



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 8, 2018 – 08:29 pm GMT

PDB ID : 5DPJ
Title : sfGFP double mutant - 133/149 p-ethynyl-L-phenylalanine
Authors : Dippel, A.B.; Olenginski, G.M.; Maurici, N.; Liskov, M.T.; Brewer, S.H.;
Phillips-Piro, C.M.
Deposited on : 2015-09-12
Resolution : 2.50 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 ⓘ symbol.

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

MolProbity : 4.02b-467
Mogul : 1.7.3 (157068), CSD as539be (2018)
Xtriage (Phenix) : 1.13
EDS : trunk30967
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac : 5.8.0158
CCP4 : 7.0 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : trunk30967

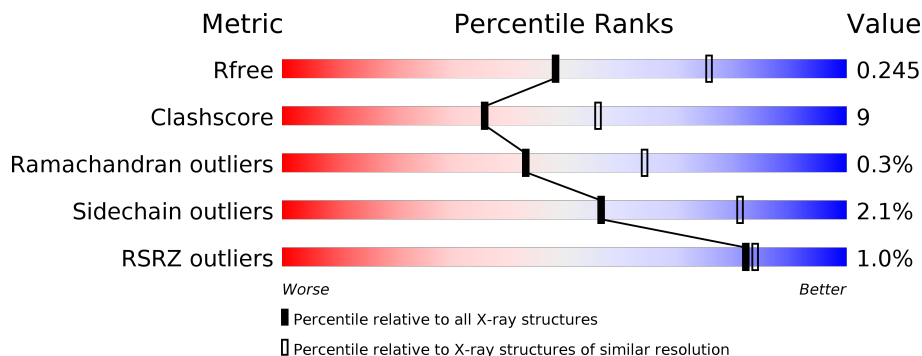
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.50 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}	111664	4155 (2.50-2.50)
Clashscore	122126	4827 (2.50-2.50)
Ramachandran outliers	120053	4735 (2.50-2.50)
Sidechain outliers	120020	4737 (2.50-2.50)
RSRZ outliers	108989	4058 (2.50-2.50)

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 ≥ 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	237	<div> <div>2%</div> <div>82% 14% .</div> </div>
1	B	237	<div> <div>80% 13% . 5%</div> </div>
1	C	237	<div> <div>73% 22% 5%</div> </div>
1	D	237	<div> <div>2%</div> <div>70% 22% . 5%</div> </div>

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 7396 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	227	Total	C	N	O	S	0	3	0
			1835	1169	313	348	5			
1	B	225	Total	C	N	O	S	0	0	0
			1791	1142	305	339	5			
1	C	225	Total	C	N	O	S	0	0	0
			1795	1144	305	341	5			
1	D	225	Total	C	N	O	S	0	0	0
			1787	1139	304	339	5			

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	MET	-	initiating methionine	UNP A0A059PIQ0
A	1	VAL	MET	engineered mutation	UNP A0A059PIQ0
A	2	SER	ARG	engineered mutation	UNP A0A059PIQ0
A	30	ARG	SER	engineered mutation	UNP A0A059PIQ0
A	66	CRO	THR	chromophore	UNP A0A059PIQ0
A	66	CRO	TYR	chromophore	UNP A0A059PIQ0
A	66	CRO	GLY	chromophore	UNP A0A059PIQ0
A	72	SER	ALA	engineered mutation	UNP A0A059PIQ0
A	80	ARG	GLN	engineered mutation	UNP A0A059PIQ0
A	133	5DW	ASP	engineered mutation	UNP A0A059PIQ0
A	149	5DW	ASN	engineered mutation	UNP A0A059PIQ0
A	206	VAL	ALA	engineered mutation	UNP A0A059PIQ0
B	0	MET	-	initiating methionine	UNP A0A059PIQ0
B	1	VAL	MET	engineered mutation	UNP A0A059PIQ0
B	2	SER	ARG	engineered mutation	UNP A0A059PIQ0
B	30	ARG	SER	engineered mutation	UNP A0A059PIQ0
B	66	CRO	THR	chromophore	UNP A0A059PIQ0
B	66	CRO	TYR	chromophore	UNP A0A059PIQ0
B	66	CRO	GLY	chromophore	UNP A0A059PIQ0
B	72	SER	ALA	engineered mutation	UNP A0A059PIQ0
B	80	ARG	GLN	engineered mutation	UNP A0A059PIQ0

