



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 07:36 AM GMT

PDB ID : 3BEZ  
Title : Crystal structure of Escherichia coli Signal peptide peptidase (SppA), SeMet crystals  
Authors : Paetzel, M.  
Deposited on : 2007-11-20  
Resolution : 2.76 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  
<http://wwpdb.org/validation/2016/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 : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk26865

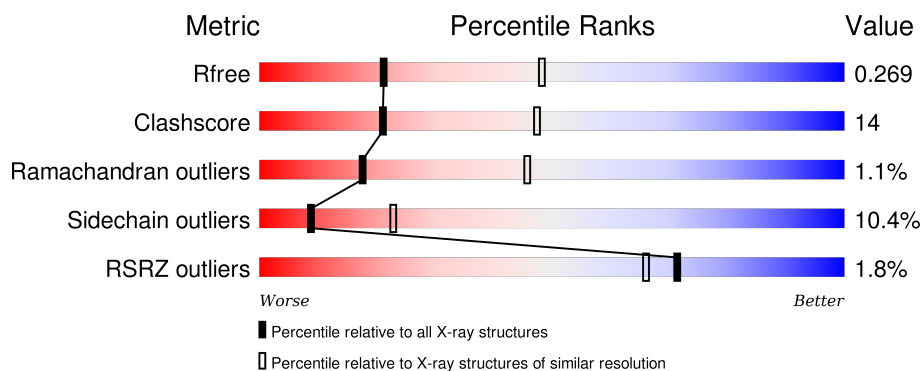
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.76 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	3340 (2.80-2.72)
Clashscore	102246	3829 (2.80-2.72)
Ramachandran outliers	100387	3767 (2.80-2.72)
Sidechain outliers	100360	3770 (2.80-2.72)
RSRZ outliers	91569	3352 (2.80-2.72)

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. 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	593	<div> <div> <div></div> <div>57%</div> <div>20%</div> <div>• •</div> <div>20%</div> </div> </div>
1	B	593	<div> <div> <div></div> <div>55%</div> <div>20%</div> <div>• •</div> <div>20%</div> </div> </div>
1	C	593	<div> <div> <div>2%</div> <div>55%</div> <div>22%</div> <div>•</div> <div>20%</div> </div> </div>
1	D	593	<div> <div> <div>2%</div> <div>59%</div> <div>20%</div> <div>•</div> <div>19%</div> </div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 14989 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 Protease 4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	473	Total	C	N	O	Se	0	0	0
			3576	2257	610	701	8			
1	B	473	Total	C	N	O	Se	0	0	0
			3572	2254	609	701	8			
1	C	476	Total	C	N	O	Se	0	0	0
			3597	2269	616	704	8			
1	D	479	Total	C	N	O	Se	0	0	0
			3617	2279	619	711	8			

There are 84 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	26	MSE	-	EXPRESSION TAG	UNP P08395
A	27	GLY	-	EXPRESSION TAG	UNP P08395
A	28	SER	-	EXPRESSION TAG	UNP P08395
A	29	SER	-	EXPRESSION TAG	UNP P08395
A	30	HIS	-	EXPRESSION TAG	UNP P08395
A	31	HIS	-	EXPRESSION TAG	UNP P08395
A	32	HIS	-	EXPRESSION TAG	UNP P08395
A	33	HIS	-	EXPRESSION TAG	UNP P08395
A	34	HIS	-	EXPRESSION TAG	UNP P08395
A	35	HIS	-	EXPRESSION TAG	UNP P08395
A	36	SER	-	EXPRESSION TAG	UNP P08395
A	37	SER	-	EXPRESSION TAG	UNP P08395
A	38	GLY	-	EXPRESSION TAG	UNP P08395
A	39	LEU	-	EXPRESSION TAG	UNP P08395
A	40	VAL	-	EXPRESSION TAG	UNP P08395
A	41	PRO	-	EXPRESSION TAG	UNP P08395
A	42	ARG	-	EXPRESSION TAG	UNP P08395
A	43	GLY	-	EXPRESSION TAG	UNP P08395
A	44	SER	-	EXPRESSION TAG	UNP P08395
A	45	HIS	-	EXPRESSION TAG	UNP P08395
A	46	MSE	-	EXPRESSION TAG	UNP P08395

