WORLDWIDE PDB PROTEIN DATA BANK

Full wwPDB X-ray Structure Validation Report ⓘ

Dec 2, 2024 – 05:59 PM EST

PDB ID	:	3M5O
Title	:	Crystal structure of HCV NS3/4A protease in complex with N-terminal product 5A5B
Authors	:	Schiffer, C.A.; Romano, K.P.
Deposited on	:	2010-03-12
Resolution	:	1.60 Å(reported)

The following versions of software and data (see references ⓘ) were used in the production of this report:

|		
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.21
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.004 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.40

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 1.60 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.
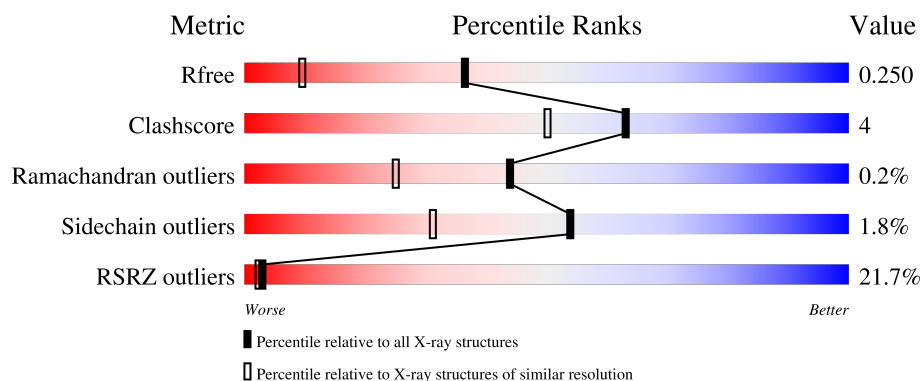
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	4274 (1.60-1.60)
Clashscore	180529	4682 (1.60-1.60)
Ramachandran outliers	177936	4583 (1.60-1.60)
Sidechain outliers	177891	4582 (1.60-1.60)
RSRZ outliers	164620	4272 (1.60-1.60)

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%

Mol	Chain	Length	Quality of chain
1	A	203	
1	B	203	
2	C	8	
2	D	8	

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 3526 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 NS3/4A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	198	Total	C	N	O	S	0	4	0
			1478	915	271	284	8			
1	B	198	Total	C	N	O	S	0	4	0
			1470	911	271	280	8			

There are 56 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	980	GLY	-	expression tag	UNP A8DG50
A	981	SER	-	expression tag	UNP A8DG50
A	982	HIS	-	expression tag	UNP A8DG50
A	983	MET	-	expression tag	UNP A8DG50
A	984	ALA	-	expression tag	UNP A8DG50
A	985	SER	-	expression tag	UNP A8DG50
A	986	MET	-	engineered mutation	UNP A8DG50
A	987	LYS	-	engineered mutation	UNP A8DG50
A	988	LYS	-	engineered mutation	UNP A8DG50
A	989	LYS	-	engineered mutation	UNP A8DG50
A	991	SER	CYS	SEE REMARK 999	UNP A8DG50
A	998	ILE	VAL	SEE REMARK 999	UNP A8DG50
A	999	ASN	ILE	SEE REMARK 999	UNP A8DG50
A	1001	SER	ALA	engineered mutation	UNP A8DG50
A	1002	GLY	PRO	engineered mutation	UNP A8DG50
A	1003	ASP	ILE	engineered mutation	UNP A8DG50
A	1013	GLU	LEU	engineered mutation	UNP A8DG50
A	1014	GLU	LEU	engineered mutation	UNP A8DG50
A	1017	GLN	ILE	engineered mutation	UNP A8DG50
A	1018	GLU	ILE	engineered mutation	UNP A8DG50
A	1021	GLN	LEU	engineered mutation	UNP A8DG50
A	1040	THR	ALA	engineered mutation	UNP A8DG50
A	1047	SER	CYS	engineered mutation	UNP A8DG50
A	1052	LEU	CYS	engineered mutation	UNP A8DG50
A	1072	THR	ILE	engineered mutation	UNP A8DG50

