wwPDB X-ray Structure Validation Summary Report (i)

Oct 23, 2023 – 10:12 pm BST

PDB ID	:	8BE5
Title	:	Crystal structure of SOS1-KRasG12V-Nanobody22-Nanobody75
Authors	:	Fischer, B.; Wohlkonig, A.; Steyaert, J.
Deposited on	:	2022-10-21
Resolution	:	3.13 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 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:

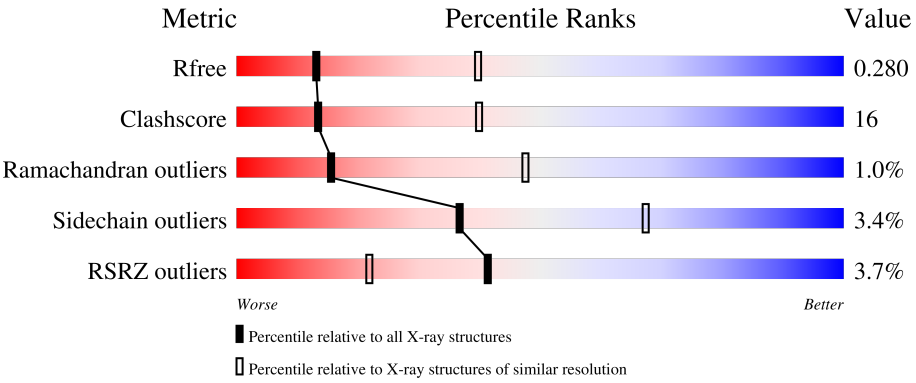
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 3.13 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	1626 (3.18-3.10)
Clashscore	141614	1735 (3.18-3.10)
Ramachandran outliers	138981	1677 (3.18-3.10)
Sidechain outliers	138945	1677 (3.18-3.10)
RSRZ outliers	127900	1588 (3.18-3.10)

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	AAZA	507	2% 84% · 12%
2	C	136	12% 38% 21% · 40%
3	R	190	% 63% 20% · 16%
4	N	124	2% 39% 44% 6% 12%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 6443 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 Son of sevenless homolog 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	AAZA	445	Total	C	N	O	S	0	0	0
			3698	2381	634	671	12			

There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAZA	543	MET	-	initiating methionine	UNP Q07889
AAZA	544	GLY	-	expression tag	UNP Q07889
AAZA	545	SER	-	expression tag	UNP Q07889
AAZA	546	SER	-	expression tag	UNP Q07889
AAZA	547	HIS	-	expression tag	UNP Q07889
AAZA	548	HIS	-	expression tag	UNP Q07889
AAZA	549	HIS	-	expression tag	UNP Q07889
AAZA	550	HIS	-	expression tag	UNP Q07889
AAZA	551	HIS	-	expression tag	UNP Q07889
AAZA	552	HIS	-	expression tag	UNP Q07889
AAZA	553	SER	-	expression tag	UNP Q07889
AAZA	554	SER	-	expression tag	UNP Q07889
AAZA	555	GLY	-	expression tag	UNP Q07889
AAZA	556	LEU	-	expression tag	UNP Q07889
AAZA	557	VAL	-	expression tag	UNP Q07889
AAZA	558	PRO	-	expression tag	UNP Q07889
AAZA	559	ARG	-	expression tag	UNP Q07889
AAZA	560	GLY	-	expression tag	UNP Q07889
AAZA	561	SER	-	expression tag	UNP Q07889
AAZA	562	HIS	-	expression tag	UNP Q07889
AAZA	563	MET	-	expression tag	UNP Q07889

- Molecule 2 is a protein called Nanobody75.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	81	Total	C	N	O	S	0	0	0
			610	380	110	117	3			

- Molecule 3 is a protein called Isoform 2B of GTPase KRas.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	R	159	Total	C	N	O	S	0	0	0
			1263	791	214	251	7			

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
R	-20	MET	-	initiating methionine	UNP P01116-2
R	-19	GLY	-	expression tag	UNP P01116-2
R	-18	SER	-	expression tag	UNP P01116-2
R	-17	SER	-	expression tag	UNP P01116-2
R	-16	HIS	-	expression tag	UNP P01116-2
R	-15	HIS	-	expression tag	UNP P01116-2
R	-14	HIS	-	expression tag	UNP P01116-2
R	-13	HIS	-	expression tag	UNP P01116-2
R	-12	HIS	-	expression tag	UNP P01116-2
R	-11	HIS	-	expression tag	UNP P01116-2
R	-10	SER	-	expression tag	UNP P01116-2
R	-9	SER	-	expression tag	UNP P01116-2
R	-8	GLY	-	expression tag	UNP P01116-2
R	-7	GLU	-	expression tag	UNP P01116-2
R	-6	ASN	-	expression tag	UNP P01116-2
R	-5	LEU	-	expression tag	UNP P01116-2
R	-4	TYR	-	expression tag	UNP P01116-2
R	-3	PHE	-	expression tag	UNP P01116-2
R	-2	GLN	-	expression tag	UNP P01116-2
R	-1	GLY	-	expression tag	UNP P01116-2
R	0	SER	-	expression tag	UNP P01116-2
R	12	VAL	GLY	engineered mutation	UNP P01116-2

- Molecule 4 is a protein called Nanobody22.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	N	109	Total	C	N	O	S	0	0	0
			853	528	158	163	4			

- Molecule 5 is water.

