WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Jun 25, 2024 – 11:59 AM EDT

|		
PDB ID	:	5ZQ5
Title	:	SidE-Ubi
Authors	:	Wang, Y.; Gao, A.; Gao, P.
Deposited on	:	2018-04-17
Resolution	:	2.49 Å(reported)

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION
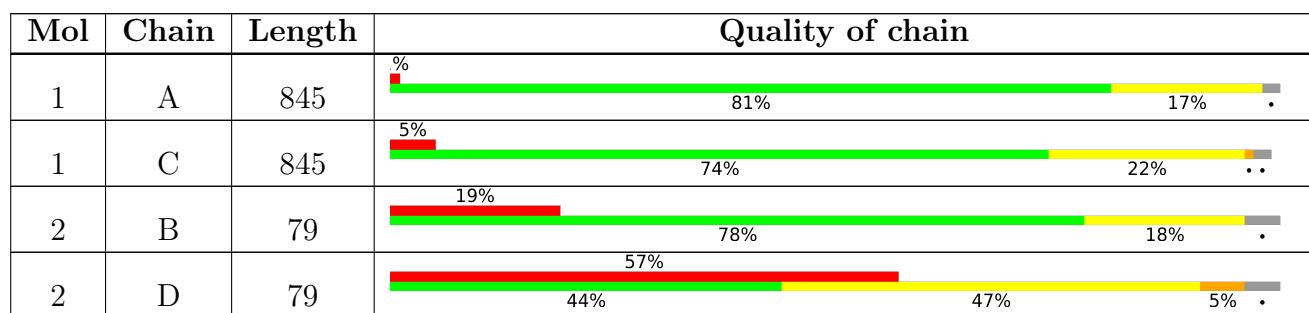
The reported resolution of this entry is 2.49 Å.

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	Percentile Ranks	Value
Rfree		0.234
Clashscore		12
Ramachandran outliers		0
Sidechain outliers		3.1%
RSRZ outliers		6.4%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	5857 (2.50-2.46)
Clashscore	141614	6594 (2.50-2.46)
Ramachandran outliers	138981	6469 (2.50-2.46)
Sidechain outliers	138945	6471 (2.50-2.46)
RSRZ outliers	127900	5738 (2.50-2.46)

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	845	% 81% 17% •
1	C	845	5% 74% 22% ••
2	B	79	19% 78% 18% •
2	D	79	57% 44% 47% 5% •

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 14624 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 SidE.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	828	Total 6606	C 4164	N 1153	O 1267	S 6	Se 16	0	0	0
1	C	824	Total 6578	C 4149	N 1149	O 1258	S 6	Se 16	0	0	0

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	221	MSE	-	initiating methionine	UNP Q6RCR1
A	1058	LEU	-	expression tag	UNP Q6RCR1
A	1059	GLU	-	expression tag	UNP Q6RCR1
A	1060	HIS	-	expression tag	UNP Q6RCR1
A	1061	HIS	-	expression tag	UNP Q6RCR1
A	1062	HIS	-	expression tag	UNP Q6RCR1
A	1063	HIS	-	expression tag	UNP Q6RCR1
A	1064	HIS	-	expression tag	UNP Q6RCR1
A	1065	HIS	-	expression tag	UNP Q6RCR1
C	221	MSE	-	initiating methionine	UNP Q6RCR1
C	1058	LEU	-	expression tag	UNP Q6RCR1
C	1059	GLU	-	expression tag	UNP Q6RCR1
C	1060	HIS	-	expression tag	UNP Q6RCR1
C	1061	HIS	-	expression tag	UNP Q6RCR1
C	1062	HIS	-	expression tag	UNP Q6RCR1
C	1063	HIS	-	expression tag	UNP Q6RCR1
C	1064	HIS	-	expression tag	UNP Q6RCR1
C	1065	HIS	-	expression tag	UNP Q6RCR1

- Molecule 2 is a protein called Ubiquitin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	76	Total 595	C 375	N 102	O 117	S 1	0	0	0

Continued on next page...