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	133	5DW	ASP	engineered mutation	UNP A0A059PIQ0
B	149	5DW	ASN	engineered mutation	UNP A0A059PIQ0
B	206	VAL	ALA	engineered mutation	UNP A0A059PIQ0
C	0	MET	-	initiating methionine	UNP A0A059PIQ0
C	1	VAL	MET	engineered mutation	UNP A0A059PIQ0
C	2	SER	ARG	engineered mutation	UNP A0A059PIQ0
C	30	ARG	SER	engineered mutation	UNP A0A059PIQ0
C	66	CRO	THR	chromophore	UNP A0A059PIQ0
C	66	CRO	TYR	chromophore	UNP A0A059PIQ0
C	66	CRO	GLY	chromophore	UNP A0A059PIQ0
C	72	SER	ALA	engineered mutation	UNP A0A059PIQ0
C	80	ARG	GLN	engineered mutation	UNP A0A059PIQ0
C	133	5DW	ASP	engineered mutation	UNP A0A059PIQ0
C	149	5DW	ASN	engineered mutation	UNP A0A059PIQ0
C	206	VAL	ALA	engineered mutation	UNP A0A059PIQ0
D	0	MET	-	initiating methionine	UNP A0A059PIQ0
D	1	VAL	MET	engineered mutation	UNP A0A059PIQ0
D	2	SER	ARG	engineered mutation	UNP A0A059PIQ0
D	30	ARG	SER	engineered mutation	UNP A0A059PIQ0
D	66	CRO	THR	chromophore	UNP A0A059PIQ0
D	66	CRO	TYR	chromophore	UNP A0A059PIQ0
D	66	CRO	GLY	chromophore	UNP A0A059PIQ0
D	72	SER	ALA	engineered mutation	UNP A0A059PIQ0
D	80	ARG	GLN	engineered mutation	UNP A0A059PIQ0
D	133	5DW	ASP	engineered mutation	UNP A0A059PIQ0
D	149	5DW	ASN	engineered mutation	UNP A0A059PIQ0
D	206	VAL	ALA	engineered mutation	UNP A0A059PIQ0

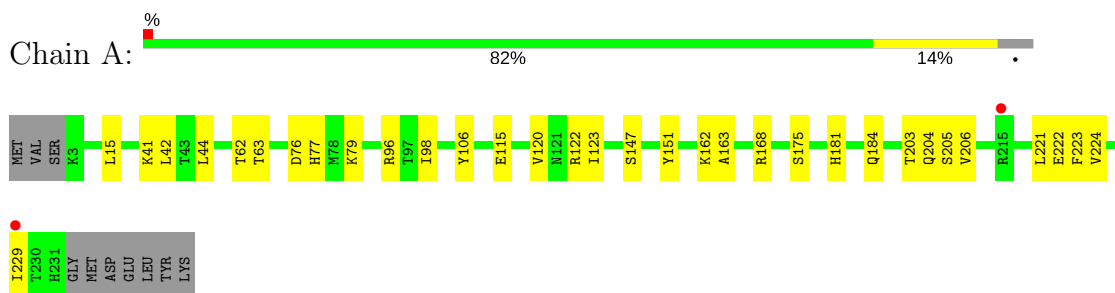
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	59	Total O 59 59	0	0
2	B	49	Total O 49 49	0	0
2	C	55	Total O 55 55	0	0
2	D	25	Total O 25 25	0	0

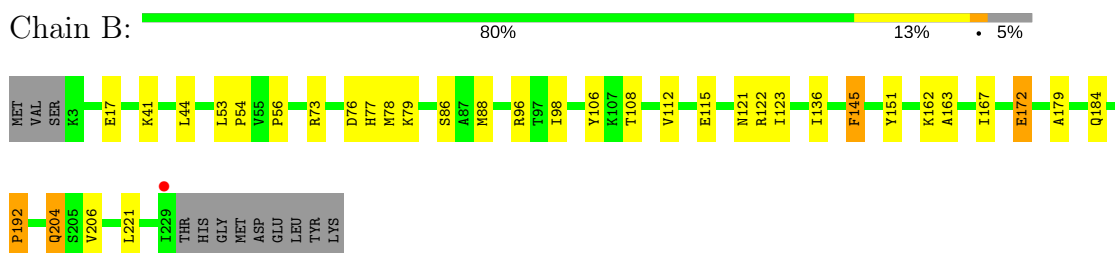
3 Residue-property plots [i](#)

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.

