



wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 31, 2016 – 10:07 PM GMT

PDB ID : 1S6P
Title : CRYSTAL STRUCTURE OF HUMAN IMMUNODEFICIENCY VIRUS
TYPE 1 REVERSE TRANSCRIPTASE (RT) IN COMPLEX WITH
JANSSEN-R100943
Authors : Das, K.; Arnold, E.
Deposited on : 2004-01-26
Resolution : 2.90 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
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>
with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) 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

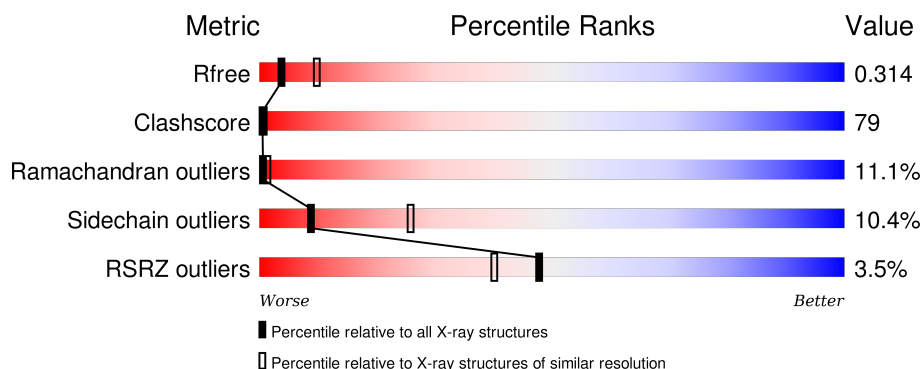
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

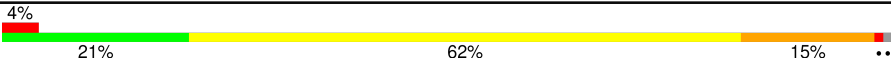
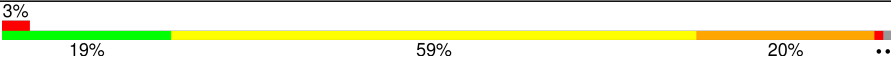
The reported resolution of this entry is 2.90 Å.

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	1451 (2.90-2.90)
Clashscore	102246	1668 (2.90-2.90)
Ramachandran outliers	100387	1630 (2.90-2.90)
Sidechain outliers	100360	1632 (2.90-2.90)
RSRZ outliers	91569	1456 (2.90-2.90)

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 $\leq 5\%$. 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	560	 4% 21% 62% 15% ..
2	B	430	 3% 19% 59% 20% ..

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 8202 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 POL polyprotein [Contains: Reverse transcriptase].

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	552	Total	C	N	O	S	18	0	0
			4498	2913	748	830	7			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	280	SER	CYS	ENGINEERED	UNP P03366

- Molecule 2 is a protein called POL polyprotein [Contains: Reverse transcriptase].

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	427	Total	C	N	O	S	17	0	0
			3529	2300	584	638	7			

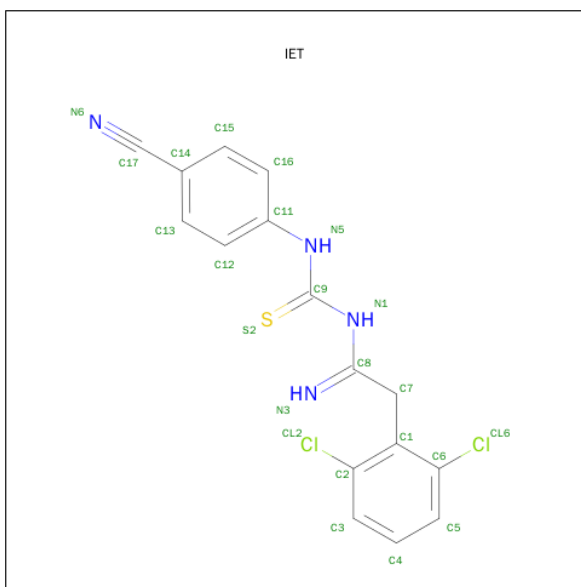
There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	280	SER	CYS	ENGINEERED	UNP P03366

