wwPDB X-ray Structure Validation Summary Report (i)

Jun 16, 2024 – 08:13 AM EDT

PDB ID	:	2DKF
Title	:	Crystal Structure of TTHA0252 from Thermus thermophilus HB8, a RNA Degradation Protein of the Metallo-beta-lactamase Superfamily
Authors	:	Ishikawa, I.; Nakagawa, N.; Kuramitsu, S.; Yokoyama, S.; Masui, R.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2006-04-10
Resolution	:	2.80 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:

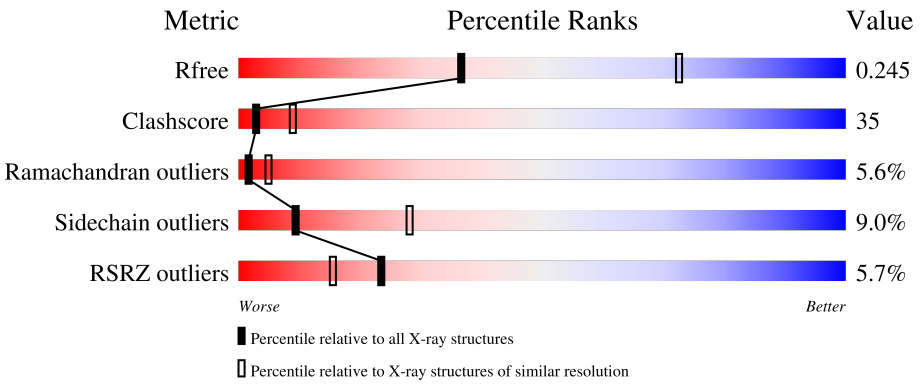
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.80 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	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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	431	48% 41% 10% •
1	B	431	% 50% 42% 7%
1	C	431	10% 45% 48% 7%
1	D	431	11% 40% 50% 9% •

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 13404 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 metallo-beta-lactamase superfamily protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	431	3326	2127	597	594	1	7	0	0	0
1	B	431	3326	2127	597	594	1	7	0	0	0
1	C	431	3326	2127	597	594	1	7	0	0	0
1	D	431	3326	2127	597	594	1	7	0	0	0

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	MET	MODIFIED RESIDUE	GB 55771634
A	32	MSE	MET	MODIFIED RESIDUE	GB 55771634
A	90	MSE	MET	MODIFIED RESIDUE	GB 55771634
A	101	MSE	MET	MODIFIED RESIDUE	GB 55771634
A	253	MSE	MET	MODIFIED RESIDUE	GB 55771634
A	307	MSE	MET	MODIFIED RESIDUE	GB 55771634
A	315	MSE	MET	MODIFIED RESIDUE	GB 55771634
B	1	MSE	MET	MODIFIED RESIDUE	GB 55771634
B	32	MSE	MET	MODIFIED RESIDUE	GB 55771634
B	90	MSE	MET	MODIFIED RESIDUE	GB 55771634
B	101	MSE	MET	MODIFIED RESIDUE	GB 55771634
B	253	MSE	MET	MODIFIED RESIDUE	GB 55771634
B	307	MSE	MET	MODIFIED RESIDUE	GB 55771634
B	315	MSE	MET	MODIFIED RESIDUE	GB 55771634
C	1	MSE	MET	MODIFIED RESIDUE	GB 55771634
C	32	MSE	MET	MODIFIED RESIDUE	GB 55771634
C	90	MSE	MET	MODIFIED RESIDUE	GB 55771634
C	101	MSE	MET	MODIFIED RESIDUE	GB 55771634
C	253	MSE	MET	MODIFIED RESIDUE	GB 55771634
C	307	MSE	MET	MODIFIED RESIDUE	GB 55771634
C	315	MSE	MET	MODIFIED RESIDUE	GB 55771634

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	1	MSE	MET	MODIFIED RESIDUE	GB 55771634
D	32	MSE	MET	MODIFIED RESIDUE	GB 55771634
D	90	MSE	MET	MODIFIED RESIDUE	GB 55771634
D	101	MSE	MET	MODIFIED RESIDUE	GB 55771634
D	253	MSE	MET	MODIFIED RESIDUE	GB 55771634
D	307	MSE	MET	MODIFIED RESIDUE	GB 55771634
D	315	MSE	MET	MODIFIED RESIDUE	GB 55771634

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	2	Total Zn 2 2	0	0
2	B	2	Total Zn 2 2	0	0
2	C	2	Total Zn 2 2	0	0
2	D	2	Total Zn 2 2	0	0

- Molecule 3 is water.