WORLDWIDE wwPDB PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	D	76	595	375	102	117	1	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-2	GLY	-	expression tag	UNP P0CG48
B	-1	GLY	-	expression tag	UNP P0CG48
B	0	SER	-	expression tag	UNP P0CG48
B	42	ALA	ARG	engineered mutation	UNP P0CG48
D	-2	GLY	-	expression tag	UNP P0CG48
D	-1	GLY	-	expression tag	UNP P0CG48
D	0	SER	-	expression tag	UNP P0CG48
D	42	ALA	ARG	engineered mutation	UNP P0CG48
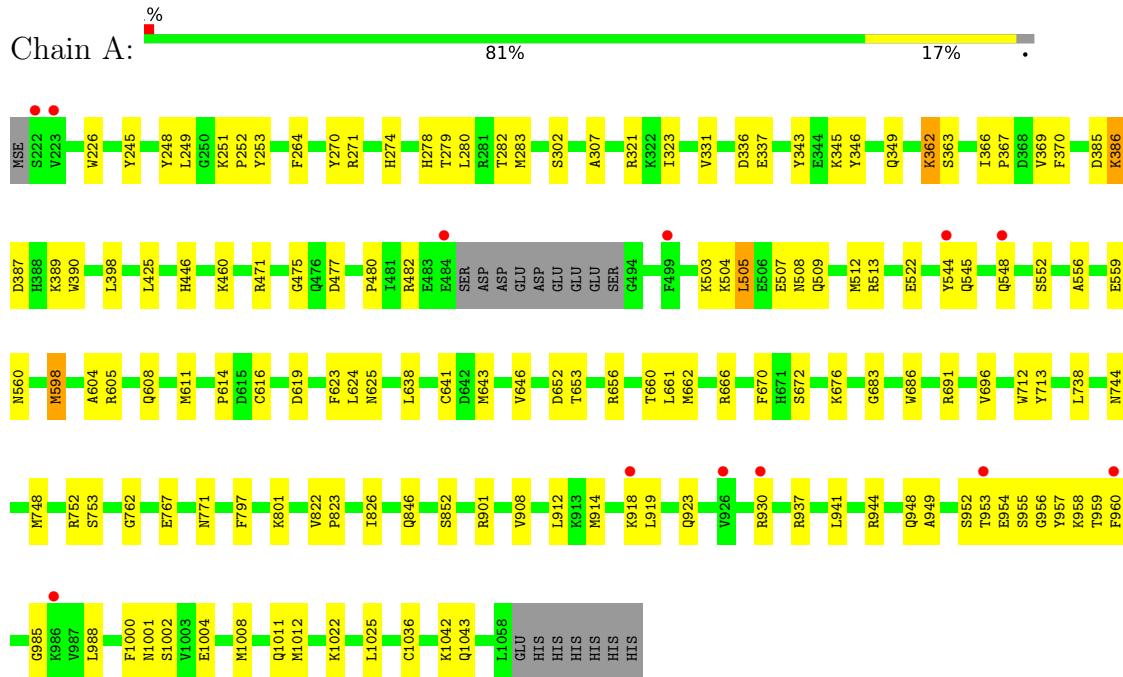
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	138	Total 138	O 138	0	0
3	B	3	Total 3	O 3	0	0
3	D	2	Total 2	O 2	0	0
3	C	107	Total 107	O 107	0	0
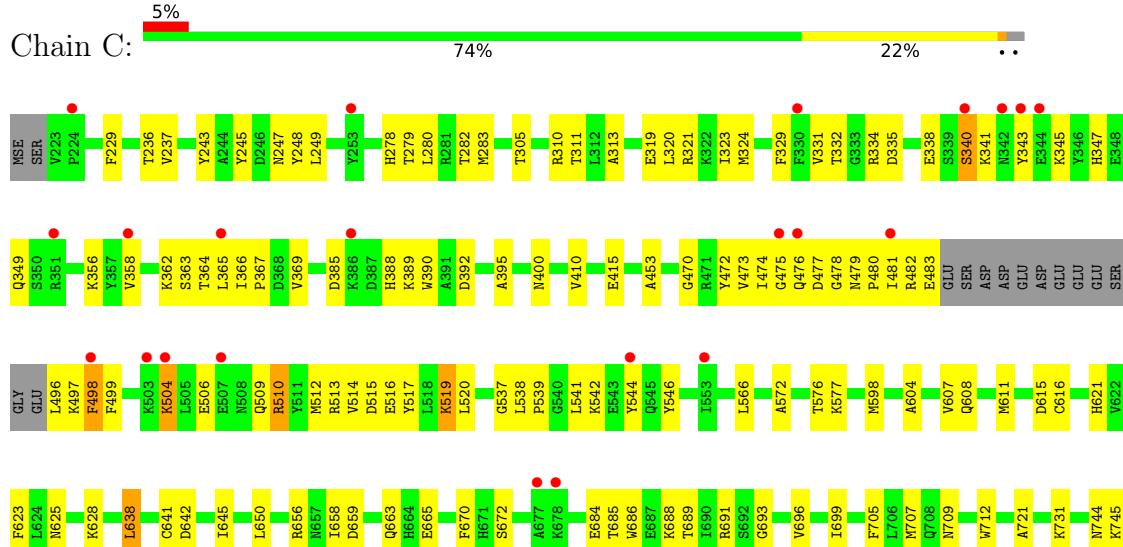
3 Residue-property plots ⓘ
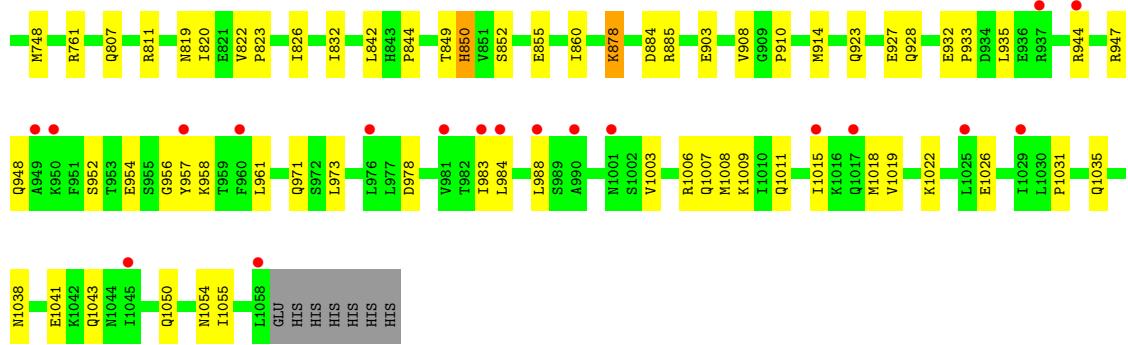
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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.
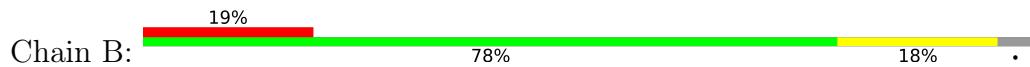
- Molecule 1: SidE

