OneDep: Unified System for Deposition, Biocuration, and Validation of Macromolecular Structures and Ligand Library based on CSD data

Atsushi NAKAGAWA

PDBj: Protein Data Bank Japan

Institute for Protein Research, Osaka University

wwpdb.org

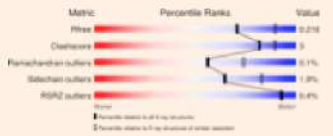
第二十七屆生物物理研討會
2023.5.18

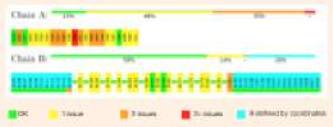
OneDep: Unified System for Deposition, Biocuration, and Validation of Macromolecular Structures

wwpdb.org

OneDep: Unified Deposition Portal for the Protein Data Bank

wwPDB Partners - RCSB PDB, PDBe, PDBj, and BMRB



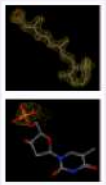


- Overall quality at a glance
- Residue-property plots
- X-ray fit of model and data
- NMR ensemble analysis
- NMR chemical shift validation

System supports
- X-ray, NMR, and 3DEM
- File re-upload
- Submissions based on existing depositions

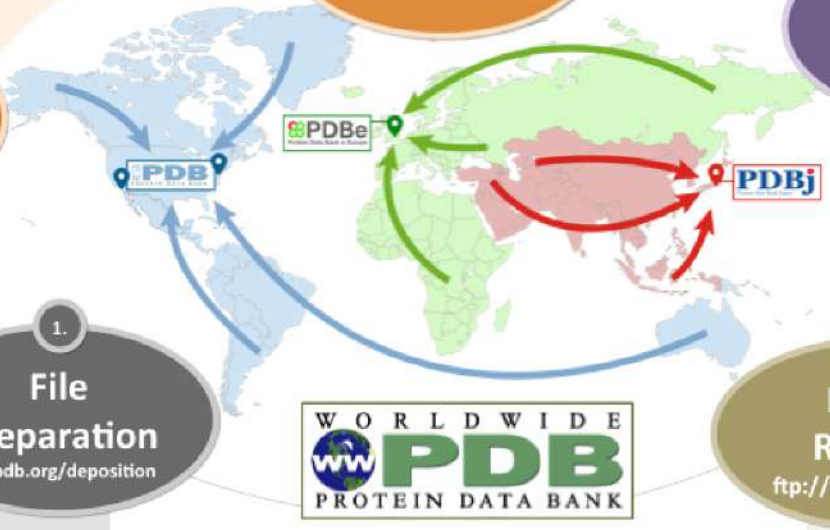
3. Deposition
deposit.wwpdb.org

- Sequence and taxonomy
- Stereochemistry
- Chemical nomenclature
- Electron density fit



4. Biocuration

2. Pre-deposition Validation
validate.wwpdb.org



- Generate atomic coordinate and experimental data files in PDBx/mmCIF format
- Assemble mandatory data items for deposition
- Complete sample polymer sequences including unobserved regions
- Source organism(s)
- Ligand and non-standard chemical details
- Macromolecular assembly and experimental support

1. File Preparation
wwpdb.org/deposition

5. Public Release
ftp://ftp.wwpdb.org

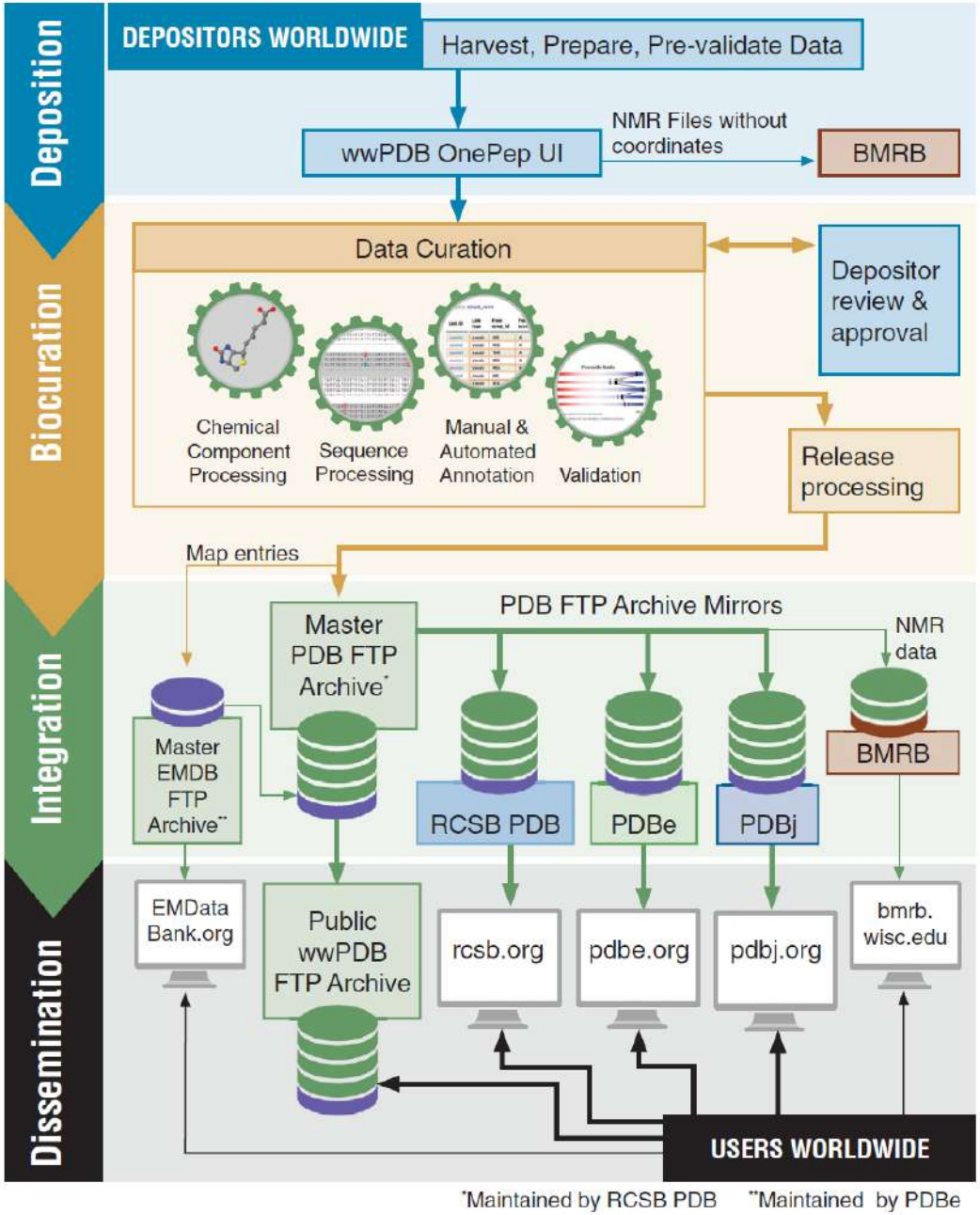
- Entries a… released… followi… Depos… instru… depo…
  - Hold until publication
  - Hold for a period of tim…
  - Release immediately
- Journal policies governing … release take precedence … requested embargo peri…

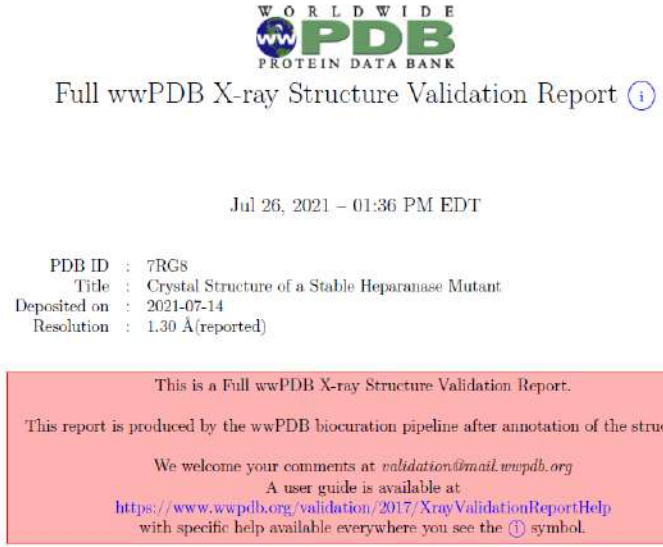
WORLDWIDE PDB PROTEIN DATA BANK

OneDep, a unified system for deposition, biocuration, and validation of experimentally determined X-ray, NMR, and 3DEM macromolecular structures, has been developed as a global project by the Worldwide PDB partners (wwPDB). The system unifies deposition, biocuration, and validation procedures across all wwPDB, EMDB, and BMRB deposition centers with a focus on simplifying and improving the process for the Depositor, improving data quality and completeness, and addressing challenges of the increasing number, size, and complexity of structures deposited to the Protein Data Bank archive.

Acknowledgements

Overview of OneDep system





WORLDWIDE PDB PROTEIN DATA BANK

Full wwPDB X-ray Structure Validation Report

Jul 26, 2021 – 01:36 PM EDT

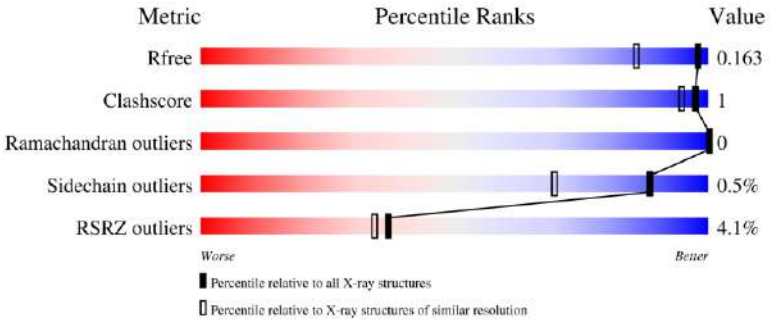
PDB ID : 7RGS
Title : Crystal Structure of a Stable Heparanase Mutant
Deposited on : 2021-07-14
Resolution : 1.30 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report.

This report is produced by the wwPDB biocuration pipeline after annotation of the structure.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the ⓘ symbol.

wwPDB validation report (PDF)

