W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Nov 26, 2024 – 02:10 PM JST

PDB ID	:	8JL2
Title	:	Crystal structure of the Green fluorescent protein SE_A277 variant at pH 9.5
Authors	:	Shin, S.C.
Deposited on	:	2023-06-02
Resolution	:	1.76 Å(reported)

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

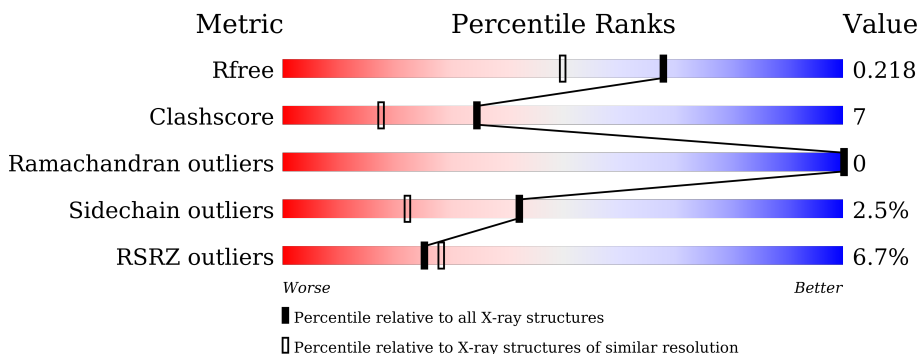|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.21
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.004 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.40

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 1.76 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	Percentile Ranks	Value
Rfree		0.218
Clashscore		7
Ramachandran outliers		0
Sidechain outliers		2.5%
RSRZ outliers		6.7%

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$_{free}$	164625	2888 (1.76-1.76)
Clashscore	180529	3097 (1.76-1.76)
Ramachandran outliers	177936	3072 (1.76-1.76)
Sidechain outliers	177891	3072 (1.76-1.76)
RSRZ outliers	164620	2887 (1.76-1.76)

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	236	6% / 84% / 11% / • 5%
1	B	236	7% / 83% / 11% / • 5%
1	C	236	5% / 84% / 10% / • 5%
1	D	236	8% / 81% / 13% / • 5%

2 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 8039 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 Green fluorescent protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	225	Total	C	N	O	S	0	0	0
			1807	1153	304	344	6			
1	B	225	Total	C	N	O	S	0	0	0
			1807	1153	304	344	6			
1	C	225	Total	C	N	O	S	0	0	0
			1807	1153	304	344	6			
1	D	225	Total	C	N	O	S	0	0	0
			1807	1153	304	344	6			

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	64	LEU	PHE	engineered mutation	UNP P42212
A	66	CRO	SER	chromophore	UNP P42212
A	66	CRO	TYR	chromophore	UNP P42212
A	66	CRO	GLY	chromophore	UNP P42212
A	80	ARG	GLN	engineered mutation	UNP P42212
A	147	ASP	SER	engineered mutation	UNP P42212
A	149	GLN	ASN	engineered mutation	UNP P42212
A	163	ALA	VAL	engineered mutation	UNP P42212
A	175	GLY	SER	engineered mutation	UNP P42212
A	202	PHE	SER	engineered mutation	UNP P42212
A	204	THR	GLN	engineered mutation	UNP P42212
A	206	THR	ALA	engineered mutation	UNP P42212
B	64	LEU	PHE	engineered mutation	UNP P42212
B	66	CRO	SER	chromophore	UNP P42212
B	66	CRO	TYR	chromophore	UNP P42212
B	66	CRO	GLY	chromophore	UNP P42212
B	80	ARG	GLN	engineered mutation	UNP P42212
B	147	ASP	SER	engineered mutation	UNP P42212
B	149	GLN	ASN	engineered mutation	UNP P42212
B	163	ALA	VAL	engineered mutation	UNP P42212
B	175	GLY	SER	engineered mutation	UNP P42212

Continued on next page...

