Full wwPDB X-ray Structure Validation Report (i)

May 17, 2020 – 01:46 pm BST

PDB ID	:	5NJ6
Title	:	Crystal structure of a thermostabilised human protease-activated receptor-2 (PAR2) in ternary complex with Fab3949 and AZ7188 at 4.0 angstrom resolution
Authors	:	Cheng, R.K.Y.; Fiez-Vandal, C.; Schlenker, O.; Edman, K.; Aggeler, B.; Brown, D.G.; Brown, G.; Cooke, R.M.; Dumelin, C.E.; Dore, A.S.; Geschwindner, S.; Grebner, C.; Hermansson, N.-O.; Jazayeri, A.; Johansson, P.; Leong, L.; Prihandoko, R.; Rappas, M.; Soutter, H.; Snijder, A.; Sundstrom, L.; Tehan, B.; Thornton, P.; Troast, D.; Wiggin, G.; Zhukov, A.; Marshall, F.H.; Dekker, N.
Deposited on	:	2017-03-28
Resolution	:	4.00 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 (i)) were used in the production of this report:

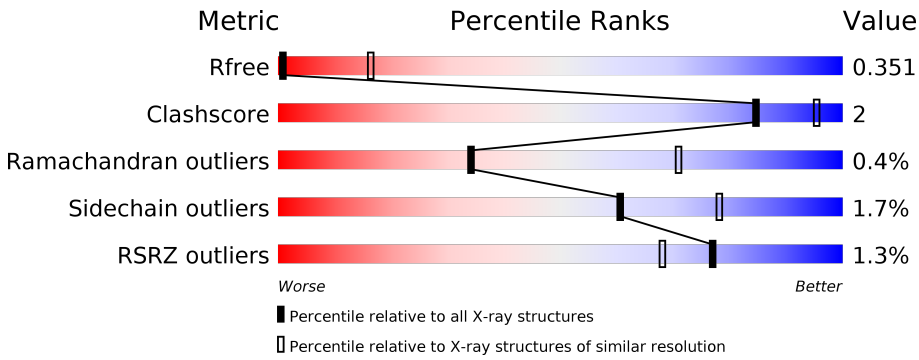
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 4.00 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	1087 (4.30-3.70)
Clashscore	141614	1148 (4.30-3.70)
Ramachandran outliers	138981	1108 (4.30-3.70)
Sidechain outliers	138945	1099 (4.30-3.70)
RSRZ outliers	127900	1028 (4.34-3.66)

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% 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	437	2% 83% 8% 8%
2	H	219	90% 10%
3	L	213	94% 6%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 6467 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 Proteinase-activated receptor 2,Soluble cytochrome b562,Proteinase-activated receptor 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	401	3172	2089	513	554	16	0	0	0

There are 27 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	54	MET	-	initiating methionine	UNP P55085
A	89	ALA	GLY	conflict	UNP P55085
A	108	ALA	HIS	conflict	UNP P55085
A	157	ALA	GLY	conflict	UNP P55085
A	166	LEU	MET	conflict	UNP P55085
A	174	ALA	TYR	conflict	UNP P55085
A	176	GLU	VAL	conflict	UNP P55085
A	222	GLN	ASN	conflict	UNP P55085
A	268	ALA	MET	conflict	UNP P55085
A	2006	TRP	MET	conflict	UNP P0ABE7
A	2101	ILE	HIS	conflict	UNP P0ABE7
A	2105	LEU	-	linker	UNP P0ABE7
A	289	ALA	ILE	conflict	UNP P55085
A	293	ALA	LEU	conflict	UNP P55085
A	378	ALA	-	expression tag	UNP P55085
A	379	ALA	-	expression tag	UNP P55085
A	380	ALA	-	expression tag	UNP P55085
A	381	HIS	-	expression tag	UNP P55085
A	382	HIS	-	expression tag	UNP P55085
A	383	HIS	-	expression tag	UNP P55085
A	384	HIS	-	expression tag	UNP P55085
A	385	HIS	-	expression tag	UNP P55085
A	386	HIS	-	expression tag	UNP P55085
A	387	HIS	-	expression tag	UNP P55085
A	388	HIS	-	expression tag	UNP P55085
A	389	HIS	-	expression tag	UNP P55085

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	390	HIS	-	expression tag	UNP P55085

- Molecule 2 is a protein called Fab3949 H.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	H	219	1657	1047	269	333	8	0	0	0

- Molecule 3 is a protein called Fab3949 L.