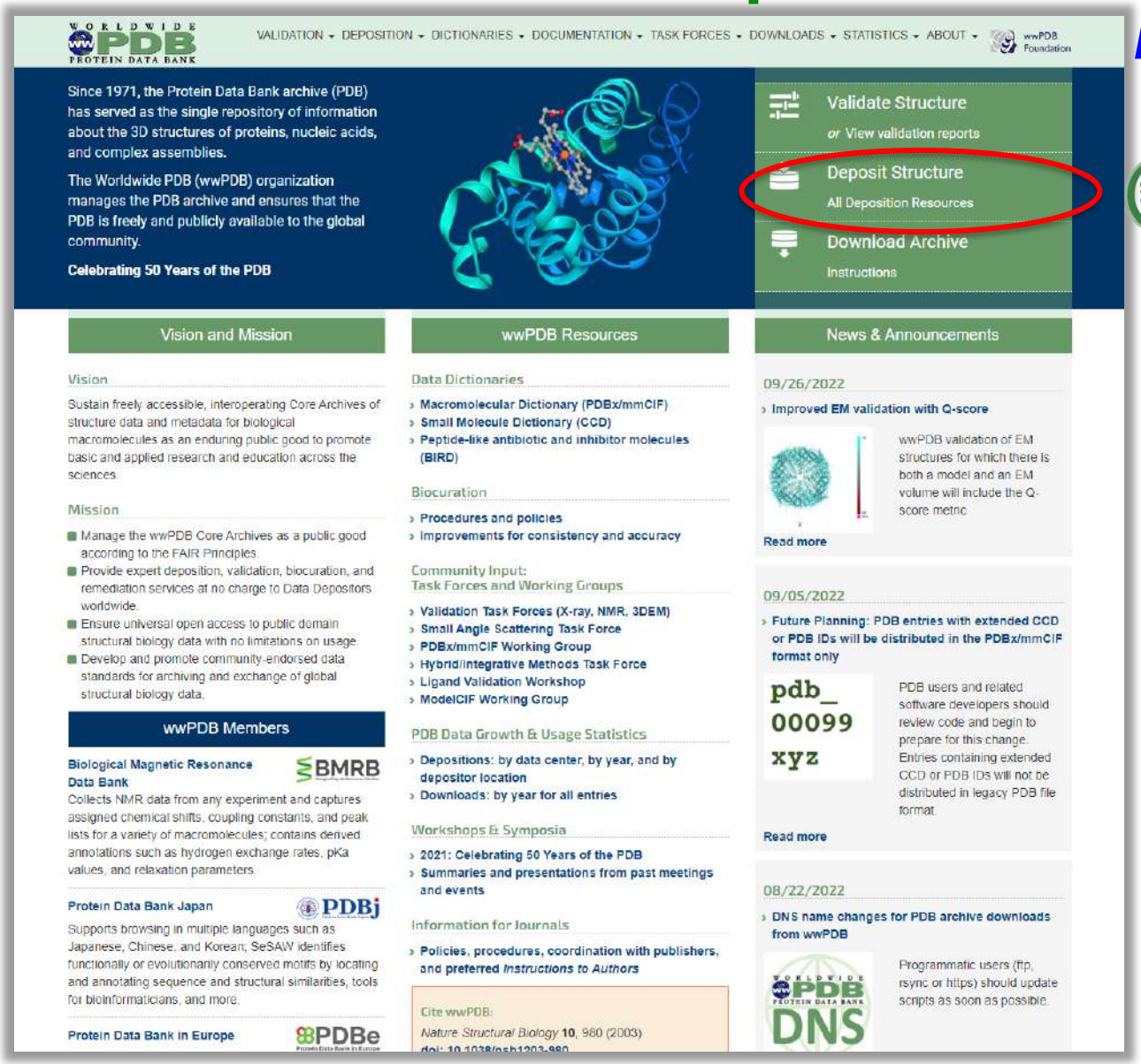


Percentile view of validation report

Young, JY et al., Structure, 25(3), 536-545, 2017

wwPDB Common Deposition & Annotation

https://wwpdb.org

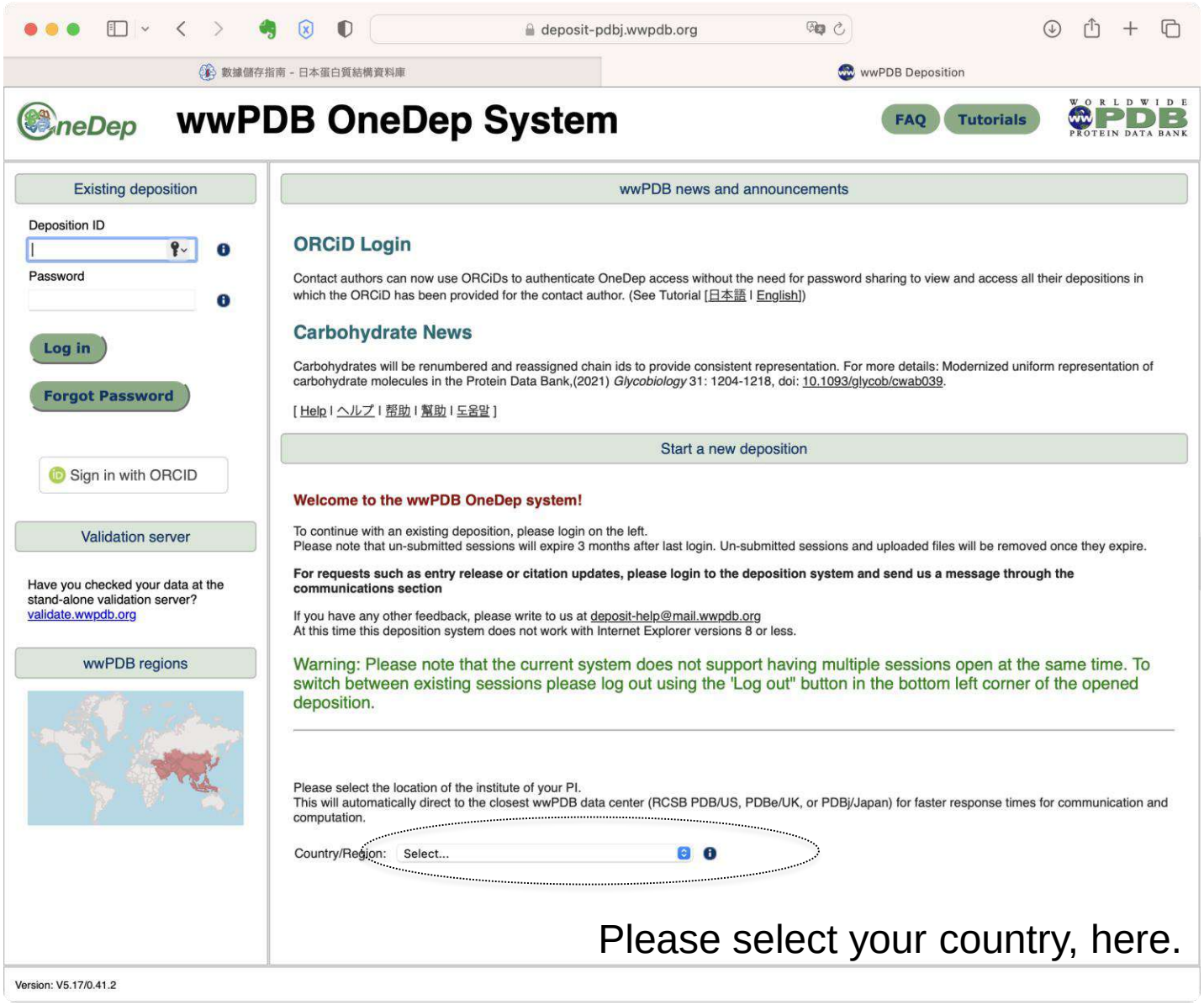


OneDep

wwPDB Common Deposition & Annotation

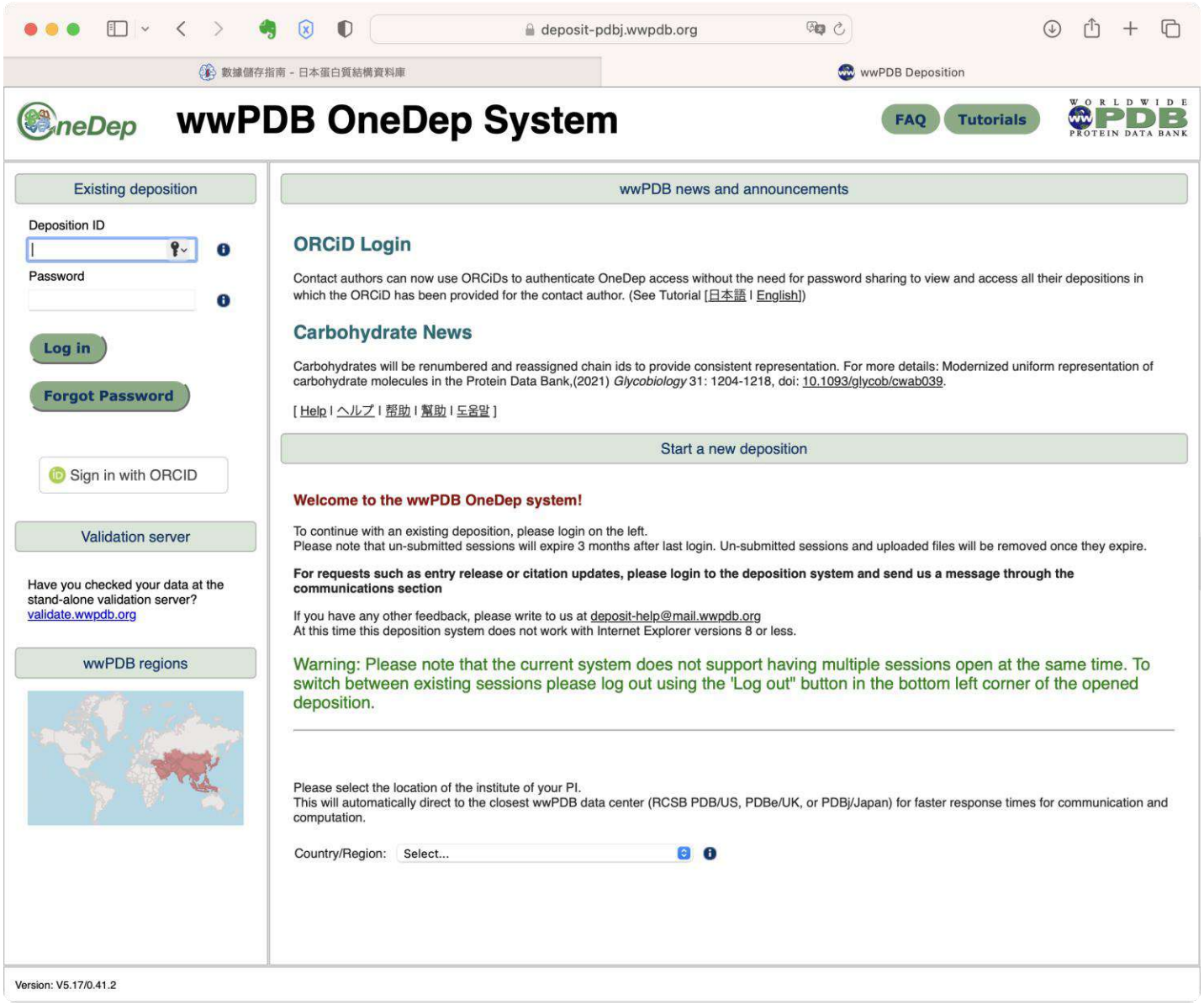
https://pdbj.org

OneDep



Please select your country, here.

After choosing your country/region

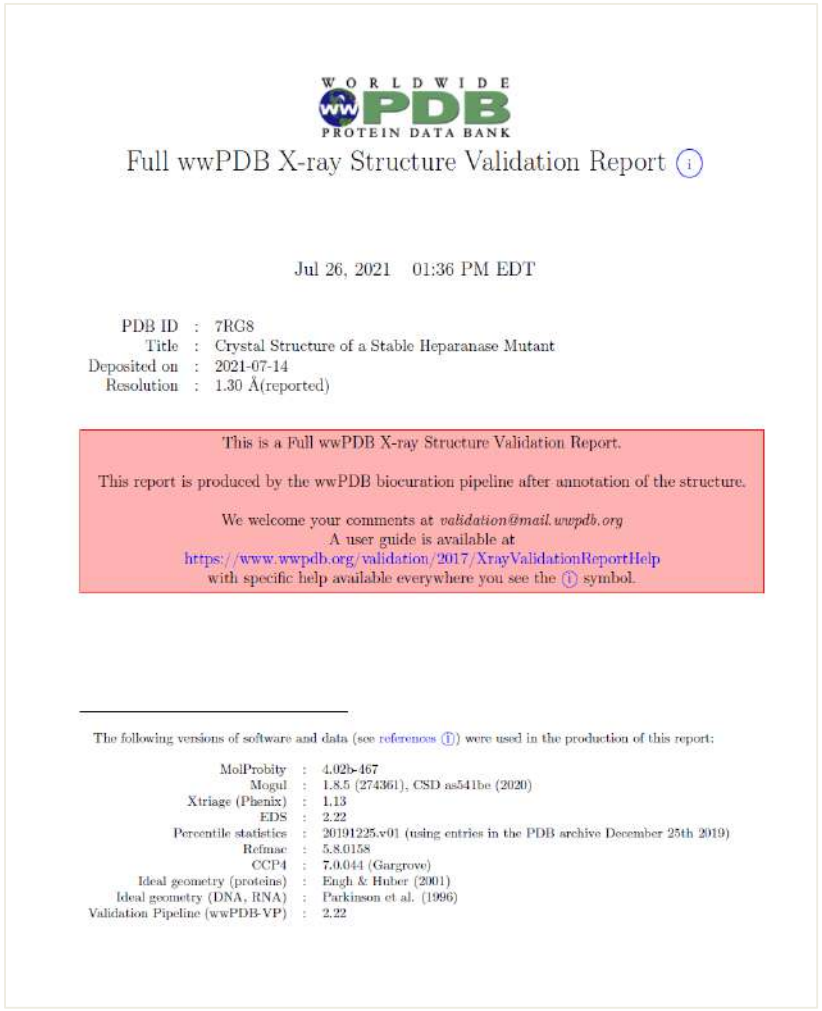


Recent updates of OneDep system

- Validation report needed for peer review
- Improvements to visualization of ligand validation and electron density maps in the wwPDB validation report
- Mandatory PDBx/mmCIF format files submission for MX depositions
- Improve your previously-released coordinates and keep your original PDB ID
- Carbohydrate remediation
- PDBx/mmCIF data files to include PI information
- Extended PDB IDs and PDB DOIs now available in PDBx/mmCIF files
- Future Planning: PDB entries with extended CCD or PDB IDs will be distributed in the PDBx/mmCIF format only

Official wwPDB Validation Report is requested for peer review

○



WORLDWIDE PDB
PROTEIN DATA BANK

Full wwPDB X-ray Structure Validation Report (i)

Jul 26, 2021 01:36 PM EDT

PDB ID : 7RG8
Title : Crystal Structure of a Stable Heparanase Mutant
Deposited on : 2021-07-14
Resolution : 1.30 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report.

