wwPDB X-ray Structure Validation Summary Report (i)

Nov 10, 2024 – 05:23 PM EST

PDB ID	:	1KTK
Title	:	Complex of Streptococcal pyrogenic enterotoxin C (SpeC) with a human T cell receptor beta chain (Vbeta2.1)
Authors	:	Sundberg, E.J.; Li, H.; Llera, A.S.; McCormick, J.K.; Tormo, J.; Karjalainen, K.; Schlievert, P.M.; Mariuzza, R.A.
Deposited on	:	2002-01-16
Resolution	:	3.00 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:

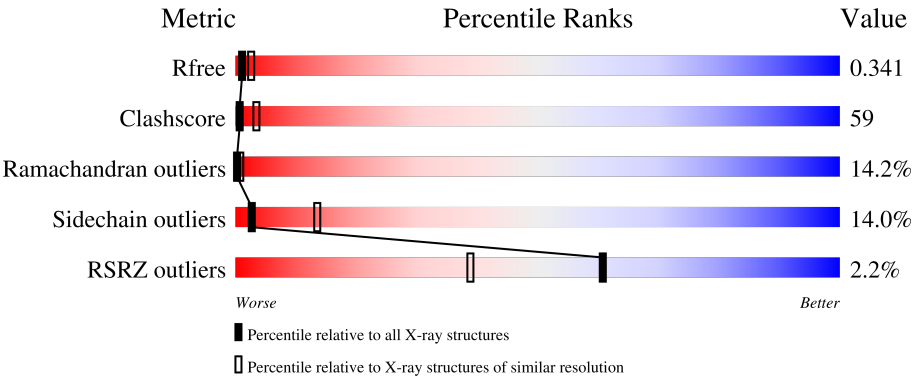
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.20.1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

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 3.00 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.



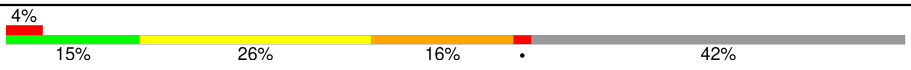
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	2511 (3.00-3.00)
Clashscore	180529	2866 (3.00-3.00)
Ramachandran outliers	177936	2778 (3.00-3.00)
Sidechain outliers	177891	2781 (3.00-3.00)
RSRZ outliers	164620	2523 (3.00-3.00)

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	208	% 31% 50% 14%
1	B	208	% 24% 59% 12%
1	C	208	% 32% 57% 10%
1	D	208	% 36% 50% 10% 5%
2	E	247	3% 21% 46% 27% 7%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	247	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 9511 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 Exotoxin type C.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	203	Total	C	N	O	S	0	0	0
			1664	1062	275	323	4			
1	B	200	Total	C	N	O	S	0	0	0
			1624	1042	267	311	4			
1	C	208	Total	C	N	O	S	0	0	0
			1699	1083	279	333	4			
1	D	198	Total	C	N	O	S	0	0	0
			1601	1028	267	303	3			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	26	ASP	ASN	conflict	UNP P13380
B	26	ASP	ASN	conflict	UNP P13380
C	26	ASP	ASN	conflict	UNP P13380
D	26	ASP	ASN	conflict	UNP P13380

- Molecule 2 is a protein called T-cell receptor beta chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	247	Total	C	N	O	S	0	0	0
			1879	1178	324	370	7			
2	F	144	Total	C	N	O	S	0	0	0
			1044	656	180	203	5			

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	10	ARG	TRP	conflict	UNP P01850
E	13	ALA	CYS	conflict	UNP P01850
E	50	ALA	ASN	conflict	UNP P01850
E	95	LEU	-	insertion	UNP P01850

Continued on next page...

Continued from previous page...

