wwPDB X-ray Structure Validation Summary Report (i)

Jan 12, 2022 – 09:25 am GMT

PDB ID	:	6ZUI
Title	:	Crystal structure of the Cys-Ser mutant of the cpYFP-based biosensor for hypochlorous acid
Authors	:	Tossounian, M.A.; Van Molle, I.; Messens, J.
Deposited on	:	2020-07-23
Resolution	:	2.20 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

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 (1)) were used in the production of this report:

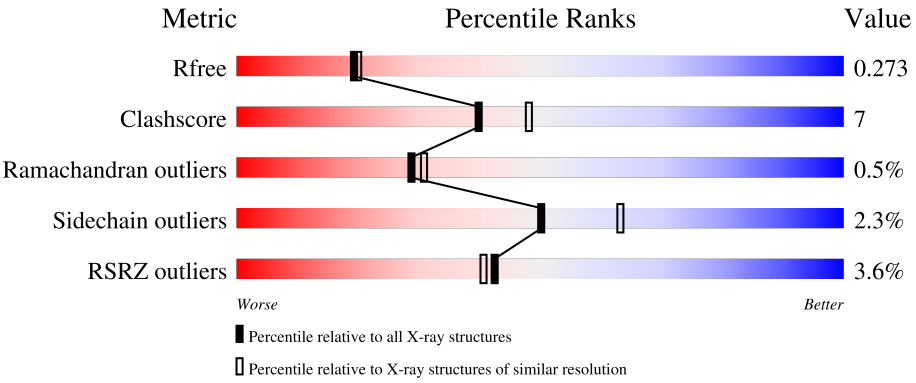
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.24
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.24

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.20 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	446	3% 80% 12% · 6%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 3503 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called HTH-type transcriptional repressor NemR,Green fluorescent protein,Green fluorescent protein,HTH-type transcriptional repressor NemR.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	421	3318	2100	573	635	10	1	1	0

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	21	SER	CYS	engineered mutation	UNP P67430
A	98	SER	CYS	engineered mutation	UNP P67430
A	103	SER	-	linker	UNP P67430
A	104	ALA	-	linker	UNP P67430
A	105	GLY	-	linker	UNP P67430
A	109	ASP	HIS	engineered mutation	UNP P42212
A	124	ALA	VAL	engineered mutation	UNP P42212
A	132	VAL	ILE	engineered mutation	UNP P42212
A	164	PHE	THR	engineered mutation	UNP P42212
A	167	VAL	ALA	engineered mutation	UNP P42212
A	192	LEU	HIS	engineered mutation	UNP P42212
A	199	ASN	-	linker	UNP P42212
A	200	VAL	-	linker	UNP P42212
A	201	ASP	-	linker	UNP P42212
A	202	GLY	-	linker	UNP P42212
A	203	GLY	-	linker	UNP P42212
A	204	SER	-	linker	UNP P42212
A	205	GLY	-	linker	UNP P42212
A	206	GLY	-	linker	UNP P42212
A	207	THR	-	linker	UNP P42212
A	208	GLY	-	linker	UNP P42212
A	253	LEU	PHE	engineered mutation	UNP P42212
A	271	LEU	PHE	engineered mutation	UNP P42212
A	273	CR2	SER	chromophore	UNP P42212
A	?	-	TYR	chromophore	UNP P42212
A	?	-	GLY	chromophore	UNP P42212

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	275	LEU	VAL	engineered mutation	UNP P42212
A	276	LYS	GLN	engineered mutation	UNP P42212
A	279	ALA	SER	engineered mutation	UNP P42212
A	336	GLY	ASP	engineered mutation	UNP P42212
A	352	GLY	-	linker	UNP P42212
A	353	THR	-	linker	UNP P42212
A	355	SER	CYS	engineered mutation	UNP P67430
A	365	SER	CYS	engineered mutation	UNP P67430
A	398	SER	CYS	engineered mutation	UNP P67430
A	402	SER	CYS	engineered mutation	UNP P67430

- Molecule 2 is water.

