Full wwPDB X-ray Structure Validation Report (i)

Feb 1, 2016 – 07:49 PM GMT

PDB ID	:	4Q2C
Title	:	Crystal structure of CRISPR-associated protein
Authors	:	Gong, B.; Shin, M.; Sun, J.; van der Oost, J.; Kim, J.-S.
Deposited on	:	2014-04-07
Resolution	:	2.50 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
http://wwpdb.org/validation/2016/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:

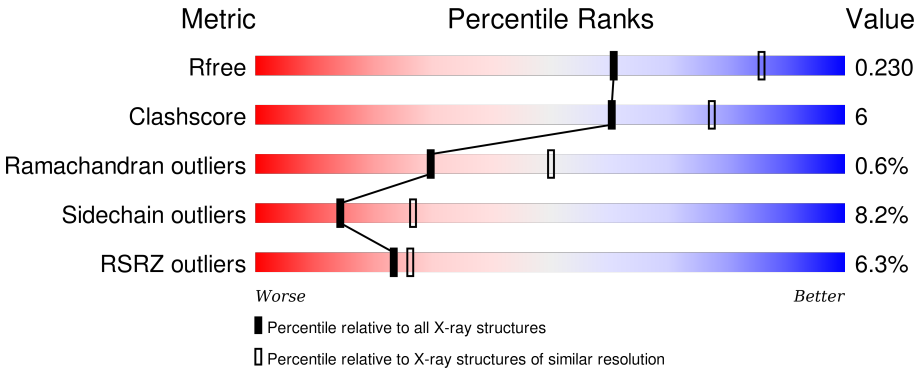
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 2.50 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}	91344	3553 (2.50-2.50)
Clashscore	102246	4242 (2.50-2.50)
Ramachandran outliers	100387	4156 (2.50-2.50)
Sidechain outliers	100360	4158 (2.50-2.50)
RSRZ outliers	91569	3562 (2.50-2.50)

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. 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	949	6% 80% 15%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	NI	A	1002	-	-	-	X

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	NI	A	1003	-	-	-	X

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 7487 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 CRISPR-associated helicase Cas3.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	926	Total	C	N	O	S	Se	0	0	0
			7202	4572	1292	1312	7	19			

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-4	GLY	-	EXPRESSION TAG	UNP D1CGD0
A	-3	GLY	-	EXPRESSION TAG	UNP D1CGD0
A	-2	GLY	-	EXPRESSION TAG	UNP D1CGD0
A	-1	GLY	-	EXPRESSION TAG	UNP D1CGD0
A	0	GLY	-	EXPRESSION TAG	UNP D1CGD0

- Molecule 2 is NICKEL (II) ION (three-letter code: NI) (formula: Ni).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	3	Total	Ni	0	0
			3	3		

- Molecule 3 is water.

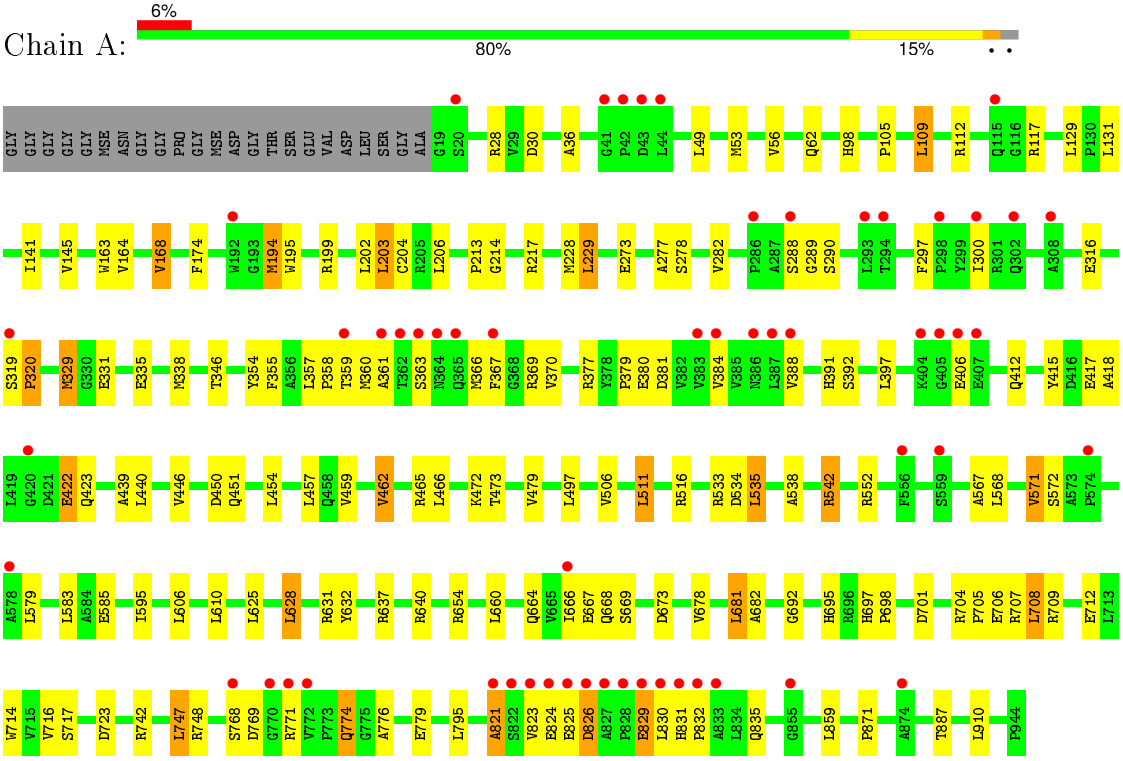
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	282	Total	O	0	0
			282	282		