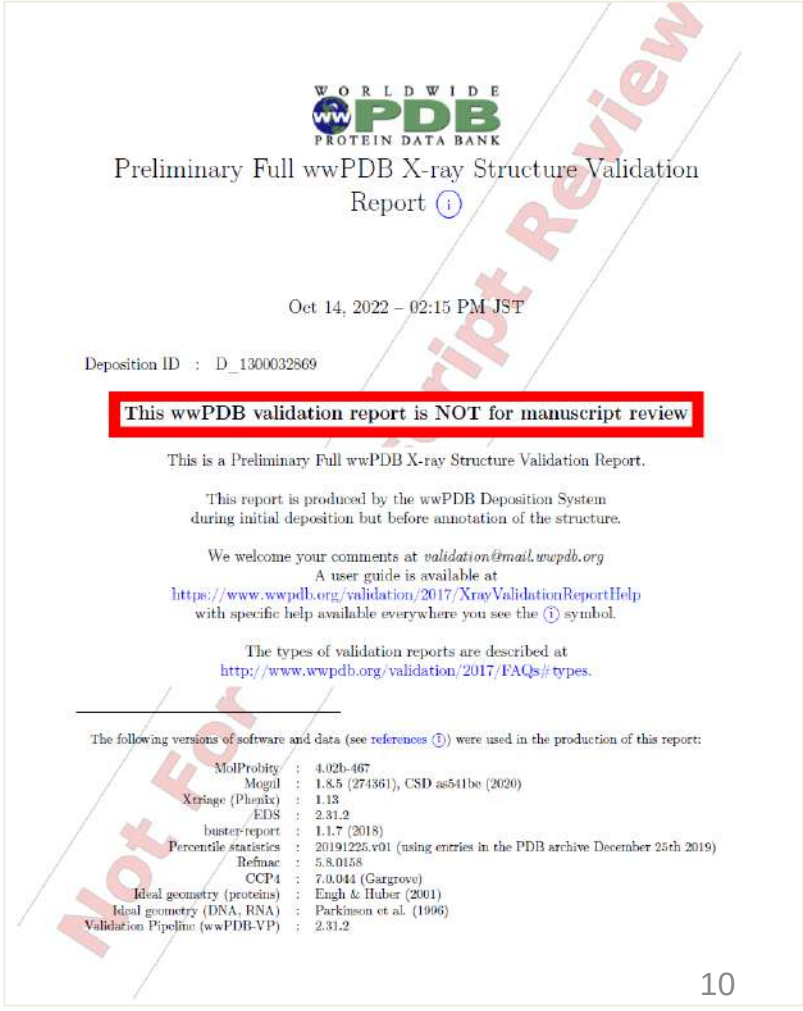
This report is produced by the wwPDB biocuration pipeline after annotation of the structure.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (i)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.22
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.22

×



WORLDWIDE PDB
PROTEIN DATA BANK

Preliminary Full wwPDB X-ray Structure Validation Report (i)

Oct 14, 2022 – 02:15 PM JST

Deposition ID : D_1300032869

This wwPDB validation report is NOT for manuscript review

This is a Preliminary Full wwPDB X-ray Structure Validation Report.

This report is produced by the wwPDB Deposition System during initial deposition but before annotation of the structure.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.31.2
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.2

Validation metrics in wwPDB validation reports

X-ray/EM/NMR

- Geometric & conformational
  - bond, angle, planarity
  - protein backbone conformation
  - protein side-chain conformation
- Atomic & molecular interaction
  - all-atom contacts
  - under packing
  - hydrogen bond quality
- Non-protein
  - nucleic acids (RNA pucker, suite)
  - carbohydrates (N-glycan core)
  - ligands (CSD)
  - ions & other solvent
- Incomplete model (e.g. CA_ONLY)

X-ray

- Structure factor & electron density
  - Wilson plot, outliers, tNCS
  - wrong symmetry
  - twinning
  - agreement (R_{free}, RSR, RSCC)

NMR

- Chemical shifts
  - completeness
  - outliers
  - estimated reference error
  - random coil index
- Structure ensembles
  - representative model (medoid)
  - domain detection

Caveat:

LLDF (Local Ligand Density Fit) has been replaced by a combination of RSR (Real-space R factor) > 0.4 and RSCC (Real-space correlations coefficient) < 0.8 since this March.

wwPDB validation report PDF

- **Standard geometry**

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\#\|Z\|>5$	RMSZ	$\#\|Z\|>5$
1	A	0.47	0/1107	0.71	0/1491

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

- **Too close contacts**

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1091	0	1106	7	0
2	A	22	0	27	2	0
3	A	100	0	0	2	0
All	All	1213	0	1133	9	0

- **Protein backbones**

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	135/137 (98%)	132 (98%)	3 (2%)	0	100 100

- **Protein sidechains**

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	123/123 (100%)	120 (98%)	3 (2%)	52 38

- **Ligand geometry**

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\|>2$	Counts	RMSZ	$\#\|Z\|>2$
2	REA	A	200	-	19,22,22	1.05	1 (5%)	26,30,30	1.02	2 (7%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	200	REA	C1-C6	2.25	1.56	1.53

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	200	REA	C11-C10-C9	-2.40	123.89	127.31
2	A	200	REA	C18-C5-C6	2.08	126.83	124.51

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	200	REA	2	0

Validation software utilized for generation of wwPDB validation report (2022)

Component Software Packages Included in the 2022 Version of the Validation Pipeline

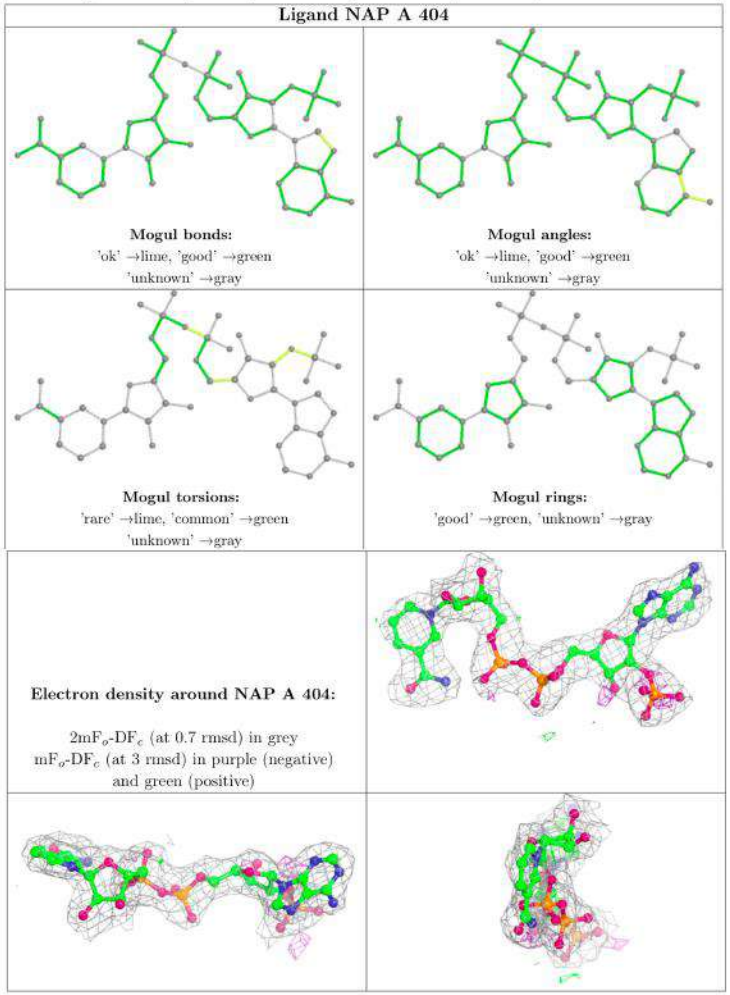
Software Package	Version
MolProbity	4.02b-467
Mogul	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	1.13
EDS	2.31.2
buster-report	1.1.7 (2018)
Percentile statistics	20191225.v01
Refmac	5.8.0158
CCP4	7.0.044 (Gargrove)
Ideal geometry (proteins)	Engh & Huber (2001)

Ligand Validation

- Adapted software from Global Phasing Ltd. under a formal agreement
- Provides geometrical quality in 2D depiction
- Provides electron density fit for X-ray
- Now mandatory at deposition: identification of Ligand/s Of Interest (LOI, author's research focus)

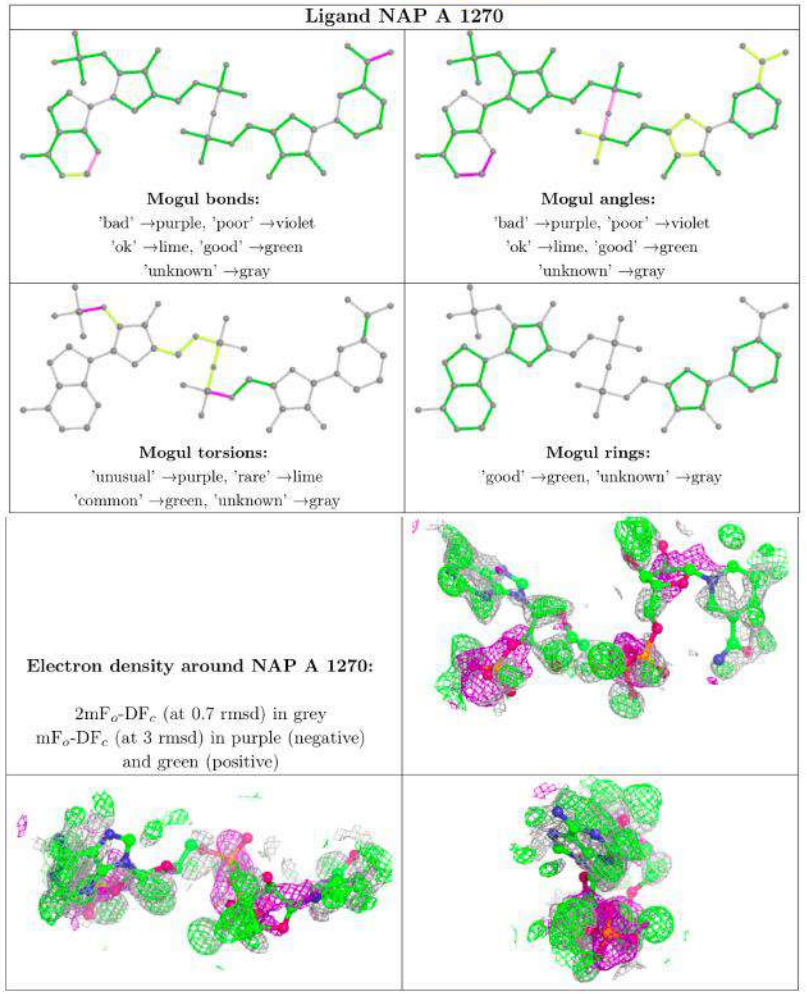
Ligand Validation- Examples on NADP Ligands

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	NAP	A	404	48/48	0.96	0.14	31,43,66,70	0



PDB entry 5zix (Better data quality)

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	NAP	A	1270	48/48	-0.06	0.67	87,96,100,100	0



PDB entry 1zk4 (Worse data quality)

Mandatory PDBx/mmCIF format files submission for MX depositions

PDB File formats from wwPDB

- (Legacy) PDB format
  - *NOT SUPPORTED !*
- **PDBx/mmCIF**
  - The canonical format of the wwPDB.
  - Ver. 5.361 released. (2022-08-31)

Coordinate Versioning

- To improve data quality in the PDB archive
- Depositors can now make corrections to existing structures in
- the PDB Core Archive by updating the atomic coordinates while preserving the original PDB identifier.
- Deployed on July 27th 2019 for all structures that were originally deposited via OneDep (Phase one)
- First coordinate replacement (PDB entry 5T26) versioned at FTP on August 7th 2019 following reviewers' comments based on the wwPDB validation reports
  - With PDB prefix and extension of 4 characters (e.g., from "1ABC" to "PDB_00001ABC")
  - Example: PDB_00001ABC_XYZ_V2-2.cif.gz

Deposition Notes

- Model coordinates
  - TER records
  - Coordinates for Chimera protein
  - ALA models
- Sample sequence
  - Complete polymer sequence
  - Non-poly residues
  - UNK
  - Source organism
- Assembly details
- Deposition efficiency
- Communication, contact authors & ORCiD
- Release

Model coordinates

TER cards must follow each protein chain,
but can only follow a protein (or nucleotide) chain.

Chain A protein
TER
Chain A ligand
Chain B protein
TER
Chain B ligand
Chain C protein
TER
Chain C ligand
Chain D ligand

ALA model

- When an amino acid residue is disordered due to low density

Side chain atoms cannot be assigned, and the residue is often modeled as:

- ALA model
- GLY model
- SER model

Etc.

However, the residue name in the coordinates should not be changed to MATCH with the sequence even without the atoms in the side chain.

Sample Sequence

Fill in a complete polymer sequence used for experiment

- Please include

HIS- or other Expression tags, Linker, Residues missing from the coordinates due to disorder

Coordinates: . . . . . LVVVTNNLR . . .RIPGIRIED . . . ITLMELILEH . . . . .

Sequence: GSHMALVVVTNNLREFERIPGIRIEDGSGITLMELILEHHHHHH

- Please don't include

Residues cleaved from the macromolecules prior to or during the experiment

DO NOT include ligands

Sample sequence is a list of the consecutive chemical components covalently linked in a linear fashion to form a polymer.

