



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 6, 2020 – 12:42 PM BST

PDB ID : 4NUF
Title : Crystal Structure of SHP/EID1
Authors : Zhi, X.; Zhou, X.E.; He, Y.; Zechner, C.; Suino-Powell, K.M.; Kliewer, S.A.;
Melcher, K.; Mangelsdorf, D.J.; Xu, H.E.
Deposited on : 2013-12-03
Resolution : 2.80 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 ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) 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 : 2.13.1
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) : 2.13.1

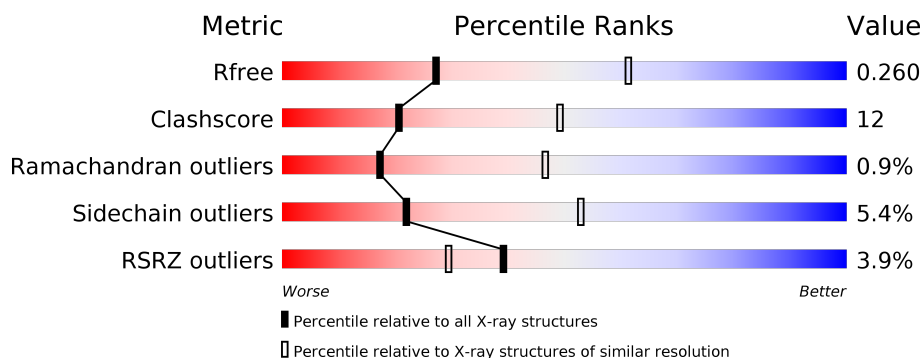
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.80 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	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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 $\leq 5\%$. 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	580	<div> <div>4%</div> <div> <div></div> <div>69%</div> <div>22%</div> <div>• 5%</div> </div> </div>
2	P	16	<div> <div>6%</div> <div> <div></div> <div>63%</div> <div>38%</div> </div> </div>
3	B	2	<div> <div></div> <div> <div>50%</div> <div>50%</div> </div> </div>

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 4493 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 Maltose ABC transporter periplasmic protein, Nuclear receptor subfamily 0 group B member 2 chimeric construct.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	549	Total	C	N	O	S	0	0	0
			4295	2769	712	798	16			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP K0BGG6
A	369	ASN	-	linker	UNP K0BGG6
A	370	ALA	-	linker	UNP K0BGG6
A	371	ALA	-	linker	UNP K0BGG6
A	372	ALA	-	linker	UNP K0BGG6
A	1053	GLU	-	linker	UNP K0BGG6
A	1054	PHE	-	linker	UNP K0BGG6
A	1085	ASP	GLU	engineered mutation	UNP Q62227
A	1126	THR	LEU	engineered mutation	UNP Q62227
A	1127	THR	GLU	engineered mutation	UNP Q62227
A	1128	ARG	GLU	engineered mutation	UNP Q62227
A	1228	ARG	LYS	engineered mutation	UNP Q62227

- Molecule 2 is a protein called EID1 peptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	P	16	Total	C	N	O	S	0	0	0
			126	80	23	22	1			

- Molecule 3 is an oligosaccharide called alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose.



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace
3	B	2	Total	C	O	0	0	0
			23	12	11			

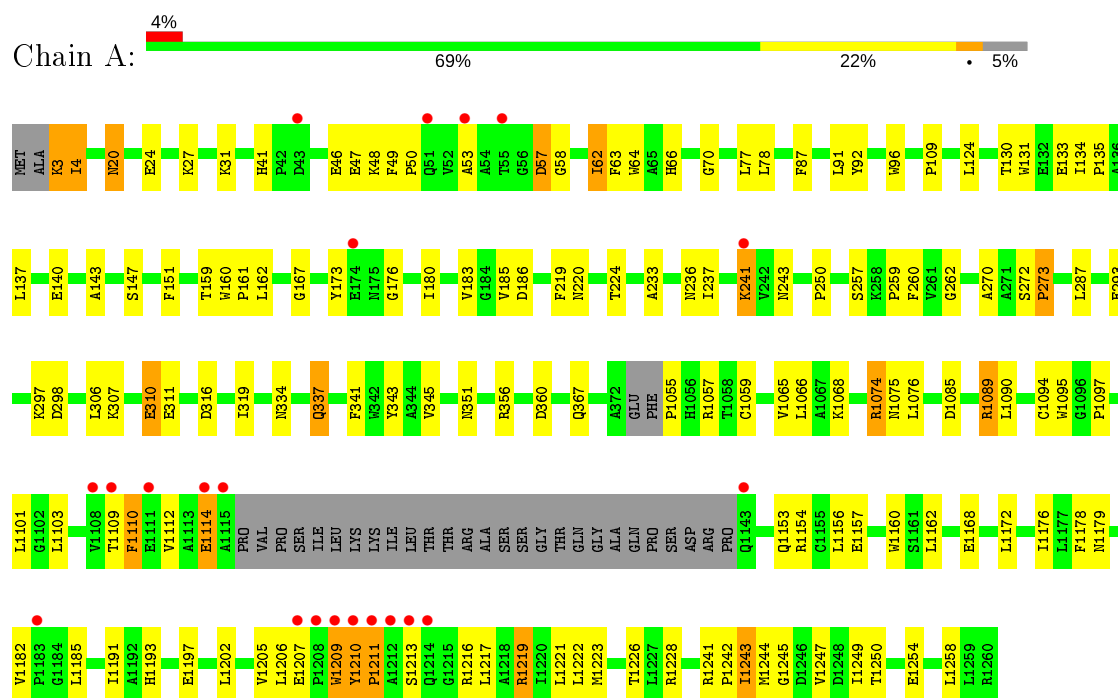
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	48	Total	O	0	0
			48	48		
4	P	1	Total	O	0	0
			1	1		