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

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	202	PHE	SER	engineered mutation	UNP P42212
B	204	THR	GLN	engineered mutation	UNP P42212
B	206	THR	ALA	engineered mutation	UNP P42212
C	64	LEU	PHE	engineered mutation	UNP P42212
C	66	CRO	SER	chromophore	UNP P42212
C	66	CRO	TYR	chromophore	UNP P42212
C	66	CRO	GLY	chromophore	UNP P42212
C	80	ARG	GLN	engineered mutation	UNP P42212
C	147	ASP	SER	engineered mutation	UNP P42212
C	149	GLN	ASN	engineered mutation	UNP P42212
C	163	ALA	VAL	engineered mutation	UNP P42212
C	175	GLY	SER	engineered mutation	UNP P42212
C	202	PHE	SER	engineered mutation	UNP P42212
C	204	THR	GLN	engineered mutation	UNP P42212
C	206	THR	ALA	engineered mutation	UNP P42212
D	64	LEU	PHE	engineered mutation	UNP P42212
D	66	CRO	SER	chromophore	UNP P42212
D	66	CRO	TYR	chromophore	UNP P42212
D	66	CRO	GLY	chromophore	UNP P42212
D	80	ARG	GLN	engineered mutation	UNP P42212
D	147	ASP	SER	engineered mutation	UNP P42212
D	149	GLN	ASN	engineered mutation	UNP P42212
D	163	ALA	VAL	engineered mutation	UNP P42212
D	175	GLY	SER	engineered mutation	UNP P42212
D	202	PHE	SER	engineered mutation	UNP P42212
D	204	THR	GLN	engineered mutation	UNP P42212
D	206	THR	ALA	engineered mutation	UNP P42212

- Molecule 2 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	196	Total 196	O 196	0	0
2	B	213	Total 213	O 213	0	0
2	C	196	Total 196	O 196	0	0
2	D	206	Total 206	O 206	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.
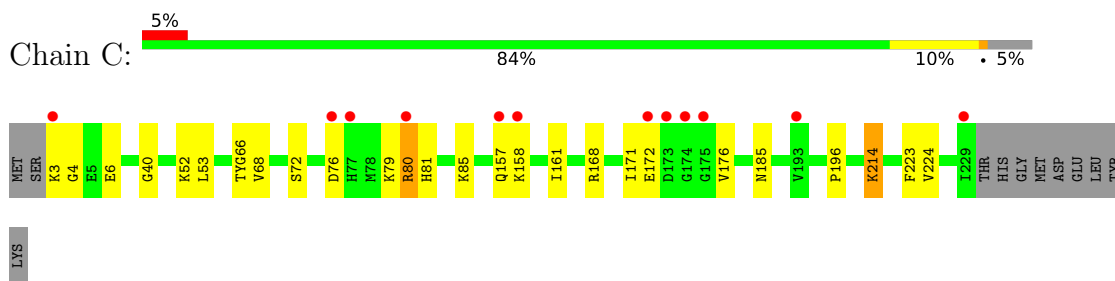
- Molecule 1: Green fluorescent protein
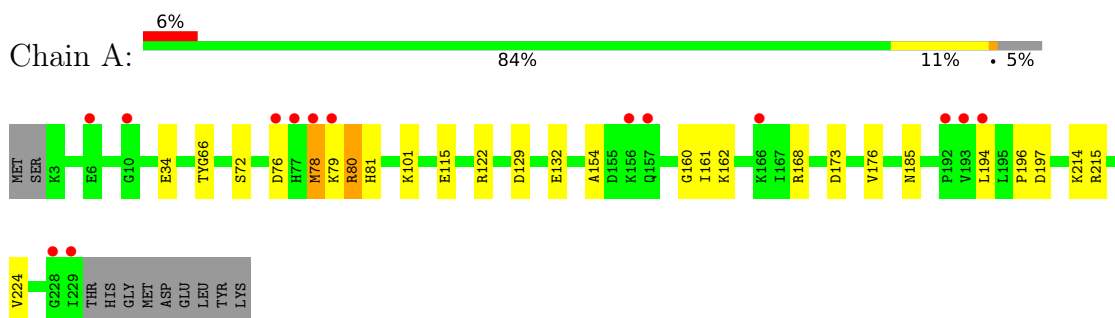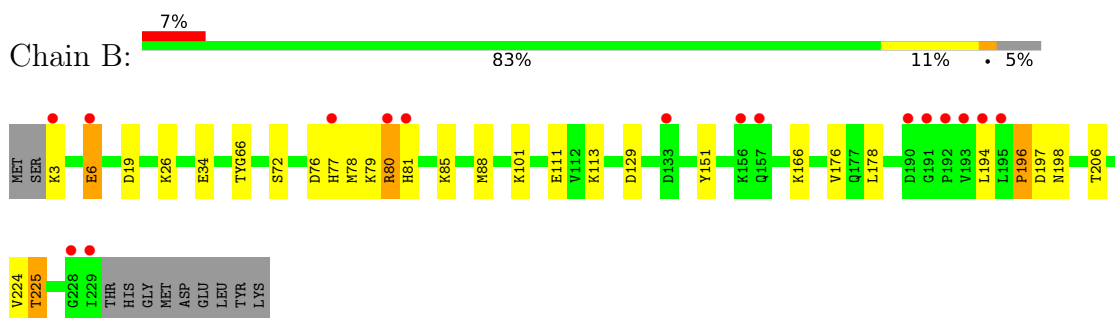
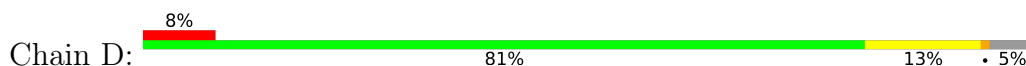

