISIO,L.MAZZARELLA,
AUTHOR      2 B.BRODSKY,A.ZAGARI,H.M.BERMAN
...
REMARK 350 BIOMOLECULE: 1
REMARK 350 APPLY THE FOLLOWING TO CHAINS: A, B, C
REMARK 350   BIOMT1   1  1.000000  0.000000  0.000000          0.00000
REMARK 350   BIOMT2   1  0.000000  1.000000  0.000000          0.00000
...
SEQRES      1  A      9  PRO PRO GLY PRO PRO GLY PRO PRO GLY
SEQRES      1  B      6  PRO PRO GLY PRO PRO GLY
SEQRES      1  C      6  PRO PRO GLY PRO PRO GLY
...
ATOM        1  N      PRO A    1          8.316  21.206  21.530  1.00 17.44      N
ATOM        2  CA     PRO A    1          7.608  20.729  20.336  1.00 17.44      C
ATOM        3  C      PRO A    1          8.487  20.707  19.092  1.00 17.44      C
ATOM        4  O      PRO A    1          9.466  21.457  19.005  1.00 17.44      O
ATOM        5  CB     PRO A    1          6.460  21.723  20.211  1.00 22.26      C
...
HETATM     130  C      ACY     401         3.682  22.541  11.236  1.00 21.19      C
HETATM     131  O      ACY     401         2.807  23.097  10.553  1.00 21.19      O
HETATM     132  OXT   ACY     401         4.306  23.101  12.291  1.00 21.19      O
...

```

PDB format

<http://www.pdb.org>

```
ATOM      1  N   GLY  A   1      15.740  11.178 -11.733  1.00  0.00           N
ATOM      2  CA  GLY  A   1      15.234  10.462 -10.556  1.00  0.00           C
ATOM      3  C   GLY  A   1      16.284   9.483  -9.998  1.00  0.00           C
ATOM      4  O   GLY  A   1      17.150   8.979 -10.709  1.00  0.00           O
ATOM      5  N   LEU  A   2      16.122   9.240  -8.705  1.00  0.00           N
ATOM      6  CA  LEU  A   2      16.803   8.164  -7.994  1.00  0.00           C
ATOM      7  C   LEU  A   2      17.902   7.481  -8.831  1.00  0.00           C
ATOM      8  O   LEU  A   2      19.057   7.424  -8.402  1.00  0.00           O
ATOM      9  CB  LEU  A   2      15.755   7.101  -7.594  1.00  0.00           C
ATOM     10  CG  LEU  A   2      14.565   7.724  -6.856  1.00  0.00           C
ATOM     11  CD1 LEU  A   2      14.958   8.214  -5.453  1.00  0.00           C
ATOM     12  CD2 LEU  A   2      13.894   8.850  -7.657  1.00  0.00           C
ATOM     13  N   SER  A   3      17.505   6.971  -9.986  1.00  0.00           N
ATOM     14  CA  SER  A   3      18.416   6.404 -10.972  1.00  0.00           C
ATOM     15  C   SER  A   3      19.535   5.484 -10.470  1.00  0.00           C
ATOM     16  O   SER  A   3      19.869   5.398  -9.293  1.00  0.00           O
ATOM     17  CB  SER  A   3      19.079   7.578 -11.738  1.00  0.00           C
ATOM     18  OG  SER  A   3      19.875   8.293 -10.785  1.00  0.00           O
ATOM     19  N   ASP  A   4      20.127   4.842 -11.478  1.00  0.00           N
ATOM     20  CA  ASP  A   4      21.268   3.953 -11.276  1.00  0.00           C
ATOM     21  C   ASP  A   4      22.229   4.485 -10.197  1.00  0.00           C
ATOM     22  O   ASP  A   4      22.347   3.897  -9.121  1.00  0.00           O
ATOM     23  CB  ASP  A   4      22.048   3.770 -12.587  1.00  0.00           C
ATOM     24  CG  ASP  A   4      21.138   3.738 -13.824  1.00  0.00           C
ATOM     25  OD1 ASP  A   4      20.964   4.838 -14.416  1.00  0.00           O
ATOM     26  OD2 ASP  A   4      20.667   2.612 -14.138  1.00  0.00           O
ATOM     27  N   GLY  A   5      22.907   5.579 -10.530  1.00  0.00           N
ATOM     28  CA  GLY  A   5      23.875   6.166  -9.594  1.00  0.00           C
ATOM     29  C   GLY  A   5      23.310   6.210  -8.163  1.00  0.00           C
ATOM     30  O   GLY  A   5      23.845   5.573  -7.256  1.00  0.00           O
ATOM     31  N   GLU  A   6      22.228   6.967  -8.038  1.00  0.00           N
ATOM     32  CA  GLU  A   6      21.545   7.090  -6.742  1.00  0.00           C
ATOM     33  C   GLU  A   6      21.419   5.695  -6.103  1.00  0.00           C
ATOM     34  O   GLU  A   6      21.902   5.451  -4.997  1.00  0.00           O
ATOM     35  CB  GLU  A   6      20.162   7.735  -6.956  1.00  0.00           C
ATOM     36  CG  GLU  A   6      19.493   8.120  -5.628  1.00  0.00           C
ATOM     37  CD  GLU  A   6      18.554   9.328  -5.804  1.00  0.00           C
ATOM     38  OE1 GLU  A   6      17.794   9.578  -4.827  1.00  0.00           O
ATOM     39  OE2 GLU  A   6      18.634   9.959  -6.891  1.00  0.00           O
ATOM     40  N   TRP  A   7      20.694   4.832  -6.811  1.00  0.00           N
ATOM     41  CA  TRP  A   7      20.454   3.477  -6.298  1.00  0.00           C
ATOM     42  C   TRP  A   7      21.756   2.847  -5.765  1.00  0.00           C
ATOM     43  O   TRP  A   7      21.824   2.453  -4.599  1.00  0.00           O
ATOM     44  CB  TRP  A   7      19.878   2.574  -7.407  1.00  0.00           C
ATOM     45  CG  TRP  A   7      18.589   3.146  -7.920  1.00  0.00           C
ATOM     46  CD1 TRP  A   7      18.358   3.602  -9.158  1.00  0.00           C
ATOM     47  CD2 TRP  A   7      17.391   3.325  -7.161  1.00  0.00           C
ATOM     48  NE1 TRP  A   7      17.049   4.030  -9.246  1.00  0.00           N
ATOM     49  CE2 TRP  A   7      16.475   3.892  -8.055  1.00  0.00           C
ATOM     50  CE3 TRP  A   7      17.004   3.075  -5.840  1.00  0.00           C
```

PDB search

<http://www.pdb.org>

The screenshot displays the RCSB PDB website homepage. At the top, a navigation bar includes links for Deposit, Search, Visualize, Analyze, Download, Learn, and More, along with a MyPDB Login button. Below this is the PDB logo and a search bar with the text "Search by PDB ID, author, macromolecule, sequence, or ligands". A sidebar on the left contains a "Welcome" section and a list of navigation options: Deposit, Search, Visualize, Analyze, Download, and Learn. The main content area is divided into several sections: "A Structural View of Biology" with introductory text and a "Structure and Health Focus: Ebola Virus Proteins" section featuring a video tour and a molecule of the month article; "March Molecule of the Month" featuring a 3D model of Phototropin; "Latest Entries" showing the crystal structure of D-threonine aldolase (4V15); "New Features" listing updates from December and November 2014; and "News" with announcements about joining the team, building 3D paper models, and changes to the release process. At the bottom, a statistics bar provides key metrics: 35471 Distinct Protein Sequences, 28030 Structures of Human Sequences, and 7595 Nucleic Acid Containing Structures.

RCSB PDB Deposit - Search - Visualize - Analyze - Download - Learn - More - MyPDB Login

RCSB PDB An Information Portal to 107620 Biological Macromolecular Structures

Search by PDB ID, author, macromolecule, sequence, or ligands Go

Advanced Search | Browse by Annotations

PD101 PDB EMDatabank Structural Biology KnowledgeBase

Facebook Twitter YouTube LinkedIn

Welcome

- Deposit
- Search
- Visualize
- Analyze
- Download
- Learn

A Structural View of Biology

This resource is powered by the Protein Data Bank archive—information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

As a member of the wwPDB, the RCSB PDB curates and annotates PDB data. The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

Structure and Health Focus: Ebola Virus Proteins

Video Tour | Molecule of the Month Article

March Molecule of the Month

Phototropin

Latest Entries

As of Tuesday, Mar 24

4V15 PDB Entry

Crystal structure of D-threonine aldolase from *Alcaligenes xylosoxidans*

View in 3D

New Features

December 2014 Release

- Improved Representation of Large Structures**
Updates to PDB Archive, Visualization, New Queries for Ribosomes and Viruses

November 2014 Release

- Redesigned Home Page**
Simple, Clean, Usable, Tools & Functions More Visible.
- Reference Management**
Quickly save references to Mendeley & Endnote.

News

Join the RCSB PDB Team

We have open positions for Web Developers at the UC San Diego site. - 03/24/15

Build 3D Paper Models of Biomolecules

Video demonstrations, PDF templates, and related activity guides are available to help build models of DNA, tRNA, HIV Capsid, and more at PDB-101. - 03/24/15

Changes to the Release Process for PDB Entries

- 02/12/15

Time-stamped Copies of the PDB Archive - 01/27/15

Integration of Large Structures with the Main PDB

PDB at a Glance | 35471 Distinct Protein Sequences | 28030 Structures of Human Sequences | 7595 Nucleic Acid Containing Structures | More Statistics

PDB search

<http://www.pdb.org>

The screenshot shows the RCSB PDB website interface. At the top, there is a navigation bar with links for Deposit, Search, Visualize, Analyze, Download, Learn, and More, along with a MyPDB Login button. Below this is the PDB logo and a search bar containing the text '4CS6'. The search bar is highlighted with a red box. To the right of the search bar is a 'Go' button. Below the search bar, there are several sections: 'A Structural View of Biology', 'March Molecule of the Month' (featuring Phototropin), 'Structure and Health Focus: Ebola Virus Proteins', 'Latest Entries' (featuring entry 4V15), 'New Features' (listing updates for December and November 2014), and 'News' (including 'Join the RCSB PDB Team' and 'Changes to the Release Process for PDB Entries'). At the bottom, there is a statistics bar showing '35471 Distinct Protein Sequences', '28030 Structures of Human Sequences', and '7595 Nucleic Acid Containing Structures'.

RCSB PDB Deposit - Search - Visualize - Analyze - Download - Learn - More - MyPDB Login

RCSB PDB An Information Portal to 107620 Biological Macromolecular Structures

4CS6

Advanced Search | Browse by Annotations

Structure and Health Focus: Ebola Virus Proteins

March Molecule of the Month

Phototropin

Latest Entries As of Tuesday, Mar 24

4V15 PDB Entry

Crystal structure of D-threonine aldolase from *Alcaligenes xylosoxidans*

View in 3D

New Features

December 2014 Release

Improved Representation of Large Structures

Updates to PDB Archive, Visualization, New Queries for Ribosomes and Viruses

November 2014 Release

Redesigned Home Page

Simple. Clean. Usable. Tools & Functions More Visible.

Reference Management

Quickly save references to Mendely & Endnote.

News Publications -

Join the RCSB PDB Team

We have open positions for Web Developers at the UC San Diego site. - 03/24/15

Build 3D Paper Models of Biomolecules

Video demonstrations, PDF templates, and related activity guides are available to help build models of DNA, tRNA, HIV Capsid, and more at PDB-101. - 03/24/15

Changes to the Release Process for PDB Entries

- 02/12/15

Time-stamped Copies of the PDB Archive - 01/27/15

Integration of Large Structures with the Main PDB

PDB at a Glance 35471 Distinct Protein Sequences | 28030 Structures of Human Sequences | 7595 Nucleic Acid Containing Structures | More Statistics

PDB search

<http://www.pdb.org>

Summary 3D View Sequence Annotations Seq. Similarity 3D Similarity Literature Biol. & Chem. Methods Links

Crystal structure of AadA - an aminoglycoside adenylyltransferase

4CS6 [Display Files](#) [Download Files](#) [Download Citation](#)

DOI:10.2210/pdb4cs6/pdb

Primary Citation

Crystal Structure of AadA-an Aminoglycoside Adenylyltransferase

Chen, Y.P., Nasvall, J.P., Andersson, D.I.P., Selmer, M.P.

Journal: To be Published

PubMed ID is not available

Molecular Description Hide

Classification: Transferase

Structure Weight: 29591.20

Molecule: AMINOGLYCOSIDE ADENYLTRANSFERASE

Polymer: 1 **Type:** protein **Length:** 264

Chains: A

Fragment: NUCLEOTIDYLTRANSFERASE DOMAIN AND ALPHA-HELICAL DOMAIN

Organism: Salmonella enterica subsp. enterica serovar Typhimurium str. LT2

UniProtKB: [Protein Feature View](#) | [Search PDB](#) | [Q6ZPX9](#)

Structure Validation Hide

View the full validation report

Metric	Percentile Ranks	Value
Rfree		0.253
Clashscore		8
Ramachandran outliers		0
Sidechain outliers		5.5%
RSRZ outliers		0

None Better

█ Percentile relative to all X-ray structures

█ Percentile relative to X-ray structures of similar resolution

MolProbity Ramachandran Plot

[Download Ramachandran Plot PDF \(from MolProbity\)](#)

Source Hide

Polymer: 1

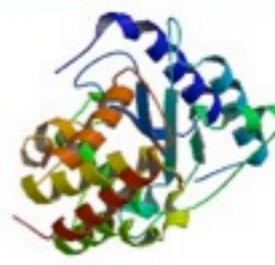
Scientific Name: Salmonella enterica subsp. enterica serovar typhimurium [Taxonomy](#)

Expression System: Escherichia coli

Modified Residues Hide

Identifier	Formula	Parent	Type
CSS Search	C ₃ H ₇ N O ₂ S ₂	CYS	iPeptideLinking

Biological Assembly Hide



[3D View](#) [More Images...](#)

Biological assembly 1 generated by PISA (software)

Downloadable viewers:

[Simple Viewer](#) [Protein Workshop](#) [Kiosk Viewer](#)

MyPDB Personal Annotations Hide

To save personal annotations, please login to your MyPDB account.

Deposition Summary Hide

Authors: Chen, Y.P., Nasvall, J.P., Andersson, D.I.P., Selmer, M.P.

Deposition: 2014-03-05

Release: 2015-03-25

Revision History Hide

No revisions since initial release

Experimental Details Hide

Method: X-RAY DIFFRACTION

Exp. Data:

[Structure Factors](#)

Resolution(Å): 2.50

R-Value: 0.196 (obs.)

R-Free: 0.252

Space Group: P 2₁ 2₁ 2₁

Unit Cell:

Length [Å]	Angles [°]
a = 48.66	α = 90.00
b = 58.64	β = 90.00
c = 104.59	γ = 90.00

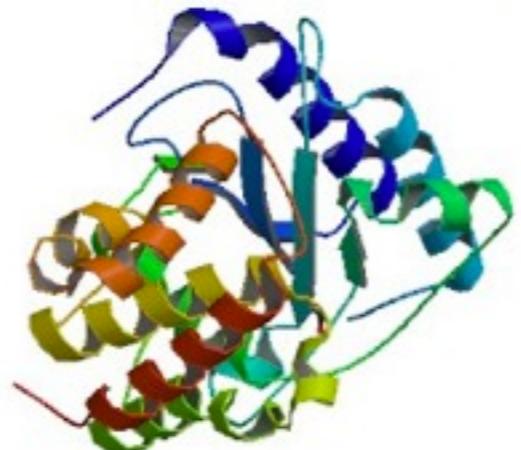
Data in orange boxes are gathered from external resources (when available).

[Reset Layout](#)

PDB search

<http://www.pdb.org>

Biological Assembly ?



3D View [More Images...](#)

Biological assembly 1 generated by PISA (software)

Downloadable viewers:

[Simple Viewer](#) [Protein Workshop](#)

[Kiosk Viewer](#)

Experimental Details Hide

Method: X-RAY DIFFRACTION

Exp. Data:

[Structure Factors](#)

Resolution[Å]: 2.50

R-Value: 0.196 (obs.)

R-Free: 0.252

Space Group: P 2₁ 2 2₁

Unit Cell:

Length [Å]	Angles [°]
a = 48.66	α = 90.00
b = 58.64	β = 90.00
c = 104.59	γ = 90.00

PDB search

<http://www.pdb.org>

```
HEADER      TRANSFERASE                               05-MAR-14  4CS6
TITLE       CRYSTAL STRUCTURE OF AADA - AN AMINOGLYCOSIDE ADENYLTRANSFERASE
COMPND     MOL_ID: 1;
COMPND     2 MOLECULE: AMINOGLYCOSIDE ADENYLTRANSFERASE;
COMPND     3 CHAIN: A;
COMPND     4 FRAGMENT: NUCLEOTIDYLTRANSFERASE DOMAIN AND ALPHA-HELICAL DOMAIN;
COMPND     5 ENGINEERED: YES
SOURCE     MOL_ID: 1;
SOURCE     2 ORGANISM_SCIENTIFIC: SALMONELLA ENTERICA SUBSP. ENTERICA SEROVAR
SOURCE     3 TYPHIMURIUM STR. LT2;
SOURCE     4 ORGANISM_TAXID: 99287;
SOURCE     5 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE     6 EXPRESSION_SYSTEM_TAXID: 469008;
SOURCE     7 EXPRESSION_SYSTEM_STRAIN: BL21(DE3);
SOURCE     8 EXPRESSION_SYSTEM_VARIANT: STAR
KEYWDS     TRANSFERASE, AMINOGLYCOSIDE ADENYLTRANSFERASE, ANT(3'')
EXPDTA     X-RAY DIFFRACTION
AUTHOR     Y.CHEN,J.NASVALL,D.I.ANDERSSON,M.SELMER
REVDAT     1 25-MAR-15 4CS6 0
JRNL       AUTH  Y.CHEN,J.NASVALL,D.I.ANDERSSON,M.SELMER
JRNL       TITL  CRYSTAL STRUCTURE OF AADA-AN AMINOGLYCOSIDE
JRNL       TITL 2 ADENYLTRANSFERASE
JRNL       REF  TO BE PUBLISHED
JRNL       REFN

REMARK     2
REMARK     2 RESOLUTION.      2.50 ANGSTROMS.
REMARK     3
REMARK     3 REFINEMENT.
REMARK     3   PROGRAM      : PHENIX (PHENIX.REFINE)
REMARK     3   AUTHORS      : PAUL ADAMS,PAVEL AFONINE,VICENT CHEN,IAN
REMARK     3                 : DAVIS,KRESHNA GOPAL,RALF GROSSE-KUNSTLEVE,
REMARK     3                 : LI-WEI HUNG,ROBERT IMMORMINO,TOM IOERGER,
REMARK     3                 : AIRLIE MCCOY,ERIK MCKEE,NIGEL MORIARTY,
REMARK     3                 : REETAL PAI,RANDY READ,JANE RICHARDSON,
REMARK     3                 : DAVID RICHARDSON,TOD ROMO,JIM SACCHETTINI,
REMARK     3                 : NICHOLAS SAUTER,JACOB SMITH,LAURENT
REMARK     3                 : STORONI,TOM TERWILLIGER,PETER ZWART
REMARK     3
REMARK     3   REFINEMENT TARGET : ML
REMARK     3
REMARK     3   DATA USED IN REFINEMENT.
REMARK     3   RESOLUTION RANGE HIGH (ANGSTROMS) : 2.502
REMARK     3   RESOLUTION RANGE LOW  (ANGSTROMS) : 39.029
REMARK     3   MIN(FOBS/SIGMA_FOBS)          : 1.34
REMARK     3   COMPLETENESS FOR RANGE         (%) : 99.66
REMARK     3   NUMBER OF REFLECTIONS          : 10816
REMARK     3
REMARK     3   FIT TO DATA USED IN REFINEMENT.
REMARK     3   R VALUE          (WORKING + TEST SET) : 0.1962
REMARK     3   R VALUE          (WORKING SET)       : 0.1935
REMARK     3   FREE R VALUE          : 0.2521
REMARK     3   FREE R VALUE TEST SET SIZE (%)      : 4.8
REMARK     3   FREE R VALUE TEST SET COUNT         : 517
REMARK     3
REMARK     3   FIT TO DATA USED IN REFINEMENT (IN BINS).
REMARK     3   BIN  RESOLUTION RANGE  COMPL.    NWORK  NFREE   RWORK  RFREE
REMARK     3   1  39.0340 - 3.9704    0.99     2685   128   0.1698  0.2001
REMARK     3   2  3.9704 - 3.1519    1.00     2571   120   0.1898  0.2684
REMARK     3   3  3.1519 - 2.7536    1.00     2543   126   0.2505  0.3378
REMARK     3   4  2.7536 - 2.5019    0.99     2500   143   0.2754  0.3924
REMARK     3
REMARK     3   BULK SOLVENT MODELLING.
REMARK     3   METHOD USED          : FLAT BULK SOLVENT MODEL
REMARK     3   SOLVENT RADIUS      : 1.11
REMARK     3   SHRINKAGE RADIUS    : 0.90
REMARK     3   K_SOL               : NULL
REMARK     3   B_SOL               : NULL
```

PDB search

<http://www.pdb.org>

Summary **3D View** Sequence Annotations Seq. Similarity 3D Similarity Literature Biol. & Chem. Methods Links

Crystal structure of AadA - an aminoglycoside adenylyltransferase **4CS6** [Display Files](#) [Download Files](#) [Download Citation](#)

NOTE: Use your mouse to drag, rotate, and zoom in and out of the structure. ?



Structure Details ?
Structure

Select Display Mode ?

Biological assembly 1 generated by PISA ?

PDB advanced search

<http://www.pdb.org>

Search Options

Advanced Search

The Advanced Search interface enables queries by specific categories. Queries can be combined with AND or OR to construct complex searches.

Search categories:

- IDs and keywords
- Structure annotation
- Deposition
- Structure features
- Sequence features
- Chemical components
- Biology
- Methods
- Publication

Start Advanced Search

Advanced Search Interface

Macromolecule Name: Actin (Result Count: 319 PDB Entries (Structures))

AND

Release Date: Search by the date the structure was released (e.g. between 2008-01-01 and 2008-12-31)

Between: 2014-01-01 and 2014-07-18 (Result Count)

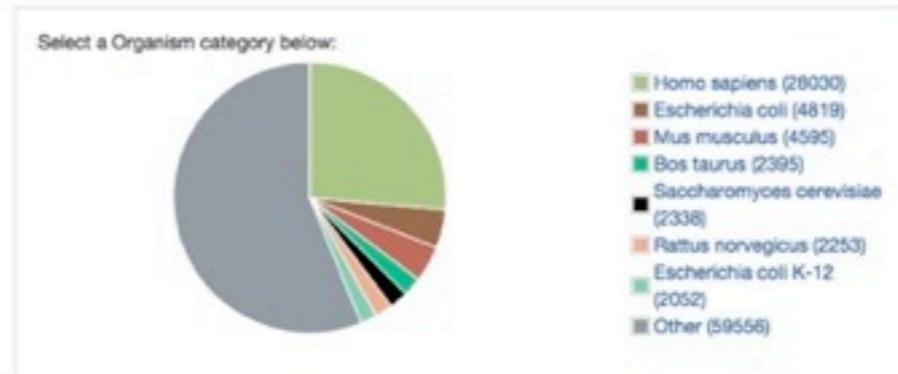
Retrieve only representatives at 100% sequence identity

Match all of the above conditions.

