



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 28, 2020 – 08:15 pm BST

PDB ID : 6KQR  
Title : A pre-assembled molecular-helical Cascade backbone of Csy3 subunits from *Zymomonas mobilis*  
Authors : Gu, D.H.; Ha, S.C.; Kim, J.S.  
Deposited on : 2019-08-18  
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
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

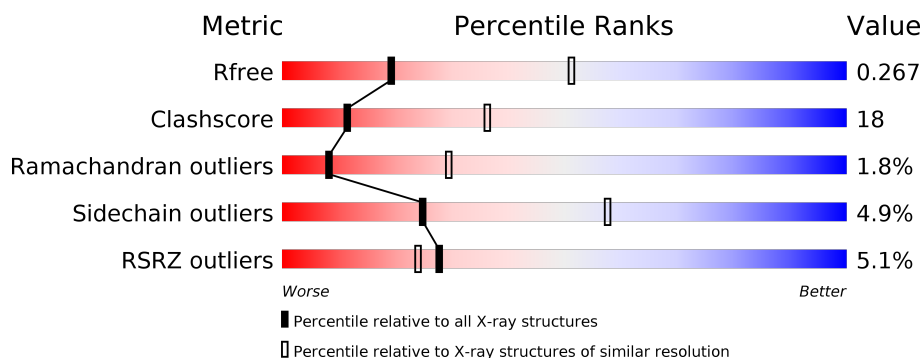
# 1 Overall quality at a glance

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	348	<div> <div>4%</div> <div> <div></div> <div>65%</div> <div>27%</div> <div>• • •</div> </div> </div>
1	B	348	<div> <div>5%</div> <div> <div></div> <div>56%</div> <div>36%</div> <div>• •</div> </div> </div>
1	C	348	<div> <div>3%</div> <div> <div></div> <div>68%</div> <div>25%</div> <div>• •</div> </div> </div>
1	D	348	<div> <div>5%</div> <div> <div></div> <div>60%</div> <div>33%</div> <div>• •</div> </div> </div>
1	E	348	<div> <div>7%</div> <div> <div></div> <div>55%</div> <div>39%</div> <div>• •</div> </div> </div>
1	F	348	<div> <div>5%</div> <div> <div></div> <div>63%</div> <div>31%</div> <div>• •</div> </div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	348	<div><div></div><div>5%</div><div>65%</div><div>28%</div><div>...</div></div>

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 18135 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 CRISPR-associated protein Csy3 family.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	333	Total	C	N	O	S	0	0	0
			2589	1643	452	491	3			
1	B	333	Total	C	N	O	S	0	0	0
			2589	1643	452	491	3			
1	C	333	Total	C	N	O	S	0	0	0
			2589	1643	452	491	3			
1	D	333	Total	C	N	O	S	0	0	0
			2589	1643	452	491	3			
1	E	334	Total	C	N	O	S	0	0	0
			2593	1645	453	492	3			
1	F	334	Total	C	N	O	S	0	0	0
			2593	1645	453	492	3			
1	G	334	Total	C	N	O	S	0	0	0
			2593	1645	453	492	3			

