Browsing, searching, and comparing 3D electron microscopy data

Hirofumi Suzuki

BSJ-2014 luncheon seminar
2014-09-27

Starting with …

This presentation is about PDBj services for electron microscopy (EM) data.

But, many of you may not be so interested in EM.

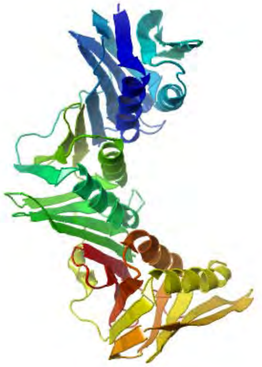
Then, I start not with EM data, but…

Function within shape

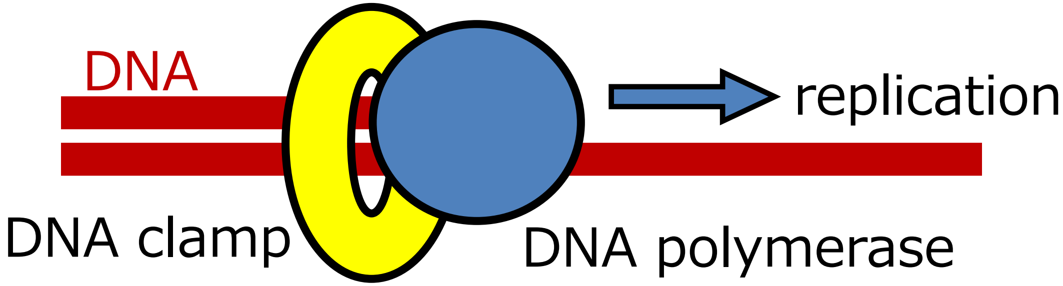
monomer

oligomer

PDB 1w60:
PCNA
from human



PDB 1vpk:
DNA polymerase beta
from bacteria



DNA

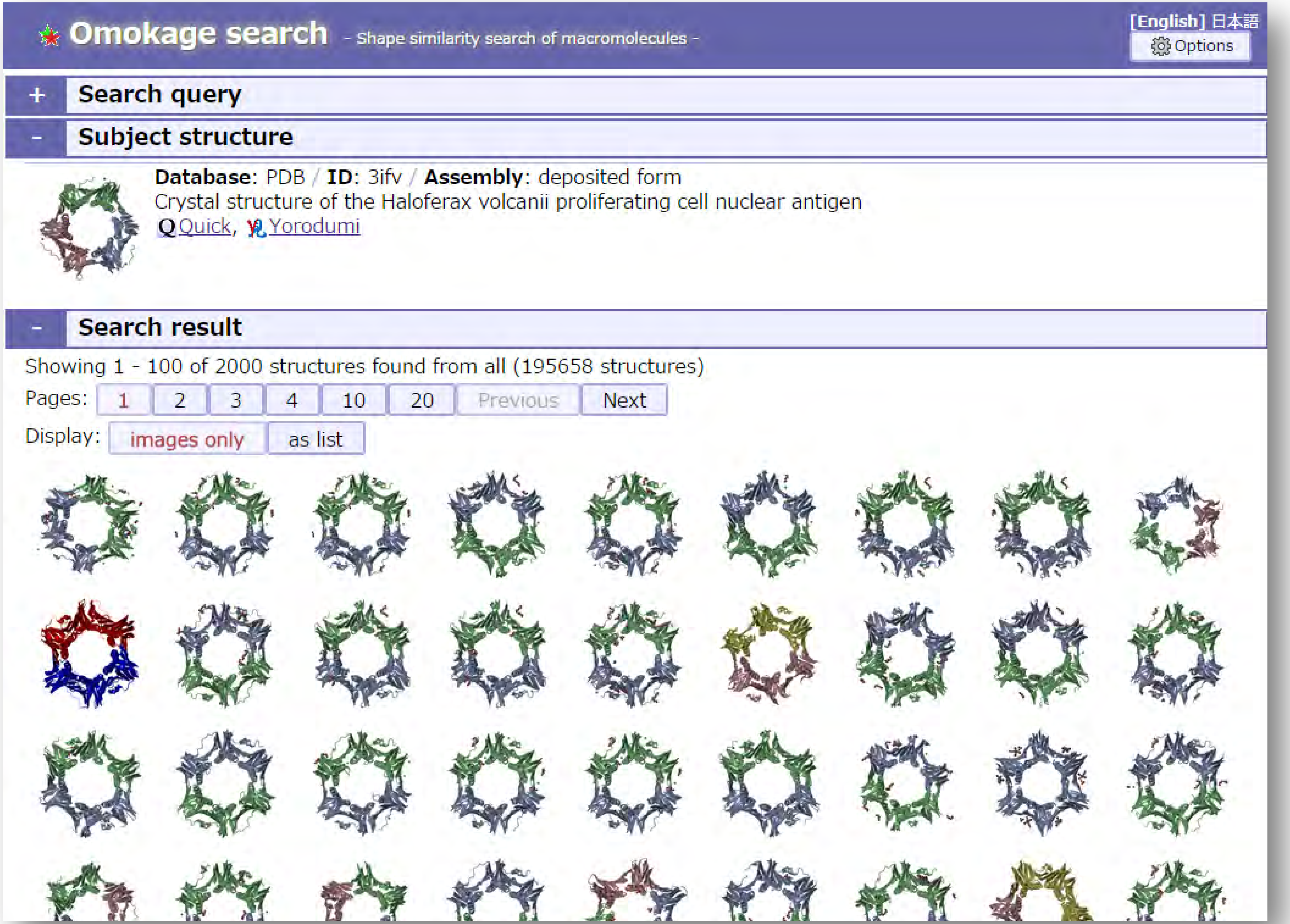
replication

DNA clamp

DNA polymerase

Shape similarity search & 3D fitting

Example #1: DNA clamp



Database: PDB / **ID**: 3ifv / **Assembly**: deposited form

Crystal structure of the Haloferax volcanii proliferating cell nuclear antigen

Quick, Yorodumi

Showing 1 - 100 of 2000 structures found from all (195658 structures)

>100 DNA clamps (2mer & 3mer) found

Omokage search

Omokage search (shape similarity search service)

Overall morphology similarity ...

URL: http://pdbj.org/emnavi/omo-search.php

Search from: ~200,000 structures

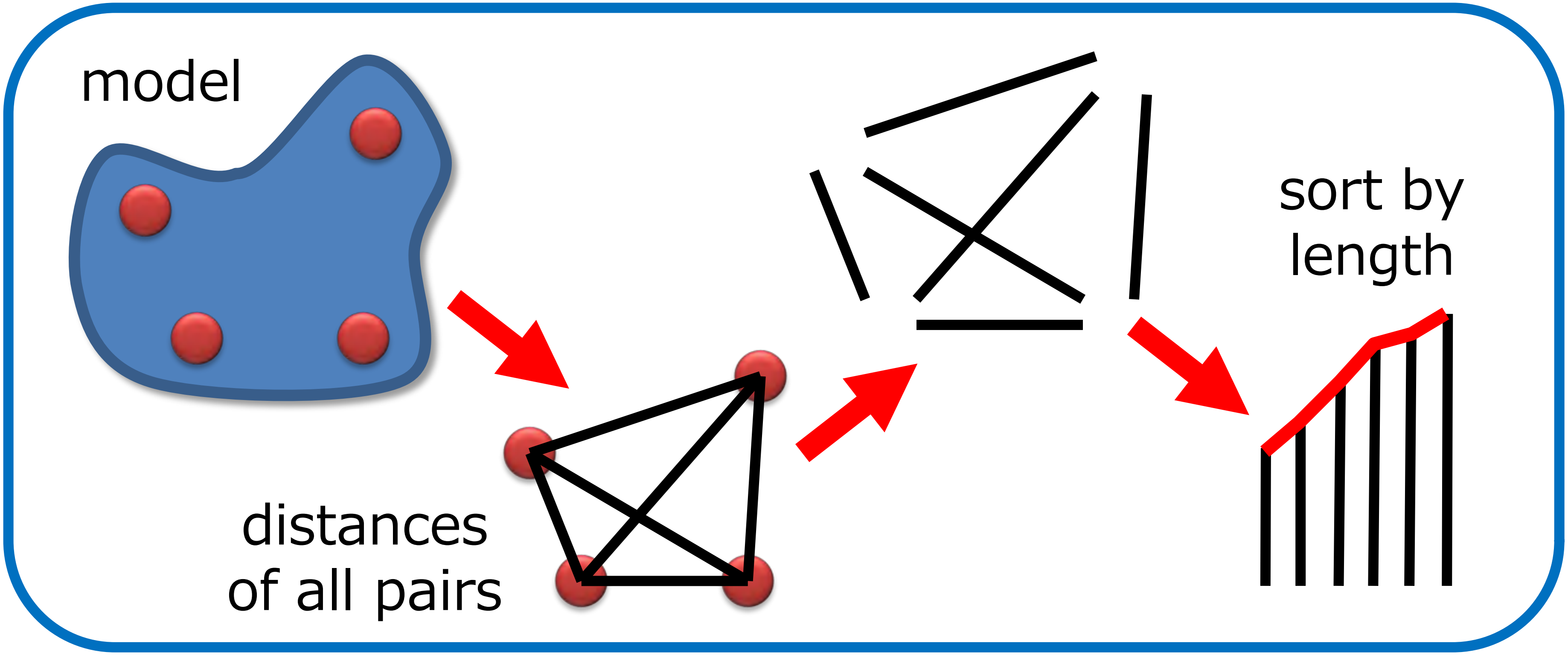
(PDB deposited units + PDB biological units
+ EMDB maps)

Search by:

- structures in EMDB/PDB
- uploaded original data
    - 3D density map (ccp4/mrc format)
    - atomic model/SAXS bead model (PDB format)