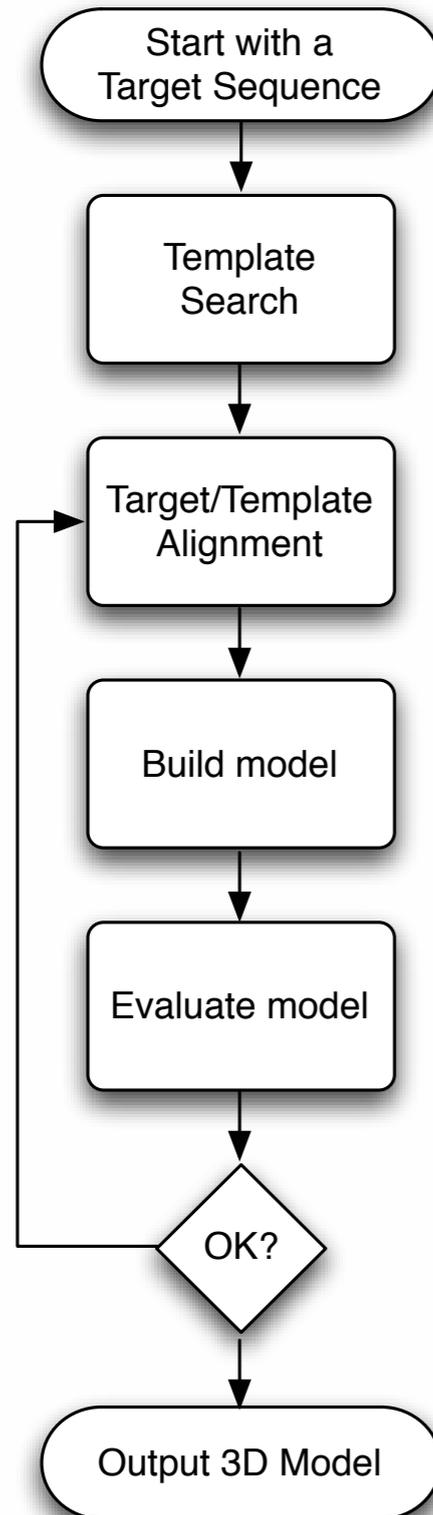
Results: Structures Clear All Parameters Submit Query

Explore the PDB Archive

- Organism
- Experimental Method
- Release Date
- Enzyme Classification
- Protein Symmetry
- Membrane Proteins
- Taxonomy
- X-ray Resolution
- Polymer Type
- SCOP Classification
- Protein Stoichiometry



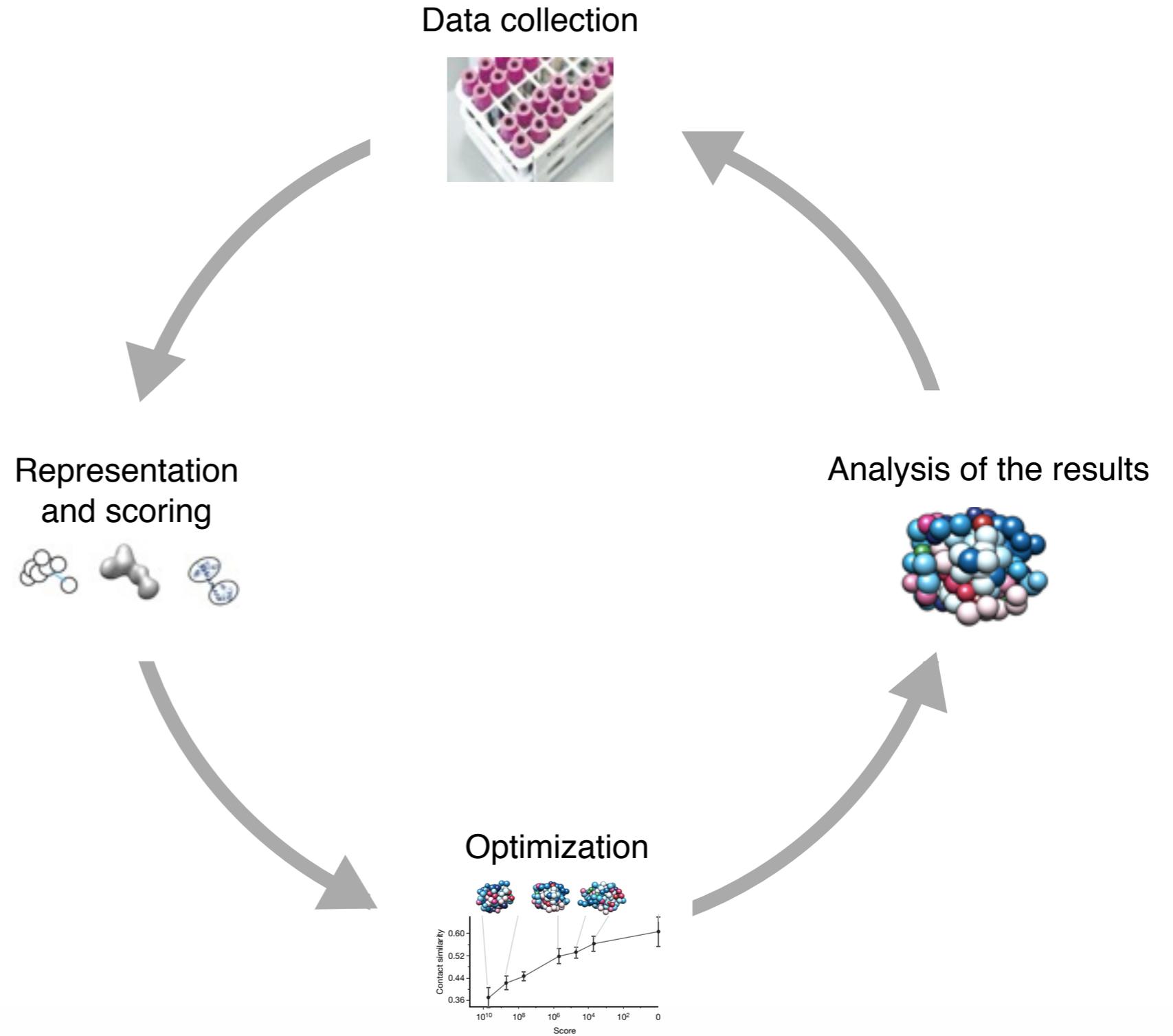
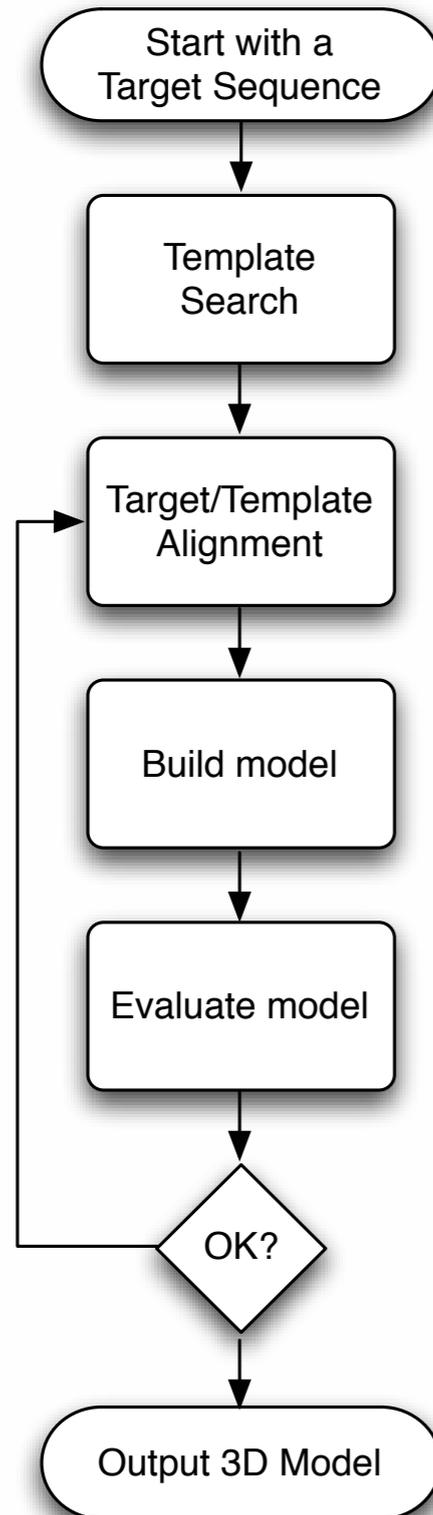
Alignments



As in any other bioinformatics problem...

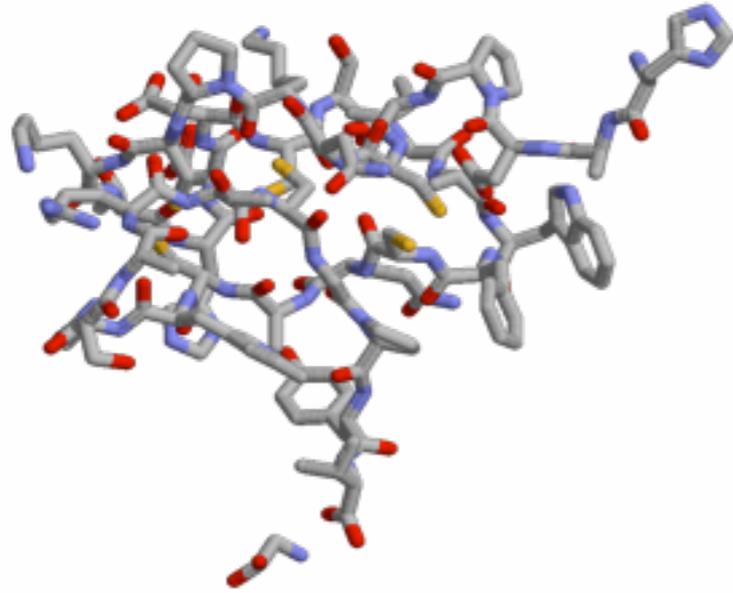
- 1. Representation**
- 2. Scoring**
- 3. Optimization**

Alignments

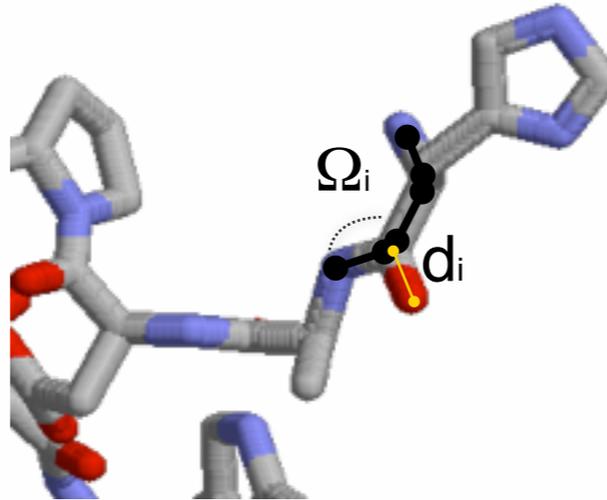


Structures

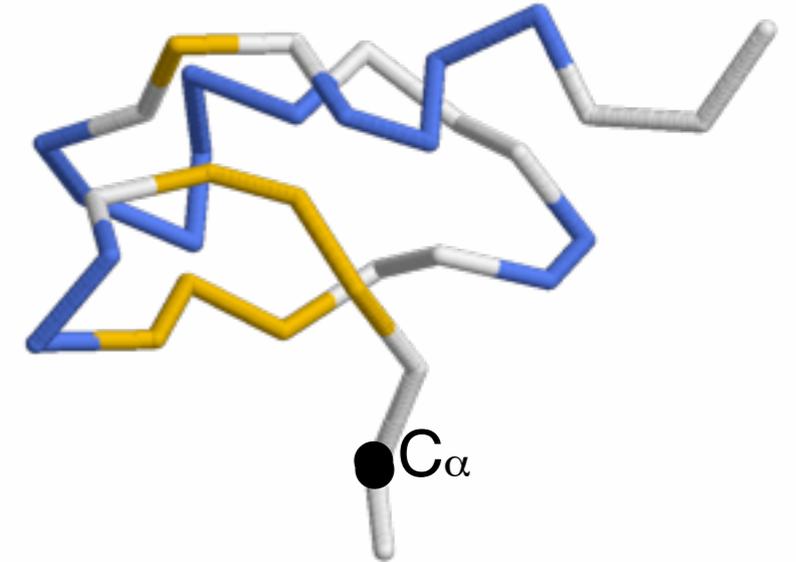
Representation



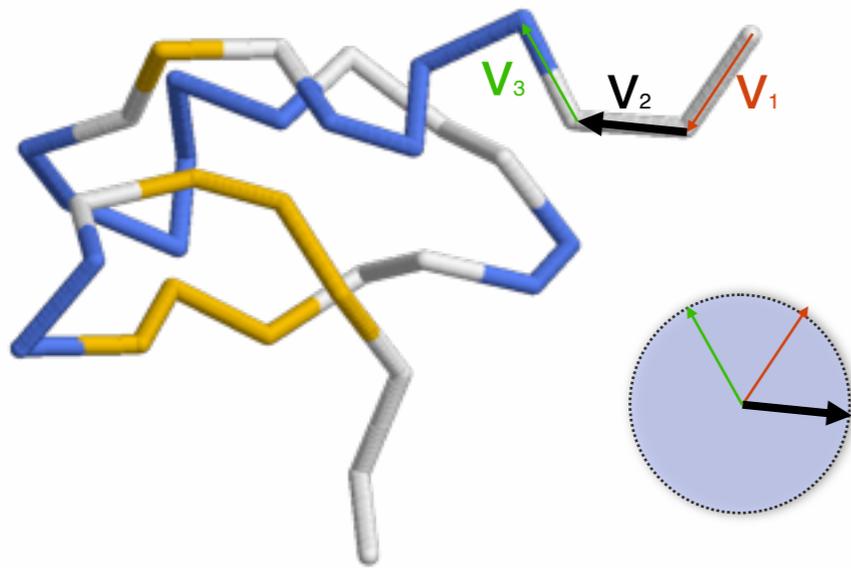
All atoms and coordinates



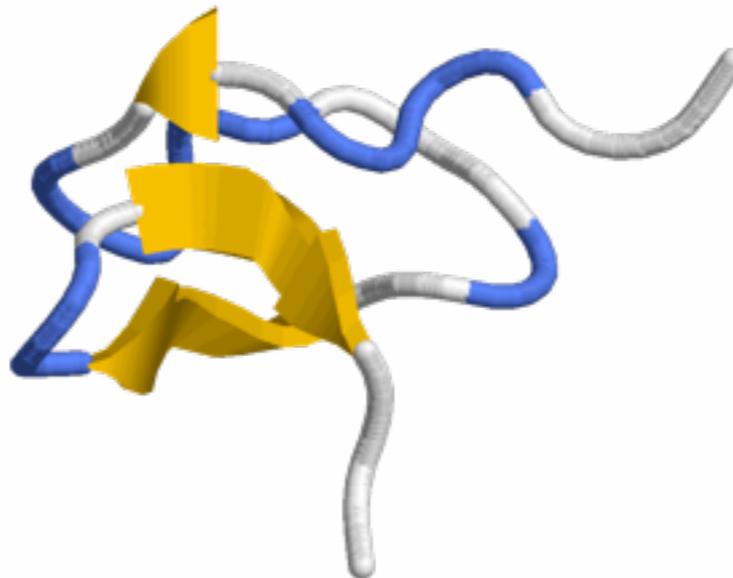
Dihedral space or distance space



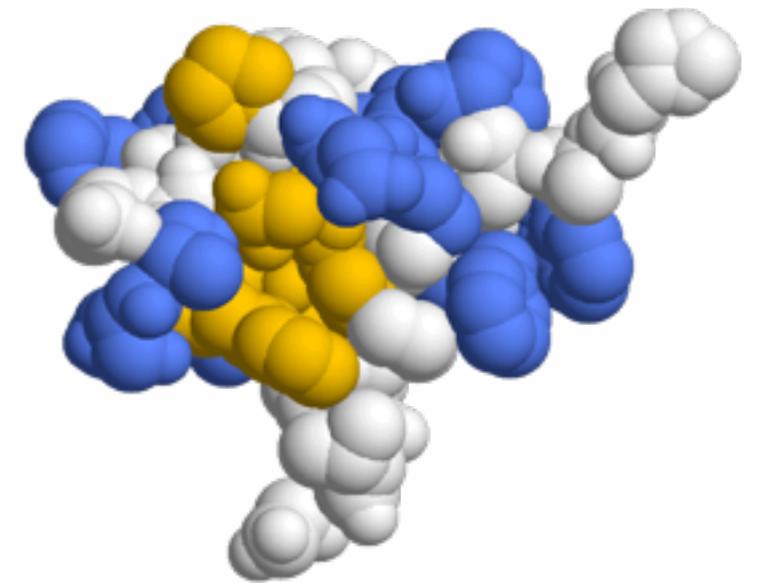
Reduced atom representation



Vector representation



Secondary Structure



Accessible surface (and others)

Raw scores

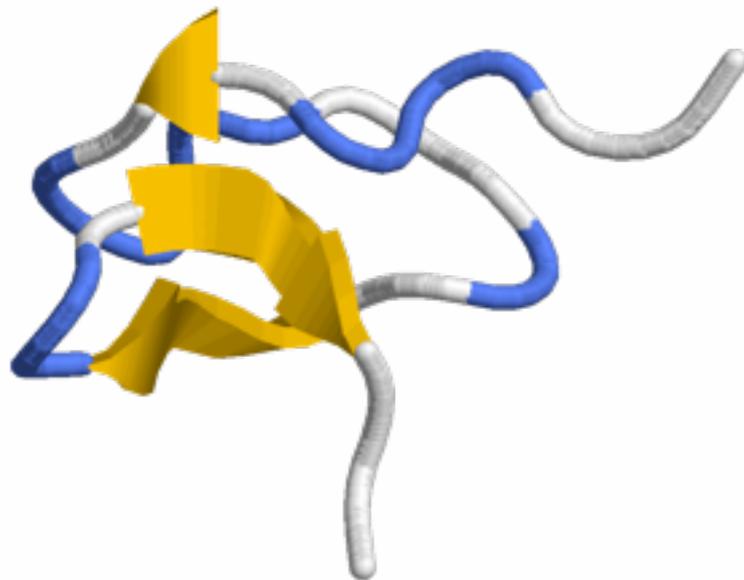
Scoring

	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W
C	9	-1	-1	-3	0	-3	-3	-3	4	-3	-3	-3	-3	-3	-3	-1	-1	-2	-2	-2
S	-1	4	1	-1	1	0	1	0	0	0	-1	-1	0	-1	-2	-2	-2	-2	-2	-3
T	-1	1	4	1	-1	1	0	1	0	0	0	-1	0	-1	-2	-2	-2	-2	-2	-3
P	-3	-1	1	4	-1	0	-1	-1	-1	-1	-2	-2	-1	-2	-3	-3	-2	-4	-3	-4
A	0	1	-1	-1	4	0	-3	-2	-1	-1	-2	-1	-1	-1	-1	-1	-2	-2	-2	-3
G	-3	0	1	-2	0	4	-2	-1	-2	-2	-2	-2	-2	-3	-4	-4	0	-1	-2	-2
N	-3	1	0	-2	-2	0	4	1	0	0	-1	0	0	-2	-3	-3	-3	-3	-3	-4
D	-3	0	1	-2	-1	-1	1	4	2	0	-1	-2	-1	-3	-3	-4	-3	-3	-3	-4
E	-4	0	0	-1	-1	-2	0	2	4	2	0	0	1	-2	-3	-3	-3	-3	-3	-4
Q	-3	0	0	-1	-1	-2	0	2	0	4	1	1	0	-3	-2	-2	-3	-3	-3	-4
H	-3	-1	0	-2	-2	-2	1	1	0	0	4	0	-1	-2	-3	-3	-2	-1	2	-2
R	-3	-1	-1	-2	-1	-2	0	-2	0	1	0	4	2	-1	-3	-3	-3	-3	-3	-4
K	-3	0	0	-1	-1	-2	0	-1	1	1	-2	0	4	-1	-2	-3	-3	-3	-3	-4
M	-1	-1	-1	-2	-1	-3	-2	-3	-2	0	-2	-1	-1	4	1	2	0	0	-1	-1
I	-1	-2	-2	-3	-1	-4	-3	-3	-3	-3	-3	-3	-3	1	4	2	1	0	-1	-3
L	-1	-2	-2	-3	-1	-4	-3	-4	-3	-3	-3	-3	-3	2	2	4	3	0	-1	-2
V	-1	-2	-2	-3	0	-3	-3	-3	-2	-2	-3	-3	-3	1	3	1	4	-1	-1	-3
F	-2	-2	-2	-3	-1	-3	-3	-3	-3	-3	-3	-3	-3	0	0	0	-1	4	3	1
Y	-2	-2	-2	-3	-2	-3	-3	-3	-3	-3	-3	-3	-3	-1	-1	-1	3	3	4	2
W	-2	-3	-3	-4	-3	-4	-4	-4	-3	-3	-3	-3	-3	-1	-1	-1	2	2	2	4

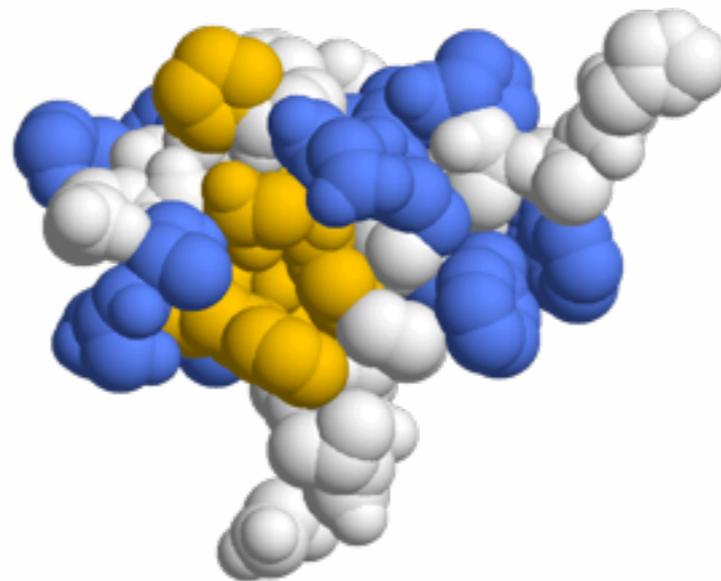
Aminoacid substitutions

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_i^N (m_i - m_i^*)^2}$$

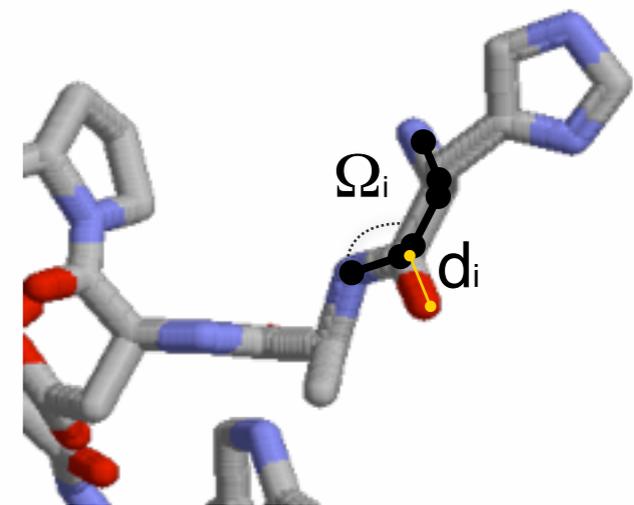
Root Mean Square Deviation



Secondary Structure (H,B,C)



Accessible surface (B,A [%])



Angles or distances

The Root Mean Square Deviation

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_i^N (m_i - m_i^*)^2}$$

Number of atoms

3D coordinates of atom i in the first model

3D coordinates of atom i in the second model

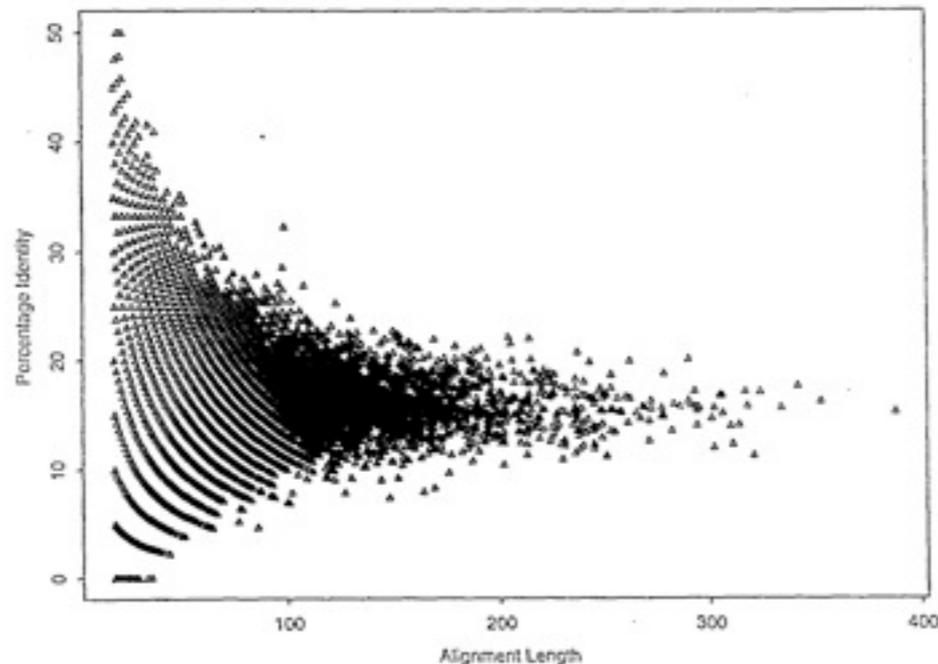
The diagram shows the mathematical formula for Root Mean Square Deviation (RMSD). The formula is $\text{RMSD} = \sqrt{\frac{1}{N} \sum_i^N (m_i - m_i^*)^2}$. Three arrows point from text labels to parts of the formula: one from 'Number of atoms' to the denominator N , one from '3D coordinates of atom i in the first model' to the term m_i , and one from '3D coordinates of atom i in the second model' to the term m_i^* .