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	26	MSE	-	EXPRESSION TAG	UNP P08395
B	27	GLY	-	EXPRESSION TAG	UNP P08395
B	28	SER	-	EXPRESSION TAG	UNP P08395
B	29	SER	-	EXPRESSION TAG	UNP P08395
B	30	HIS	-	EXPRESSION TAG	UNP P08395
B	31	HIS	-	EXPRESSION TAG	UNP P08395
B	32	HIS	-	EXPRESSION TAG	UNP P08395
B	33	HIS	-	EXPRESSION TAG	UNP P08395
B	34	HIS	-	EXPRESSION TAG	UNP P08395
B	35	HIS	-	EXPRESSION TAG	UNP P08395
B	36	SER	-	EXPRESSION TAG	UNP P08395
B	37	SER	-	EXPRESSION TAG	UNP P08395
B	38	GLY	-	EXPRESSION TAG	UNP P08395
B	39	LEU	-	EXPRESSION TAG	UNP P08395
B	40	VAL	-	EXPRESSION TAG	UNP P08395
B	41	PRO	-	EXPRESSION TAG	UNP P08395
B	42	ARG	-	EXPRESSION TAG	UNP P08395
B	43	GLY	-	EXPRESSION TAG	UNP P08395
B	44	SER	-	EXPRESSION TAG	UNP P08395
B	45	HIS	-	EXPRESSION TAG	UNP P08395
B	46	MSE	-	EXPRESSION TAG	UNP P08395
C	26	MSE	-	EXPRESSION TAG	UNP P08395
C	27	GLY	-	EXPRESSION TAG	UNP P08395
C	28	SER	-	EXPRESSION TAG	UNP P08395
C	29	SER	-	EXPRESSION TAG	UNP P08395
C	30	HIS	-	EXPRESSION TAG	UNP P08395
C	31	HIS	-	EXPRESSION TAG	UNP P08395
C	32	HIS	-	EXPRESSION TAG	UNP P08395
C	33	HIS	-	EXPRESSION TAG	UNP P08395
C	34	HIS	-	EXPRESSION TAG	UNP P08395
C	35	HIS	-	EXPRESSION TAG	UNP P08395
C	36	SER	-	EXPRESSION TAG	UNP P08395
C	37	SER	-	EXPRESSION TAG	UNP P08395
C	38	GLY	-	EXPRESSION TAG	UNP P08395
C	39	LEU	-	EXPRESSION TAG	UNP P08395
C	40	VAL	-	EXPRESSION TAG	UNP P08395
C	41	PRO	-	EXPRESSION TAG	UNP P08395
C	42	ARG	-	EXPRESSION TAG	UNP P08395
C	43	GLY	-	EXPRESSION TAG	UNP P08395
C	44	SER	-	EXPRESSION TAG	UNP P08395
C	45	HIS	-	EXPRESSION TAG	UNP P08395
C	46	MSE	-	EXPRESSION TAG	UNP P08395