Search method: see our poster, 3P004

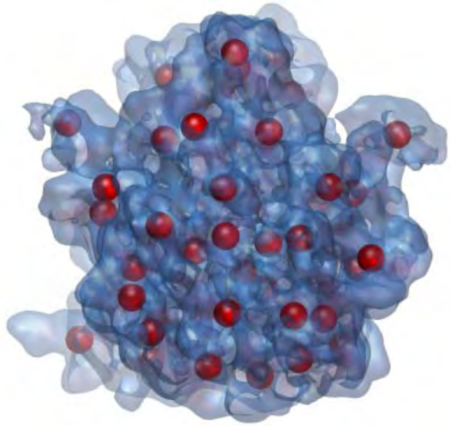
Omokage comparison



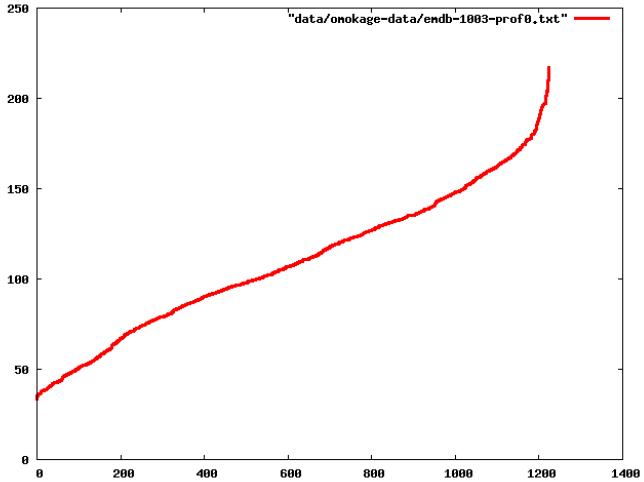
model

distances
of all pairs

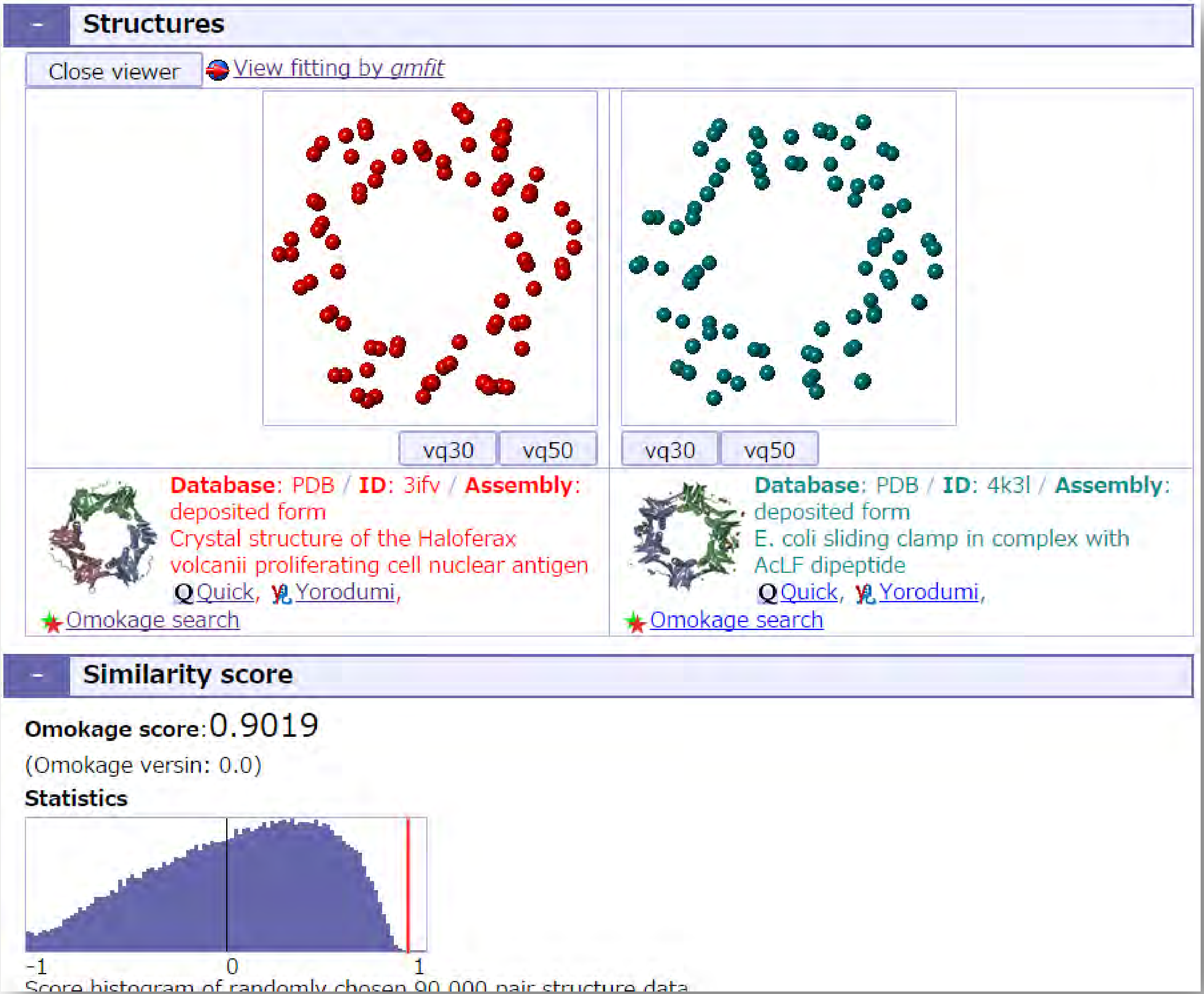
sort by
length



50 dots
1225 pairs



Seeing comparison results



Structures

Database: PDB / **ID**: 3ifv / **Assembly**: deposited form
Crystal structure of the Haloferax volcanii proliferating cell nuclear antigen
Quick, Yorodumi,
Omokage search

Database: PDB / **ID**: 4k3l / **Assembly**: deposited form
E. coli sliding clamp in complex with AcLF dipeptide
Quick, Yorodumi,
Omokage search

Similarity score

Omokage score: 0.9019

(Omokage versin: 0.0)

Statistics

Score histogram of randomly chosen 90,000 pair structure data

Shapes (by Jmol) and statistics are shown

Similar shapes ?

#118 Score: 0.8064 gmfit

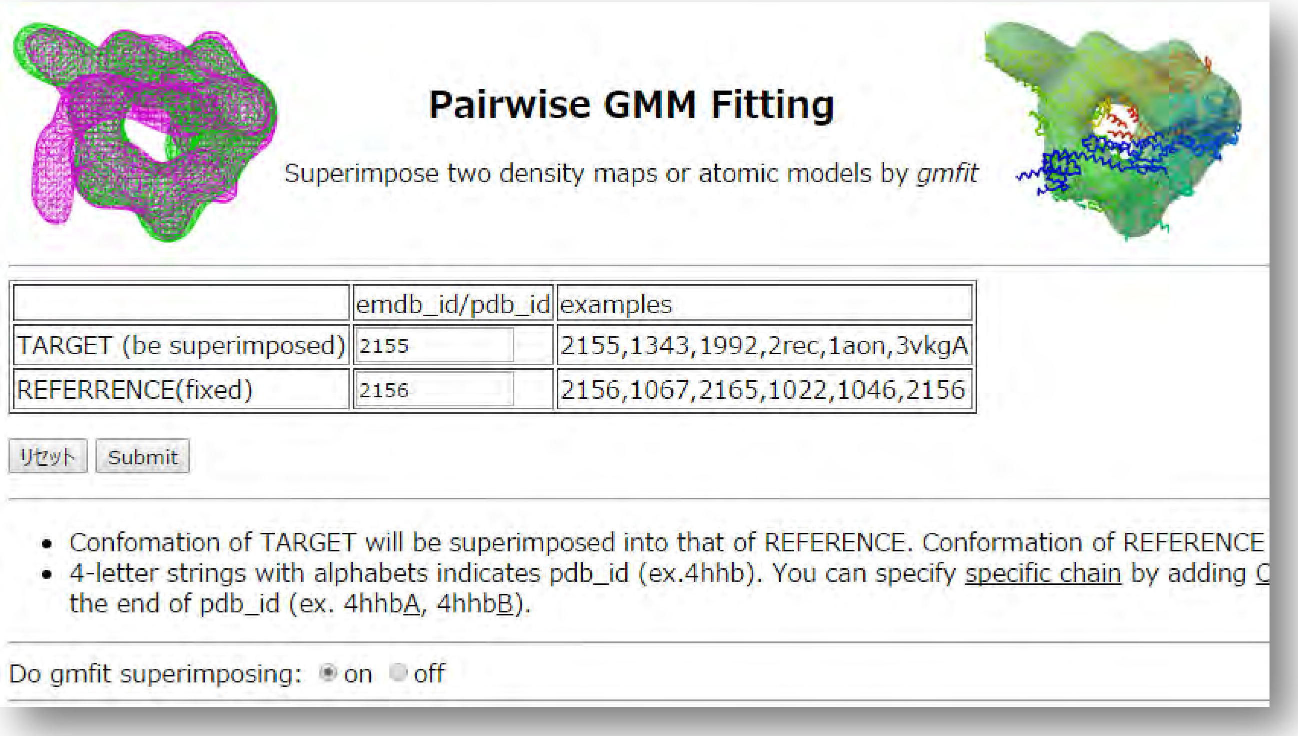
Database: PDB / **ID**: 4hkh / **Assembly**: deposited form

Structure of the Hcp1 protein from E. coli EAEC 042 pathovar, mutants N93W-S158W

Quick, Yorodumi, Omokage search

External link to *"gmfit"*

gmfit



Pairwise GMM Fitting

Superimpose two density maps or atomic models by *gmfit*

	emdb_id/pdb_id	examples
TARGET (be superimposed)	2155	2155,1343,1992,2rec,1aon,3vkgA
REFERRENCE(fixed)	2156	2156,1067,2165,1022,1046,2156

- Confomation of TARGET will be superimposed into that of REFERENCE. Conformation of REFERENCE
- 4-letter strings with alphabets indicates pdb_id (ex.4hhb). You can specify specific chain by adding the end of pdb_id (ex. 4hhbA, 4hhbB).