Significance of an alignment (score)

Scoring

Probability that the optimal alignment of two random sequences/structures of the same length and composition as the aligned sequences/structures have at least as good a score as the evaluated alignment.

Empirical:



Analytic:

$$P(S > x) = 1 - \exp(-Kmn e^{-\lambda e})$$

Global dynamic programming alignment

Optimizer

Needleman-Wunsch

match = 1 mismatch = -1 gap = -1

		G	C	A	T	G	C	U
	0	-1	-2	-3	-4	-5	-6	-7
G	-1	1	0	-1	-2	-3	-4	-5
A	-2	0	0	1	0	-1	-2	-3
T	-3	-1	-1	0	2	1	0	-1
T	-4	-2	-2	-1	1	1	0	-1
A	-5	-3	-3	-1	0	0	0	-1
C	-6	-4	-2	-2	-1	-1	1	0
A	-7	-5	-3	-1	-1	-1	0	0

Backtracking to get the best alignment

Sequences

Best Alignments

GCATGCU

GCATG-CU

GCA-TGCU

GCAT-GCU

GATTACA

G-ATTACA

G-ATTACA

G-ATTACA

Local dynamic programming alignment

Optimizer

Smith-Waterman

match = 2 mismatch = -1 gap = -1

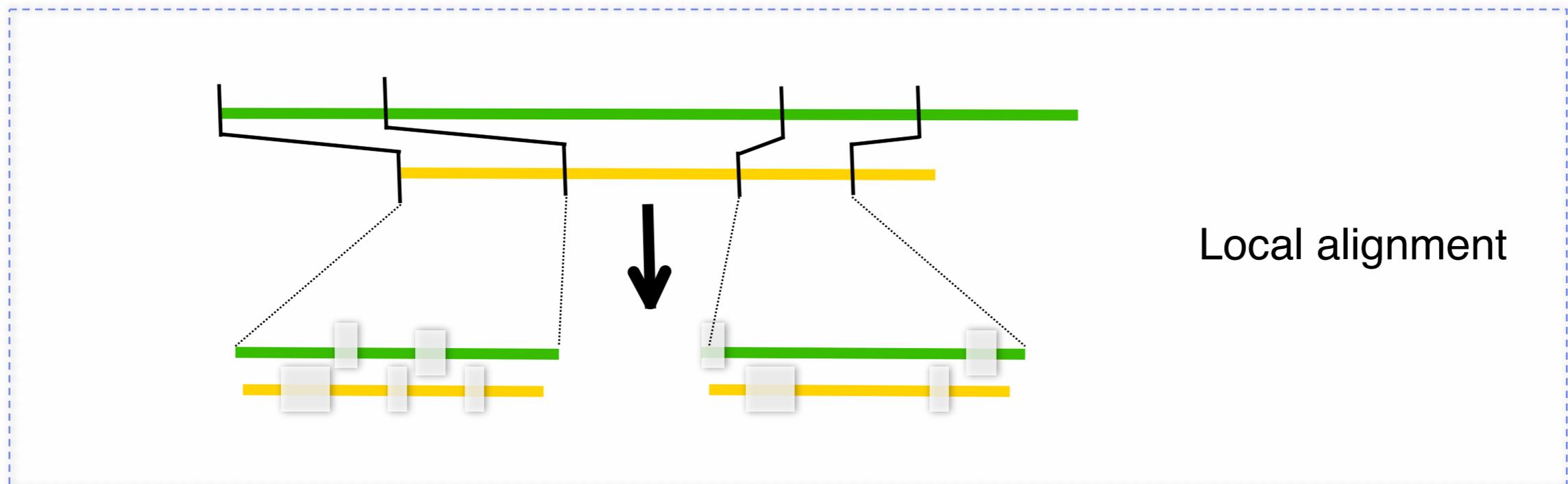
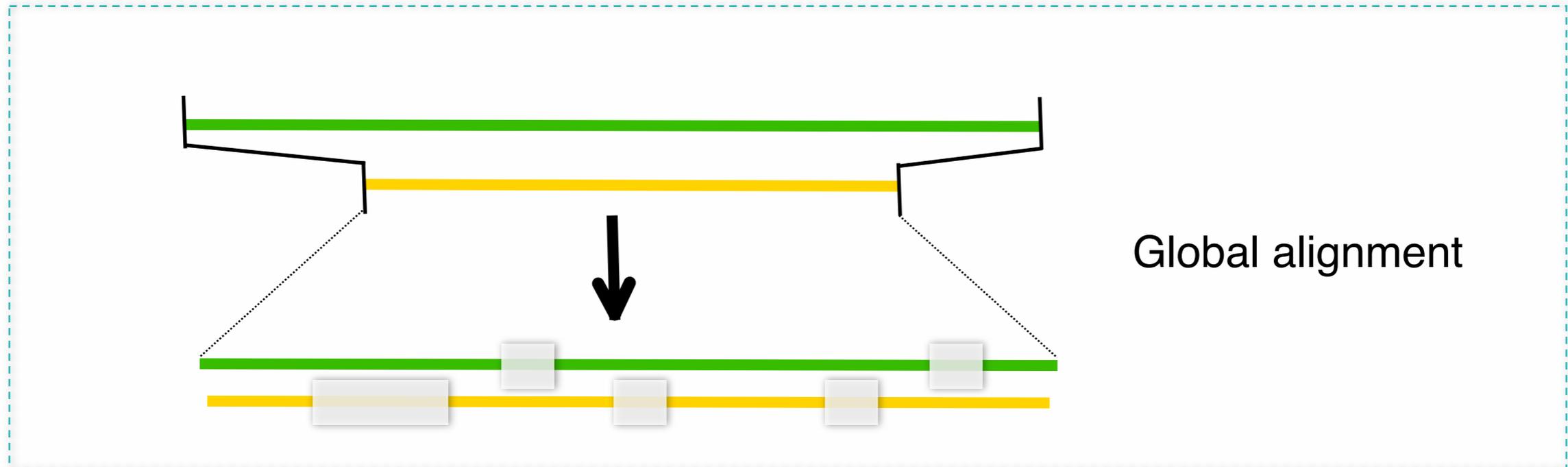
		A	C	A	C	A	C	T	A
	0	0	0	0	0	0	0	0	0
A	0	2	1	2	1	2	1	0	2
G	0	1	1	1	1	1	1	0	1
C	0	0	3	2	3	2	3	2	1
A	0	2	2	5	4	5	4	3	4
C	0	1	4	4	7	6	7	6	5
A	0	2	3	6	6	9	8	7	8
C	0	1	4	5	8	8	11	10	9
A	0	2	3	6	7	10	10	10	12

Backtracking to get the best alignment

Sequences	Best Alignment
-----	-----
ACACACTA	A-CACACTA
AGCACACA	AGCACAC-A

Global vs local alignment

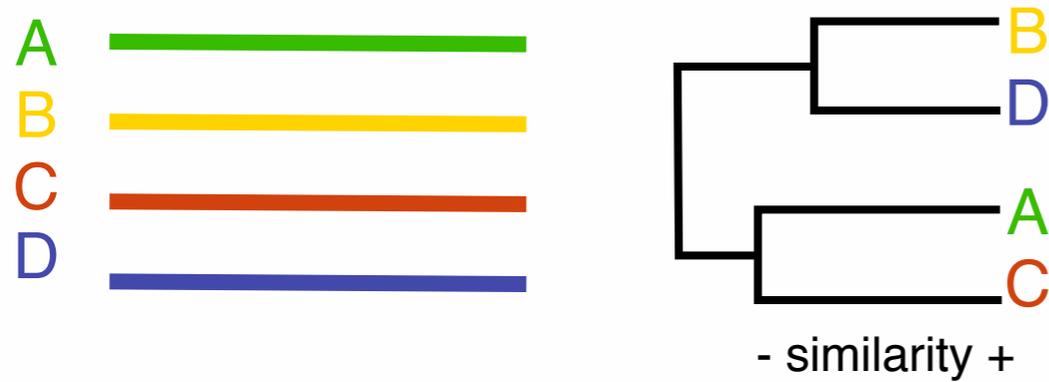
Optimizer



Multiple alignment Optimizer

Pairwise alignments

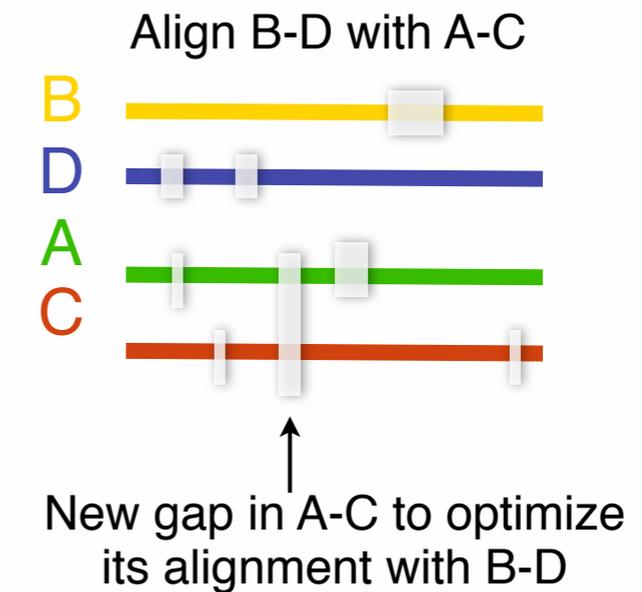
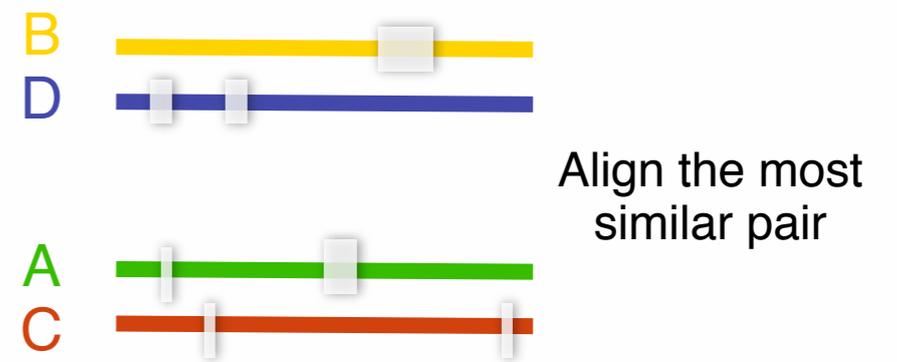
Example: 4 sequences A, B, C, D



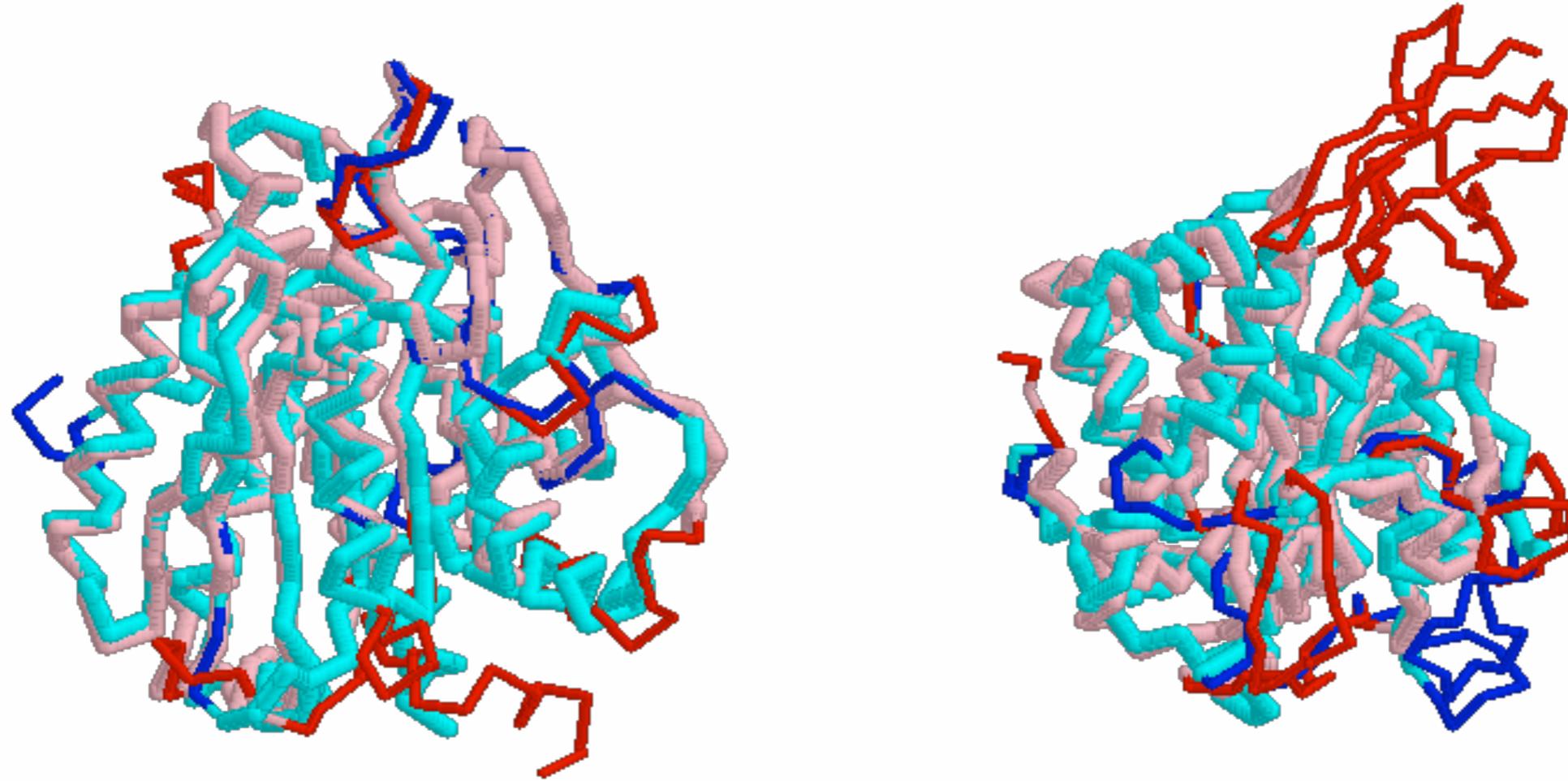
6 pairwise comparisons
then cluster analysis

Multiple alignments

Following the tree from step 1



Coverage vs Accuracy

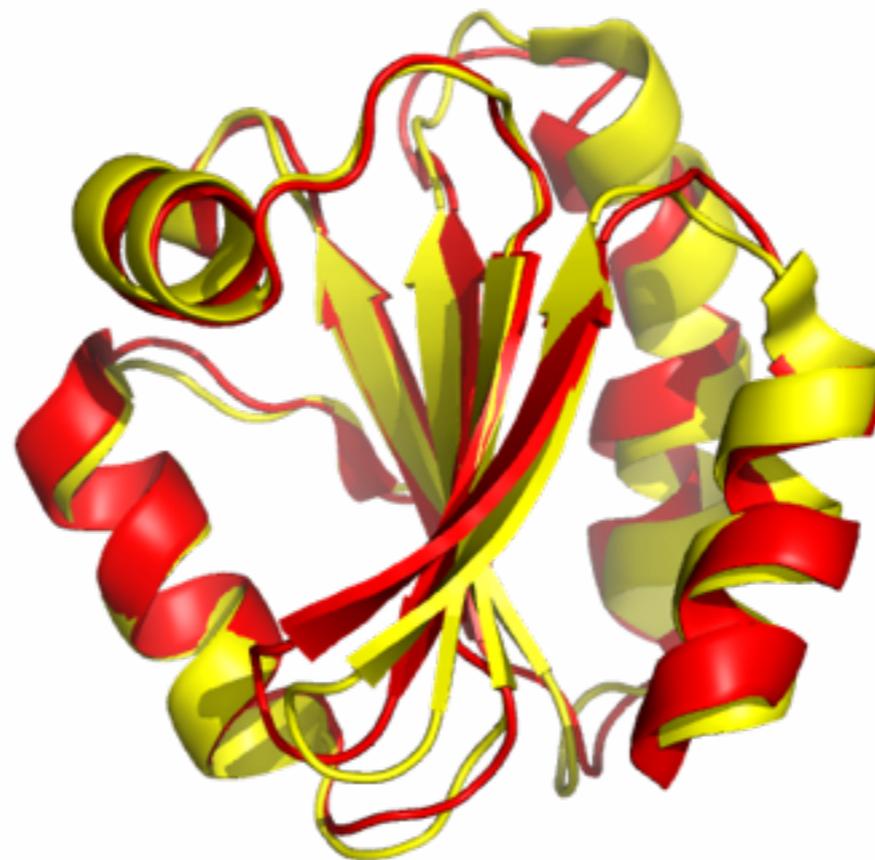


Same RMSD ~ 2.5Å

Coverage ~90% C α

Coverage ~75% C α

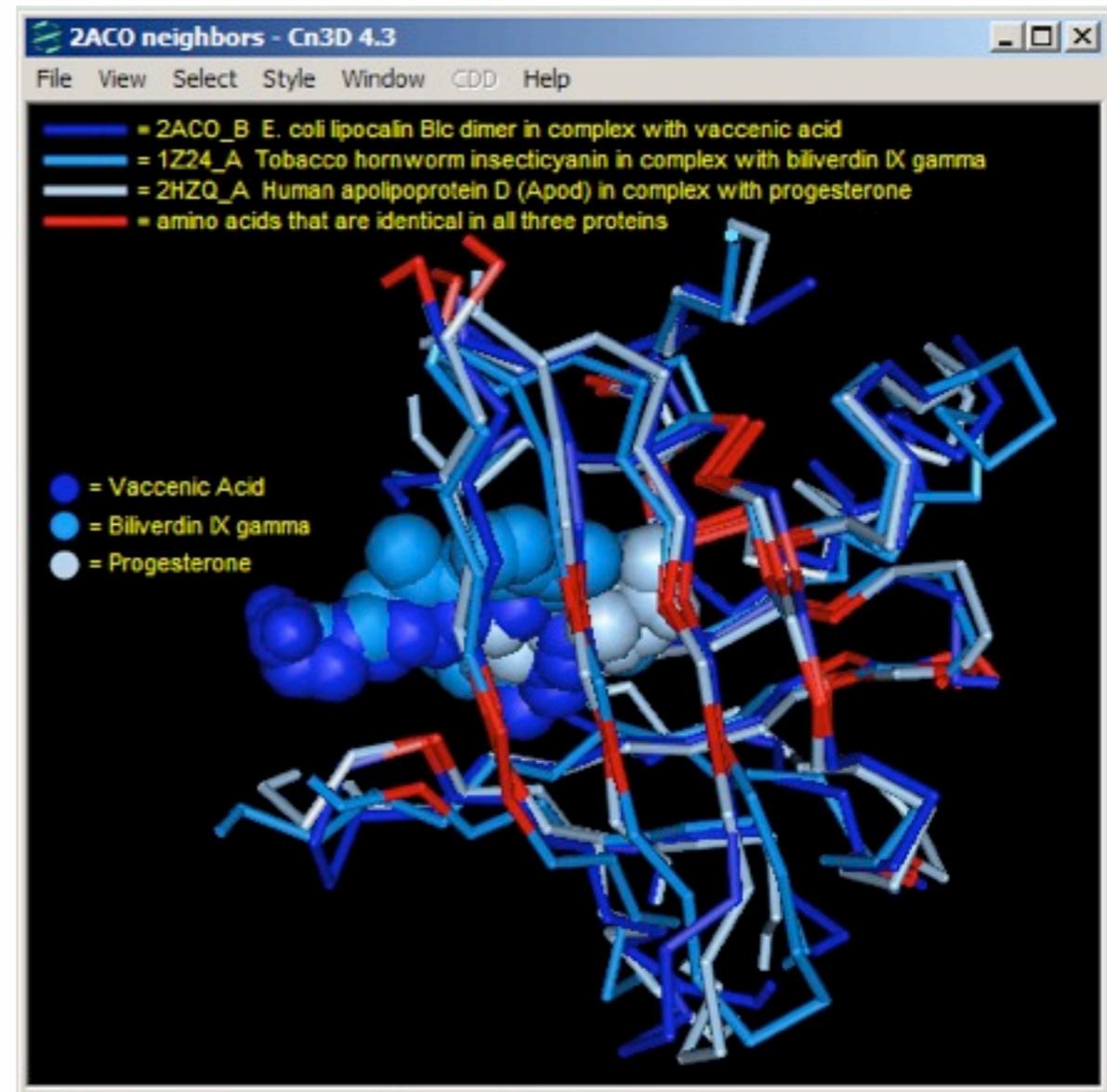
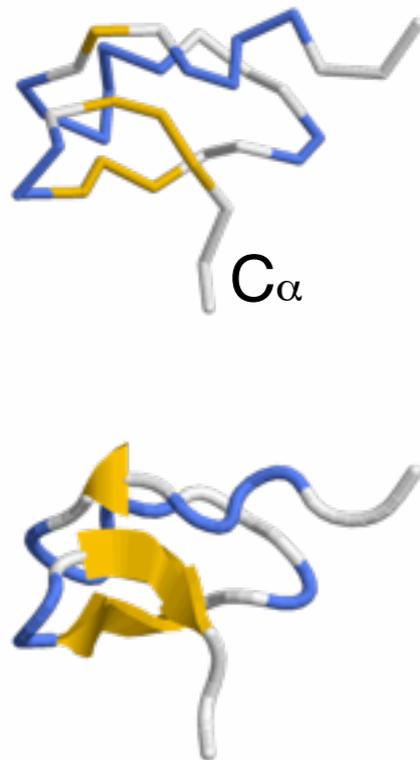
Structural alignment



Vector Alignment Search Tool (VAST)

<http://www.ncbi.nlm.nih.gov/Structure/VAST/vast.shtml>

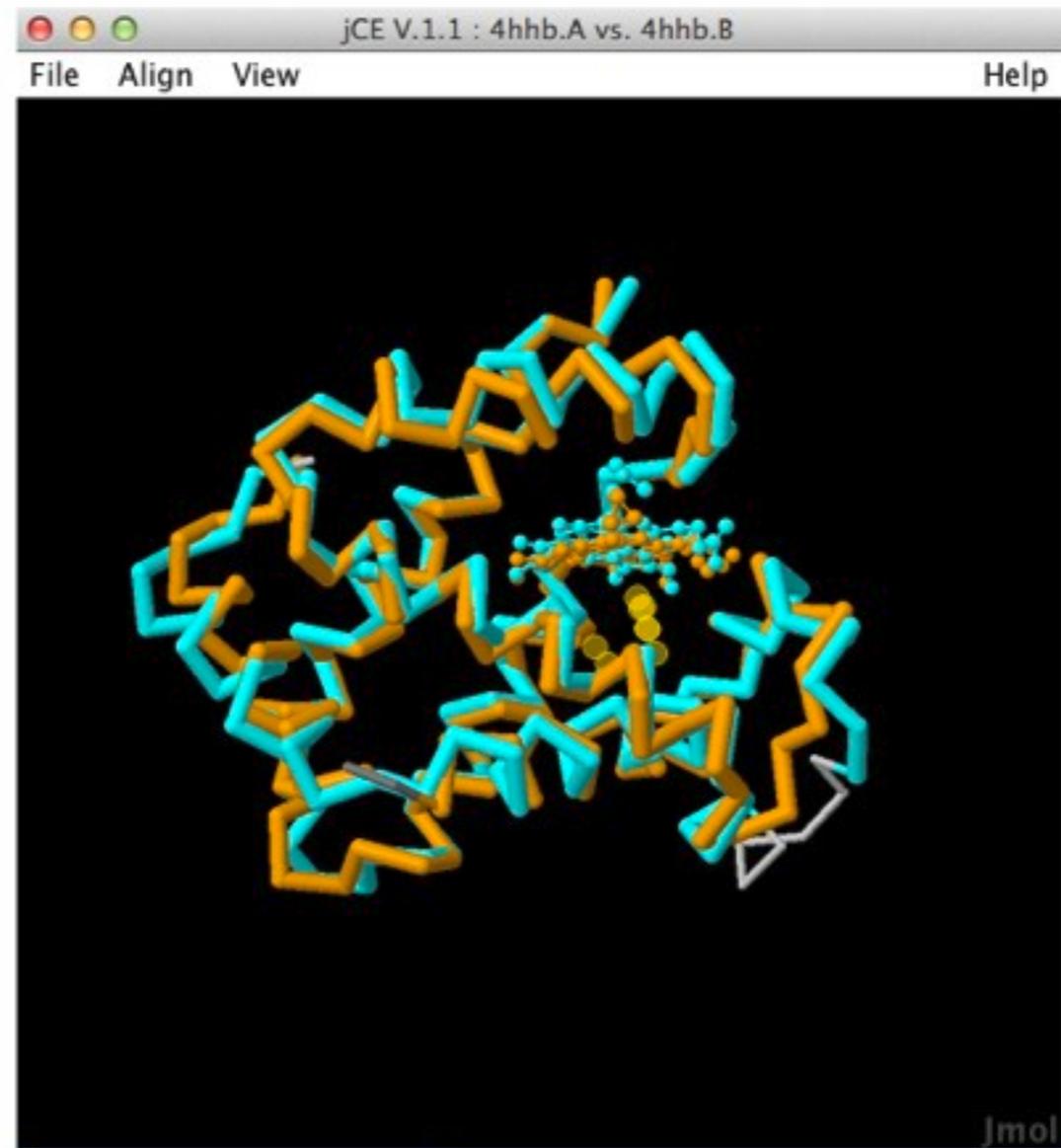
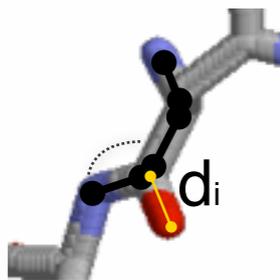
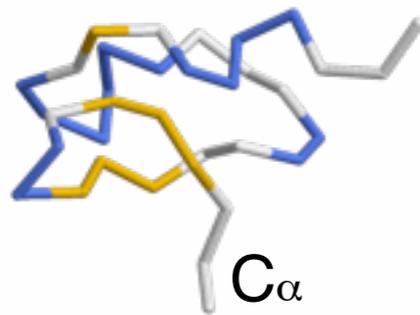
Vectorial representation of secondary structure elements



Incremental combinatorial extension (CE)

<http://source.rcsb.org/jfatcatserver/ceHome.jsp>

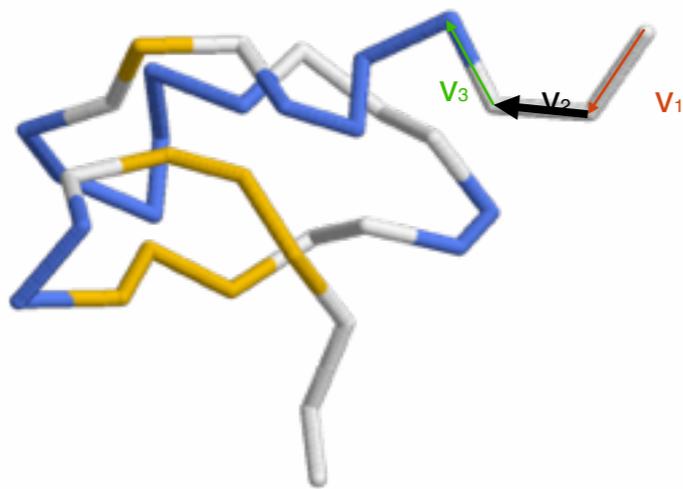
8 residues peptides



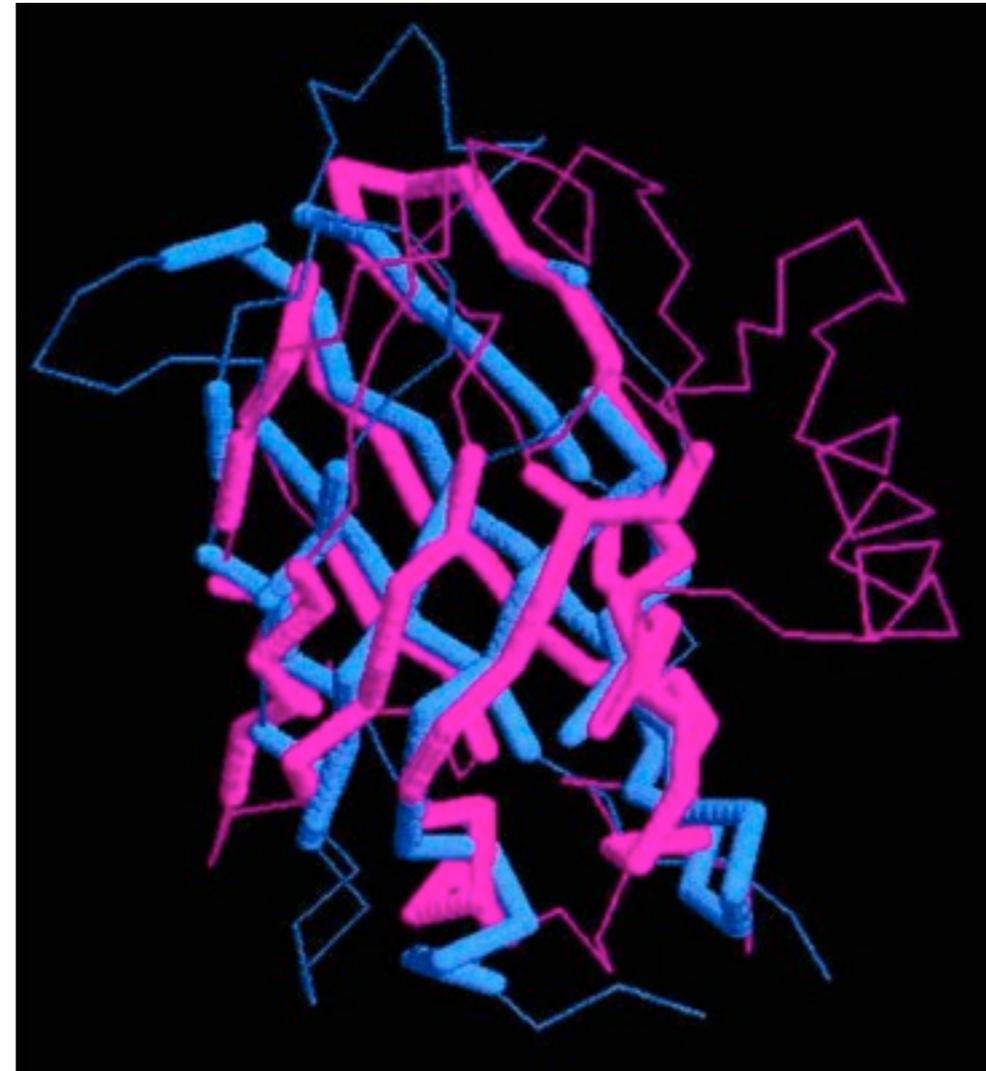
Matching molecular models obtained from theory (MAMMOTH)

<http://ub.cbm.uam.es/software/mammoth.php>

7 residues peptides



URMS instead of RMSD



Structural alignment in the PDB

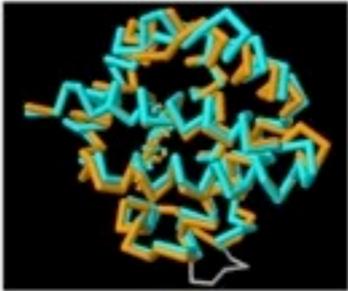
RCSB PDB [Deposit](#) [Search](#) [Visualize](#) [Analyze](#) [Download](#) [Learn](#) [More](#) [MyPDB Login](#)

Analyze Options

Sequence & Structure Alignment

RCSB PDB's Comparison Tool calculates pairwise sequence (blast2seq, Needleman-Wunsch, and Smith-Waterman) and structure alignments (jFATCAT, CE, Mammoth, TM-Align, TopMatch).

