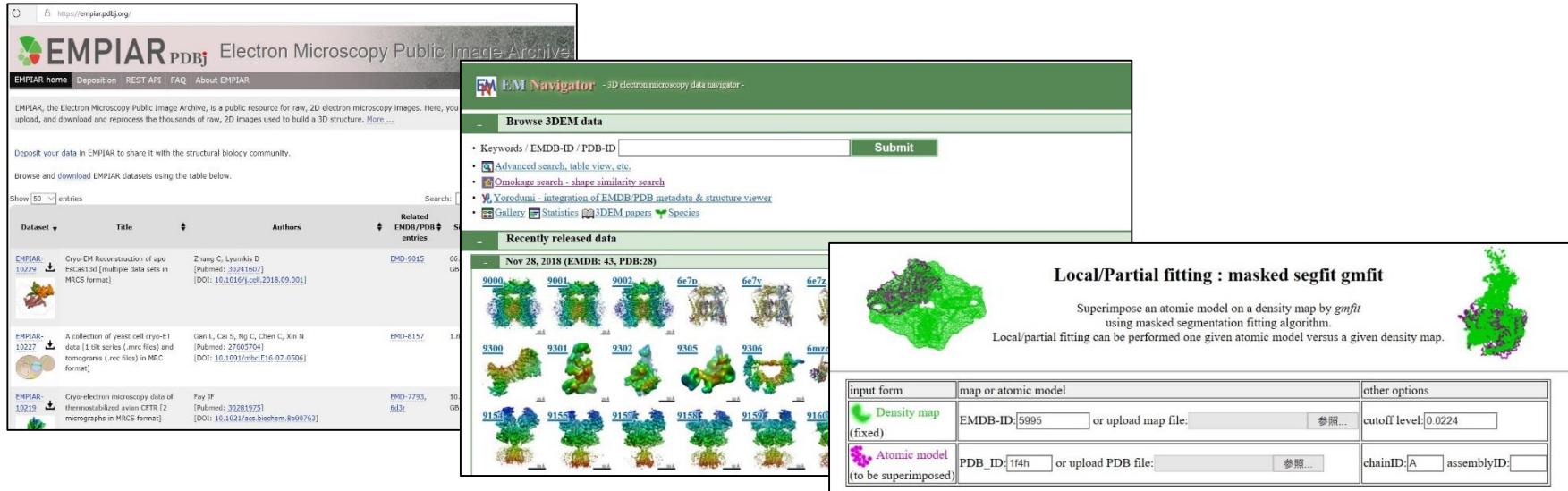


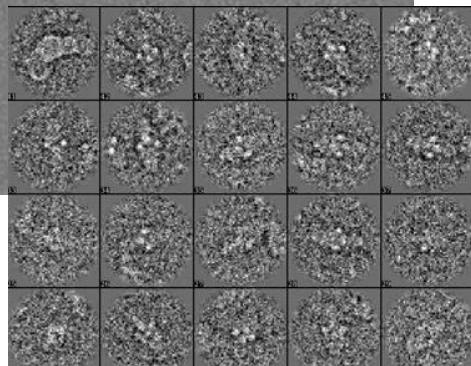
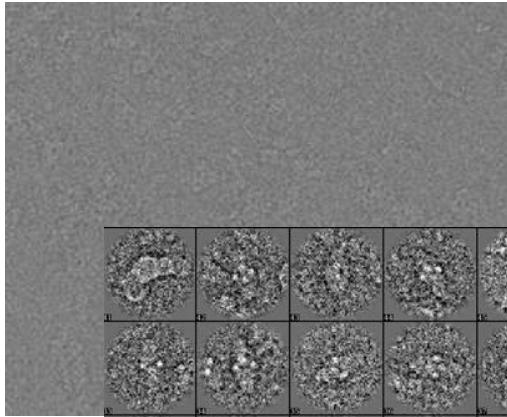
Databases and Web services of **PDBj** for Electron Microscopy

Kawabata, Takeshi; Suzuki, Hirofumi; Kurisu, Genji
Institute for Protein Research, Osaka University



The figure displays two screenshots of scientific databases. On the left is the EMPIAR PDBj website, featuring a search interface and a table of datasets. The table includes columns for Dataset ID, Title, Authors, and Related EMDB/PDB entries. Datasets shown include Cryo-EM Reconstruction of apo Escherichia coli [multiple data sets in MRCG format] by Zhang C, Lymirek D, and A collection of yeast cell cryo-ET data [1 tilt series (.mrc files) and tomograms (.rec files) in MRCG format] by Gan L, Cai S, Ng C, Chen C, Xin N. On the right is the EM Navigator interface, which allows users to browse 3DEM data. It includes a search bar, a list of recently released data (Nov 28, 2018), and a detailed view of a density map labeled "Local/Partial fitting : masked segfit gmfit". This view shows a green density map and a purple atomic model, with input fields for "Density map (fixed)" and "Atomic model (to be superimposed)".

Data processing for EM Single Particle Analysis



2D images

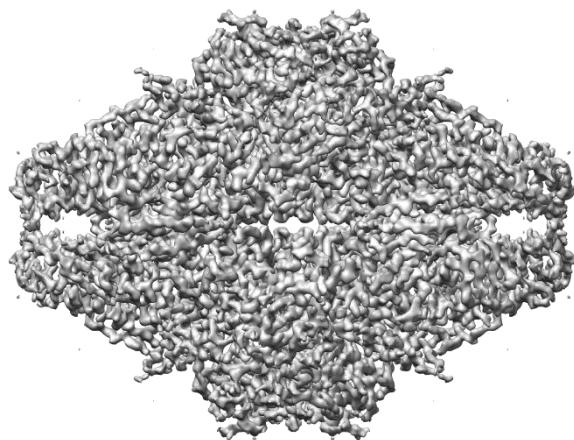
509 DM4 movies (7420 x 7676 x 38; 32 bit)

EMPIAR-10013

402.2GB

EMPIAR

Beta-galactosidase



3D reconstruction

3D image

340 x 340 x 340 voxels

emd_5995.map

150 MB 3.2Å

EMDB

Atomic modeling

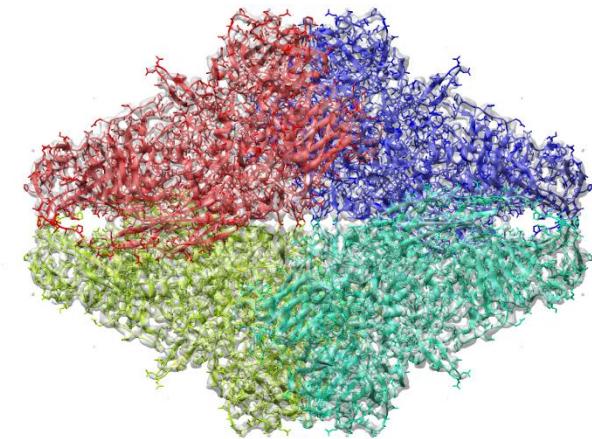
Atomic model

32878 atoms

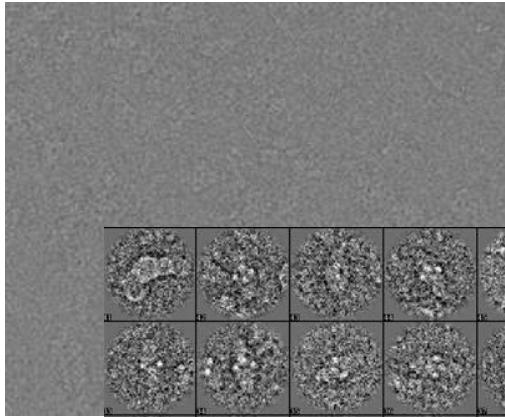
pdb3j7h.ent

2.6 MB

wwPDB



Data processing for EM Single Particle Analysis



2D images

509 DM4 movies (7420 x 7676 x 38; 32 bit)

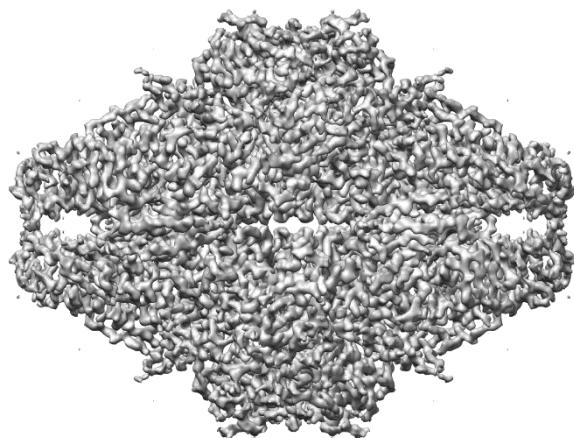
EMPIAR-10013

402.2GB

EMPIAR

EMPIAR-PDBj

Beta-galactosidase



3D reconstruction

3D image

340 x 340 x 340 voxels

emd_5995.map

150 MB

3.2Å

Atomic modeling

Atomic model

32878 atoms

pdb3j7h.ent

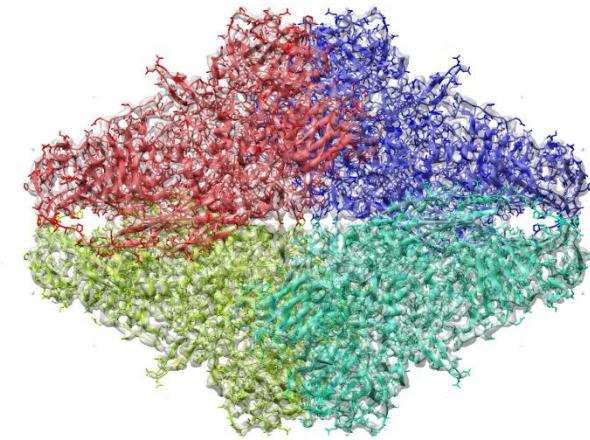
2.6 MB

PDBj

EMDB

EM Navigator

Omokage search
gmfit



EMPIAR : Archive of *raw* 2D images of EM

European Bioinformatics Institute [GB] <https://www.ebi.ac.uk/pdbe/emdb/empiar/>



EMPIAR home

Deposition

REST API

FAQ

About

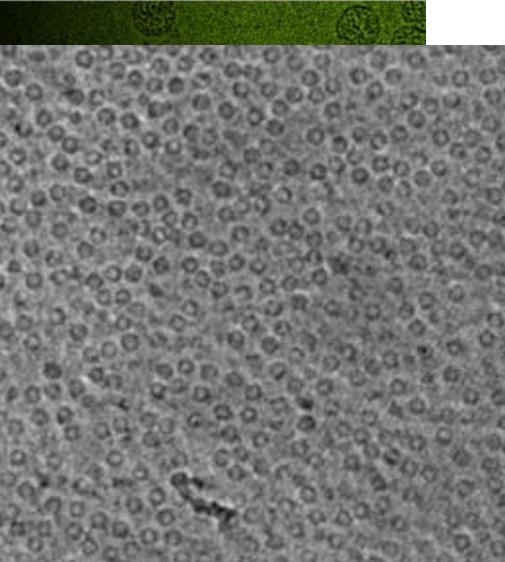
EMPIAR, the Electron Microscopy Public Image Archive, is a public archive where users can browse, upload and download the raw images used to build a 3D density map.

Deposit your data in EMPIAR to share it with the structural biology community.

Browse and download EMPIAR datasets using the table below.

