wwPDB X-ray Structure Validation Summary Report (i)

Oct 16, 2021 – 06:07 PM EDT

PDB ID	:	1Q2V
Title	:	Crystal structure of the chaperonin from Thermococcus strain KS-1 (nucleotide-free form)
Authors	:	Shomura, Y.; Yoshida, T.; Iizuka, R.; Yohda, M.; Maruyama, T.; Miki, K.
Deposited on	:	2003-07-26
Resolution	:	2.40 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 following versions of software and data (see references (i)) were used in the production of this report:

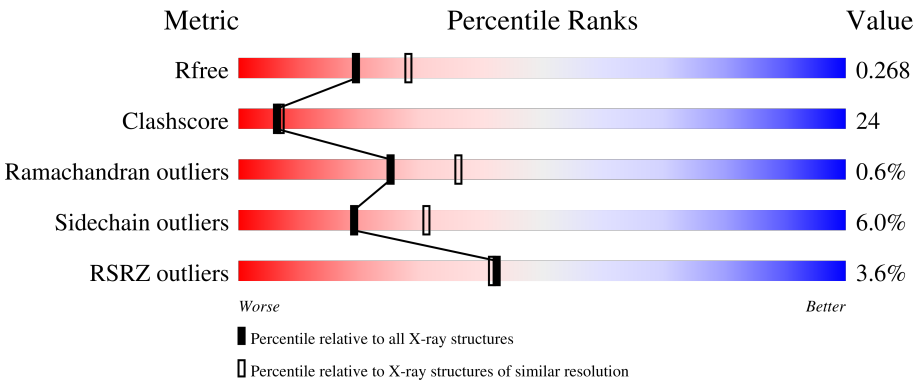
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 2.40 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	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

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	548	3% 56% 36% • 5%
1	B	548	4% 57% 33% • 5%
1	C	548	3% 57% 35% • 5%
1	D	548	3% 58% 32% • 5%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 15963 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.

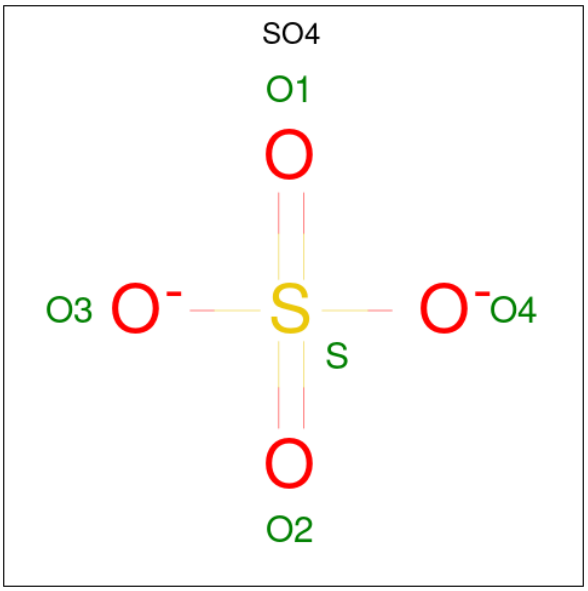
- Molecule 1 is a protein called Thermosome alpha subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	518	3947	2485	673	772	17	0	0	0
1	B	518	3947	2485	673	772	17	0	0	0
1	C	518	3947	2485	673	772	17	0	0	0
1	D	518	3947	2485	673	772	17	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	65	CYS	GLY	engineered mutation	UNP O24729
A	125	THR	ILE	engineered mutation	UNP O24729
B	65	CYS	GLY	engineered mutation	UNP O24729
B	125	THR	ILE	engineered mutation	UNP O24729
C	65	CYS	GLY	engineered mutation	UNP O24729
C	125	THR	ILE	engineered mutation	UNP O24729
D	65	CYS	GLY	engineered mutation	UNP O24729
D	125	THR	ILE	engineered mutation	UNP O24729

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	O	S		
2	A	1	5	4	1	0	0
2	B	1	5	4	1	0	0
2	C	1	5	4	1	0	0
2	D	1	5	4	1	0	0

- Molecule 3 is water.