Comparisons can be made for any protein in the PDB archive and for customized or local files not in the PDB. Special features include support for both rigid-body and flexible alignments and detection of circular permutations.



Enter First PDB ID ↔ Enter Second PDB ID

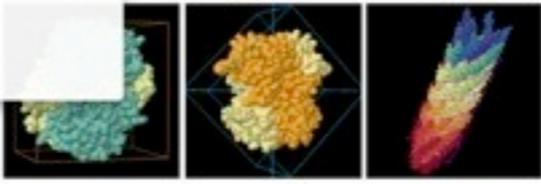
Select Associated Chain ID

Select Associated Chain ID

- Select Comparison Method -

- Sequence Alignment
 - blast2seq
 - Needleman-Wunsch
 - Smith-Waterman
- Structure Alignment
 - jFATCAT - rigid
 - jFATCAT - flexible
 - JCE algorithm
 - JCE Circular Permutation
 - External: FatCat
 - External: Mammoth
 - External: TM-Align
 - External: TopMatch

among subunits. The view displays the symmetry axes, a polyhedron that reflects the symmetry, and a color



PDB ID: 4HHB PDB ID: 1STP PDB ID: 1FO

Enter a PDB ID

Structural alignment in the PDB

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB Login

Structure Alignment View

Pre-calculated jCE results for 1EL7.A vs. 1E7D.A

This page provides a summary view of the protein structure alignment.

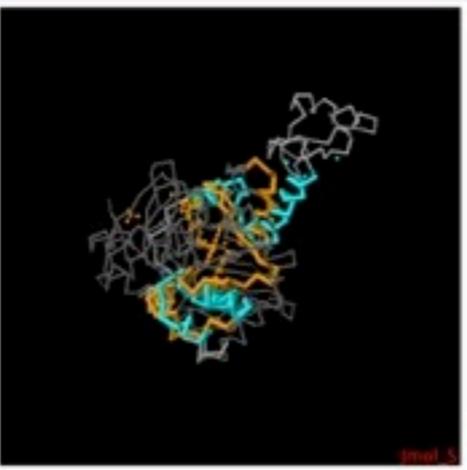
Structure Alignment Results

Alignment Details:	Query: (orange/dark grey) SARCOSINE OXIDASE	Subject: (cyan/light grey) RECOMBINATION ENDONUCLEASE VZI
Z-score: 2.30	 PDB ID: 1EL7	 PDB ID: 1E7D
Score: 397.60	Chain ID: A	Chain ID: A
RMSD: 7.19	Length: 385	Length: 157
%Id: 3.3%	Similarity: 19%	Similarity: 46%
	EC number: 1.5.3.1	EC number:

Comparison Method

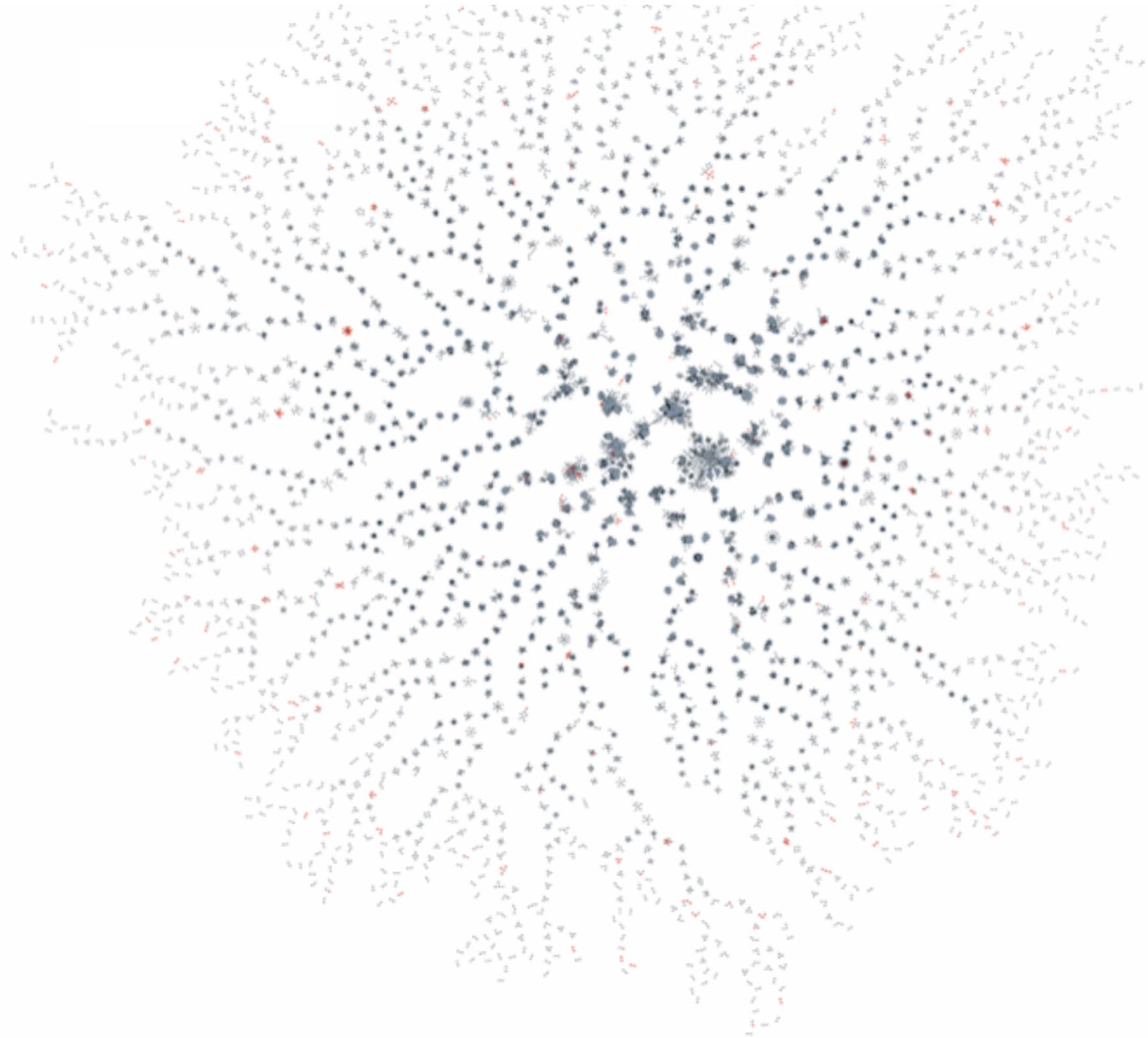
Select these two chains for other comparison: [Click here to align other protein chains. Back to the all vs. all search results for 1EL7.A or 1E7D.A](#)

Jmol



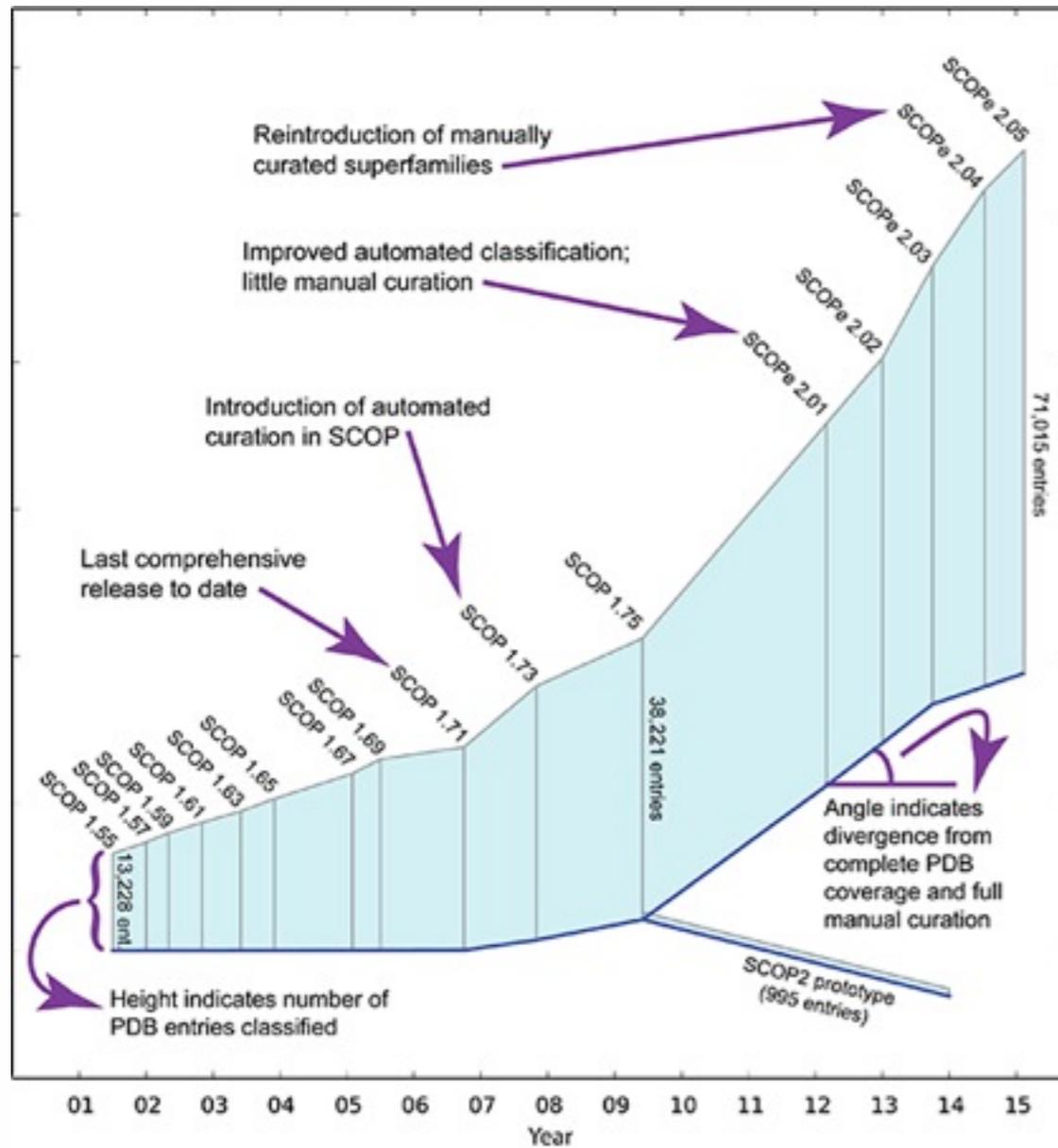
Tip: right-mouse click on Jmol to get access to additional Jmol functionality.
You may also drag the right-bottom corner of the Jmol area to resize it.

Classification of the structural space



SCOPe2.05 database

<http://scop.berkeley.edu/statistics/ver=2.05>



Class	Number of folds	Number of superfamilies	Number of families
All alpha proteins	286	509	1037
All beta proteins	176	359	931
Alpha and beta proteins (a/b)	148	245	965
Alpha and beta proteins (a+b)	381	558	1301
Multi-domain proteins	68	68	109
Membrane and cell surface proteins	57	113	153
Small proteins	92	132	260
Total	1208	1984	4756

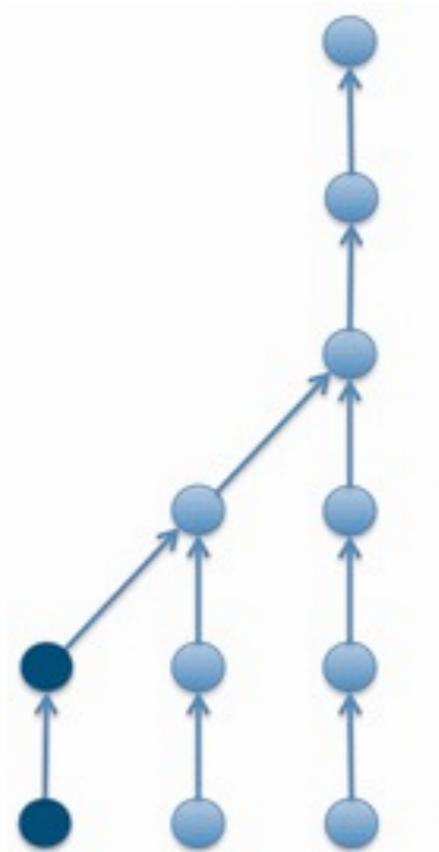
Clear classification of structures in:

- CLASS
- FOLD
- SUPER-FAMILY
- FAMILY

SCOP2 database

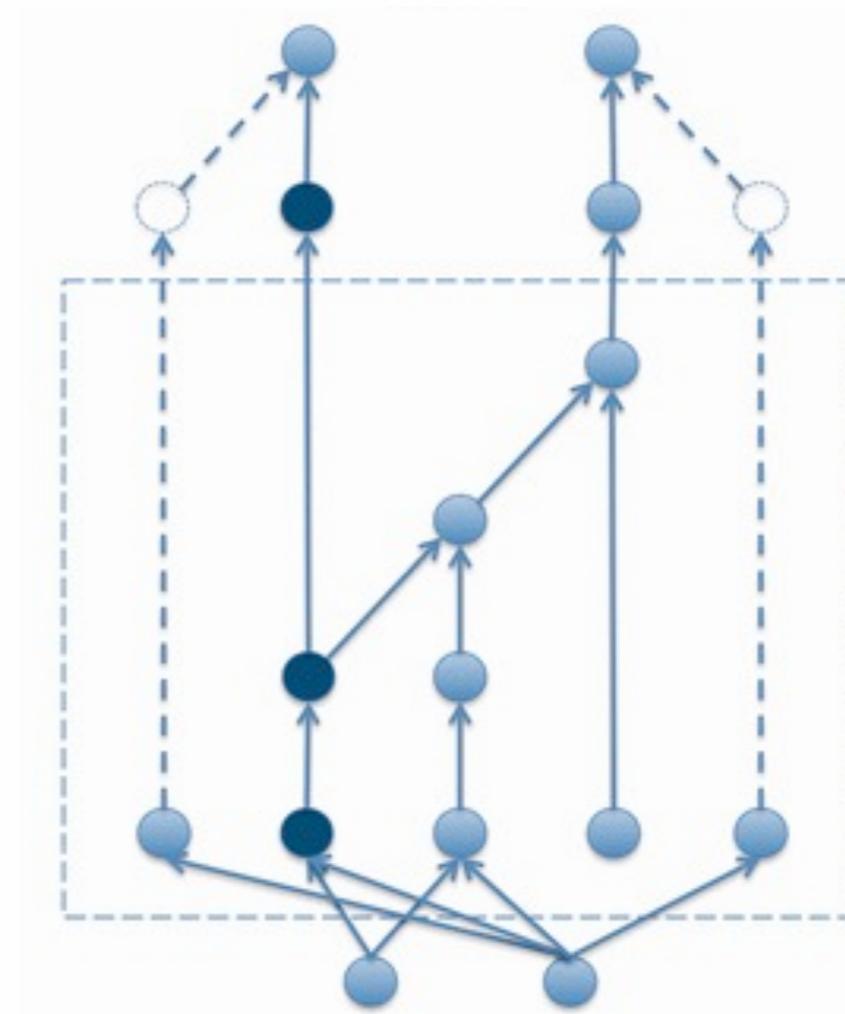
<http://scop2.mrc-lmb.cam.ac.uk/>

SCOP



SCOP2

Class
Fold
Super-Family
Family
Protein
Species



SCOP2 database

<http://scop2.mrc-lmb.cam.ac.uk/>

The screenshot displays the SCOP2 database interface with several key components and annotations:

- Alternative ways for database access:** A box pointing to the top navigation menu containing links for [About](#), [Browser](#), [Graph](#), [Download](#), and [Support](#).
- Main navigation menu:** The top header area with the SCOP2 logo and the text "Structural Classification of Proteins 2".
- Action buttons:** A box pointing to the "Configure", "Share", "Reset", and "Help" buttons located below the search bar.
- Navigation tabs:** A box pointing to the tabs in the pop-up window: "Structure", "Parents", "Children", and "Comments".
- Action buttons (pop-up):** A box pointing to the "Focus", "Remove", and "Browse" buttons within the pop-up window.
- Links to external sequence and structural resources:** A box pointing to the links for "View in PDB: SCRD", "View in PDB: SCRD", and "View in UniProt: P03040".
- Click to bring up the node pop up window:** A box pointing to the "Fold change in the lambda C1 repressor DNA-binding domain-like superfamily" node in the main graph.

The main interface includes a "Colour Legend" with options for "Configure", "Share", "Reset", and "Help". On the left, there are sections for "Relations" (is_a, part_of, occurred_in) and "Nodes" (Category, Attribute, Type, Class, Hyperfamily, Fold, Superfamily, Family, Protein, Species, Event, Interrelationship). The central graph shows a hierarchical structure of protein families and classes, with a highlighted node for the "Fold change in the lambda C1 repressor DNA-binding domain-like superfamily".

At the bottom, the footer contains the text "SCOP2 - July 2013 © MRC Laboratory of Molecular Biology" and navigation links for "Contact", "Sitemap", and "Top of page".

CATH4 database

<http://www.cathdb.info/>

CATH / Gene3D
26 million protein domains classified into 2,738 superfamilies

Buttons: [Browse](#), [Search](#), [Download](#), [Take the Tour](#)

What is CATH?
CATH is a classification of protein structures downloaded from the Protein Data Bank. We group protein domains into superfamilies when there is sufficient evidence they have diverged from a common ancestor.