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	1086	GLN	PRO	engineered mutation	UNP A8DG50
A	1139	ALA	SER	engineered mutation	UNP A8DG50
A	1159	SER	CYS	engineered mutation	UNP A8DG50
B	980	GLY	-	expression tag	UNP A8DG50
B	981	SER	-	expression tag	UNP A8DG50
B	982	HIS	-	expression tag	UNP A8DG50
B	983	MET	-	expression tag	UNP A8DG50
B	984	ALA	-	expression tag	UNP A8DG50
B	985	SER	-	expression tag	UNP A8DG50
B	986	MET	-	engineered mutation	UNP A8DG50
B	987	LYS	-	engineered mutation	UNP A8DG50
B	988	LYS	-	engineered mutation	UNP A8DG50
B	989	LYS	-	engineered mutation	UNP A8DG50
B	991	SER	CYS	SEE REMARK 999	UNP A8DG50
B	998	ILE	VAL	SEE REMARK 999	UNP A8DG50
B	999	ASN	ILE	SEE REMARK 999	UNP A8DG50
B	1001	SER	ALA	engineered mutation	UNP A8DG50
B	1002	GLY	PRO	engineered mutation	UNP A8DG50
B	1003	ASP	ILE	engineered mutation	UNP A8DG50
B	1013	GLU	LEU	engineered mutation	UNP A8DG50
B	1014	GLU	LEU	engineered mutation	UNP A8DG50
B	1017	GLN	ILE	engineered mutation	UNP A8DG50
B	1018	GLU	ILE	engineered mutation	UNP A8DG50
B	1021	GLN	LEU	engineered mutation	UNP A8DG50
B	1040	THR	ALA	engineered mutation	UNP A8DG50
B	1047	SER	CYS	engineered mutation	UNP A8DG50
B	1052	LEU	CYS	engineered mutation	UNP A8DG50
B	1072	THR	ILE	engineered mutation	UNP A8DG50
B	1086	GLN	PRO	engineered mutation	UNP A8DG50
B	1139	ALA	SER	engineered mutation	UNP A8DG50
B	1159	SER	CYS	engineered mutation	UNP A8DG50

- Molecule 2 is a protein called TEDVVCC peptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	8	Total	C	N	O	S	0	0	0
			54	31	7	14	2			
2	D	8	Total	C	N	O	S	0	0	0
			54	31	7	14	2			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	0	ACE	-	acetylation	UNP B3TKQ3
D	0	ACE	-	acetylation	UNP B3TKQ3

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O_4S).
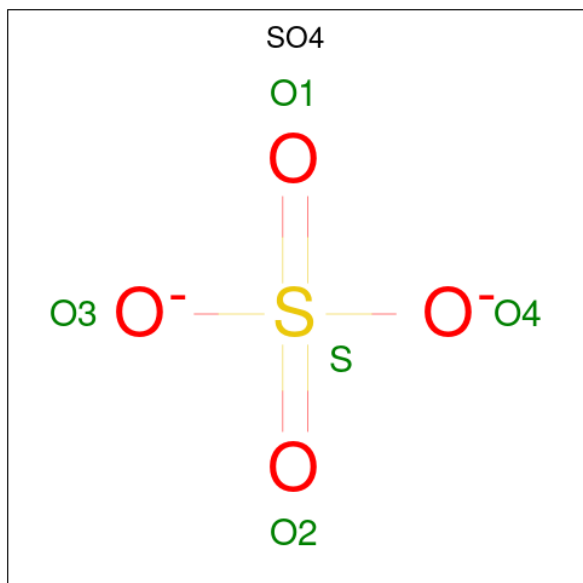
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	O	S	0	0
			5	4	1		