Show 25 entries

Dataset	Title	Author
EMPIAR-10229	Cryo-EM Reconstruction of apo EsCas13d [multiple data sets in MRCs format]	Zhang C, Lyumkis L, et al. [PubMed: 30241131 DOI: 10.1016/j.celrep.2020.321111]
EMPIAR-10227	A collection of yeast cell cryo-ET data [1 tilt series (.mrc files) and tomograms (.rec files) in MRC format]	Gan L, Cai S, Ng K, et al. [PubMed: 27605311 DOI: 10.1093/mnras/stw2611]



Established by Ardan Patwardhan (EBI)



2D images for published 3D maps (most are in EMDB)

Purposes

- 1) Validation of 3D density maps in EMDB
- 2) Enhancing developments of new software
- 3) Training and education

Data size is huge ! (**110 TB** in total, 187 entries)

We are building the mirror site in Japan (**EMPIAR-PDBj**).

Mirror site of EMPIAR in PDBj

Cambridge(UK)

Osaka(Japan)

Time to transfer **1TB**

Osaka -> Tokyo : 6-7 hours

UK -> Japan : 15 hours

(using **Aspera**)

The screenshot shows the EMPIAR homepage with a search bar and a table of datasets. One entry is highlighted:

Dataset	Title	Authors	Related EMDB/PDB entries	Size	Resolution
EMPIAR-10229	Cryo-EM Reconstruction of apo EsCas13d [multiple data sets in MRCS format]	Zhang C, Lyumkis D [Pubmed: 30241607] [DOI: 10.1016/j.cell.2018.09.001]	EMD-9015	66.9 GB	6.5

<https://www.ebi.ac.uk/pdbe/emdb/empiar/>

The screenshot shows the EMPIAR-PDBj mirror site with a search bar and a table of datasets. One entry is highlighted:

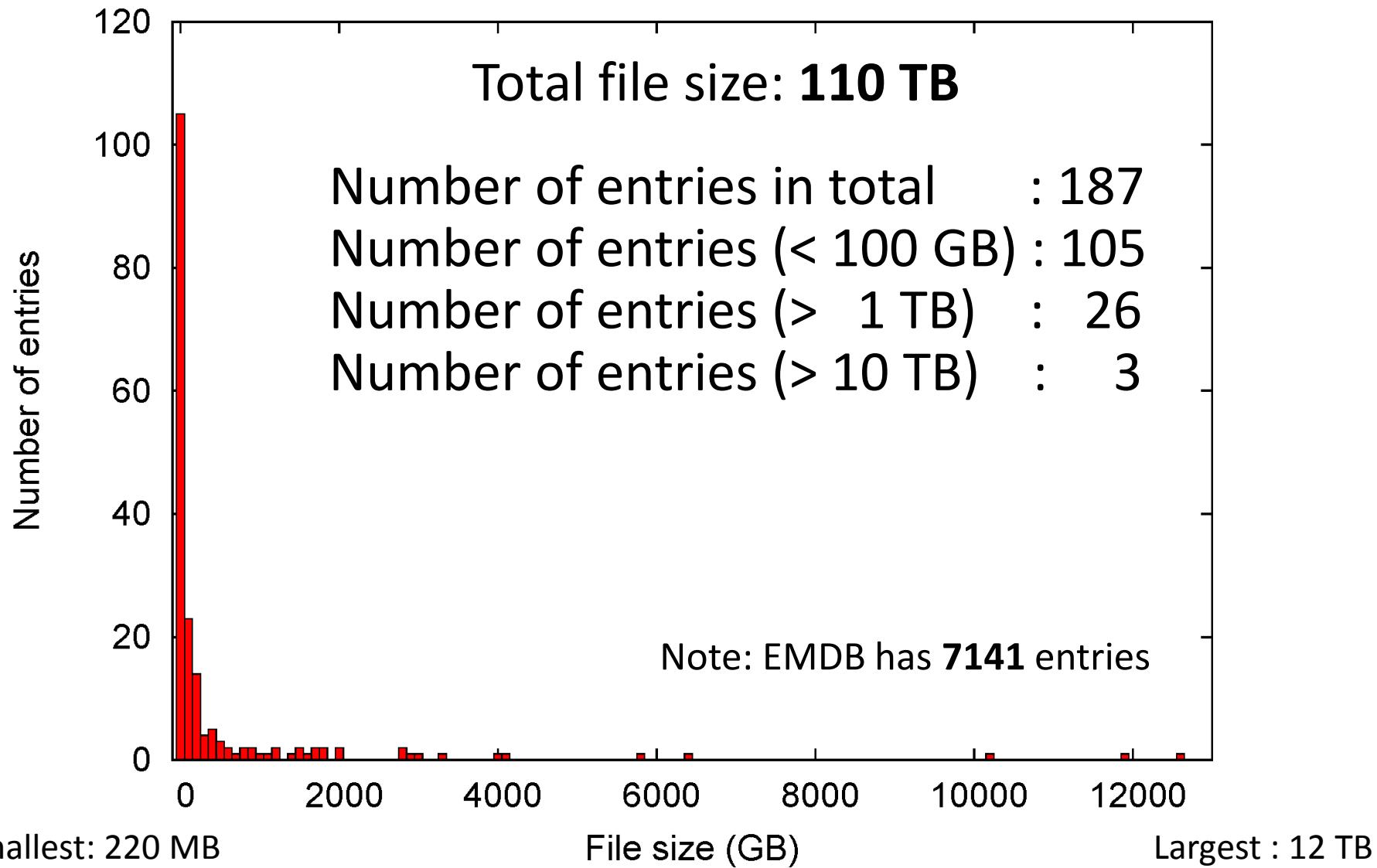
Dataset	Title	Authors	Related EMDB/PDB entries	Size	Resolution
EMPIAR-10229	Cryo-EM Reconstruction of apo EsCas13d [multiple data sets in MRCS format]	Zhang C, Lyumkis D [Pubmed: 30241607] [DOI: 10.1016/j.cell.2018.09.001]	EMD-9015	66.9 GB	6.5

<https://empiar.pdbj.org>

EMPIAR-PDBj provides a faster access for Asia and Oceania.

We also plan to open the ***data-deposition site*** in Japan.

File size distribution of EMPIAR



EMPIAR-10127

Images for Picked Particles (84.4 GB)

169 MRCs (320 x 320 x ~100; 32 bit)
Human TRPM4 ion channel in a lipid nanodisc in a calcium-bound state

Publication:

Structure of the human TRPM4 ion channel in a lipid nanodisc

Autzen HE  , Cheng Y 

Science 359 228-232 (2017)

PMID: 29217581

DOI: 10.1126/science.aar4510

2.2 Å

Related PDB entry:

6bqv

Related EMDB entry:

EMD-7133

Deposited:

2017-11-30

Released:

2018-01-22

Last modified:

2018-01-22

Dataset size:

84.4 GB  

Dataset DOI:

10.6019/EMPIAR-10127

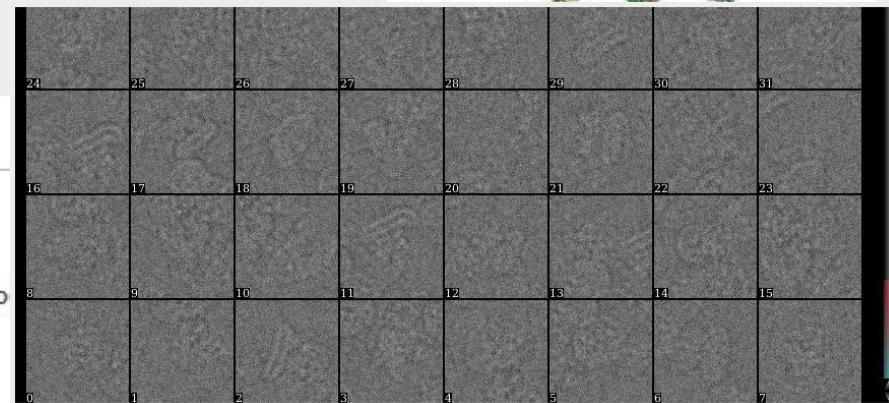
84.4 GB

Many small 2D images

Contains:



picked particles



+ Image set

+ particle stacks of TRPM4 particles post 2D clean-up

Category:

picked particles - multiframe - pro

Image format:

MRCs

No. of images or tilt series:

221100

Frames per image:

169

Image size:

(320, 320)

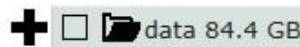
169 MRCs (320 x 320 x ~100; 32 bit)

Pixel type:

32 BIT FLOAT

Details:

Particle positions in the corresponding .star files. Files have variable number of frames with maximum number being 169.



data 84.4 GB

Necessary to validate *resolution* of 3D map by FSC.

EMPIAR-10061

Multi-frame images (movies) (12.4 TB)

2.2 Å resolution cryo-EM structure of beta-
with a cell-permeant inhibitor

Publication:

2.2 Å resolution cryo-EM sti

with a cell-permeant inhibitor

Bartesaghi A, Merk A, Banerjee S, Matthies D, Wu X, Milne JL,

Subramaniam S

Science 348 1147-1151 (2015)

PMID: 25953817

DOI: 10.1126/science.aab1576

Related PDB entry:

5a1a

Related EMDB entry:

EMD-2984

Deposited:

2016-04-11

Released:

2016-04-15

Last modified:

2016-05-13

Dataset size:

12.4 TB



Dataset DOI:

10.6019/EMPIAR-10061

Version history:

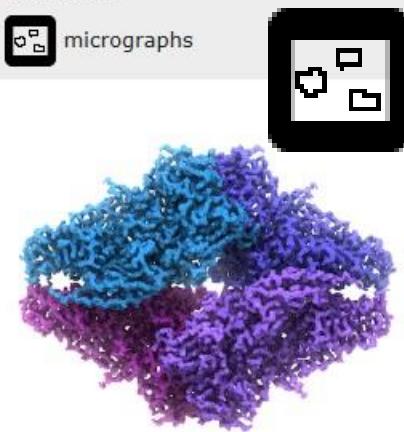
1 2016-05-13

Added the particle co
2984_boxes.tbz archive

2.2 Å

12.4 TB

micrographs



Averages of aligned movie frames

Category: micrographs - single frame

Image format: MRC

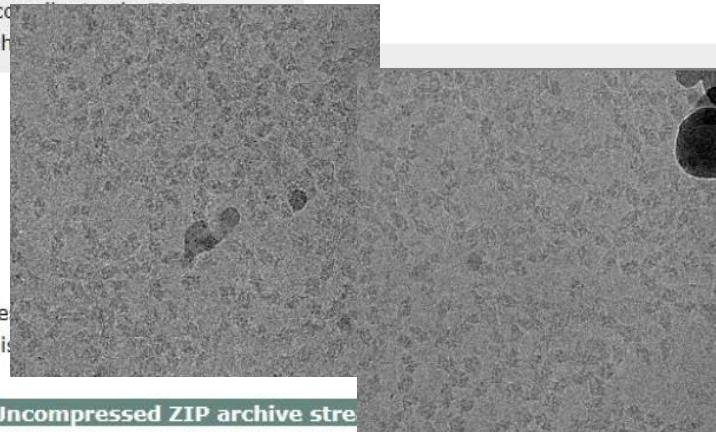
No. of images or tilt series: 1539

Image size: (7676, 7420)

Pixel type: 32 BIT FLOAT

Details: Micrographs are the average of frame

EMD-2984_boxes.tbz archive. There is
in EMAN's box format.