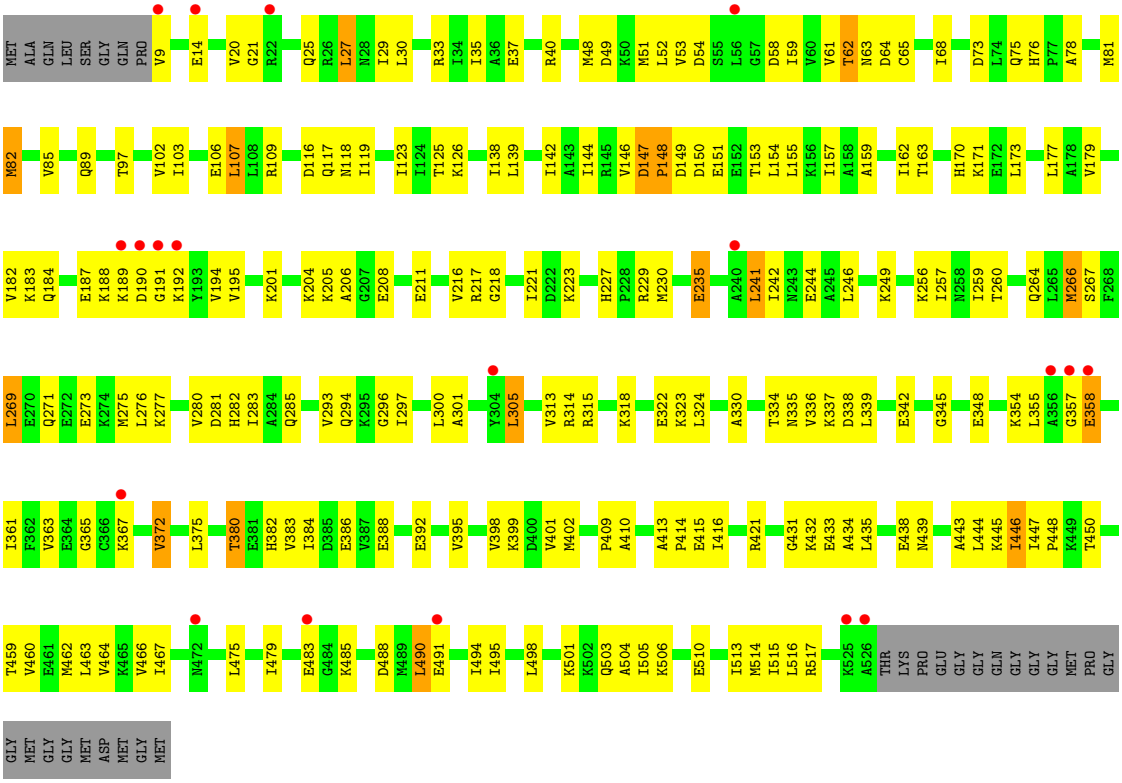
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	43	43	43	0	0
3	B	40	40	40	0	0
3	C	36	36	36	0	0
3	D	36	36	36	0	0

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.