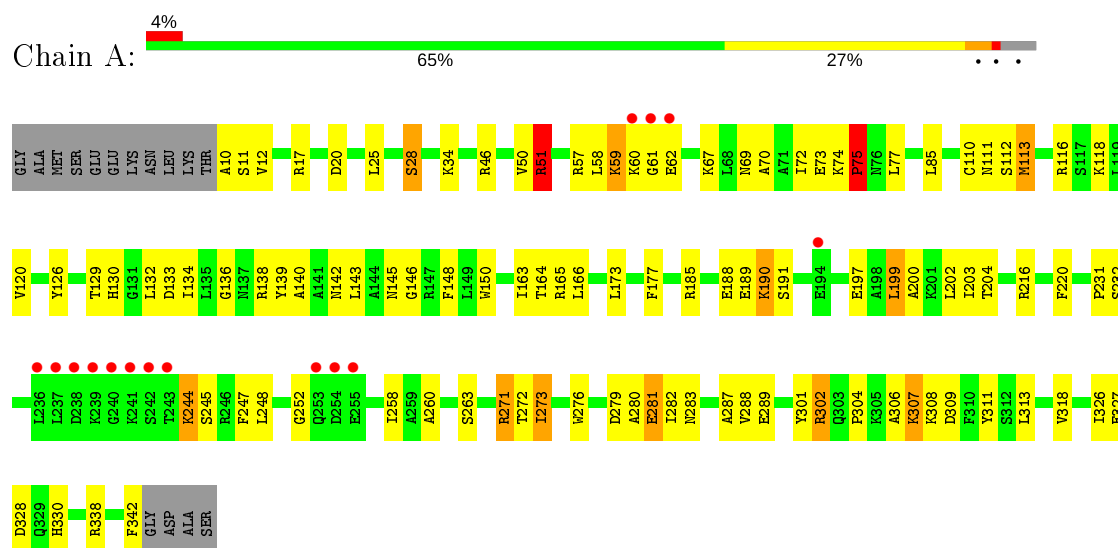
There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	expression tag	UNP A0A2R4RXD4
A	0	ALA	-	expression tag	UNP A0A2R4RXD4
B	-1	GLY	-	expression tag	UNP A0A2R4RXD4
B	0	ALA	-	expression tag	UNP A0A2R4RXD4
C	-1	GLY	-	expression tag	UNP A0A2R4RXD4
C	0	ALA	-	expression tag	UNP A0A2R4RXD4
D	-1	GLY	-	expression tag	UNP A0A2R4RXD4
D	0	ALA	-	expression tag	UNP A0A2R4RXD4
E	-1	GLY	-	expression tag	UNP A0A2R4RXD4
E	0	ALA	-	expression tag	UNP A0A2R4RXD4
F	-1	GLY	-	expression tag	UNP A0A2R4RXD4
F	0	ALA	-	expression tag	UNP A0A2R4RXD4
G	-1	GLY	-	expression tag	UNP A0A2R4RXD4
G	0	ALA	-	expression tag	UNP A0A2R4RXD4

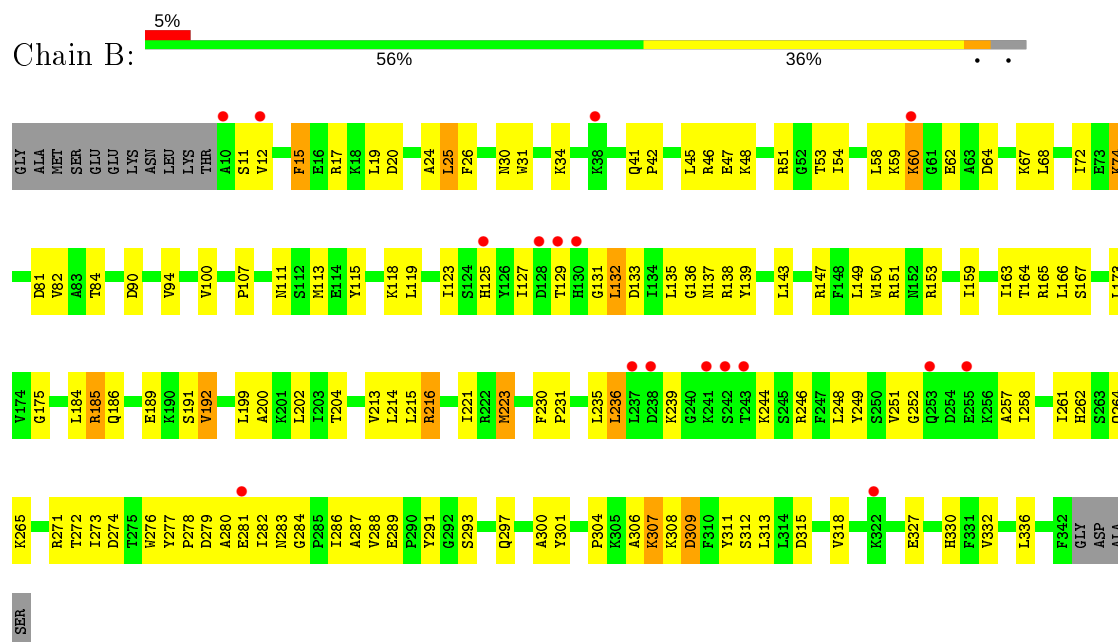
### 3 Residue-property plots [i](#)

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.