+ Micrographs 326.5 GB

Uncompressed ZIP archive stream

Raw movie frames

1539 MRC movies (7676 x 7420 x 38; 32 bit)

Category: micrographs - multiframe

Image format: MRC

No. of images or tilt series: 1539

Frames per image: 38

Image size: (7676, 7420)

Pixel type: 32 BIT FLOAT

Real *raw* data (too *raw*?)

+ Movies 12.1 TB

Uncompressed ZIP archive streamed via HTTP

(3710 x 3838 x 1338; 16 bit)

The first reconstruction of beta-galactosidase so

Publication:First data of beta-galactosidase for validation of the state-of-the-art-cryo EM, named CRYOARM200
Kato T, Terahara N, Namba K**Related EMDB entry:**

EMD-6840

Deposited:

2018-08-01

Released:

2018-08-15

Last modified:

2018-08-17

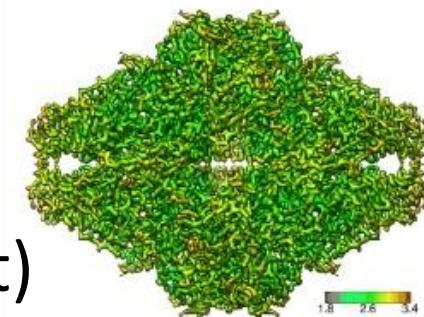
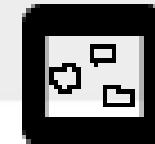
Dataset size:321.4 GB  

2.6 Å

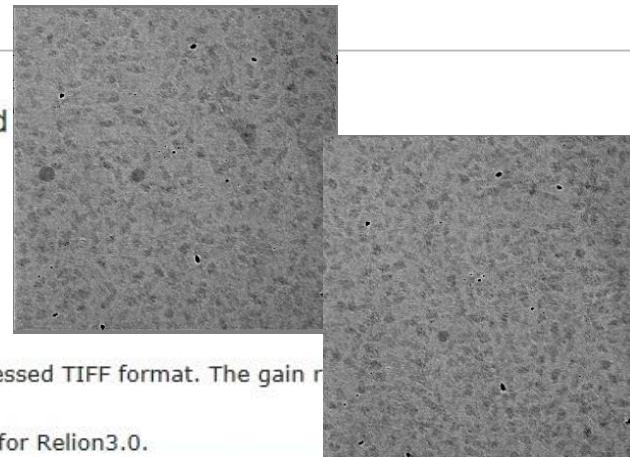
321.4 GB

Dataset DOI:[10.6019/EMPIAR-10204](https://doi.org/10.6019/EMPIAR-10204)**Contains:**

micrographs



1338 TIFF images (3710 x 3838; 16 bit)
with a gain-reference file (K2-gain170629.mrc)

+ Image set**+ Unaligned multi-frame micrographs of beta-galactosidase recorded****Category:** micrographs - multiframe**Image format:** TIFF**No. of images or tilt series:** 1338**Frames per image:** 49**Image size:** (3710, 3838)**Pixel type:** UNSIGNED 16 BIT INTEGER**Details:** Raw, non aligned micrograph moves in compressed TIFF format. The gain reference file is in MRC format. It is located in the same directory in MRC format.**+  data 321.4 GB****Uncompressed ZIP archive streamed via HTTP**

EMPIAR-10013

Multi-frame images (movie) (442 GB)

(7420x 7676 x 38; 32 bit)

Structure of β -galactosidase at 3.2- \AA resolution of electron microscopy

Publication:

Structure of beta-galactosidase at 3.2- \AA resolution obtained by cryo-electron microscopy

Bartesaghi A, Matthies D, Banerjee S, Merk A, Subramaniam S

Proc. Natl. Acad. Sci. USA **111** 11709-11714 (2014)

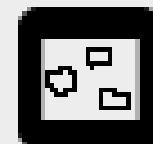
PMID: [25071206](#)

3.2 \AA

Contains:



micrographs



Related PDB entry:

[3j7h](#)

Related EMDB entry:

EMD-5995

Deposited:

2014-07-23

Released:

2014-08-07

Last modified:

2014-08-07

Dataset size:

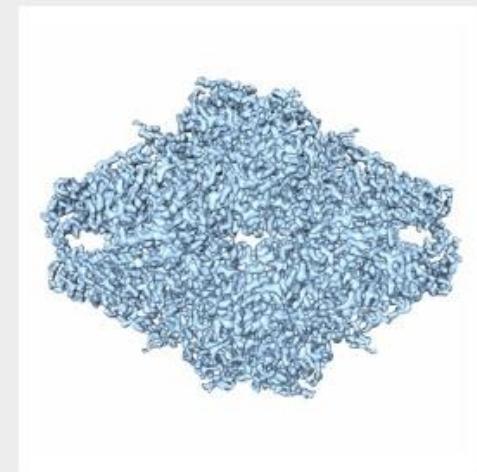
442.5 GB

Dataset DOI:

[10.6019/EMPIAR-10013](#)

Version history:

1 2015-07-01 Directory structure reorganized. Entry has now full set of micrographs including frame alignment shifts, particle locations and defoci.



+ Image sets

509 MRC images (7420 x 7676 x 38; 32 bit)

+ Full set of multi-frame micrographs

Category: micrographs - multiframe

Image format: MRC

No. of images or tilt series: 509

Frames per image: 38

Image size: (7420, 7676)

Pixel type: 32 BIT FLOAT

Details: Full set of multiframe micrographs related to the dataset:10012. Related PDB: 3J7H.

Please note that the number of frames varies between [5, 10, 25 and 38]!

Single-frame Images (Still image) (108 GB)

EMPIAR-10012

(7420x 7676 x 1; 32 bit)

Structure of β -galactosidase at 3.2- \AA resolution obtained by cryo-electron microscopy (frame-averaged micrographs)

Publication:

Structure of beta-galactosidase at 3.2- \AA resolution obtained by cryo-electron microscopy

3.2 \AA

Bartesaghi A, Matthes D, Banerjee S, Merk A, Subramaniam S

Proc. Natl. Acad. Sci. USA **111** 11709-11714 (2014)

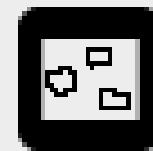
PMID: 25071206

DOI: 10.1073/pnas.1402809111

108.0 GB

Contains:

 micrographs



Related PDB entry:

3J7H

Related EMDB entry:

EMD-5995

Deposited:

2014-07-22

Released:

2014-08-07

Last modified:

2014-08-07

Dataset size:

108.0 GB  

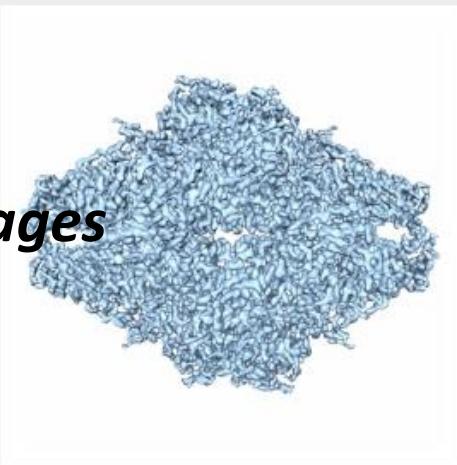
Dataset DOI:

10.6019/EMPIAR-10012

Version history:

1 2015-10-28 Restructured the entry directory to have "data" and ".private".

Motion-corrected still images



+ Image set

509 MRC images (7420 x 7676 x 1; 32 bit)

+ Beta Galactosidase micrographs

Category: micrographs - single frame

Image format: MRC

No. of images or tilt series: 509

Image size: (7420, 7676)

Pixel type: 32 BIT FLOAT

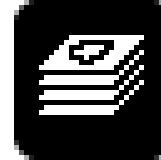
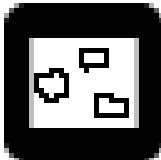
Details: These images were generated by aligning and then averaging multiframe images. All frames (1-38) from the unprocessed data were used. Related PDB: 3J7H.

+  data 108.0 GB

 Uncompressed ZIP archive streamed via HTTP

Various Types of data in EMPIAR

1) Micrographs or Particle images



2) Multiple-frame or Single-frame (Movie) (Still image)

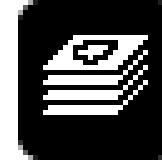
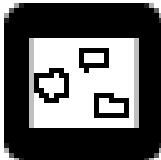
3) Gray scale depth: 32 bit or 16 bit

Different opinions about what is *raw* data

We accept everything you can provide.

Various Types of data in EMPIAR

1) Micrographs or Particle images



2) Multiple-frame or Single-frame (Movie) (Still image)

3) Gray scale depth: 32 bit or 16 bit

Different opinions about what is *raw* data

We accept everything you can provide.

Let's submit your 2D images to EMPIAR !!

EM Navigator :lovely EMDB server, user-friendly interface



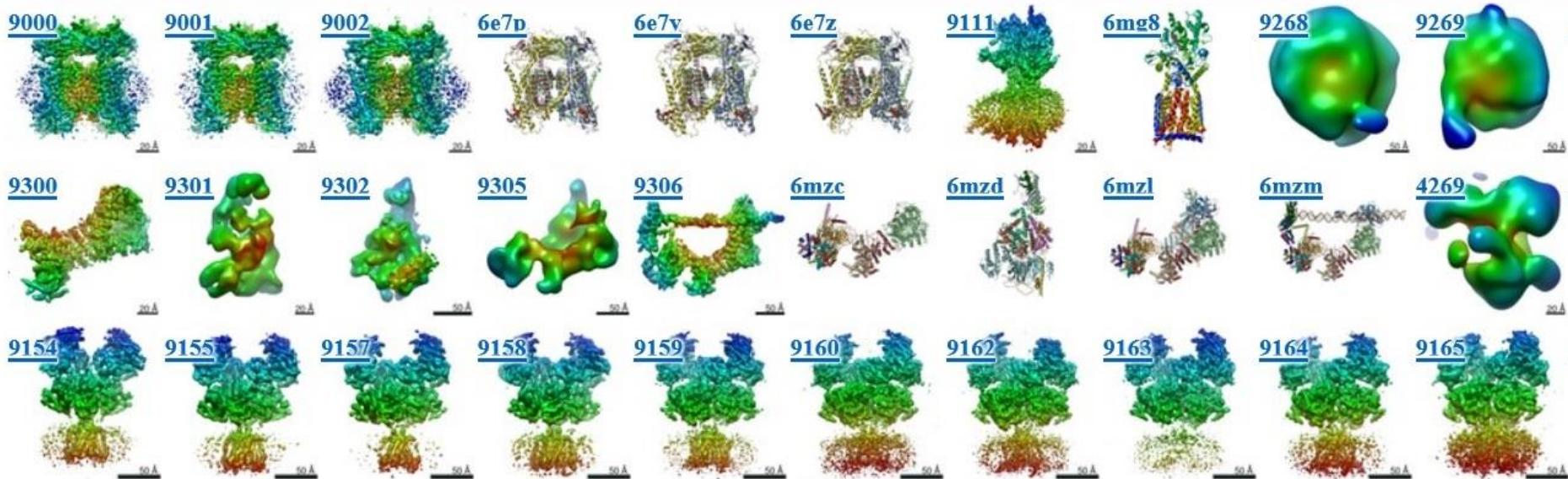
Browse 3DEM data

- Keywords / EMDB-ID / PDB-ID **Submit**
- [Advanced search, table view, etc.](#)
- [Omokage search - shape similarity search](#)
- [Yorodumi - integration of EMDB/PDB metadata & structure viewer](#)
- [Gallery](#) [Statistics](#) [3DEM papers](#) [Species](#)

Recently released data

Nov 28, 2018 (EMDB: 43, PDB: 1)

The latest released data are shown in the top page.