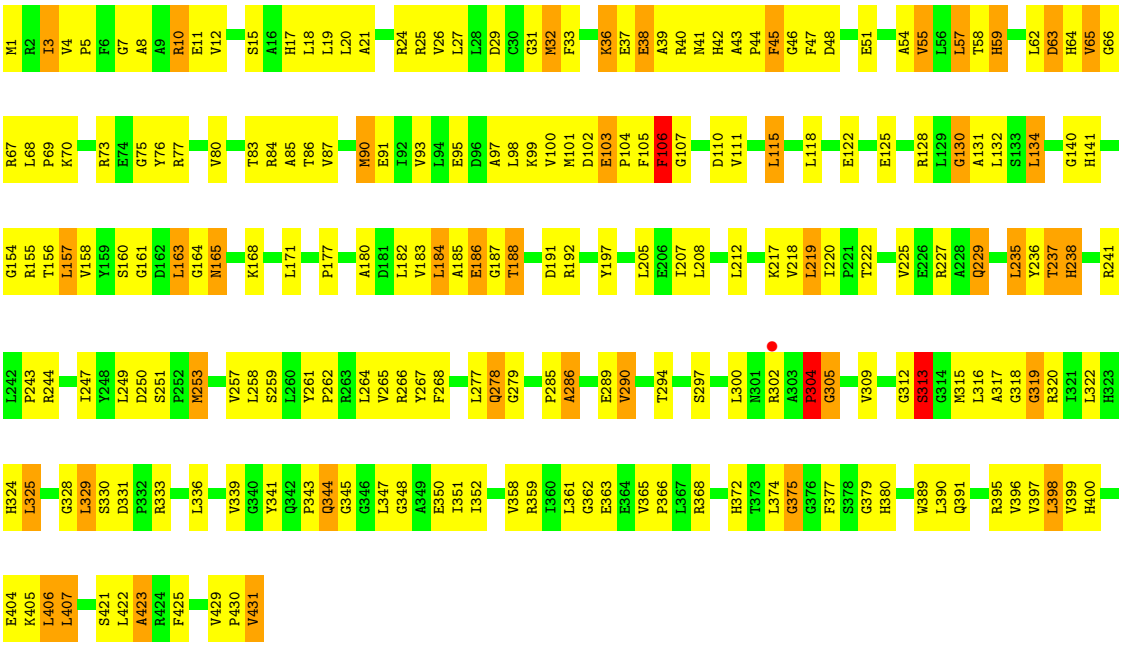
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	32	Total O 32 32	0	0
3	B	28	Total O 28 28	0	0
3	C	16	Total O 16 16	0	0
3	D	16	Total O 16 16	0	0

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.

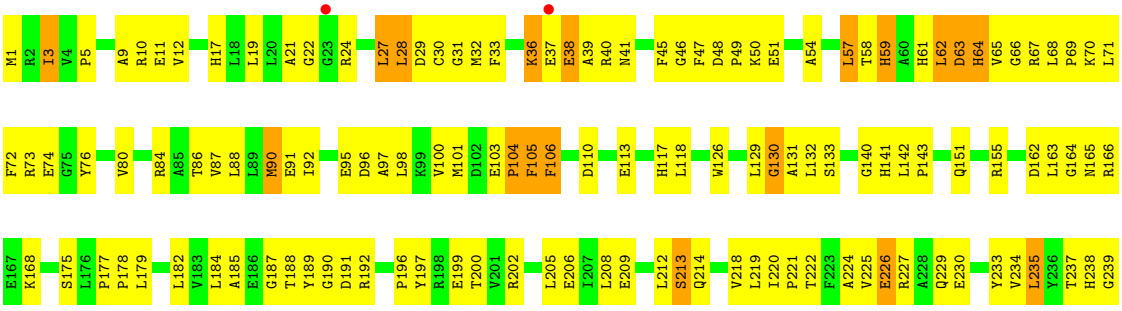
- Molecule 1: metallo-beta-lactamase superfamily protein