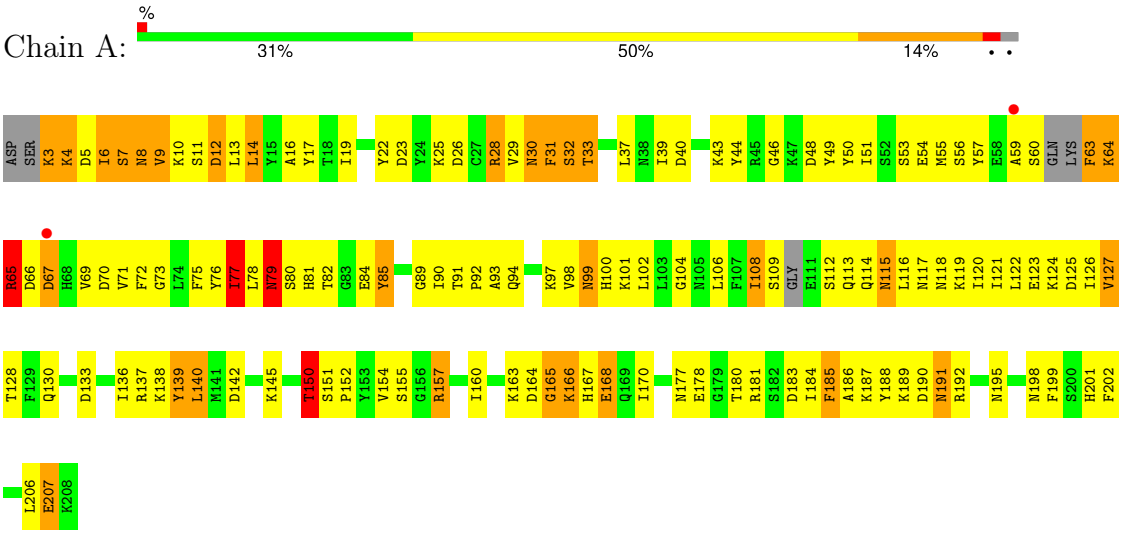
Chain	Residue	Modelled	Actual	Comment	Reference
E	96	ALA	ARG	conflict	UNP P01850
E	99	GLY	GLU	conflict	UNP P01850
E	99	GLY	THR	conflict	UNP P01850
E	101	SER	-	insertion	UNP P01850
E	102	THR	-	insertion	UNP P01850
E	?	-	PRO	deletion	UNP P01850
E	?	-	LYS	deletion	UNP P01850
E	?	-	ASN	deletion	UNP P01850
E	105	THR	GLU	conflict	UNP P01850
E	107	TYR	PHE	conflict	UNP P01850
E	?	-	VAL	deletion	UNP P01850
E	191	ALA	CYS	conflict	UNP P01850
F	10	ARG	TRP	conflict	UNP P01850
F	13	ALA	CYS	conflict	UNP P01850
F	50	ALA	ASN	conflict	UNP P01850
F	96	LEU	-	insertion	UNP P01850
F	97	ALA	ARG	conflict	UNP P01850
F	98	GLY	GLU	conflict	UNP P01850
F	100	GLY	THR	conflict	UNP P01850
F	102	SER	-	insertion	UNP P01850
F	103	THR	-	insertion	UNP P01850
F	?	-	PRO	deletion	UNP P01850
F	?	-	LYS	deletion	UNP P01850
F	?	-	ASN	deletion	UNP P01850
F	105	THR	GLU	conflict	UNP P01850
F	107	TYR	PHE	conflict	UNP P01850
F	?	-	VAL	deletion	UNP P01850
F	191	ALA	CYS	conflict	UNP P01850

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.