Omokage Search

Fast shape similarity search against EMDB and PDB

★ Omokage search - Shape similarity search of macromolecules -

- Search query

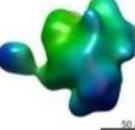
Query structure data: [Registered data structure in databanks](#) [Upload your original/modified data](#)

ID of EMDB, PDB or SASBDB: [Search](#) ?

Database: [EMDB / ID: 2190](#) [▶ Movie](#) [Structure viewer](#)

human RNA polymerase II in complex with AluRA RNA [Search](#)

Samples: [Recommended](#) [EMDB](#) [PDB](#) [SASBDB](#)



50 Å

EMDB ID / PDB ID can be used for the query shape.

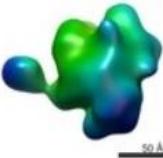
Omokage Search

Fast shape similarity search against EMDB and PDB

★ **Omokage search** - Shape similarity search of macromolecules -

+ **Search query**
- **Subject structure**

Database: EMDB / ID: 2190 Movie Structure viewer
human RNA polymerase II in complex with AluRA RNA

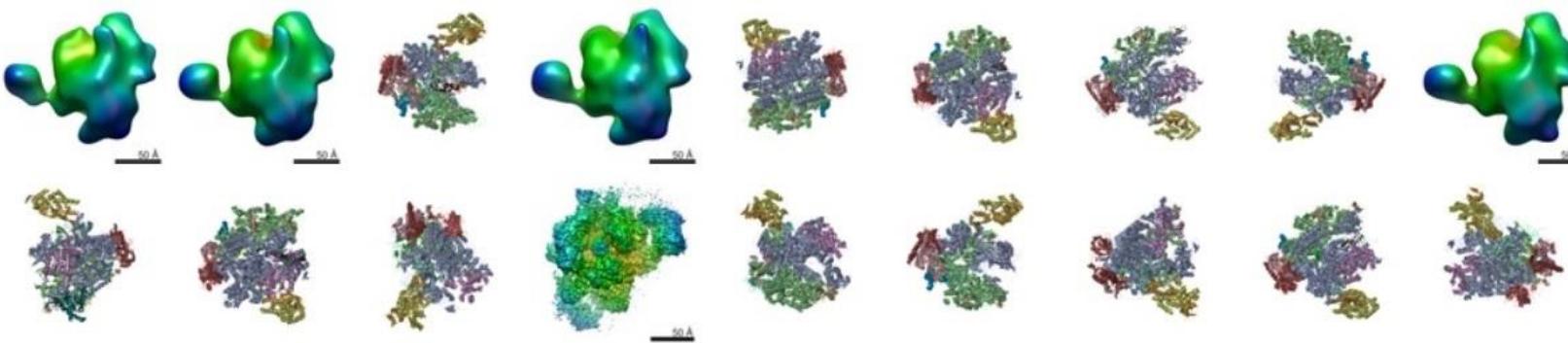


- **Search result**

Within a minute, similar-shape maps and atomic models are shown.

Summary Display Filter Refine

Showing 1 - 100 of 2,000 items found from all 280,812 structures

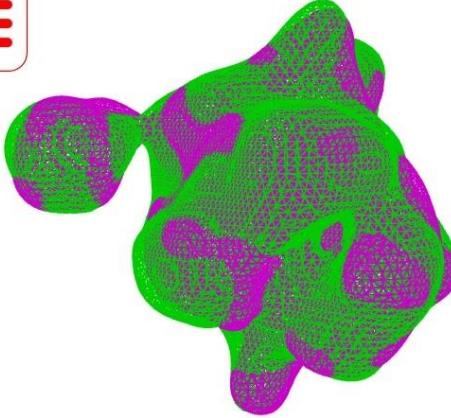


Map vs Map

Map vs Atomic model

Pairwise fitting of target [emdb_2191](#) on reference [emdb_2190](#) by [gmfit](#)(PID:18906).

RANK[1] Corr.Coeff:0.972 [JSmol] [Jmol_S] [Molmil]



Fitted by **gmfit**

Visualized by **MolMil**
WebGL-based viewer

TARGET(emdb_2191) human RNA polymerase II in complex with B2 RNA [Ngauss:20]
[Download superimposed target map \(CCP4\) \(emdb_2191\)](#)

REFERENCE(emdb_2190) human RNA polymerase II in complex with AluRA RNA [Ngauss:20]
[Download the reference map](#)

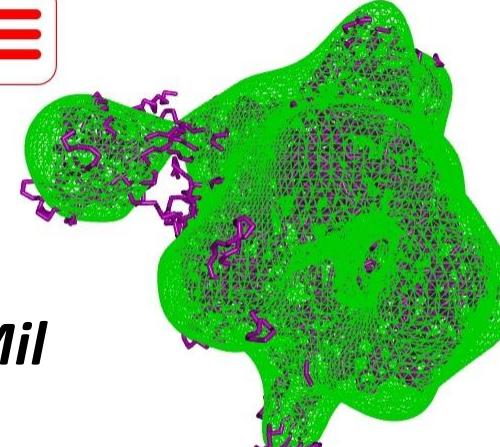
rank	1	2	3	4	5	6	7	8	9	10
Corr.Coeff.	0.972	0.972	0.750	0.712	0.704	0.704	0.702	0.700	0.689	0.666

Summary Display Filter Refine

Showing 1 - 100 of 2,000 items found from all 280,812 structures

Pairwise fitting of target [pdb_5c44](#) on reference [emdb_2190](#) by [gmfit](#)(PID:18951).

RANK[1] Corr.Coeff:0.896 [JSmol] [Jmol_S] [Molmil]



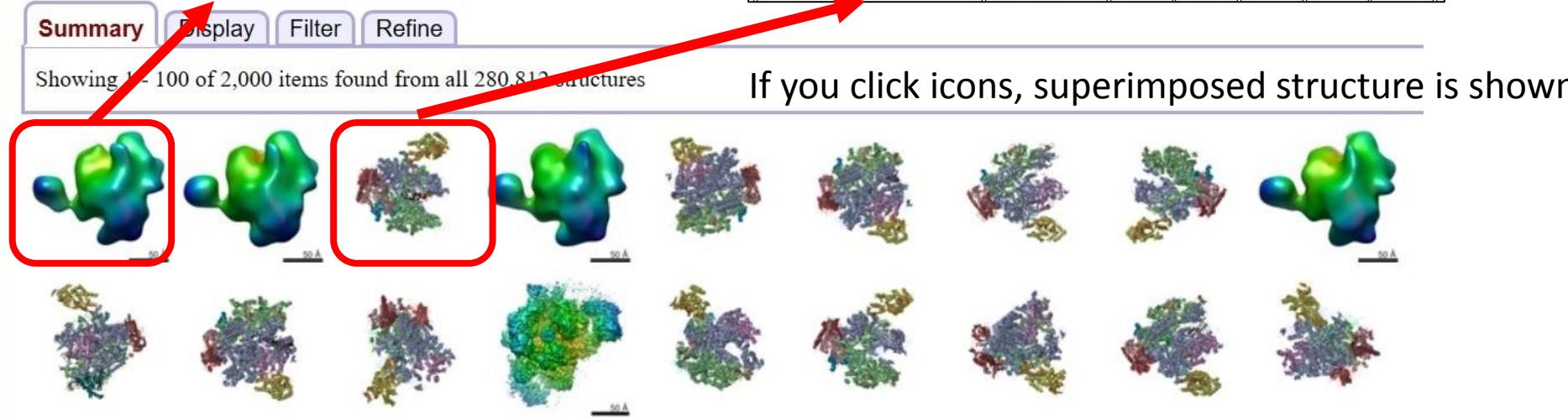
Fitted by **gmfit**

Visualized by **MolMil**
WebGL-based viewer

TARGET(pdb_5c44) CRYSTAL STRUCTURE OF A TRANSCRIBING RNA POLYMERASE
[Download superimposed target atoms \(PDB\)\(pdb_5c44\)](#)

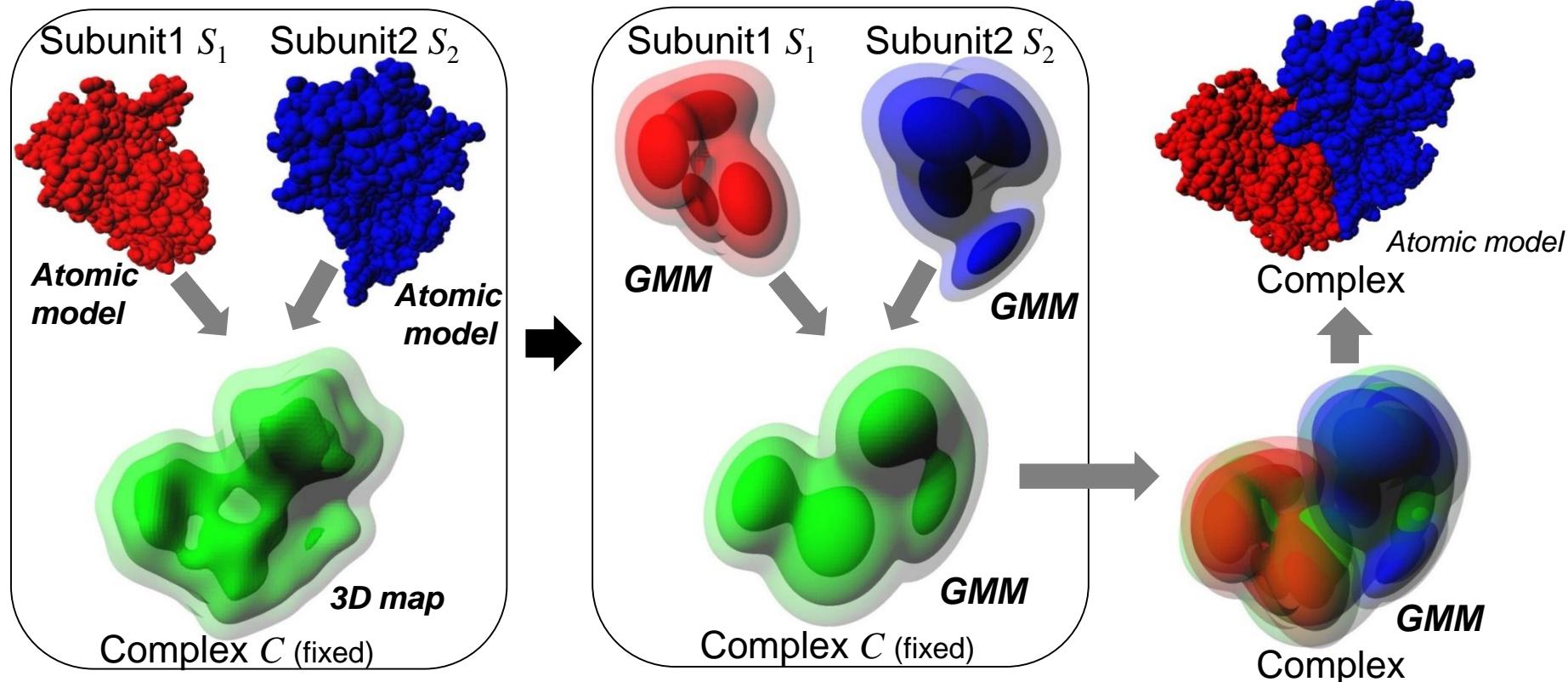
REFERENCE(emdb_2190) human RNA polymerase II in complex with AluRA RNA [Ngauss:20]
[Download the reference map](#)