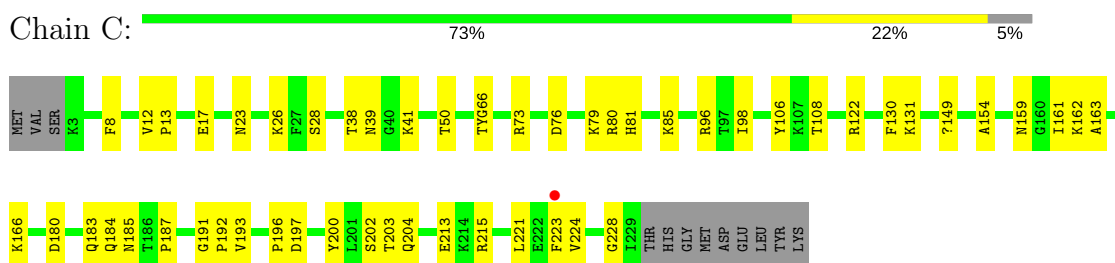
- Molecule 1: Green fluorescent protein



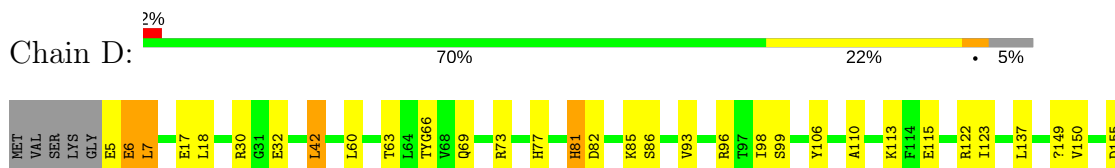
- Molecule 1: Green fluorescent protein

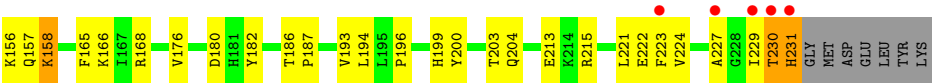


- Molecule 1: Green fluorescent protein



- Molecule 1: Green fluorescent protein





4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	46.29Å 112.94Å 93.41Å 90.00° 104.09° 90.00°	Depositor
Resolution (Å)	47.92 – 2.50 47.92 – 2.50	Depositor EDS
% Data completeness (in resolution range)	99.3 (47.92-2.50) 99.4 (47.92-2.50)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.88 (at 2.48Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.180 , 0.244 0.182 , 0.245	Depositor DCC
R_{free} test set	1640 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å ²)	39.6	Xtriage
Anisotropy	0.761	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 44.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.031 for h,-k,-h-l	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	7396	wwPDB-VP
Average B, all atoms (Å ²)	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.85% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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

Bond lengths and bond angles in the following residue types are not validated in this section: 5DW, 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	0.47	0/1833	0.62	0/2475
1	B	0.46	1/1779 (0.1%)	0.60	0/2404
1	C	0.42	0/1783	0.59	0/2409
1	D	0.46	0/1776	0.61	1/2404 (0.0%)
All	All	0.45	1/7171 (0.0%)	0.61	1/9692 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	192	PRO	N-CD	5.05	1.54	1.47

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	42	LEU	CA-CB-CG	5.09	127.00	115.30

There are no chirality outliers.

There are no planarity outliers.

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	1835	0	1776	25	0
1	B	1791	0	1724	20	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	1795	0	1728	36	0
1	D	1787	0	1705	56	0
2	A	59	0	0	3	0
2	B	49	0	0	1	0
2	C	55	0	0	3	0
2	D	25	0	0	0	0
All	All	7396	0	6933	123	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 9.

The worst 5 of 123 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:6:GLU:HB3	1:D:7:LEU:HD12	1.09	1.05
1:D:6:GLU:HB3	1:D:7:LEU:CD1	1.91	1.00
1:D:6:GLU:CB	1:D:7:LEU:HD12	2.01	0.89
1:D:66:CRO:HB2	1:D:69:GLN:HE21	1.52	0.74
1:D:7:LEU:HD12	1:D:7:LEU:N	2.03	0.72

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	223/237 (94%)	222 (100%)	1 (0%)	0	100	100
1	B	218/237 (92%)	210 (96%)	7 (3%)	1 (0%)	31	51
1	C	218/237 (92%)	205 (94%)	13 (6%)	0	100	100
1	D	218/237 (92%)	199 (91%)	17 (8%)	2 (1%)	19	34
All	All	877/948 (92%)	836 (95%)	38 (4%)	3 (0%)	43	64

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	6	GLU
1	B	172	GLU
1	D	157	GLN

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	197/205 (96%)	196 (100%)	1 (0%)	90	97
1	B	190/205 (93%)	185 (97%)	5 (3%)	49	76
1	C	191/205 (93%)	189 (99%)	2 (1%)	78	92
1	D	189/205 (92%)	181 (96%)	8 (4%)	32	57
All	All	767/820 (94%)	751 (98%)	16 (2%)	56	81

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	184	GLN
1	D	7	LEU
1	D	158	LYS
1	C	80	ARG
1	D	186	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