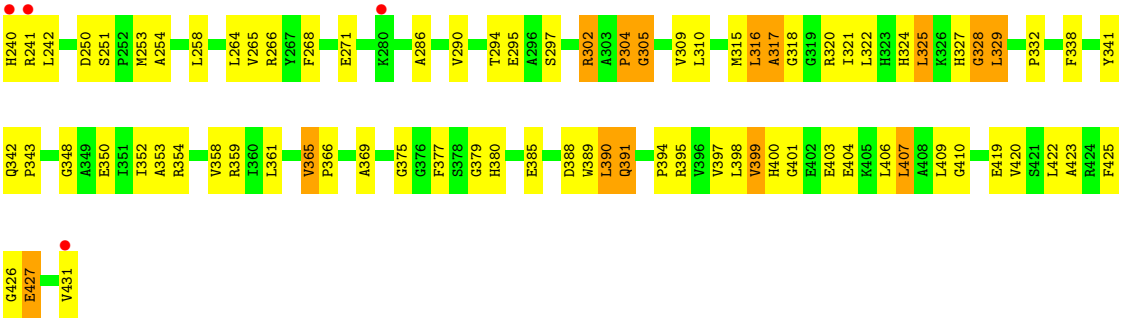
Chain A:



- Molecule 1: metallo-beta-lactamase superfamily protein

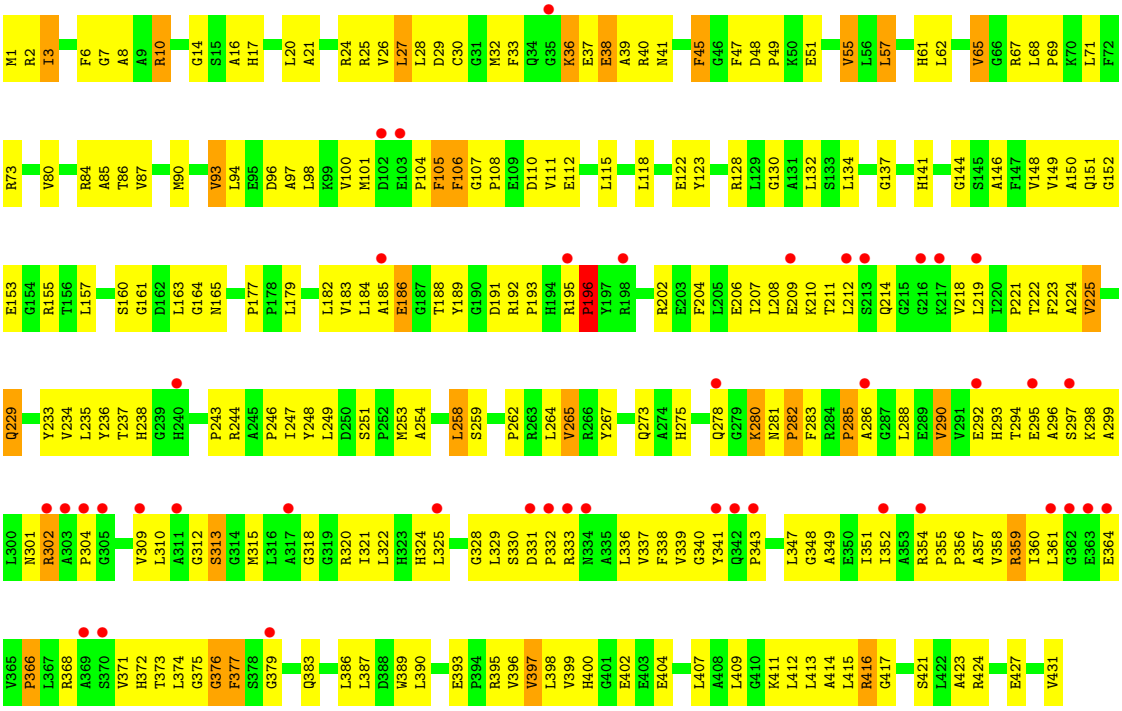
Chain B:





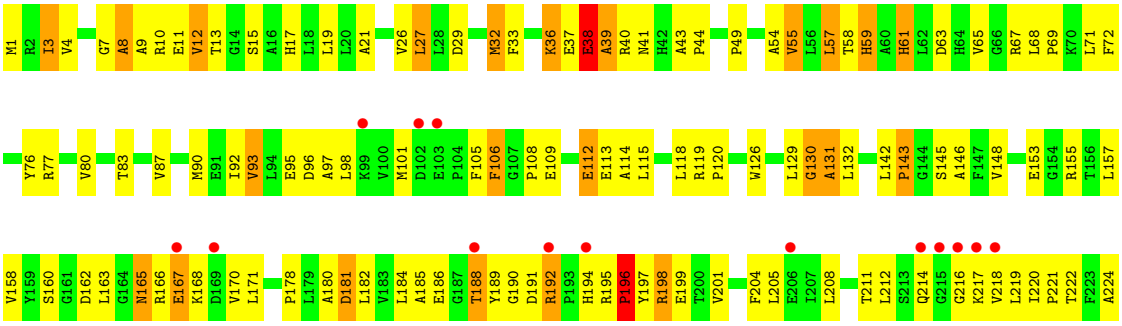
- Molecule 1: metallo-beta-lactamase superfamily protein