rank	1	2	3	4	5	6
Corr.Coeff.	0.896	0.895	0.714	0.707	0.688	0.666



Gmfit : rigid-body fitting using GMM

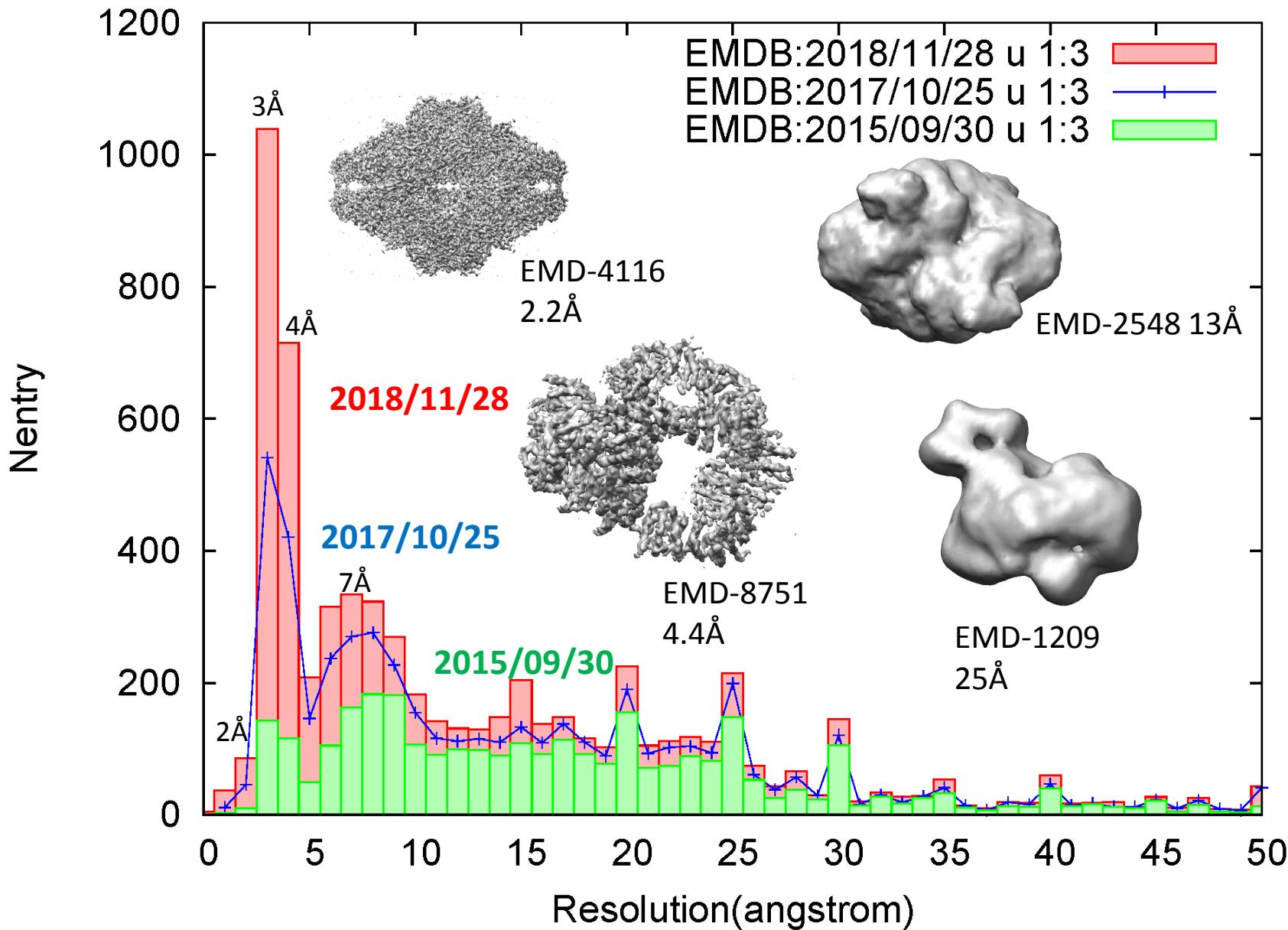
GMM : Gaussian mixture model



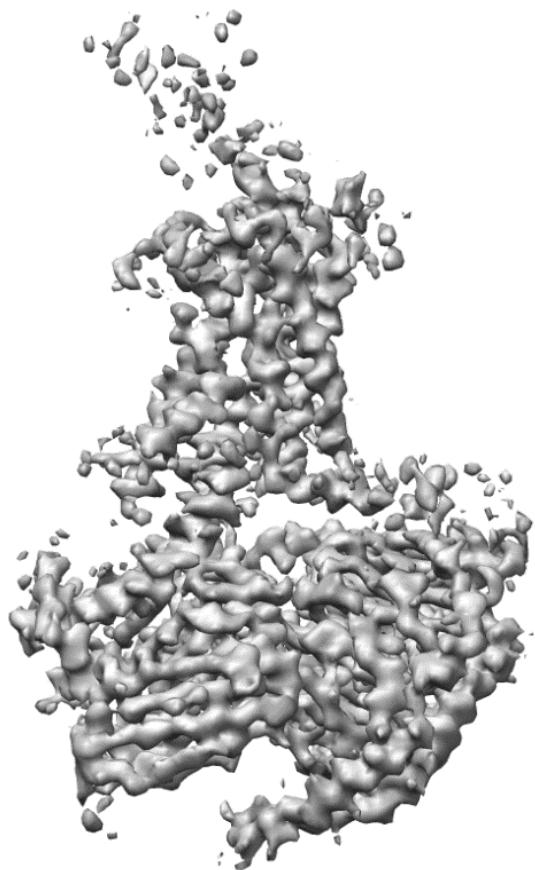
Subunit1: CyclinA (1finB) , Subunit2:Cyclin-dependent kinase(1finA)

Kawabata (2008), Biophys. J.

Resolution Revolution of EMDB



Resolution revolution of GMM

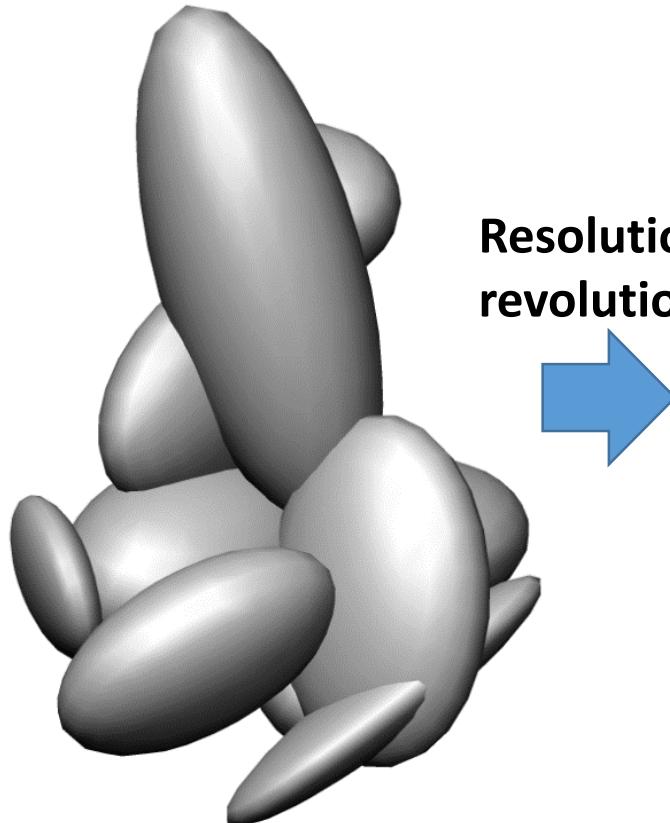


3D density map

200^3 voxels

Resolution: 4.1 Å

EMD-8623



GMM

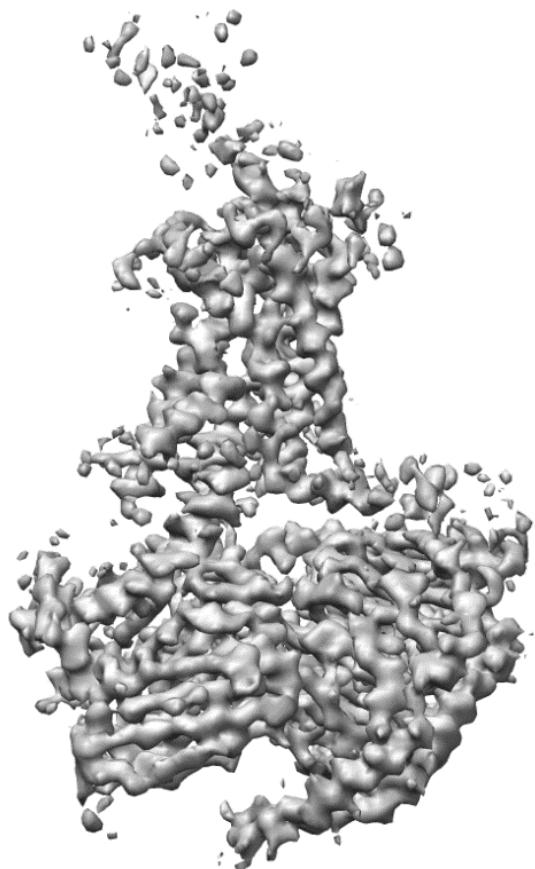
Ngauss=10

EM algorithm

Kawabata (2008), Biophys. J.

Resolution
revolution !