- Search CATH by text, ID or keyword
- Search CATH by protein sequence (FASTA)
- Search CATH by PDB structure
- Browse CATH Hierarchy
- CATH Release Notes
- CATH Tutorials

Latest Release Statistics

CATH v4.0		based on PDB dated March 26, 2013
235,858	CATH Domains	
2,738	CATH Superfamilies	
69,058	Annotated PDBs	

Gene3D v12		Released March 18, 2012
6,131	Cellular Genomes	
21,662,155	Protein Sequences	
25,615,754	CATH Domain Predictions	

Example pages

- PDB "Zoop"
- Domain "1oukA01"
- Relatives of "1oukA01"
- Superfamily "HUPs"
- Functional Family
- FunFam Alignment
- Search for "enclose"
- Superfamily Comparison

Class: SS composition (mostly alpha, mostly beta, mixed alpha/beta or few secondary structures)

Architecture: overall shape SSe orientations in 3D space but ignores the connectivity between them

Topology (fold family): fold groups depending on both the overall shape and connectivity of the SSe

Homologous superfamily: groups together homologous protein domains

Class	Architecture	Topology	Homologous Superfamily	S35 Family	S60 Family	S95 Family	S100 Family	Domains
Class 1	5	397	907	3879	5118	6737	13368	48121
Class 2	20	241	547	3650	5221	8373	14526	58944
Class 3	14	626	1158	9171	13415	17047	35313	125772
Class 4	1	111	126	233	293	410	651	3021
TOTAL	40	1375	2738	16933	24047	32567	63858	235858

Orengo, C.A., et al. (1997) *Structure*. 5. 1093-1108.

PFAM

<http://pfam.xfam.org/>

Protein: ABL1_HUMAN (P00519)

Summary

This is the summary of UniProt entry *ABL1_HUMAN* (P00519).

Description: Tyrosine-protein kinase ABL1 EC=2.7.10.2

Source organism: *Homo sapiens (human)* (NCBI taxonomy ID 9606)

Length: 1130 amino acids

Pfam domains

