



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 25, 2020 – 02:05 am BST

PDB ID : 5BV1  
Title : Crystal Structure of a Vps33-Vps16 Complex from Chaetomium thermophilum  
Authors : Baker, R.W.; Jeffrey, P.D.; Hughson, F.M.  
Deposited on : 2015-06-04  
Resolution : 2.90 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](mailto: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 ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) 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	:	2.11
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)	:	2.11

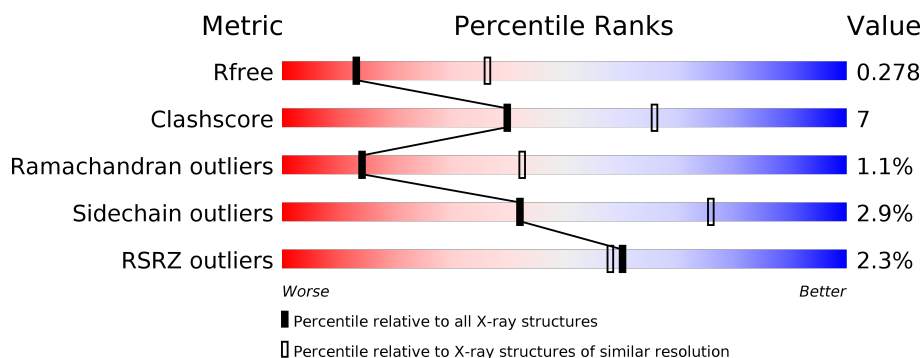
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.90 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	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	669	<div> <div>2%</div> <div> <div></div> <div>71%</div> <div>18%</div> <div>10%</div> </div> </div>
1	C	669	<div> <div>%</div> <div> <div></div> <div>72%</div> <div>16%</div> <div>10%</div> </div> </div>
2	B	333	<div> <div>2%</div> <div> <div></div> <div>67%</div> <div>21%</div> <div>11%</div> </div> </div>
2	D	333	<div> <div>5%</div> <div> <div></div> <div>73%</div> <div>15%</div> <div>11%</div> </div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 14259 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 VPS33.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	601	Total	C	N	O	S	0	0	0
			4752	3007	836	898	11			
1	C	602	Total	C	N	O	S	0	0	0
			4759	3012	837	899	11			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	expression tag	UNP G0SCM5
C	-1	GLY	-	expression tag	UNP G0SCM5

- Molecule 2 is a protein called Putative vacuolar protein sorting-associated protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	297	Total	C	N	O	S	0	0	0
			2365	1487	423	446	9			
2	D	297	Total	C	N	O	S	0	0	0
			2365	1487	423	446	9			

There are 44 discrepancies between the modelled and reference sequences:

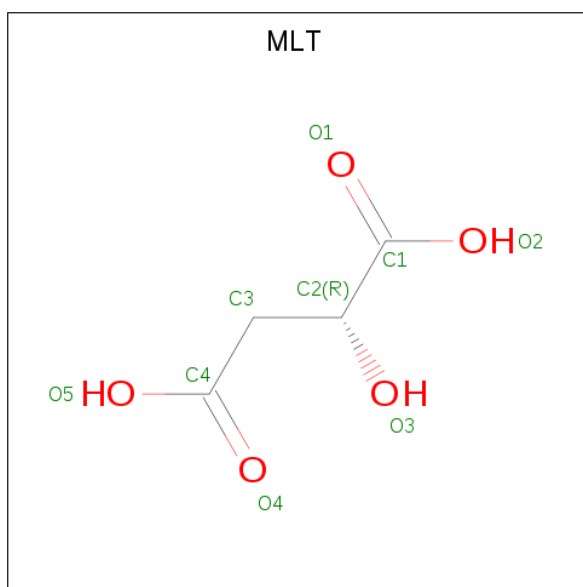
Chain	Residue	Modelled	Actual	Comment	Reference
B	502	MET	-	initiating methionine	UNP G0S6M7
B	503	GLY	-	expression tag	UNP G0S6M7
B	504	SER	-	expression tag	UNP G0S6M7
B	672	ARG	-	insertion	UNP G0S6M7
B	673	MET	-	insertion	UNP G0S6M7
B	674	GLN	-	insertion	UNP G0S6M7
B	675	GLU	-	insertion	UNP G0S6M7
B	676	THR	-	insertion	UNP G0S6M7
B	677	PHE	-	insertion	UNP G0S6M7
B	678	GLU	-	insertion	UNP G0S6M7