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Mg	0	0
			1	1		

- Molecule 4 is 1-(4-CYANO-PHENYL)-3-[2-(2,6-DICHLORO-PHENYL)-1-IMINO-ETHYL]-THIOUREA (three-letter code: IET) (formula: C₁₆H₁₂Cl₂N₄S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	A	1	Total	C	Cl	N	S	0	0
			23	16	2	4	1		

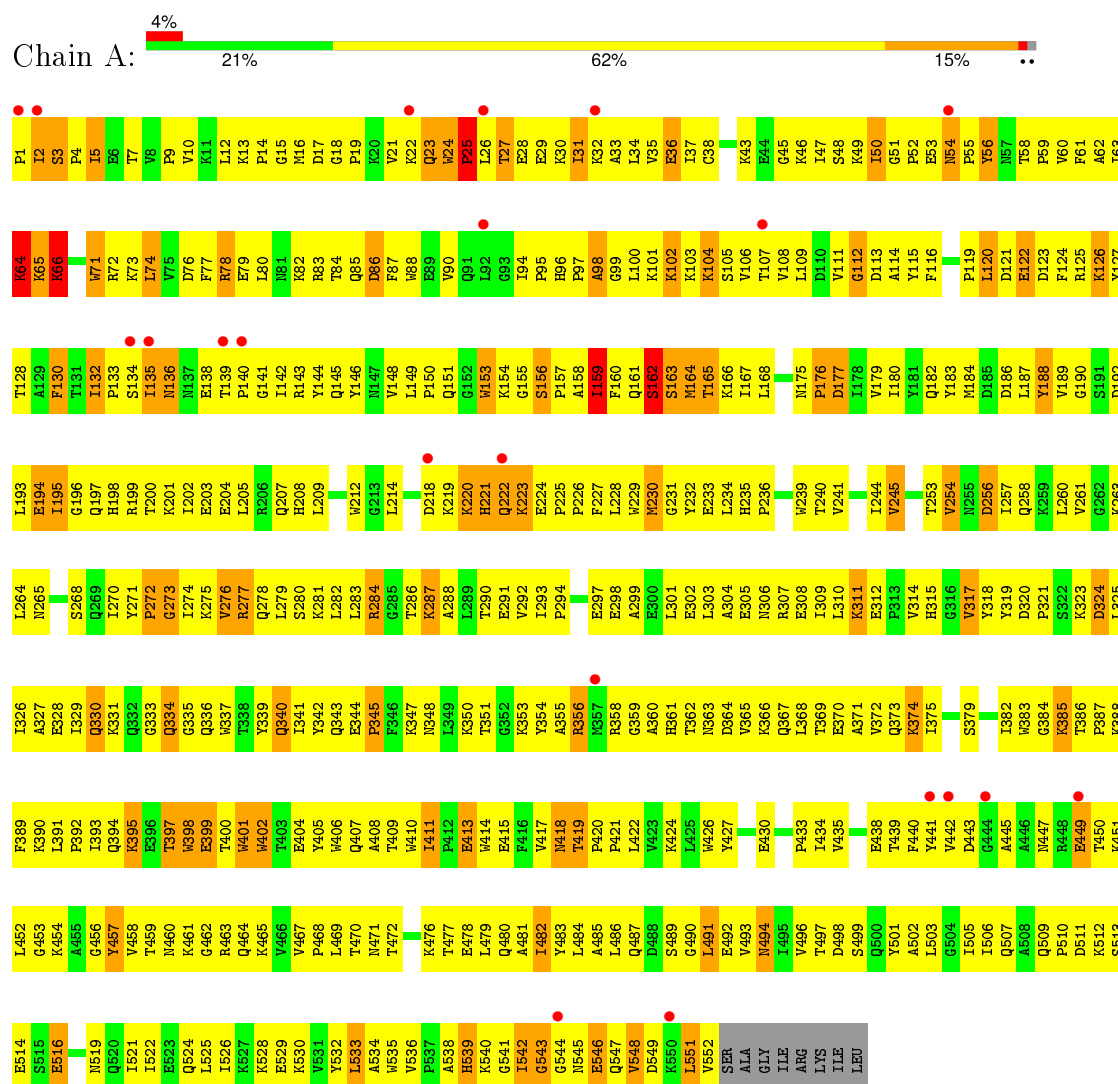
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	81	Total	O	0	0
			81	81		
5	B	70	Total	O	0	0
			70	70		