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
D	26	MSE	-	EXPRESSION TAG	UNP P08395
D	27	GLY	-	EXPRESSION TAG	UNP P08395
D	28	SER	-	EXPRESSION TAG	UNP P08395
D	29	SER	-	EXPRESSION TAG	UNP P08395
D	30	HIS	-	EXPRESSION TAG	UNP P08395
D	31	HIS	-	EXPRESSION TAG	UNP P08395
D	32	HIS	-	EXPRESSION TAG	UNP P08395
D	33	HIS	-	EXPRESSION TAG	UNP P08395
D	34	HIS	-	EXPRESSION TAG	UNP P08395
D	35	HIS	-	EXPRESSION TAG	UNP P08395
D	36	SER	-	EXPRESSION TAG	UNP P08395
D	37	SER	-	EXPRESSION TAG	UNP P08395
D	38	GLY	-	EXPRESSION TAG	UNP P08395
D	39	LEU	-	EXPRESSION TAG	UNP P08395
D	40	VAL	-	EXPRESSION TAG	UNP P08395
D	41	PRO	-	EXPRESSION TAG	UNP P08395
D	42	ARG	-	EXPRESSION TAG	UNP P08395
D	43	GLY	-	EXPRESSION TAG	UNP P08395
D	44	SER	-	EXPRESSION TAG	UNP P08395
D	45	HIS	-	EXPRESSION TAG	UNP P08395
D	46	MSE	-	EXPRESSION TAG	UNP P08395

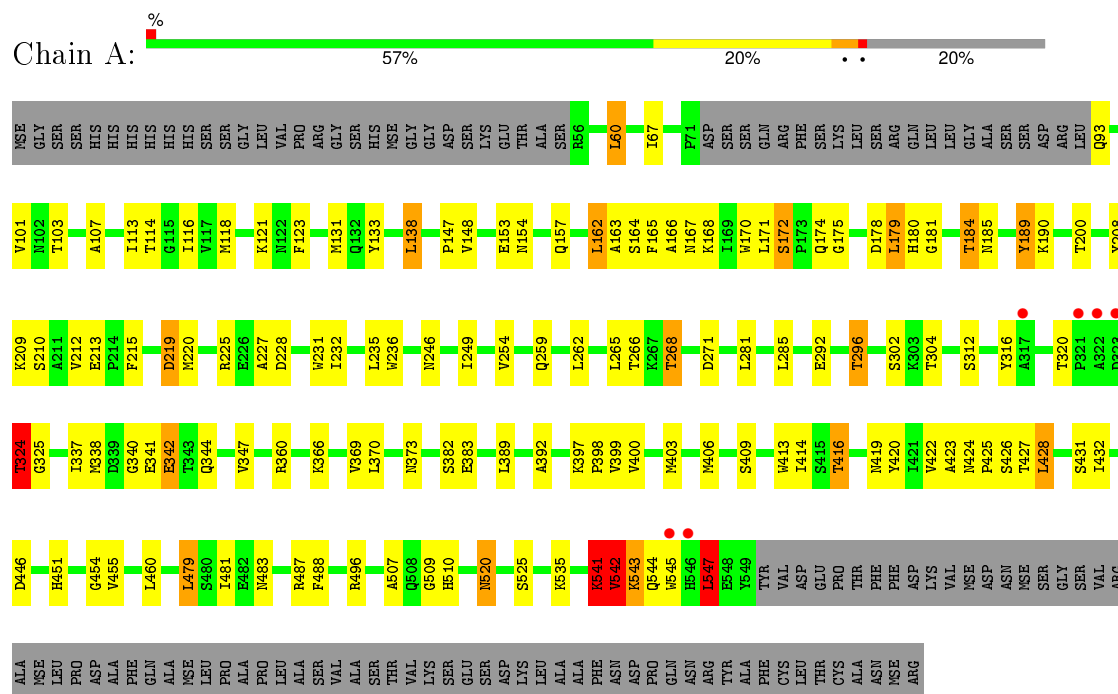
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	149	Total	O	0	0
			149	149		
2	B	166	Total	O	0	0
			166	166		
2	C	139	Total	O	0	0
			139	139		
2	D	173	Total	O	0	0
			173	173		

