wwPDB X-ray Structure Validation Summary Report (i)

Mar 23, 2016 – 12:16 AM EDT

PDB ID	:	4YO5
Title	:	EAEC T6SS TssA-Cterminus
Authors	:	Durand, E.; Zoued, A.; Spinelli, S.; Douzi, B.; Brunet, Y.R.; Bebeacua, C.; Legrand, P.; Journet, L.; Mignot, T.; Cambillau, C.; Cascales, E.
Deposited on	:	2015-03-11
Resolution	:	3.35 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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

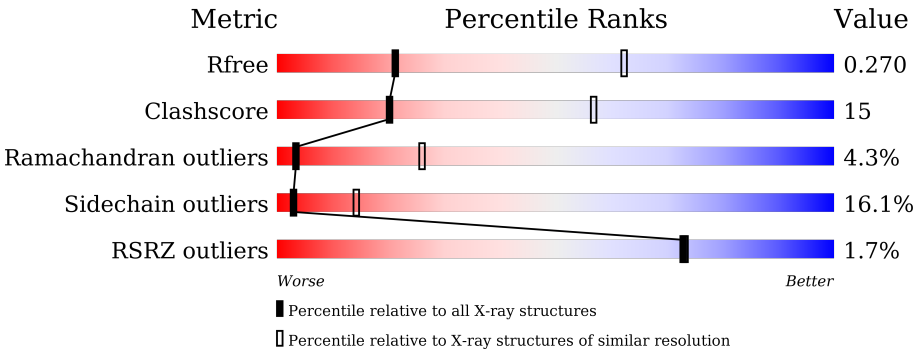
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027107
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0122
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027107

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 3.35 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.



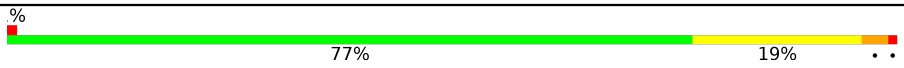
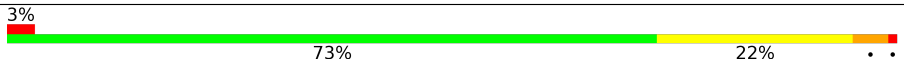
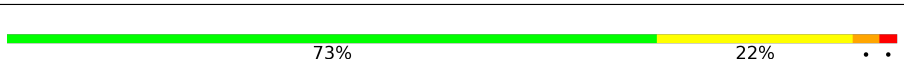
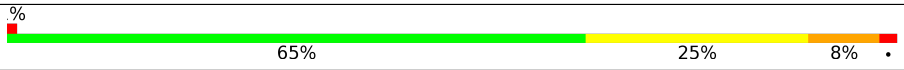
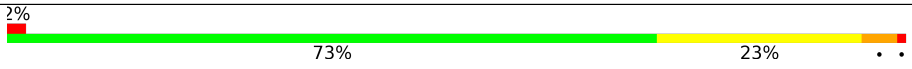
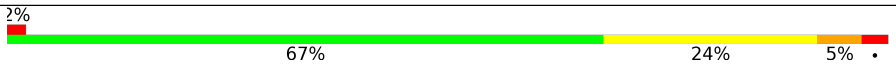
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	1005 (3.42-3.30)
Clashscore	102246	1076 (3.42-3.30)
Ramachandran outliers	100387	1059 (3.42-3.30)
Sidechain outliers	100360	1058 (3.42-3.30)
RSRZ outliers	91569	1010 (3.42-3.30)

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	131	3% 76% 20%
1	B	131	67% 23% 7%
1	C	131	3% 58% 30% 9%
1	D	131	2% 76% 21%
1	E	131	% 76% 18% 5%
1	F	131	3% 77% 20%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	131	
1	H	131	
1	I	131	
1	J	131	
1	K	131	
1	L	131	