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	679	ARG	-	insertion	UNP G0S6M7
B	680	ASP	-	insertion	UNP G0S6M7
B	681	LEU	-	insertion	UNP G0S6M7
B	682	THR	-	insertion	UNP G0S6M7
B	683	ASP	-	insertion	UNP G0S6M7
B	684	SER	-	insertion	UNP G0S6M7
B	685	PHE	-	insertion	UNP G0S6M7
B	686	VAL	-	insertion	UNP G0S6M7
B	687	GLY	-	insertion	UNP G0S6M7
B	688	LEU	-	insertion	UNP G0S6M7
B	689	SER	-	insertion	UNP G0S6M7
B	690	VAL	LEU	engineered mutation	UNP G0S6M7
D	502	MET	-	initiating methionine	UNP G0S6M7
D	503	GLY	-	expression tag	UNP G0S6M7
D	504	SER	-	expression tag	UNP G0S6M7
D	672	ARG	-	insertion	UNP G0S6M7
D	673	MET	-	insertion	UNP G0S6M7
D	674	GLN	-	insertion	UNP G0S6M7
D	675	GLU	-	insertion	UNP G0S6M7
D	676	THR	-	insertion	UNP G0S6M7
D	677	PHE	-	insertion	UNP G0S6M7
D	678	GLU	-	insertion	UNP G0S6M7
D	679	ARG	-	insertion	UNP G0S6M7
D	680	ASP	-	insertion	UNP G0S6M7
D	681	LEU	-	insertion	UNP G0S6M7
D	682	THR	-	insertion	UNP G0S6M7
D	683	ASP	-	insertion	UNP G0S6M7
D	684	SER	-	insertion	UNP G0S6M7
D	685	PHE	-	insertion	UNP G0S6M7
D	686	VAL	-	insertion	UNP G0S6M7
D	687	GLY	-	insertion	UNP G0S6M7
D	688	LEU	-	insertion	UNP G0S6M7
D	689	SER	-	insertion	UNP G0S6M7
D	690	VAL	LEU	engineered mutation	UNP G0S6M7

- Molecule 3 is D-MALATE (three-letter code: MLT) (formula: C<sub>4</sub>H<sub>6</sub>O<sub>5</sub>).