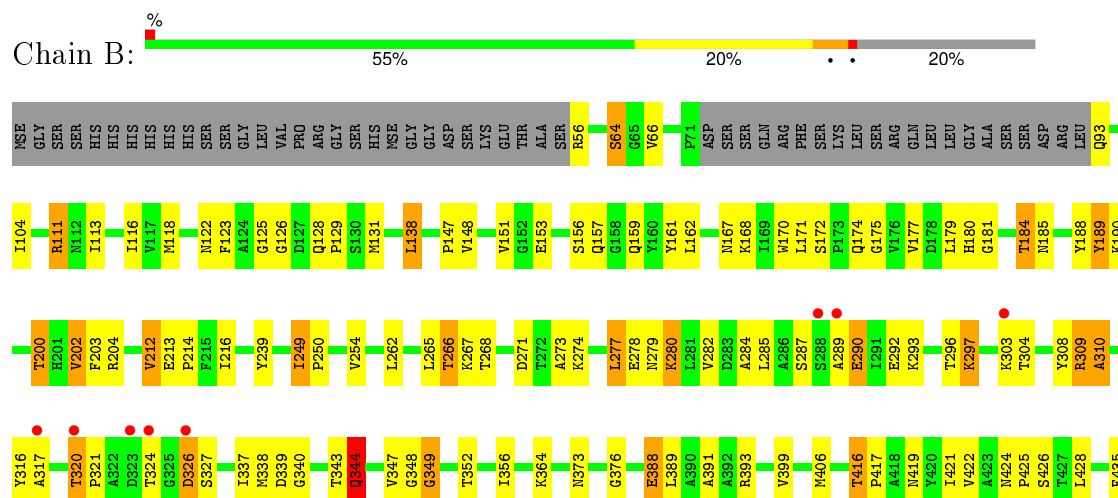
### 3 Residue-property plots

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.