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 12255 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 TssA.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	131	Total	C	N	O	S	Se	0	0	0
			1016	634	188	190	1	3			
1	B	131	Total	C	N	O	S	Se	0	0	0
			1010	631	185	190	1	3			
1	C	131	Total	C	N	O	S	Se	0	0	0
			995	625	178	188	1	3			
1	D	131	Total	C	N	O	S	Se	0	0	0
			1007	629	184	190	1	3			
1	E	131	Total	C	N	O	S	Se	0	0	0
			1010	631	185	190	1	3			
1	F	131	Total	C	N	O	S	Se	0	0	0
			1010	631	185	190	1	3			
1	G	131	Total	C	N	O	S	Se	0	0	0
			1010	631	185	190	1	3			
1	H	131	Total	C	N	O	S	Se	0	0	0
			1010	631	185	190	1	3			
1	I	131	Total	C	N	O	S	Se	0	0	0
			1011	632	185	190	1	3			
1	J	131	Total	C	N	O	S	Se	0	0	0
			1010	631	185	190	1	3			
1	K	131	Total	C	N	O	S	Se	0	0	0
			1010	631	185	190	1	3			
1	L	131	Total	C	N	O	S	Se	0	0	0
			1010	631	185	190	1	3			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	399	ALA	-	expression tag	UNP B7LFT5
A	400	ALA	-	expression tag	UNP B7LFT5
B	399	ALA	-	expression tag	UNP B7LFT5
B	400	ALA	-	expression tag	UNP B7LFT5
C	399	ALA	-	expression tag	UNP B7LFT5

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	400	ALA	-	expression tag	UNP B7LFT5
D	399	ALA	-	expression tag	UNP B7LFT5
D	400	ALA	-	expression tag	UNP B7LFT5
E	399	ALA	-	expression tag	UNP B7LFT5
E	400	ALA	-	expression tag	UNP B7LFT5
F	399	ALA	-	expression tag	UNP B7LFT5
F	400	ALA	-	expression tag	UNP B7LFT5
G	399	ALA	-	expression tag	UNP B7LFT5
G	400	ALA	-	expression tag	UNP B7LFT5
H	399	ALA	-	expression tag	UNP B7LFT5
H	400	ALA	-	expression tag	UNP B7LFT5
I	399	ALA	-	expression tag	UNP B7LFT5
I	400	ALA	-	expression tag	UNP B7LFT5
J	399	ALA	-	expression tag	UNP B7LFT5
J	400	ALA	-	expression tag	UNP B7LFT5
K	399	ALA	-	expression tag	UNP B7LFT5
K	400	ALA	-	expression tag	UNP B7LFT5
L	399	ALA	-	expression tag	UNP B7LFT5
L	400	ALA	-	expression tag	UNP B7LFT5

- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	11	Total O 11 11	0	0
2	B	14	Total O 14 14	0	0
2	C	11	Total O 11 11	0	0
2	D	9	Total O 9 9	0	0
2	E	14	Total O 14 14	0	0
2	F	5	Total O 5 5	0	0
2	G	12	Total O 12 12	0	0
2	H	11	Total O 11 11	0	0
2	I	27	Total O 27 27	0	0
2	J	15	Total O 15 15	0	0

Continued on next page...

Continued from previous page...

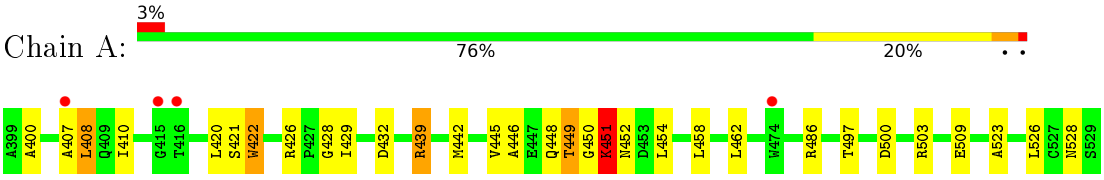
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	K	10	Total O 10 10	0	0
2	L	7	Total O 7 7	0	0

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.

