wwPDB X-ray Structure Validation Summary Report (i)

Jul 21, 2024 – 01:47 am BST

PDB ID	:	8QKR
Title	:	Plasmodium falciparum reticulocyte-binding protein homologue 5 (PfRH5) bound to R5.251
Authors	:	Wright, N.D.; Barrett, J.R.; Bradshaw, W.J.; Paterson, N.G.; MacLean, E.M.; Ferreira, L.; McHugh, K.; Koekemoer, L.; Draper, S.J.
Deposited on	:	2023-09-16
Resolution	:	3.23 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 (1)) were used in the production of this report:

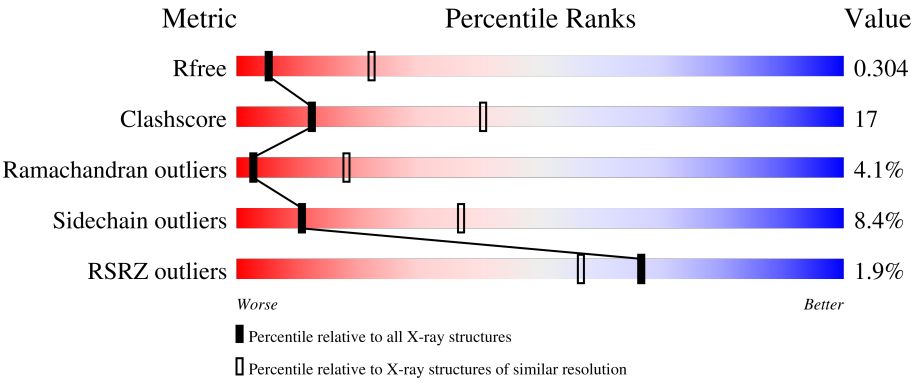
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 3.23 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	1619 (3.28-3.20)
Clashscore	141614	1755 (3.28-3.20)
Ramachandran outliers	138981	1728 (3.28-3.20)
Sidechain outliers	138945	1727 (3.28-3.20)
RSRZ outliers	127900	1567 (3.28-3.20)

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	A	325	50% 34% 6% 9%
1	D	325	55% 31% 9%
2	B	234	62% 31%
2	E	234	62% 28% 5%
3	C	220	61% 32% 5%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	220	

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 22982 atoms, of which 11213 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 Reticulocyte-binding protein-like protein 5.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	295	Total	C	H	N	O	S	0	0	0
			4947	1613	2447	419	453	15			
1	D	295	Total	C	H	N	O	S	0	2	0
			4972	1619	2461	423	454	15			

There are 102 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	77	ALA	THR	conflict	UNP A0A8F2YHP6
A	?	-	ASP	deletion	UNP A0A8F2YHP6
A	?	-	ASP	deletion	UNP A0A8F2YHP6
A	?	-	SER	deletion	UNP A0A8F2YHP6
A	?	-	TYR	deletion	UNP A0A8F2YHP6
A	?	-	ARG	deletion	UNP A0A8F2YHP6
A	?	-	TYR	deletion	UNP A0A8F2YHP6
A	?	-	ASP	deletion	UNP A0A8F2YHP6
A	?	-	ILE	deletion	UNP A0A8F2YHP6
A	?	-	SER	deletion	UNP A0A8F2YHP6
A	?	-	GLU	deletion	UNP A0A8F2YHP6
A	?	-	GLU	deletion	UNP A0A8F2YHP6
A	?	-	ILE	deletion	UNP A0A8F2YHP6
A	?	-	ASP	deletion	UNP A0A8F2YHP6
A	?	-	ASP	deletion	UNP A0A8F2YHP6
A	?	-	LYS	deletion	UNP A0A8F2YHP6
A	?	-	SER	deletion	UNP A0A8F2YHP6
A	?	-	GLU	deletion	UNP A0A8F2YHP6
A	?	-	GLU	deletion	UNP A0A8F2YHP6
A	?	-	THR	deletion	UNP A0A8F2YHP6
A	?	-	ASP	deletion	UNP A0A8F2YHP6
A	?	-	ASP	deletion	UNP A0A8F2YHP6
A	?	-	GLU	deletion	UNP A0A8F2YHP6
A	?	-	THR	deletion	UNP A0A8F2YHP6
A	?	-	GLU	deletion	UNP A0A8F2YHP6