- Molecule 1: Green fluorescent protein

- Molecule 1: Green fluorescent protein

- Molecule 1: Green fluorescent protein

MET SER K3 G4 E5 K52 P56 TYG66 S72 R73 Y74 P75 D76 H77 M78 K79 R80 H81 K85 M88 P89 T97 K107 D117 D133 I136 D155 K156 Q157 K158 K162 K166 I167 R168 G174 G175 V176 D190 G191 P192 V193 L194 L195 P196

D197 K214 R215 V224 T225 G228 I229 THR HIS GLY MET ASP GLU LEU TYR LYS
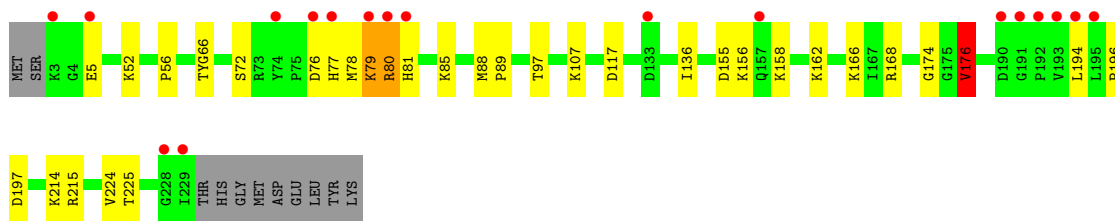
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	76.34Å 80.55Å 179.78Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	33.12 – 1.76 33.12 – 1.76	Depositor EDS
% Data completeness (in resolution range)	98.8 (33.12-1.76) 98.9 (33.12-1.76)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.98 (at 1.76Å)	Xtriage
Refinement program	PHENIX (1.19.2_4158: ???)	Depositor
R, R_{free}	0.191 , 0.218 0.195 , 0.218	Depositor DCC
R_{free} test set	108499 reflections (1.81%)	wwPDB-VP
Wilson B-factor (Å^2)	17.1	Xtriage
Anisotropy	0.191	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 41.1	EDS
L-test for twinning[2]	$< |L| > = 0.45, < L^2 > = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	8039	wwPDB-VP
Average B, all atoms (Å^2)	22.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.16% 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: CRO

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	0.50	1/1826 (0.1%)	0.72	2/2467 (0.1%)
1	B	0.52	1/1826 (0.1%)	0.67	0/2467
1	C	0.45	0/1826	0.67	0/2467
1	D	0.55	2/1826 (0.1%)	0.70	2/2467 (0.1%)
All	All	0.51	4/7304 (0.1%)	0.69	4/9868 (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
All	All	0	2

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	173	ASP	CB-CG	-7.15	1.36	1.51
1	D	176	VAL	CB-CG1	-5.92	1.40	1.52
1	B	176	VAL	CB-CG1	-5.38	1.41	1.52
1	D	174	GLY	C-O	-5.37	1.15	1.23

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	173	ASP	CB-CG-OD1	-8.35	110.78	118.30
1	D	79	LYS	N-CA-C	8.08	132.82	111.00
1	A	78	MET	CB-CA-C	-6.67	97.06	110.40

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	79	LYS	CB-CA-C	-5.90	98.60	110.40

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	78	MET	Peptide
1	B	196	PRO	Peptide

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	1807	0	1745	19	0
1	B	1807	0	1745	32	0
1	C	1807	0	1745	19	0
1	D	1807	0	1745	39	0
2	A	196	0	0	5	0
2	B	213	0	0	3	0
2	C	196	0	0	2	0
2	D	206	0	0	3	0
All	All	8039	0	6980	105	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 7.