3 Residue-property plots

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: POL polyprotein [Contains: Reverse transcriptase]



K383	G384	K385	T386	P387	K388	F389	K390	L391	P392	I393	Q394	K395	K398	E399	T400	W401	W402	T403	E404	Y405	W406	T409	W410	I411	P412	F416	W417	M418	T419	P420	P421	L422	V423	K424	L425	W426	Y427	GLN	LEU	GLU																			
Y319	D320	P321	K322	K323	D324	L325	I326	A327	E328	I329	Q330	S268	K331	Q332	G335	Q336	K337	T338	Y339	Q340	I341	Y342	Q343	E344	P345	F346	L349	K350	T351	Y354	A355	R358	G359	A360	H361	N362	N363	D364	V365	K366	Q367	L368	T369	E370	A371	V372	Q373	K374	I375	T376	T377	E378	S379	I380	W381	I382			
D256	T257	K258	L259	V261	Q262	K263	L264	N265	W266	A267	S268	Q269	I270	Y271	Q272	Q273	L274	K275	V276	R277	Q278	L279	S280	K281	L282	L283	R284	Q285	L289	T290	E291	V292	L293	P294	L295	T296	E297	E298	A299	E300	L301	E302	E305	N306	R307	E308	I309	L310	K311	E312	P313	V314	R315	G316	V317	Y318			
G196	Q197	H198	R199	K201	L202	E203	E204	L205	R206	Q207	H208	L209	L210	W211	G212	G213	L214	T215	T216	D217	D218	K219	K220	H221	Q222	K223	E224	P225	P226	F227	L228	W229	M230	G231	Y232	E233	L234	H235	P236	D237	K238	W239	T240	V241	Q242	P243	I244	V245	L246	P247	E248	K249	D250	S251	W252	T253	V254	N255	
S134	I135	T139	P140	G141	L142	R143	Y144	Q145	Y146	N147	W148	L149	P150	Q151	G152	W153	K154	G155	S156	P157	A158	I159	F160	Q161	S162	S163	M164	T165	K166	L167	L168	E169	P170	F171	K172	K173	Q174	N175	P176	D177	I178	V179	I180	Y181	Q182	Y183	M184	D185	D186	L187	F188	Y189	G190	S191	D192	L193	E194	P195	I196
S68	T69	K70	W71	T72	K73	L74	V75	D76	V77	E78	L79	K80	N81	K82	R83	T84	Q85	D86	F87	H88	W89	T90	Q91	L92	H96	P97	L100	K101	K102	K103	K104	T107	V108	L109	D110	V111	G112	S113	E114	P115	Y116	F117	P118	P119	E122	N123	F124	R125	K126	Y127	F130	T131	I132	P133					

4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	226.70Å 67.40Å 104.30Å 90.00° 107.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.90 19.95 – 2.82	Depositor EDS
% Data completeness (in resolution range)	89.8 (20.00-2.90) 87.2 (19.95-2.82)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.59 (at 2.83Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.245 , 0.312 0.252 , 0.314	Depositor DCC
R_{free} test set	1584 reflections (5.11%)	DCC
Wilson B-factor (Å ²)	67.3	Xtriage
Anisotropy	0.215	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 97.7	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.29$	Xtriage
Outliers	0 of 31794 reflections	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	8202	wwPDB-VP
Average B, all atoms (Å ²)	76.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.75% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

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

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	0.67	1/4616 (0.0%)	0.90	8/6271 (0.1%)
2	B	0.66	2/3634 (0.1%)	0.95	9/4940 (0.2%)
All	All	0.67	3/8250 (0.0%)	0.92	17/11211 (0.2%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	1	0
2	B	0	1
All	All	1	1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	31	ILE	CG1-CD1	23.29	3.11	1.50
2	B	224	GLU	C-N	-5.33	1.24	1.34
2	B	225	PRO	N-CD	-5.15	1.40	1.47