Do gmfit superimposing: ● on ○ off

gmfit:

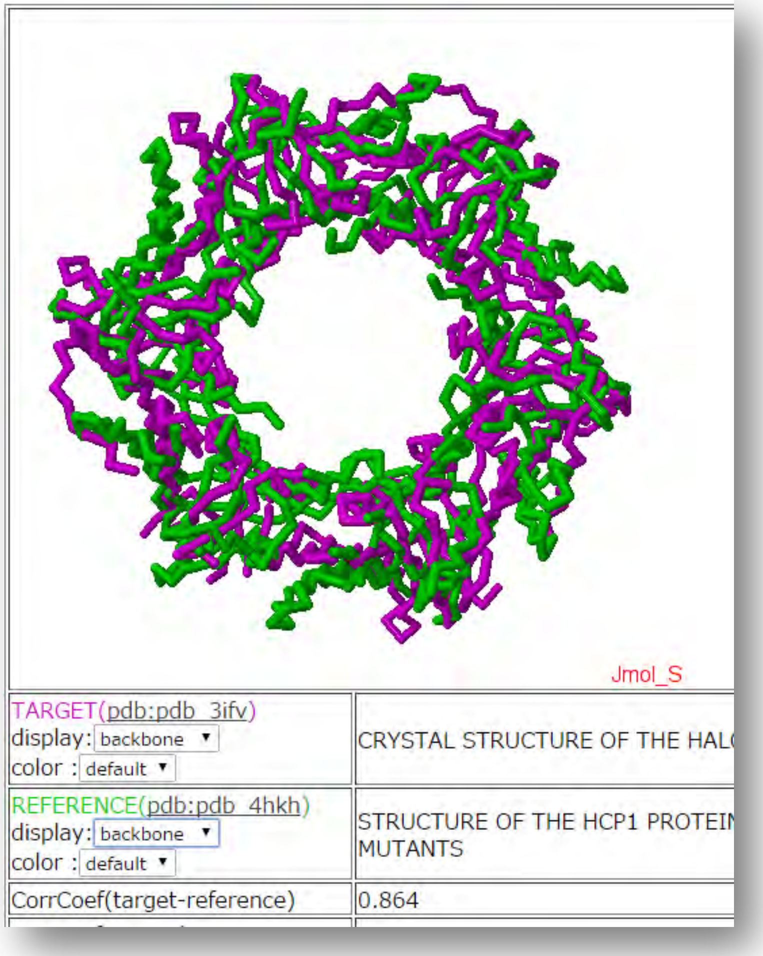
- URL: http://pdbj.org/gmfit/

- “a program for fitting subunits into density map of complex using GMM (Gaussian Mixture Model)”

- developed by Dr. Kawabata

- see our poster, 3P005

Well-fitted!

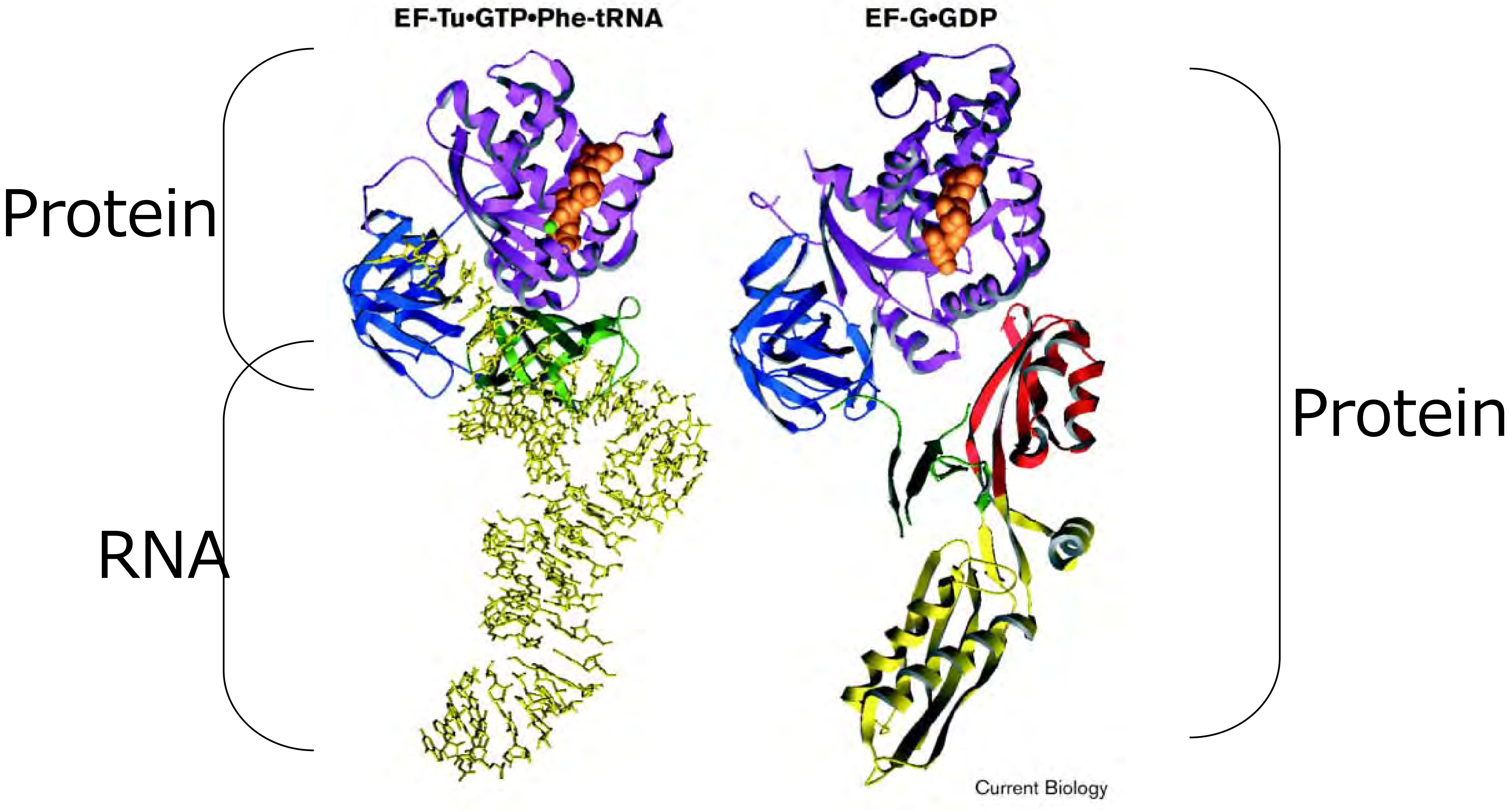


TARGET(pdb:pdb_3ifv) display: backbone ▾ color : default ▾	CRYSTAL STRUCTURE OF THE HAL
REFERENCE(pdb:pdb_4hkh) display: backbone ▾ color : default ▾	STRUCTURE OF THE HCP1 PROTEI MUTANTS
CorrCoef(target-reference)	0.864

Side view

Fitting of PDB-3ify + PDB-4hkh by *gmfit*
(Jmol: interactive 3D structure viewer)