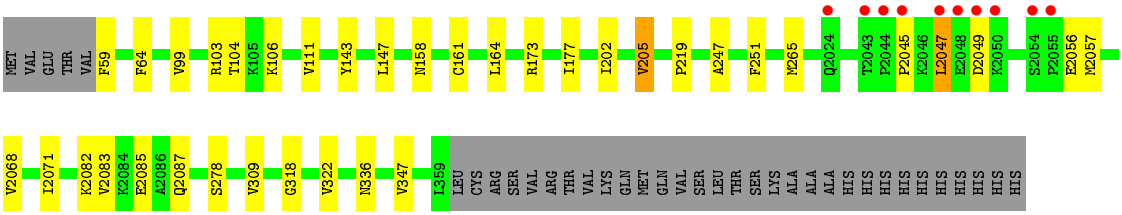
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	L	213	1638	1019	276	335	8	0	0	0

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 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.

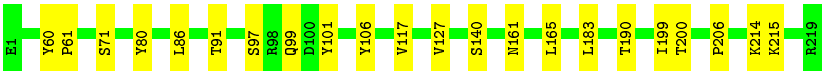
• Molecule 1: Proteinase-activated receptor 2,Soluble cytochrome b562,Proteinase-activated receptor 2

Chain A:



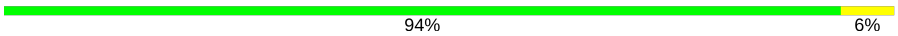
• Molecule 2: Fab3949 H

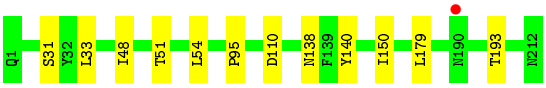
Chain H:



• Molecule 3: Fab3949 L

Chain L:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	164.14Å 38.52Å 159.66Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.13 – 4.00 45.13 – 4.00	Depositor EDS
% Data completeness (in resolution range)	91.6 (45.13-4.00) 84.7 (45.13-4.00)	Depositor EDS
R_{merge}	0.32	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.61 (at 4.00Å)	Xtriage
Refinement program	BUSTER 2.11.6	Depositor
R, R_{free}	0.263 , 0.319 0.296 , 0.351	Depositor DCC
R_{free} test set	395 reflections (4.69%)	wwPDB-VP
Wilson B-factor (Å^2)	66.8	Xtriage
Anisotropy	0.271	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.24 , -8.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.43$, $< L^2 > = 0.26$	Xtriage
Estimated twinning fraction	0.021 for l,-k,h	Xtriage
F_o,F_c correlation	0.79	EDS
Total number of atoms	6467	wwPDB-VP
Average B, all atoms (Å^2)	71.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.94% 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)

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	0.29	0/3246	0.43	0/4415
2	H	0.35	0/1699	0.54	0/2319
3	L	0.34	0/1678	0.53	0/2278
All	All	0.32	0/6623	0.49	0/9012

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	3172	0	3265	20	0
2	H	1657	0	1615	8	0
3	L	1638	0	1565	5	0
All	All	6467	0	6445	30	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 2.

All (30) 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:59:PHE:O	1:A:219:PRO:HD2	1.90	0.70