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 errors 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: CRISPR-associated helicase Cas3

Chain A:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	104.27Å 214.56Å 102.20Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	93.78 – 2.50 93.78 – 2.50	Depositor EDS
% Data completeness (in resolution range)	93.4 (93.78-2.50) 93.4 (93.78-2.50)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.91 (at 2.51Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.1_1168)	Depositor
R, R_{free}	0.191 , 0.229 0.188 , 0.230	Depositor DCC
R_{free} test set	1870 reflections (5.00%)	DCC
Wilson B-factor (Å^2)	41.6	Xtriage
Anisotropy	0.566	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 53.7	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.32$	Xtriage
Outliers	0 of 37404 reflections	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	7487	wwPDB-VP
Average B, all atoms (Å^2)	51.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.27% 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, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

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

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	0.24	0/7361	0.47	0/10000

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	A	7202	0	7156	81	0
2	A	3	0	0	0	0
3	A	282	0	0	6	0
All	All	7487	0	7156	81	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 6.

All (81) 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:418:ALA:HB3	1:A:826:ASP:HB2	1.65	0.79

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:829:GLU:HA	1:A:830:LEU:HB2	1.66	0.77
1:A:572:SER:HA	1:A:717:SER:HB3	1.70	0.74
1:A:701:ASP:OD1	1:A:704:ARG:NH1	2.21	0.72
1:A:109:LEU:O	1:A:117:ARG:NH1	2.23	0.71
1:A:129:LEU:HD23	1:A:194:MSE:HB3	1.73	0.69
1:A:415:TYR:CE2	1:A:422:GLU:HG2	2.29	0.67
1:A:98:HIS:HA	1:A:168:VAL:HG13	1.77	0.66
1:A:748:ARG:NH1	1:A:774:GLN:OE1	2.29	0.65
1:A:288:SER:OG	1:A:289:GLY:O	2.16	0.63
1:A:53:MSE:HE1	1:A:202:LEU:HB3	1.81	0.63
1:A:316:GLU:OE1	1:A:552:ARG:NH2	2.30	0.61
1:A:871:PRO:HG3	1:A:910:LEU:HD12	1.83	0.60
1:A:361:ALA:HA	1:A:669:SER:OG	2.04	0.58
1:A:62:GLN:NE2	1:A:273:GLU:OE2	2.37	0.58
1:A:567:ALA:HB3	1:A:712:GLU:HG2	1.86	0.57
1:A:701:ASP:O	1:A:709:ARG:NH1	2.36	0.57
1:A:511:LEU:O	1:A:516:ARG:NH1	2.38	0.56
1:A:821:ALA:H	1:A:823:VAL:HG23	1.70	0.56
1:A:705:PRO:HG2	1:A:708:LEU:HB2	1.87	0.56
1:A:357:LEU:HD12	1:A:363:SER:HA	1.87	0.56
1:A:654:ARG:NH2	1:A:673:ASP:O	2.39	0.56
1:A:109:LEU:HD21	1:A:131:LEU:HG	1.88	0.56
1:A:821:ALA:HA	1:A:823:VAL:H	1.71	0.55
1:A:632:TYR:HB2	1:A:637:ARG:HG2	1.87	0.55
1:A:199:ARG:NH1	3:A:1214:HOH:O	2.39	0.55
1:A:338:MSE:HE1	1:A:370:VAL:HA	1.90	0.54
1:A:535:LEU:H	1:A:535:LEU:HD13	1.74	0.53
1:A:454:LEU:HD23	1:A:457:LEU:HD12	1.89	0.53
1:A:771:ARG:NH2	3:A:1157:HOH:O	2.38	0.53
1:A:640:ARG:NH2	3:A:1124:HOH:O	2.40	0.53
1:A:289:GLY:HA3	1:A:377:ARG:HH12	1.72	0.53
1:A:742:ARG:NH1	3:A:1380:HOH:O	2.37	0.52
1:A:329:MSE:SE	1:A:329:MSE:H	2.42	0.52
1:A:538:ALA:O	1:A:542:ARG:NH1	2.44	0.51
1:A:174:PHE:HB2	1:A:824:GLU:HG3	1.92	0.51
1:A:706:GLU:HA	1:A:709:ARG:HG3	1.92	0.50
1:A:459:VAL:O	1:A:462:VAL:HG22	2.10	0.50
1:A:367:PHE:CE2	1:A:391:HIS:HA	2.47	0.50
1:A:392:SER:HB3	1:A:439:ALA:HB2	1.93	0.50
1:A:329:MSE:HG3	1:A:692:GLY:HA3	1.93	0.50
1:A:681:LEU:HD22	1:A:682:ALA:H	1.77	0.50