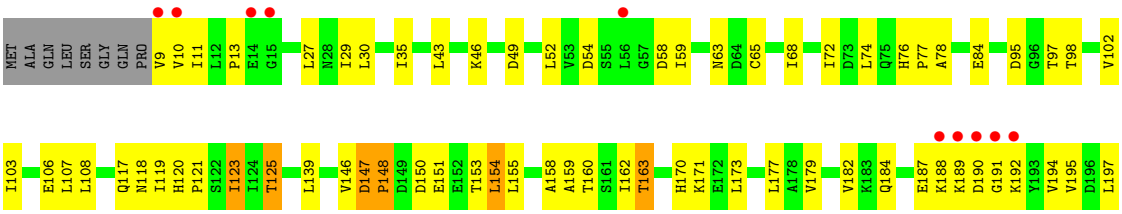
• Molecule 1: Thermosome alpha subunit

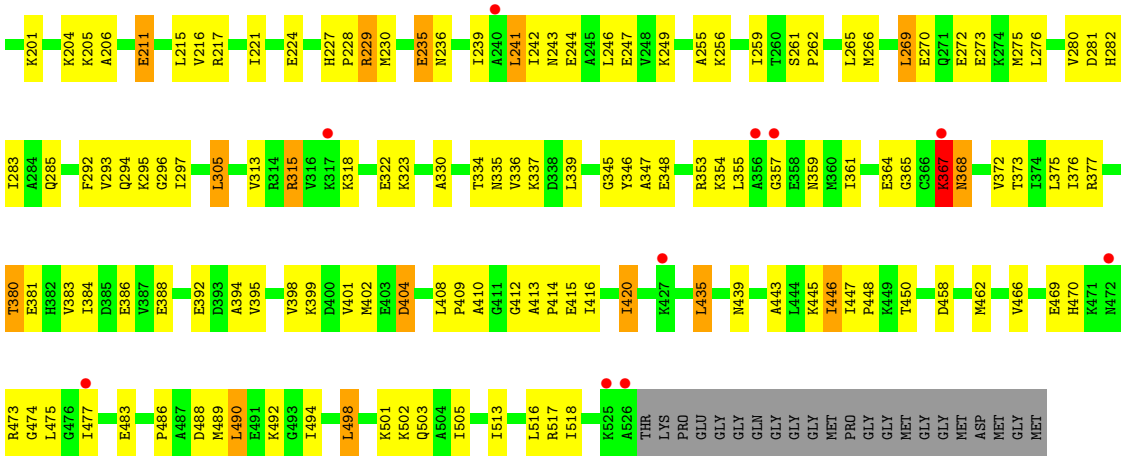
Chain A: 3% 56% 36% • 5%



• Molecule 1: Thermosome alpha subunit

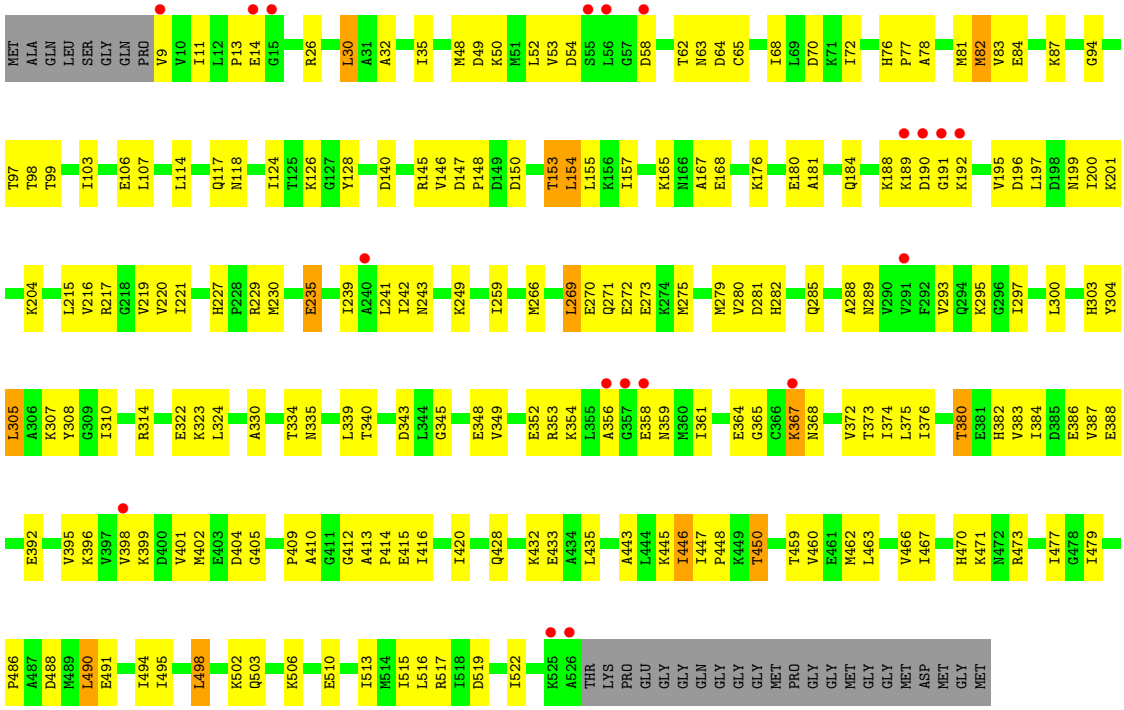
Chain B: 4% 57% 33% • 5%





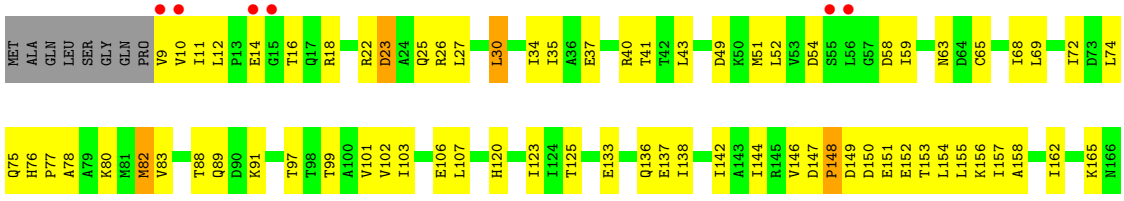
• Molecule 1: Thermosome alpha subunit

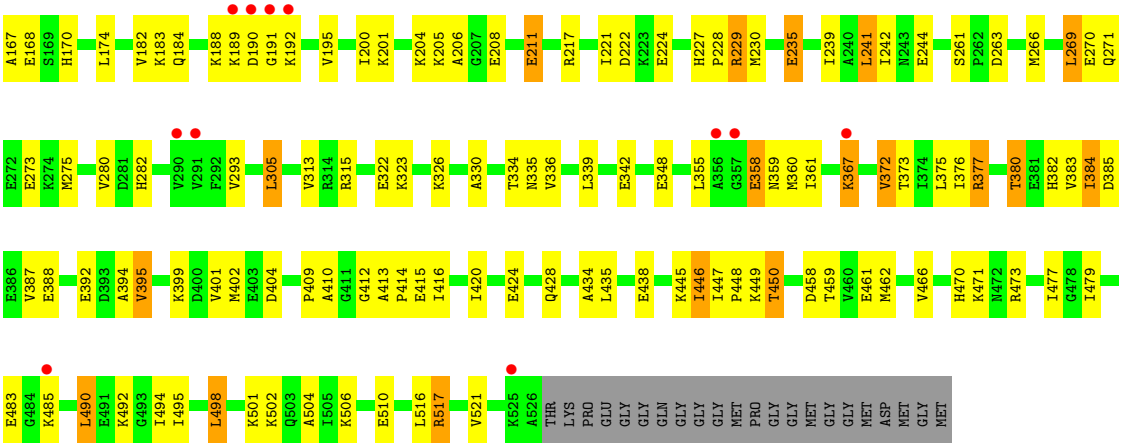
Chain C:



• Molecule 1: Thermosome alpha subunit

Chain D:




A167 E168 S169 H170 L174 V182 K183 Q184 K188 K189 D190 G191 K192 V195 I200 K201 K204 K205 A206 G207 E208 E211 R217 I221 D222 K223 E224 H227 P228 R229 M230 E235 I239 A240 L241 I242 N243 E244 S261 P262 D263 M266 L269 E270 Q271
E272 E273 K274 M275 V280 D281 H282 V290 V291 F292 V293 L305 V313 R314 R315 E322 K323 K326 A330 T334 N335 V336 L339 E342 E348 L355 A356 G357 E358 M359 M360 I361 K367 V372 T373 I374 L375 I376 R377 T380 E381 H382 V383 I384 D385
E386 V387 E388 E392 D393 A394 V395 K399 D400 V401 M402 E403 D404 P409 A410 G411 G412 A413 P414 E415 I416 I420 E424 Q428 A434 L435 E438 K445 I446 I447 P448 K449 T450 D458 T459 V460 E461 M462 V466 H470 K471 N472 R473 I477 G478 I479
E483 G484 K485 L490 E491 K492 G493 I494 I495 L498 K501 K502 Q503 A504 I505 K506 E510 L516 R517 V521 K525 A526 THR LYS PRO GLU GLY GLY GLN GLY GLY GLY MET PRO GLY GLY MET GLY GLY MET ASP MET GLY MET

