W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 11, 2024 – 04:14 AM EST

PDB ID	:	3BF0
Title	:	Crystal structure of Escherichia coli Signal peptide peptidase (SppA), Native crystals
Authors	:	Paetzel, M.
Deposited on	:	2007-11-20
Resolution	:	2.55 Å(reported)

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

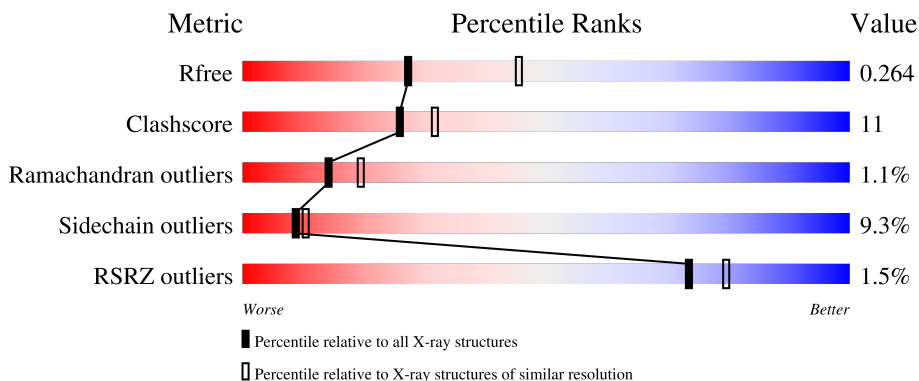|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation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

1   Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.55 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	Percentile Ranks	Value
Rfree		0.264
Clashscore		11
Ramachandran outliers		1.1%
Sidechain outliers		9.3%
RSRZ outliers		1.5%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1284 (2.56-2.52)
Clashscore	141614	1332 (2.56-2.52)
Ramachandran outliers	138981	1315 (2.56-2.52)
Sidechain outliers	138945	1315 (2.56-2.52)
RSRZ outliers	127900	1272 (2.56-2.52)

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	593	.% 61% 17% • 20%
1	B	593	.% 62% 15% • 20%
1	C	593	2% 63% 15% • 19%
1	D	593	.% 60% 17% • 20%

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

2 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 14865 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	474	Total 3587	C 2263	N 614	O 702	S 8	0	0	0
1	B	474	Total 3583	C 2260	N 613	O 702	S 8	0	0	0
1	C	478	Total 3606	C 2273	N 617	O 708	S 8	0	0	0
1	D	476	Total 3601	C 2270	N 616	O 707	S 8	0	0	0

There are 84 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	26	MET	-	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	MET	-	expression tag	UNP P08395

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	26	MET	-	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	MET	-	expression tag	UNP P08395
C	26	MET	-	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	MET	-	expression tag	UNP P08395

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	26	MET	-	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	MET	-	expression tag	UNP P08395

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	116	Total 116	O 116	0	0
2	B	127	Total 127	O 127	0	0
2	C	144	Total 144	O 144	0	0
2	D	101	Total 101	O 101	0	0