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:723:ASP:OD2	1:A:776:ALA:HB1	2.12	0.49
1:A:358:PRO:HG2	1:A:361:ALA:HB2	1.94	0.49
1:A:628:LEU:HD13	1:A:666:ILE:HD11	1.94	0.49
1:A:319:SER:HB2	1:A:320:PRO:HD3	1.93	0.49
1:A:667:GLU:HG2	1:A:668:GLN:HG3	1.94	0.48
1:A:695:HIS:HA	1:A:704:ARG:NH2	2.29	0.48
1:A:204:CYS:SG	3:A:1136:HOH:O	2.60	0.48
1:A:628:LEU:HB2	1:A:660:LEU:HD11	1.95	0.47
1:A:379:PRO:O	1:A:381:ASP:N	2.48	0.47
1:A:164:VAL:O	1:A:168:VAL:HB	2.16	0.46
1:A:367:PHE:CE2	1:A:446:VAL:HB	2.51	0.46
1:A:571:VAL:HG13	1:A:716:VAL:HA	1.98	0.45
1:A:354:TYR:HB2	1:A:472:LYS:HD3	1.99	0.45
1:A:213:PRO:HA	1:A:214:GLY:HA2	1.62	0.45
1:A:277:ALA:HA	1:A:282:VAL:HG22	1.98	0.45
1:A:288:SER:HA	1:A:289:GLY:HA3	1.69	0.45
1:A:451:GLN:OE1	1:A:465:ARG:NH2	2.50	0.45
1:A:678:VAL:HG22	1:A:714:TRP:HB2	1.99	0.45
1:A:829:GLU:HG2	1:A:832:PRO:HD2	1.99	0.44
1:A:297:PHE:O	1:A:300:ILE:HG22	2.17	0.44
1:A:56:VAL:HG21	1:A:98:HIS:CE1	2.53	0.43
1:A:821:ALA:HA	1:A:823:VAL:N	2.34	0.43
1:A:335:GLU:OE1	1:A:369:ARG:NH1	2.52	0.43
1:A:831:HIS:HB2	1:A:832:PRO:HD3	2.01	0.42
1:A:105:PRO:HD3	1:A:195:TRP:CE2	2.54	0.42
1:A:459:VAL:HG12	1:A:830:LEU:H	1.84	0.42
1:A:141:ILE:O	1:A:145:VAL:HG13	2.18	0.42
1:A:288:SER:HB2	1:A:346:THR:HG23	2.01	0.42
1:A:203:LEU:HA	1:A:203:LEU:HD12	1.88	0.42
1:A:36:ALA:HB2	1:A:49:LEU:HA	2.01	0.42
1:A:595:ILE:HG21	1:A:606:LEU:HD21	2.01	0.42
1:A:701:ASP:HA	1:A:704:ARG:HH11	1.85	0.41
1:A:654:ARG:HD3	3:A:1132:HOH:O	2.20	0.41
1:A:300:ILE:HD12	1:A:300:ILE:HA	1.97	0.41
1:A:479:VAL:HG21	1:A:506:VAL:HG13	2.02	0.41
1:A:747:LEU:HA	1:A:747:LEU:HD12	1.91	0.41
1:A:229:LEU:HD12	1:A:229:LEU:HA	1.85	0.41
1:A:422:GLU:HG3	1:A:422:GLU:H	1.47	0.40
1:A:697:HIS:HA	1:A:698:PRO:HD2	1.87	0.40