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	?	-	GLU	deletion	UNP A0A8F2YHP6
A	?	-	VAL	deletion	UNP A0A8F2YHP6
A	?	-	GLU	deletion	UNP A0A8F2YHP6
A	?	-	ASP	deletion	UNP A0A8F2YHP6
A	?	-	SER	deletion	UNP A0A8F2YHP6
A	?	-	ILE	deletion	UNP A0A8F2YHP6
A	?	-	GLN	deletion	UNP A0A8F2YHP6
A	?	-	ASP	deletion	UNP A0A8F2YHP6
A	?	-	THR	deletion	UNP A0A8F2YHP6
A	?	-	ASP	deletion	UNP A0A8F2YHP6
A	?	-	SER	deletion	UNP A0A8F2YHP6
A	?	-	ASN	deletion	UNP A0A8F2YHP6
A	?	-	HIS	deletion	UNP A0A8F2YHP6
A	?	-	THR	deletion	UNP A0A8F2YHP6
A	?	-	PRO	deletion	UNP A0A8F2YHP6
A	?	-	SER	deletion	UNP A0A8F2YHP6
A	?	-	ASN	deletion	UNP A0A8F2YHP6
A	?	-	LYS	deletion	UNP A0A8F2YHP6
A	?	-	LYS	deletion	UNP A0A8F2YHP6
A	?	-	LYS	deletion	UNP A0A8F2YHP6
A	?	-	ASN	deletion	UNP A0A8F2YHP6
A	?	-	ASP	deletion	UNP A0A8F2YHP6
A	?	-	LEU	deletion	UNP A0A8F2YHP6
A	?	-	MET	deletion	UNP A0A8F2YHP6
A	?	-	ASN	deletion	UNP A0A8F2YHP6
A	111	ALA	THR	conflict	UNP A0A8F2YHP6
D	77	ALA	THR	conflict	UNP A0A8F2YHP6
D	?	-	ASP	deletion	UNP A0A8F2YHP6
D	?	-	ASP	deletion	UNP A0A8F2YHP6
D	?	-	SER	deletion	UNP A0A8F2YHP6
D	?	-	TYR	deletion	UNP A0A8F2YHP6
D	?	-	ARG	deletion	UNP A0A8F2YHP6
D	?	-	TYR	deletion	UNP A0A8F2YHP6
D	?	-	ASP	deletion	UNP A0A8F2YHP6
D	?	-	ILE	deletion	UNP A0A8F2YHP6
D	?	-	SER	deletion	UNP A0A8F2YHP6
D	?	-	GLU	deletion	UNP A0A8F2YHP6
D	?	-	GLU	deletion	UNP A0A8F2YHP6
D	?	-	ILE	deletion	UNP A0A8F2YHP6
D	?	-	ASP	deletion	UNP A0A8F2YHP6
D	?	-	ASP	deletion	UNP A0A8F2YHP6
D	?	-	LYS	deletion	UNP A0A8F2YHP6