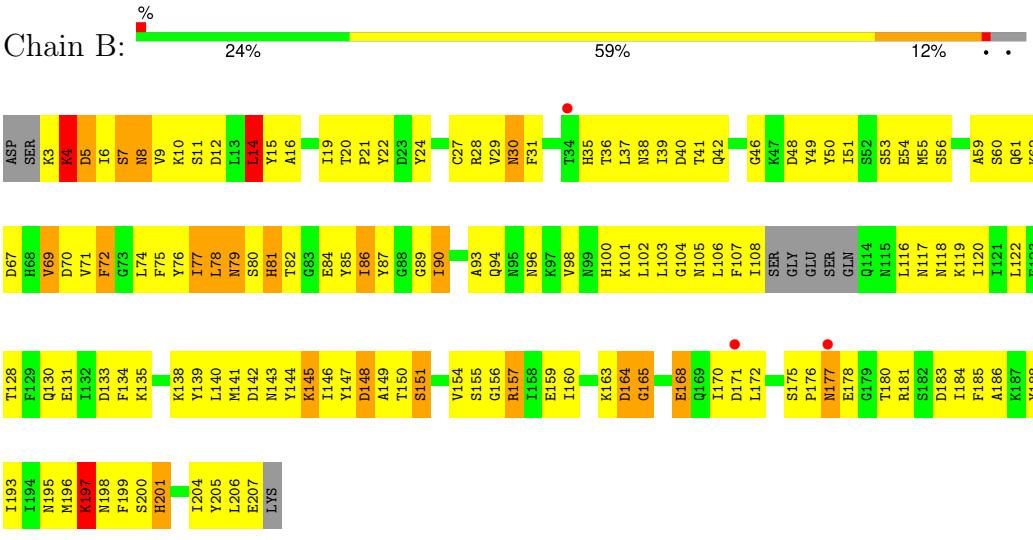
• Molecule 1: Exotoxin type C

Chain A:



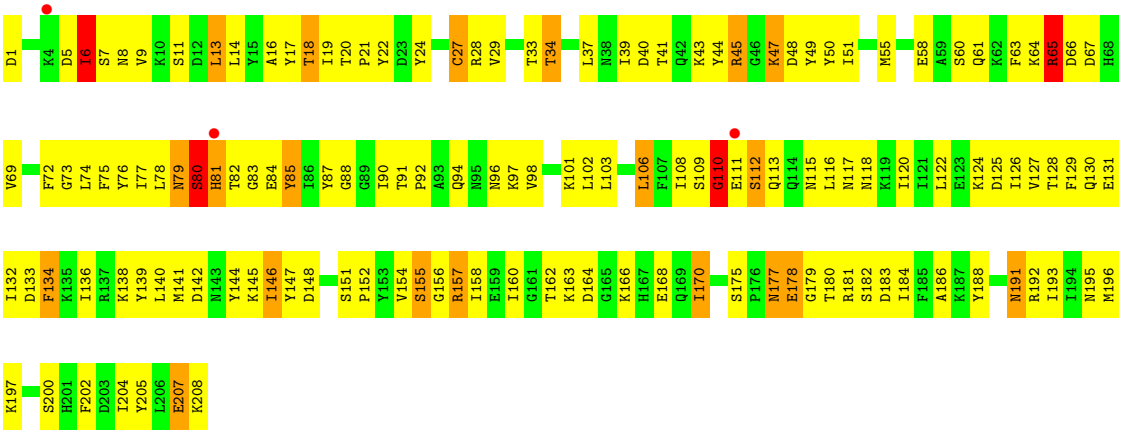
• Molecule 1: Exotoxin type C

Chain B:



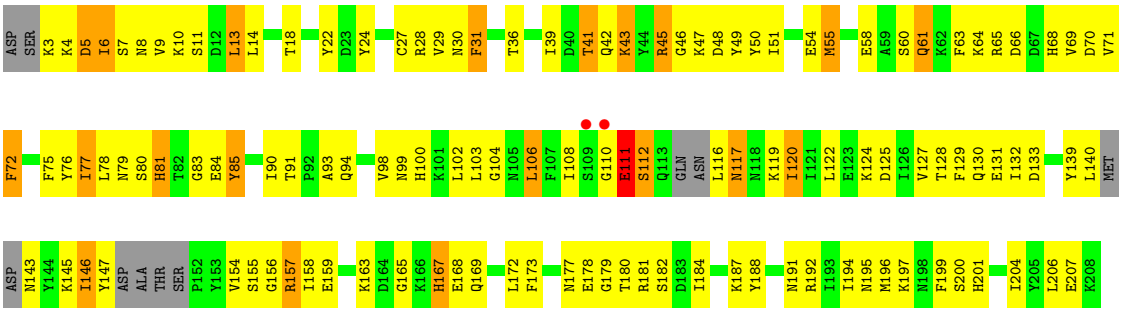
• Molecule 1: Exotoxin type C

Chain C:



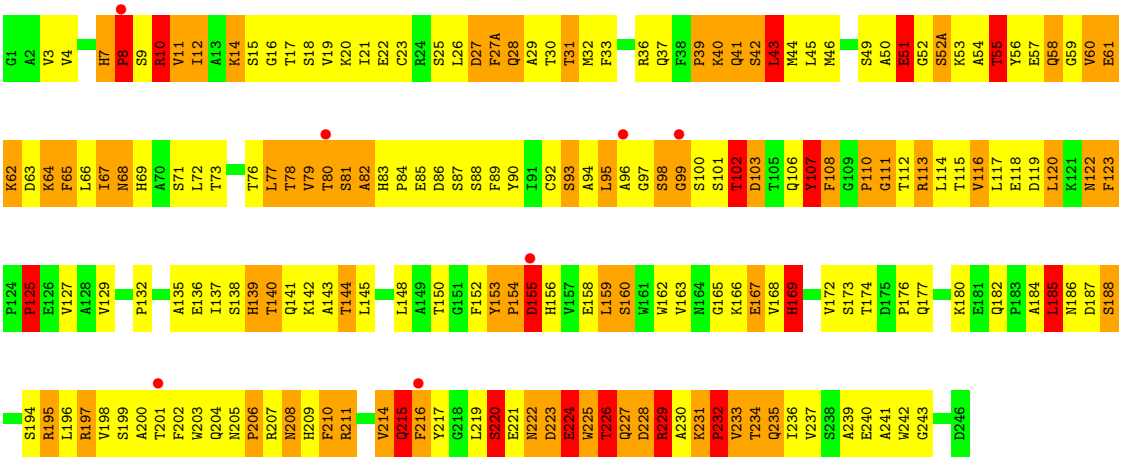
• Molecule 1: Exotoxin type C

Chain D:



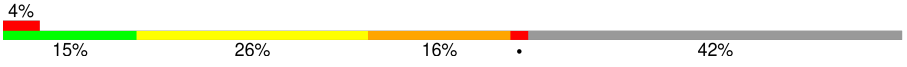
• Molecule 2: T-cell receptor beta chain

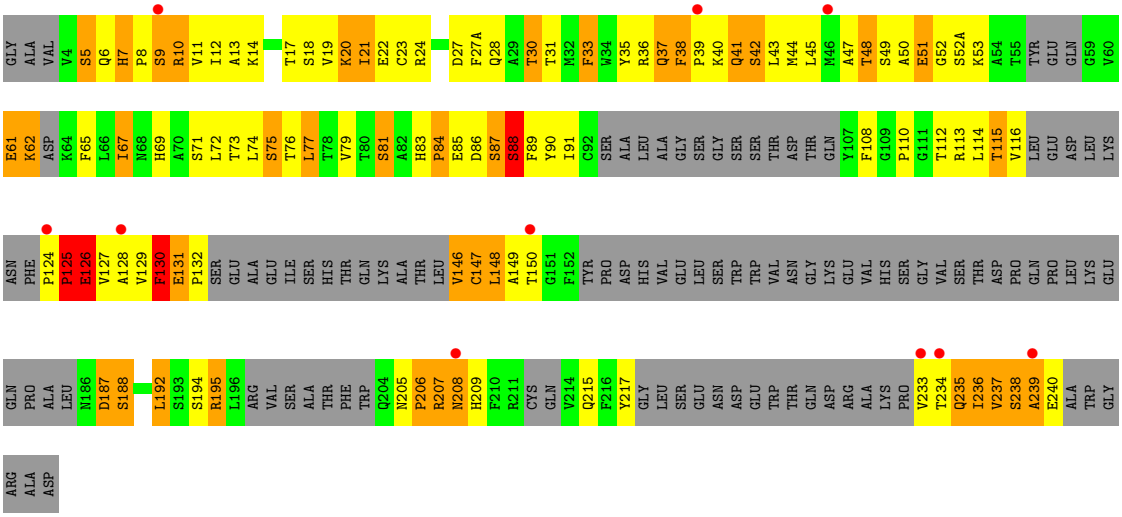
Chain E:



