wwPDB X-ray Structure Validation Summary Report (i)

May 26, 2020 – 11:56 am BST

PDB ID	:	4S1X
Title	:	Crystal structure of HA2-Del-L2seM, Central Coiled-Coil from Influenza Hemagglutinin HA2 without Heptad Repeat Stutter
Authors	:	Malashkevich, V.N.; Higgins, C.D.; Lai, J.R.; Almo, S.C.
Deposited on	:	2015-01-15
Resolution	:	1.90 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:

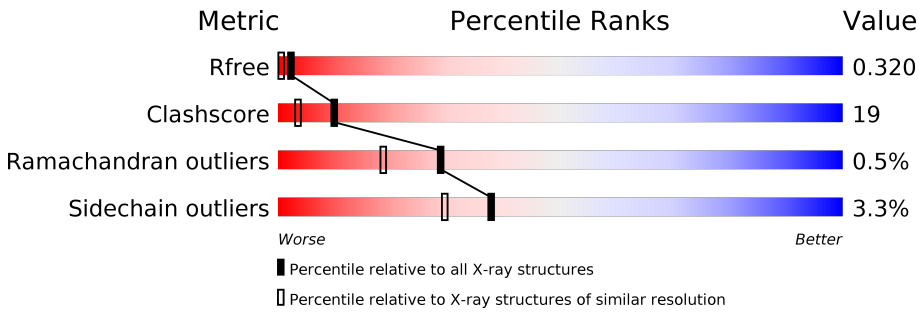
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 1.90 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	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)

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 on 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%

Mol	Chain	Length	Quality of chain
1	A	38	82% 18%
1	B	38	74% 24% .
1	C	38	18% 32% 8% 42%
1	D	38	76% 24%
1	E	38	87% 11% .
1	F	38	26% 24% 8% 42%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 1722 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 Truncated hemagglutinin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	Se			
1	A	38	324	203	56	63	2	0	0	0
1	B	37	316	197	55	62	2	0	1	0
1	C	22	190	121	34	34	1	0	0	0
1	D	38	324	203	56	63	2	0	0	0
1	E	37	318	197	57	62	2	0	1	0
1	F	22	190	121	34	34	1	0	0	0

There are 54 discrepancies between the modelled and reference sequences:

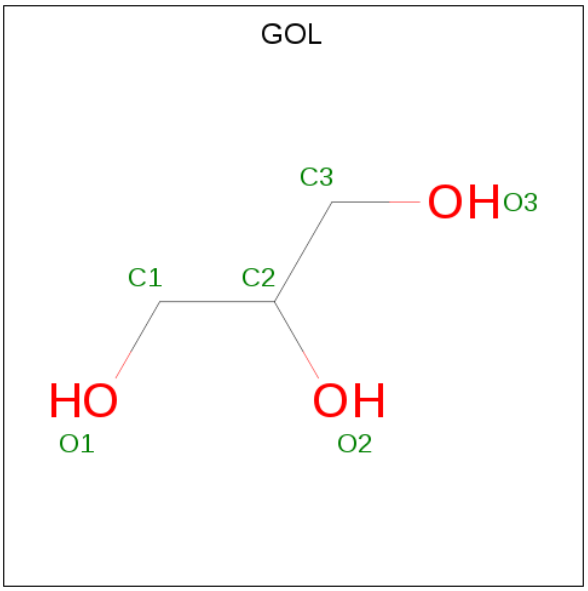
Chain	Residue	Modelled	Actual	Comment	Reference
A	42	TRP	-	EXPRESSION TAG	UNP A8TXJ0
A	43	GLY	-	EXPRESSION TAG	UNP A8TXJ0
A	44	SER	-	EXPRESSION TAG	UNP A8TXJ0
A	52	MSE	LEU	CONFLICT	UNP A8TXJ0
A	?	-	GLU	DELETION	UNP A8TXJ0
A	?	-	LYS	DELETION	UNP A8TXJ0
A	?	-	THR	DELETION	UNP A8TXJ0
A	?	-	ASN	DELETION	UNP A8TXJ0
A	80	MSE	LEU	CONFLICT	UNP A8TXJ0
B	42	TRP	-	EXPRESSION TAG	UNP A8TXJ0
B	43	GLY	-	EXPRESSION TAG	UNP A8TXJ0
B	44	SER	-	EXPRESSION TAG	UNP A8TXJ0
B	52	MSE	LEU	CONFLICT	UNP A8TXJ0
B	?	-	GLU	DELETION	UNP A8TXJ0
B	?	-	LYS	DELETION	UNP A8TXJ0
B	?	-	THR	DELETION	UNP A8TXJ0
B	?	-	ASN	DELETION	UNP A8TXJ0