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	?	-	SER	deletion	UNP A0A8F2YHP6
D	?	-	GLU	deletion	UNP A0A8F2YHP6
D	?	-	GLU	deletion	UNP A0A8F2YHP6
D	?	-	THR	deletion	UNP A0A8F2YHP6
D	?	-	ASP	deletion	UNP A0A8F2YHP6
D	?	-	ASP	deletion	UNP A0A8F2YHP6
D	?	-	GLU	deletion	UNP A0A8F2YHP6
D	?	-	THR	deletion	UNP A0A8F2YHP6
D	?	-	GLU	deletion	UNP A0A8F2YHP6
D	?	-	GLU	deletion	UNP A0A8F2YHP6
D	?	-	VAL	deletion	UNP A0A8F2YHP6
D	?	-	GLU	deletion	UNP A0A8F2YHP6
D	?	-	ASP	deletion	UNP A0A8F2YHP6
D	?	-	SER	deletion	UNP A0A8F2YHP6
D	?	-	ILE	deletion	UNP A0A8F2YHP6
D	?	-	GLN	deletion	UNP A0A8F2YHP6
D	?	-	ASP	deletion	UNP A0A8F2YHP6
D	?	-	THR	deletion	UNP A0A8F2YHP6
D	?	-	ASP	deletion	UNP A0A8F2YHP6
D	?	-	SER	deletion	UNP A0A8F2YHP6
D	?	-	ASN	deletion	UNP A0A8F2YHP6
D	?	-	HIS	deletion	UNP A0A8F2YHP6
D	?	-	THR	deletion	UNP A0A8F2YHP6
D	?	-	PRO	deletion	UNP A0A8F2YHP6
D	?	-	SER	deletion	UNP A0A8F2YHP6
D	?	-	ASN	deletion	UNP A0A8F2YHP6
D	?	-	LYS	deletion	UNP A0A8F2YHP6
D	?	-	LYS	deletion	UNP A0A8F2YHP6
D	?	-	LYS	deletion	UNP A0A8F2YHP6
D	?	-	ASN	deletion	UNP A0A8F2YHP6
D	?	-	ASP	deletion	UNP A0A8F2YHP6
D	?	-	LEU	deletion	UNP A0A8F2YHP6
D	?	-	MET	deletion	UNP A0A8F2YHP6
D	?	-	ASN	deletion	UNP A0A8F2YHP6
D	111	ALA	THR	conflict	UNP A0A8F2YHP6

- Molecule 2 is a protein called R5251VHCH.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	224	Total	C	H	N	O	S	0	2	0
			3307	1078	1599	285	337	8			
2	E	224	Total	C	H	N	O	S	0	6	0
			3354	1095	1626	286	339	8			

- Molecule 3 is a protein called R5251VLCL.

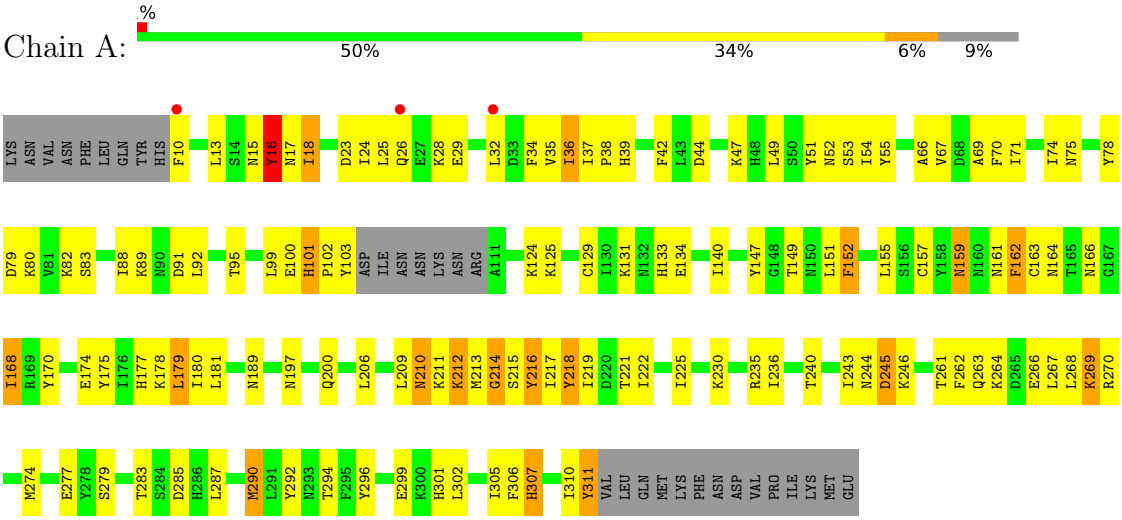
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	C	215	Total	C	H	N	O	S	0	0	0
			3201	1035	1540	281	339	6			
3	F	215	Total	C	H	N	O	S	0	0	0
			3201	1035	1540	281	339	6			

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.