This image shows the arrangement of the Pfam domains that we found on this sequence. Clicking on a domain will take you to the page describing that Pfam entry. The table below gives the domain boundaries for each of the domains. [More...](#)

Source	Domain	Start	End
low_complexity	n/a	5	23
disorder	n/a	53	54
Pfam A	SH3_1	67	113
Pfam A	SH2	127	202
Pfam A	Kinase_Tyr	242	493
low_complexity	n/a	249	260
disorder	n/a	489	490
disorder	n/a	503	511
disorder	n/a	513	719
low_complexity	n/a	605	616
low_complexity	n/a	629	639
low_complexity	n/a	701	710
disorder	n/a	724	993
Pfam B	Pfam-B_13565	864	1003
low_complexity	n/a	896	915
low_complexity	n/a	977	993
disorder	n/a	1009	1010
disorder	n/a	1013	1021
Pfam A	F_actin_bind	1022	1130
disorder	n/a	1023	1027

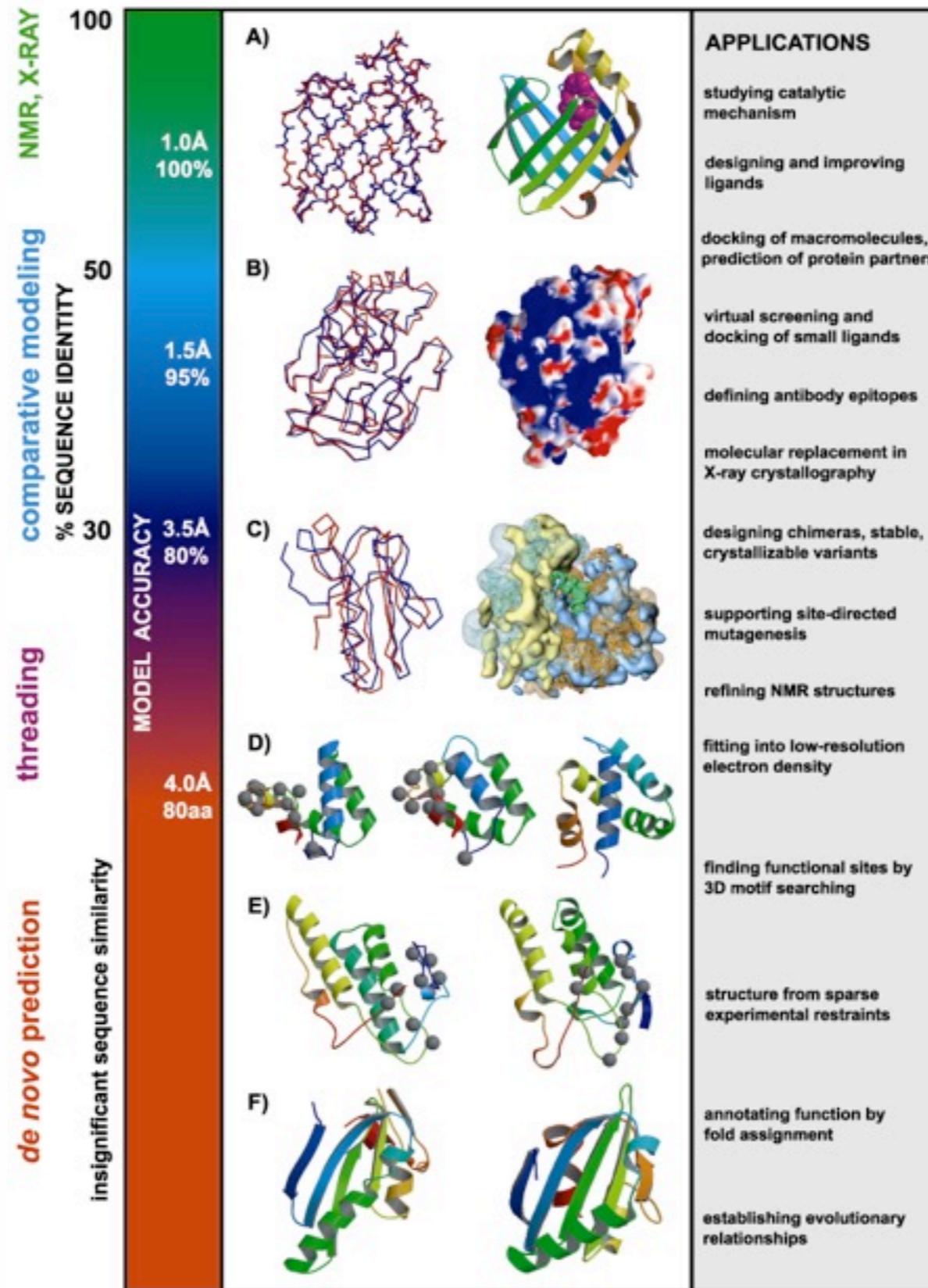
Pfam domains:



Source	Domain	Start	End
disorder	n/a	1	306
low_complexity	n/a	2	43
low_complexity	n/a	51	89
low_complexity	n/a	117	134
low_complexity	n/a	205	224
low_complexity	n/a	252	307
coiled_coil	n/a	254	276
Pfam A	zf-H2C2_2	386	411
Pfam A	zf-C2H2	430	453
low_complexity	n/a	457	469
disorder	n/a	477	496
low_complexity	n/a	489	500
disorder	n/a	547	549
disorder	n/a	559	562
disorder	n/a	564	570

Comments or questions on the site? Send a mail to pfam-help@sanger.ac.uk. Our cookie policy.
The Wellcome Trust

Utility of protein structure models, despite errors



Take home message

Protein and Nucleic acids structures are stored in publicly available databases

Proteins are aligned by sequence and by structure

Structural alignments might identify distant homologs that cannot be recognized by sequence comparison

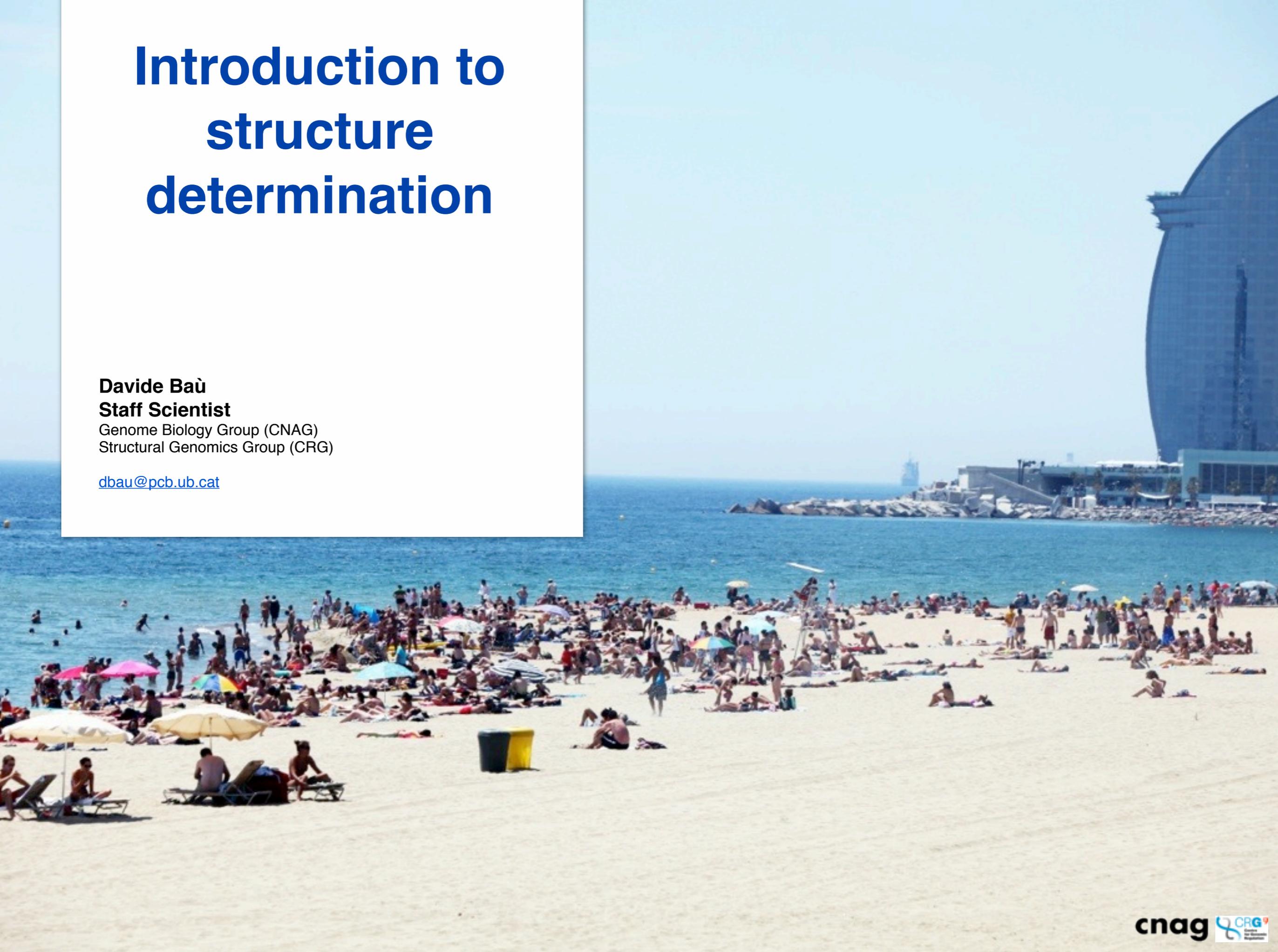
There are several databases that classify protein structures

Introduction to structure determination