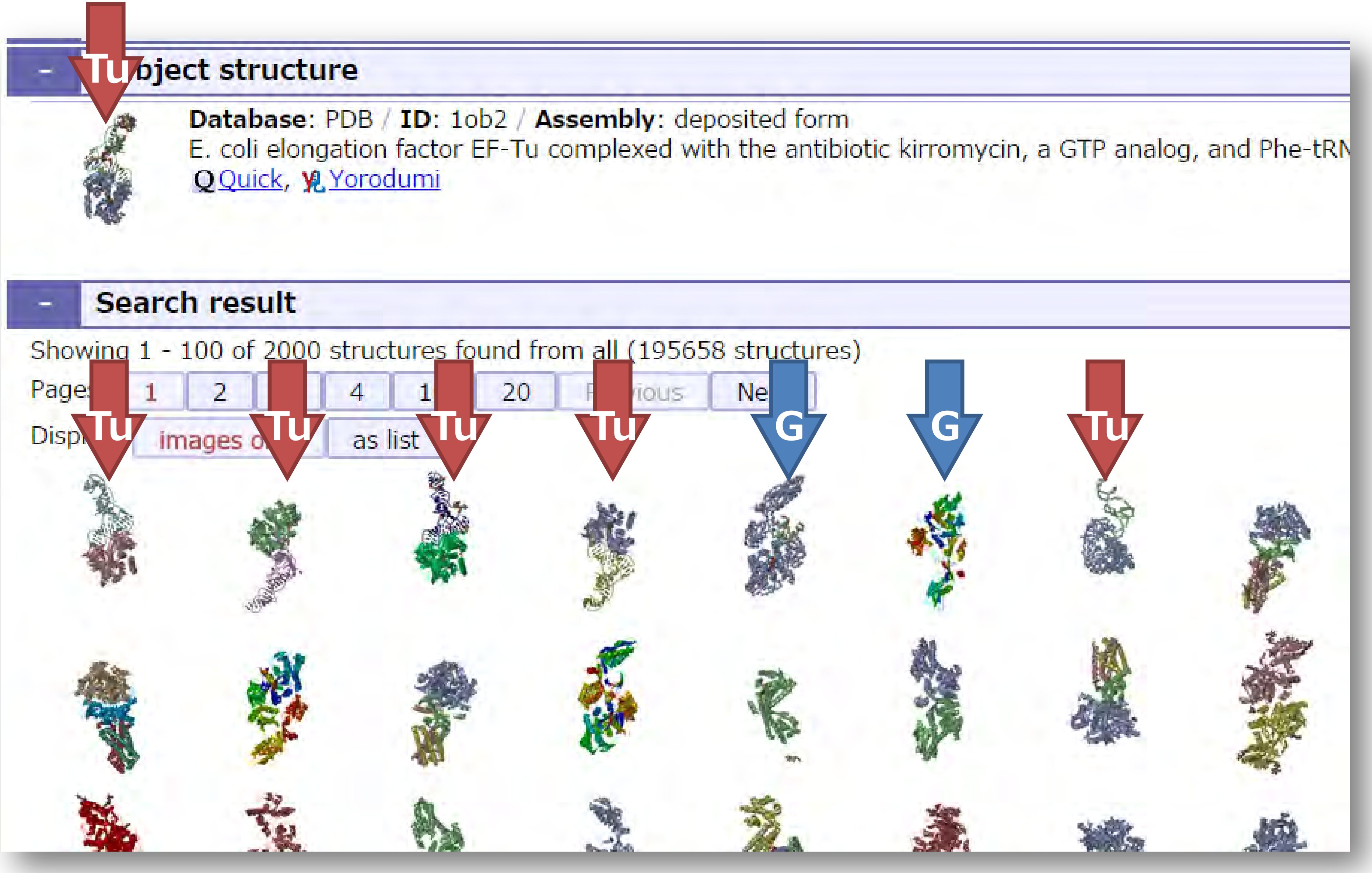
Example #2: Molecular mimicry



Green, *Curr Biol.*, 2000

Molecular mimicry
分子擬態

Example #2: Molecular mimicry



bject structure

Database: PDB / **ID**: 1ob2 / **Assembly**: deposited form

E. coli elongation factor EF-Tu complexed with the antibiotic kirromycin, a GTP analog, and Phe-tRN

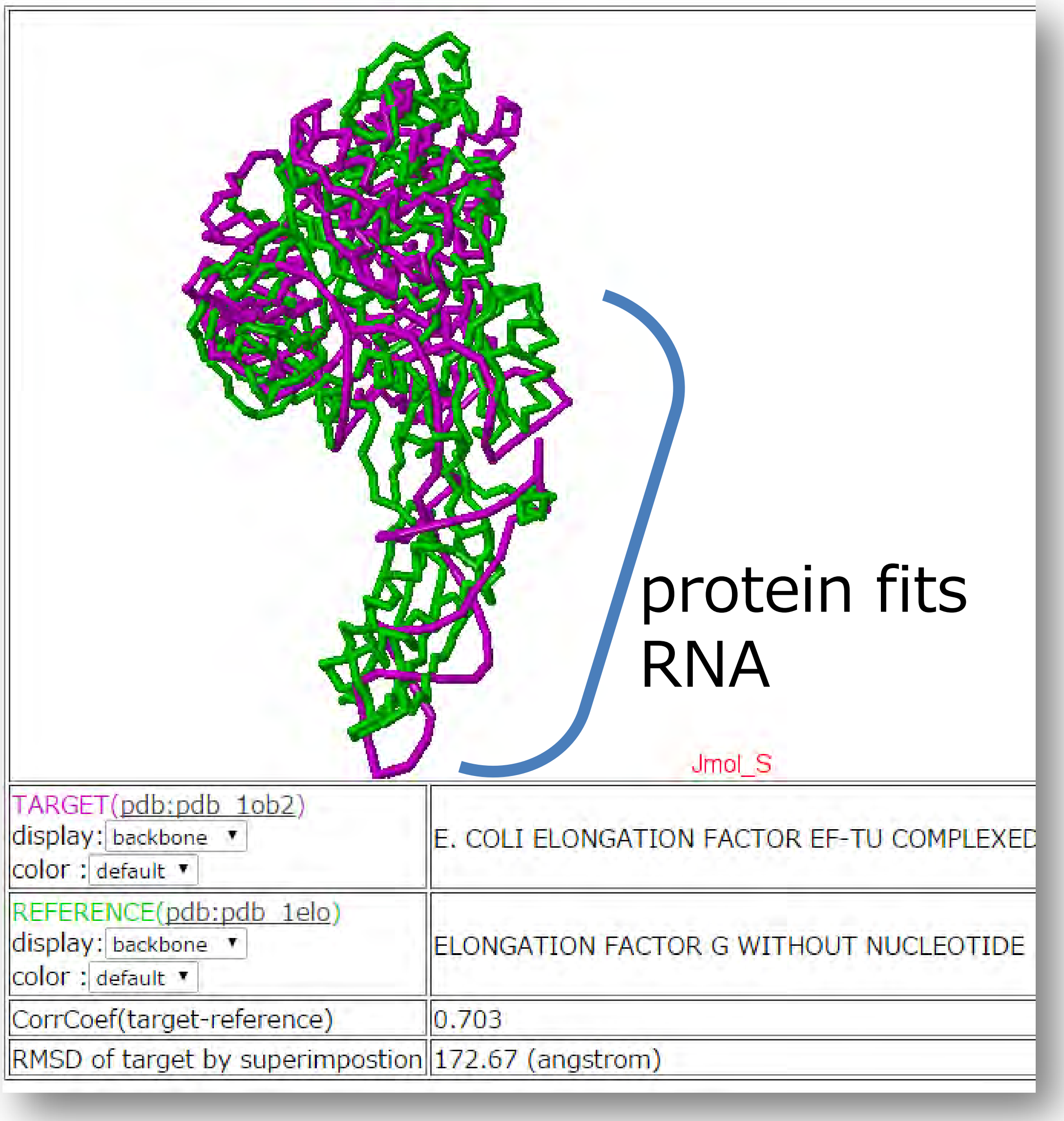
Quick, Yorodumi

Search result

Showing 1 - 100 of 2000 structures found from all (195658 structures)

EF-G structures found by search for **EF-Tu** complex

Example #2: Molecular mimicry



protein fits RNA

TARGET(pdb:pdb_1ob2) display: backbone ▼ color : default ▼	E. COLI ELONGATION FACTOR EF-TU COMPLEXED
REFERENCE(pdb:pdb_1elo) display: backbone ▼ color : default ▼	ELONGATION FACTOR G WITHOUT NUCLEOTIDE
CorrCoef(target-reference)	0.703
RMSD of target by superimpostion	172.67 (angstrom)

Fitting of PDB 1ob2 + PDB-1elo by *gmfit*

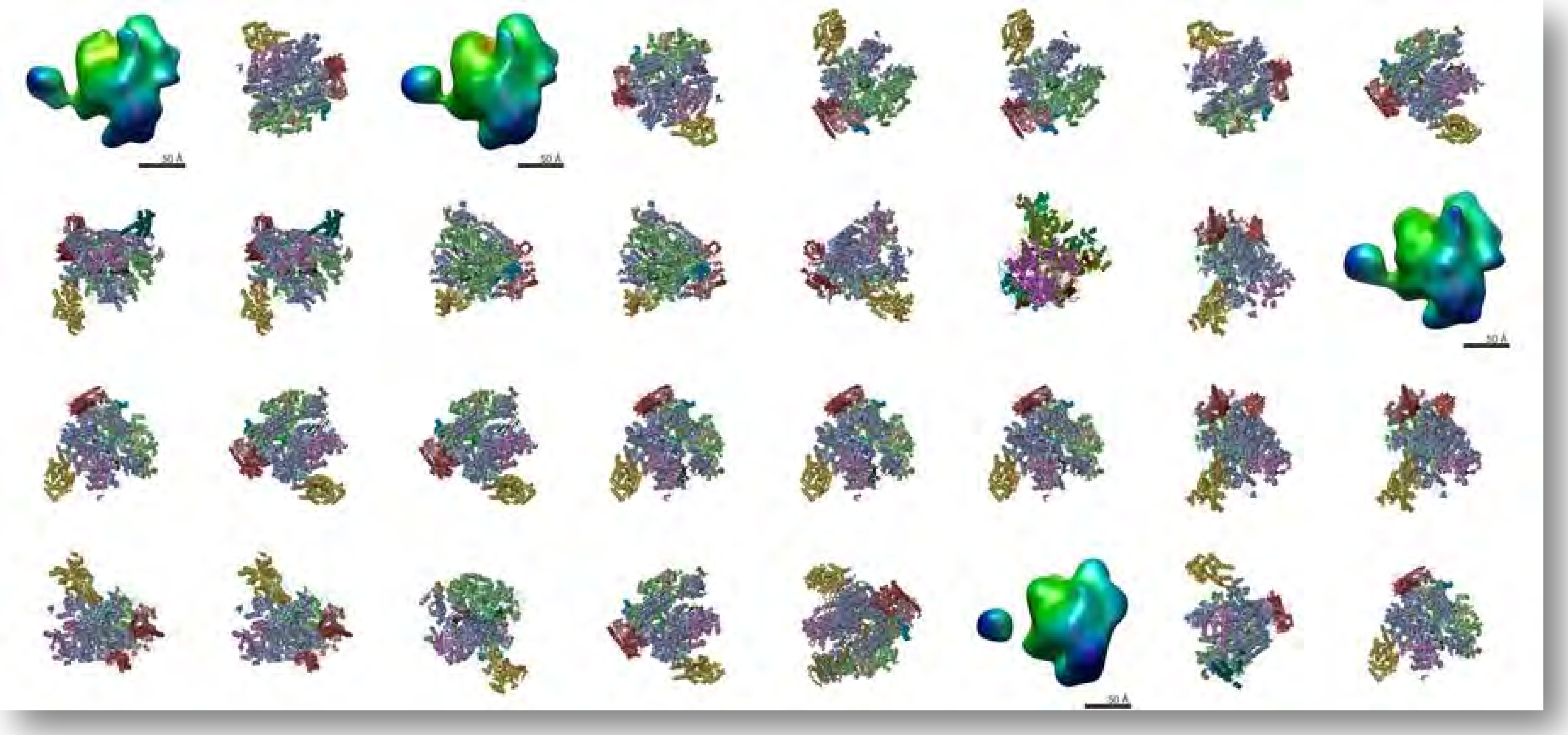
Example #3: low-resolution 3DEM data

EMDB 2189:
RNA polymerase II
from human
at 25Å resolution

Electron Microscopy Databank (EMDB):