• Molecule 2: T-cell receptor beta chain

Chain F:




GLY ALA VAL V4 S5 Q6 H7 P8 S9 R10 V11 I12 A13 K14 T17 S18 V19 K20 I21 E22 C23 R24 D27 F27A Q28 A29 T30 T31 M32 F33 W34 Y35 R36 Q37 F38 P39 K40 Q41 S42 L43 M44 L45 M46 A47 T48 S49 A50 E51 G52 S52A K53 A54 T55 TYR GLU GLN G59 V60
E61 K62 ASP K64 F65 L66 I67 N68 H69 A70 S71 L72 T73 L74 S75 T76 L77 T78 V79 T80 S81 A82 H83 P84 E85 D86 S87 S88 F89 Y90 I91 C92 SER ALA LEU ALA GLY SER GLY SER SER THR ASP THR GLN Y107 F108 G109 P110 G111 T112 R113 L114 T115 V116 LEU GLU ASP LEU LYS
ASN PHE P124 P125 E126 V127 A128 V129 F130 E131 P132 SER GLU ALA GLU ILE SER HIS THR GLN LYS ALA THR LEU V146 C147 L148 A149 T150 G151 F152 TYR PRO ASP HIS VAL GLU LEU SER TRP TRP VAL ASN GLY LYS GLU VAL HIS SER GLY VAL SER THR ASP PRO GLN PRO LEU LYS GLU
GLN PRO ALA LEU N186 D187 S188 L192 S193 S194 R195 L196 ARG VAL SER ALA THR PHE TRP Q204 N205 P206 R207 N208 H209 F210 R211 CYS GLN V214 Q215 F216 Y217 GLY LEU SER GLU ASN ASP GLU TRP THR GLN ASP ARG ALA LYS PRO V233 T234 Q235 I236 V237 S238 A239 E240 ALA TRP GLY
ARG ALA ASP

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	57.38Å 146.76Å 135.69Å 90.00° 98.61° 90.00°	Depositor
Resolution (Å)	6.00 – 3.00 6.00 – 3.00	Depositor EDS
% Data completeness (in resolution range)	94.1 (6.00-3.00) 85.4 (6.00-3.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	5.09 (at 3.01Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.326 , 0.334 0.334 , 0.341	Depositor DCC
R_{free} test set	1941 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å^2)	78.6	Xtriage
Anisotropy	0.543	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.52 , 115.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.82	EDS
Total number of atoms	9511	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 2.93% 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	0.65	0/1697	0.88	4/2282 (0.2%)
1	B	0.48	0/1659	0.80	3/2240 (0.1%)
1	C	0.64	1/1735 (0.1%)	0.85	3/2338 (0.1%)
1	D	0.49	0/1633	0.82	2/2197 (0.1%)
2	E	0.55	0/1924	1.02	11/2620 (0.4%)
2	F	0.48	0/1055	0.96	5/1421 (0.4%)
All	All	0.56	1/9703 (0.0%)	0.89	28/13098 (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
2	F	0	1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	27	CYS	CB-SG	-5.29	1.73	1.81