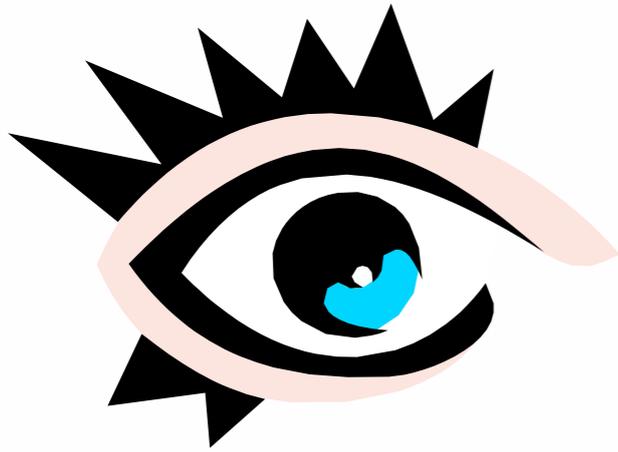
Davide Baù
Staff Scientist

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

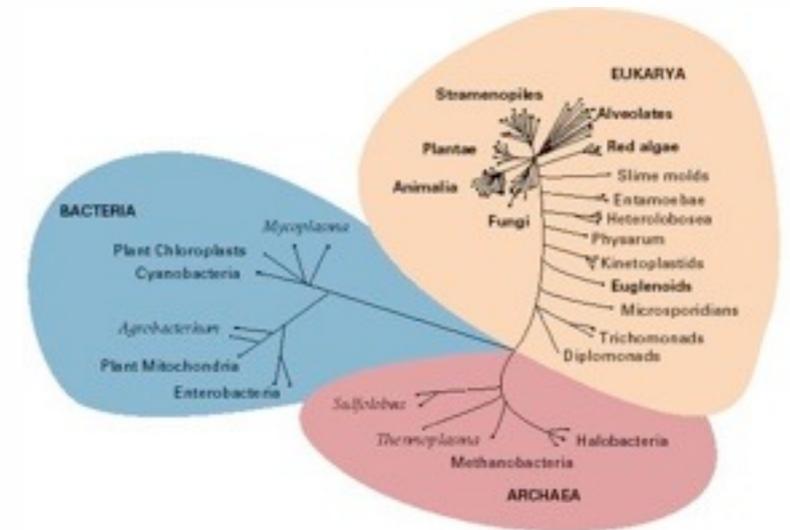
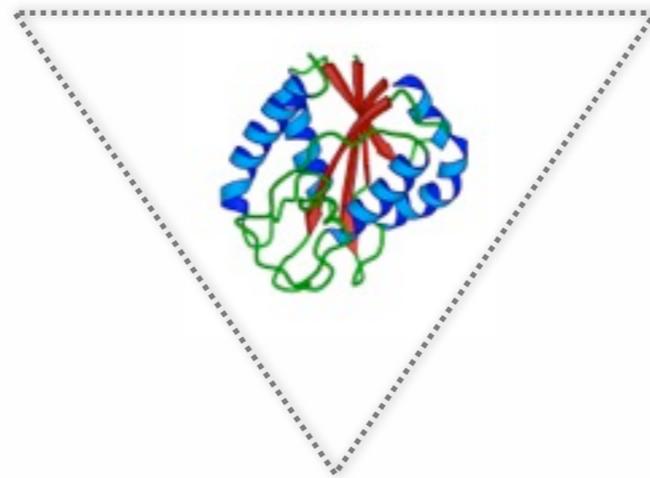
dbau@pcb.ub.cat



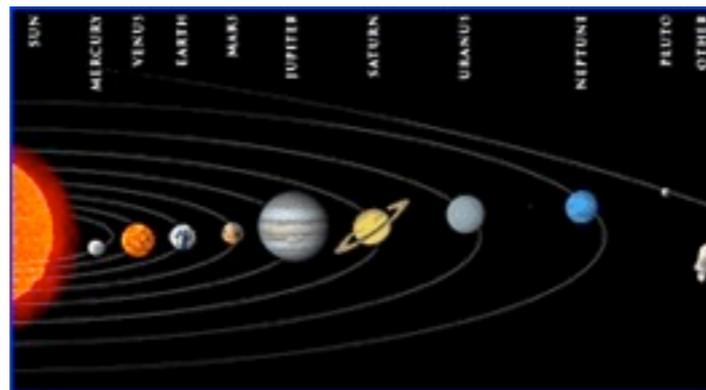
Data groups



Experimental observations

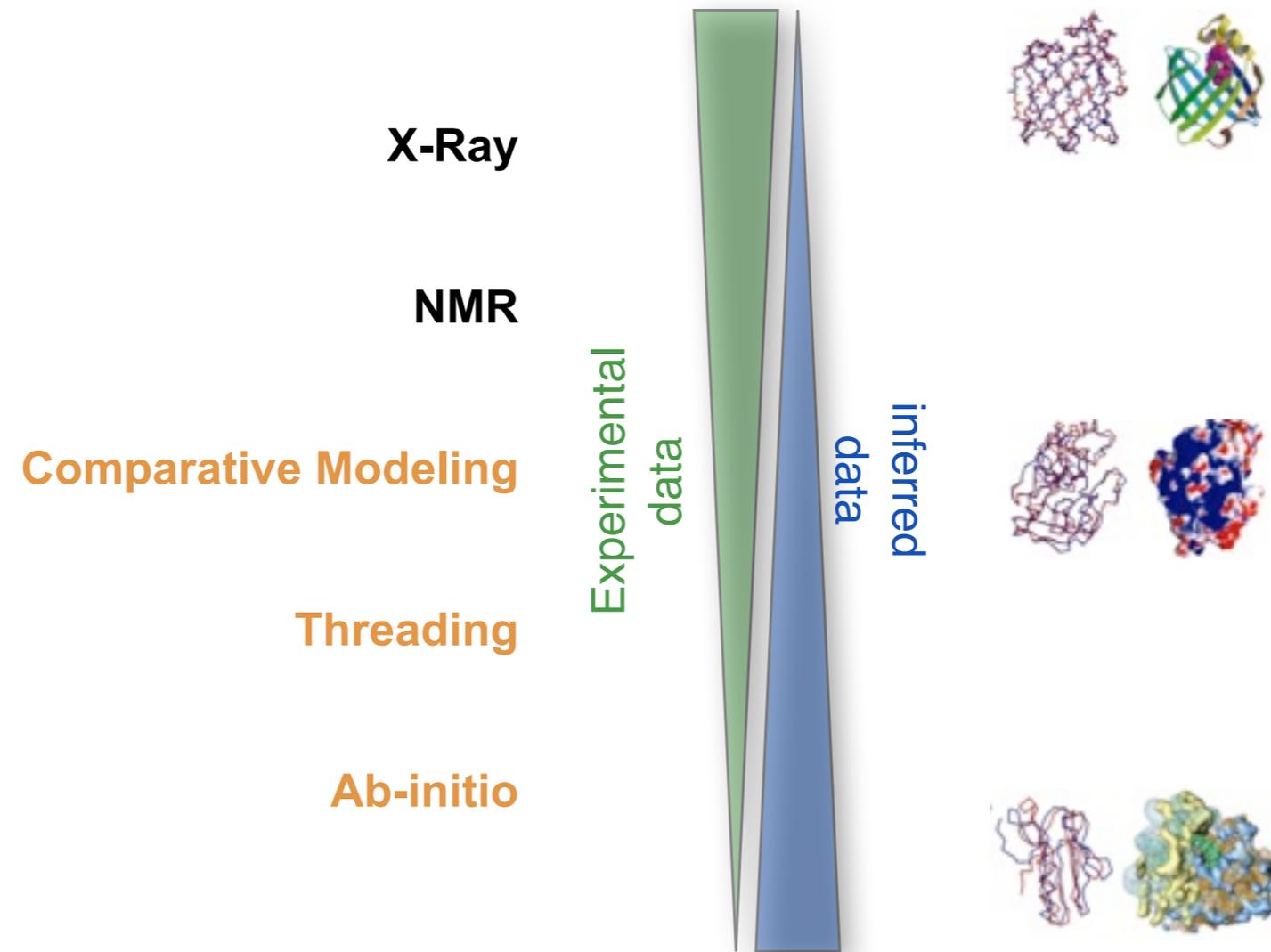


Statistical rules

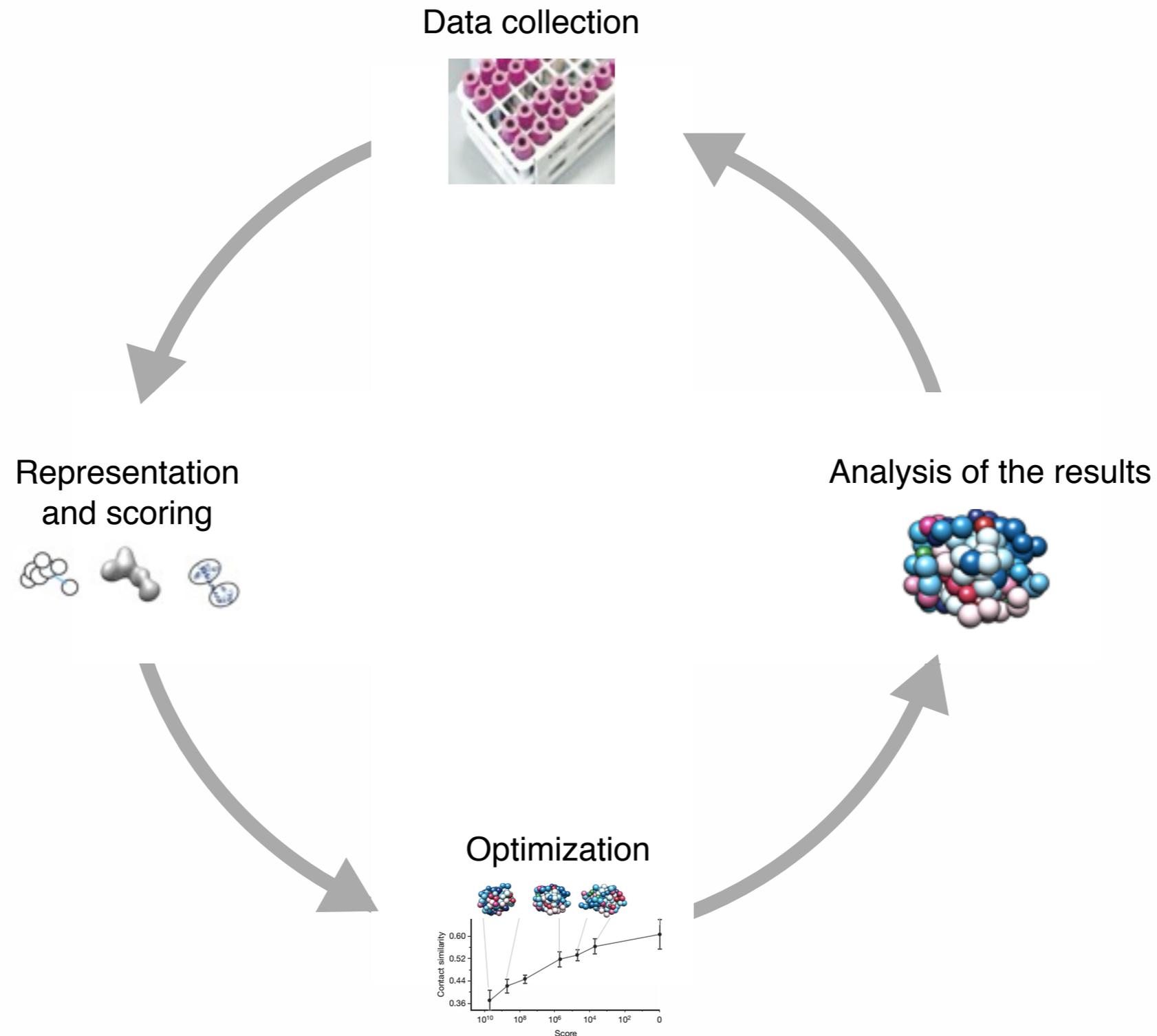


Laws of physics

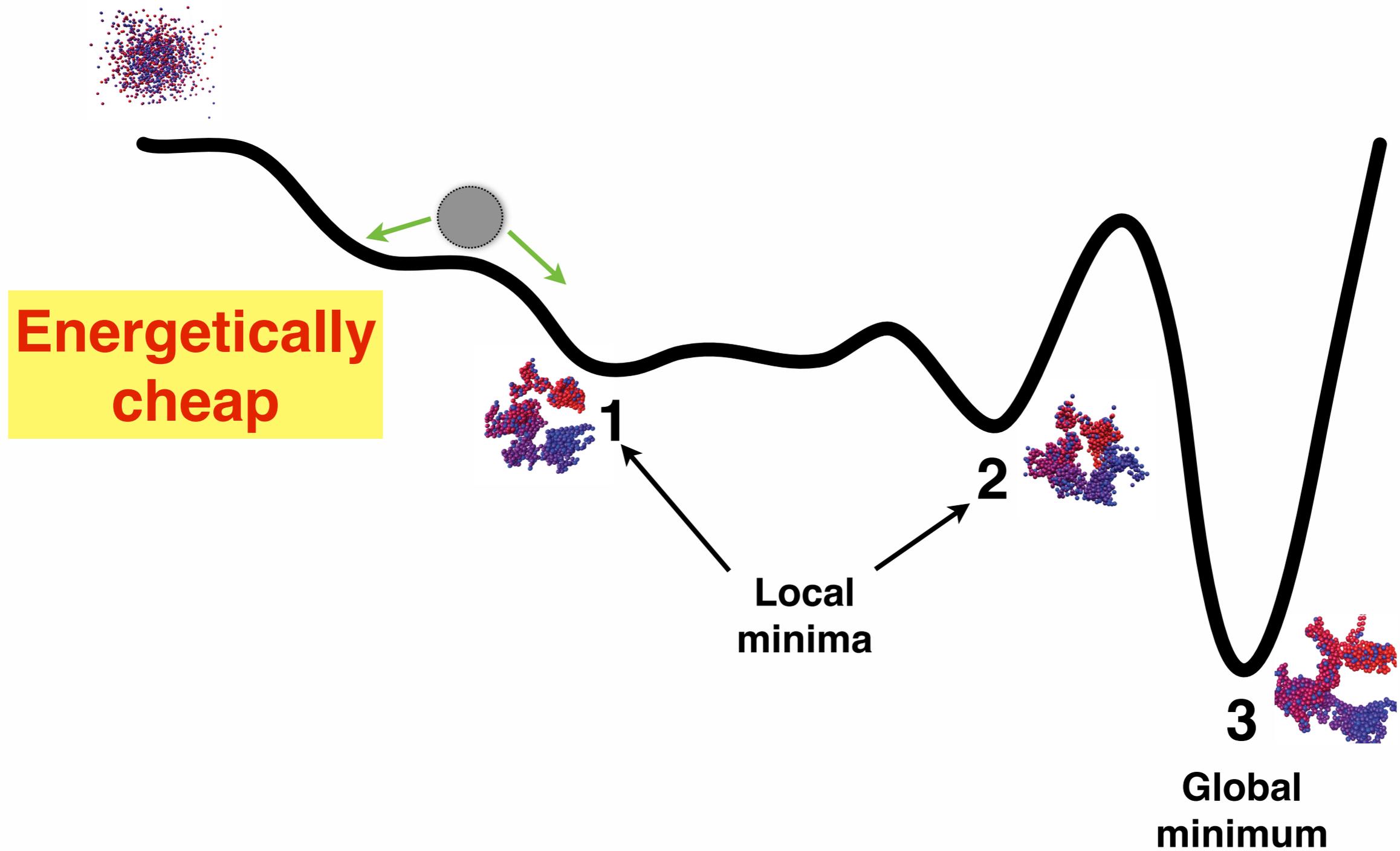
Structure prediction vs determination



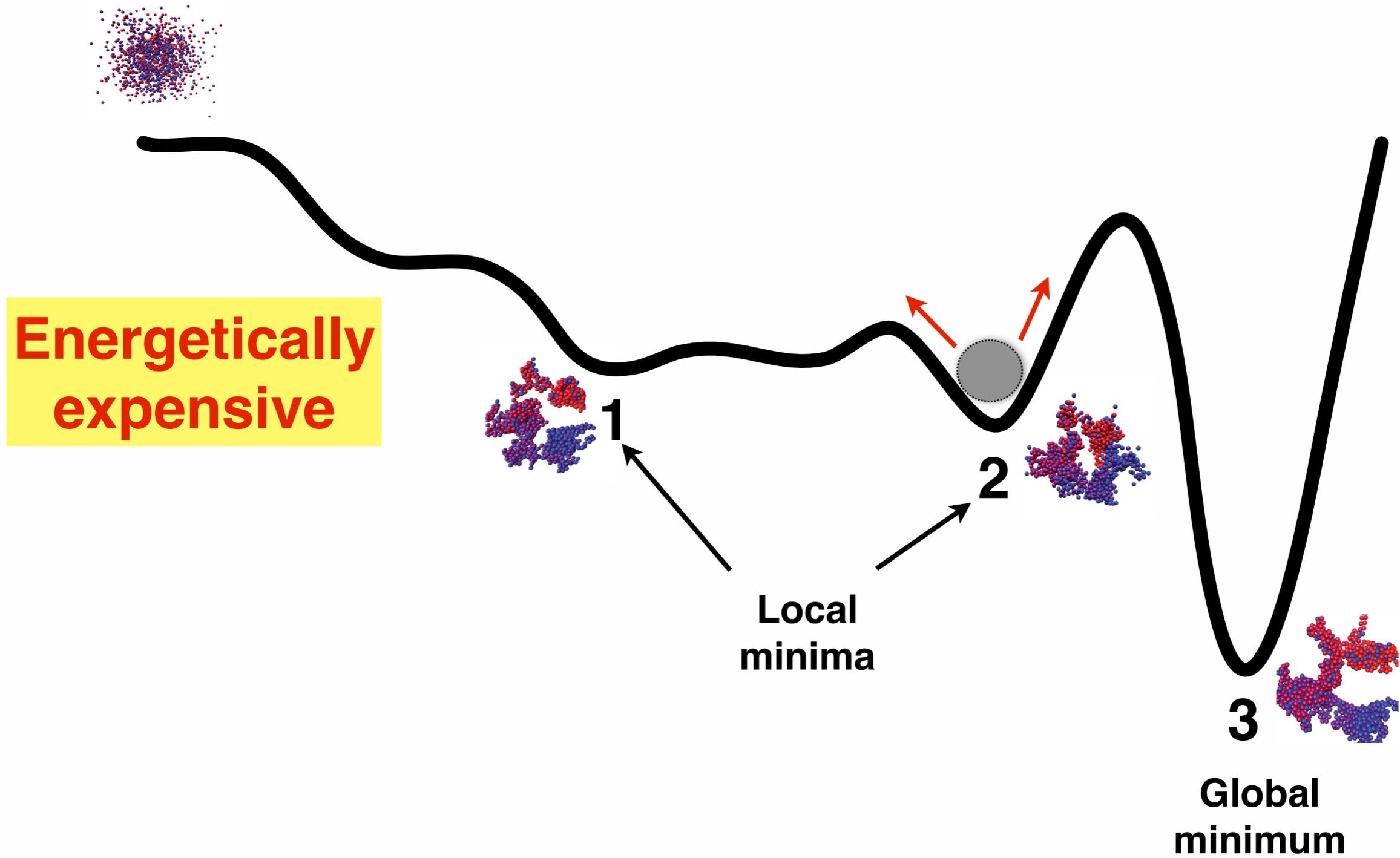
The four stages of integrative modeling



Energy landscape

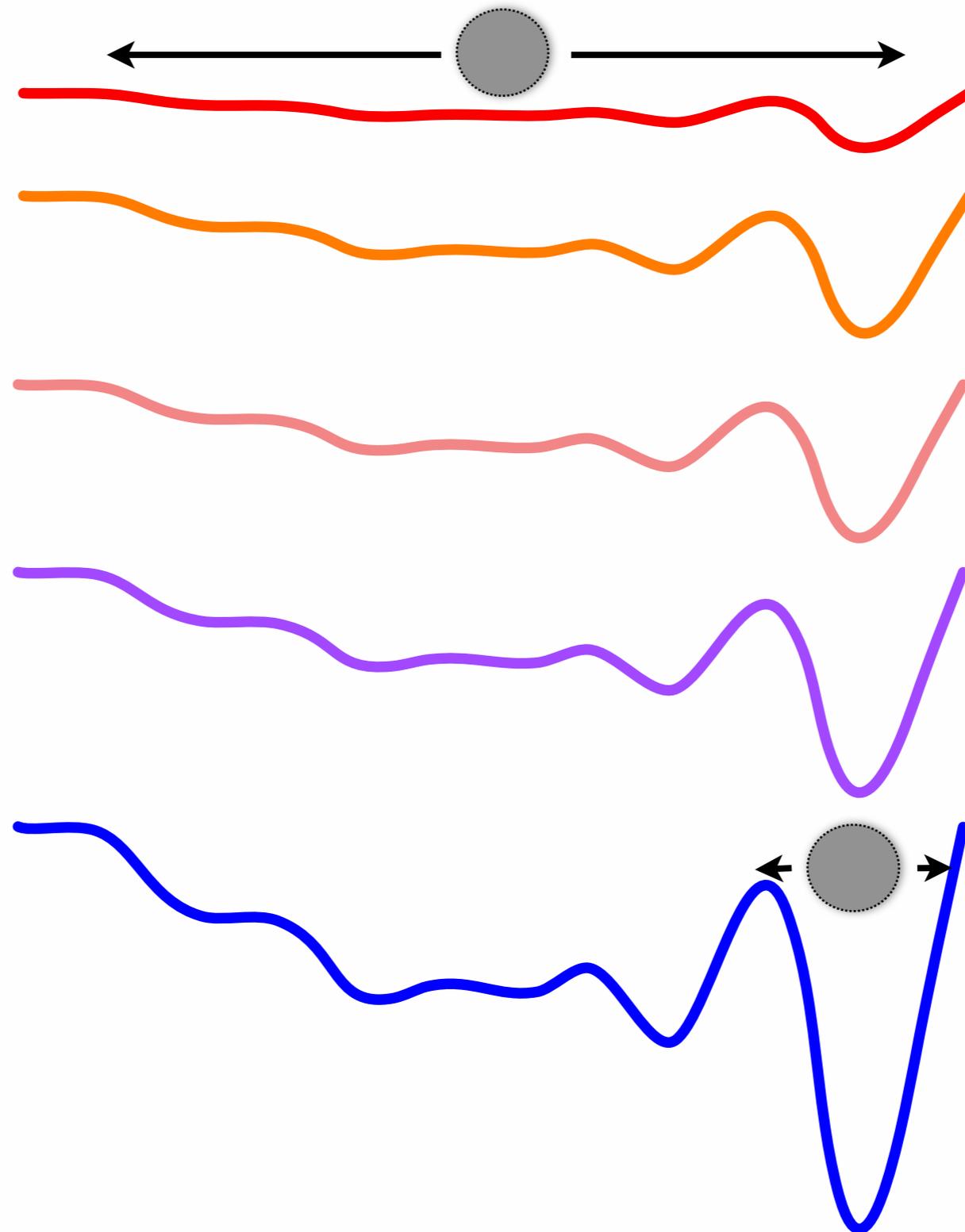


Energy landscape

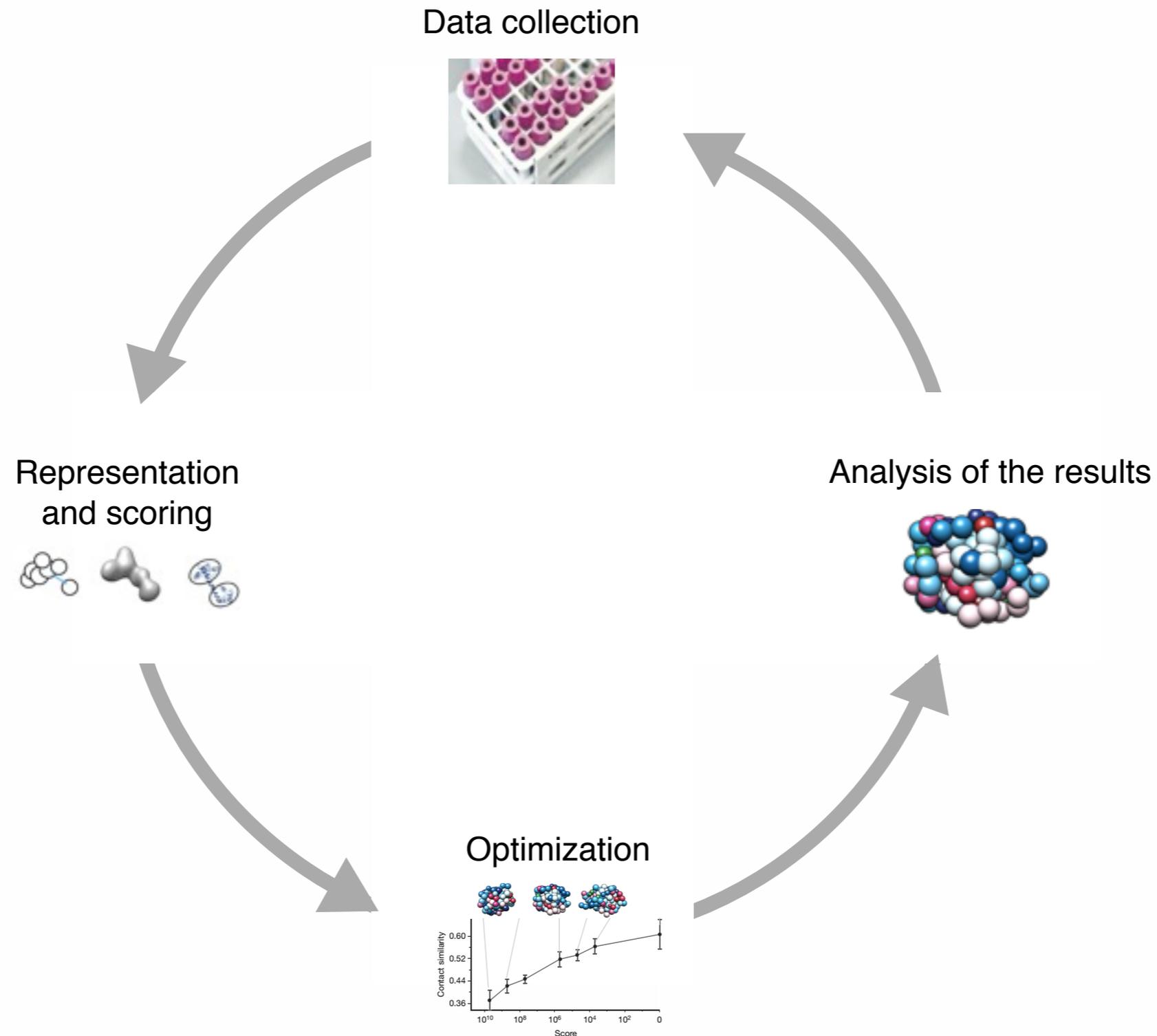


The simulating annealing procedure

Temperature
Movements



The four stages of integrative modeling

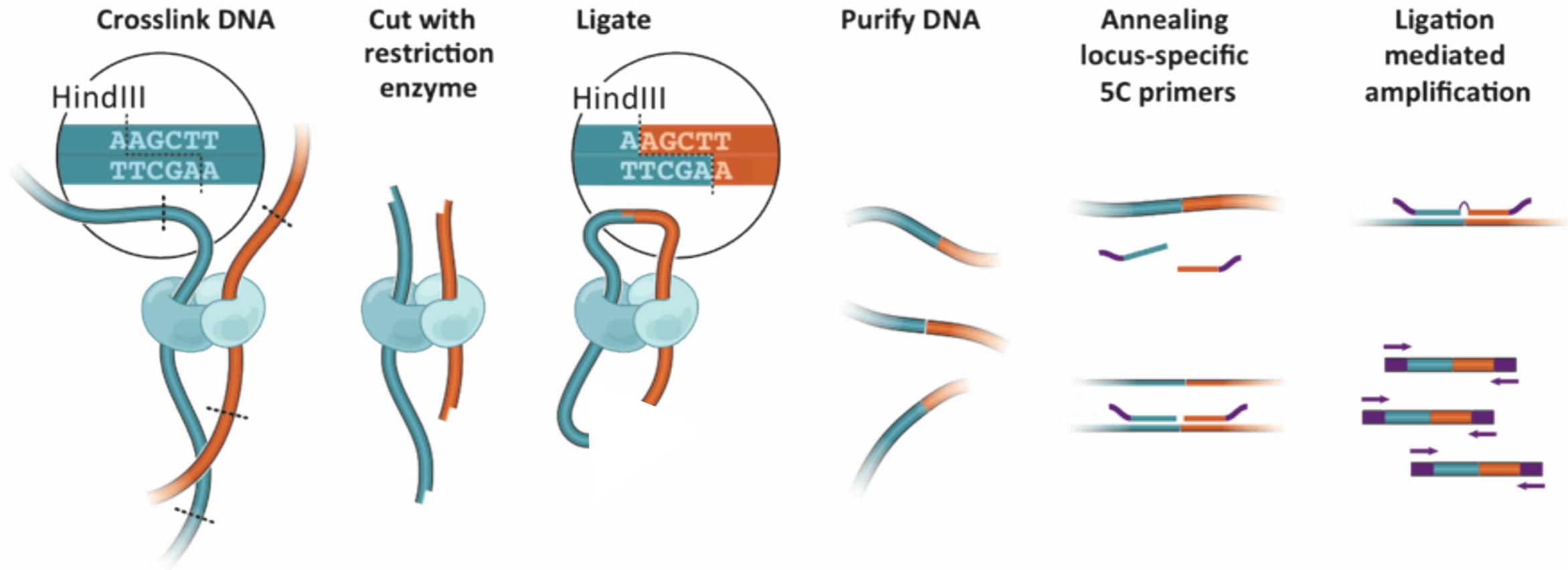




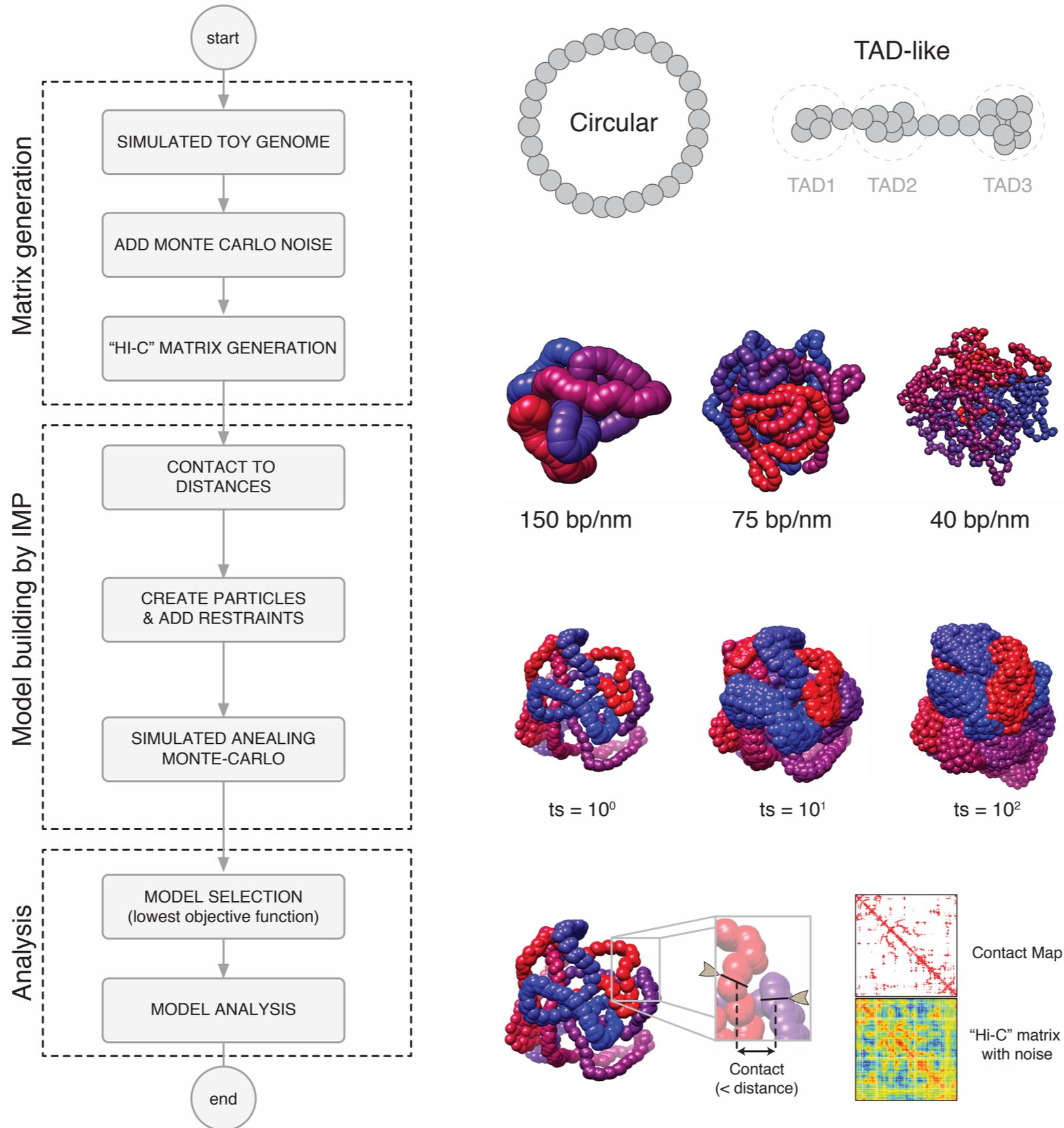
Job Dekker

5C technology

<http://my5C.umassmed.edu>

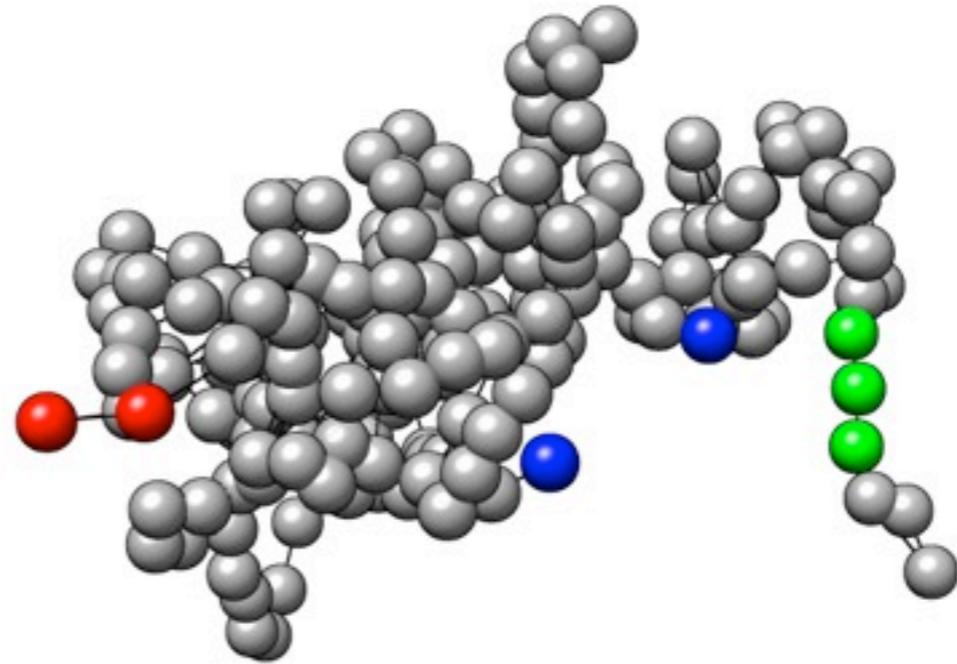
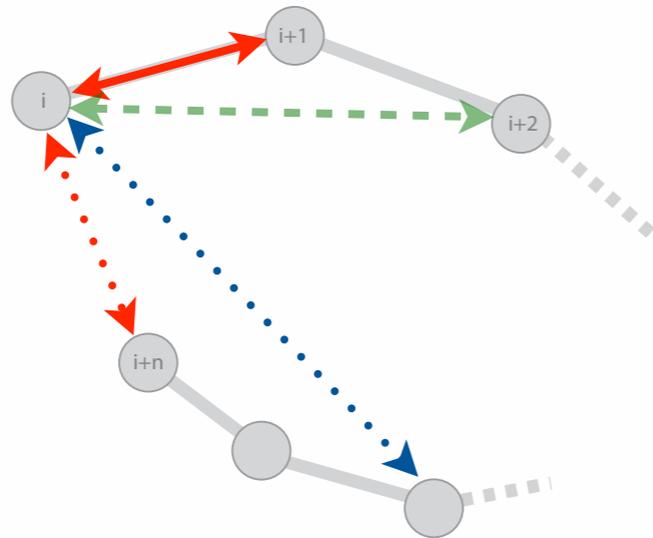


Toy models



Model representation and scoring

Constituent parts of the molecule



$d < d_0$



$d = d_0$

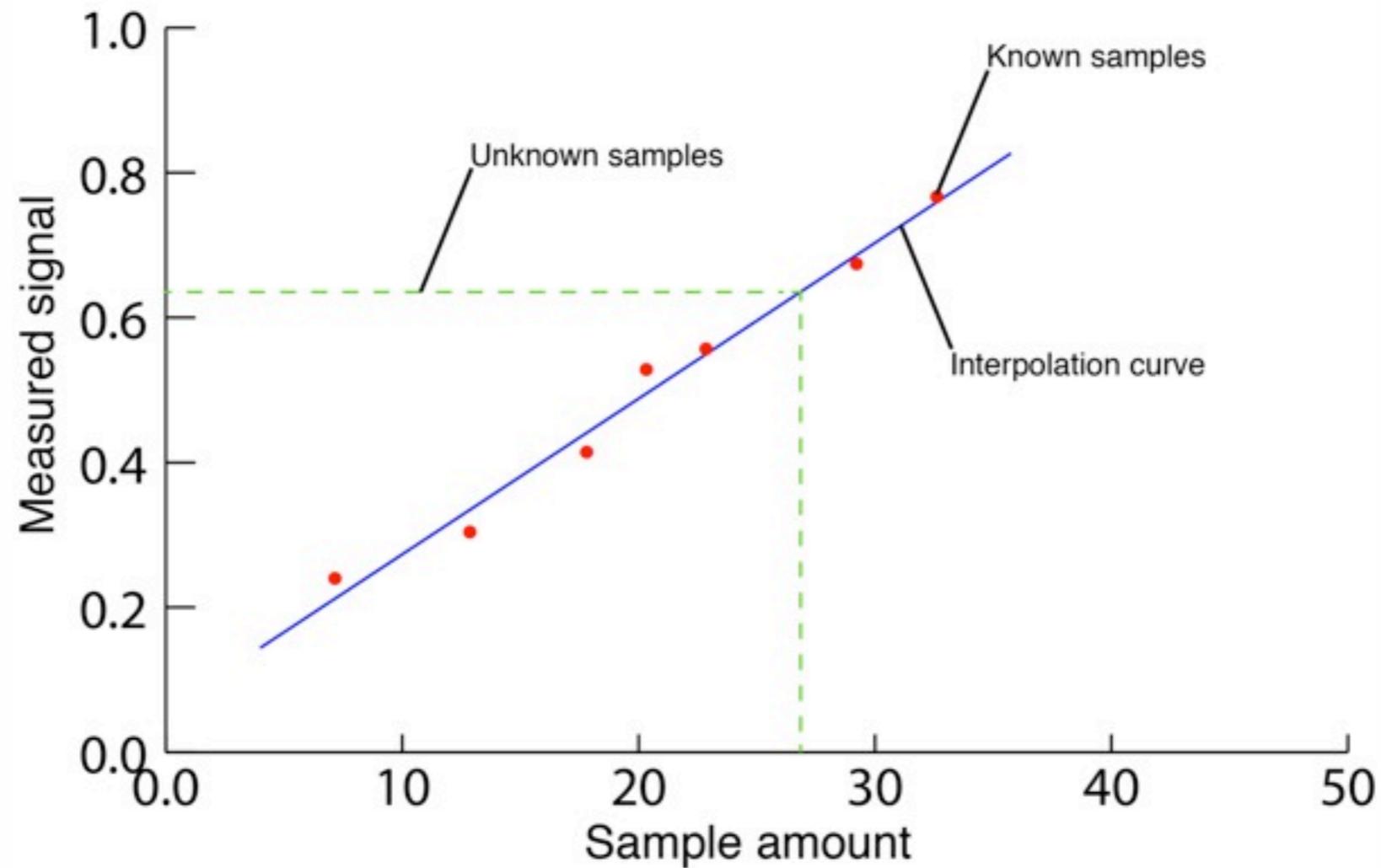


$d > d_0$

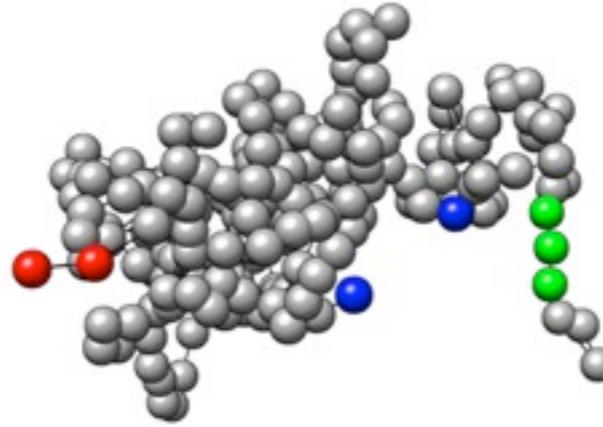


From 5C data to spatial distances

The sample curve

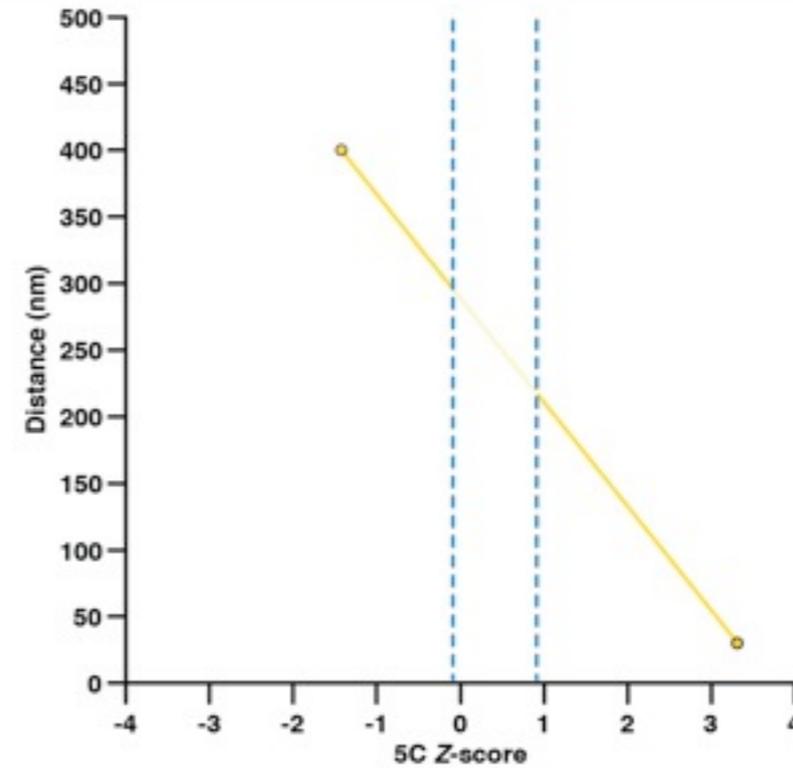
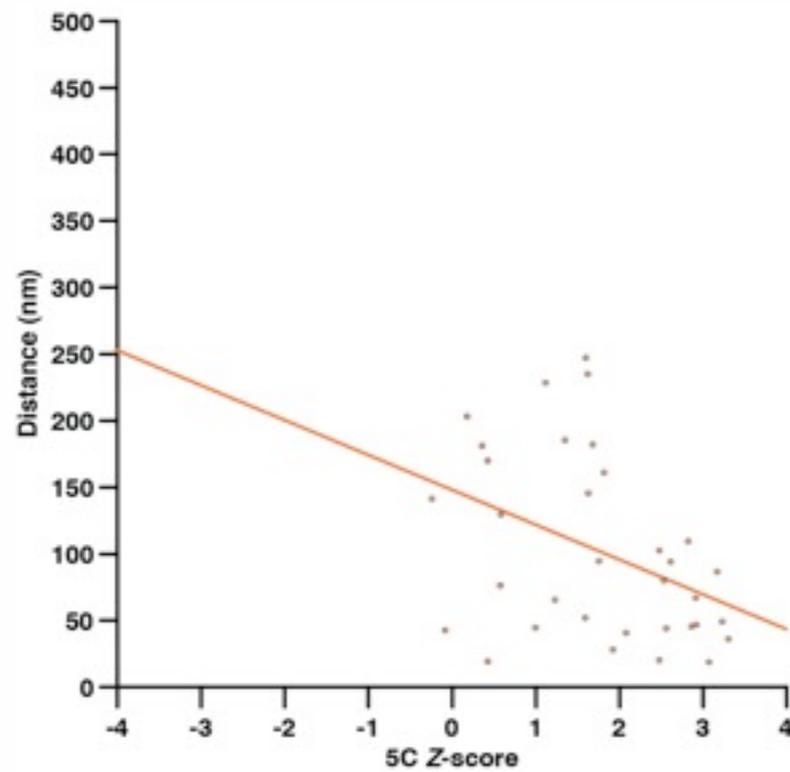


From 5C data to spatial distances

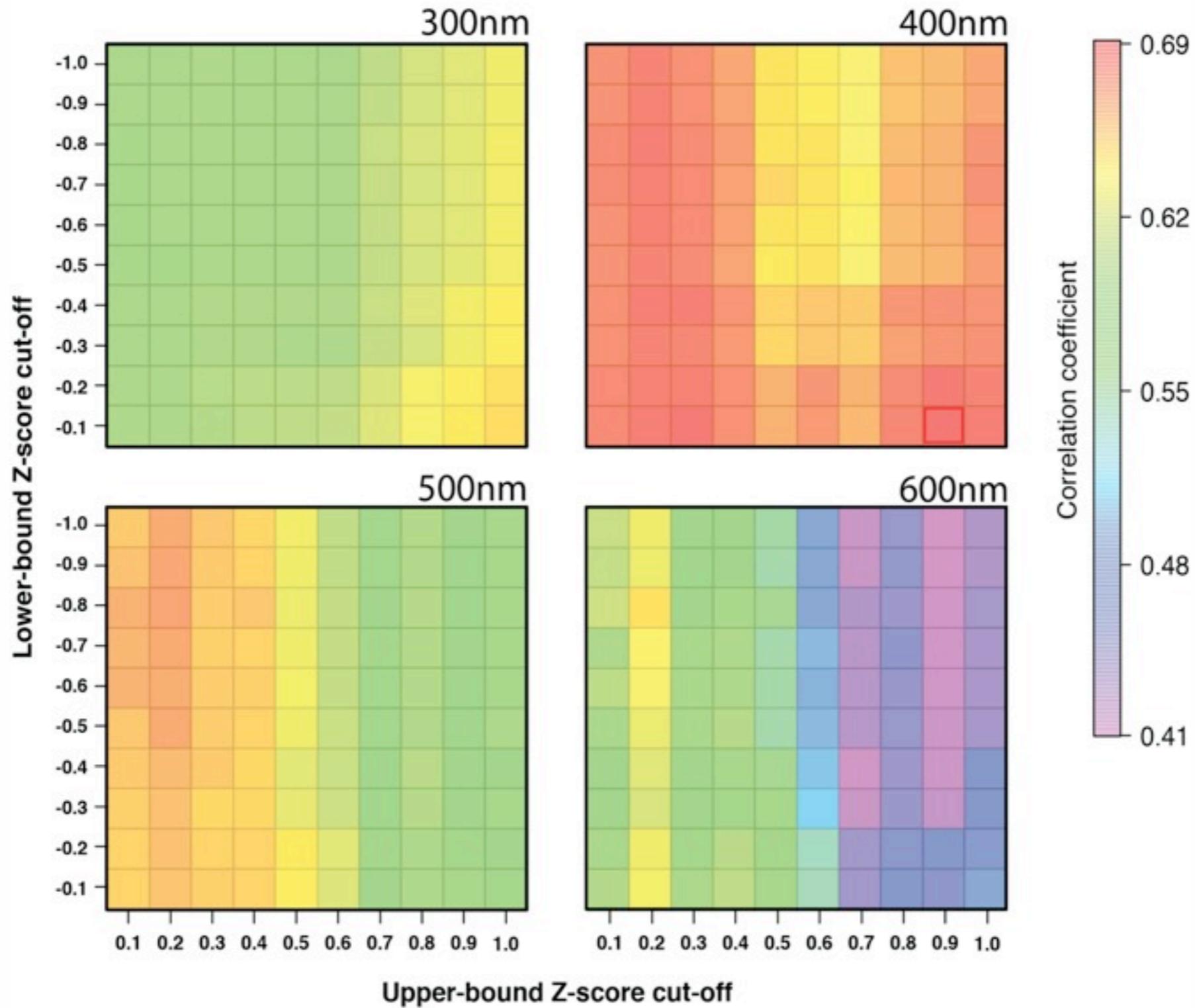


Neighbor fragments

Non-Neighbor fragments

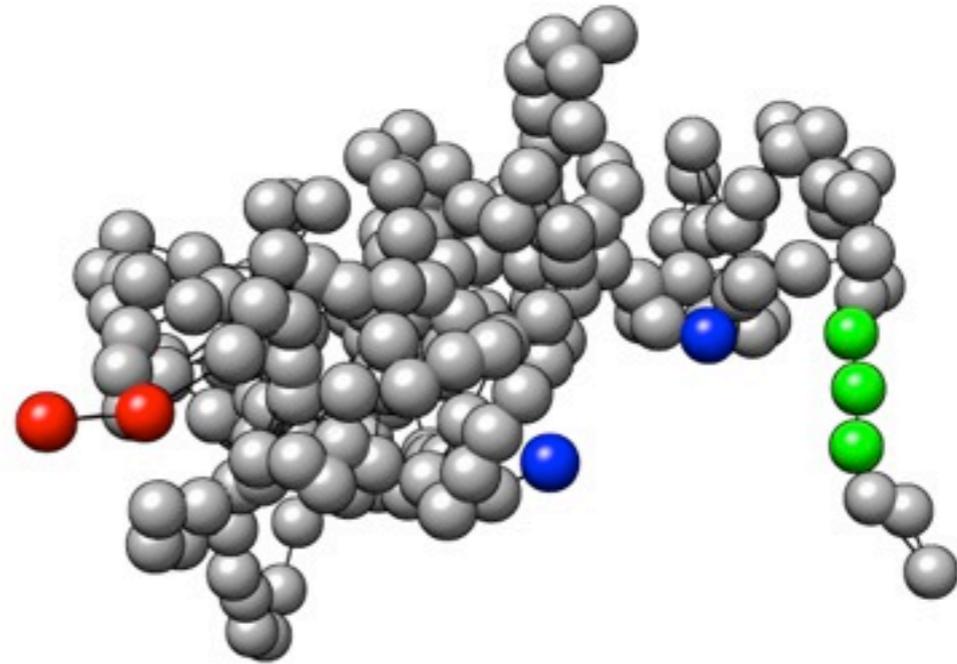
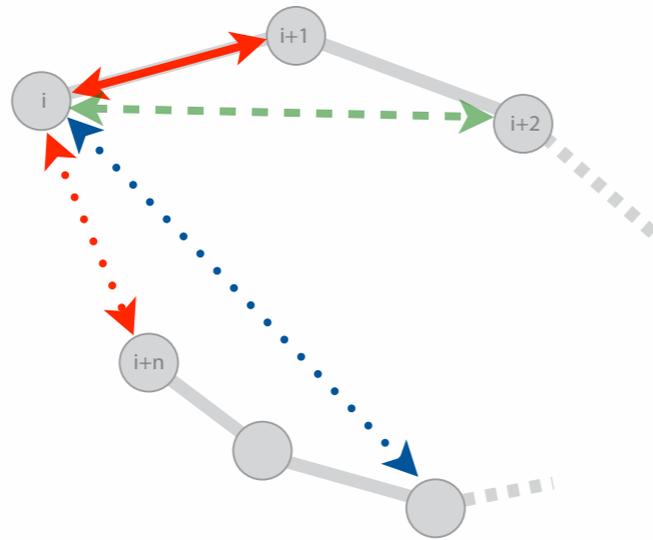


Parameter optimization



Model representation and scoring

Constituent parts of the molecule



$d < d_0$



$d = d_0$