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	924/949 (97%)	889 (96%)	29 (3%)	6 (1%)	30 50

All (6) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	290	SER
1	A	769	ASP
1	A	380	GLU
1	A	821	ALA
1	A	30	ASP
1	A	320	PRO

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	745/736 (101%)	684 (92%)	61 (8%)	14 27

All (61) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	28	ARG
1	A	109	LEU
1	A	112	ARG
1	A	163	TRP
1	A	168	VAL

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	194	MSE
1	A	203	LEU
1	A	206	LEU
1	A	217	ARG
1	A	228	MSE
1	A	229	LEU
1	A	278	SER
1	A	329	MSE
1	A	331	GLU
1	A	355	PHE
1	A	359	THR
1	A	360	MSE
1	A	366	MSE
1	A	384	VAL
1	A	388	VAL
1	A	397	LEU
1	A	406	GLU
1	A	412	GLN
1	A	417	GLU
1	A	422	GLU
1	A	423	GLN
1	A	440	LEU
1	A	450	ASP
1	A	462	VAL
1	A	466	LEU
1	A	473	THR
1	A	497	LEU
1	A	511	LEU
1	A	533	ARG
1	A	534	ASP
1	A	535	LEU
1	A	542	ARG
1	A	568	LEU
1	A	571	VAL
1	A	579	LEU
1	A	583	LEU
1	A	585	GLU
1	A	610	LEU
1	A	625	LEU
1	A	628	LEU
1	A	631	ARG
1	A	664	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	681	LEU
1	A	707	ARG
1	A	708	LEU
1	A	747	LEU
1	A	768	SER
1	A	774	GLN
1	A	779	GLU
1	A	795	LEU
1	A	825	GLU
1	A	826	ASP
1	A	829	GLU
1	A	835	GLN
1	A	859	LEU
1	A	887	THR

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)

Of 3 ligands modelled in this entry, 3 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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, 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	A	907/949 (95%)	0.61	57 (6%)	23	26	24, 45, 100, 175	0

All (57) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	828	PRO	15.4
1	A	827	ALA	14.4
1	A	826	ASP	13.3
1	A	831	HIS	9.7
1	A	832	PRO	8.8
1	A	364	ASN	7.9
1	A	830	LEU	7.7
1	A	829	GLU	6.1
1	A	822	SER	6.0
1	A	823	VAL	5.1
1	A	406	GLU	5.0
1	A	41	GLY	4.9
1	A	824	GLU	4.9
1	A	363	SER	4.7
1	A	43	ASP	4.7
1	A	772	VAL	4.6
1	A	387	LEU	4.2
1	A	362	THR	4.0
1	A	855	GLY	4.0
1	A	825	GLU	3.8
1	A	833	ALA	3.7
1	A	405	GLY	3.7
1	A	361	ALA	3.6
1	A	821	ALA	3.4
1	A	420	GLY	3.3
1	A	365	GLN	3.3
1	A	383	VAL	3.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	359	THR	3.2
1	A	115	GLN	3.2
1	A	574	PRO	3.0
1	A	302	GLN	3.0
1	A	42	PRO	2.9
1	A	666	ILE	2.9
1	A	407	GLU	2.8
1	A	768	SER	2.7
1	A	367	PHE	2.6
1	A	286	PRO	2.6
1	A	319	SER	2.5
1	A	298	PRO	2.5
1	A	294	THR	2.5
1	A	388	VAL	2.5
1	A	288	SER	2.5
1	A	308	ALA	2.5
1	A	578	ALA	2.5
1	A	300	ILE	2.4
1	A	386	ASN	2.4
1	A	44	LEU	2.4
1	A	556	PHE	2.3
1	A	559	SER	2.3
1	A	192	TRP	2.3
1	A	874	ALA	2.2
1	A	293	LEU	2.2
1	A	404	LYS	2.2
1	A	20	SER	2.1
1	A	771	ARG	2.1
1	A	770	GLY	2.1
1	A	384	VAL	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

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95^{th} 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	LLDF	B-factors(Å^2)	Q<0.9
2	NI	A	1003	1/1	0.95	0.22	3.14	57,57,57,57	0
2	NI	A	1002	1/1	0.99	0.21	2.66	30,30,30,30	0
2	NI	A	1001	1/1	0.99	0.16	-0.22	34,34,34,34	0

6.5 Other polymers

There are no such residues in this entry.