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 4 21 2	Depositor
Cell constants a, b, c, α, β, γ	209.27Å 209.27Å 156.62Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.99 – 2.40 19.99 – 2.40	Depositor EDS
% Data completeness (in resolution range)	99.2 (19.99-2.40) 99.3 (19.99-2.40)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.06	Depositor
$< I/\sigma(I) >$ [1]	1.97 (at 2.41Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.246 , 0.269 0.245 , 0.268	Depositor DCC
R_{free} test set	6737 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	38.4	Xtriage
Anisotropy	0.142	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 35.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	15963	wwPDB-VP
Average B, all atoms (Å^2)	39.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.94% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: SO4

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	0.36	0/3982	0.63	0/5367
1	B	0.38	1/3982 (0.0%)	0.64	2/5367 (0.0%)
1	C	0.37	0/3982	0.64	0/5367
1	D	0.37	1/3982 (0.0%)	0.62	0/5367
All	All	0.37	2/15928 (0.0%)	0.63	2/21468 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	11	ILE	C-N	-6.78	1.18	1.34
1	D	158	ALA	C-N	-6.17	1.19	1.34

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	404	ASP	N-CA-C	-5.11	97.19	111.00
1	B	236	ASN	N-CA-C	-5.09	97.24	111.00

There are no chirality outliers.