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	80	MSE	LEU	CONFLICT	UNP A8TXJ0
C	42	TRP	-	EXPRESSION TAG	UNP A8TXJ0
C	43	GLY	-	EXPRESSION TAG	UNP A8TXJ0
C	44	SER	-	EXPRESSION TAG	UNP A8TXJ0
C	52	MSE	LEU	CONFLICT	UNP A8TXJ0
C	?	-	GLU	DELETION	UNP A8TXJ0
C	?	-	LYS	DELETION	UNP A8TXJ0
C	?	-	THR	DELETION	UNP A8TXJ0
C	?	-	ASN	DELETION	UNP A8TXJ0
C	80	MSE	LEU	CONFLICT	UNP A8TXJ0
D	42	TRP	-	EXPRESSION TAG	UNP A8TXJ0
D	43	GLY	-	EXPRESSION TAG	UNP A8TXJ0
D	44	SER	-	EXPRESSION TAG	UNP A8TXJ0
D	52	MSE	LEU	CONFLICT	UNP A8TXJ0
D	?	-	GLU	DELETION	UNP A8TXJ0
D	?	-	LYS	DELETION	UNP A8TXJ0
D	?	-	THR	DELETION	UNP A8TXJ0
D	?	-	ASN	DELETION	UNP A8TXJ0
D	80	MSE	LEU	CONFLICT	UNP A8TXJ0
E	42	TRP	-	EXPRESSION TAG	UNP A8TXJ0
E	43	GLY	-	EXPRESSION TAG	UNP A8TXJ0
E	44	SER	-	EXPRESSION TAG	UNP A8TXJ0
E	52	MSE	LEU	CONFLICT	UNP A8TXJ0
E	?	-	GLU	DELETION	UNP A8TXJ0
E	?	-	LYS	DELETION	UNP A8TXJ0
E	?	-	THR	DELETION	UNP A8TXJ0
E	?	-	ASN	DELETION	UNP A8TXJ0
E	80	MSE	LEU	CONFLICT	UNP A8TXJ0
F	42	TRP	-	EXPRESSION TAG	UNP A8TXJ0
F	43	GLY	-	EXPRESSION TAG	UNP A8TXJ0
F	44	SER	-	EXPRESSION TAG	UNP A8TXJ0
F	52	MSE	LEU	CONFLICT	UNP A8TXJ0
F	?	-	GLU	DELETION	UNP A8TXJ0
F	?	-	LYS	DELETION	UNP A8TXJ0
F	?	-	THR	DELETION	UNP A8TXJ0
F	?	-	ASN	DELETION	UNP A8TXJ0
F	80	MSE	LEU	CONFLICT	UNP A8TXJ0

- Molecule 2 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
2	E	1	6	3	3	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	14	14	14	0	0
			Total	O		
3	B	15	15	15	0	0
			Total	O		
3	C	2	2	2	0	0
			Total	O		
3	D	8	8	8	0	0
			Total	O		
3	E	12	12	12	0	0
			Total	O		
3	F	3	3	3	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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. 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. 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: Truncated hemagglutinin

Chain A:

- Molecule 1: Truncated hemagglutinin

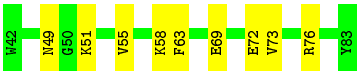
Chain B:

- Molecule 1: Truncated hemagglutinin

Chain C:

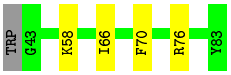
- Molecule 1: Truncated hemagglutinin

Chain D:



- Molecule 1: Truncated hemagglutinin

Chain E:



- Molecule 1: Truncated hemagglutinin

Chain F:

TRP
GLY
SER
I45
D46
Q47
I48
K51
M52
N53
R54
V55
I56
H64
Q65
I66
E67
F70
SER
GLU
VAL
GLU
GLY
ARG
ILE
GLN
ASP
MSE
GLU
LYS
TYR