$d > d_0$



Model representation and scoring

Constituent parts of the molecule

Harmonic

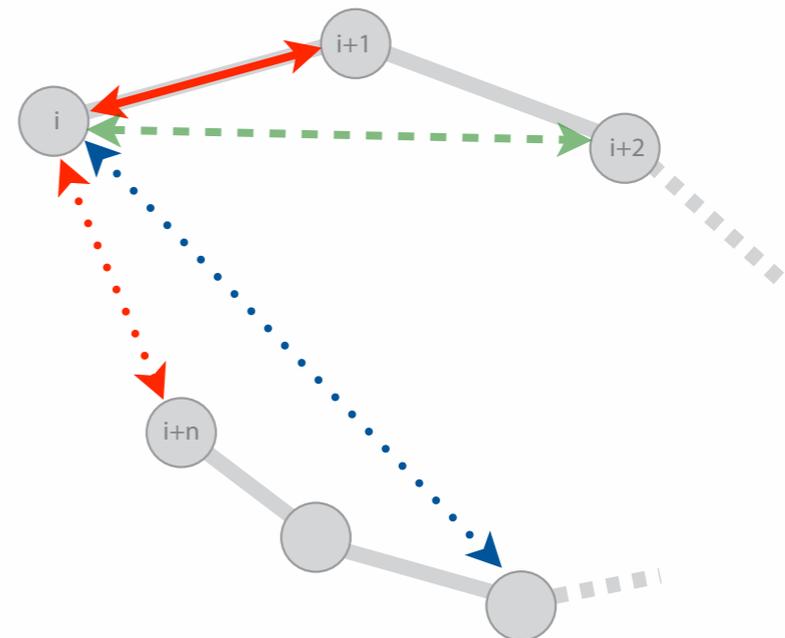
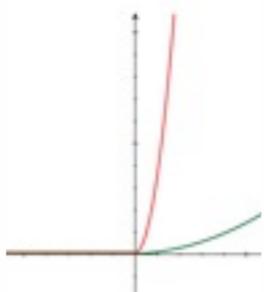
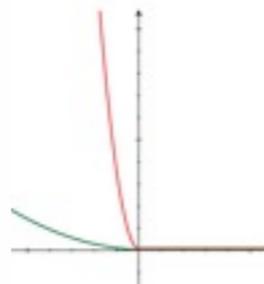
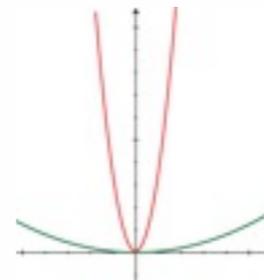
$$H_{i,j} = k(d_{i,j} - d_{i,j}^0)^2$$

Harmonic Lower Bound

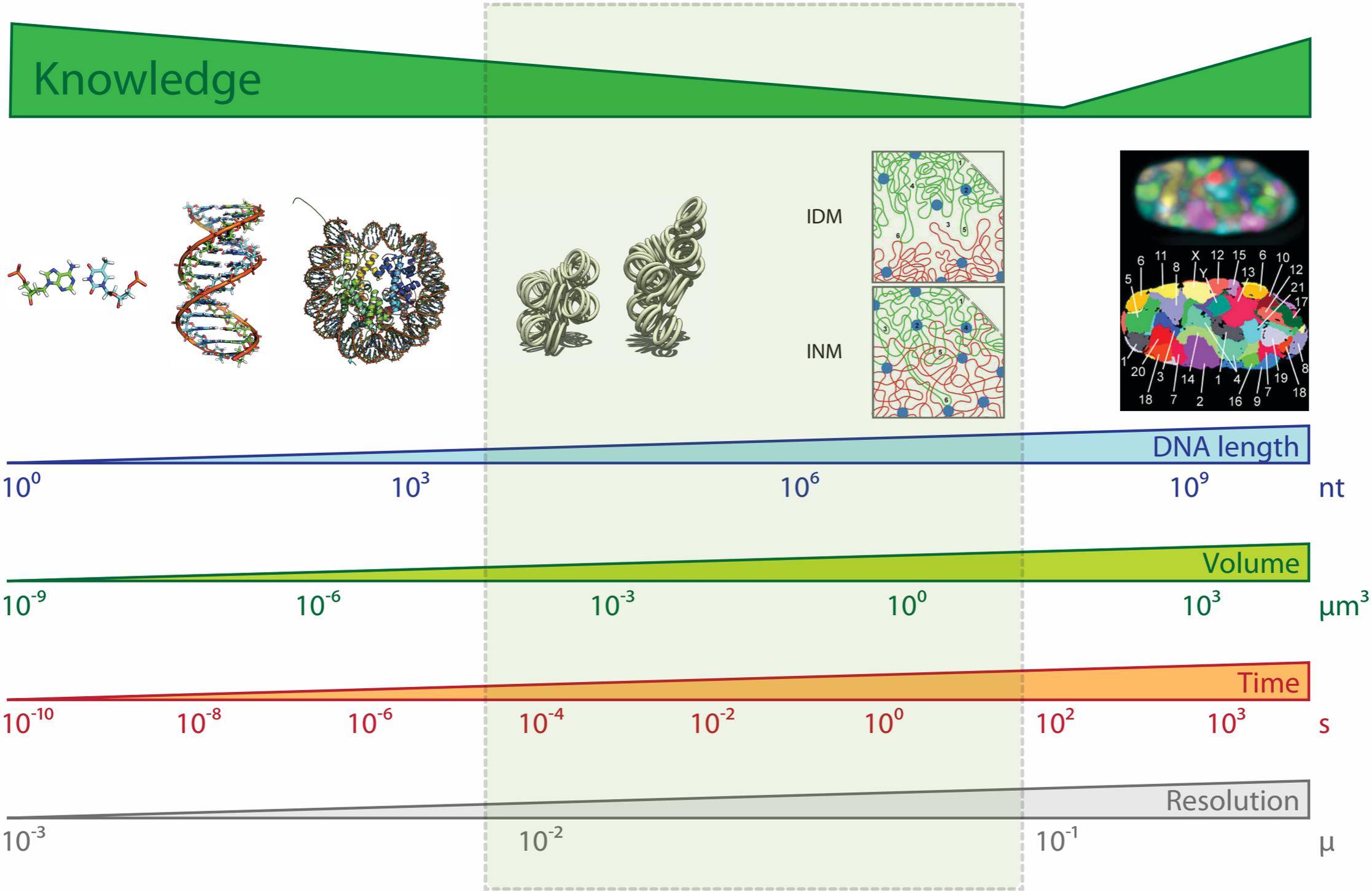
$$\begin{cases} \text{if } d_{i,j} \leq d_{i,j}^0; & lbH_{i,j} = k(d_{i,j} - d_{i,j}^0)^2 \\ \text{if } d_{i,j} > d_{i,j}^0; & lbH_{i,j} = 0 \end{cases}$$

Harmonic Upper Bound

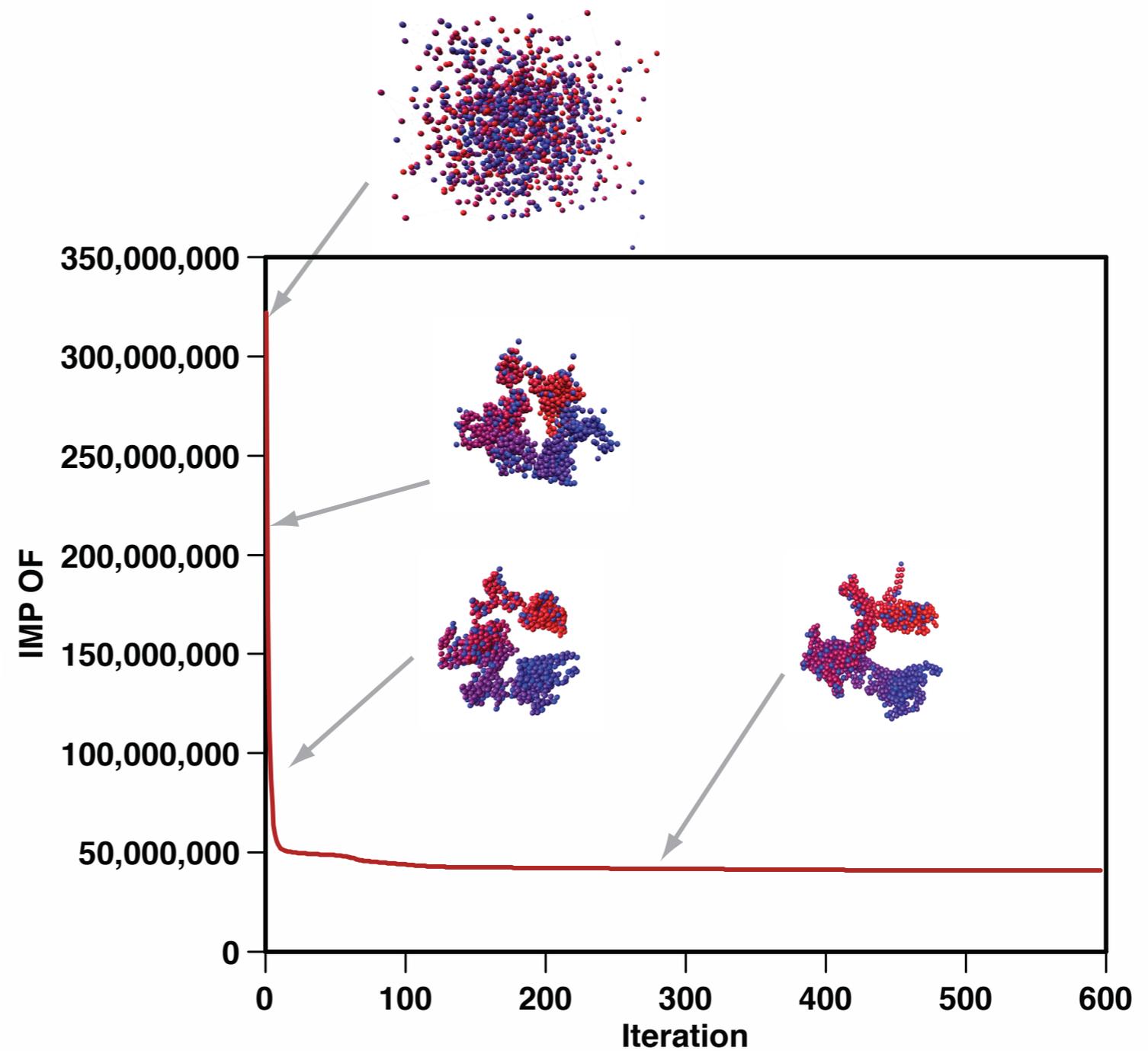
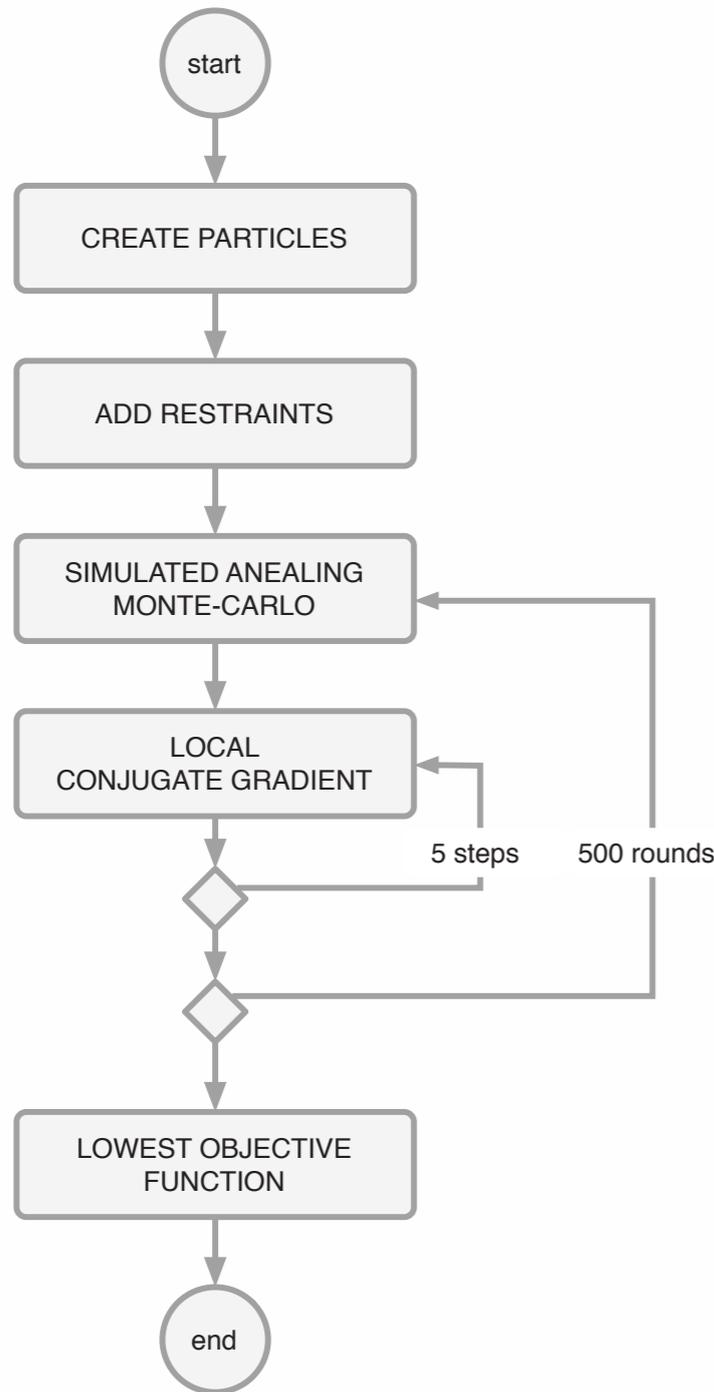
$$\begin{cases} \text{if } d_{i,j} \geq d_{i,j}^0; & ubH_{i,j} = k(d_{i,j} - d_{i,j}^0)^2 \\ \text{if } d_{i,j} < d_{i,j}^0; & ubH_{i,j} = 0 \end{cases}$$



The resolution gap

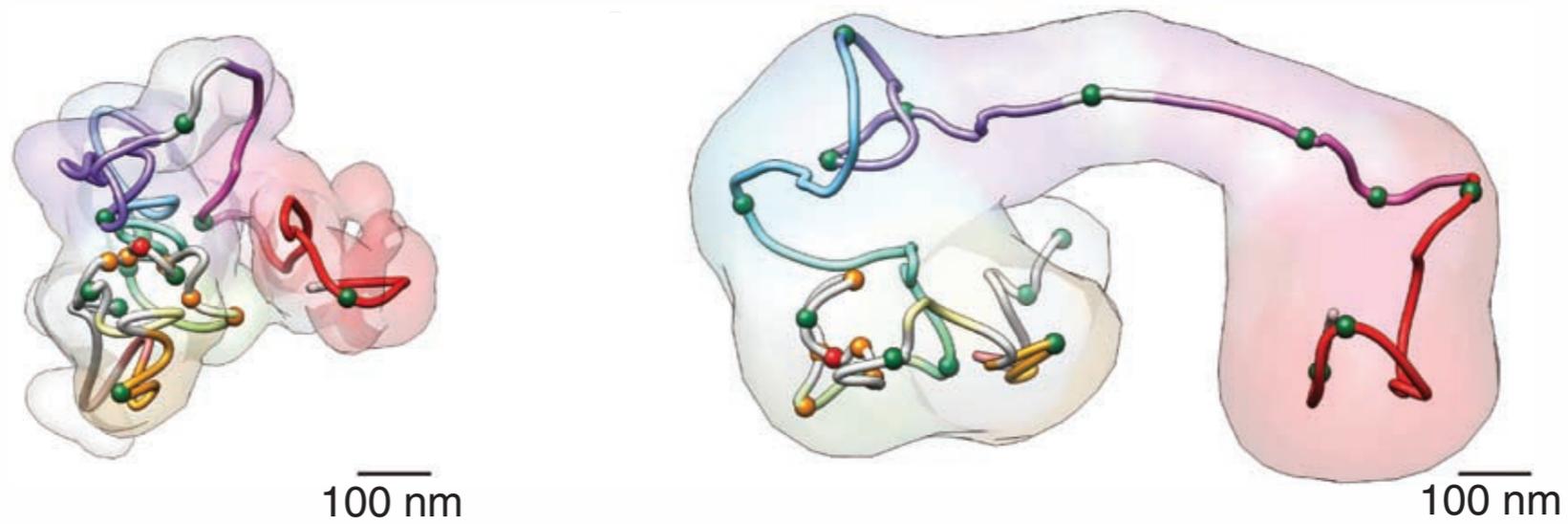


Optimization of the scoring function



Chromatin globules

Frequency contact map differences

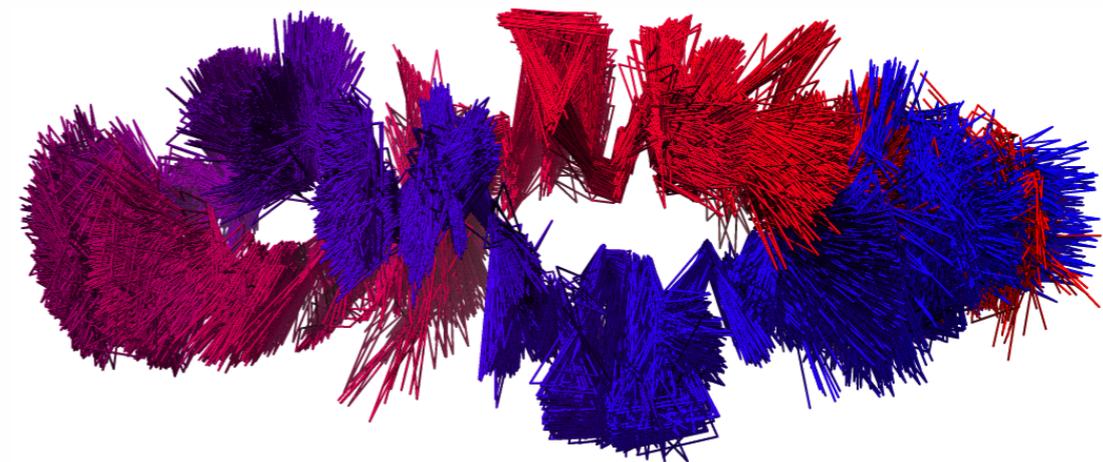
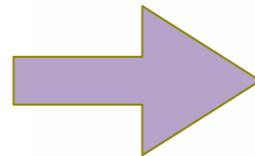
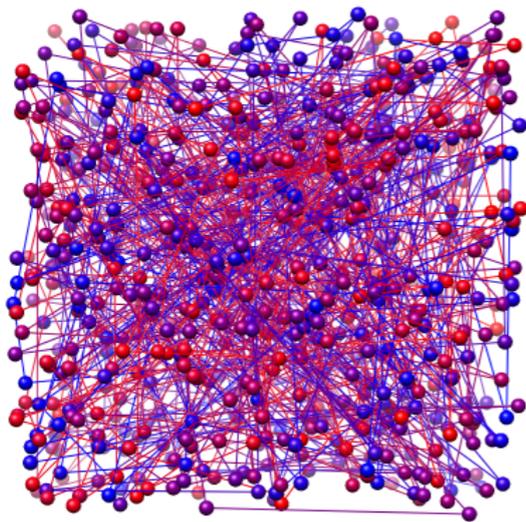
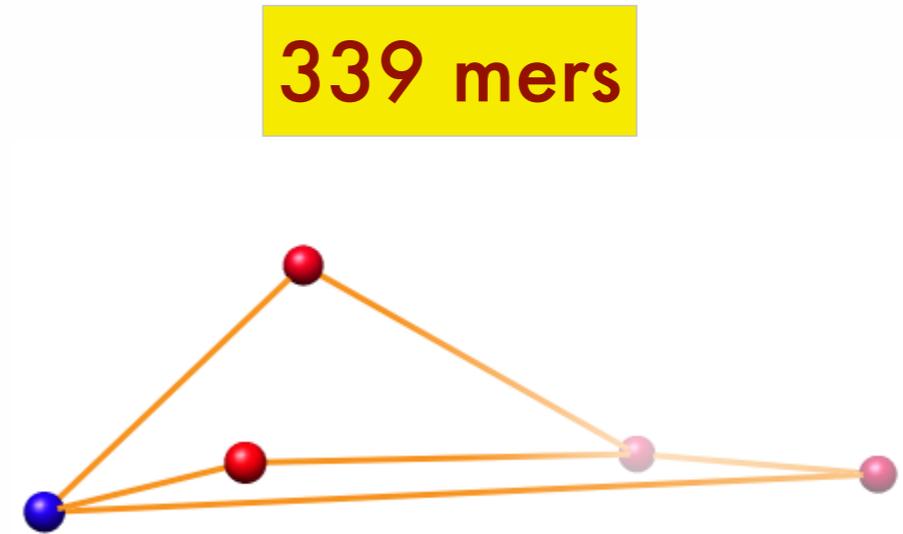
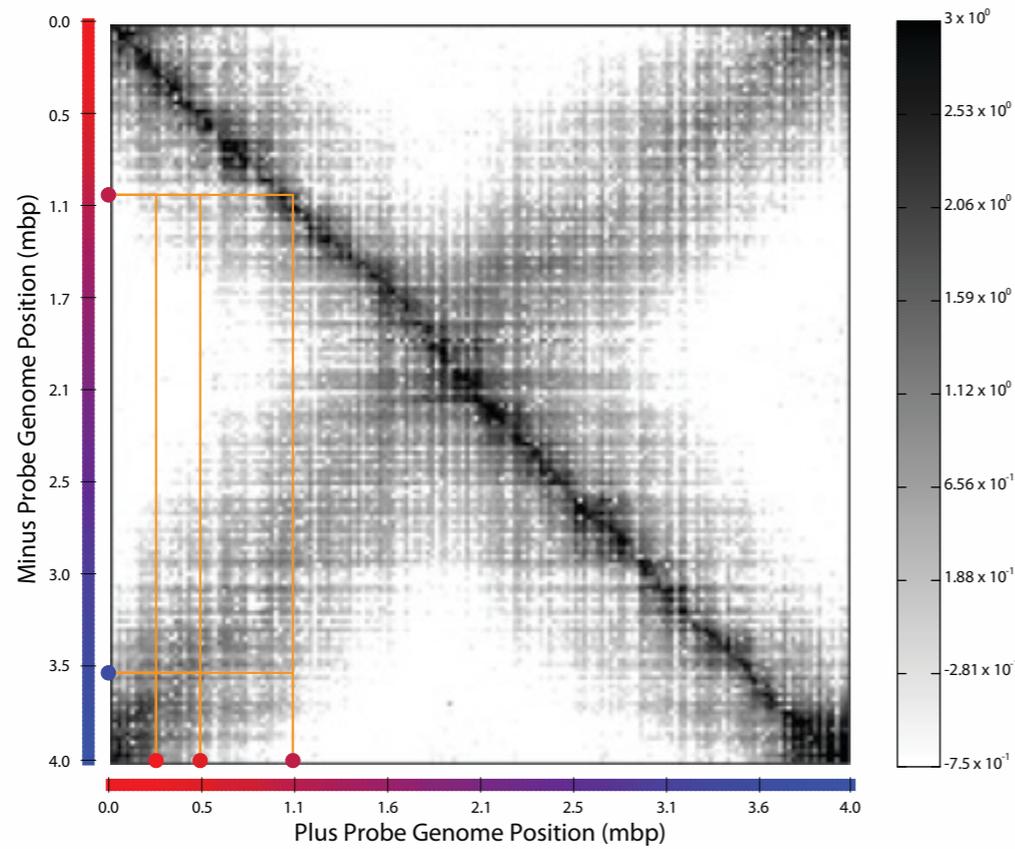


Increased in GM12878



Increased in K562

3D model building with the 5C + IMP approach

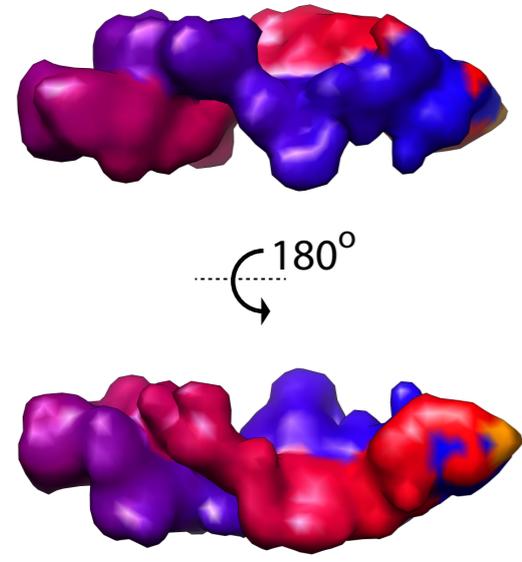


Genome organization in *Caulobacter Crescentus*

Arms are helical

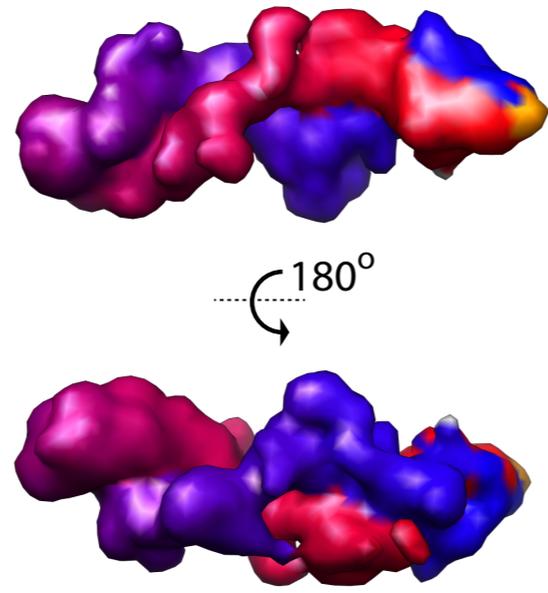


Cluster 1



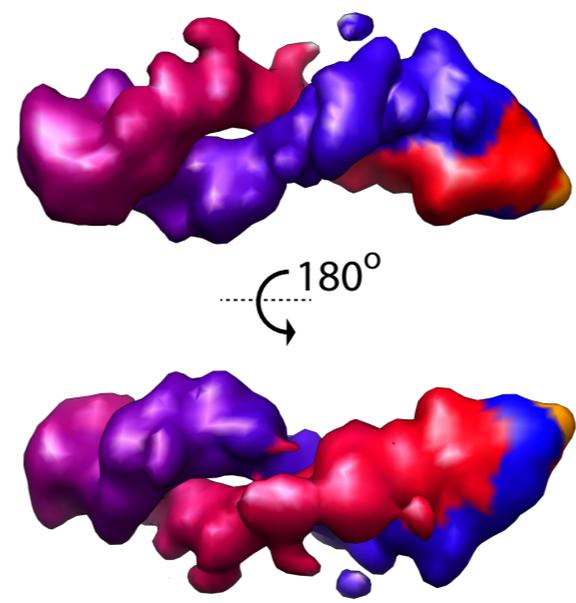
500 nm

Cluster 2



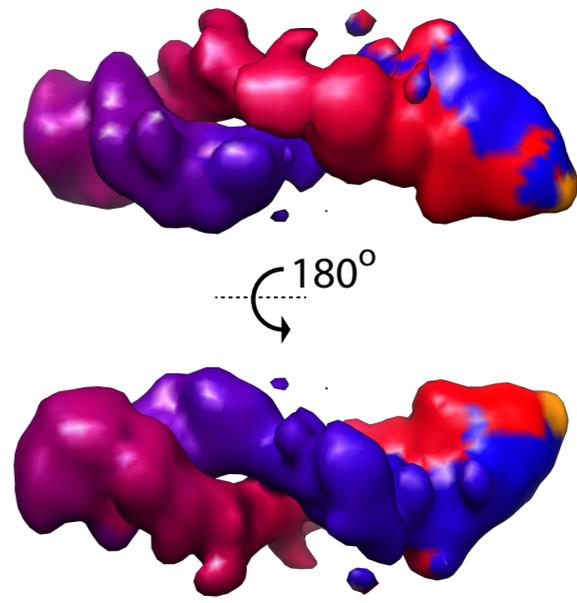
500 nm

Cluster 3



500 nm

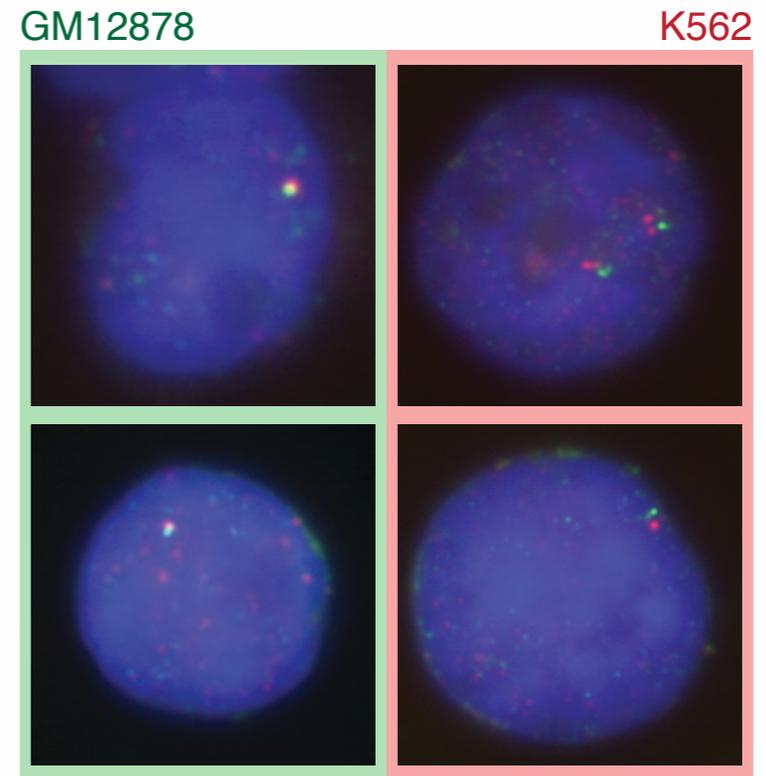
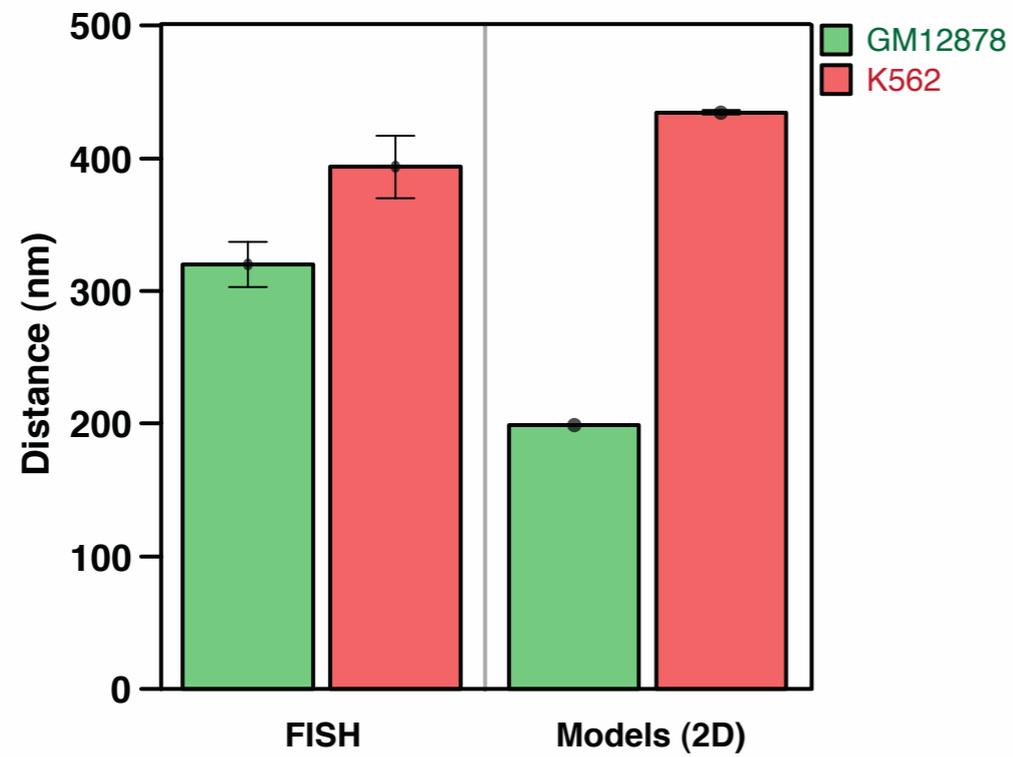
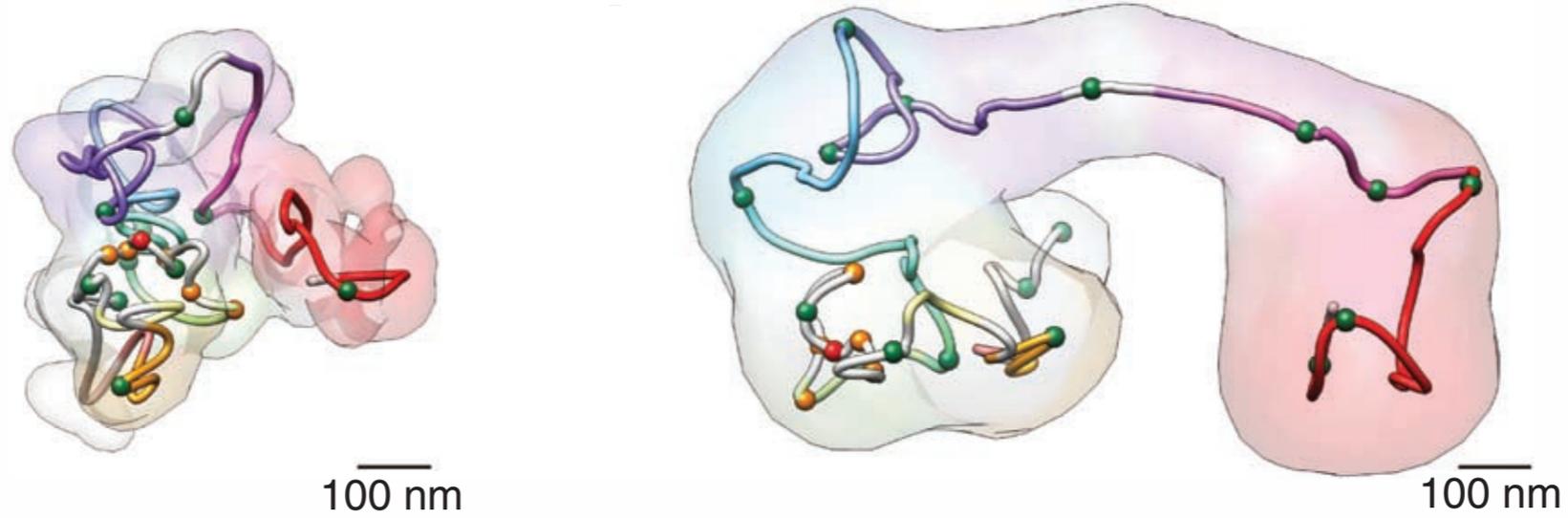
Cluster 4



500 nm

MIRRORS!

Model validation



Take home message