# 3 Residue-property plots ⓘ

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.
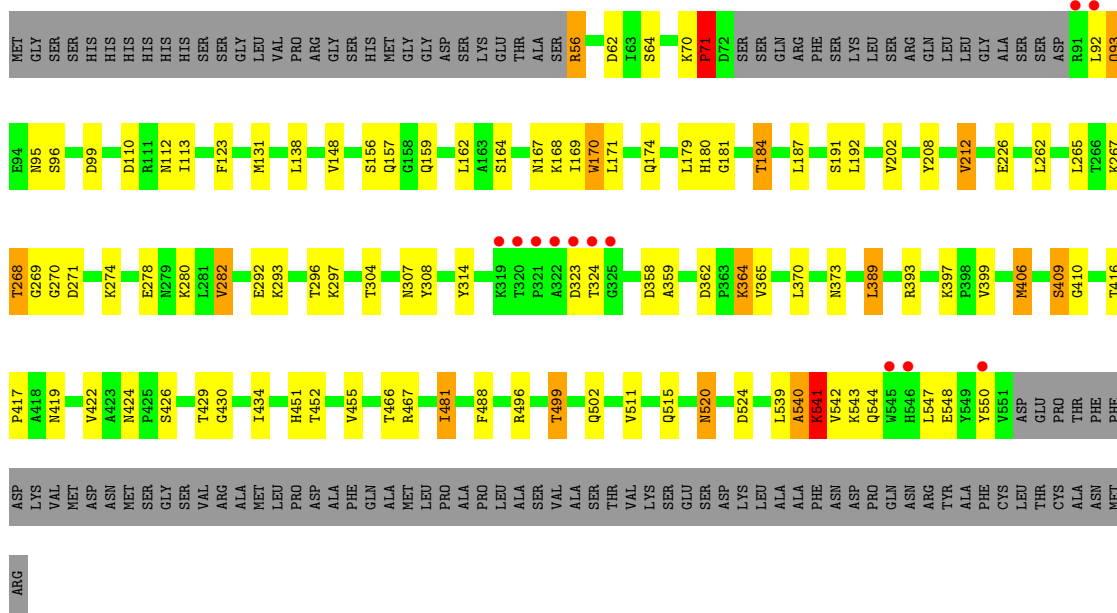
• Molecule 1: Protease 4
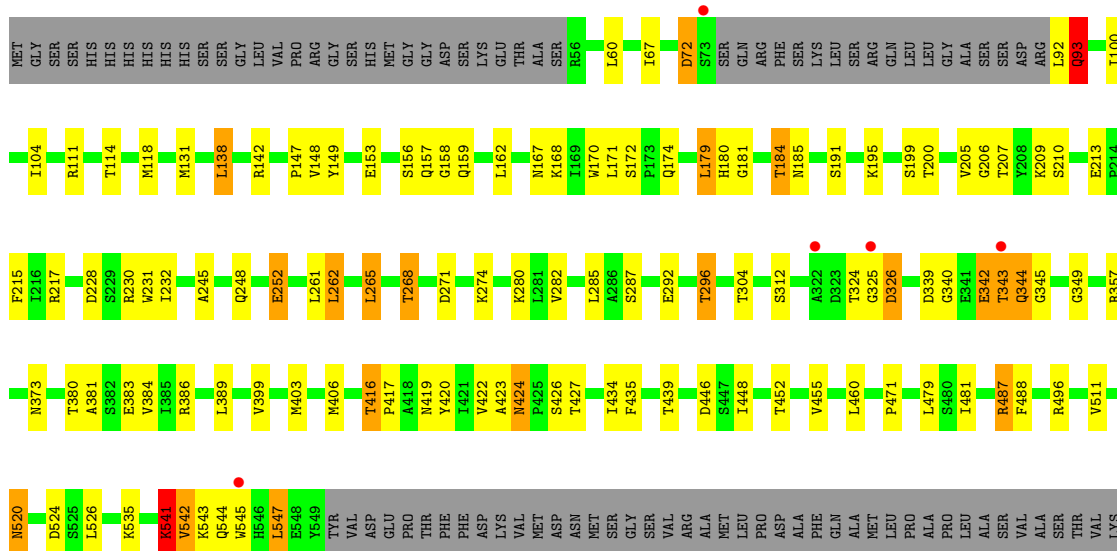

• Molecule 1: Protease 4

- Molecule 1: Protease 4

Chain C:

- Molecule 1: Protease 4

Chain D:
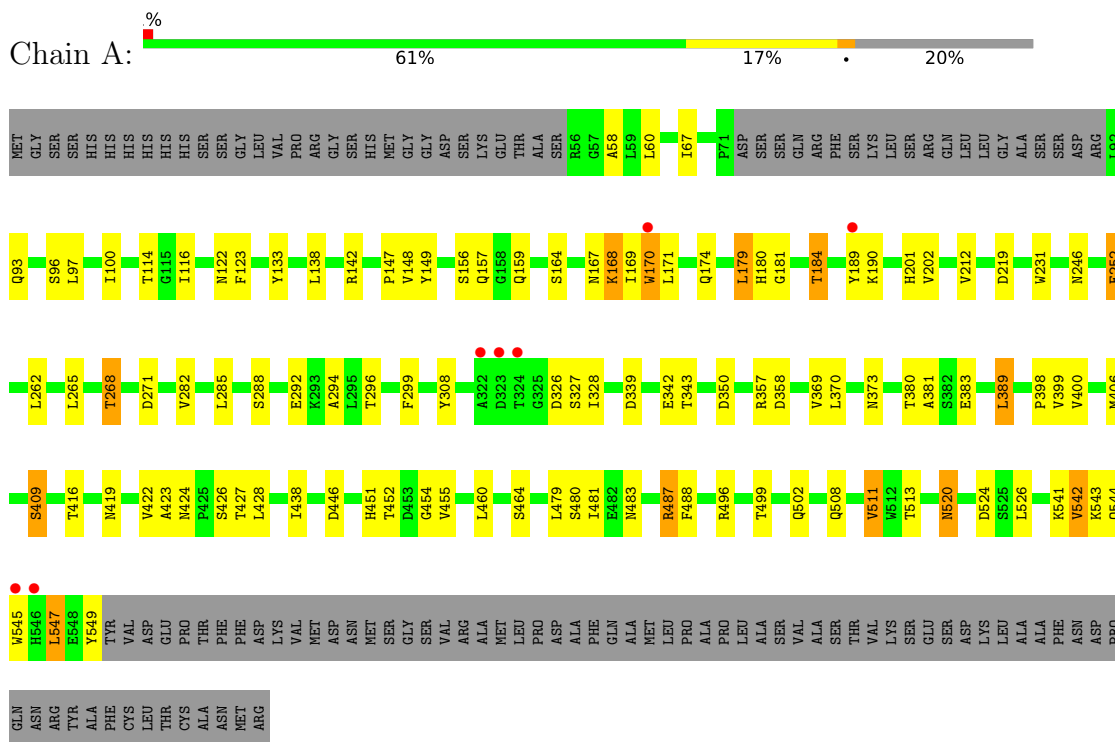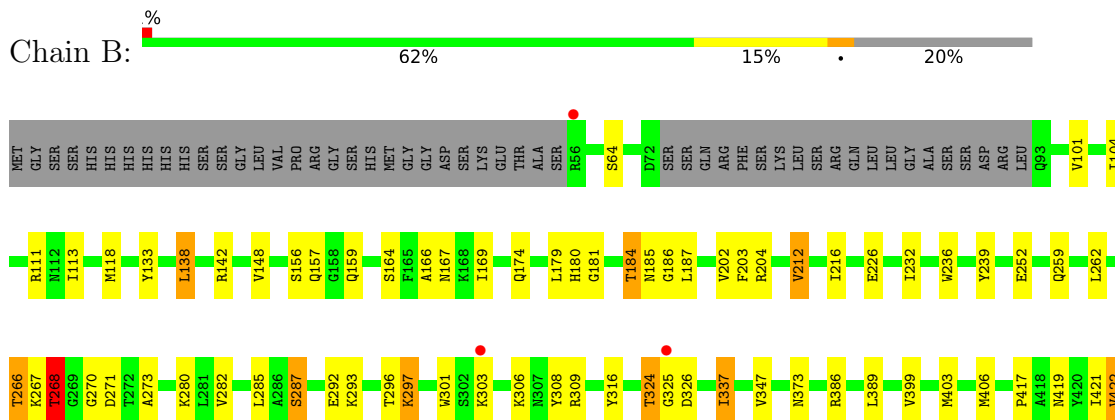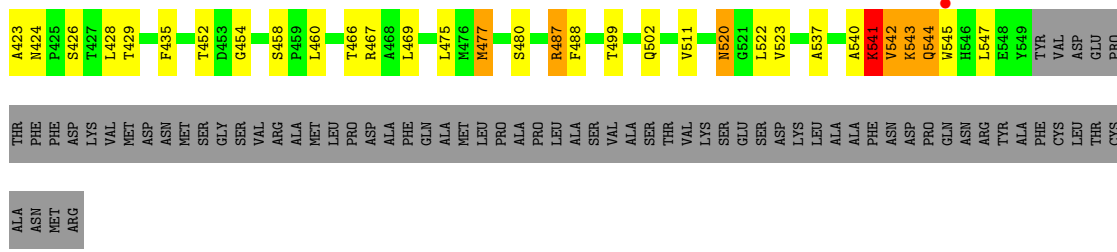
SER GLU SER ASP LYS LEU ALA ALA PHE ASN ASP PRO GLN ASN ARG TYR ALA PHE CYS LEU THR CYS ALA ASN MET ARG