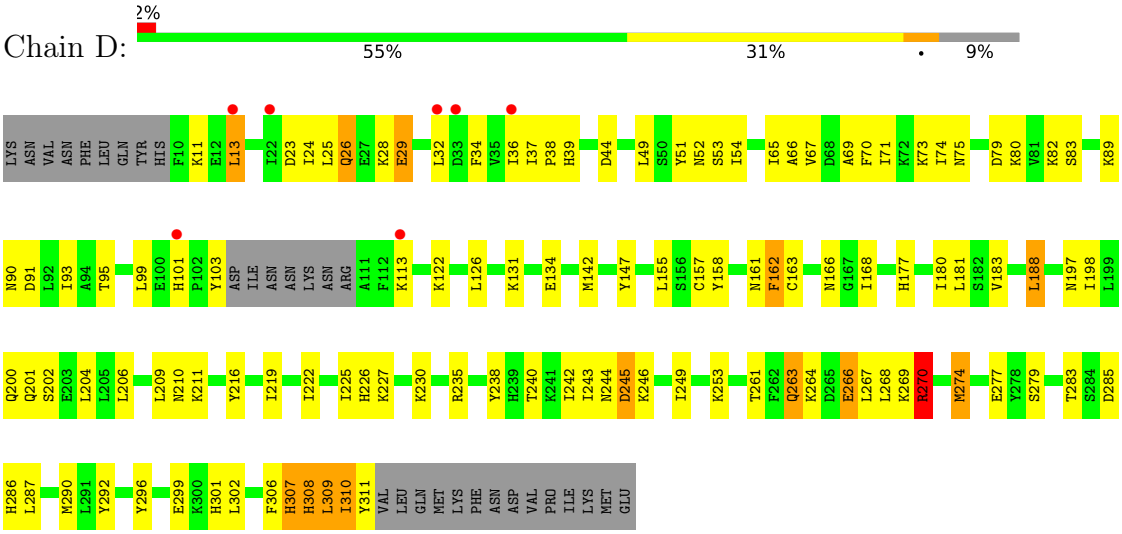
• Molecule 1: Reticulocyte-binding protein-like protein 5

Chain A:



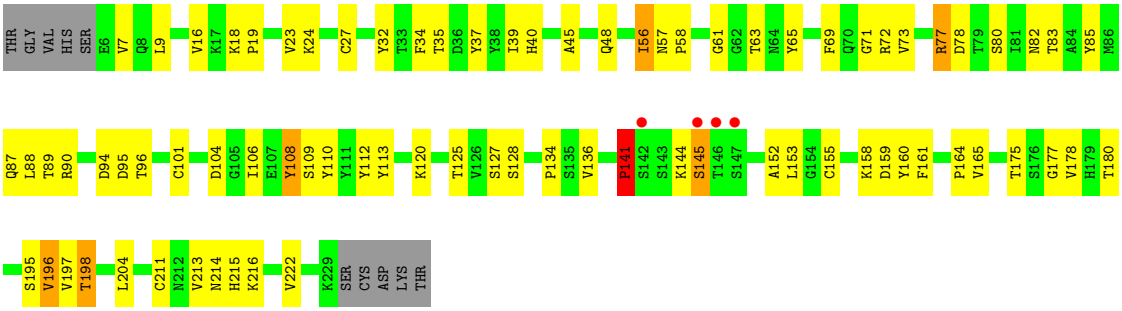
• Molecule 1: Reticulocyte-binding protein-like protein 5