- Databank for **3D density map data** analyzed by **EM**

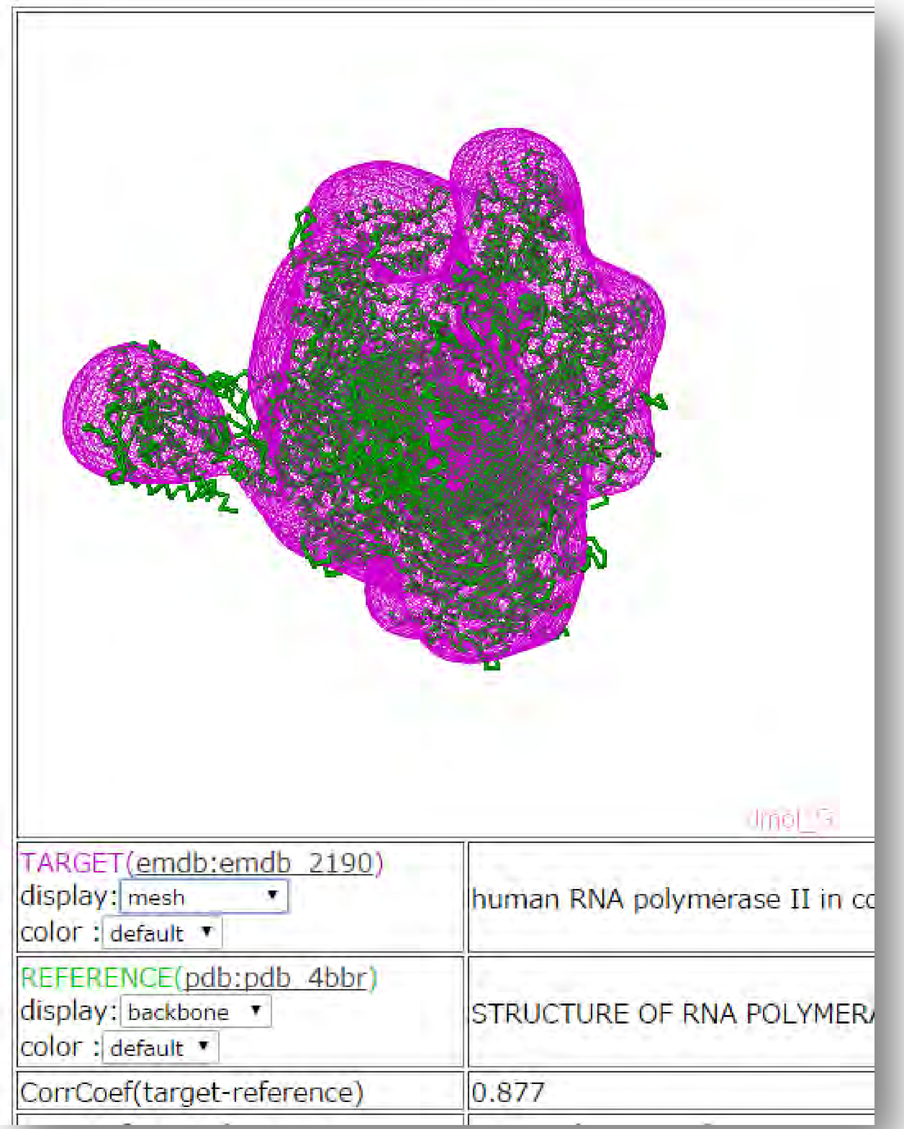
Example #3: low-resolution 3DEM data



many RNA polymerase II structures found irrespective of

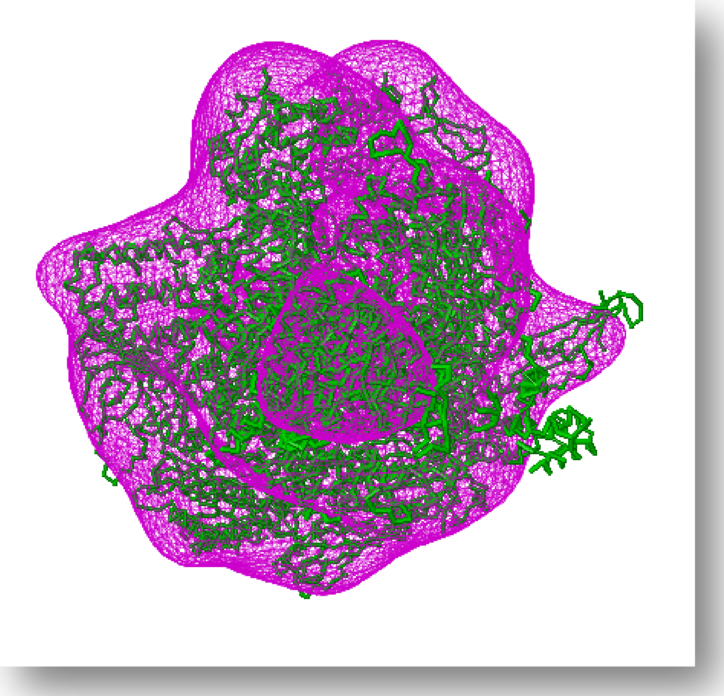
- data types (density maps / atomic models)
- resolutions

Example #3: low-resolution 3DEM data



TARGET(emdb:emdb_2190) display: mesh ▾ color : default ▾	human RNA polymerase II in co
REFERENCE(pdb:pdb_4bbr) display: backbone ▾ color : default ▾	STRUCTURE OF RNA POLYMERA
CorrCoef(target-reference)	0.877