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	31	ILE	CB-CG1-CD1	-20.89	55.42	113.90
1	A	72	ARG	NE-CZ-NH2	7.40	124.00	120.30
2	B	242	GLN	N-CA-C	7.12	130.24	111.00
1	A	287	LYS	O-C-N	6.82	133.61	122.70
2	B	83	ARG	NE-CZ-NH2	6.67	123.63	120.30

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	A	31	ILE	CB

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	130	PHE	Sidechain

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	4498	0	4560	697	0
2	B	3529	0	3568	615	0
3	A	1	0	0	0	0
4	A	23	0	12	5	0
5	A	81	0	0	8	0
5	B	70	0	0	5	0
All	All	8202	0	8140	1277	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 79.

The worst 5 of 1277 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:64:LYS:O	1:A:66:LYS:HG2	1.34	1.25
2:B:139:THR:CG2	2:B:140:PRO:HD2	1.66	1.24
1:A:222:GLN:O	1:A:224:GLU:HG3	1.41	1.16
1:A:497:THR:HG22	1:A:499:SER:H	1.09	1.16
2:B:358:ARG:HG2	2:B:359:GLY:H	1.09	1.16

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	550/560 (98%)	404 (74%)	92 (17%)	54 (10%)	1	2
2	B	425/430 (99%)	277 (65%)	94 (22%)	54 (13%)	0	1
All	All	975/990 (98%)	681 (70%)	186 (19%)	108 (11%)	0	1

5 of 108 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	5	ILE
1	A	25	PRO
1	A	52	PRO
1	A	65	LYS
1	A	112	GLY

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	494/500 (99%)	449 (91%)	45 (9%)	12	34
2	B	389/392 (99%)	342 (88%)	47 (12%)	6	18
All	All	883/892 (99%)	791 (90%)	92 (10%)	9	26

5 of 92 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	522	ILE

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	B	67	ASP
2	B	362	THR
1	A	533	LEU
2	B	24	TRP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 33 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	507	GLN
2	B	136	ASN
2	B	340	GLN
1	A	519	ASN
2	B	23	GLN

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 2 ligands modelled in this entry, 1 is 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
4	IET	A	701	-	22,24,24	4.00	11 (50%)	29,32,32	2.19	6 (20%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	IET	A	701	-	-	0/12/14/14	0/2/2/2

The worst 5 of 11 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	701	IET	C6-CL6	-9.16	1.51	1.73
4	A	701	IET	C2-CL2	-9.12	1.51	1.73
4	A	701	IET	C7-C8	-9.11	1.39	1.50
4	A	701	IET	C9-N1	-2.92	1.34	1.39
4	A	701	IET	C4-C3	2.03	1.43	1.38

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	701	IET	C9-N1-C8	-5.83	121.42	128.84
4	A	701	IET	C7-C1-C6	-3.19	116.63	121.62
4	A	701	IET	S2-C9-N5	-2.54	116.50	124.22
4	A	701	IET	C7-C1-C2	3.04	126.39	121.62
4	A	701	IET	N5-C9-N1	5.03	123.15	115.28

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	701	IET	5	0

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 [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	552/560 (98%)	-0.09	21 (3%)	44 37	37, 83, 109, 111	5 (0%)
2	B	427/430 (99%)	-0.16	13 (3%)	54 47	19, 69, 110, 111	4 (0%)
All	All	979/990 (98%)	-0.12	34 (3%)	48 40	19, 78, 110, 111	9 (0%)

The worst 5 of 34 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	231	GLY	6.8
1	A	444	GLY	5.8
2	B	3	SER	4.1
1	A	107	THR	3.7
2	B	4	PRO	3.6

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](#)

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, 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	LLDF	B-factors(\AA^2)	Q<0.9
4	IET	A	701	23/23	0.86	0.24	0.34	63,78,84,92	0
3	MG	A	601	1/1	0.90	0.14	-	79,79,79,79	0

6.5 Other polymers [i](#)

There are no such residues in this entry.