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:2045:PRO:HG2	1:A:278:SER:HA	1.74	0.69
1:A:173:ARG:HD2	1:A:347:VAL:HG13	1.85	0.59
2:H:71:SER:HB3	2:H:80:TYR:HB2	1.86	0.58
1:A:99:VAL:HA	1:A:103:ARG:HB2	1.89	0.55
2:H:91:THR:HG22	2:H:117:VAL:H	1.72	0.53
1:A:2047:LEU:HD22	1:A:2056:GLU:HB3	1.90	0.52
3:L:48:ILE:HG12	3:L:54:LEU:HD23	1.93	0.51
2:H:97:SER:HB2	2:H:106:TYR:HB3	1.93	0.50
2:H:200:THR:HG22	2:H:215:LYS:HA	1.93	0.49
1:A:143:TYR:HB3	1:A:147:LEU:HD23	1.94	0.49
1:A:318:GLY:HA2	3:L:31:SER:HB2	1.94	0.49
2:H:61:PRO:HB3	3:L:95:PRO:HB3	1.95	0.47
1:A:2082:LYS:HB3	1:A:2085:GLU:HB2	1.96	0.47
1:A:202:ILE:HA	1:A:205:VAL:HG12	1.96	0.47
3:L:110:ASP:HA	3:L:140:TYR:HB3	1.97	0.46
1:A:64:PHE:HB2	2:H:101:TYR:HE1	1.82	0.44
1:A:309:VAL:HG12	1:A:322:VAL:HG21	1.99	0.44
1:A:2068:VAL:HA	1:A:2071:ILE:HD12	2.00	0.44
1:A:106:LYS:HD3	1:A:2049:ASP:HB2	1.99	0.43
1:A:177:ILE:HD12	1:A:265:MET:HG3	2.00	0.43
3:L:150:ILE:HD11	3:L:179:LEU:HD21	1.99	0.43
1:A:104:THR:HG21	1:A:111:VAL:HG21	2.00	0.43
1:A:164:LEU:HD12	1:A:202:ILE:HD11	2.01	0.42
1:A:247:ALA:O	1:A:251:PHE:HB3	2.20	0.41
1:A:2083:VAL:HG12	1:A:2087:GLN:HE21	1.85	0.41
2:H:161:ASN:HD21	2:H:199:ILE:HA	1.86	0.41
1:A:158:ASN:HD21	1:A:336:ASN:HD21	1.68	0.40
2:H:127:VAL:HB	2:H:214:LYS:HD3	2.03	0.40
1:A:161:CYS:SG	1:A:202:ILE:HD12	2.61	0.40

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	397/437 (91%)	382 (96%)	15 (4%)	0	100	100
2	H	217/219 (99%)	201 (93%)	15 (7%)	1 (0%)	29	67
3	L	211/213 (99%)	202 (96%)	7 (3%)	2 (1%)	17	55
All	All	825/869 (95%)	785 (95%)	37 (4%)	3 (0%)	34	71

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	L	51	THR
3	L	138	ASN
2	H	206	PRO

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	342/375 (91%)	339 (99%)	3 (1%)	78	88
2	H	189/189 (100%)	182 (96%)	7 (4%)	34	60
3	L	187/187 (100%)	185 (99%)	2 (1%)	73	85
All	All	718/751 (96%)	706 (98%)	12 (2%)	60	78

All (12) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	205	VAL
1	A	2047	LEU
1	A	2057	MET
2	H	60	TYR
2	H	86	LEU
2	H	99	GLN
2	H	140	SER
2	H	165	LEU
2	H	183	LEU
2	H	190	THR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	L	33	LEU
3	L	193	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	158	ASN
1	A	2087	GLN
2	H	161	ASN
3	L	138	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 carbohydrates 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	2085:GLU	C	2086:ALA	N	3.15

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	401/437 (91%)	-0.14	10 (2%) 57 47	50, 76, 106, 121	0
2	H	219/219 (100%)	-0.39	0 100 100	42, 64, 79, 96	0
3	L	213/213 (100%)	-0.40	1 (0%) 91 85	41, 59, 93, 101	0
All	All	833/869 (95%)	-0.27	11 (1%) 77 68	41, 69, 100, 121	0

All (11) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	2044	PRO	6.5
1	A	2048	GLU	4.7
1	A	2055	PRO	4.5
1	A	2054	SER	4.2
1	A	2049	ASP	3.5
1	A	2045	PRO	3.2
1	A	2043	THR	3.1
1	A	2047	LEU	3.0
1	A	2024	GLN	2.7
3	L	190	ASN	2.4
1	A	2050	LYS	2.0

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)

There are no carbohydrates in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.