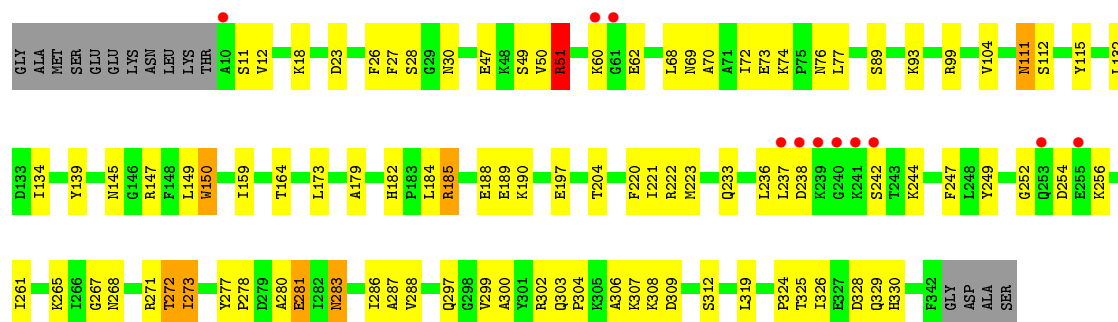
- Molecule 1: CRISPR-associated protein Csy3 family



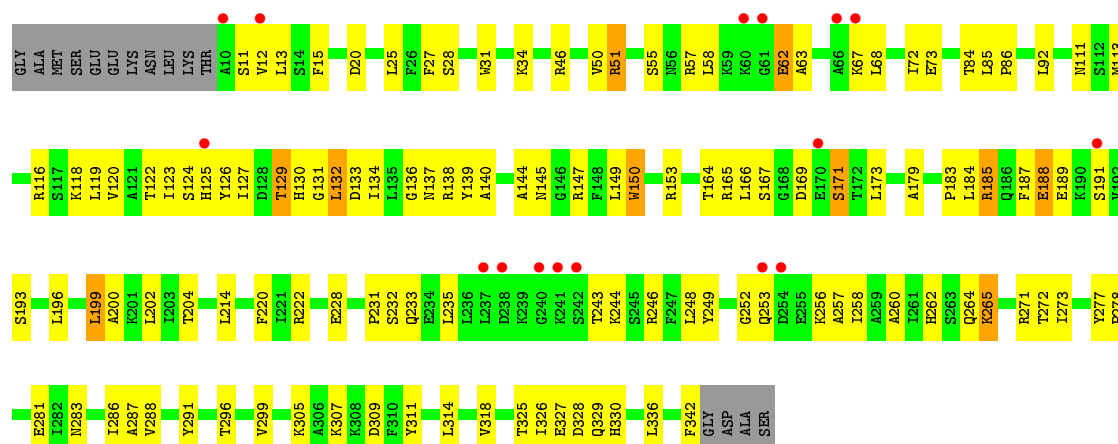
- Molecule 1: CRISPR-associated protein Csy3 family



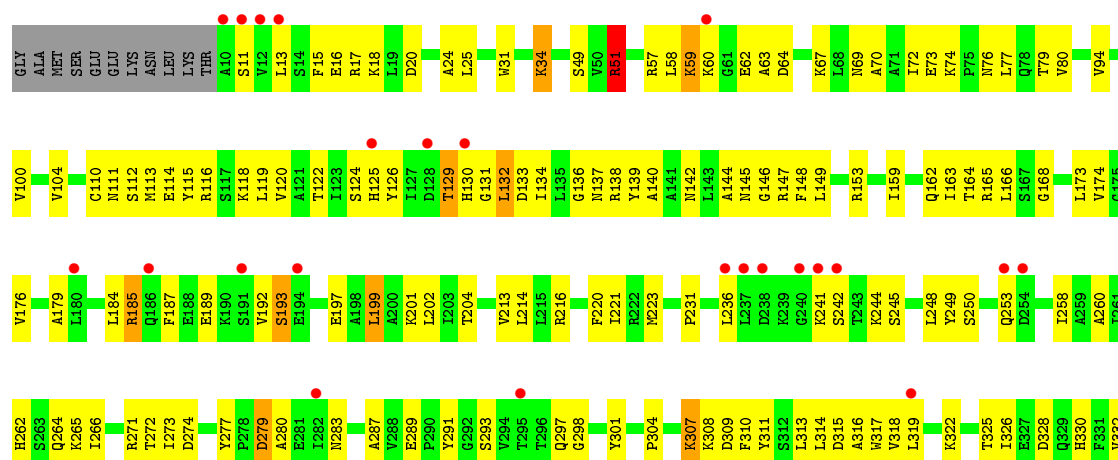
- Molecule 1: CRISPR-associated protein Csy3 family

