wwPDB X-ray Structure Validation Summary Report (i)

Nov 11, 2023 – 08:56 am GMT

PDB ID	:	8C4I
Title	:	Ligand-free Crystal Structure of the decameric Sulfofructose Transaldolase BmSF-TAL
Authors	:	Snow, A.J.D.; Sharma, M.; Davies, G.J.
Deposited on	:	2023-01-04
Resolution	:	3.20 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:

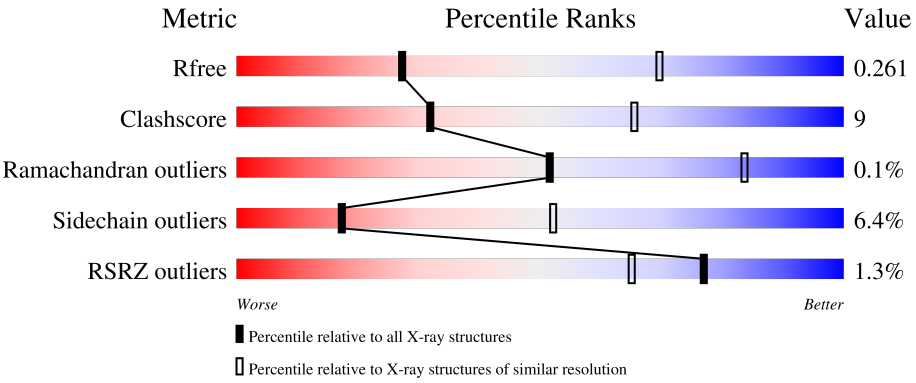
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 3.20 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.



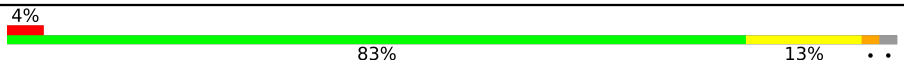
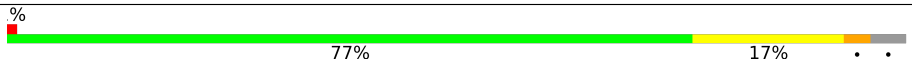
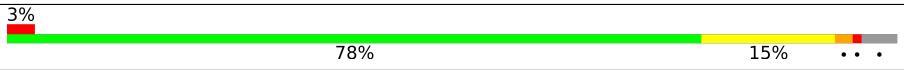
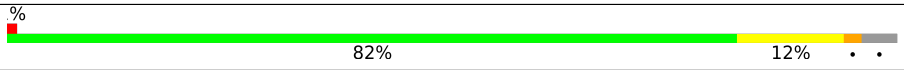
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1133 (3.20-3.20)
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)
RSRZ outliers	127900	1095 (3.20-3.20)

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	226	85% 10%
1	B	226	84% 12%
1	C	226	83% 15%
1	D	226	72% 21%
1	E	226	74% 19%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	226	
1	G	226	
1	H	226	
1	J	226	
1	K	226	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 32872 atoms, of which 15987 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 BmSF-TAL.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	218	Total	C	H	N	O	S	48	0	0
			3359	1094	1652	280	328	5			
1	B	221	Total	C	H	N	O	S	49	0	0
			3418	1112	1688	281	332	5			
1	C	222	Total	C	H	N	O	S	54	0	0
			3376	1101	1660	281	329	5			
1	D	219	Total	C	H	N	O	S	51	0	0
			3372	1094	1666	283	324	5			
1	E	217	Total	C	H	N	O	S	53	0	0
			3306	1079	1621	280	321	5			
1	F	221	Total	C	H	N	O	S	76	0	0
			3145	1037	1514	275	314	5			
1	G	218	Total	C	H	N	O	S	68	0	0
			3170	1039	1538	270	318	5			
1	H	218	Total	C	H	N	O	S	79	0	0
			3049	1009	1463	263	309	5			
1	J	218	Total	C	H	N	O	S	61	0	0
			3202	1051	1554	273	319	5			
1	K	219	Total	C	H	N	O	S	53	0	0
			3331	1087	1631	281	327	5			

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	22	Total	O	0	0
			22	22		
2	B	24	Total	O	0	0
			24	24		
2	C	18	Total	O	0	0
			18	18		
2	D	22	Total	O	0	0
			22	22		

Continued on next page...

Continued from previous page...