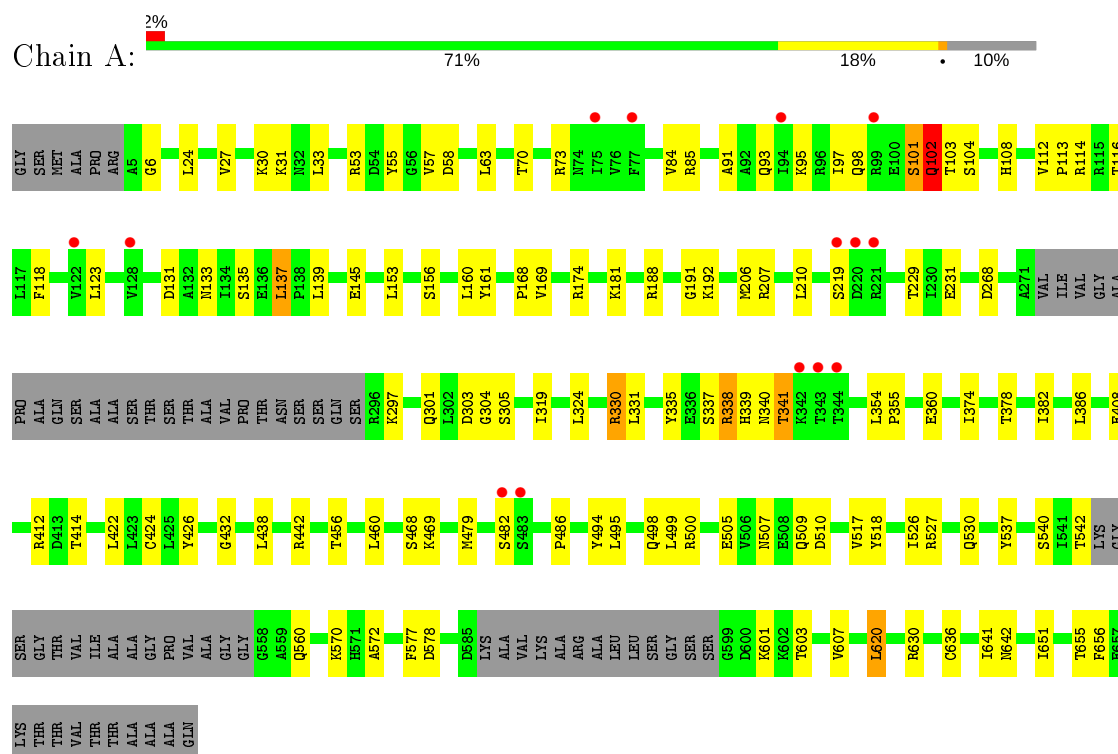


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	B	1	Total	C	O	0	0
			9	4	5		
3	D	1	Total	C	O	0	0
			9	4	5		

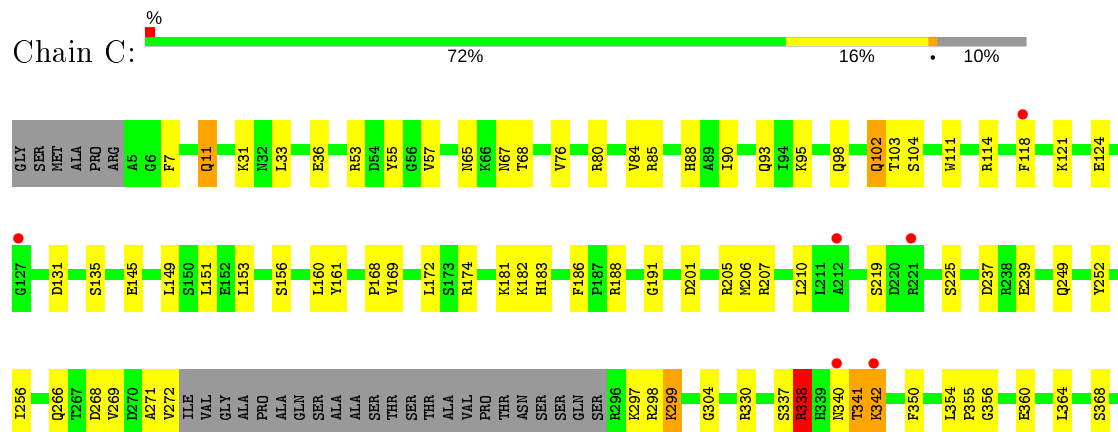
### 3 Residue-property plots [i](#)

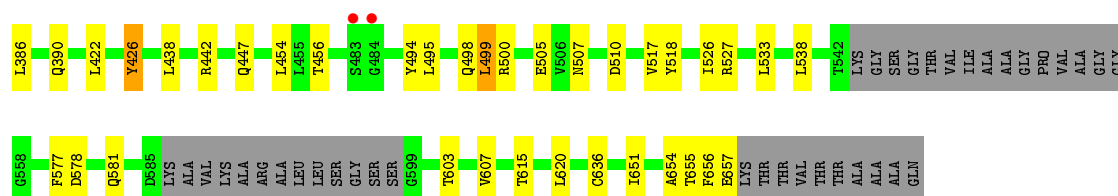
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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: VPS33