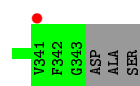


• Molecule 1: CRISPR-associated protein Csy3 family

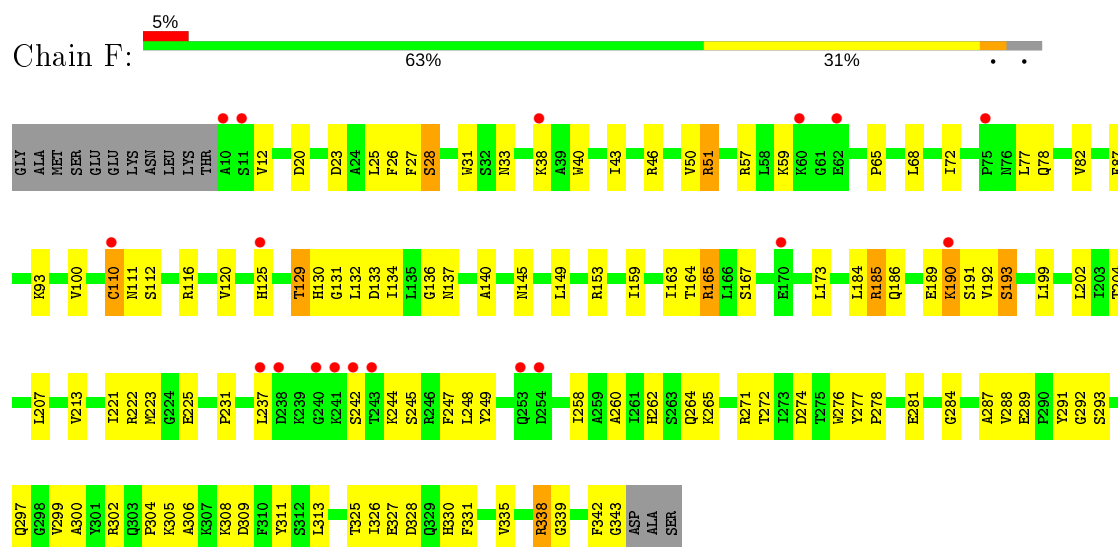


• Molecule 1: CRISPR-associated protein Csy3 family





• Molecule 1: CRISPR-associated protein Csy3 family



• Molecule 1: CRISPR-associated protein Csy3 family



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	82.18Å 116.16Å 115.94Å 99.37° 103.12° 103.35°	Depositor
Resolution (Å)	15.00 – 2.90 14.99 – 2.90	Depositor EDS
% Data completeness (in resolution range)	98.2 (15.00-2.90) 98.2 (14.99-2.90)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.15 (at 2.91Å)	Xtriage
Refinement program	PHENIX 1.11.1_2575	Depositor
R, $R_{free}$	0.214 , 0.267 0.214 , 0.267	Depositor DCC
$R_{free}$ test set	3417 reflections (3.99%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	67.3	Xtriage
Anisotropy	0.783	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 59.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.012 for -h,-l,-k	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	18135	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	77.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.79% 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

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	0.60	1/2640 (0.0%)	0.73	2/3575 (0.1%)
1	B	0.46	0/2640	0.67	1/3575 (0.0%)
1	C	0.54	0/2640	0.70	1/3575 (0.0%)
1	D	0.67	1/2640 (0.0%)	0.74	2/3575 (0.1%)
1	E	0.50	0/2644	0.69	1/3580 (0.0%)
1	F	0.52	0/2644	0.73	1/3580 (0.0%)
1	G	0.52	1/2644 (0.0%)	0.68	2/3580 (0.1%)
All	All	0.55	3/18492 (0.0%)	0.70	10/25040 (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	C	0	1
1	D	0	1
All	All	0	2

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	86	PRO	N-CD	-22.37	1.16	1.47
1	A	75	PRO	N-CD	-16.40	1.24	1.47
1	G	75	PRO	N-CD	-9.85	1.34	1.47