- Molecule 1: SidE

M748 R761 Q807 R811 N819 I820 E821 V822 P823 I826 I832 L842 H843 P844 T849 H850 V851 S852 E855 I860 K878 D884 R885 E903 V908 G909 P910 M914 Q923 E927 Q928 E932 P933 D934 L935 E936 R937 R944 R947

Q948 A949 K950 F951 S952 T953 E954 S955 G956 Y957 K958 T959 F960 L961 Q971 S972 L973 L976 L977 D978 V981 T982 I983 L984 L988 S989 A990 N1001 S1002 V1003 R1006 Q1007 M1008 K1009 I1010 Q1011 I1015 K1016 Q1017 M1018 V1019 K1022 L1025 E1026 I1029 L1030 P1031 Q1035

N1038 E1041 K1042 Q1043 N1044 I1045 Q1050 N1054 I1055 L1058 GLU HIS HIS HIS HIS HIS HIS

• Molecule 2: Ubiquitin

Chain B:

19%

78% 18% •

GLY GLY SER M1 Q2 I3 F4 V5 L8 T9 G10 K11 T12 I13 T14 L15 E16 V17 E18 T22 I23 E24 N25 I30 K33 P37 P38 D39 L43 D52 Y59 Q62 V70 G76

• Molecule 2: Ubiquitin

Chain D:

57%

44% 47% 5% •

GLY GLY SER M1 Q2 I3 F4 V5 K6 T7 L8 T9 G10 K11 T12 I13 T14 L15 E16 V17 E18 P19 S20 D21 T22 I23 E24 N25 V26 K27 A28 K29 I30 Q31 D32 K33 E34 G35 I36 P37 Q40 Q41 A42 L43 I44 F45 Q49 L50 E51 D52 G53 R54 T55 L56 S57 D58 Y59 N60

I61 Q62 K63 E64 S65 T66 L67 V70 L71 R72 L73 R74 G75 G76

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	92.67Å 129.08Å 191.51Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.69 – 2.49 48.69 – 2.49	Depositor EDS
% Data completeness (in resolution range)	99.5 (48.69-2.49) 99.5 (48.69-2.49)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.28 (at 2.48Å)	Xtriage
Refinement program	PHENIX 1.11.1_2575	Depositor
R, R_{free}	0.190 , 0.233 0.194 , 0.234	Depositor DCC
R_{free} test set	4061 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	46.1	Xtriage
Anisotropy	0.334	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 49.4	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	14624	wwPDB-VP
Average B, all atoms (Å^2)	62.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.48% 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 ⓘ

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	0.41	0/6720	0.50	1/9053 (0.0%)
1	C	0.46	0/6692	0.52	0/9016
2	B	0.38	0/601	0.52	0/809
2	D	0.53	0/601	0.70	0/809
All	All	0.44	0/14614	0.52	1/19687 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	598	MSE	N-CA-CB	-6.05	99.70	110.60

There are no chirality outliers.

There are no planarity outliers.

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	6606	0	6542	106	0
1	C	6578	0	6522	181	0
2	B	595	0	621	9	0
2	D	595	0	621	47	0
3	A	138	0	0	6	0
3	B	3	0	0	0	0
3	C	107	0	0	4	0
3	D	2	0	0	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	14624	0	14306	334	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 334 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:474:ILE:HG21	1:C:510:ARG:NH2	1.35	1.39
1:C:474:ILE:CG2	1:C:510:ARG:HG3	1.66	1.25
1:C:482:ARG:HB2	1:C:496:LEU:HA	1.18	1.18
2:D:5:VAL:HG12	2:D:67:LEU:HB2	1.24	1.17
1:C:482:ARG:CB	1:C:496:LEU:HA	1.81	1.09

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	824/845 (98%)	804 (98%)	20 (2%)	0	100	100
1	C	820/845 (97%)	806 (98%)	14 (2%)	0	100	100
2	B	74/79 (94%)	73 (99%)	1 (1%)	0	100	100
2	D	74/79 (94%)	71 (96%)	3 (4%)	0	100	100
All	All	1792/1848 (97%)	1754 (98%)	38 (2%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains ⓘ

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	719/719 (100%)	697 (97%)	22 (3%)	40	64
1	C	716/719 (100%)	694 (97%)	22 (3%)	40	64
2	B	67/68 (98%)	66 (98%)	1 (2%)	65	83
2	D	67/68 (98%)	63 (94%)	4 (6%)	19	34
All	All	1569/1574 (100%)	1520 (97%)	49 (3%)	40	64

5 of 49 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	341	LYS
1	C	506	GLU
1	C	363	SER
1	C	392	ASP
1	C	519	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	843	HIS
1	C	657	ASN
2	D	25	ASN
1	A	1011	GLN
1	C	388	HIS

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 ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

5.7 Other polymers ⓘ

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 ⓘ

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, 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	812/845 (96%)	0.15	12 (1%)	73	75	26, 48, 93, 122	0
1	C	808/845 (95%)	0.44	41 (5%)	28	29	28, 63, 105, 141	0
2	B	76/79 (96%)	1.22	15 (19%)	1	0	46, 71, 93, 103	0
2	D	76/79 (96%)	2.79	45 (59%)	0	0	62, 121, 171, 178	0
All	All	1772/1848 (95%)	0.44	113 (6%)	19	19	26, 58, 109, 178	0

The worst 5 of 113 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	59	TYR	8.8
2	D	36	ILE	8.0
2	D	53	GLY	7.8
2	D	20	SER	7.8
1	C	475	GLY	6.5

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

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

6.3   Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4   Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.