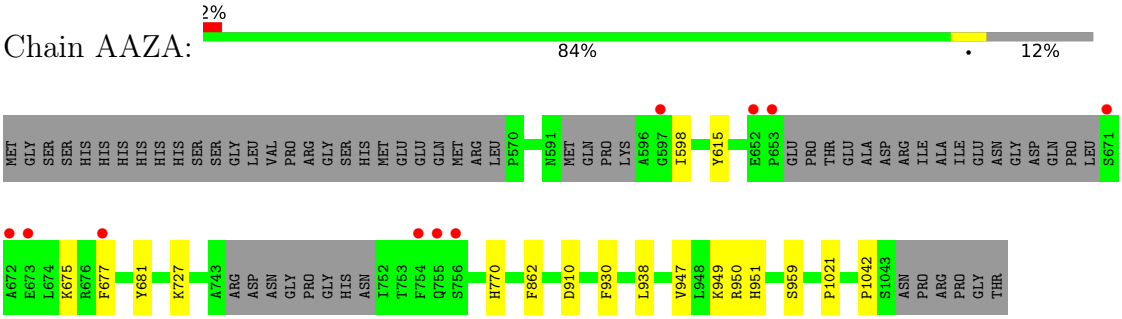
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	AAZA	10	Total O 10 10	0	0
5	C	2	Total O 2 2	0	0
5	R	6	Total O 6 6	0	0
5	N	1	Total O 1 1	0	0

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.

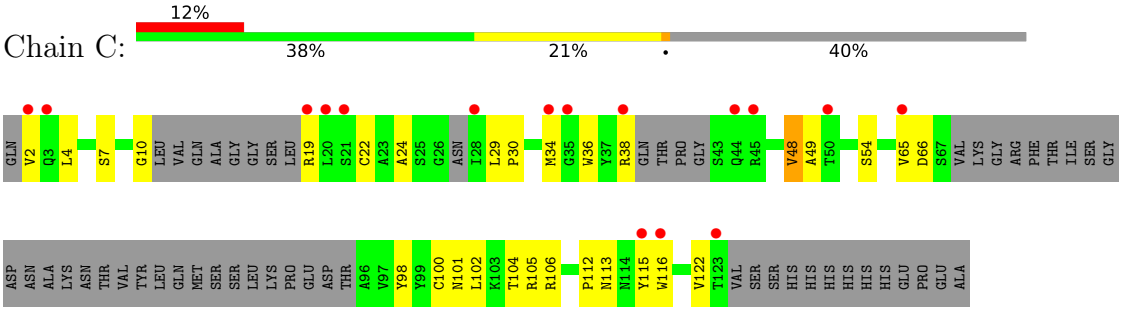
• Molecule 1: Son of sevenless homolog 1

Chain AAZA:



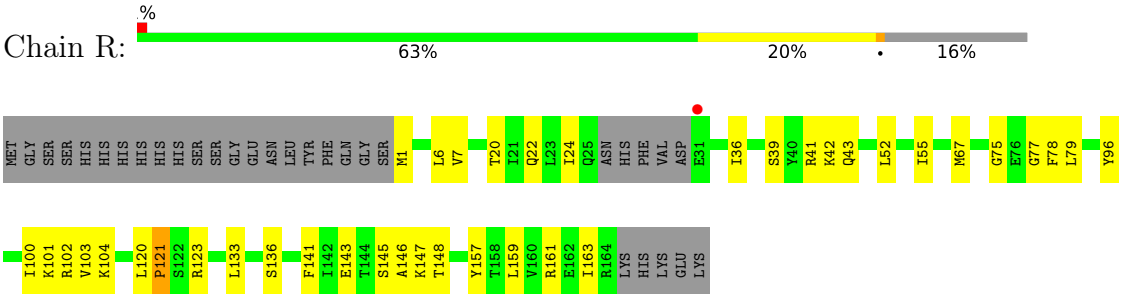
• Molecule 2: Nanobody75

Chain C:



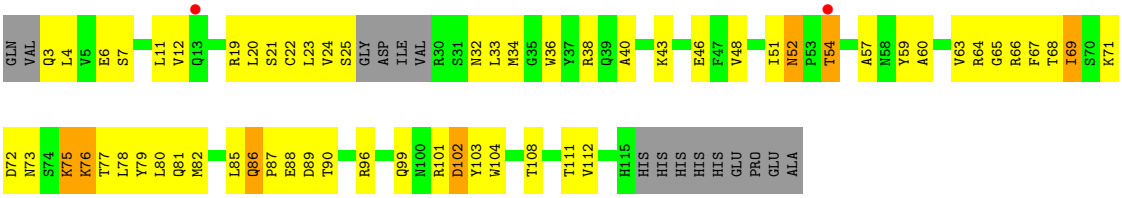
• Molecule 3: Isoform 2B of GTPase KRas

Chain R:



• Molecule 4: Nanobody22

Chain N:


GLN VAL Q3 L4 V5 E6 S7 L11 V12 Q13 R19 L20 S21 C22 L23 V24 S25 GLY ASP ILE VAL R30 S31 N32 L33 M34 G35 W36 Y37 R38 Q39 A40 K43 E46 F47 V48 I51 N52 P53 T54 A57 N58 Y59 A60 V63 R64 G65 R66 F67 T68 I69 S70 K71
D72 N73 S74 K75 K76 T77 L78 Y79 L80 Q81 M82 L85 Q86 P87 E88 D89 T90 R96 Q99 N100 R101 D102 Y103 W104 T108 T111 V112 H115 HIS HIS HIS HIS GLU PRO GLU ALA

4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 4 2 2	Depositor
Cell constants a, b, c, α, β, γ	124.08Å 124.08Å 408.81Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.91 – 3.13 29.91 – 3.13	Depositor EDS
% Data completeness (in resolution range)	99.5 (29.91-3.13) 99.5 (29.91-3.13)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.96 (at 3.11Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.226 , 0.278 0.227 , 0.280	Depositor DCC
R_{free} test set	1427 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	99.4	Xtriage
Anisotropy	0.296	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 63.1	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	0.93	EDS
Total number of atoms	6443	wwPDB-VP
Average B, all atoms (Å^2)	93.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.20% 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	AAZA	0.56	0/3787	0.72	0/5120
2	C	0.48	0/619	0.72	0/839
3	R	0.56	0/1280	0.74	0/1727
4	N	0.57	0/867	0.79	0/1168
All	All	0.55	0/6553	0.73	0/8854

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

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	AAZA	3698	0	0	0	0
2	C	610	0	586	23	0
3	R	1263	0	1252	27	0
4	N	853	0	835	51	0
5	AAZA	10	0	0	0	0
5	C	2	0	0	0	0
5	N	1	0	0	0	0
5	R	6	0	0	0	0
All	All	6443	0	2673	101	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 16.

The worst 5 of 101 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)
2:C:65:VAL:HG13	2:C:66:ASP:H	1.34	0.92
3:R:120:LEU:HB2	3:R:121:PRO:HD3	1.51	0.92
2:C:65:VAL:HG13	2:C:66:ASP:N	1.93	0.83
2:C:29:LEU:HD12	2:C:104:THR:HB	1.60	0.81
4:N:57:ALA:CB	4:N:69:ILE:HD11	2.15	0.75

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	AAZA	437/507 (86%)	396 (91%)	37 (8%)	4 (1%)	17	50
2	C	71/136 (52%)	56 (79%)	14 (20%)	1 (1%)	11	39
3	R	155/190 (82%)	142 (92%)	11 (7%)	2 (1%)	12	41
4	N	105/124 (85%)	90 (86%)	14 (13%)	1 (1%)	15	47
All	All	768/957 (80%)	684 (89%)	76 (10%)	8 (1%)	15	47

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	N	76	LYS
1	AAZA	677	PHE
3	R	147	LYS
2	C	54	SER
1	AAZA	727	LYS

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	AAZA	416/469 (89%)	402 (97%)	14 (3%)	37 67
2	C	64/111 (58%)	62 (97%)	2 (3%)	40 69
3	R	140/168 (83%)	139 (99%)	1 (1%)	84 93
4	N	92/105 (88%)	85 (92%)	7 (8%)	13 39
All	All	712/853 (84%)	688 (97%)	24 (3%)	37 67

5 of 24 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	7	SER
4	N	33	LEU
3	R	67	MET
4	N	52	ASN
1	AAZA	910	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
2	C	3	GLN
2	C	61	ASN
4	N	73	ASN

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 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)

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, 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	AAZA	445/507 (87%)	-0.22	10 (2%) 62 42	54, 84, 133, 158	0
2	C	81/136 (59%)	0.81	16 (19%) 1 0	83, 131, 147, 166	0
3	R	159/190 (83%)	-0.27	1 (0%) 89 80	56, 76, 108, 131	0
4	N	109/124 (87%)	0.02	2 (1%) 68 49	77, 108, 129, 136	0
All	All	794/957 (82%)	-0.09	29 (3%) 41 22	54, 89, 139, 166	0

The worst 5 of 29 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	AAZA	653	PRO	5.7
2	C	38	ARG	4.5
2	C	21	SER	4.0
1	AAZA	754	PHE	3.5
1	AAZA	652	GLU	3.1

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 (i)

There are no such residues in this entry.