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	Zn	0	0
			1	1		
4	B	1	Total	Zn	0	0
			1	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	257	Total	O	0	0
			257	257		
5	B	186	Total	O	0	0
			186	186		
5	C	10	Total	O	0	0
			10	10		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	D	10	Total 10	O 10	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. 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. 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.
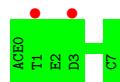
• Molecule 1: NS3/4A

Chain A:
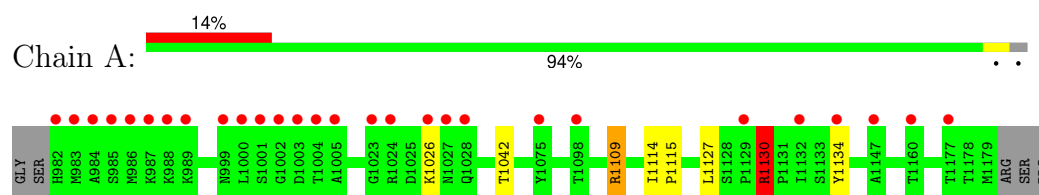
• Molecule 1: NS3/4A
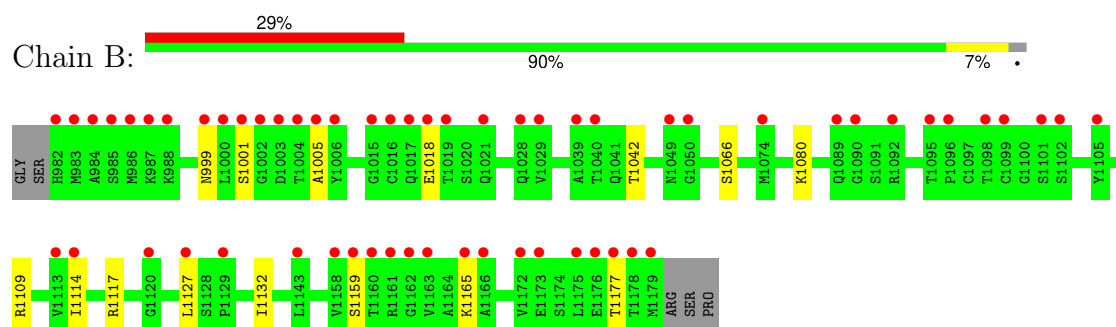
Chain B:

• Molecule 2: TEDVVCC peptide

Chain C:

There are no outlier residues recorded for this chain.

• Molecule 2: TEDVVCC peptide

Chain D:

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	47.27Å 58.88Å 67.06Å 90.00° 99.22° 90.00°	Depositor
Resolution (Å)	50.00 – 1.60 50.00 – 1.60	Depositor EDS
% Data completeness (in resolution range)	94.4 (50.00-1.60) 94.5 (50.00-1.60)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	0.07	Depositor
$< I/\sigma(I) >$ [1]	8.39 (at 1.60Å)	Xtriage
Refinement program	REFMAC 5.5.0102	Depositor
R, R_{free}	0.180 , 0.226 0.220 , 0.250	Depositor DCC
R_{free} test set	2289 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å^2)	14.0	Xtriage
Anisotropy	0.046	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 38.5	EDS
L-test for twinning[2]	$< |L| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	3526	wwPDB-VP
Average B, all atoms (Å^2)	23.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.34% 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 ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, SO4, ACE

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	0.47	0/1515	0.65	1/2056 (0.0%)
1	B	0.41	0/1507	0.65	0/2046
2	C	0.47	0/51	0.56	0/68
2	D	0.31	0/51	0.61	0/68
All	All	0.44	0/3124	0.65	1/4238 (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	1130	ARG	NE-CZ-NH1	5.53	123.07	120.30

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	1478	0	1487	11	0
1	B	1470	0	1479	13	0
2	C	54	0	48	0	0
2	D	54	0	48	0	0
3	A	5	0	0	0	0
4	A	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	B	1	0	0	0	0
5	A	257	0	0	0	0
5	B	186	0	0	1	0
5	C	10	0	0	0	0
5	D	10	0	0	0	0
All	All	3526	0	3062	24	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 4.