Please don't include:

- Metal ions, Chemical components or groups covalently linked to side-chains (in peptides)

- Floating Metal ions, Chemical components or groups

incorrect sequence: NLREFERIPG(NAG)IRIEDYTYITLMELILEHHH(NAG)(ZN)

(Should be) NLREFERIPGIRIEDYTYITLMELILEHHH

Use "UNK” or ”N" ONLY in TWO special cases

UNK : unknown amino acid

N or DN : unknown nucleotide

1. You don't know the sequence
2. You know the sequence but don't know how the coordinates align with the sequence

(Example)

You know the sequence but are not sure if the first residue seen in the density is really the first residue of the sequence.

(Note)

Please use "UNK" or ”N" in the coordinates only when you use "UNK" or ”N" in sequence.

Provide details of source information

For an exact sequence database reference match,

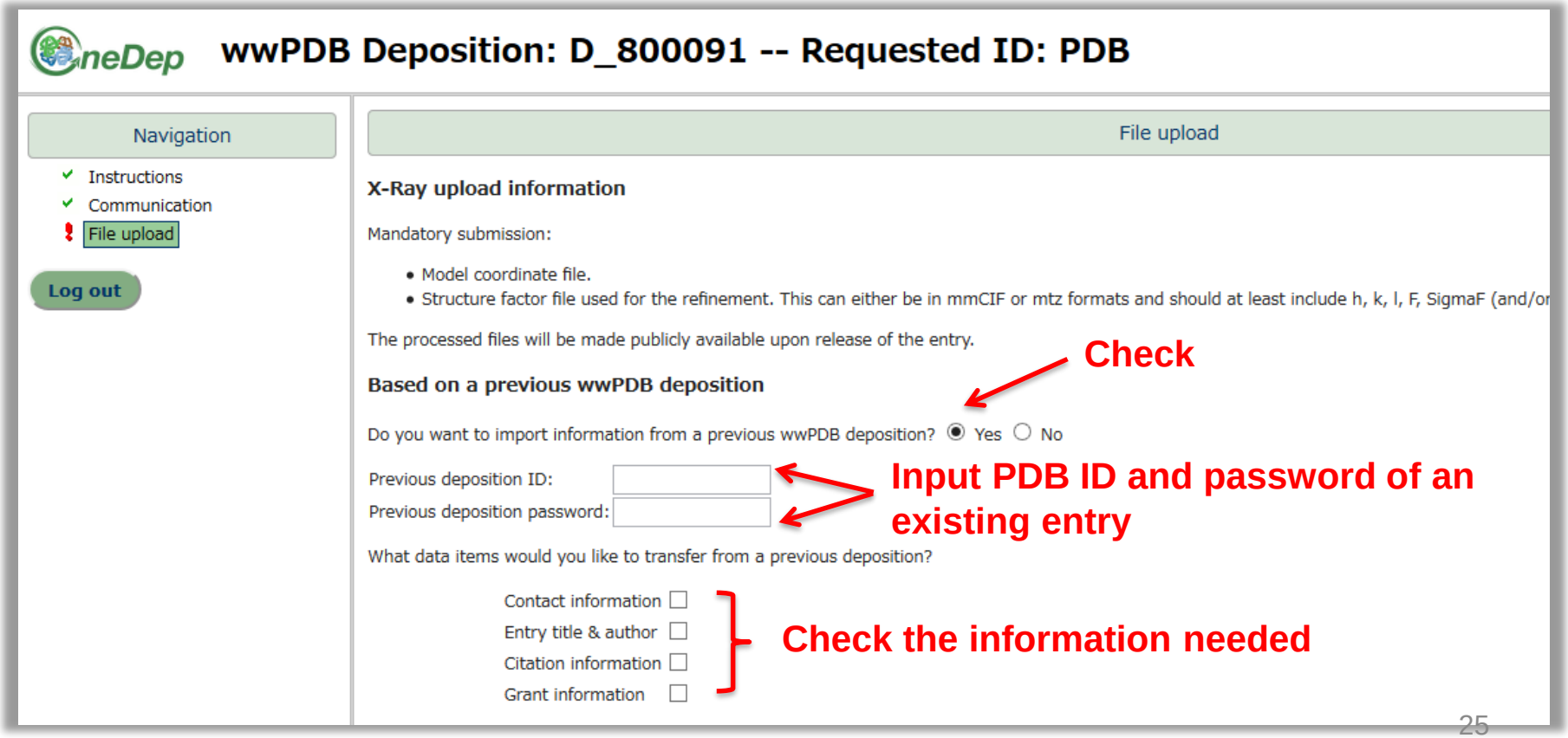
please provide detail information (e.g. strain)
in addition to source organism name.

Example)
Escherichia coli K-12
Saccharomyces cerevisiae S288c

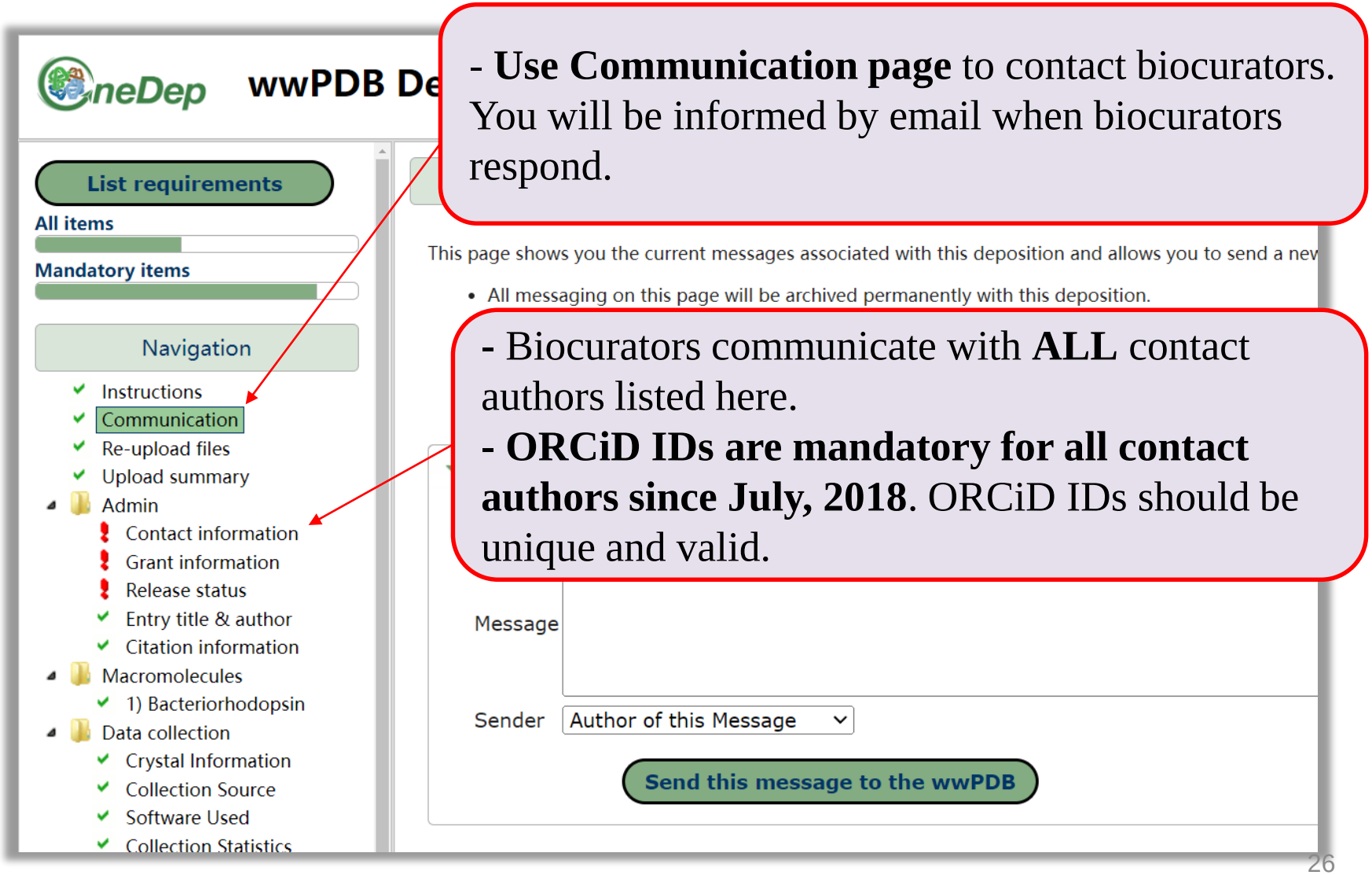
Deposition efficiency

- Deposit Multiple related entries
- Deposit an entry to supersede an existing entry
- Re-submit withdrawn entries

OneDep provides an option to copy information from a deposited entry



Communication, Contact authors & ORCiD



- **Use Communication page** to contact biocurators. You will be informed by email when biocurators respond.

- Biocurators communicate with **ALL** contact authors listed here.
- **ORCiD IDs are mandatory for all contact authors since July, 2018**. ORCiD IDs should be unique and valid.

Release

- Do not ask for replacement of coordinates just before the release
- When you refer to PDB in your paper, please refer as “**the coordinates are deposited to the wwPDB**”.
  - Your entries are processed by PDBj, RCSB or PDBe
- Release instructions (REL, HPUB, HOLD) cannot be changed by depositors
  - Ask biocurators after submission

DOI Landing Page Layout

Ligand Library based on Cambridge Structural Database (CSD) data

Refinement of structures

It is usually difficult to obtain a sufficient number of observations for refinement of atom parameters due to low resolution in X-ray crystallography

Observations/parameter ratio（isotropic B-factor: 4 parameters/atom)
5Å … 0.14
3.5Å … 0.4
2.5Å … 1
2Å … 2.6
1.6Å … 4
1.4Å … 20
1.2Å … 35

Require restraints

Geometry restraints … refmac5, buster
Energy restraints … phenix.refine

Refinement of structures

(eg. refmac5, BUSTER)

- Restrained parameter
  - Stereochemical restrains are combined with X-ray term
    - $f_{\text{total}} = f_{\text{geom}} + w \cdot f_{\text{xray}}$
      - (eg. $\sum_{bond} \frac{1}{\sigma_b^2}(b_m - b_i)$)

(eg. phenix.refine)

- Energy minimization (with simulated annealing)
  - Observations (structure factors) are treated as a part of energy term
    - $T_{xyz} = wxc_{\text{scale}} \cdot T_{exp} + wc \cdot T_{xyz\_restraints}$

Ideal (target) geometries obtained by CSD data （Engh & Huber, 1991)

Restraints Library（dictionary）

refmac5

- $CLIBD/standard_geometry.cif
  - Amino acids
    - Engh and Huber, 1999
      - » Engh, R.A. & Huber, R. (2001). International Tables for Crystallography, Vol F, pp 382-392.
    - 他…（水素原子）
  - Nucleic acids
    - Clowney *et al.*, 1996
      - » Clowney *et al.* (1996), J. Am. Chem. Soc., 118, 509-518.
      - » Parkinson *et al.* (1996) Acta Cryst. D 52, 57-64.
    - NDB (Nucleic Acid Database)
      - » http://ndbserver.rutgers.edu/
- $CLIBD/monomers/…
  - Ligands

phenix.refine

- Conformation Dependent Library (CDL)
- $CLIBD/monomers/…
  - Ligands

Preparation of Ligand Library

AceDRG

F. Long *et al*. ACEDRG: A stereo-chemical description generator for ligands.
Acta Crystallogr. Sect. D Biol. Crystallogr., **73**, 112–122 (2017).

eLBOW

N. W. Moriarty, R. W. Grosse-Kunstleve, P. D. Adams. Electronic ligand builder and optimization workbench (eLBOW): a tool for ligand coordinate and restraint generation.
Acta Crystallogr. D Biol. Crystallogr. **65**, 1074–1080 (2009)

Grade

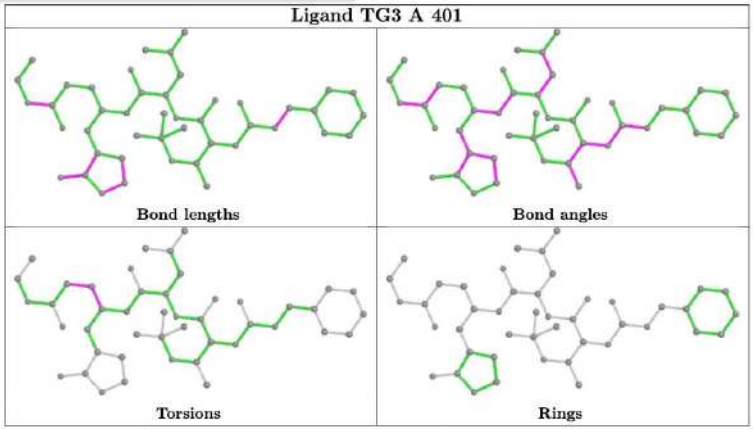
Smart, O. S., Womack, T. O., Sharff, A., Flensburg, C., Keller, P., Paciorek, W., Vonrhein, C. and Bricogne, G. (2011).
Grade, version 1.2.20. Cambridge, United Kingdom, Global Phasing Ltd.