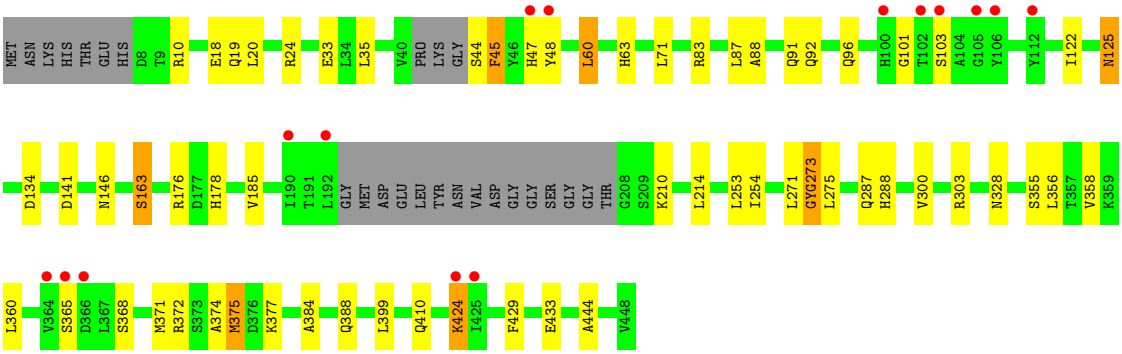
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	185	Total	O	0	0
			185	185		

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

• Molecule 1: HTH-type transcriptional repressor NemR,Green fluorescent protein,Green fluorescent protein,HTH-type transcriptional repressor NemR

Chain A:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	90.23Å 95.44Å 106.25Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.72 – 2.20 47.72 – 2.20	Depositor EDS
% Data completeness (in resolution range)	94.2 (47.72-2.20) 94.2 (47.72-2.20)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.78 (at 2.20Å)	Xtriage
Refinement program	PHENIX 1.11.1_2575	Depositor
R, R_{free}	0.199 , 0.273 0.199 , 0.273	Depositor DCC
R_{free} test set	1136 reflections (5.10%)	wwPDB-VP
Wilson B-factor (Å^2)	34.5	Xtriage
Anisotropy	0.913	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	3503	wwPDB-VP
Average B, all atoms (Å^2)	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.65% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: CR2

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\#\|Z\| > 5$	RMSZ	$\#\|Z\| > 5$
1	A	0.39	0/3366	0.55	0/4541

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	303	ARG	Peptide

5.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	3318	0	3257	47	0
2	A	185	0	0	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	3503	0	3257	47	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 7.

The worst 5 of 47 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:273:CR2:CA3	1:A:275:LEU:N	2.21	1.02
1:A:372:ARG:HH11	1:A:372:ARG:HB3	1.61	0.66
1:A:360:LEU:HB3	1:A:375:MET:HE1	1.78	0.64
1:A:273:CR2:HA31	1:A:275:LEU:N	2.13	0.64
1:A:24:ARG:NH1	1:A:33:GLU:OE2	2.21	0.63

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	413/446 (93%)	404 (98%)	7 (2%)	2 (0%)	29 31

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	365	SER
1	A	45	PHE

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar

resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	353/371 (95%)	344 (98%)	9 (2%)	47 60

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	424[A]	LYS
1	A	424[B]	LYS
1	A	141	ASP
1	A	163	SER
1	A	375	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	19	GLN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

1 non-standard protein/DNA/RNA residue is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
1	CR2	A	273	1	20,20,21	3.95	5 (25%)	25,27,29	4.42	11 (44%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral

centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
1	CR2	A	273	1	-	2/6/25/26	0/2/2/2

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	273	CR2	CB2-CA2	15.96	1.48	1.35
1	A	273	CR2	CA2-C2	-4.83	1.43	1.48
1	A	273	CR2	C2-N3	-2.97	1.32	1.39
1	A	273	CR2	O2-C2	2.57	1.28	1.23
1	A	273	CR2	C1-N2	2.54	1.37	1.32

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	273	CR2	O2-C2-CA2	-15.24	122.40	130.96
1	A	273	CR2	CA2-C2-N3	12.57	109.32	103.37
1	A	273	CR2	C2-N3-C1	-4.69	105.70	107.99
1	A	273	CR2	O3-C3-CA3	-3.53	115.73	126.39
1	A	273	CR2	C2-CA2-N2	-3.30	106.62	108.93

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	A	273	CR2	C3-CA3-N3-C2
1	A	273	CR2	C3-CA3-N3-C1

There are no ring outliers.

1 monomer is involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	A	273	CR2	4	0

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	A	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	273:CR2	C3	275:LEU	N	1.71

6 Fit of model and data

6.1 Protein, DNA and RNA chains

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95^{th} percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å^2)	Q<0.9
1	A	420/446 (94%)	0.36	15 (3%) 42 41	26, 41, 57, 72	16 (3%)

The worst 5 of 15 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	102	THR	5.2
1	A	47	HIS	4.2
1	A	106	TYR	4.0
1	A	192	LEU	3.4
1	A	100	HIS	3.1

6.2 Non-standard residues in protein, DNA, RNA chains

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
1	CR2	A	273	19/20	0.94	0.15	23,29,33,35	0

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.