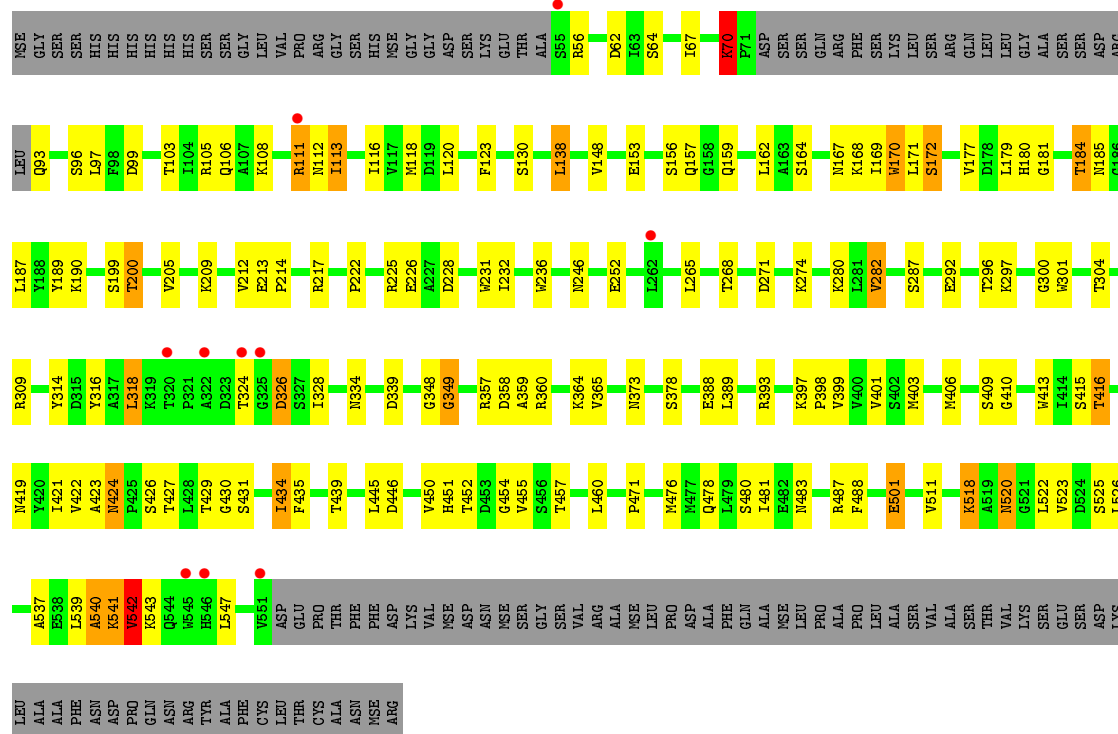
#### • Molecule 1: Protease 4



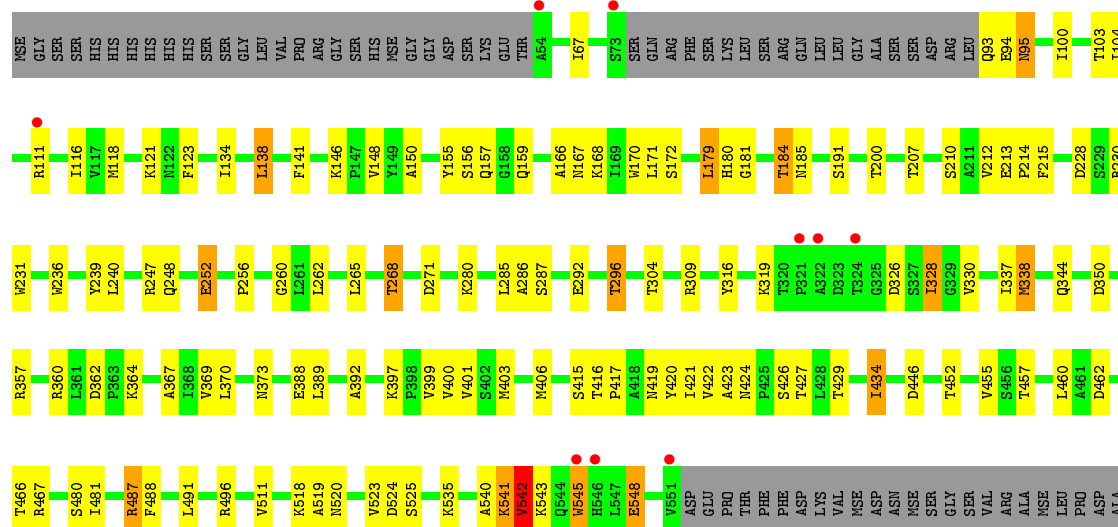
#### • Molecule 1: Protease 4



- Molecule 1: Protease 4



- Molecule 1: Protease 4



PHE	GLN	ALA	MTS	LEU	PRO	ALA	PRO	LEU	ALA	SER	VAL	ALA	SER	THR	VAL	LYS	SER	GLU	SER	ASP	LYS	LEU	ALA	ALA	PHE	ASN	ASP	PRO	GLN	ASN	ARG	TYR	ALA	PHE	CYS	LEU	THR	CYS	ALA	ASN	MTS	ARG
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	93.02Å 153.04Å 100.21Å 90.00° 104.17° 90.00°	Depositor
Resolution (Å)	46.63 – 2.76 46.85 – 2.76	Depositor EDS
% Data completeness (in resolution range)	100.0 (46.63-2.76) 98.6 (46.85-2.76)	Depositor EDS
$R_{merge}$	0.18	Depositor
$R_{sym}$	0.18	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.84 (at 2.77Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.200 , 0.259 0.200 , 0.269	Depositor DCC
$R_{free}$ test set	3500 reflections (5.33%)	DCC
Wilson B-factor (Å <sup>2</sup> )	29.4	Xtriage
Anisotropy	0.387	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 47.5	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 69169 reflections	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	14989	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	25.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.78% 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, 0.375 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	0.80	1/3635 (0.0%)	0.85	5/4922 (0.1%)
1	B	0.81	1/3631 (0.0%)	0.87	1/4918 (0.0%)
1	C	0.81	0/3656	0.88	2/4950 (0.0%)
1	D	0.80	2/3676 (0.1%)	0.86	2/4977 (0.0%)
All	All	0.80	4/14598 (0.0%)	0.86	10/19767 (0.1%)