Grade2

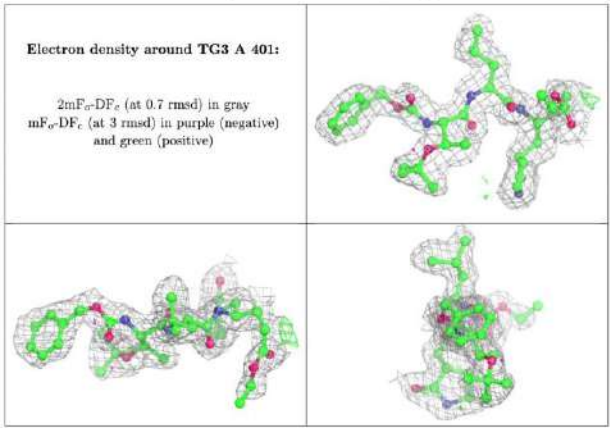
Smart, O.S., Sharff A., Holstein, J., Womack, T.O., Flensburg, C., Keller, P., Paciorek, W., Vonrhein, C. and Bricogne G. (2021).
Grade2 version 1.0.0. Cambridge, United Kingdom: Global Phasing Ltd.

wwPDB Validation Improvement: Ligand

- Partnership with Global Phasing Ltd.
- Ligand of Interest (LOI) highlighted
- 2D views of geometrical quality
- 3D views of electron density fits for X-ray



Ligand ID TG3



PDB ID 7JT7

Feng, Z. *et al.*, Structure, 2021 doi: 10.1016/j.str.2021.02.004

wwPDB validation report uses buster-report (by Global Phasing Inc.)

06/11/2019

Improvements to visualization of ligand validation and electron density maps in the wwPDB validation report

Our recent update to the wwPDB validation reports provides much clearer validation information for ligands.

We now include 2-dimensional diagrams of ligands, highlighting geometric validation criteria and, for structures determined by crystallography, 3-dimensional views of electron density.

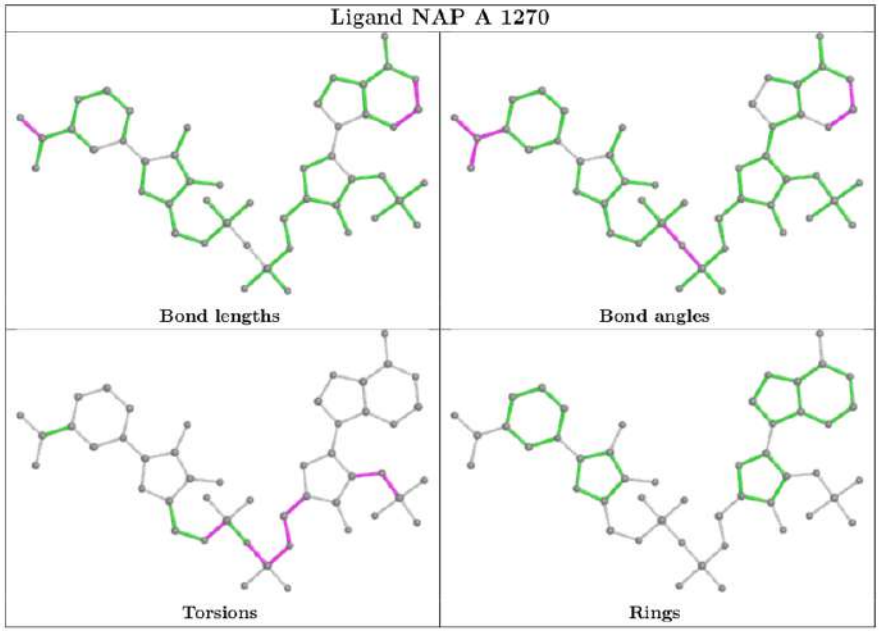
We also provide calculated electron density map coefficients which were used to generate the analysis in the validation reports.

Ligand Validation

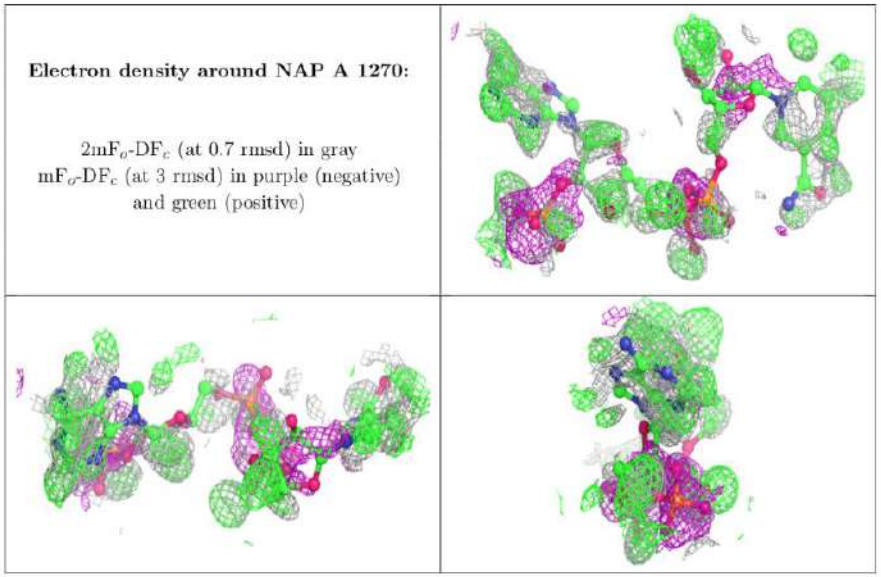
We have collaborated with **Global Phasing Ltd** to integrate the ligand visualization from **buster-report** into the wwPDB validation report, as recommended by the **wwPDB/CCDC/D3R Ligand Validation Workshop**. The ligand visualization will be available for ligands that have been designated as "Ligand of Interest" by the depositor and ligands with a molecular weight greater than 250 Daltons that have outliers.

The following ligand instance of NAP was chosen intentionally as a representative of sub-optimal quality in both the ligand model and its agreement with the X-ray data.

Geometric analysis provided by CCDC Mogul will be highlighted on a 2D diagram of the ligand, as shown below.



In addition to geometric validation for ligands, for X-ray diffraction PDB entries the wwPDB validation report also presents images displaying the ligand and the surrounding electron density map.



Electron Density Map Coefficient Files

We are now providing depositors with electron density map coefficient files (2mFo-DFc and mFo-DFc) from the wwPDB validation pipeline alongside the wwPDB validation report. The electron density map coefficients generated for wwPDB validation reports will be made available to end users in the PDB archive as new entries are released and for existing entries when validation reports for the PDB archive are recalculated.

We hope that these changes to the wwPDB validation pipeline will help depositors to interpret the validation information provided for PDB entries more easily. If you have any queries, please contact the wwPDB at **deposit-help@mail.wwpdb.org**.

BUSTER

- BUSTER is a refinement package developed by Global Phasing Ltd. (Director: Gérard Bricogne)
- s a maximum-likelihood macromolecular refinement package based on TNT (Dale Trounrud and Lyn TenEyck) (requires TNT license)
- BUSTER assembles the structural model, scales observed and calculated structure-factor amplitudes and computes the model likelihood, whilst TNT handles the stereochemistry and NCS restraints/constraints and shifts the atomic coordinates, B factors and occupancies.
- Requires CCP4 program suite

BUSTER

- grade uses the RM1 semiempirical method as implemented in the fdynamo library written by Martin Field and co-workers.
- visualise-geometry-coot can visualize the results of a BUSTER refinement
- grade (and grade2) typically runs the CCP4 libcheck program to produce an initial dictionary for a given input SMILES string or structure.
- This initial dictionary is then improved: grade runs the CCDC mogul program (in batch mode) to obtain ideal bond lengths/sigma, bond angles/sigmas and planar torsions from similar structures in the CSD.
- Using OpenBabel (http://openbabel.org) to convert a PDB file containing 3D coordinates to a MOL2 file fills in the atom types and bond orders.
- reduce from Molprobily is used to add all hydrogens explicitly and to perform HQN flip

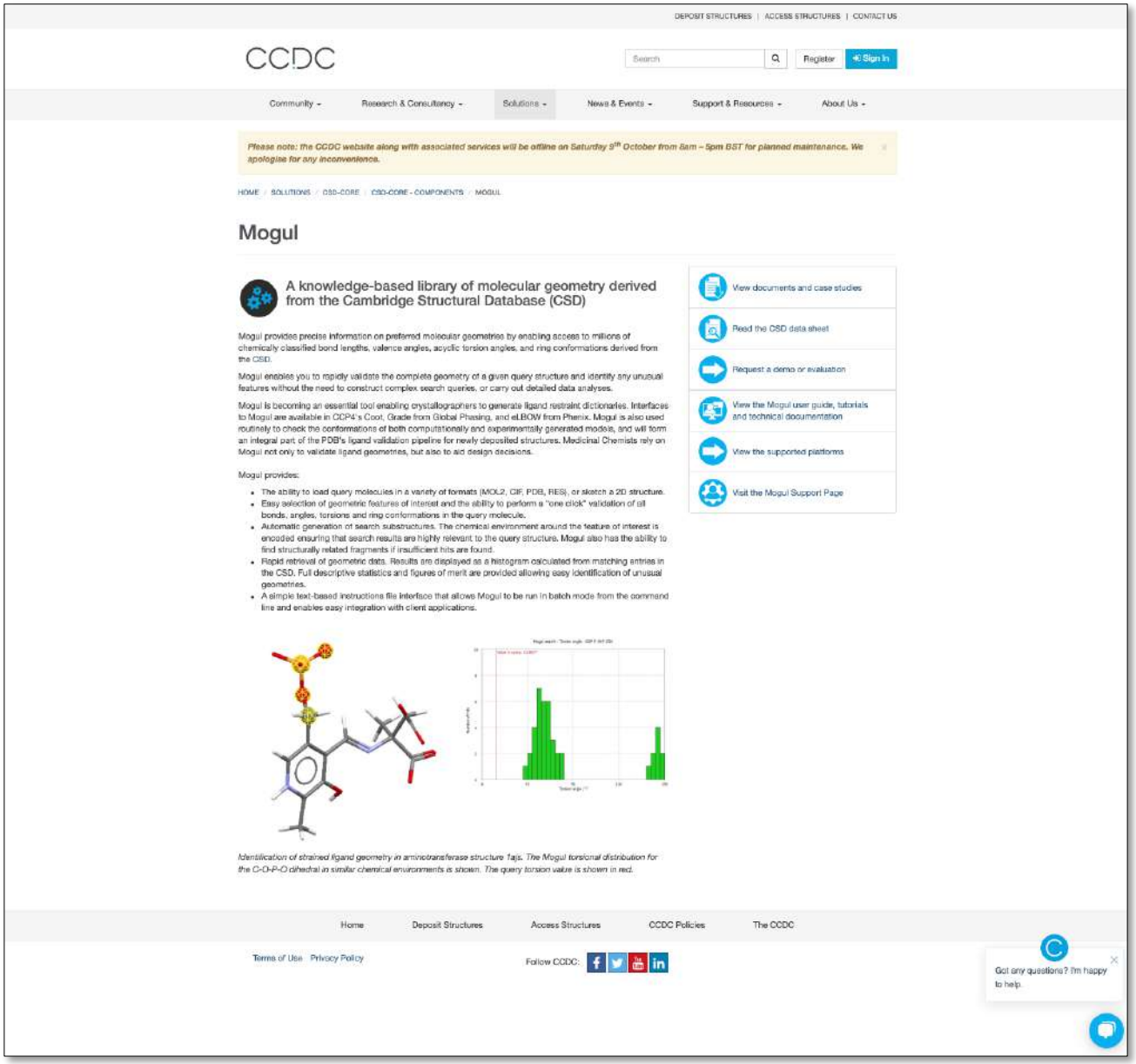
BUSTER Input data

- PDB … checked by pdbchk (in BUSTER)
  - each atom should have a chain identifier (e.g A, B and C for protein chains and W for water)
  - a correct CRYST1 card is required (see PDB format guide), especially the space group symbol.
  - although not enforced by the PDB standard, it seems sensible to use letters (A, B, C etc) in column 17 of ATOM/HETATM records to denote alternate conformations and numbers (1, 2, 3 etc) in column 27 of ATOM/HETATM records to denote insertion code.
  - standard SSBOND and LINK records are supported and often required (to describe correctly the molecular connectivity).

BUSTER Input data

- MTZ
  - normal MTZ file with F/SIGF columns (any column name is possible, but the column types have to be F/Q - which they nearly always are anyway, unless something went really wrong)
  - the cell parameters for the refinement are taken from the MTZ file header (please note that it does not yet handle different cell entries for different datasets). The assumption is that the MTZ file usually contains only a single dataset.
  - if the MTZ file contains a set of columns with Hendrickson-Lattmann coefficients (usually named HLA, HLB, HLC and HLD) these can be used as additional, external phase information (unless the MTZ file is actually the output of a previous BUSTER run - which would not be a good idea). The user needs to set the parameter autoBUSTER_hls to the four column names, e.g. with 'refine autoBUSTER_hls="HLA HLB HLC HLD" ... '.