Resolution revolution of GMM

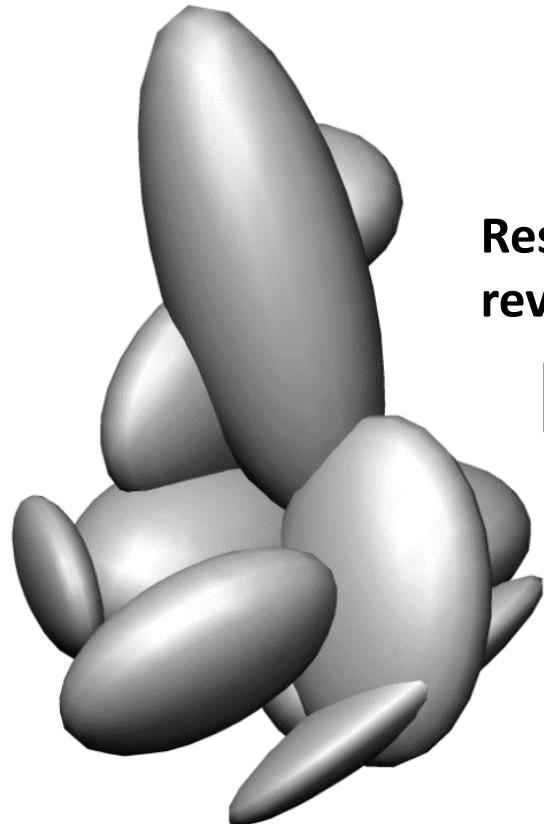


3D density map

200^3 voxels

Resolution: 4.1\AA

EMD-8623



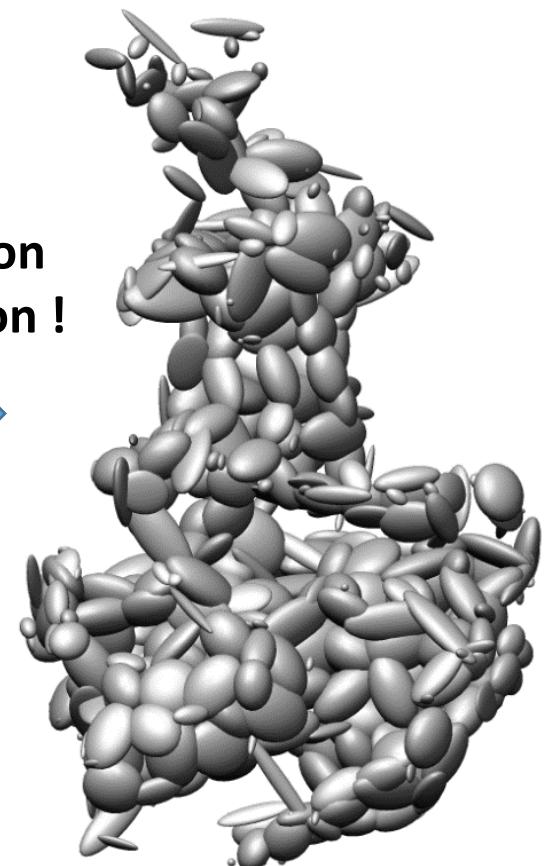
GMM

Ngauss=10

EM algorithm

Kawabata (2008), Biophys. J.

Resolution
revolution !



GMM

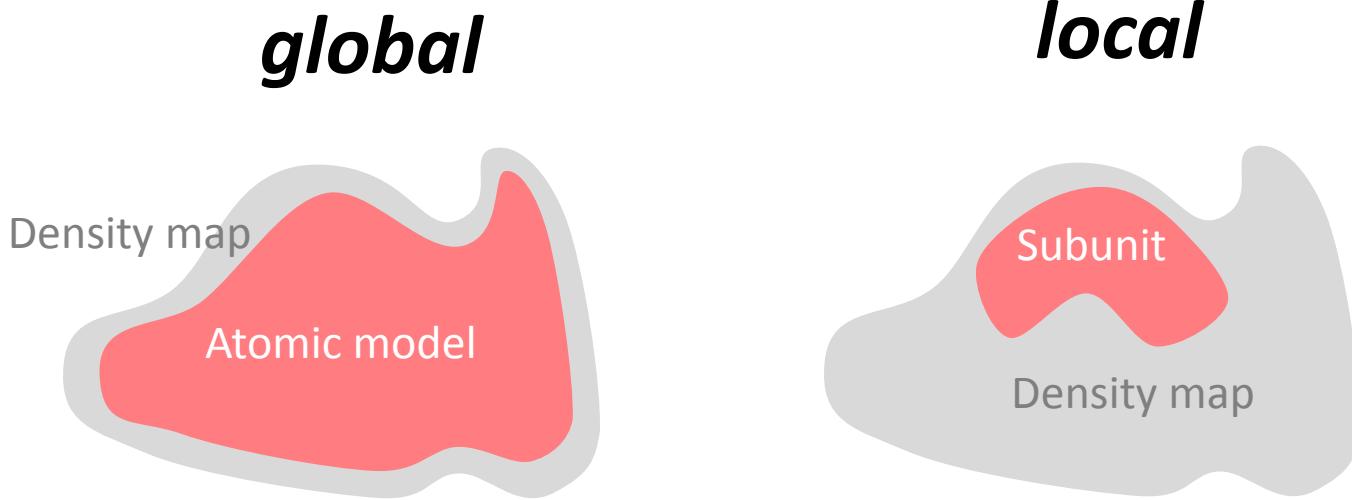
Ngauss=546

Down-sampling Gaussians

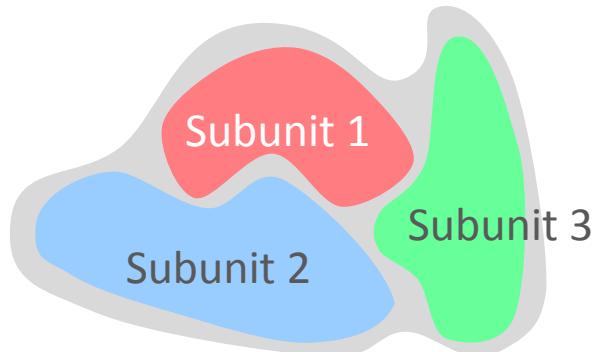
Kawabata (2018), J. Str. Biol.

Types of rigid fitting problem

single
subunit

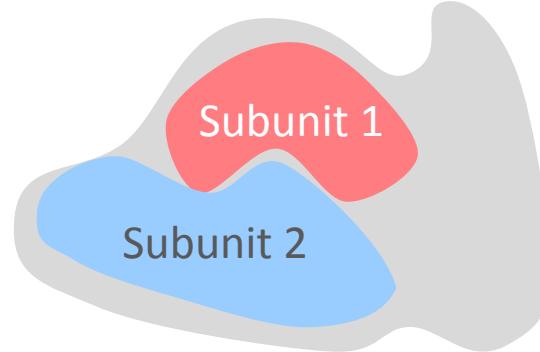
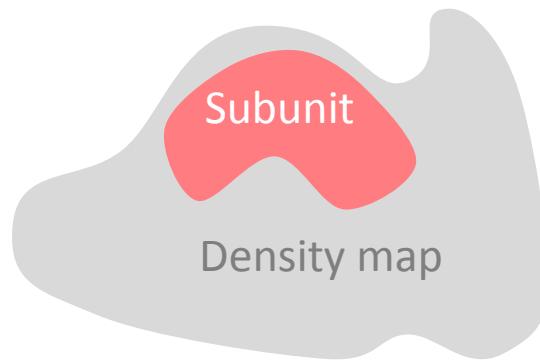