There are no planarity outliers.

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	3947	0	4127	213	0
1	B	3947	0	4126	219	0
1	C	3947	0	4127	189	0
1	D	3947	0	4126	193	0
2	A	5	0	0	0	0
2	B	5	0	0	0	0
2	C	5	0	0	0	0
2	D	5	0	0	0	0
3	A	43	0	0	3	0
3	B	40	0	0	4	0
3	C	36	0	0	1	0
3	D	36	0	0	1	0
All	All	15963	0	16506	782	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 24.

The worst 5 of 782 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:53:VAL:HG22	1:A:59:ILE:HD13	1.22	1.14
1:B:120:HIS:HB3	1:B:123:ILE:HD13	1.23	1.13
1:D:120:HIS:HB3	1:D:123:ILE:HD13	1.32	1.10
1:D:52:LEU:HD22	1:D:72:ILE:HD11	1.33	1.09
1:B:473:ARG:HB2	1:B:477:ILE:HD11	1.35	1.07

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	516/548 (94%)	495 (96%)	18 (4%)	3 (1%)	25 36

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	516/548 (94%)	499 (97%)	14 (3%)	3 (1%)	25	36
1	C	516/548 (94%)	501 (97%)	13 (2%)	2 (0%)	34	48
1	D	516/548 (94%)	500 (97%)	12 (2%)	4 (1%)	19	29
All	All	2064/2192 (94%)	1995 (97%)	57 (3%)	12 (1%)	25	36

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	367	LYS
1	C	235	GLU
1	A	235	GLU
1	B	235	GLU
1	B	367	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	423/441 (96%)	398 (94%)	25 (6%)	19	32
1	B	423/441 (96%)	398 (94%)	25 (6%)	19	32
1	C	423/441 (96%)	404 (96%)	19 (4%)	27	44
1	D	423/441 (96%)	391 (92%)	32 (8%)	13	20
All	All	1692/1764 (96%)	1591 (94%)	101 (6%)	19	31

5 of 101 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	155	LEU
1	D	30	LEU
1	D	516	LEU
1	C	305	LEU
1	C	459	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 53 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	76	HIS
1	C	285	GLN
1	D	282	HIS
1	C	89	GLN
1	C	184	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)

4 ligands are modelled in this entry.

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	$\#\|Z\| > 2$
2	SO4	D	4527	-	4,4,4	0.27	0	6,6,6	0.18	0
2	SO4	B	2527	-	4,4,4	0.26	0	6,6,6	0.14	0
2	SO4	C	3527	-	4,4,4	0.25	0	6,6,6	0.12	0
2	SO4	A	1527	-	4,4,4	0.26	0	6,6,6	0.20	0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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
1	D	1
1	B	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	D	158:ALA	C	159:ALA	N	1.19
1	B	11:ILE	C	12:LEU	N	1.18

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	518/548 (94%)	-0.14	19 (3%)	41	41	23, 37, 58, 93	0
1	B	518/548 (94%)	-0.09	20 (3%)	39	38	22, 39, 58, 93	0
1	C	518/548 (94%)	-0.15	19 (3%)	41	41	20, 35, 55, 96	0
1	D	518/548 (94%)	-0.14	17 (3%)	46	45	24, 38, 58, 96	0
All	All	2072/2192 (94%)	-0.13	75 (3%)	42	42	20, 37, 57, 96	0

The worst 5 of 75 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	56	LEU	5.3
1	C	189	LYS	5.2
1	B	357	GLY	4.9
1	A	526	ALA	4.8
1	B	189	LYS	4.7

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 monosaccharides 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. 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	B-factors(Å^2)	Q<0.9
2	SO4	D	4527	5/5	0.97	0.10	34,35,36,37	0
2	SO4	B	2527	5/5	0.99	0.09	36,37,38,40	0
2	SO4	C	3527	5/5	0.99	0.08	33,33,35,35	0
2	SO4	A	1527	5/5	0.99	0.07	31,32,33,34	0

6.5 Other polymers (i)

There are no such residues in this entry.