Mogul in CSD


DEPOSIT STRUCTURES | ACCESS STRUCTURES | CONTACT US
CCDC
Search
Register
Sign In
Community
Research & Consultancy
Solutions
News & Events
Support & Resources
About Us
Please note: the CCDC website along with associated services will be offline on Saturday 9th October from 8am – 5pm BST for planned maintenance. We apologise for any inconvenience.
HOME / SOLUTIONS / CSD-CORE / CSD-CORE - COMPONENTS / MOGUL
Mogul
A knowledge-based library of molecular geometry derived from the Cambridge Structural Database (CSD)
Mogul provides precise information on preferred molecular geometries by enabling access to millions of chemically classified bond lengths, valence angles, acyclic torsion angles, and ring conformations derived from the CSD.
Mogul enables you to rapidly validate the complete geometry of a given query structure and identify any unusual features without the need to construct complex search queries, or carry out detailed data analyses.
Mogul is becoming an essential tool enabling crystallographers to generate ligand restraint dictionaries. Interfaces to Mogul are available in CCP4's Coot, Grade from Global Phasing, and eLBOW from Phenix. Mogul is also used routinely to check the conformations of both computationally and experimentally generated models, and will form an integral part of the PDB's ligand validation pipeline for newly deposited structures. Medicinal Chemists rely on Mogul not only to validate ligand geometries, but also to aid design decisions.
Mogul provides:
The ability to load query molecules in a variety of formats (MOL2, CIF, PDB, RES), or sketch a 2D structure.
Easy selection of geometric features of interest and the ability to perform a "one click" validation of all bonds, angles, torsions and ring conformations in the query molecule.
Automatic generation of search substructures. The chemical environment around the feature of interest is encoded ensuring that search results are highly relevant to the query structure. Mogul also has the ability to find structurally related fragments if insufficient hits are found.
Rapid retrieval of geometric data. Results are displayed as a histogram calculated from matching entries in the CSD. Full descriptive statistics and figures of merit are provided allowing easy identification of unusual geometries.
A simple text-based instructions file interface that allows Mogul to be run in batch mode from the command line and enables easy integration with client applications.
View documents and case studies
Read the CSD data sheet
Request a demo or evaluation
View the Mogul user guide, tutorials and technical documentation
View the supported platforms
Visit the Mogul Support Page
Identification of strained ligand geometry in aminotransferase structure 1ajs. The Mogul torsional distribution for the C-O-P-O dihedral in similar chemical environments is shown. The query torsion value is shown in red.
Home
Deposit Structures
Access Structures
CCDC Policies
The CCDC
Terms of Use
Privacy Policy
Follow CCDC:
Got any questions? I'm happy to help.

Grade Web Server

http://grade.globalphasing.org/cgi-bin/grade/server.cgi

GΦL Grade Web Serve

Introduction

Grade generates geometrical restraints for novel small molecules. It does so by querying the Cambridge Structural Database (CSD) as a main source of restraint information, and then invoking quantum chemical procedures to obtain whatever information the CSD is unable to supply. Grade takes a SMILES string or a mol2 file containing 3D coordinates of all atoms as input. For compounds that are already in the PDB chemical components dictionary, Grade can produce restraints given the PDB 3-letter code. For novel large and complex systems the preparation and upload of a mol2 file is recommended. Please see preferred input formats help page for further information. Please note that there are limitations regarding which element types can be handled by Grade.

News:

- Updated to latest Grade release (16 Jul 2021) and CSD Data Update May21. *16 July 2021*
- Upgrade to Grade (29 Nov 2019, including multiple improvements and bug fixes), added electron-cloud option and updated CSD database. *11 December 2019*'
  Please note that the option to generate dictionaries without the use of the CSD has been removed.
- Maintenance: Latest versions of Grade (27 Nov 2018) installed and updated to CSD 2019. *14 December 2018*
- Maintenance: updated to CSD 2018. *3 March 2018*
- Maintenance: Latest versions of Grade (20 Sep 2017) installed and CSD database updated. When giving a PDB compound identifier for which no ideal coordinates are available, we now fall back to the experimental coordinates to avoid a failure for those compounds. *22 September 2017*
- Maintenance: Latest versions of Grade (7th Dec 2016) and Mogul (CSD 2017) installed. *17 February 2017*
- Maintenance: Latest versions of Grade and Mogul installed: details.. *29 March 2015*
- Bug fixes. *20 February 2014*
- Major upgrade to Grade Web Server with new features and improvements. *20 December 2013*

For further details please see Grade Web Server Help Page on the BUSTER wiki.

Run Grade

Name email:

Use of the Grade Web Server is subject to you agreeing to the conditions of use

Tick box to confirm that you have read and agree to the conditions of use.

Submit details Clear

Acknowledgements

Grade is developed at Global Phasing Ltd.

Big thanks to:

- CCDC for permission to use Mogul on the Grade Web Server
- Garib Murshudov and CCP4 for libcheck currently used to interpret SMILES strings
- Martin Field for the fdynamo implementation of the RM1 semiempirical QM method
- PyMOL and Open Babel
- Members of the Global Phasing Consortium and European Commission project SILVER (FP7-HEALTH-F3-2010-260644) for financial support
- BUSTER and grade users for suggestions, discussion, feedback and bug reports. Particular thanks to all the Grade web Server pre-release testers for their help!

Current usage statistics (updated every 10 min):

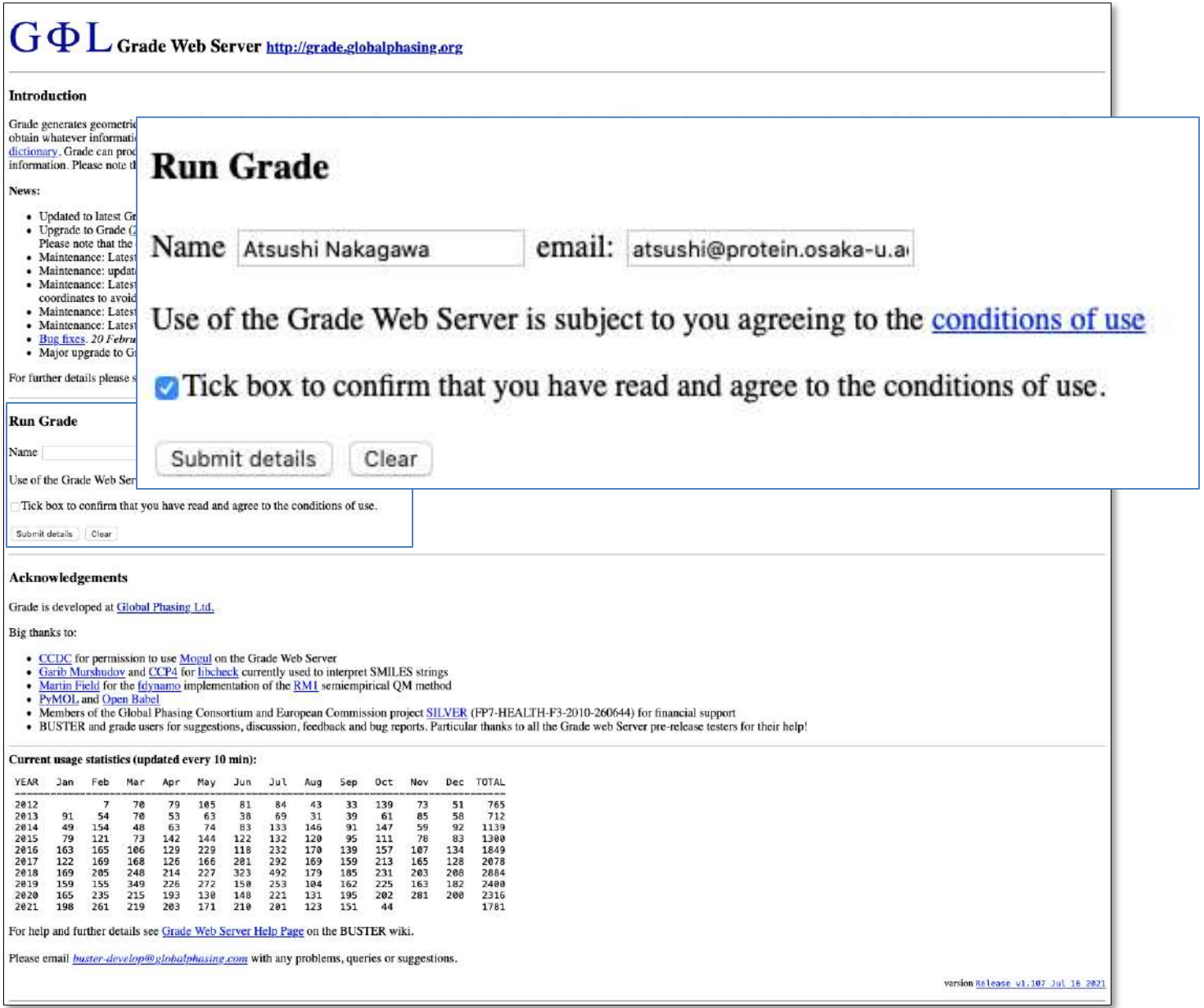
YEAR	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	TOTAL
2012		7	70	79	105	81	84	43	33	139	73	51	765
2013	91	54	70	53	63	38	69	31	39	61	85	58	712
2014	49	154	48	63	74	83	133	146	91	147	59	92	1139
2015	79	121	73	142	144	122	132	120	95	111	78	83	1300
2016	163	165	106	129	229	118	232	170	139	157	107	134	1849
2017	122	169	168	126	166	201	292	169	159	213	165	128	2078
2018	169	205	248	214	227	323	492	179	185	231	203	208	2884
2019	159	155	349	226	272	150	253	104	162	225	163	182	2400
2020	165	235	215	193	130	148	221	131	195	202	281	200	2316
2021	198	261	219	203	171	210	201	123	151	44			1781

For help and further details see Grade Web Server Help Page on the BUSTER wiki.

Please email buster-develop@globalphasing.com with any problems, queries or suggestions.

version Release v1.107 Jul 16 2021

Grade Web Server



GΦL Grade Web Server http://grade.globalphasing.org

Introduction

Grade generates geometri
obtain whatever informati
dictionary. Grade can proc
information. Please note t

News:

- Updated to latest Gr
- Upgrade to Grade (2
  Please note that the
- Maintenance: Latest
- Maintenance: updat
- Maintenance: Latest
  coordinates to avoid
- Maintenance: Latest
- Maintenance: Latest
- Bug fixes. *20 Febru*
- Major upgrade to G

For further details please s

Run Grade

Name

Use of the Grade Web Ser

Tick box to confirm that you have read and agree to the conditions of use.

Submit details Clear

Run Grade

Name Atsushi Nakagawa email: atsushi@protein.osaka-u.a

Use of the Grade Web Server is subject to you agreeing to the conditions of use

Tick box to confirm that you have read and agree to the conditions of use.

Submit details Clear

Acknowledgements

Grade is developed at Global Phasing Ltd.

Big thanks to:

- CCDC for permission to use Mogul on the Grade Web Server
- Garib Murshudov and CCP4 for libcheck currently used to interpret SMILES strings
- Martin Field for the fdynamo implementation of the RM1 semiempirical QM method
- PyMOL and Open Babel
- Members of the Global Phasing Consortium and European Commission project SILVER (FP7-HEALTH-F3-2010-260644) for financial support
- BUSTER and grade users for suggestions, discussion, feedback and bug reports. Particular thanks to all the Grade web Server pre-release testers for their help!