The worst 5 of 105 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:D:76:ASP:HA	1:D:79:LYS:HD2	1.48	0.95
1:D:5:GLU:OE2	1:D:85:LYS:HB2	1.68	0.93
1:C:76:ASP:HA	1:C:79:LYS:HG3	1.56	0.88
1:D:80:ARG:O	1:D:194:LEU:HD13	1.73	0.88
1:D:81:HIS:CE1	1:D:197:ASP:H	1.91	0.87

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	220/236 (93%)	215 (98%)	5 (2%)	0	100	100
1	B	220/236 (93%)	216 (98%)	4 (2%)	0	100	100
1	C	220/236 (93%)	216 (98%)	4 (2%)	0	100	100
1	D	220/236 (93%)	216 (98%)	4 (2%)	0	100	100
All	All	880/944 (93%)	863 (98%)	17 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	193/205 (94%)	188 (97%)	5 (3%)	41	21
1	B	193/205 (94%)	187 (97%)	6 (3%)	35	15
1	C	193/205 (94%)	190 (98%)	3 (2%)	58	42
1	D	193/205 (94%)	188 (97%)	5 (3%)	41	21
All	All	772/820 (94%)	753 (98%)	19 (2%)	42	22

5 of 19 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	52	LYS
1	D	214	LYS
1	D	225	THR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	D	176	VAL
1	B	101	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	81	HIS
1	D	81	HIS

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

4 non-standard protein/DNA/RNA residues 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	$\#|Z| > 2$
1	CRO	D	66	1	23,23,24	6.06	10 (43%)	30,32,34	3.18	10 (33%)
1	CRO	B	66	1	23,23,24	6.06	10 (43%)	30,32,34	3.19	10 (33%)
1	CRO	C	66	1	23,23,24	6.06	10 (43%)	30,32,34	3.16	9 (30%)
1	CRO	A	66	1	23,23,24	6.07	11 (47%)	30,32,34	3.26	10 (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
1	CRO	D	66	1	-	1/12/31/32	0/2/2/2
1	CRO	B	66	1	-	1/12/31/32	0/2/2/2

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
1	CRO	C	66	1	-	1/12/31/32	0/2/2/2
1	CRO	A	66	1	-	1/12/31/32	0/2/2/2

The worst 5 of 41 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	66	CRO	CB2-CA2	20.80	1.52	1.35
1	B	66	CRO	CB2-CA2	20.77	1.52	1.35
1	D	66	CRO	CB2-CA2	20.73	1.52	1.35
1	A	66	CRO	CB2-CA2	20.66	1.52	1.35
1	A	66	CRO	CA2-C2	-12.83	1.36	1.48

The worst 5 of 39 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	66	CRO	CA2-C2-N3	9.42	107.82	103.37
1	B	66	CRO	CA2-C2-N3	9.34	107.78	103.37
1	D	66	CRO	CA2-C2-N3	9.32	107.78	103.37
1	C	66	CRO	CA2-C2-N3	9.28	107.76	103.37
1	B	66	CRO	O2-C2-CA2	-7.65	126.67	130.96

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	A	66	CRO	C1-CA1-CB1-CG1
1	B	66	CRO	C1-CA1-CB1-CG1
1	C	66	CRO	C1-CA1-CB1-CG1
1	D	66	CRO	C1-CA1-CB1-CG1

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates (i)

There are no oligosaccharides 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 ⓘ

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(Å^2)	Q<0.9
1	A	224/236 (94%)	0.16	14 (6%)	27	30	11, 21, 44, 53	0
1	B	224/236 (94%)	0.02	16 (7%)	23	26	10, 18, 45, 64	0
1	C	224/236 (94%)	0.03	12 (5%)	32	36	10, 19, 41, 51	0
1	D	224/236 (94%)	0.02	18 (8%)	20	22	10, 18, 44, 62	0
All	All	896/944 (94%)	0.06	60 (6%)	25	28	10, 19, 45, 64	0

The worst 5 of 60 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	229	ILE	6.6
1	D	194	LEU	6.0
1	D	229	ILE	5.7
1	B	77	HIS	5.5
1	D	81	HIS	5.3

6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

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
1	CRO	B	66	22/23	0.89	0.13	17,22,24,25	0
1	CRO	C	66	22/23	0.89	0.11	17,22,24,25	0
1	CRO	D	66	22/23	0.89	0.12	17,22,24,25	0
1	CRO	A	66	22/23	0.93	0.10	17,22,24,25	0

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.