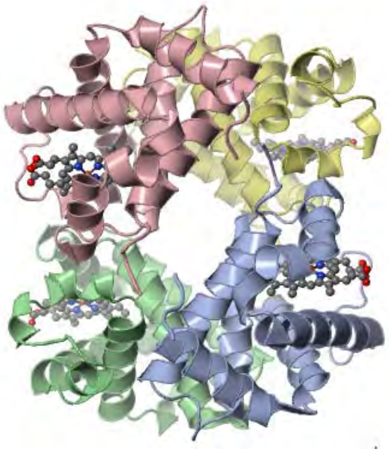
Fitting by *gmfit*

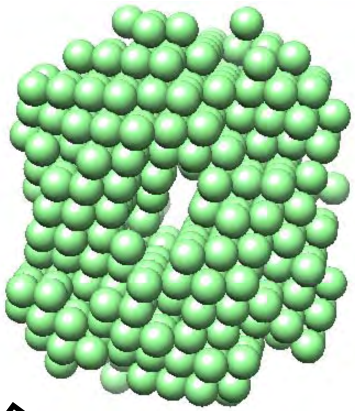


Side view

Example #4: SAXS bead model

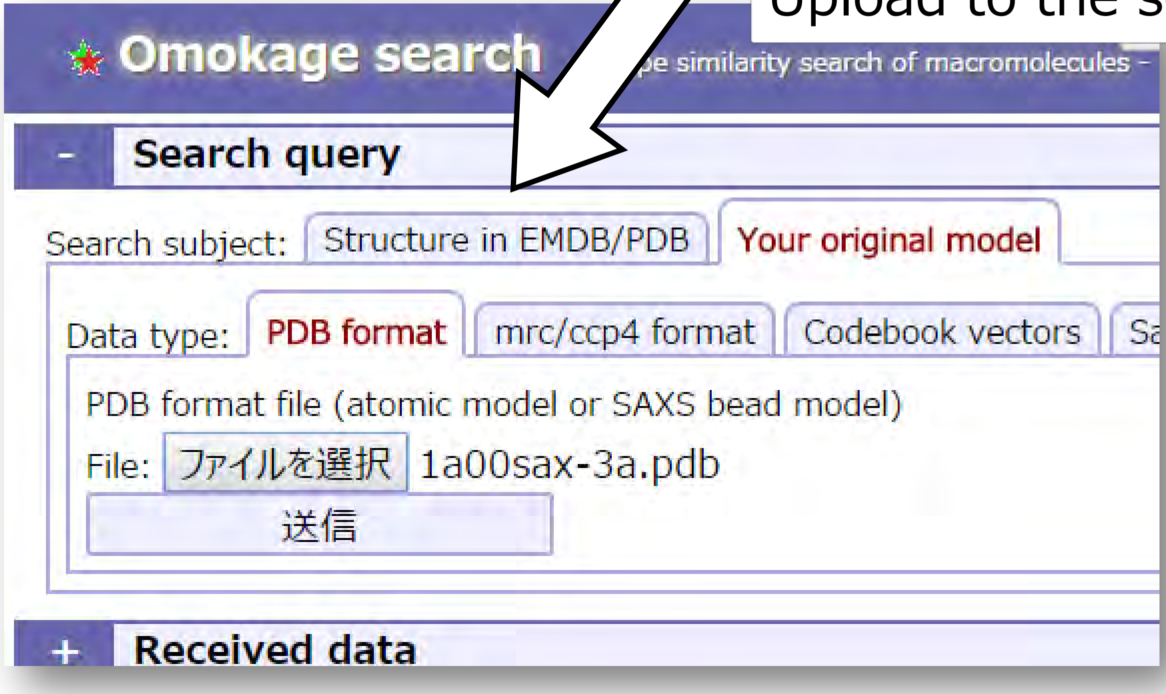
PDB-1a00
hemoglobin





Simulated SAXS
bead model

Upload to the server



Omokage search …pe similarity search of macromolecules –

Search query

Search subject: Structure in EMDB/PDB | Your original model

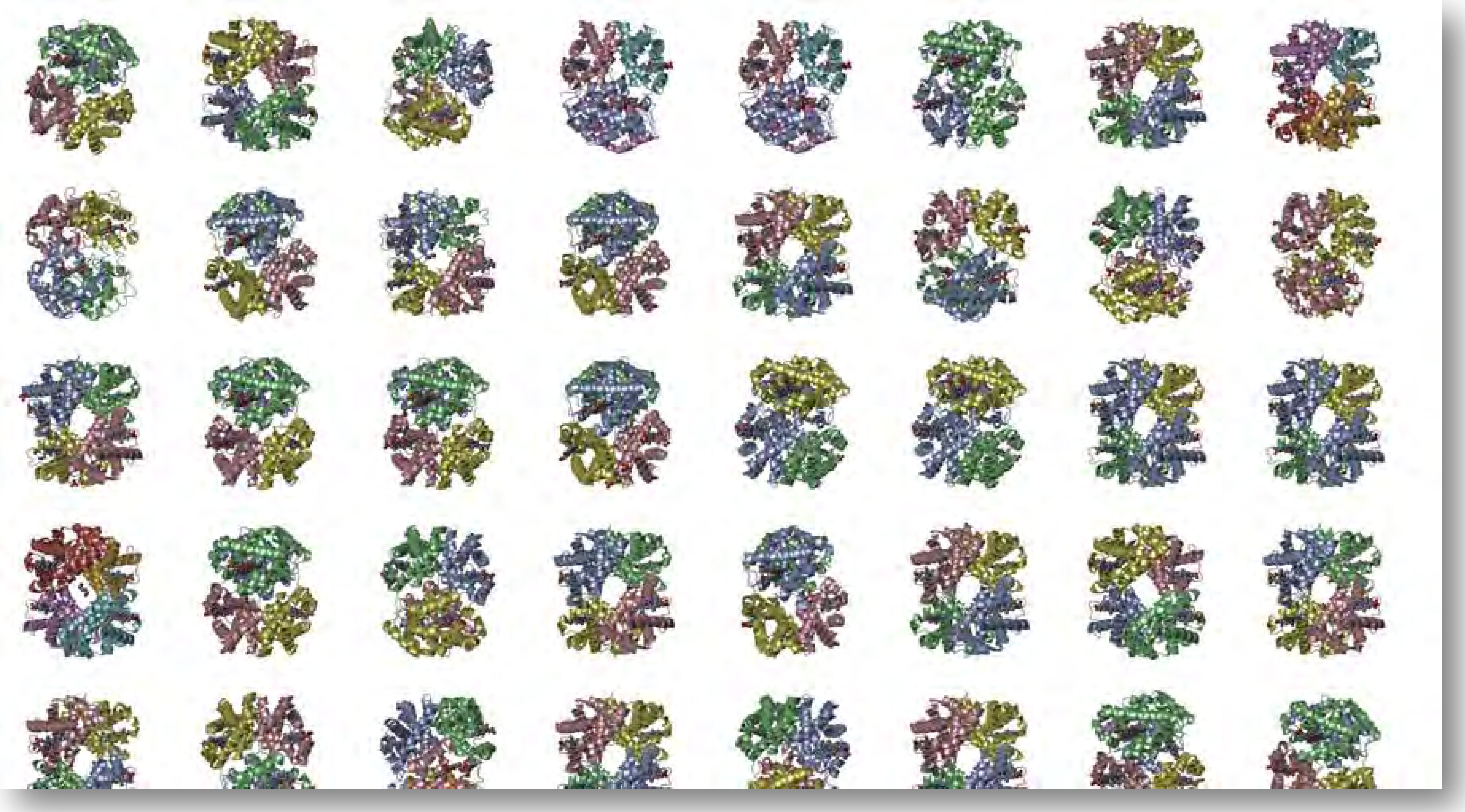
Data type: PDB format | mrc/ccp4 format | Codebook vectors | Sa

PDB format file (atomic model or SAXS bead model)

File: ファイルを選択 1a00sax-3a.pdb

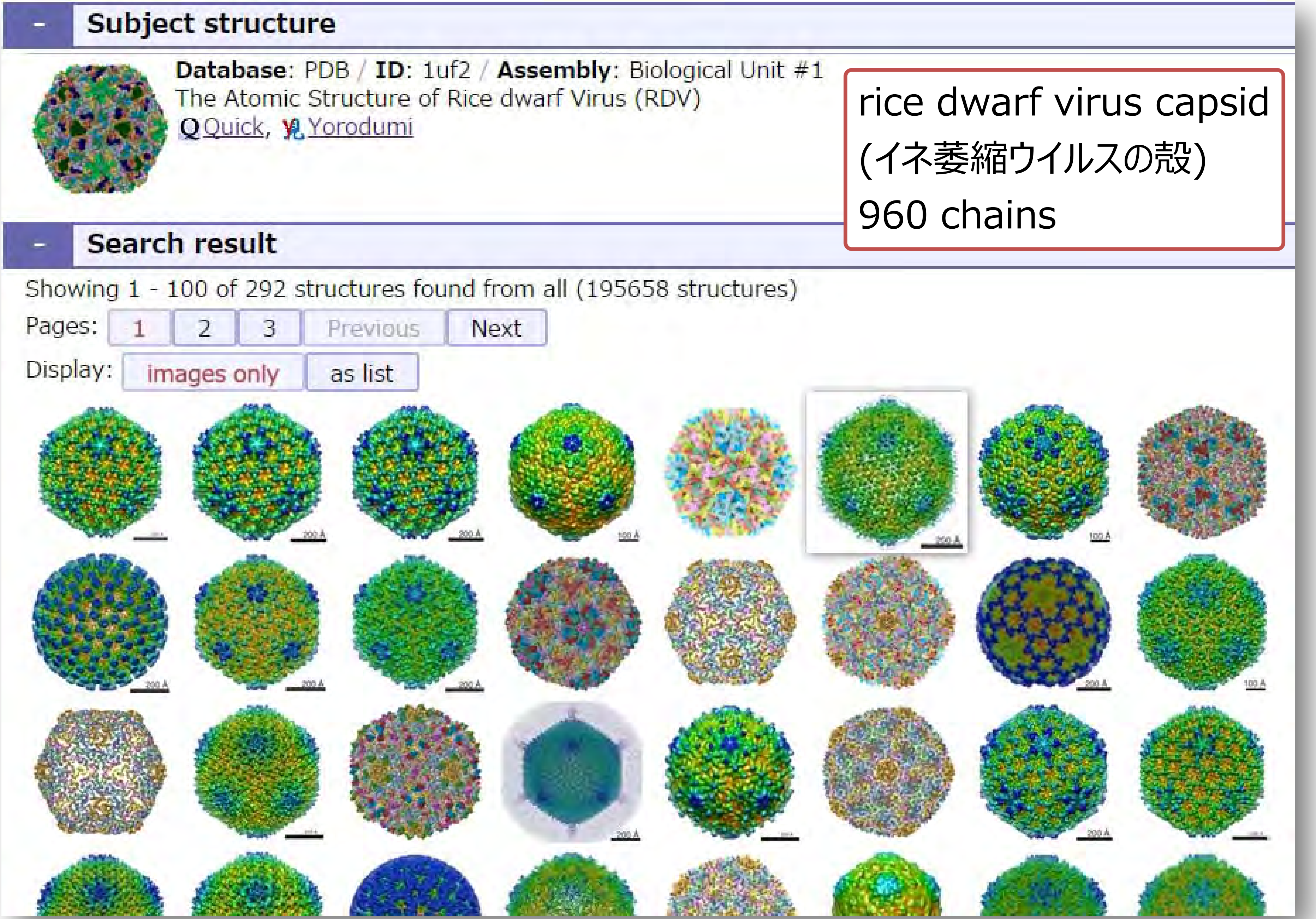
Received data

Example #4: SAXS bead model



Hundreds hemoglobin structures found

Example #5: Huge structures



Subject structure

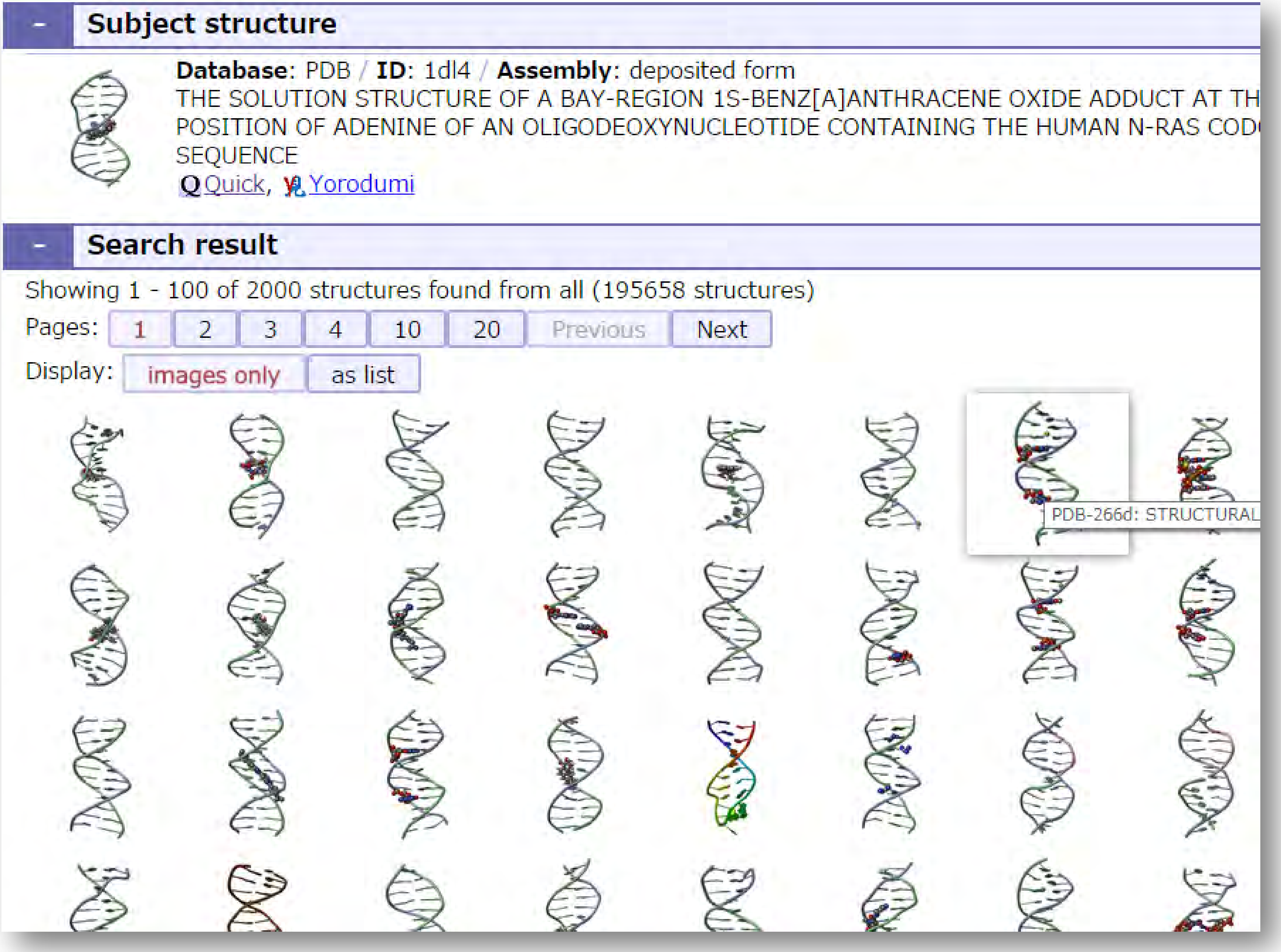
Database: PDB / **ID**: 1uf2 / **Assembly**: Biological Unit #1
The Atomic Structure of Rice dwarf Virus (RDV)
Quick, Yorodumi