Current usage statistics (updated every 10 min):

YEAR	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	TOTAL
2012		7	70	79	105	81	84	43	33	139	73	51	765
2013	91	54	70	53	63	38	69	31	39	61	85	58	712
2014	49	154	48	63	74	83	133	146	91	147	59	92	1139
2015	79	121	73	142	144	122	132	120	95	111	78	83	1300
2016	163	165	106	129	229	118	232	170	139	157	107	134	1849
2017	122	169	168	126	166	201	292	169	159	213	165	128	2078
2018	169	205	248	214	227	323	492	179	185	231	203	208	2884
2019	159	155	349	226	272	150	253	104	162	225	163	182	2400
2020	165	235	215	193	130	148	221	131	195	202	281	200	2316
2021	198	261	219	203	171	210	201	123	151	44			1781

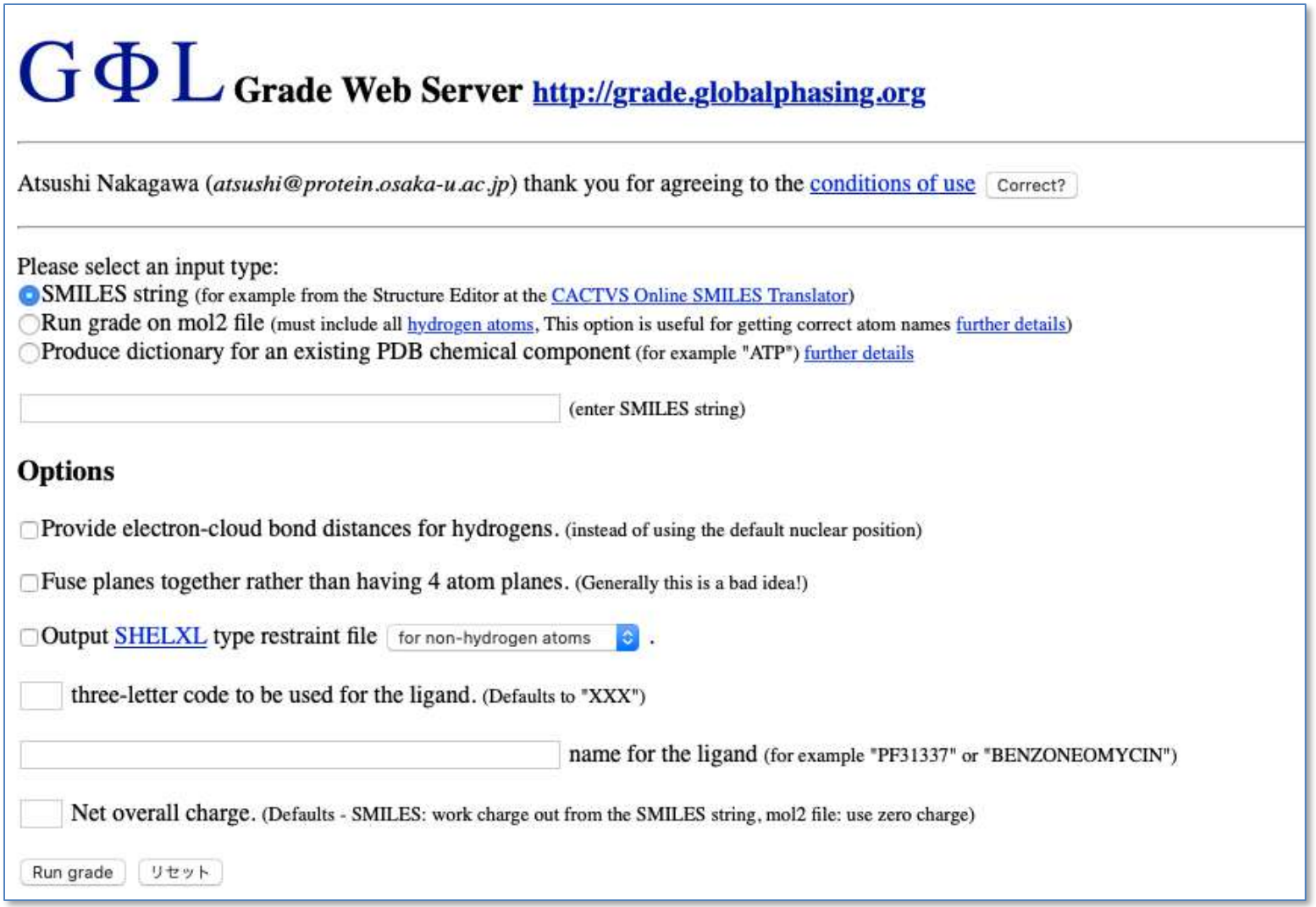
For help and further details see Grade Web Server Help Page on the BUSTER wiki.

Please email *buster-develop@globalphasing.com* with any problems, queries or suggestions.

version Release v1.107 Jul 16 2021

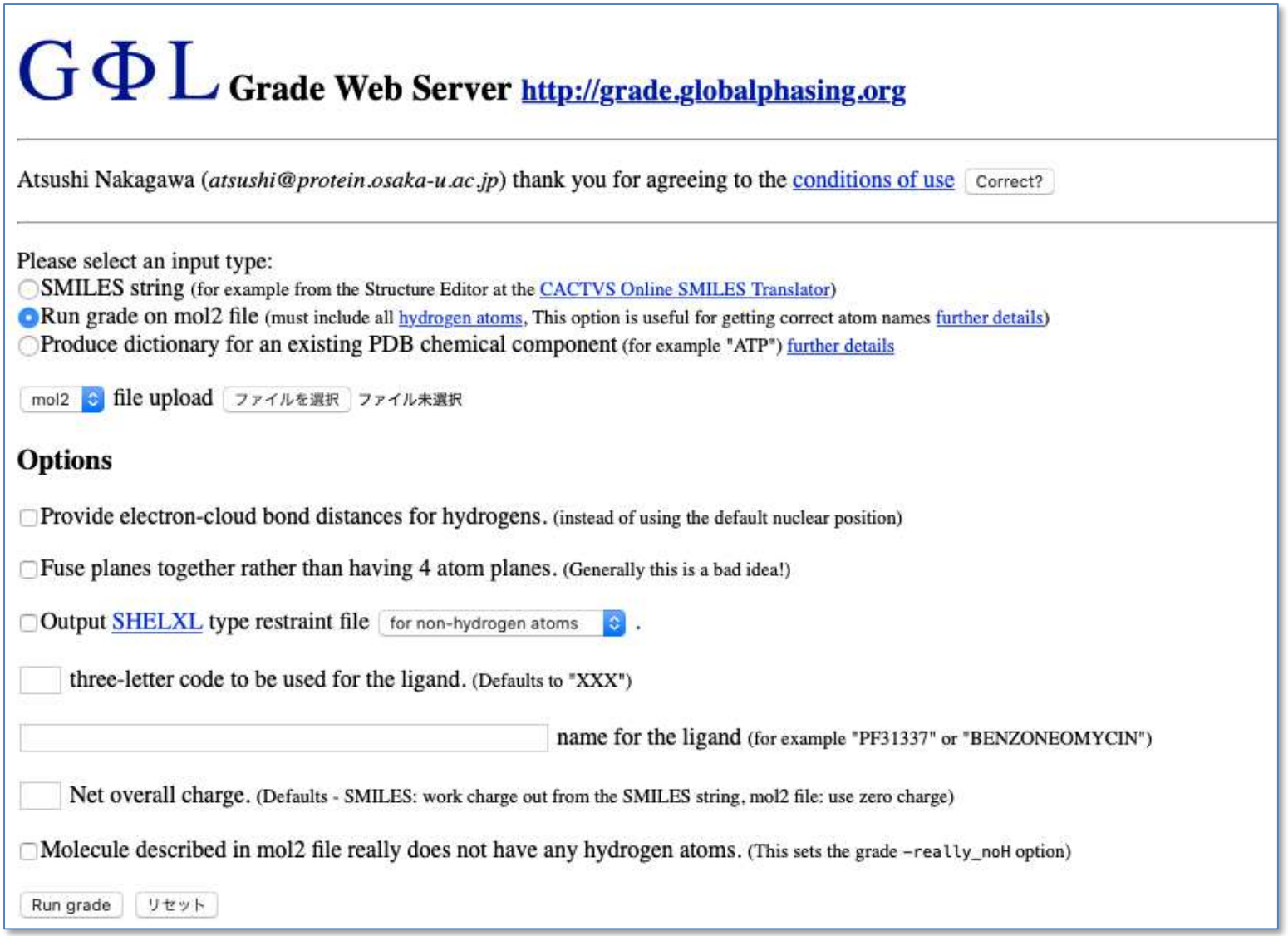
Grade Web Server

SMILES input

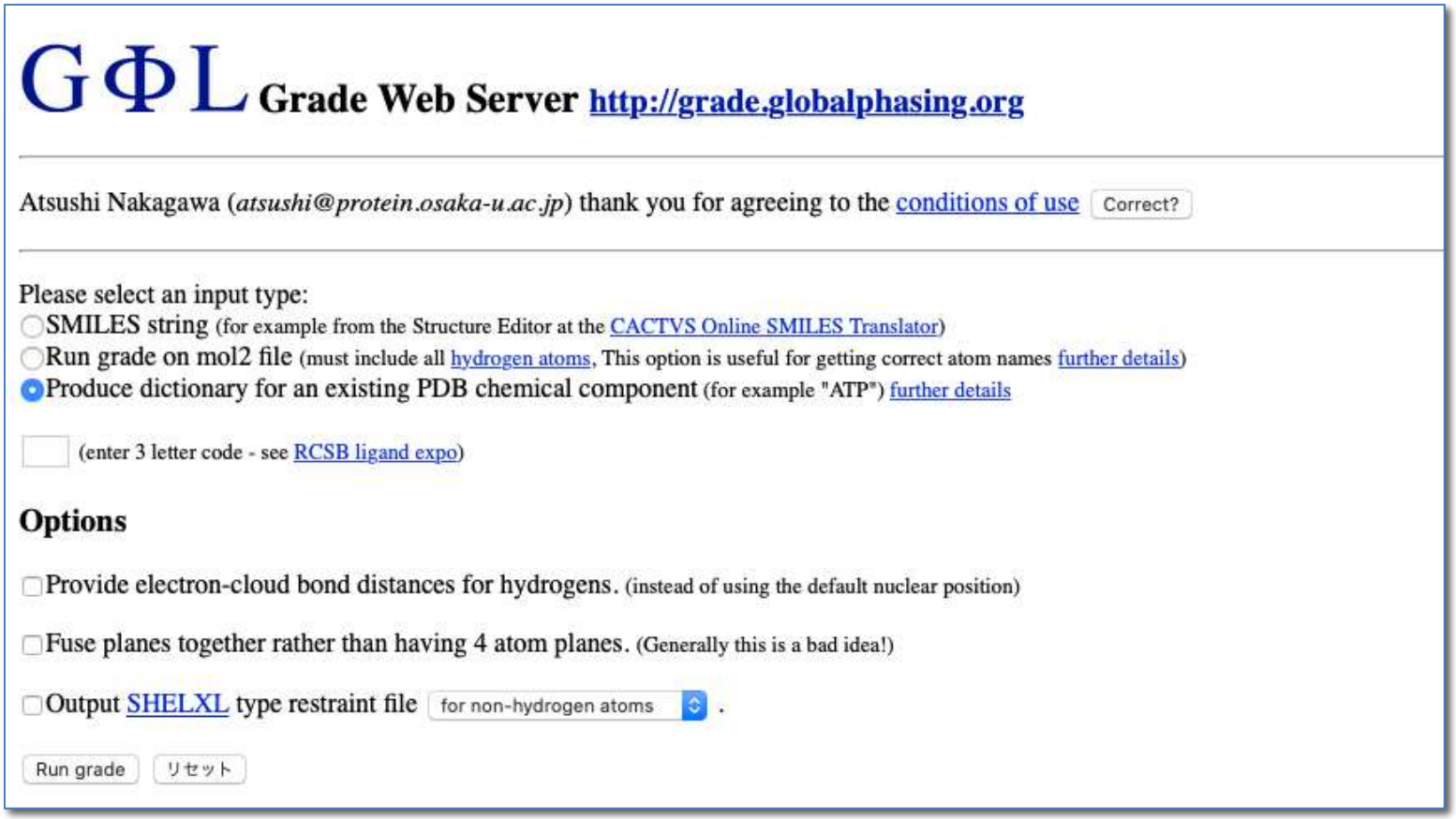


Grade Web Server

MOL2 input



Grade Web Server PDB input



Grade2

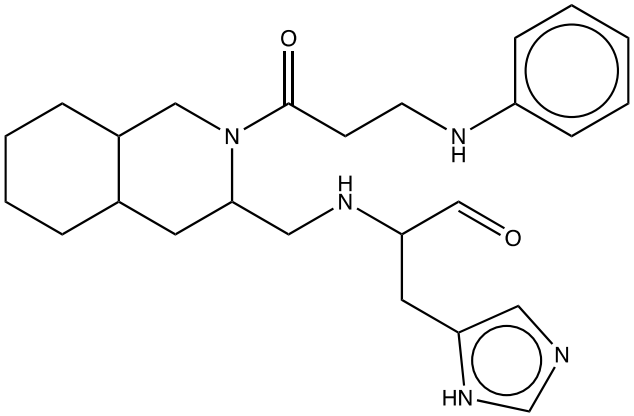
- The initial release of the new restraint generation tool, written by Oliver Smart.
- Grade2 is a clean-slate reimplementation using ideas from Grade, using the CSD Python API and the RDKit toolkit.
- Grade2 already has several advantages compared to Grade:
  - Grade2 makes better use of CSD information for rings and planes, using the CSD Python API to extend Mogul analysis.
  - Grade2 can handle a wider range of elements.
  - Grade2 is more robust. In a recent test of generating restraints-dictionaries for 2893 chemical components recently added to wwPDB, Grade2 produced dictionaries for all but 6 (0.2% failure) whereas Grade failed in 336 cases (12%).
  - Grade2 protonates or deprotonates groups commonly charged at neutral pH such as carboxylic acids, phosphates and alkylamines.