4   Data and refinement statistics   ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	94.17Å 153.49Å 100.66Å 90.00° 104.25° 90.00°	Depositor
Resolution (Å)	50.00 – 2.55 48.24 – 2.55	Depositor EDS
% Data completeness (in resolution range)	96.2 (50.00-2.55) 96.2 (48.24-2.55)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	0.12	Depositor
$< I/\sigma(I) >$ [1]	4.07 (at 2.54Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.214 , 0.252 0.208 , 0.264	Depositor DCC
R_{free} test set	4338 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	29.4	Xtriage
Anisotropy	0.551	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 39.0	EDS
L-test for twinning[2]	$< \lvert L \rvert > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	14865	wwPDB-VP
Average B, all atoms (Å^2)	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.40% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< \lvert L \rvert >$, $< L^2 >$ 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	0.73	1/3654 (0.0%)	0.78	2/4960 (0.0%)
1	B	0.74	0/3650	0.82	0/4956
1	C	0.79	3/3673 (0.1%)	0.80	0/4988
1	D	0.71	0/3668	0.76	0/4979
All	All	0.74	4/14645 (0.0%)	0.79	2/19883 (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	A	0	1
1	B	0	1
1	C	0	1
1	D	0	2
All	All	0	5

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	208	TYR	CD2-CE2	-5.94	1.30	1.39
1	A	252	GLU	CG-CD	5.51	1.60	1.51
1	C	226	GLU	CG-CD	5.51	1.60	1.51
1	C	208	TYR	CD1-CE1	-5.16	1.31	1.39

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	219	ASP	CB-CG-OD1	5.14	122.93	118.30
1	A	542	VAL	N-CA-C	5.14	124.87	111.00

There are no chirality outliers.

WORLDWIDE PDB
PROTEIN DATA BANK

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	541	LYS	Peptide
1	B	541	LYS	Peptide
1	C	541	LYS	Peptide
1	D	340	GLY	Peptide
1	D	541	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	3587	0	3550	76	0
1	B	3583	0	3539	86	0
1	C	3606	0	3549	66	0
1	D	3601	0	3559	92	0
2	A	116	0	0	4	0
2	B	127	0	0	3	0
2	C	144	0	0	6	0
2	D	101	0	0	4	0
All	All	14865	0	14197	301	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 11.

The worst 5 of 301 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:542:VAL:HG12	1:B:543:LYS:N	1.40	1.18
1:D:403:MET:CE	1:D:423:ALA:HB2	1.72	1.17
1:B:542:VAL:CG1	1:B:543:LYS:N	2.08	1.12
1:B:542:VAL:CG1	1:B:543:LYS:H	1.54	1.10
1:D:403:MET:HE2	1:D:423:ALA:HB2	1.07	1.07

There are no symmetry-related clashes.

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	470/593 (79%)	440 (94%)	28 (6%)	2 (0%)	34	46
1	B	470/593 (79%)	442 (94%)	22 (5%)	6 (1%)	12	16
1	C	474/593 (80%)	447 (94%)	22 (5%)	5 (1%)	14	19
1	D	472/593 (80%)	442 (94%)	22 (5%)	8 (2%)	9	11
All	All	1886/2372 (80%)	1771 (94%)	94 (5%)	21 (1%)	14	19

5 of 21 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	508	GLN
1	B	541	LYS
1	B	542	VAL
1	C	324	THR
1	C	540	ALA

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	374/475 (79%)	339 (91%)	35 (9%)	8	10
1	B	373/475 (78%)	339 (91%)	34 (9%)	9	11
1	C	374/475 (79%)	337 (90%)	37 (10%)	8	9
1	D	376/475 (79%)	343 (91%)	33 (9%)	10	12
All	All	1497/1900 (79%)	1358 (91%)	139 (9%)	9	10

5 of 139 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	232	ILE
1	D	268	THR
1	D	460	LEU
1	B	266	THR
1	B	262	LEU

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	185	ASN
1	C	451	HIS
1	D	424	ASN
1	C	237	GLN
1	C	346	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 (i)

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, 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(Å²)	Q<0.9
1	A	474/593 (79%)	-0.21	7 (1%)	73	79	14, 28, 51, 65	0
1	B	474/593 (79%)	-0.26	4 (0%)	86	89	14, 25, 48, 59	0
1	C	478/593 (80%)	-0.21	12 (2%)	57	63	14, 25, 49, 62	0
1	D	476/593 (80%)	-0.21	5 (1%)	80	85	14, 27, 49, 60	0
All	All	1902/2372 (80%)	-0.22	28 (1%)	73	79	14, 26, 49, 65	0

The worst 5 of 28 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	545	TRP	4.8
1	D	545	TRP	4.3
1	D	343	THR	4.2
1	C	550	TYR	4.0
1	C	545	TRP	3.7

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 ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

6.5 Other polymers ⓘ

There are no such residues in this entry.