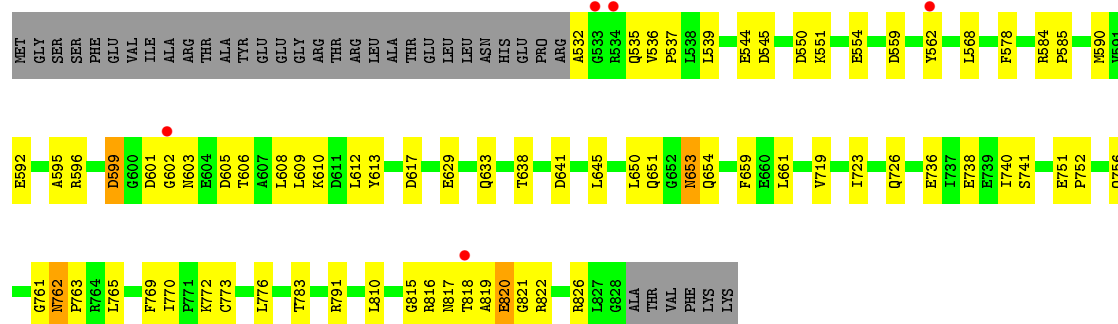


#### • Molecule 1: VPS33

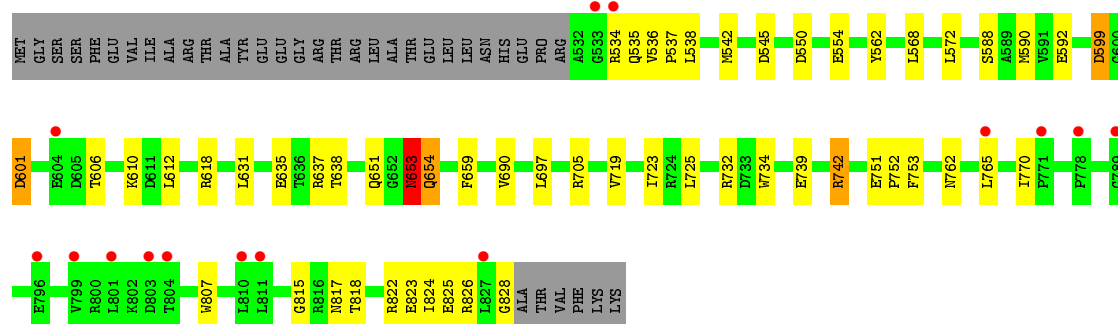
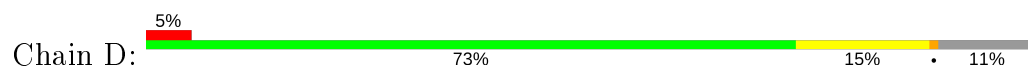




- Molecule 2: Putative vacuolar protein sorting-associated protein



- Molecule 2: Putative vacuolar protein sorting-associated protein



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	79.28Å 159.90Å 97.39Å 90.00° 90.21° 90.00°	Depositor
Resolution (Å)	61.59 – 2.90 83.18 – 2.90	Depositor EDS
% Data completeness (in resolution range)	98.9 (61.59-2.90) 98.9 (83.18-2.90)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.15 (at 2.91Å)	Xtriage
Refinement program	PHENIX	Depositor
R, $R_{free}$	0.211 , 0.276 0.215 , 0.278	Depositor DCC
$R_{free}$ test set	2699 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	65.3	Xtriage
Anisotropy	0.307	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 32.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.155 for h,-k,-l	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	14259	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	66.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.06% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  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

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

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	0/4827	0.54	0/6512
1	C	0.33	0/4834	0.54	0/6522
2	B	0.32	0/2404	0.50	0/3242
2	D	0.32	0/2404	0.54	0/3242
All	All	0.33	0/14469	0.54	0/19518

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	A	0	2
1	C	0	2
All	All	0	4

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	340	ASN	Peptide
1	A	341	THR	Peptide
1	C	340	ASN	Peptide
1	C	341	THR	Peptide

## 5.2 Too-close contacts ⓘ

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	4752	0	4794	76	0
1	C	4759	0	4803	68	1
2	B	2365	0	2368	44	0
2	D	2365	0	2368	33	1
3	B	9	0	4	1	0
3	D	9	0	4	1	0
All	All	14259	0	14341	210	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 7.