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	B	0	2
1	C	0	2
1	D	0	1
All	All	0	7

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	542	VAL	CA-CB	6.76	1.69	1.54
1	A	542	VAL	CA-CB	6.10	1.67	1.54
1	D	542	VAL	CA-CB	5.75	1.66	1.54
1	D	545	TRP	CE3-CZ3	5.07	1.47	1.38

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	350	ASP	CB-CG-OD1	6.16	123.84	118.30
1	A	547	LEU	CA-CB-CG	5.94	128.96	115.30
1	A	219	ASP	CB-CG-OD1	5.83	123.55	118.30
1	C	542	VAL	N-CA-C	5.72	126.43	111.00

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	178	ASP	CB-CG-OD2	-5.30	113.53	118.30

There are no chirality outliers.

5 of 7 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	541	LYS	Peptide
1	A	544	GLN	Peptide
1	B	340	GLY	Peptide
1	B	541	LYS	Peptide
1	C	70	LYS	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	3576	0	3537	106	0
1	B	3572	0	3526	110	0
1	C	3597	0	3554	127	0
1	D	3617	0	3571	93	0
2	A	149	0	0	0	0
2	B	166	0	0	6	0
2	C	139	0	0	5	0
2	D	173	0	0	2	0
All	All	14989	0	14188	409	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 14.

The worst 5 of 409 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:123:PHE:CZ	1:B:131:MSE:HE2	1.73	1.21
1:A:403:MSE:HE2	1:A:423:ALA:HB2	1.27	1.17
1:C:403:MSE:HE2	1:C:423:ALA:HB2	1.28	1.15

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:541:LYS:HD3	1:A:542:VAL:HG23	1.24	1.14
1:D:121:LYS:HE2	2:D:691:HOH:O	1.55	1.05

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	469/593 (79%)	426 (91%)	39 (8%)	4 (1%)	21	52
1	B	469/593 (79%)	428 (91%)	32 (7%)	9 (2%)	10	28
1	C	472/593 (80%)	433 (92%)	33 (7%)	6 (1%)	15	40
1	D	475/593 (80%)	437 (92%)	37 (8%)	1 (0%)	52	83
All	All	1885/2372 (80%)	1724 (92%)	141 (8%)	20 (1%)	17	46

5 of 20 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	542	VAL
1	B	326	ASP
1	B	344	GLN
1	B	542	VAL
1	C	326	ASP

### 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	373/460 (81%)	340 (91%)	33 (9%)	12	32
1	B	372/460 (81%)	328 (88%)	44 (12%)	6	17
1	C	374/460 (81%)	333 (89%)	41 (11%)	8	21
1	D	377/460 (82%)	339 (90%)	38 (10%)	9	24
All	All	1496/1840 (81%)	1340 (90%)	156 (10%)	9	23

5 of 156 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	511	VAL
1	C	252	GLU
1	D	446	ASP
1	B	525	SER
1	C	111	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 55 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	424	ASN
1	C	167	ASN
1	D	346	ASN
1	B	483	ASN
1	C	128	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 [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](#)

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	465/593 (78%)	-0.34	6 (1%) 79 75	9, 25, 46, 60	0
1	B	465/593 (78%)	-0.30	8 (1%) 73 68	11, 23, 50, 62	0
1	C	468/593 (78%)	-0.32	10 (2%) 67 61	9, 21, 43, 60	0
1	D	471/593 (79%)	-0.31	9 (1%) 70 64	10, 23, 44, 55	0
All	All	1869/2372 (78%)	-0.32	33 (1%) 71 66	9, 23, 47, 62	0

The worst 5 of 33 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	322	ALA	4.7
1	C	324	THR	4.7
1	D	551	VAL	4.1
1	C	325	GLY	3.9
1	C	545	TRP	3.6

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers ⓘ

There are no such residues in this entry.