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	189	GLU	N-CA-C	7.00	129.91	111.00
1	F	23	ASP	CB-CG-OD1	6.97	124.58	118.30
1	G	74	LYS	CD-CE-NZ	6.85	127.45	111.70
1	A	51	ARG	NE-CZ-NH1	-6.68	116.96	120.30

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	E	51	ARG	NE-CZ-NH1	-6.10	117.25	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	272	THR	Peptide
1	D	188	GLU	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	2589	0	2591	95	1
1	B	2589	0	2591	121	0
1	C	2589	0	2591	76	0
1	D	2589	0	2591	94	1
1	E	2593	0	2594	128	0
1	F	2593	0	2594	108	0
1	G	2593	0	2594	97	0
All	All	18135	0	18146	639	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 18.

The worst 5 of 639 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance ( $\text{\AA}$ )	Clash overlap ( $\text{\AA}$ )
1:G:74:LYS:CD	1:G:75:PRO:HD2	1.62	1.30
1:G:74:LYS:HD2	1:G:75:PRO:CD	1.63	1.27
1:A:74:LYS:CD	1:A:75:PRO:HD2	1.79	1.13
1:B:47:GLU:OE2	1:B:244:LYS:NZ	1.82	1.11
1:A:74:LYS:HD2	1:A:75:PRO:HD2	1.19	1.09

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:A:138:ARG:NH2	1:D:281:GLU:OE1[1_565]	2.19	0.01

## 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	331/348 (95%)	307 (93%)	15 (4%)	9 (3%)	5	19
1	B	331/348 (95%)	309 (93%)	18 (5%)	4 (1%)	13	40
1	C	331/348 (95%)	301 (91%)	26 (8%)	4 (1%)	13	40
1	D	331/348 (95%)	302 (91%)	21 (6%)	8 (2%)	6	22
1	E	332/348 (95%)	308 (93%)	17 (5%)	7 (2%)	7	26
1	F	332/348 (95%)	312 (94%)	17 (5%)	3 (1%)	17	48
1	G	332/348 (95%)	304 (92%)	22 (7%)	6 (2%)	8	29
All	All	2320/2436 (95%)	2143 (92%)	136 (6%)	41 (2%)	8	29

5 of 41 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	190	LYS
1	A	281	GLU
1	A	307	LYS
1	B	192	VAL
1	B	280	ALA

### 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	274/286 (96%)	260 (95%)	14 (5%)	24	56
1	B	274/286 (96%)	262 (96%)	12 (4%)	28	61
1	C	274/286 (96%)	262 (96%)	12 (4%)	28	61
1	D	274/286 (96%)	263 (96%)	11 (4%)	31	65
1	E	274/286 (96%)	259 (94%)	15 (6%)	21	53
1	F	274/286 (96%)	258 (94%)	16 (6%)	20	50
1	G	274/286 (96%)	260 (95%)	14 (5%)	24	56
All	All	1918/2002 (96%)	1824 (95%)	94 (5%)	25	57

5 of 94 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	132	LEU
1	E	51	ARG
1	G	185	ARG
1	D	171	SER
1	D	265	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	76	ASN
1	E	111	ASN
1	F	233	GLN
1	E	69	ASN
1	F	186	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	333/348 (95%)	-0.11	15 (4%) 33 29	46, 66, 122, 160	0
1	B	333/348 (95%)	0.01	17 (5%) 28 24	45, 78, 126, 167	0
1	C	333/348 (95%)	-0.19	11 (3%) 46 41	43, 66, 108, 156	0
1	D	333/348 (95%)	-0.08	16 (4%) 30 27	43, 63, 125, 179	0
1	E	334/348 (95%)	0.06	24 (7%) 15 11	45, 77, 127, 171	0
1	F	334/348 (95%)	-0.04	18 (5%) 25 22	47, 70, 131, 164	0
1	G	334/348 (95%)	0.04	19 (5%) 23 19	48, 79, 128, 199	0
All	All	2334/2436 (95%)	-0.05	120 (5%) 28 24	43, 72, 126, 199	0

The worst 5 of 120 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	10	ALA	10.7
1	D	10	ALA	9.8
1	F	242	SER	7.1
1	G	241	LYS	6.8
1	E	241	LYS	6.3

### 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 [i](#)

There are no such residues in this entry.