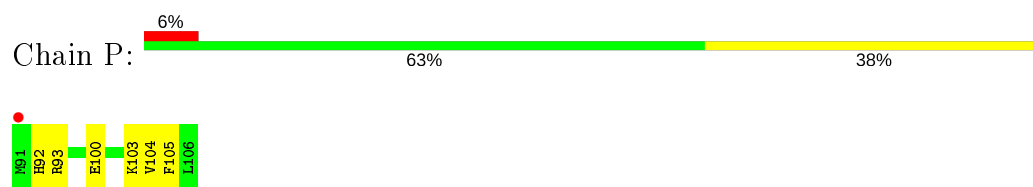
3 Residue-property plots

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: Maltose ABC transporter periplasmic protein, Nuclear receptor subfamily 0 group B member 2 chimeric construct



- Molecule 2: EID1 peptide



- Molecule 3: alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	56.39Å 105.15Å 136.28Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 2.80 29.77 – 2.80	Depositor EDS
% Data completeness (in resolution range)	83.3 (30.00-2.80) 94.3 (29.77-2.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.09 (at 2.80Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.196 , 0.245 0.213 , 0.260	Depositor DCC
R_{free} test set	1410 reflections (6.82%)	wwPDB-VP
Wilson B-factor (Å ²)	33.5	Xtriage
Anisotropy	0.593	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 44.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	4493	wwPDB-VP
Average B, all atoms (Å ²)	33.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.84% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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: GLC

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	0.45	0/4401	0.55	0/5982
2	P	0.48	0/127	0.53	0/169
All	All	0.45	0/4528	0.55	0/6151

There are no bond length outliers.

There are no bond angle outliers.

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	4295	0	4268	106	0
2	P	126	0	128	5	0
3	B	23	0	21	0	0
4	A	48	0	0	4	0
4	P	1	0	0	0	0
All	All	4493	0	4417	109	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 12.

The worst 5 of 109 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:1101:LEU:HD22	1:A:1176:ILE:HD12	1.65	0.77
1:A:1210:TYR:HB2	1:A:1211:PRO:HD3	1.68	0.75
1:A:124:LEU:HD11	1:A:137:LEU:HD21	1.70	0.73
1:A:337:GLN:H	1:A:337:GLN:NE2	1.88	0.70
1:A:186:ASP:OD1	1:A:367:GLN:HG3	1.92	0.70

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	543/580 (94%)	505 (93%)	33 (6%)	5 (1%)	17	46
2	P	14/16 (88%)	13 (93%)	1 (7%)	0	100	100
All	All	557/596 (94%)	518 (93%)	34 (6%)	5 (1%)	17	46

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	1210	TYR
1	A	1211	PRO
1	A	167	GLY
1	A	4	ILE
1	A	1112	VAL

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	448/474 (94%)	423 (94%)	25 (6%)	21	51
2	P	13/13 (100%)	13 (100%)	0	100	100
All	All	461/487 (95%)	436 (95%)	25 (5%)	22	53

5 of 25 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	360	ASP
1	A	1074	ARG
1	A	1221	LEU
1	A	1066	LEU
1	A	1089	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	220	ASN
1	A	243	ASN
1	A	369	ASN
1	A	203	ASN
1	A	357	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates ⓘ

2 monosaccharides are 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	$\# Z > 2$
3	GLC	B	1	3	12,12,12	0.86	0	17,17,17	0.50	0
3	GLC	B	2	3	11,11,12	1.17	1 (9%)	15,15,17	0.92	1 (6%)

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
3	GLC	B	1	3	-	0/2/22/22	0/1/1/1
3	GLC	B	2	3	-	0/2/19/22	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	2	GLC	C2-C3	3.01	1.57	1.52

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	2	GLC	C1-O5-C5	2.20	115.17	112.19

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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 ⓘ

There are no chain breaks in this entry.

6 Fit of model and data [i](#)

6.1 Protein, DNA and RNA chains [i](#)

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, 95th 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	549/580 (94%)	-0.15	21 (3%) 40 30	15, 30, 57, 107	0
2	P	16/16 (100%)	-0.16	1 (6%) 20 12	19, 28, 57, 67	0
All	All	565/596 (94%)	-0.15	22 (3%) 39 29	15, 30, 57, 107	0

The worst 5 of 22 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1210	TYR	6.7
1	A	1209	TRP	5.3
1	A	1115	ALA	4.7
1	A	1211	PRO	4.5
1	A	241	LYS	3.8

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

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

6.3 Carbohydrates [i](#)

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, 95th 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
3	GLC	B	2	11/12	0.96	0.19	12,18,20,20	0
3	GLC	B	1	12/12	0.96	0.21	22,23,28,29	0

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.