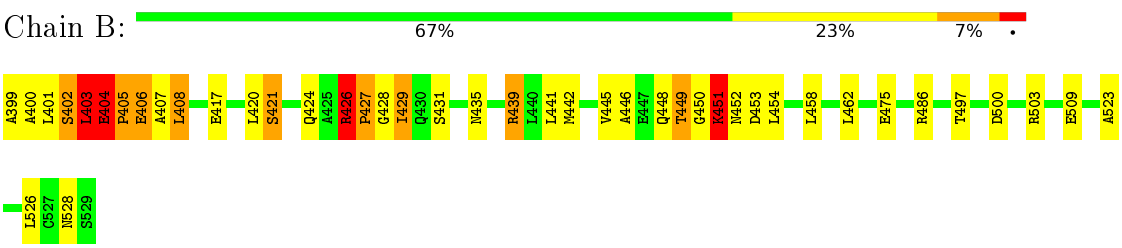
- Molecule 1: TssA

Chain A:



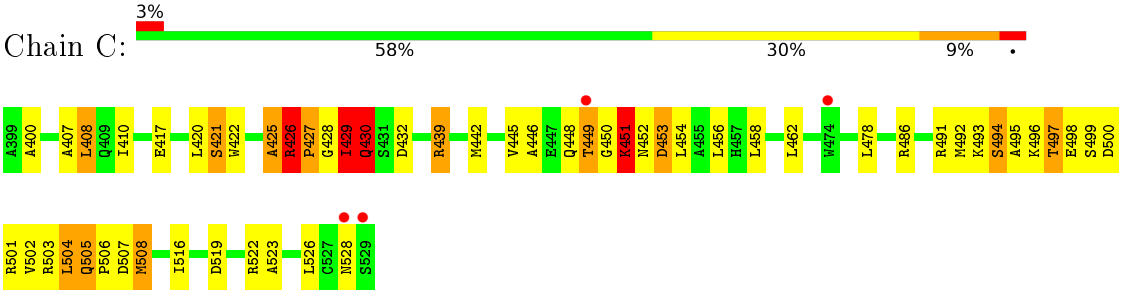
- Molecule 1: TssA

Chain B:



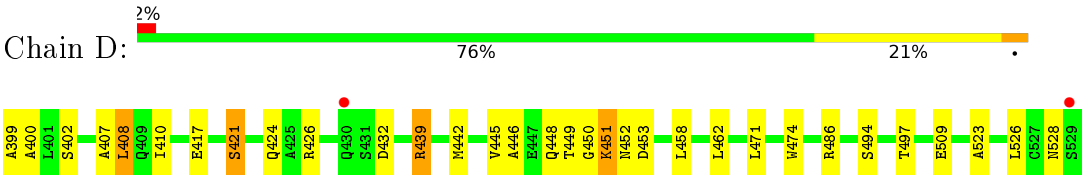
- Molecule 1: TssA

Chain C:



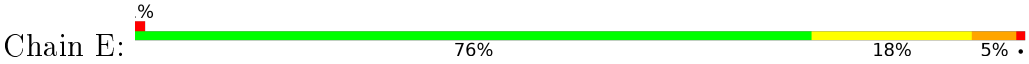
- Molecule 1: TssA

Chain D:



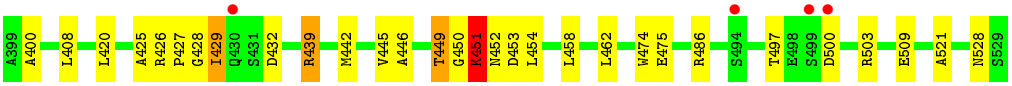
- Molecule 1: TssA

Chain E:



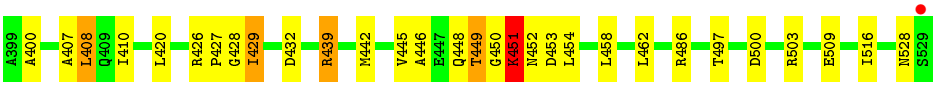
• Molecule 1: TssA

Chain F:



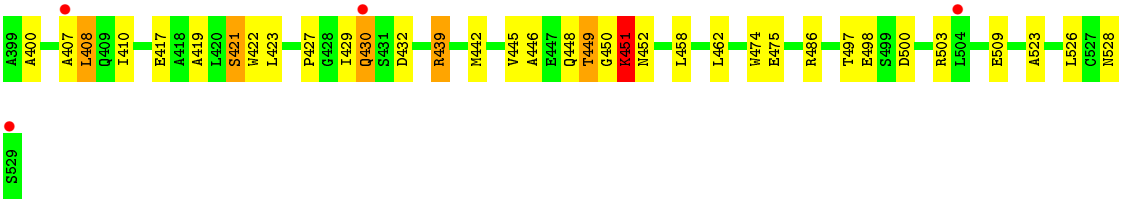
• Molecule 1: TssA

Chain G:



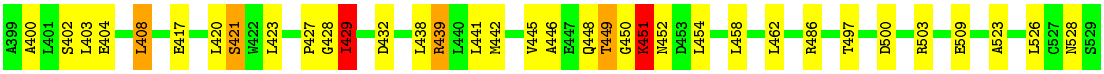
• Molecule 1: TssA

Chain H:



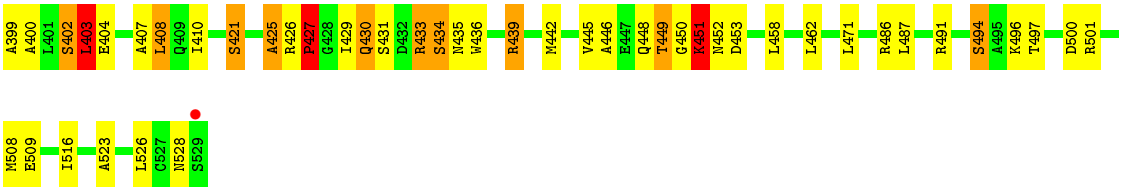
• Molecule 1: TssA

Chain I:



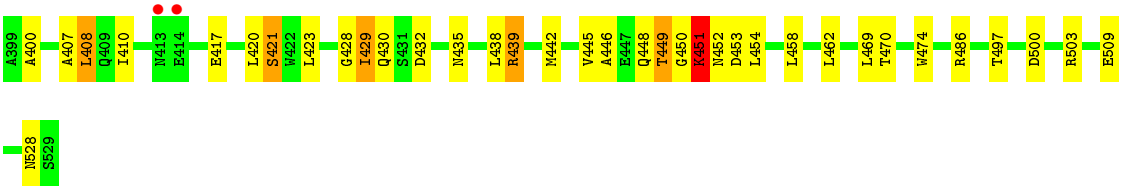
• Molecule 1: TssA

Chain J:



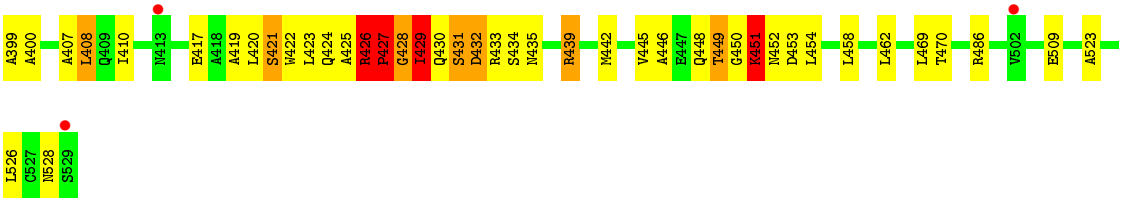
• Molecule 1: TssA

Chain K:



- Molecule 1: TssA

Chain L:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	112.28Å 90.63Å 120.09Å 90.00° 101.96° 90.00°	Depositor
Resolution (Å)	46.20 – 3.35 42.31 – 3.35	Depositor EDS
% Data completeness (in resolution range)	99.3 (46.20-3.35) 99.3 (42.31-3.35)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.87 (at 3.32Å)	Xtriage
Refinement program	BUSTER 2.10.1	Depositor
R, R_{free}	0.225 , 0.260 0.239 , 0.270	Depositor DCC
R_{free} test set	1698 reflections (5.26%)	DCC
Wilson B-factor (Å^2)	99.4	Xtriage
Anisotropy	0.159	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 64.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Outliers	0 of 33971 reflections	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	12255	wwPDB-VP
Average B, all atoms (Å^2)	105.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.37% 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

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	0/1025	0.66	0/1382
1	B	0.56	1/1019 (0.1%)	0.75	2/1375 (0.1%)
1	C	0.57	0/1004	0.76	2/1357 (0.1%)
1	D	0.48	0/1016	0.70	0/1372
1	E	0.49	0/1019	0.68	0/1375
1	F	0.47	0/1019	0.66	0/1375
1	G	0.49	0/1019	0.71	0/1375
1	H	0.49	0/1019	0.67	0/1374
1	I	0.54	0/1020	0.71	0/1376
1	J	0.57	0/1019	0.77	1/1375 (0.1%)
1	K	0.47	0/1019	0.67	0/1375
1	L	0.53	0/1019	0.70	2/1375 (0.1%)
All	All	0.51	1/12217 (0.0%)	0.71	7/16486 (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
1	B	0	3
1	C	0	6
1	J	0	2
1	L	0	5
All	All	0	16

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	427	PRO	N-CD	5.23	1.55	1.47

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	404	GLU	C-N-CD	6.89	142.88	128.40
1	L	426	ARG	C-N-CD	6.30	141.63	128.40
1	J	426	ARG	C-N-CD	6.29	141.61	128.40
1	C	426	ARG	C-N-CD	6.21	141.44	128.40
1	C	505	GLN	C-N-CD	6.07	141.15	128.40

There are no chirality outliers.

5 of 16 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	403	LEU	Peptide
1	B	404	GLU	Peptide
1	B	426	ARG	Peptide
1	C	425	ALA	Peptide
1	C	426	ARG	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	1016	0	1038	19	0
1	B	1010	0	1027	52	0
1	C	995	0	999	68	0
1	D	1007	0	1018	9	0
1	E	1010	0	1027	18	0
1	F	1010	0	1027	18	0
1	G	1010	0	1027	17	0
1	H	1010	0	1025	27	0
1	I	1011	0	1029	29	0
1	J	1010	0	1027	60	0
1	K	1010	0	1027	26	0
1	L	1010	0	1027	52	0
2	A	11	0	0	0	0
2	B	14	0	0	1	0
2	C	11	0	0	0	0
2	D	9	0	0	0	0
2	E	14	0	0	0	0
2	F	5	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	G	12	0	0	1	0
2	H	11	0	0	0	0
2	I	27	0	0	0	0
2	J	15	0	0	0	0
2	K	10	0	0	0	0
2	L	7	0	0	0	0
All	All	12255	0	12298	359	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 15.

The worst 5 of 359 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:B:429:ILE:HD13	1:H:430:GLN:NE2	1.22	1.49
1:I:429:ILE:CD1	1:L:430:GLN:HE22	1.23	1.48
1:J:429:ILE:HD11	1:J:434:SER:C	1.43	1.39
1:I:429:ILE:CD1	1:L:430:GLN:NE2	1.86	1.37
1:J:429:ILE:CD1	1:J:434:SER:C	2.01	1.28

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	129/131 (98%)	117 (91%)	7 (5%)	5 (4%)	4 29
1	B	129/131 (98%)	118 (92%)	5 (4%)	6 (5%)	3 23
1	C	129/131 (98%)	114 (88%)	10 (8%)	5 (4%)	4 29
1	D	129/131 (98%)	117 (91%)	7 (5%)	5 (4%)	4 29
1	E	129/131 (98%)	115 (89%)	8 (6%)	6 (5%)	3 23