Search result

Showing 1 - 100 of 292 structures found from all (195658 structures)

rice dwarf virus capsid
(イネ萎縮ウイルスの殻)
960 chains

Example #6: small DNA fragments



Subject structure

Database: PDB / **ID**: 1dl4 / **Assembly**: deposited form
THE SOLUTION STRUCTURE OF A BAY-REGION 1S-BENZ[A]ANTHRACENE OXIDE ADDUCT AT TH
POSITION OF ADENINE OF AN OLIGODEOXYNUCLEOTIDE CONTAINING THE HUMAN N-RAS COD
SEQUENCE

Search result

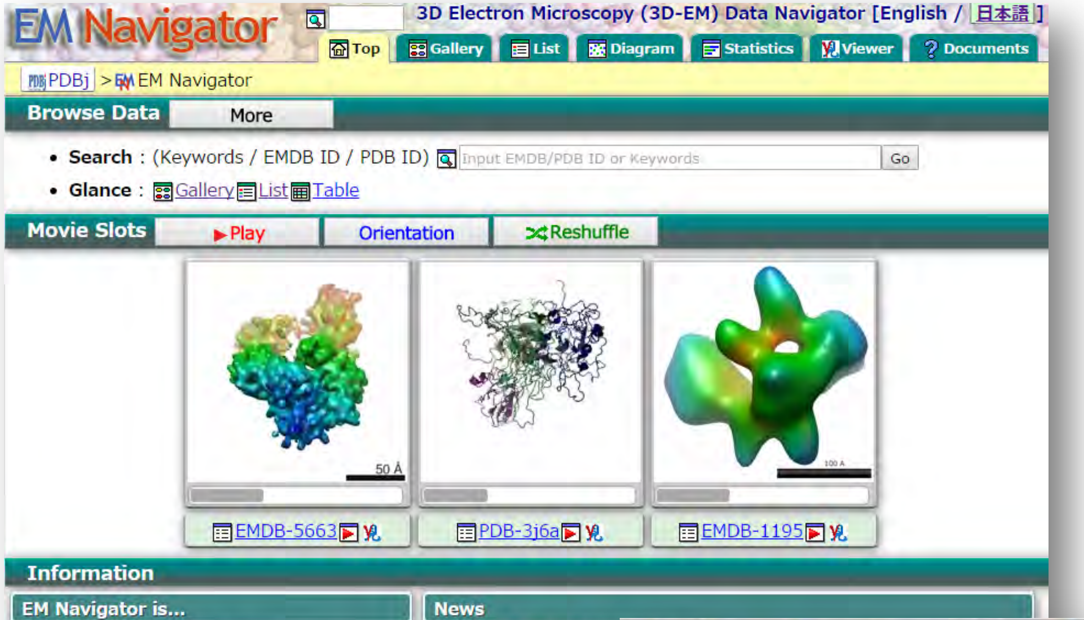
Showing 1 - 100 of 2000 structures found from all (195658 structures)

That’s Omokage & *gmfit*

Enjoy the shapes
of macromolecules
with Omokage & *gmfit*!

Other services related to EM data

EM Navigator & Yorodumi



3D Electron Microscopy (3D-EM) Data Navigator [English / 日本語]

PDBj > EM Navigator

Browse Data

- **Search** : (Keywords / EMDB ID / PDB ID) Input EMDB/PDB ID or Keywords
- **Glance** : Gallery List Table

Movie Slots

EMDB-5663 PDB-3j6a EMDB-1195

Information

EM Navigator is...

- web site to browse 3D electron microscopy (3D-EM) data of biological molecules and assemblies,
- based on EMDB and PDB data, (statistics)

News

- **2014-09-10:** Lat

“EM Navigator”

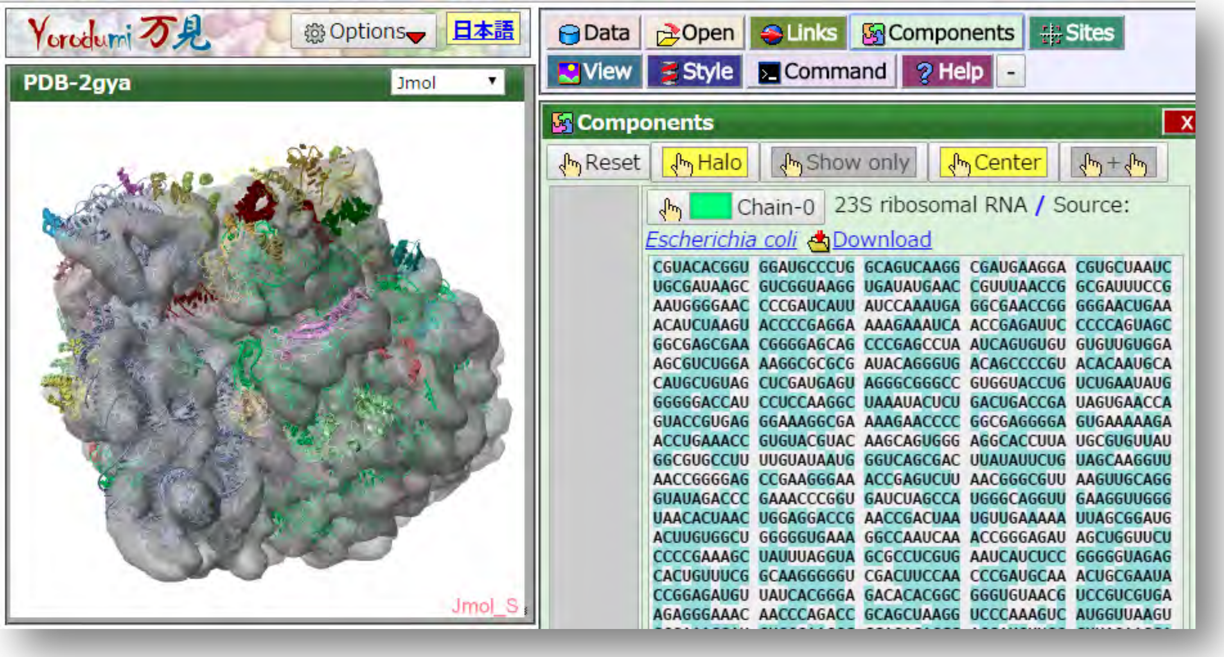
EM data explorer

http://pdbj.org/emnavi

“Yorodumi”