The worst 5 of 210 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:102:GLN:HE21	1:A:103:THR:H	1.20	0.89
1:C:153:LEU:HB2	1:C:156:SER:HB3	1.65	0.77
1:A:338:ARG:O	1:A:341:THR:OG1	2.02	0.75
1:A:31:LYS:NZ	1:A:55:TYR:O	2.18	0.75
1:A:231:GLU:OE2	1:A:560:GLN:NE2	2.22	0.73

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)
1:C:342:LYS:NZ	2:D:535:GLN:OE1[1_455]	2.19	0.01

## 5.3 Torsion angles ⓘ

### 5.3.1 Protein backbone ⓘ

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	593/669 (89%)	552 (93%)	35 (6%)	6 (1%)	15	45
1	C	594/669 (89%)	548 (92%)	40 (7%)	6 (1%)	15	45
2	B	295/333 (89%)	267 (90%)	23 (8%)	5 (2%)	9	31
2	D	295/333 (89%)	266 (90%)	26 (9%)	3 (1%)	15	45
All	All	1777/2004 (89%)	1633 (92%)	124 (7%)	20 (1%)	14	42

5 of 20 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	338	ARG
1	A	482	SER
2	B	653	ASN
2	B	654	GLN
1	C	338	ARG

### 5.3.2 Protein sidechains ⓘ

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	515/564 (91%)	504 (98%)	11 (2%)	53	81
1	C	516/564 (92%)	500 (97%)	16 (3%)	40	74
2	B	248/280 (89%)	240 (97%)	8 (3%)	39	73
2	D	248/280 (89%)	239 (96%)	9 (4%)	35	69
All	All	1527/1688 (90%)	1483 (97%)	44 (3%)	42	76

5 of 44 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	11	GLN
1	C	207	ARG
2	D	739	GLU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	C	102	GLN
1	C	131	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	183	HIS

### 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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

2 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	$\# Z  > 2$
3	MLT	D	901	-	2,8,8	0.67	0	3,10,10	1.54	1 (33%)
3	MLT	B	901	-	2,8,8	0.75	0	3,10,10	2.49	1 (33%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the

Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	MLT	D	901	-	-	0/2/8/8	-
3	MLT	B	901	-	-	2/2/8/8	-

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
3	B	901	MLT	C3-C2-C1	-4.22	105.73	111.10
3	D	901	MLT	C3-C2-C1	-2.23	108.26	111.10

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	B	901	MLT	C1-C2-C3-C4
3	B	901	MLT	O3-C2-C3-C4

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	D	901	MLT	1	0
3	B	901	MLT	1	0

## 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 [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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	601/669 (89%)	0.16	14 (2%) 60 58	30, 54, 120, 170	0
1	C	602/669 (89%)	0.15	8 (1%) 77 77	30, 55, 113, 152	0
2	B	297/333 (89%)	0.12	5 (1%) 70 69	48, 65, 106, 134	0
2	D	297/333 (89%)	0.44	15 (5%) 28 24	42, 72, 141, 165	0
All	All	1797/2004 (89%)	0.20	42 (2%) 60 58	30, 60, 123, 170	0

The worst 5 of 42 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	75	ILE	6.0
2	D	533	GLY	5.1
1	C	221	ARG	4.4
1	C	483	SER	4.1
1	A	219	SER	4.1

### 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 carbohydrates 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, 95<sup>th</sup> 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( $\text{\AA}^2$ )	Q<0.9
3	MLT	D	901	9/9	0.88	0.17	72,75,89,95	0
3	MLT	B	901	9/9	0.88	0.25	58,70,89,99	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.