12 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	5DW	A	133	1	13,13,14	1.27	1 (7%)	15,16,18	0.61	0
1	5DW	A	149	1	13,13,14	1.36	2 (15%)	15,16,18	1.30	1 (6%)
1	CRO	A	66	1	23,23,24	2.69	5 (21%)	27,32,34	3.26	12 (44%)
1	5DW	B	133	1	13,13,14	1.11	1 (7%)	15,16,18	0.86	0
1	5DW	B	149	1	13,13,14	1.11	1 (7%)	15,16,18	1.11	1 (6%)
1	CRO	B	66	1	23,23,24	2.94	6 (26%)	27,32,34	3.41	12 (44%)
1	5DW	C	133	1	13,13,14	1.14	1 (7%)	15,16,18	1.00	0
1	5DW	C	149	1	13,13,14	1.26	2 (15%)	15,16,18	1.02	1 (6%)
1	CRO	C	66	1	23,23,24	2.85	6 (26%)	27,32,34	3.55	11 (40%)
1	5DW	D	133	1	13,13,14	1.35	3 (23%)	15,16,18	1.03	2 (13%)
1	5DW	D	149	1	13,13,14	1.15	1 (7%)	15,16,18	1.06	0
1	CRO	D	66	1	23,23,24	2.70	7 (30%)	27,32,34	3.29	11 (40%)

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	5DW	A	133	1	-	0/4/8/10	0/1/1/1
1	5DW	A	149	1	-	0/4/8/10	0/1/1/1
1	CRO	A	66	1	-	0/12/31/32	0/2/2/2
1	5DW	B	133	1	-	0/4/8/10	0/1/1/1
1	5DW	B	149	1	-	0/4/8/10	0/1/1/1
1	CRO	B	66	1	-	0/12/31/32	0/2/2/2
1	5DW	C	133	1	-	0/4/8/10	0/1/1/1
1	5DW	C	149	1	-	0/4/8/10	0/1/1/1
1	CRO	C	66	1	-	0/12/31/32	0/2/2/2
1	5DW	D	133	1	-	0/4/8/10	0/1/1/1
1	5DW	D	149	1	-	0/4/8/10	0/1/1/1
1	CRO	D	66	1	-	0/12/31/32	0/2/2/2

The worst 5 of 36 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	66	CRO	CA1-N1	-2.21	1.40	1.47
1	D	133	5DW	CA-N	-2.16	1.40	1.47
1	D	66	CRO	CA1-C1	2.01	1.54	1.51
1	C	149	5DW	CA-C	2.04	1.52	1.50
1	A	149	5DW	CA-C	2.08	1.53	1.50

The worst 5 of 51 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	66	CRO	O2-C2-CA2	-10.71	125.28	130.97
1	B	66	CRO	O2-C2-CA2	-10.53	125.38	130.97
1	D	66	CRO	O2-C2-CA2	-8.76	126.32	130.97
1	A	66	CRO	O2-C2-CA2	-8.66	126.37	130.97
1	C	66	CRO	C2-CA2-N2	-6.02	104.68	108.93

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

4 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	C	149	5DW	1	0
1	C	66	CRO	1	0
1	D	149	5DW	2	0
1	D	66	CRO	3	0

5.5 Carbohydrates [i](#)

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 ⓘ

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	224/237 (94%)	-0.16	2 (0%) 84 85	29, 38, 52, 73	0
1	B	222/237 (93%)	-0.09	1 (0%) 90 91	29, 43, 61, 87	0
1	C	222/237 (93%)	-0.05	1 (0%) 90 91	30, 44, 64, 76	0
1	D	222/237 (93%)	0.12	5 (2%) 60 63	35, 50, 74, 103	0
All	All	890/948 (93%)	-0.05	9 (1%) 82 84	29, 44, 66, 103	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	230	THR	6.9
1	D	231	HIS	3.5
1	C	223	PHE	3.4
1	D	229	ILE	3.0
1	D	227	ALA	2.8

6.2 Non-standard residues in protein, DNA, RNA chains [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
1	5DW	D	133	13/14	0.91	0.27	50,57,64,67	0
1	5DW	A	149	13/14	0.95	0.14	31,37,40,44	0
1	CRO	B	66	22/23	0.96	0.15	28,35,37,43	0
1	5DW	D	149	13/14	0.96	0.15	44,47,51,53	0
1	CRO	D	66	22/23	0.96	0.17	35,39,41,44	0
1	5DW	A	133	13/14	0.96	0.19	37,40,43,44	0
1	5DW	C	133	13/14	0.96	0.20	39,42,46,51	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
1	CRO	C	66	22/23	0.96	0.14	28,35,39,44	0
1	CRO	A	66	22/23	0.96	0.16	29,32,36,41	0
1	5DW	C	149	13/14	0.97	0.13	34,41,48,51	0
1	5DW	B	149	13/14	0.97	0.18	34,39,41,42	0
1	5DW	B	133	13/14	0.98	0.18	33,35,37,44	0

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.