The worst 5 of 28 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	E	7	HIS	C-N-CD	-15.24	87.06	120.60
2	F	85	GLU	N-CA-C	-9.54	85.23	111.00
1	B	4	LYS	N-CA-C	8.53	134.03	111.00
1	D	80	SER	N-CA-C	-8.38	88.37	111.00
1	A	79	ASN	N-CA-C	7.51	131.29	111.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	F	130	PHE	Sidechain

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	1664	0	1608	183	0
1	B	1624	0	1554	212	0
1	C	1699	0	1634	143	0
1	D	1601	0	1539	158	0
2	E	1879	0	1775	308	1
2	F	1044	0	965	144	0
All	All	9511	0	9075	1104	1

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 59.

The worst 5 of 1104 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)
2:E:57:GLU:OE2	2:E:63:ASP:HB2	1.43	1.17
1:A:180:THR:HG22	2:E:54:ALA:HB2	1.30	1.10
2:E:15:SER:HA	2:E:82:ALA:O	1.52	1.09
2:E:79:VAL:HG13	2:E:80:THR:H	1.17	1.08
2:F:86:ASP:HB3	2:F:115:THR:HA	1.15	1.08

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:E:141:GLN:OE1	2:E:223:ASP:OD2[1_655]	2.18	0.02

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	197/208 (95%)	145 (74%)	33 (17%)	19 (10%)	0 2
1	B	196/208 (94%)	146 (74%)	32 (16%)	18 (9%)	0 2
1	C	206/208 (99%)	152 (74%)	37 (18%)	17 (8%)	0 3
1	D	190/208 (91%)	147 (77%)	25 (13%)	18 (10%)	0 2
2	E	245/247 (99%)	143 (58%)	42 (17%)	60 (24%)	0 0
2	F	122/247 (49%)	59 (48%)	31 (25%)	32 (26%)	0 0
All	All	1156/1326 (87%)	792 (68%)	200 (17%)	164 (14%)	0 1

5 of 164 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	6	ILE
1	A	9	VAL
1	A	65	ARG
1	A	66	ASP
1	A	80	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	183/192 (95%)	153 (84%)	30 (16%)	2 9
1	B	175/192 (91%)	156 (89%)	19 (11%)	5 22
1	C	186/192 (97%)	166 (89%)	20 (11%)	5 22

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	D	171/192 (89%)	157 (92%)	14 (8%)	9 34
2	E	201/213 (94%)	162 (81%)	39 (19%)	1 6
2	F	108/213 (51%)	87 (81%)	21 (19%)	1 6
All	All	1024/1194 (86%)	881 (86%)	143 (14%)	3 13

5 of 143 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	E	222	ASN
2	E	229	ARG
2	F	67	ILE
1	B	197	LYS
1	B	193	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 38 such sidechains are listed below:

Mol	Chain	Res	Type
2	E	37	GLN
2	E	215	GLN
2	E	68	ASN
2	E	156	HIS
2	F	68	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 oligosaccharides 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
2	F	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	F	129:VAL	C	130:PHE	N	2.76

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	203/208 (97%)	-0.58	2 (0%)	79	60	29, 56, 87, 119	0
1	B	200/208 (96%)	-0.24	3 (1%)	71	50	49, 81, 103, 124	0
1	C	208/208 (100%)	-0.60	3 (1%)	73	52	34, 57, 84, 104	0
1	D	198/208 (95%)	-0.27	2 (1%)	79	60	50, 74, 110, 130	0
2	E	247/247 (100%)	-0.02	7 (2%)	55	33	42, 80, 119, 130	0
2	F	144/247 (58%)	0.58	10 (6%)	24	14	69, 112, 137, 149	0
All	All	1200/1326 (90%)	-0.22	27 (2%)	61	39	29, 74, 123, 149	0

The worst 5 of 27 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	9	SER	4.0
1	D	110	GLY	3.6
2	F	239	ALA	3.3
2	F	208	ASN	3.2
2	F	233	VAL	3.2

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.