Chain D:



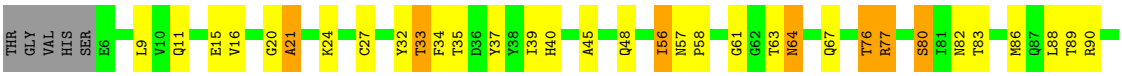
• Molecule 2: R5251VHCH

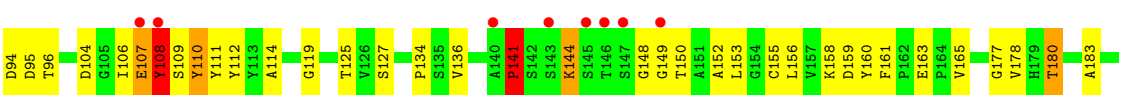
Chain B:



• Molecule 2: R5251VHCH

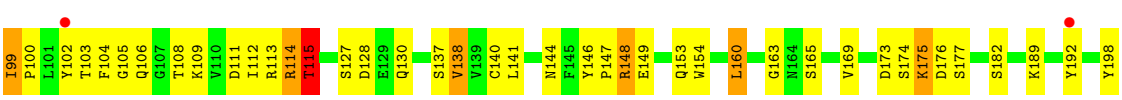
Chain E:





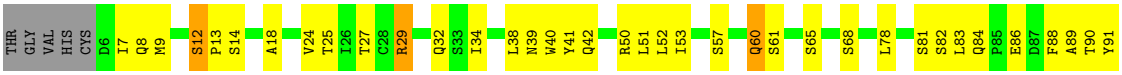
• Molecule 3: R5251VLCL

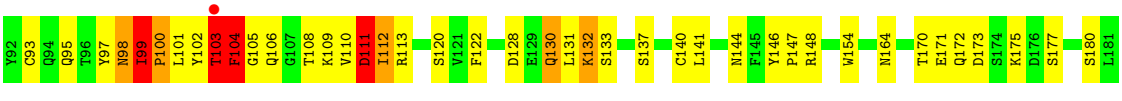
Chain C:

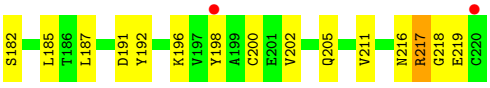


• Molecule 3: R5251VLCL

Chain F:







4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	69.93Å 92.02Å 92.49Å 60.53° 72.13° 79.82°	Depositor
Resolution (Å)	59.10 – 3.23 59.10 – 3.23	Depositor EDS
% Data completeness (in resolution range)	98.5 (59.10-3.23) 89.4 (59.10-3.23)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.11 (at 3.26Å)	Xtriage
Refinement program	REFMAC 5.8.0425	Depositor
R, R_{free}	0.254 , 0.327 0.236 , 0.304	Depositor DCC
R_{free} test set	1403 reflections (4.66%)	wwPDB-VP
Wilson B-factor (Å^2)	67.9	Xtriage
Anisotropy	0.168	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 24.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.37$, $< L^2 > = 0.20$	Xtriage
Estimated twinning fraction	0.189 for -h,k-l,-l	Xtriage
Reported twinning fraction	0.791 for H, K, L 0.209 for -H, K-L, -L	Depositor
Outliers	0 of 30108 reflections	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	22982	wwPDB-VP
Average B, all atoms (Å^2)	73.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.00% 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	A	0.34	0/2553	0.79	2/3424 (0.1%)
1	D	0.35	0/2572	0.80	3/3449 (0.1%)
2	B	0.36	0/1756	0.82	1/2393 (0.0%)
2	E	0.36	0/1796	0.83	0/2449
3	C	0.36	0/1696	0.79	0/2304
3	F	0.36	0/1696	0.79	0/2304
All	All	0.36	0/12069	0.80	6/16323 (0.0%)

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
3	C	0	4
3	F	0	4
All	All	0	8

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	142	MET	CG-SD-CE	6.43	110.48	100.20
1	A	274	MET	CG-SD-CE	5.80	109.47	100.20
1	A	290	MET	CG-SD-CE	5.74	109.39	100.20
1	D	270	ARG	NE-CZ-NH1	5.62	123.11	120.30
2	B	108	TYR	N-CA-CB	5.31	120.15	110.60

There are no chirality outliers.