All (24) 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:1114:ILE:HD11	1:A:1134:TYR:HE2	1.46	0.80
1:B:1042:THR:HG21	1:B:1109:ARG:HH22	1.45	0.79
1:B:1117:ARG:CG	1:B:1127:LEU:HD21	2.13	0.79
1:B:1042:THR:HG22	1:B:1109:ARG:HH12	1.49	0.76
1:B:1117:ARG:HG2	1:B:1127:LEU:HD21	1.74	0.69
1:B:1042:THR:CG2	1:B:1109:ARG:HH22	2.09	0.66
1:A:1042:THR:HG21	1:A:1109[B]:ARG:HH22	1.63	0.63
1:A:1042:THR:HG22	1:A:1109[B]:ARG:HH12	1.67	0.60
1:B:1132:ILE:HG21	1:B:1159:SER:HB2	1.82	0.59
1:A:1114:ILE:HD11	1:A:1134:TYR:CE2	2.33	0.58
1:A:1115:PRO:HB2	1:A:1127:LEU:HD22	1.89	0.54
1:A:1042:THR:CG2	1:A:1109[B]:ARG:HH22	2.20	0.53
1:B:1117:ARG:HG3	1:B:1127:LEU:HD21	1.91	0.51
1:B:1018:GLU:N	1:B:1018:GLU:OE1	2.45	0.49
1:B:1042:THR:HG22	1:B:1109:ARG:NH1	2.24	0.48
1:B:1005:ALA:HB3	5:B:446:HOH:O	2.15	0.46
1:A:1130:ARG:HH11	1:A:1130:ARG:HG2	1.82	0.44
1:A:1114:ILE:CD1	1:A:1134:TYR:HE2	2.21	0.44
1:B:1080:LYS:NZ	1:B:1177:THR:OG1	2.45	0.43
1:A:1109[A]:ARG:HE	1:A:1109[A]:ARG:HB2	1.59	0.43
1:A:1130:ARG:HH11	1:A:1130:ARG:CG	2.32	0.42
1:A:1127:LEU:HD12	1:A:1127:LEU:N	2.34	0.42
1:B:999:ASN:O	1:B:1005:ALA:HA	2.20	0.41
1:B:1042:THR:HG21	1:B:1109:ARG:NH2	2.22	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	200/203 (98%)	198 (99%)	2 (1%)	0	100	100
1	B	200/203 (98%)	195 (98%)	4 (2%)	1 (0%)	25	10
2	C	6/8 (75%)	6 (100%)	0	0	100	100
2	D	6/8 (75%)	6 (100%)	0	0	100	100
All	All	412/422 (98%)	405 (98%)	6 (2%)	1 (0%)	44	25

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	1001	SER

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	161/164 (98%)	157 (98%)	4 (2%)	42	19
1	B	159/164 (97%)	156 (98%)	3 (2%)	52	29
2	C	7/7 (100%)	7 (100%)	0	100	100
2	D	7/7 (100%)	7 (100%)	0	100	100
All	All	334/342 (98%)	327 (98%)	7 (2%)	54	25

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1026	LYS
1	A	1109[A]	ARG
1	A	1109[B]	ARG
1	A	1130	ARG
1	B	1066	SER
1	B	1114	ILE
1	B	1165	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 oligosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 3 ligands modelled in this entry, 2 are monoatomic - leaving 1 for Mogul analysis.

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	SO4	A	1	-	4,4,4	0.27	0	6,6,6	0.18	0

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 ⓘ

Warning: The R factor obtained from EDS is 0.2317, which does not match the depositor's R factor of 0.17982. Please interpret the results in this section carefully.