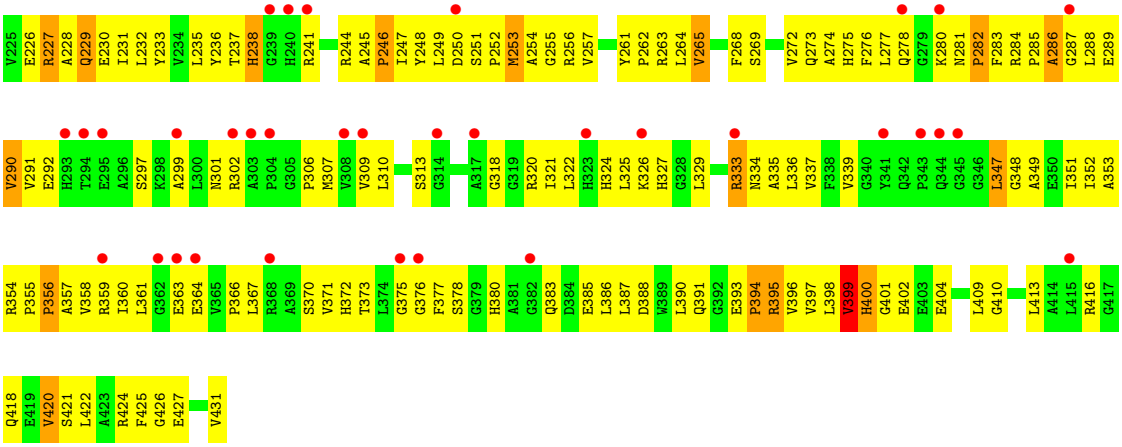
Chain C:



- Molecule 1: metallo-beta-lactamase superfamily protein

Chain D:




V225 E226 R227 A228 Q229 E230 I231 L232 Y233 V234 L235 Y236 T237 H238 G239 H240 R241 R244 A245 P246 I247 Y248 L249 D250 S251 P252 M253 A254 G255 R256 V257 Y261 P262 R263 L264 V265 F268 S269 V272 Q273 A274 H275 F276 L277 Q278 G279 K280 N281 P282 F283 R284 P285 A286 G287 L288 E289
V290 V291 E292 H293 T294 E295 A296 S297 K298 A299 L300 N301 R302 A303 P304 G305 P306 M307 V308 V309 L310 S313 G314 A317 G318 G319 R320 I321 L322 H323 H324 L325 K326 H327 G328 L329 R333 N334 A335 L336 V337 F338 V339 G340 Y341 Q342 P343 Q344 G345 G346 L347 G348 A349 E350 I351 I352 A353
R354 P355 P356 A357 V358 R359 I360 L361 G362 E363 E364 V365 P366 L367 R368 A369 S370 V371 H372 T373 L374 G375 G376 F377 S378 G379 H380 A381 G382 Q383 D384 E385 L386 L387 D388 W389 L390 Q391 G392 E393 P394 R395 V396 V397 L398 V399 H400 G401 E402 E403 E404 L409 G410 L413 A414 L415 R416 G417
Q418 E419 V420 S421 L422 A423 R424 F425 G426 E427 V431

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	143.22Å 147.10Å 121.23Å 90.00° 109.25° 90.00°	Depositor
Resolution (Å)	19.99 – 2.80 19.99 – 2.80	Depositor EDS
% Data completeness (in resolution range)	96.0 (19.99-2.80) 96.1 (19.99-2.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	28.81 (at 2.79Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.242 , 0.285 0.246 , 0.245	Depositor DCC
R_{free} test set	2847 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å^2)	59.7	Xtriage
Anisotropy	0.048	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 61.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	13404	wwPDB-VP
Average B, all atoms (Å^2)	67.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.99% 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: ZN

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	0.59	4/3401 (0.1%)	0.84	5/4603 (0.1%)
1	B	0.59	1/3401 (0.0%)	0.80	2/4603 (0.0%)
1	C	0.47	0/3401	0.67	2/4603 (0.0%)
1	D	0.45	1/3401 (0.0%)	0.65	0/4603
All	All	0.53	6/13604 (0.0%)	0.75	9/18412 (0.0%)