5 of 8 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	C	103	THR	Peptide
3	C	12	SER	Peptide
3	C	98	ASN	Peptide
3	C	99	ILE	Peptide
3	F	12	SER	Peptide

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	2500	2447	2519	83	0
1	D	2511	2461	2521	82	0
2	B	1708	1599	1669	55	0
2	E	1728	1626	1687	65	0
3	C	1661	1540	1608	43	0
3	F	1661	1540	1608	72	0
All	All	11769	11213	11612	386	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 17.

The worst 5 of 386 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)
3:F:7:ILE:HB	3:F:102:TYR:CE2	1.89	1.06
3:F:95:GLN:HG2	3:F:102:TYR:CE2	2.05	0.92
1:A:24:ILE:HG22	1:A:32:LEU:HD11	1.53	0.88
1:D:24:ILE:HG22	1:D:32:LEU:HD11	1.56	0.87
1:A:102:PRO:C	1:A:218:TYR:OH	2.15	0.84

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	291/325 (90%)	253 (87%)	25 (9%)	13 (4%)	2 16
1	D	293/325 (90%)	258 (88%)	29 (10%)	6 (2%)	7 34
2	B	224/234 (96%)	193 (86%)	29 (13%)	2 (1%)	17 52
2	E	228/234 (97%)	196 (86%)	24 (10%)	8 (4%)	3 21
3	C	213/220 (97%)	171 (80%)	27 (13%)	15 (7%)	1 7
3	F	213/220 (97%)	171 (80%)	26 (12%)	16 (8%)	1 6
All	All	1462/1558 (94%)	1242 (85%)	160 (11%)	60 (4%)	3 18

5 of 60 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	16	TYR
1	A	162	PHE
1	A	210	ASN
1	A	213	MET
1	A	215	SER

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	284/314 (90%)	263 (93%)	21 (7%)	13 43
1	D	286/314 (91%)	262 (92%)	24 (8%)	11 37
2	B	191/198 (96%)	178 (93%)	13 (7%)	16 46

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
2	E	195/198 (98%)	177 (91%)	18 (9%)	9 32
3	C	191/195 (98%)	172 (90%)	19 (10%)	8 29
3	F	191/195 (98%)	175 (92%)	16 (8%)	11 37
All	All	1338/1414 (95%)	1227 (92%)	111 (8%)	11 38

5 of 111 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	82	LYS
3	F	141	LEU
1	D	263	GLN
3	F	132	LYS
3	F	68	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 30 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	160	ASN
3	F	166	GLN
1	D	210	ASN
3	F	195	HIS
3	F	42	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 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	A	295/325 (90%)	-0.08	3 (1%)	82	75	40, 68, 111, 144	0
1	D	295/325 (90%)	0.04	7 (2%)	59	47	35, 72, 114, 166	0
2	B	224/234 (95%)	-0.08	4 (1%)	68	58	27, 62, 106, 171	0
2	E	224/234 (95%)	0.16	8 (3%)	42	31	32, 62, 120, 233	0
3	C	215/220 (97%)	0.18	3 (1%)	75	66	38, 75, 118, 135	0
3	F	215/220 (97%)	0.07	3 (1%)	75	66	44, 74, 114, 139	0
All	All	1468/1558 (94%)	0.04	28 (1%)	66	57	27, 68, 117, 233	0

The worst 5 of 28 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	E	146	THR	7.7
2	E	145	SER	6.1
1	D	32	LEU	5.5
3	F	220	CYS	5.2
2	B	142	SER	5.1

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.