Structure viewer

http://pdbj.org/yorodumi



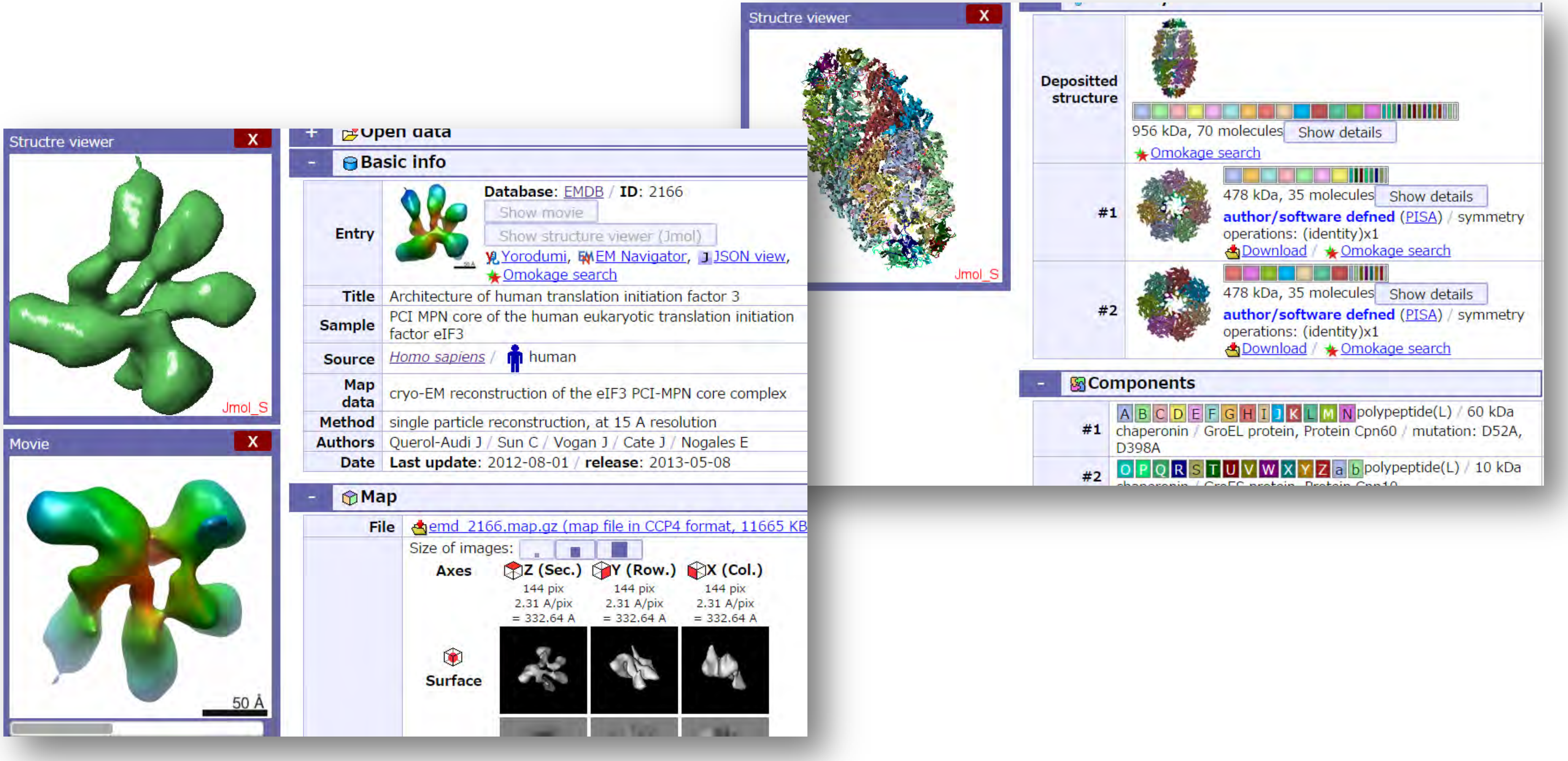
PDB-2gya

Components

Chain-0 23S ribosomal RNA / Source: *Escherichia coli* Download

```
CGUACACGGU GGAUGCCCUG GCAGUCAAGG CGAUGAAGGA CGUGCUAAUC
UGCGAUAAGC GUCGGUAAGG UGAUAUGAAC CGUUUAACCG GCGAUUUCCG
AAUGGGGAAC CCCGAUCAUU AUCCAAAUGA GGCGAACCGG GGGAACUGAA
ACAUCUAAGU ACCCCGAGGA AAAGAAAUCA ACCGAGAUUC CCCCAGUAGC
GGCGAGCGAA CGGGGAGCAG CCCGAGCCUA AUCAGUGUGU GUGUUGUGGA
AGCGUCUGGA AAGGCGCGCG AUACAGGGUG ACAGCCCCGU ACACAAUGCA
CAUGCUGUAG CUCGAUGAGU AGGGCGGGCC GUGGUACCUG UCUGAAUAUG
GGGGGACCAU CCUCCAAGGC UAAAUACUCU GACUGACCGA UAGUGAACCA
GUACCGUGAG GGAAAGGCGA AAAGAACCCC GGCGAGGGGA GUGAAAAAGA
ACCUGAAACC GUGUACGUAC AAGCAGUGGG AGGCACCUUA UGCGUGUUAU
GGCGUGCCUU UUGUAUAAUG GGUCAGCGAC UUAUAUUCUG UAGCAAGGUU
AACCGGGGAG CCGAAGGGAA ACCGAGUCUU AACGGGCGUU AAGUUGCAGG
GUAUAGACCC GAAACCCGGU GAUCUAGCCA UGGGCAGGUU GAAGGUUGGG
UAACACUAAC UGGAGGACCG AACCGACUAA UGUUGAAAAA UUAGCGGAUG
ACUUGUGGCU GGGGGUGAAA GGCCAAUCAA ACCGGGAGAU AGCUGGUUCU
CCCCGAAAGC UAUUUAGGUA GCGCCUCGUG AAUCAUCUCC GGGGGUAGAG
CACUGUUUCG GCAAGGGGGU CGACUUCCAA CCCGAUGCAA ACUGCGAAUA
CCGGAGAUGU UAUCACGGGA GACACACGGC GGGUGUAACG UCCGUCGUGA
AGAGGGAAAC AACCCAGACC GCAGCUAAGG UCCCAAAGUC AUGGUUAAGU
```

New EM Navigator + Yorodumi



Entry	**Database**: EMDB / **ID**: 2166 Yorodumi, EM Navigator, JSON view, Omokage search
Title	Architecture of human translation initiation factor 3
Sample	PCI MPN core of the human eukaryotic translation initiation factor eIF3
Source	*Homo sapiens* / human
Map data	cryo-EM reconstruction of the eIF3 PCI-MPN core complex
Method	single particle reconstruction, at 15 A resolution
Authors	Querol-Audi J / Sun C / Vogan J / Cate J / Nogales E
Date	**Last update**: 2012-08-01 / **release**: 2013-05-08

File emd_2166.map.gz (map file in CCP4 format, 11665 KB

Size of images:

Axes	**Z (Sec.)**	**Y (Row.)**	**X (Col.)**
	144 pix 2.31 A/pix = 332.64 A	144 pix 2.31 A/pix = 332.64 A	144 pix 2.31 A/pix = 332.64 A
Surface			

Depositted structure	956 kDa, 70 molecules Omokage search
#1	478 kDa, 35 molecules **author/software defned** (PISA) / symmetry operations: (identity)x1 Download / Omokage search
#2	478 kDa, 35 molecules **author/software defned** (PISA) / symmetry operations: (identity)x1 Download / Omokage search

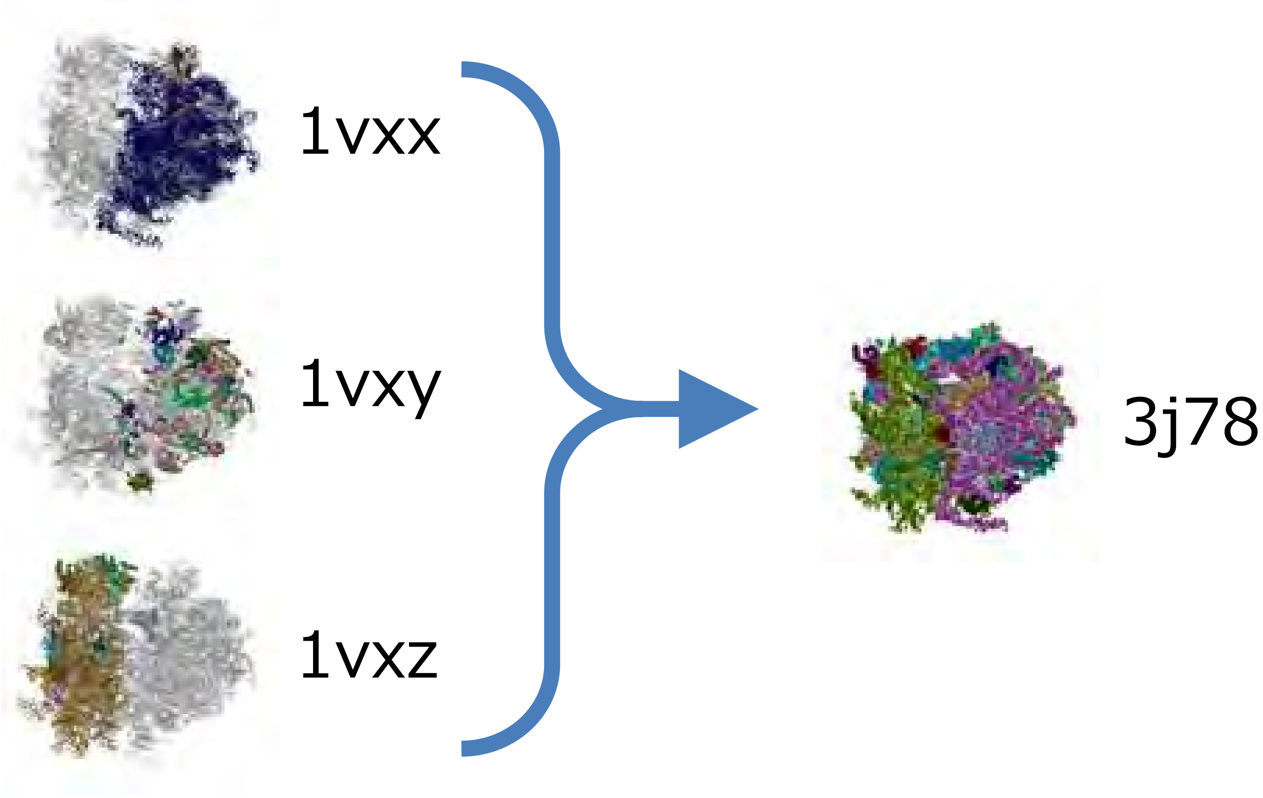
#1	A B C D E F G H I J K L M N polypeptide(L) / 60 kDa chaperonin / GroEL protein, Protein Cpn60 / mutation: D52A, D398A
#2	O P Q R S T U V W X Y Z a b polypeptide(L) / 10 kDa

- simpler and unified interface
- integration with Omokage services
- much better support for mobile devices
- support for “large structure”, etc.

coming soon!

Announcements

PDB format will be obsolete



1vxx

1vxy

1vxz

3j78

"split" entries
in PDB format
=> obsolete

"large structure" entries
in PDBx/mmCIF format
=> official

See news from PDBj / wwPDB for details