multiple
subunits

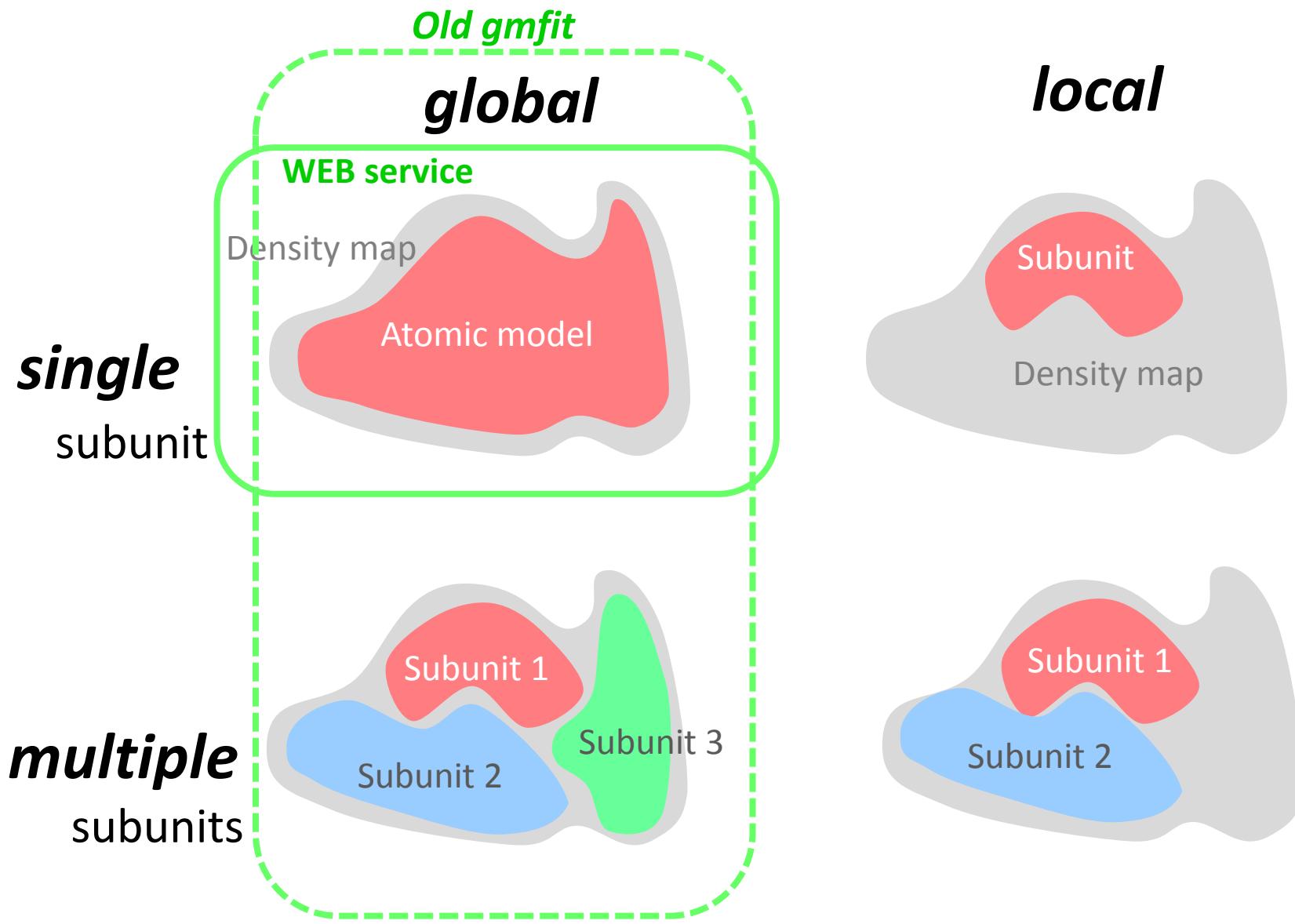


global

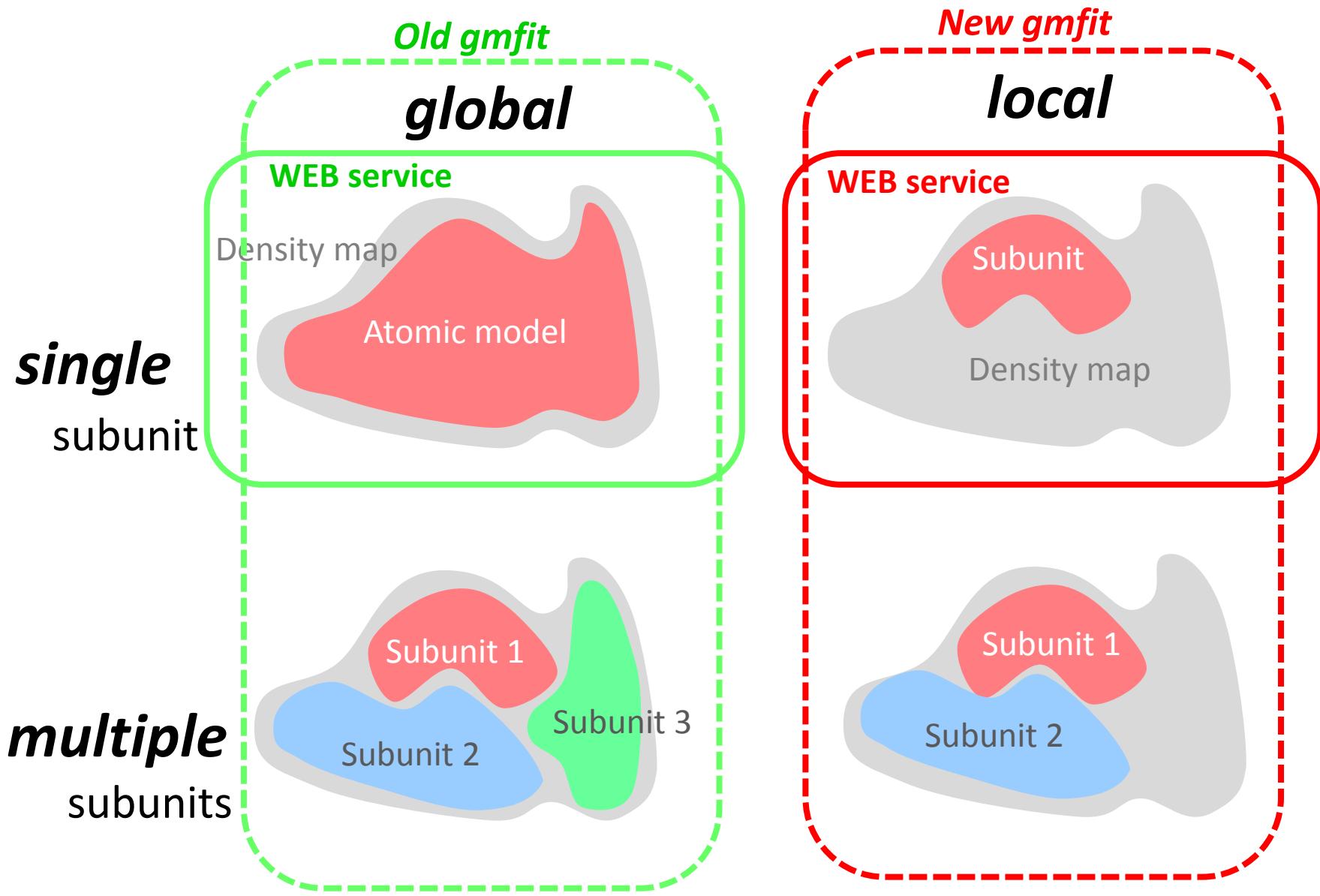
local



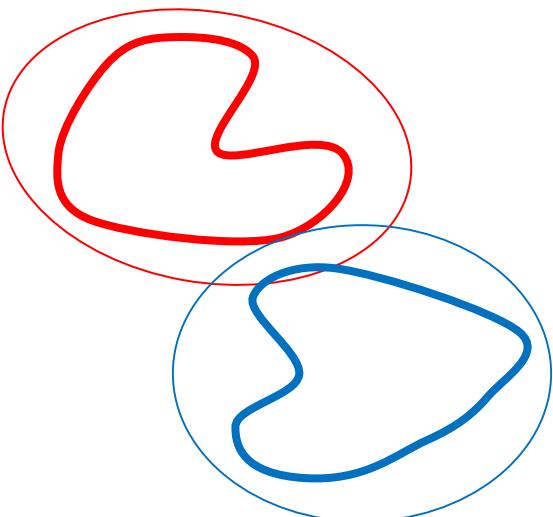
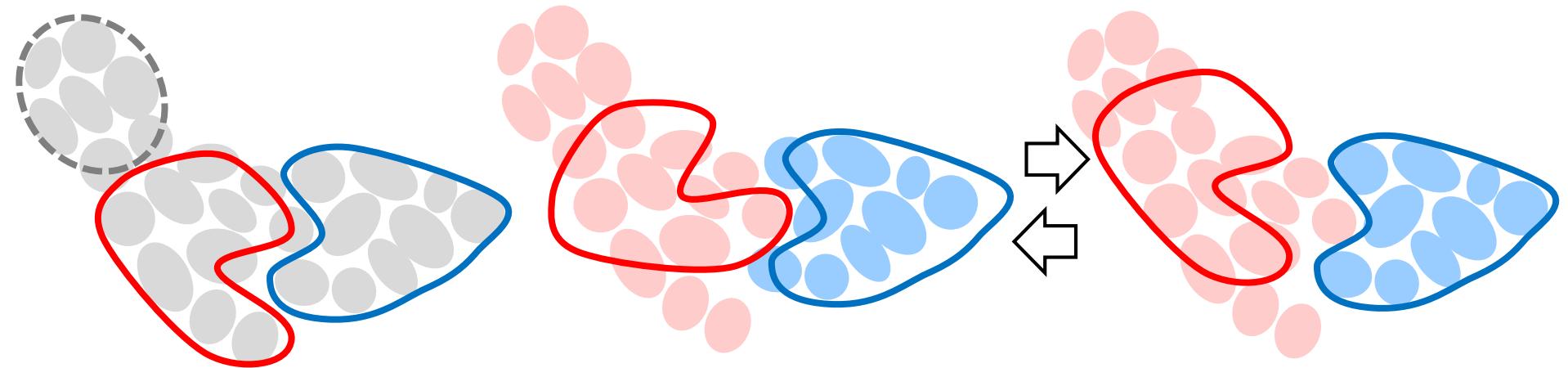
Types of rigid fitting problem



Types of rigid fitting problem

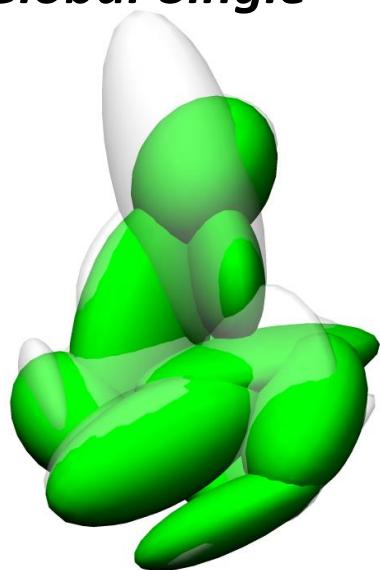


Masked Segmentation & Fitting (MaskSegFit)

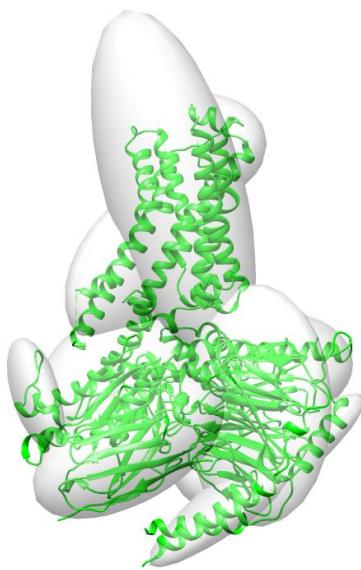


Example of rigid fittings

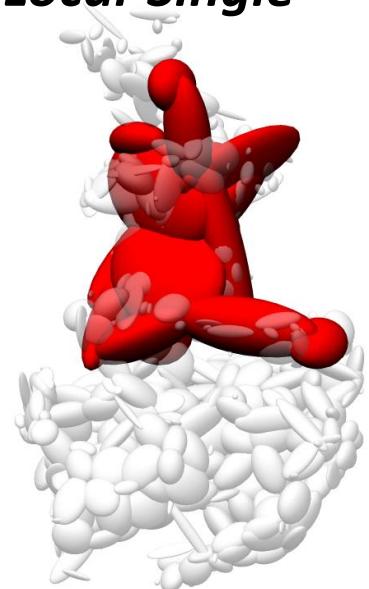
Global-Single



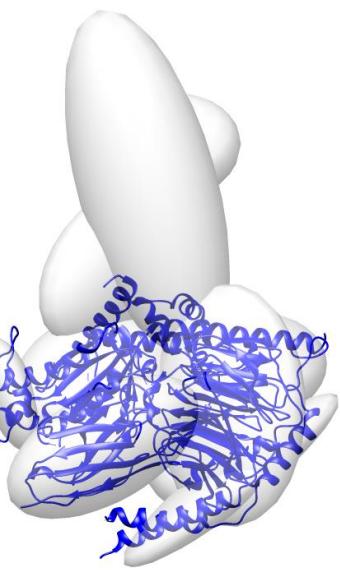
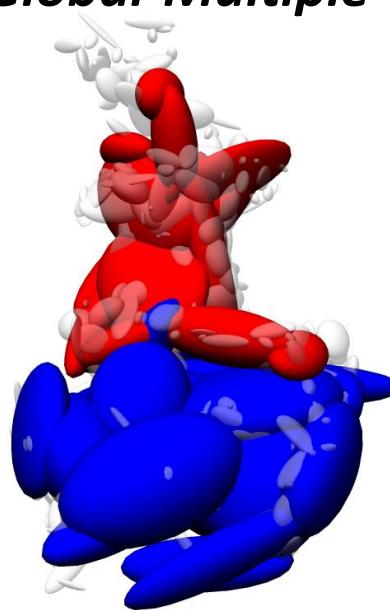
Local-Single