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	253	MSE	CG-SE	-6.77	1.72	1.95
1	A	90	MSE	SE-CE	-6.12	1.59	1.95
1	B	90	MSE	CG-SE	-6.07	1.74	1.95
1	A	90	MSE	CG-SE	-5.91	1.75	1.95
1	A	32	MSE	SE-CE	-5.21	1.64	1.95

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	375	GLY	N-CA-C	-8.06	92.94	113.10
1	B	375	GLY	N-CA-C	-6.94	95.75	113.10
1	C	161	GLY	N-CA-C	-5.78	98.66	113.10
1	A	161	GLY	N-CA-C	-5.76	98.69	113.10
1	C	105	PHE	N-CA-C	-5.18	97.00	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	3326	0	3351	206	0
1	B	3326	0	3351	224	0
1	C	3326	0	3351	242	0
1	D	3326	0	3351	272	0
2	A	2	0	0	0	0
2	B	2	0	0	0	0
2	C	2	0	0	0	0
2	D	2	0	0	0	0
3	A	32	0	0	3	0
3	B	28	0	0	0	0
3	C	16	0	0	1	0
3	D	16	0	0	0	0
All	All	13404	0	13404	933	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 35.

The worst 5 of 933 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:C:37:GLU:HB3	1:C:40:ARG:HH11	0.97	1.13
1:C:37:GLU:HB3	1:C:40:ARG:NH1	1.76	1.00
1:D:227:ARG:HB2	1:D:227:ARG:HH21	1.25	0.99
1:C:73:ARG:HE	1:C:106:PHE:HA	1.24	0.98
1:C:235:LEU:HD23	1:C:247:ILE:HD13	1.45	0.97

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	429/431 (100%)	361 (84%)	45 (10%)	23 (5%)	2 5
1	B	429/431 (100%)	364 (85%)	45 (10%)	20 (5%)	2 7
1	C	429/431 (100%)	340 (79%)	64 (15%)	25 (6%)	1 4
1	D	429/431 (100%)	333 (78%)	68 (16%)	28 (6%)	1 3
All	All	1716/1724 (100%)	1398 (82%)	222 (13%)	96 (6%)	2 5

5 of 96 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	38	GLU
1	A	59	HIS
1	A	106	PHE
1	A	286	ALA
1	A	304	PRO

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	342/335 (102%)	299 (87%)	43 (13%)	4 14
1	B	342/335 (102%)	313 (92%)	29 (8%)	10 31
1	C	342/335 (102%)	317 (93%)	25 (7%)	14 38
1	D	342/335 (102%)	316 (92%)	26 (8%)	13 36
All	All	1368/1340 (102%)	1245 (91%)	123 (9%)	9 28

5 of 123 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	184	LEU
1	D	115	LEU
1	B	427	GLU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	D	112	GLU
1	D	229	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 35 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	194	HIS
1	D	214	GLN
1	D	278	GLN
1	B	229	GLN
1	B	165	ASN

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)

Of 8 ligands modelled in this entry, 8 are monoatomic - leaving 0 for Mogul analysis.

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)

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, 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	424/431 (98%)	-0.29	1 (0%)	95	94	21, 44, 66, 81	0
1	B	424/431 (98%)	-0.19	6 (1%)	75	70	22, 45, 77, 93	0
1	C	424/431 (98%)	0.41	42 (9%)	7	4	33, 78, 134, 141	0
1	D	424/431 (98%)	0.57	48 (11%)	5	3	36, 90, 146, 160	0
All	All	1696/1724 (98%)	0.13	97 (5%)	23	15	21, 56, 136, 160	0

The worst 5 of 97 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	333	ARG	8.2
1	D	216	GLY	6.4
1	D	215	GLY	6.3
1	C	332	PRO	4.7
1	C	341	TYR	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	ZN	D	432	1/1	0.81	0.23	59,59,59,59	1
2	ZN	B	433	1/1	0.86	0.18	42,42,42,42	1
2	ZN	A	433	1/1	0.86	0.21	37,37,37,37	1
2	ZN	C	432	1/1	0.89	0.20	48,48,48,48	1
2	ZN	D	433	1/1	0.90	0.17	73,73,73,73	1
2	ZN	B	432	1/1	0.91	0.22	28,28,28,28	1
2	ZN	C	433	1/1	0.95	0.19	44,44,44,44	1
2	ZN	A	432	1/1	0.97	0.12	19,19,19,19	1

6.5 Other polymers (i)

There are no such residues in this entry.