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	F	129/131 (98%)	116 (90%)	6 (5%)	7 (5%)	2 19
1	G	129/131 (98%)	114 (88%)	10 (8%)	5 (4%)	4 29
1	H	129/131 (98%)	114 (88%)	11 (8%)	4 (3%)	5 35
1	I	129/131 (98%)	114 (88%)	10 (8%)	5 (4%)	4 29
1	J	129/131 (98%)	115 (89%)	7 (5%)	7 (5%)	2 19
1	K	129/131 (98%)	117 (91%)	8 (6%)	4 (3%)	5 35
1	L	129/131 (98%)	114 (88%)	8 (6%)	7 (5%)	2 19
All	All	1548/1572 (98%)	1385 (90%)	97 (6%)	66 (4%)	3 26

5 of 66 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	451	LYS
1	A	452	ASN
1	B	406	GLU
1	B	426	ARG
1	B	451	LYS

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	104/104 (100%)	87 (84%)	17 (16%)	3 14
1	B	103/104 (99%)	85 (82%)	18 (18%)	2 11
1	C	99/104 (95%)	80 (81%)	19 (19%)	2 7
1	D	102/104 (98%)	87 (85%)	15 (15%)	4 18
1	E	103/104 (99%)	87 (84%)	16 (16%)	3 16
1	F	103/104 (99%)	88 (85%)	15 (15%)	4 18
1	G	103/104 (99%)	88 (85%)	15 (15%)	4 18
1	H	103/104 (99%)	86 (84%)	17 (16%)	3 13
1	I	103/104 (99%)	86 (84%)	17 (16%)	3 13

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	J	103/104 (99%)	82 (80%)	21 (20%)	1 6
1	K	103/104 (99%)	87 (84%)	16 (16%)	3 16
1	L	103/104 (99%)	91 (88%)	12 (12%)	7 28
All	All	1232/1248 (99%)	1034 (84%)	198 (16%)	3 14

5 of 198 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	445	VAL
1	G	497	THR
1	K	500	ASP
1	F	453	ASP
1	G	408	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 22 such sidechains are listed below:

Mol	Chain	Res	Type
1	G	435	ASN
1	H	435	ASN
1	L	430	GLN
1	G	448	GLN
1	H	430	GLN

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 carbohydrates in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

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	128/131 (97%)	0.12	4 (3%) 52 53	77, 113, 148, 160	0
1	B	128/131 (97%)	-0.04	0 100 100	70, 100, 141, 179	0
1	C	128/131 (97%)	-0.04	4 (3%) 52 53	70, 111, 150, 160	0
1	D	128/131 (97%)	-0.07	2 (1%) 74 75	69, 109, 149, 162	0
1	E	128/131 (97%)	-0.04	1 (0%) 87 88	63, 99, 142, 174	0
1	F	128/131 (97%)	0.05	4 (3%) 52 53	70, 111, 158, 164	0
1	G	128/131 (97%)	-0.06	1 (0%) 87 88	62, 101, 144, 175	0
1	H	128/131 (97%)	0.13	4 (3%) 52 53	78, 115, 155, 176	0
1	I	128/131 (97%)	-0.08	0 100 100	54, 82, 124, 162	0
1	J	128/131 (97%)	-0.13	1 (0%) 87 88	51, 90, 134, 147	0
1	K	128/131 (97%)	0.02	2 (1%) 74 75	63, 109, 153, 161	0
1	L	128/131 (97%)	-0.01	3 (2%) 64 64	65, 101, 152, 171	0
All	All	1536/1572 (97%)	-0.01	26 (1%) 73 73	51, 103, 149, 179	0

The worst 5 of 26 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	529	SER	10.1
1	G	529	SER	5.4
1	C	529	SER	4.7
1	F	499	SER	3.7
1	A	415	GLY	3.6

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 (i)

There are no carbohydrates in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.