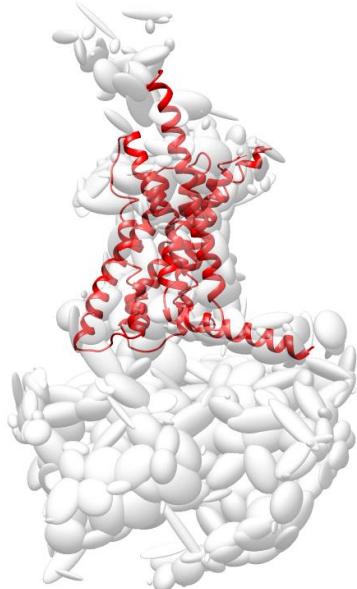
Local-Single



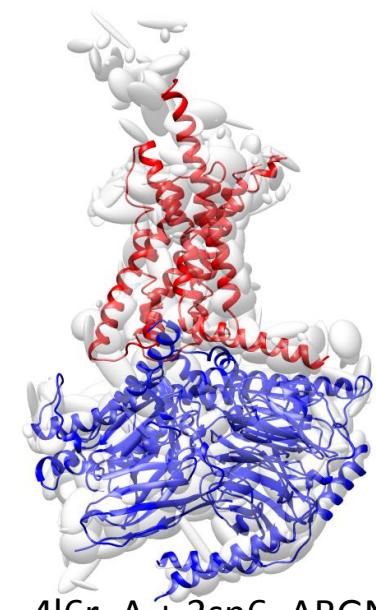
Global-Multiple



3sn6_ABGN

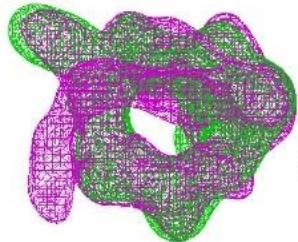


4l6r_A



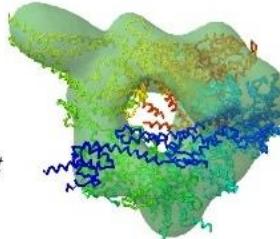
4l6r_A + 3sn6_ABGN

Gmfit WEB : 1-to-1 global fitting

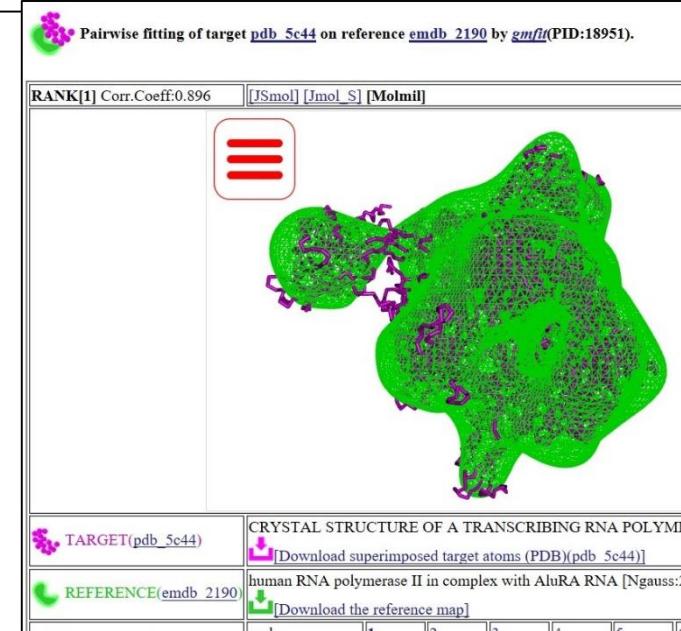
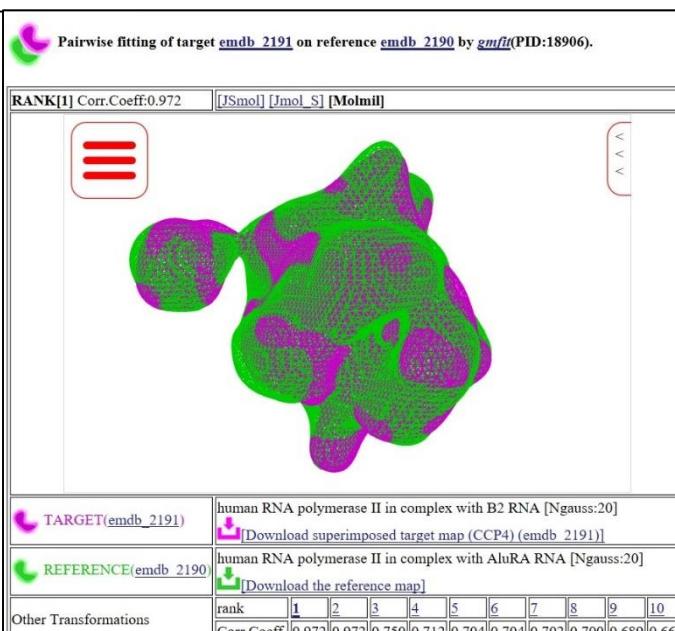


Pairwise gmfit

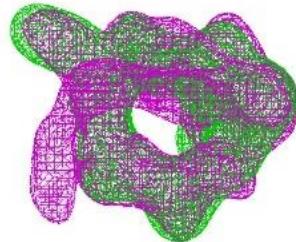
Superimpose two density maps or atomic models by *gmfit*



	DENSITY MAP			ATOMIC MODEL			
	Input ID	Upload Map File		Input ID	Upload Pdb File		
	emdb id		level	pdb_id	chain_id	assembly_id	pdb file upload
TARGET (be superimposed)	 <input type="text" value=""/>	<input type="button" value="参照..."/>	<input type="text" value=""/>	 <input type="text" value="1f4h"/>	<input type="text" value="A"/>	<input type="text" value=""/>	<input type="button" value="参照..."/>
REFERENCE (fixed)	 <input type="text" value="5995"/>	<input type="button" value="参照..."/>	<input type="text" value="0.0224"/>	 <input type="text" value=""/>	<input type="text" value=""/>	<input type="text" value=""/>	<input type="button" value="参照..."/>

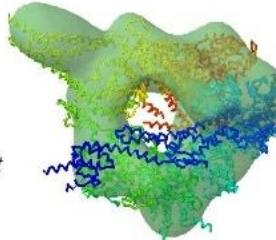


Gmfit WEB : bad for 1-to-1 local fitting..



Pairwise gmfit

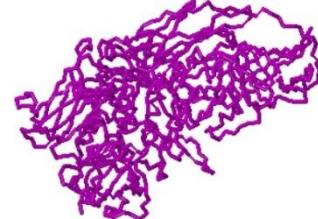
Superimpose two density maps or atomic models by *gmfit*



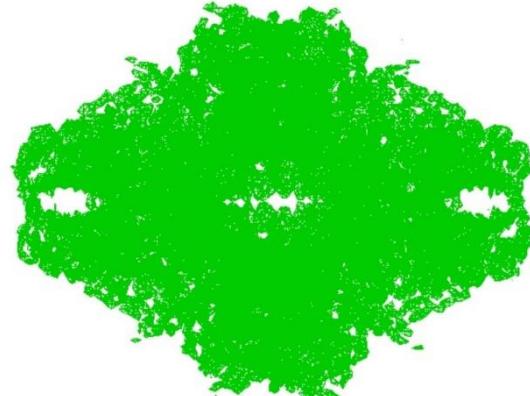
	DENSITY MAP			ATOMIC MODEL			
	Input ID	Upload Map File		Input ID	Upload Pdb File		
	emdb id		level	pdb_id	chain_id	assembly_id	pdb file upload
TARGET (be superimposed)	 <input type="text" value="1f4h"/>	<input type="button" value="参照..."/>	<input type="text"/>	 <input type="text" value="1f4h"/>	<input type="text" value="A"/>	<input type="text"/>	<input type="button" value="参照..."/>
REFERENCE (fixed)	 <input type="text" value="5995"/>	<input type="button" value="参照..."/>	0.0224	 <input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="button" value="参照..."/>

Fit one subunit on the map of homo tetramer

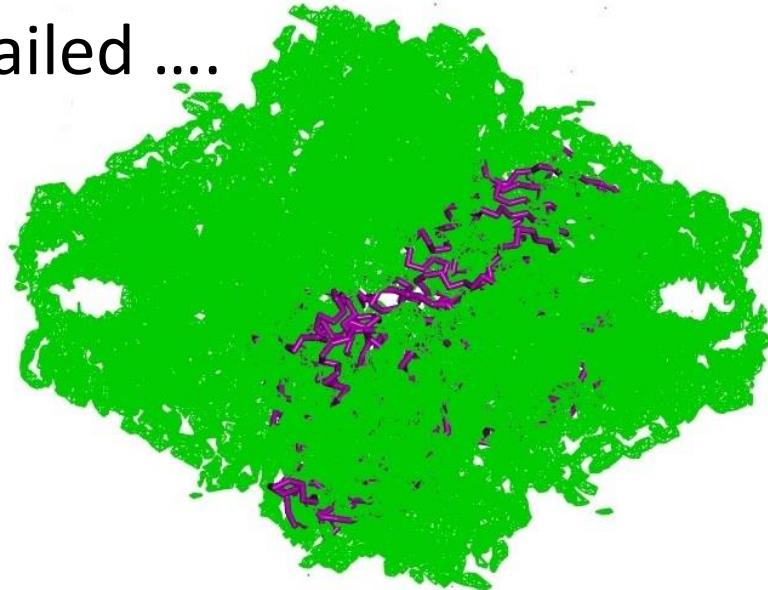
Fit



on

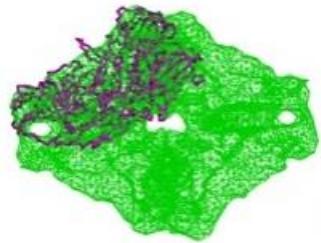


Failed



Ngauss=20

New Gmfit WEB : 1-to-1 local fitting



Local/Partial fitting : masked segfit gmfit

Superimpose an atomic model on a density map by *gmfit*
using masked segmentation fitting algorithm.



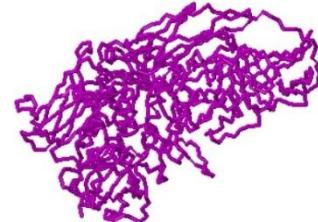
Local/partial fitting can be performed one given atomic model versus a given density map.

This new Web service will be available soon !

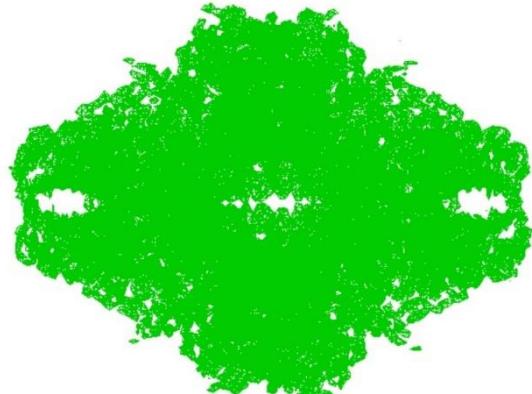
input form	map or atomic model	other options
Density map (fixed)	EMDB-ID: <input type="text" value="5995"/> or upload map file: <input type="button" value="参照..."/>	cutoff level: <input type="text" value="0.0224"/>
Atomic model (to be superimposed)	PDB_ID: <input type="text" value="1f4h"/> or upload PDB file: <input type="button" value="参照..."/>	chainID: <input type="text" value="A"/> assemblyID: <input type="text"/>

Fit one subunit on the map of homo tetramer

Fit



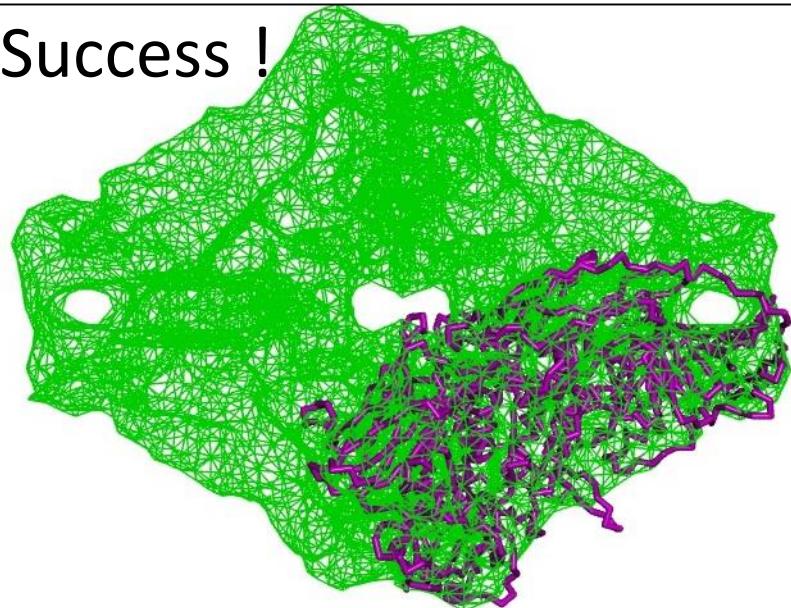
on



Ngauss=20

Ngauss=100

Success !



Acknowledgements

PDBj staff



Kawabata

gmfit

EMPIAR-PDBj



Suzuki

EM Navigator



Gert-jan

MolMil



Kurisu

Head of PDBj



Ardan Patwardhan



Andrii Iuidan