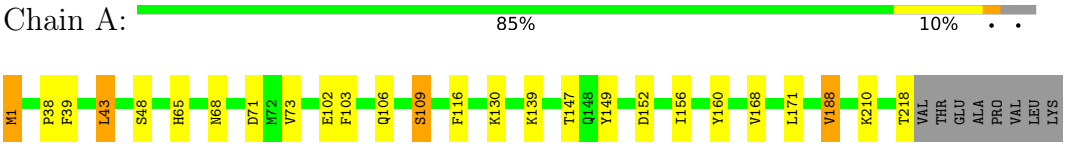
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	E	11	Total O 11 11	0	0
2	F	8	Total O 8 8	0	0
2	G	6	Total O 6 6	0	0
2	H	7	Total O 7 7	0	0
2	J	11	Total O 11 11	0	0
2	K	15	Total O 15 15	0	0

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.

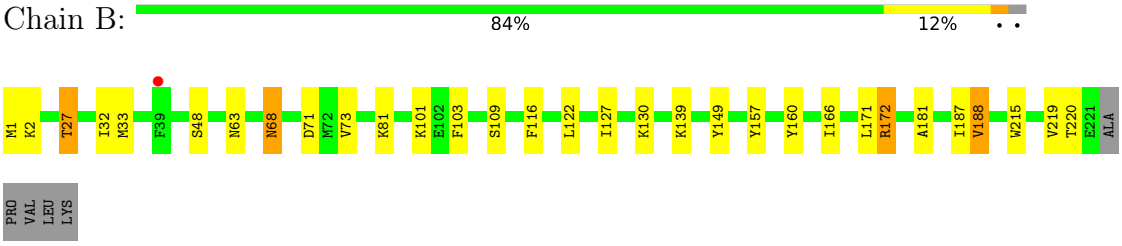
• Molecule 1: BmSF-TAL

Chain A:



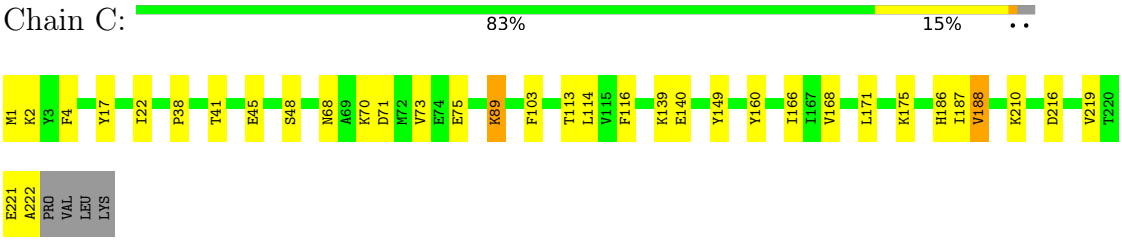
• Molecule 1: BmSF-TAL

Chain B:



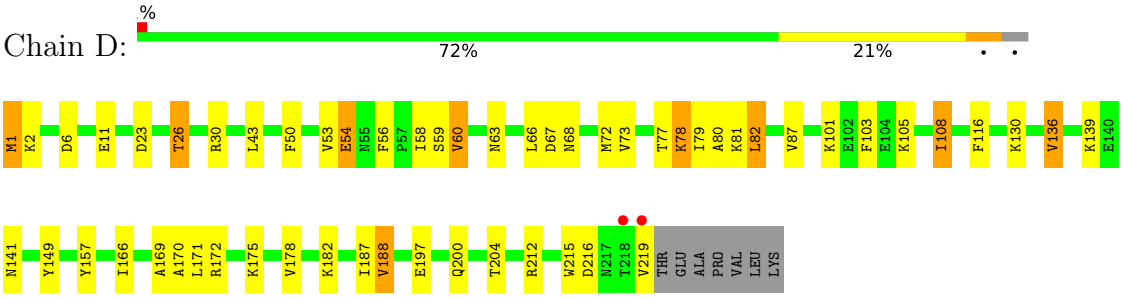
• Molecule 1: BmSF-TAL

Chain C:



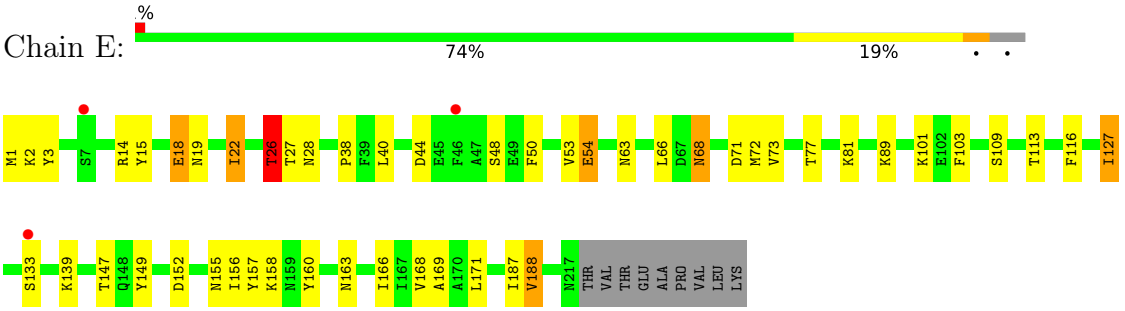
• Molecule 1: BmSF-TAL

Chain D:



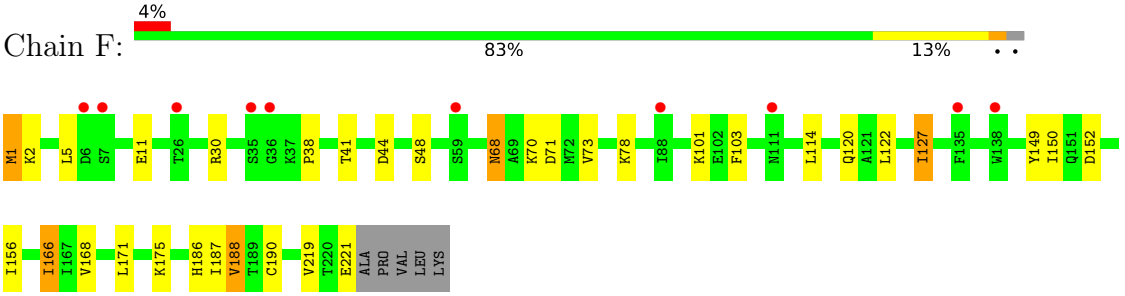
• Molecule 1: BmSF-TAL

Chain E:



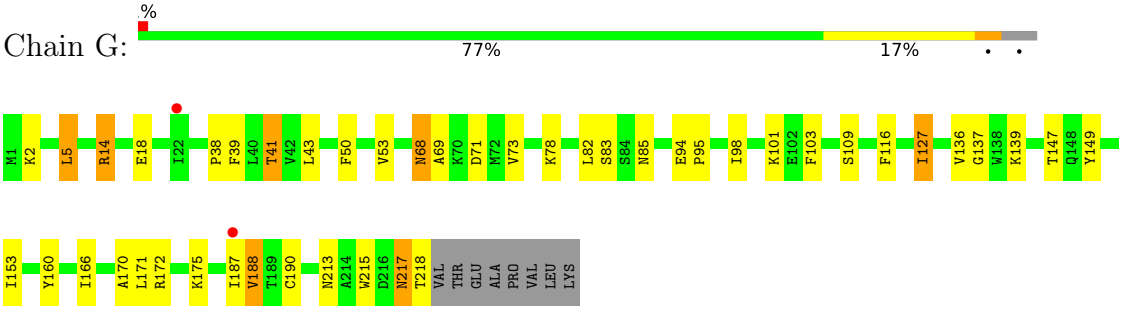
• Molecule 1: BmSF-TAL

Chain F:



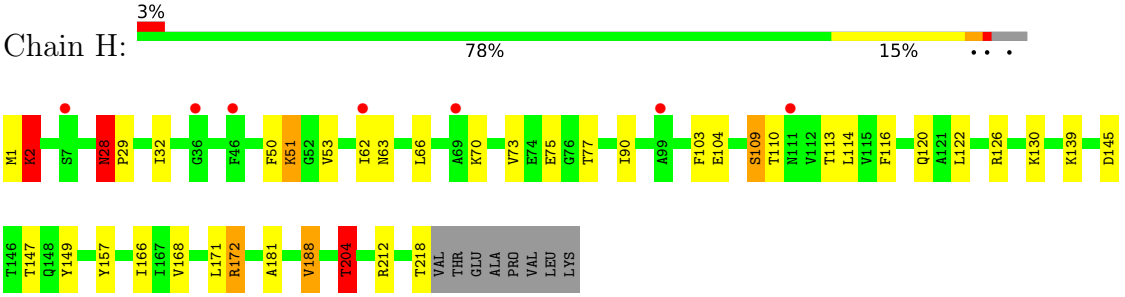
• Molecule 1: BmSF-TAL

Chain G:



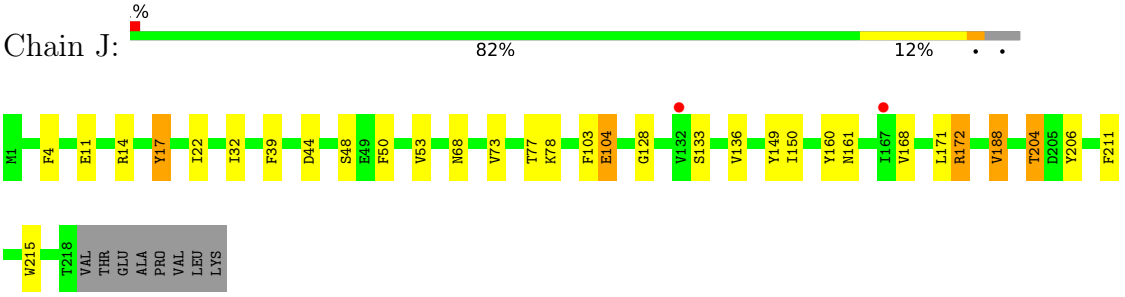
• Molecule 1: BmSF-TAL

Chain H:



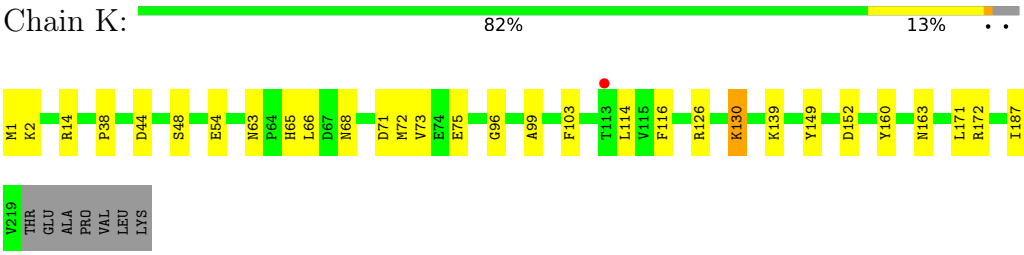
• Molecule 1: BmSF-TAL

Chain J:



• Molecule 1: BmSF-TAL

Chain K:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	73.89Å 120.84Å 152.64Å 90.00° 93.40° 90.00°	Depositor
Resolution (Å)	68.08 – 3.20 67.99 – 3.20	Depositor EDS
% Data completeness (in resolution range)	99.6 (68.08-3.20) 99.2 (67.99-3.20)	Depositor EDS
R_{merge}	0.18	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.39 (at 3.19Å)	Xtriage
Refinement program	REFMAC 5.8.0403, REFMAC 5.8.0403	Depositor
R, R_{free}	0.235 , 0.268 0.236 , 0.261	Depositor DCC
R_{free} test set	2089 reflections (4.75%)	wwPDB-VP
Wilson B-factor (Å^2)	48.3	Xtriage
Anisotropy	0.744	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 48.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.53$, $< L^2 > = 0.36$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.87	EDS
Total number of atoms	32872	wwPDB-VP
Average B, all atoms (Å^2)	50.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.57% 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	0/1748	0.67	0/2378
1	B	0.37	0/1771	0.69	0/2407
1	C	0.35	0/1757	0.67	0/2393
1	D	0.35	0/1746	0.69	0/2374
1	E	0.35	0/1726	0.69	2/2346 (0.1%)
1	F	0.36	0/1670	0.69	1/2281 (0.0%)
1	G	0.36	0/1669	0.69	0/2274
1	H	0.34	0/1624	0.71	3/2224 (0.1%)
1	J	0.35	0/1688	0.73	4/2304 (0.2%)
1	K	0.35	0/1741	0.67	0/2369
All	All	0.35	0/17140	0.69	10/23350 (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	1
1	D	0	2
1	F	0	1
1	H	0	2
1	K	0	1
All	All	0	7

There are no bond length outliers.

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	F	188	VAL	CA-CB-CG1	6.84	121.16	110.90
1	J	17	TYR	CB-CG-CD2	6.23	124.74	121.00
1	J	17	TYR	CB-CG-CD1	-6.03	117.38	121.00

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	H	2	LYS	CB-CA-C	-5.41	99.58	110.40
1	H	28	ASN	CB-CA-C	-5.38	99.64	110.40

There are no chirality outliers.