Grade2 Input

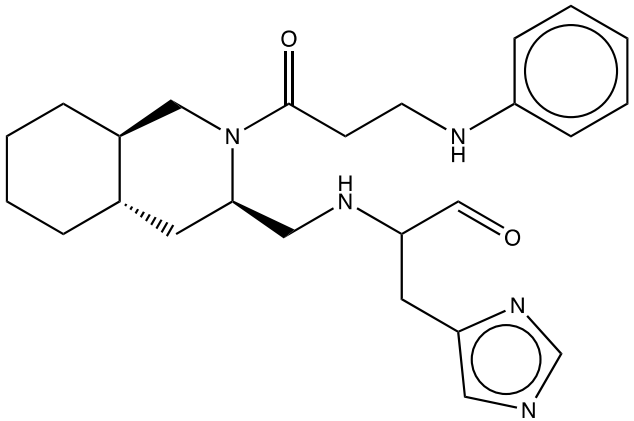
- PDB 3-letter code
  grade2 -P <pdb id>
- SMILES
  grade2 'Oc1ccccc1' --resname LIG
- mol/sdf
  grade2 --in Conformer3D_CID_135398744.sdf
- Tripos mol2
- Restraint dictionary CIF
- wwPDB CCD CIF
  - wwPDB Chemical Component Dictionary

SLH phenyl-beta-alanyl (R,S)-N-decalin type inhibitor

(2S)-3-(1H-imidazol-5-yl)-2-({[(3R,4aS,8aR)-2-(N-phenyl-beta-alanyl)decahydroisoquinolin-3-yl]methyl}amino)propanal



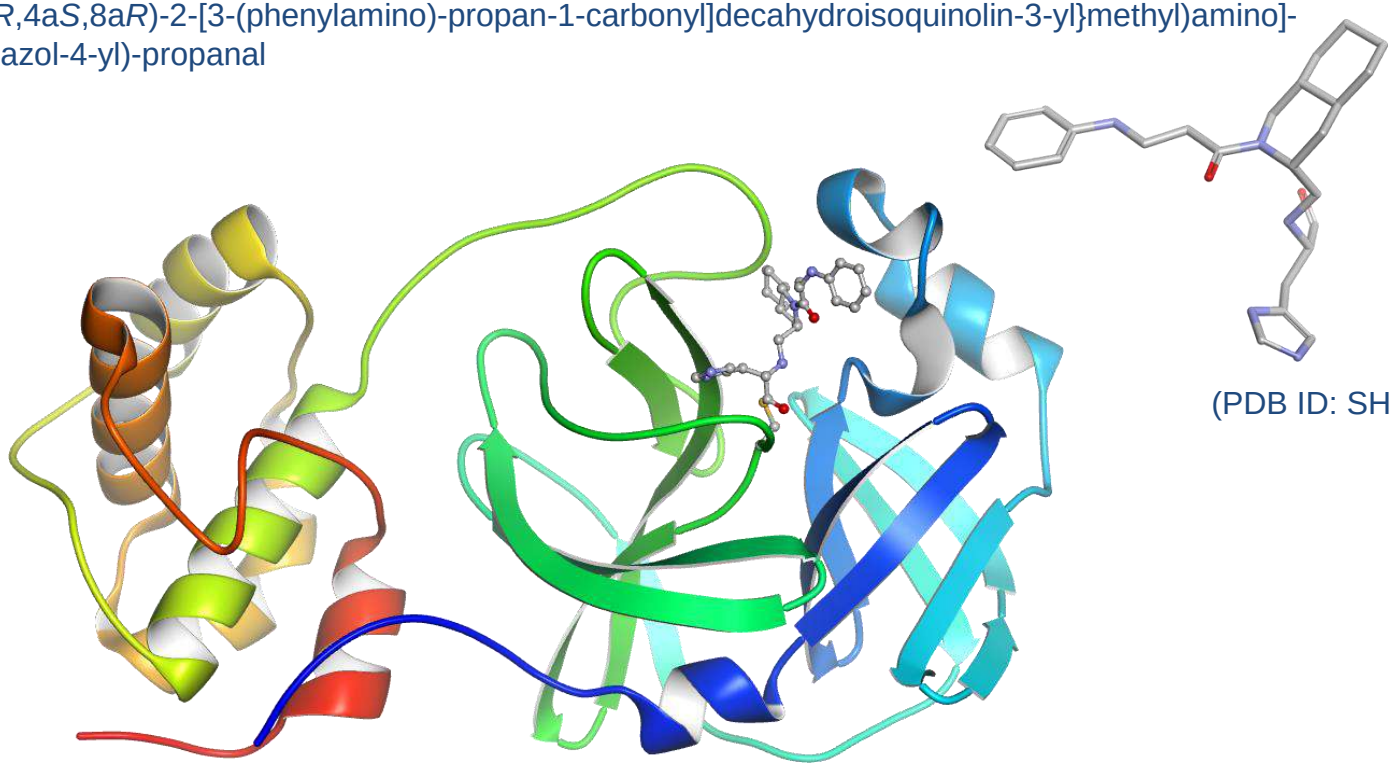
SMILES: 3C1C(CCCC1)CC(CNC(C=O)Cc2ncnc2)N3C(=O)CCNc4ccccc4



SMILES: O=C[C@H](Cc1[nH]cnc1)NC[C@H]2C[C@@H]3CCCC[C@H]3CN2C(=O)CCNc4ccccc4

Example 1:
SARS-3CL protease complex with a phenyl-beta-alanyl (*R*)-N-decalin type inhibitor (PDB ID: 5c5n)

(*S*)-2-[({(3*R*,4a*S*,8a*R*)-2-[3-(phenylamino)-propan-1-carbonyl]decahydroisoquinolin-3-yl}methyl)amino]-3-(1H-imidazol-4-yl)-propanal



(PDB ID: SHL)

Teruya, K. *et al.* Structural basis for the development of SARS 3CL protease inhibitors from a peptide mimic to an aza-decaline scaffold. *Biopolymers* **106**, 391-403, doi:10.1002/bip.22773 (2016).

Ligand Library by grade2

```
# GEN: Generated by grade2 version 1.0.0
# GEN: from PDB_chemical_components_definition_file ftp://ftp.ebi.ac.uk/pub/databases/msd/pdbechem_v2/S/SLH/SLH.cif
# GEN: using RDKit 2021.03.4
# GEN: ((DatabaseID PDB-SLH))
#
# BUSTER-KEYWORD TRUSTCOORD
# BUSTER-KEYWORD TRUSTTORS
#
data_comp_list
#
_chem_comp.id                       SLH
_chem_comp.three_letter_code        SLH
_chem_comp.name                     "(2S)-3-(1H-imidazol-5-yl)-2-({[(3R,4aS,8aR)-2-(N-phenyl-beta-
alanyl)decahydroisoquinolin-3-yl]methyl}amino)propanal (CHARGED)"
_chem_comp.group                    .
_chem_comp.number_atoms_all         68
_chem_comp.number_atoms_nh          32
_chem_comp.desc_level               .


#
data_comp_SLH
#
loop_
_chem_comp_atom.comp_id
_chem_comp_atom.atom_id
_chem_comp_atom.type_symbol
_chem_comp_atom.type_energy
_chem_comp_atom.partial_charge
_chem_comp_atom.charge
_chem_comp_atom.x
_chem_comp_atom.y
_chem_comp_atom.z
SLH   C1 C  CH2  0.024 0 -0.654 -1.761 -0.422
SLH   C2 C  CH1 -0.021 0 -1.592 -2.611  0.430
SLH   C3 C  CH2 -0.048 0 -1.998 -3.896 -0.292
SLH   N4 N  NH1 -0.385 0  4.527 -1.137 -0.603
SLH   C5 C  CH2 -0.053 0 -4.230 -3.881  0.870
SLH   C6 C  CH2 -0.050 0 -3.831 -2.604  1.604
SLH   C9 C  CH1  0.081 0 -1.375  0.300  0.791
```

Example 1: 5c5n SARS-3CL

complexed with a phenyl-beta-alanyl (*R*)-N-decalin type inhibitor

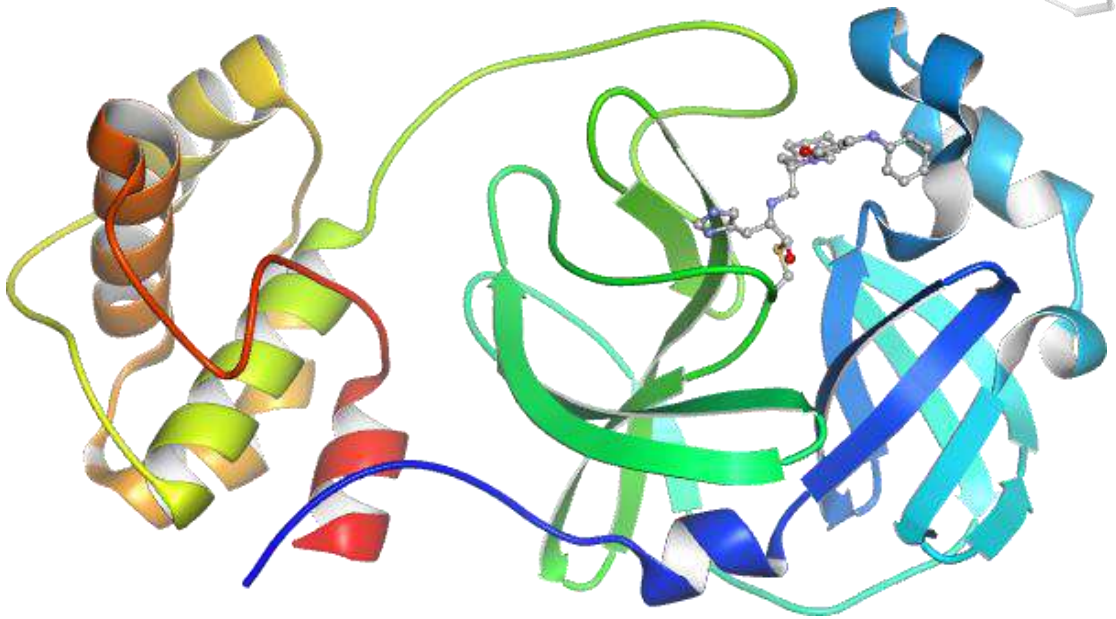
(1.69Å)

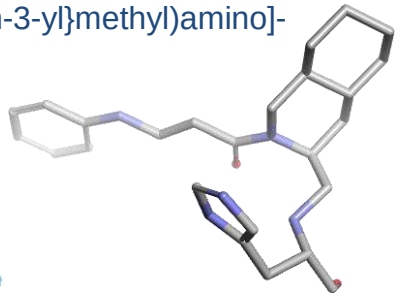
	CCP4	grade2
R	0.279517	0.279434
Rfree	0.304315	0.304152
Rxpct	0.254654	0.254535
Rxpctfree	0.282053	0.272112
BOND (Å)	0.0082	0.0082
ANGL (deg.)	1.03	1.02
TORS	15.63	15.46
SINTOR	15.63	15.46
PLAN (Å)	0.0180	0.0150
BCOR ($Å^2$)	0.879	0.879

Bricogne G., Blanc E., Brandl M., Flensburg C., Keller P., Paciorek W., Roversi P, Sharff A., Smart O.S., Vonrhein C., Womack T.O. (2017). BUSTER version 2.10.4. Cambridge, United Kingdom: Global Phasing Ltd.

Example 2:
SARS-3CL protease complex with a phenyl-beta-alanyl (*S*)-N-decalin type inhibitor (PDB ID: 5c5o)

(*S*)-2-[({(3*S*,4a*S*,8a*S*)-2-[3-(phenylamino)-propan-1-carbonyl]decahydroisoquinolin-3-yl}methyl)amino]-3-(1H-imidazol-4-yl)-propanal





(PDB ID: SDJ)

Teruya, K. *et al.* Structural basis for the development of SARS 3CL protease inhibitors from a peptide mimic to an aza-decaline scaffold. *Biopolymers* **106**, 391-403, doi:10.1002/bip.22773 (2016).

Example 2: 5c5o SARS-3CL

complexed with a phenyl-beta-alanyl (*S*)-N-decalin type inhibitor

(1.50Å)

	CCP4	grade2
R	0.232363	0.232599
Rfree	0.256229	0.255772
Rxpct	0.223368	0.223642
Rxpctfree	0.244443	0.244695
BOND (Å)	0.082	0.082
ANGL (deg.)	0.98	0.97
TORS	15.07	14.97
SINTOR	15.07	14.97
PLAN (Å)	0.0201	0.0168
BCOR ($Å^2$)	0.964	0.964

Bricogne G., Blanc E., Brandl M., Flensburg C., Keller P., Paciorek W., Roversi P, Sharff A., Smart O.S., Vonrhein C., Womack T.O. (2017). BUSTER version 2.10.4. Cambridge, United Kingdom: Global Phasing Ltd.

Please use CSD data to prepare Ligand Library

- ACEDRG … CCP4
  - COD: Crystallography Open Database
  - http://www.crystallography.net/cod/
- eLBOW … phenix
  - Mogul
- Grade, Grade2 … Global Phasing Inc.
  - Mogul
- Pyrogen … *Coot* (CCP4)
  - Mogul

Acknowledgements

- Thanks to all PDBj members and wwPDB consortium