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	34.45Å 98.52Å 34.55Å 90.00° 89.96° 90.00°	Depositor
Resolution (Å)	28.29 – 1.90 28.29 – 1.90	Depositor EDS
% Data completeness (in resolution range)	98.7 (28.29-1.90) 98.6 (28.29-1.90)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.46 (at 1.89Å)	Xtriage
Refinement program	REFMAC 5.8.0073	Depositor
R, R_{free}	0.208 , 0.245 0.280 , 0.320	Depositor DCC
R_{free} test set	874 reflections (4.84%)	wwPDB-VP
Wilson B-factor (Å^2)	39.6	Xtriage
Anisotropy	0.125	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 53.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.55$, $< L^2 > = 0.40$	Xtriage
Estimated twinning fraction	0.157 for -l,k,h 0.026 for h,-k,-l 0.011 for -l,-k,-h	Xtriage
Reported twinning fraction	0.374 for H, K, L 0.141 for -h,-k,l 0.098 for L, -K, H 0.388 for -L, K, H	Depositor
Outliers	0 of 18070 reflections	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	1722	wwPDB-VP
Average B, all atoms (Å^2)	41.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.59% 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 (i)

5.1 Standard geometry (i)

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

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	0.73	1/327 (0.3%)	0.57	0/429
1	B	0.64	1/319 (0.3%)	0.61	0/414
1	C	0.42	0/191	0.87	2/250 (0.8%)
1	D	0.78	1/327 (0.3%)	0.67	0/429
1	E	0.63	1/322 (0.3%)	0.62	0/420
1	F	0.51	0/191	0.65	0/250
All	All	0.65	4/1677 (0.2%)	0.65	2/2192 (0.1%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	58	LYS	C-N	10.60	1.58	1.34
1	A	58	LYS	C-N	10.54	1.58	1.34
1	E	58	LYS	C-N	6.85	1.49	1.34
1	B	58	LYS	C-N	6.81	1.49	1.34

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	58	LYS	O-C-N	-7.88	110.10	122.70
1	C	58	LYS	CA-C-N	5.03	128.27	117.20

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts (i)

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	324	0	311	6	0
1	B	316	0	314	5	0
1	C	190	0	191	31	0
1	D	324	0	311	8	0
1	E	318	0	314	5	0
1	F	190	0	190	20	0
2	E	6	0	8	0	0
3	A	14	0	0	0	0
3	B	15	0	0	1	0
3	C	2	0	0	0	0
3	D	8	0	0	0	0
3	E	12	0	0	1	0
3	F	3	0	0	0	0
All	All	1722	0	1639	63	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 19.

The worst 5 of 63 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:C:66:ILE:HG22	1:C:70:PHE:CE1	1.13	1.62
1:C:66:ILE:CG2	1:C:70:PHE:CE1	2.02	1.41
1:C:66:ILE:CG2	1:C:70:PHE:HE1	1.42	1.22
1:C:52:MSE:HE3	1:C:56:ILE:HD12	1.33	1.10
1:C:66:ILE:O	1:C:70:PHE:HD1	1.34	1.07

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	36/38 (95%)	36 (100%)	0	0	100	100
1	B	36/38 (95%)	35 (97%)	1 (3%)	0	100	100
1	C	20/38 (53%)	19 (95%)	1 (5%)	0	100	100
1	D	36/38 (95%)	35 (97%)	1 (3%)	0	100	100
1	E	36/38 (95%)	36 (100%)	0	0	100	100
1	F	20/38 (53%)	17 (85%)	2 (10%)	1 (5%)	2	0
All	All	184/228 (81%)	178 (97%)	5 (3%)	1 (0%)	29	18

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	F	67	GLU

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	35/33 (106%)	35 (100%)	0	100	100
1	B	35/33 (106%)	33 (94%)	2 (6%)	20	11
1	C	21/33 (64%)	19 (90%)	2 (10%)	8	3
1	D	35/33 (106%)	35 (100%)	0	100	100
1	E	35/33 (106%)	35 (100%)	0	100	100
1	F	21/33 (64%)	19 (90%)	2 (10%)	8	3
All	All	182/198 (92%)	176 (97%)	6 (3%)	38	29

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	52	MSE
1	F	56	ILE
1	C	68	LYS
1	B	71	SER

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	F	52	MSE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

1 ligand 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	$\#\|Z\| > 2$
2	GOL	E	101	-	5,5,5	0.25	0	5,5,5	0.34	0

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
2	GOL	E	101	-	-	1/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	E	101	GOL	C1-C2-C3-O3

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Fit of model and data (i)

6.1 Protein, DNA and RNA chains (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers (i)

Unable to reproduce the depositors R factor - this section is therefore empty.