5 of 7 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	172	ARG	Sidechain
1	D	212	ARG	Sidechain
1	D	30	ARG	Sidechain
1	F	30	ARG	Sidechain
1	H	126	ARG	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	1707	1652	1638	13	0
1	B	1730	1688	1674	26	0
1	C	1716	1660	1637	32	0
1	D	1706	1666	1646	64	0
1	E	1685	1621	1599	45	0
1	F	1631	1514	1446	22	0
1	G	1632	1538	1486	39	0
1	H	1586	1463	1388	40	0
1	J	1648	1554	1514	21	0
1	K	1700	1631	1611	28	0
2	A	22	0	0	1	0
2	B	24	0	0	1	0
2	C	18	0	0	5	0
2	D	22	0	0	2	0
2	E	11	0	0	2	0
2	F	8	0	0	0	0
2	G	6	0	0	1	0
2	H	7	0	0	1	0
2	J	11	0	0	0	0
2	K	15	0	0	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	16885	15987	15639	287	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 9.

The worst 5 of 287 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:D:136:VAL:HG11	1:D:171:LEU:HD23	1.31	1.06
1:B:219:VAL:HG13	1:E:38:PRO:HB3	1.37	1.05
1:D:136:VAL:HG11	1:D:171:LEU:CD2	1.86	1.03
1:D:79:ILE:HA	1:D:82:LEU:HD22	1.41	1.00
1:D:103:PHE:HD1	1:D:108:ILE:HG21	1.25	0.98

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	216/226 (96%)	213 (99%)	3 (1%)	0	100	100
1	B	219/226 (97%)	215 (98%)	4 (2%)	0	100	100
1	C	220/226 (97%)	216 (98%)	4 (2%)	0	100	100
1	D	217/226 (96%)	214 (99%)	3 (1%)	0	100	100
1	E	215/226 (95%)	211 (98%)	4 (2%)	0	100	100
1	F	219/226 (97%)	216 (99%)	3 (1%)	0	100	100
1	G	216/226 (96%)	213 (99%)	3 (1%)	0	100	100
1	H	216/226 (96%)	207 (96%)	7 (3%)	2 (1%)	17	56
1	J	216/226 (96%)	212 (98%)	4 (2%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	K	217/226 (96%)	214 (99%)	3 (1%)	0	100 100
All	All	2171/2260 (96%)	2131 (98%)	38 (2%)	2 (0%)	51 83

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	H	2	LYS
1	H	51	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	181/193 (94%)	170 (94%)	11 (6%)	18 54
1	B	184/193 (95%)	175 (95%)	9 (5%)	25 61
1	C	179/193 (93%)	174 (97%)	5 (3%)	43 74
1	D	179/193 (93%)	166 (93%)	13 (7%)	14 46
1	E	174/193 (90%)	160 (92%)	14 (8%)	12 42
1	F	152/193 (79%)	139 (91%)	13 (9%)	10 38
1	G	158/193 (82%)	145 (92%)	13 (8%)	11 41
1	H	147/193 (76%)	138 (94%)	9 (6%)	18 54
1	J	165/193 (86%)	152 (92%)	13 (8%)	12 43
1	K	177/193 (92%)	168 (95%)	9 (5%)	24 60
All	All	1696/1930 (88%)	1587 (94%)	109 (6%)	17 52

5 of 109 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	68	ASN
1	G	109	SER
1	J	188	VAL
1	F	122	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	G	5	LEU

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

Mol	Chain	Res	Type
1	H	85	ASN
1	J	68	ASN
1	K	217	ASN
1	C	186	HIS
1	C	159	ASN

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 monosaccharides 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	218/226 (96%)	-0.01	0 100 100	24, 37, 58, 84	0
1	B	221/226 (97%)	0.03	1 (0%) 91 86	23, 38, 62, 102	0
1	C	222/226 (98%)	-0.07	0 100 100	22, 40, 59, 74	0
1	D	219/226 (96%)	0.20	2 (0%) 84 75	33, 54, 83, 104	0
1	E	217/226 (96%)	0.28	3 (1%) 75 63	28, 49, 73, 115	0
1	F	221/226 (97%)	0.36	10 (4%) 33 21	28, 54, 83, 99	0
1	G	218/226 (96%)	0.17	2 (0%) 84 75	36, 58, 80, 88	0
1	H	218/226 (96%)	0.37	7 (3%) 47 31	37, 63, 88, 101	0
1	J	218/226 (96%)	0.19	2 (0%) 84 75	30, 52, 79, 96	0
1	K	219/226 (96%)	0.14	1 (0%) 91 86	23, 41, 62, 83	0
All	All	2191/2260 (96%)	0.17	28 (1%) 77 65	22, 49, 77, 115	0

The worst 5 of 28 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	7	SER	4.1
1	F	59	SER	3.7
1	H	7	SER	3.3
1	H	99	ALA	2.9
1	D	219	VAL	2.6

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 monosaccharides 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.