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	198/203 (97%)	1.08	28 (14%)	7	6	9, 18, 44, 81	4 (2%)
1	B	198/203 (97%)	1.56	59 (29%)	1	1	11, 22, 46, 79	4 (2%)
2	C	7/8 (87%)	1.12	0	100	100	18, 19, 26, 26	0
2	D	7/8 (87%)	1.78	2 (28%)	1	1	23, 24, 40, 41	0
All	All	410/422 (97%)	1.32	89 (21%)	3	2	9, 20, 44, 81	8 (1%)

All (89) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1003	ASP	7.4
1	B	985	SER	6.5
1	A	985	SER	6.3
1	B	1016	CYS	5.4
1	A	1004	THR	5.2
1	B	984	ALA	5.2
1	B	1015	GLY	5.2
1	A	982	HIS	5.1
1	A	984	ALA	5.1
1	B	1002	GLY	5.1
1	A	1000	LEU	5.1
1	A	1001	SER	5.0
1	B	1003	ASP	5.0
1	A	1134	TYR	4.7
1	B	982	HIS	4.7
1	A	983	MET	4.5
1	A	986	MET	4.5
1	B	986	MET	4.3
1	B	983	MET	4.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	1002	GLY	4.2
1	A	1005	ALA	4.2
1	B	1004	THR	4.1
1	A	987	LYS	4.0
1	B	1001	SER	3.8
1	B	1120	GLY	3.7
1	B	1176	GLU	3.6
1	B	1179	MET	3.6
1	B	1172	VAL	3.6
1	B	1099	CYS	3.5
1	B	1127	LEU	3.5
1	B	987	LYS	3.4
1	B	988	LYS	3.3
1	B	1049	ASN	3.3
1	A	999	ASN	3.3
1	B	999	ASN	3.3
1	B	1177	THR	3.2
1	B	1178	THR	3.2
1	B	1017	GLN	3.1
1	A	988	LYS	3.0
1	B	1006	TYR	3.0
1	A	1023	GLY	3.0
1	B	1175	LEU	3.0
1	B	1160	THR	2.9
1	B	1000	LEU	2.8
1	A	1026	LYS	2.7
2	D	3	ASP	2.7
1	B	1040	THR	2.7
1	B	1089	GLN	2.7
1	B	1161	ARG	2.7
1	B	1095	THR	2.6
1	B	1173	GLU	2.6
1	A	989	LYS	2.6
1	B	1021	GLN	2.6
1	B	1005	ALA	2.6
1	B	1018	GLU	2.6
1	B	1029	VAL	2.6
1	B	1050	GLY	2.6
1	A	1028	GLN	2.5
1	B	1158	VAL	2.5
1	A	1075	TYR	2.5
1	B	1098	THR	2.5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	1074	MET	2.4
2	D	1	THR	2.4
1	B	1166	ALA	2.4
1	B	1102	SER	2.4
1	B	1163	VAL	2.4
1	B	1105	TYR	2.4
1	A	1024	ARG	2.4
1	B	1092	ARG	2.4
1	A	1147	ALA	2.4
1	B	1129	PRO	2.3
1	A	1132	ILE	2.3
1	B	1159	SER	2.3
1	A	1129	PRO	2.2
1	B	1019	THR	2.2
1	B	1143	LEU	2.2
1	A	1098	THR	2.1
1	A	1177	THR	2.1
1	B	1090	GLY	2.1
1	B	1101[A]	SER	2.1
1	B	1113	VAL	2.1
1	B	1114	ILE	2.1
1	B	1028	GLN	2.1
1	B	1039	ALA	2.0
1	A	1160	THR	2.0
1	B	1165	LYS	2.0
1	B	1096	PRO	2.0
1	A	1027	ASN	2.0
1	B	1162	GLY	2.0

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 ⓘ

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	SO4	A	1	5/5	0.87	0.19	32,34,35,35	0
4	ZN	B	2	1/1	0.96	0.08	23,23,23,23	0
4	ZN	A	1183	1/1	0.99	0.03	17,17,17,17	0

6.5 Other polymers ⓘ

There are no such residues in this entry.

W O R L D W I D E
wwPDB
PROTEIN DATA BANK