



wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 7, 2023 – 02:38 AM EST

PDB ID : 4W6R
Title : Crystal Structure of Full-Length Split GFP Mutant D102C Disulfide Dimer,
P 1 Space Group
Authors : Leibly, D.J.; Waldo, G.S.; Yeates, T.O.
Deposited on : 2014-08-20
Resolution : 3.47 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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

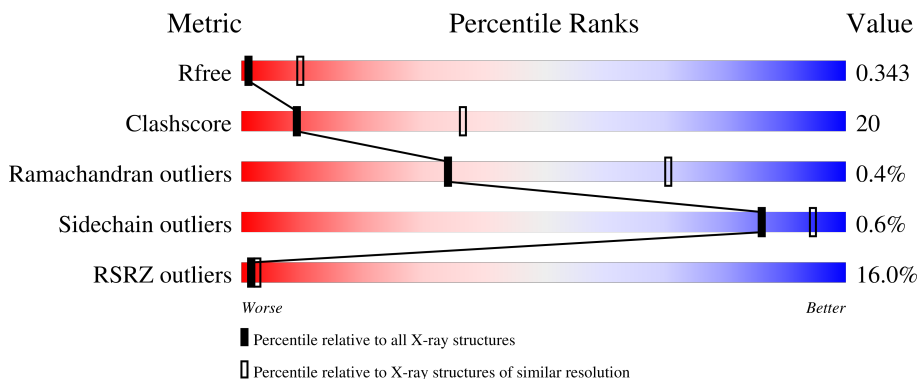
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.47 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	1379 (3.56-3.40)
Clashscore	141614	1461 (3.56-3.40)
Ramachandran outliers	138981	1424 (3.56-3.40)
Sidechain outliers	138945	1425 (3.56-3.40)
RSRZ outliers	127900	1289 (3.56-3.40)

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 $\leq 5\%$. 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	234	 13% 59% 31% 8%
1	B	234	 10% 58% 29% 12%
1	C	234	 13% 59% 29% 12%
1	D	234	 8% 53% 28% 17%
1	E	234	 9% 66% 23% 11%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	234	
1	G	234	
1	H	234	
1	I	234	
1	J	234	
1	K	234	
1	L	234	
1	M	234	
1	N	234	
1	O	234	
1	P	234	

2 Entry composition [i](#)

There is only 1 type of molecule in this entry. The entry contains 25002 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.

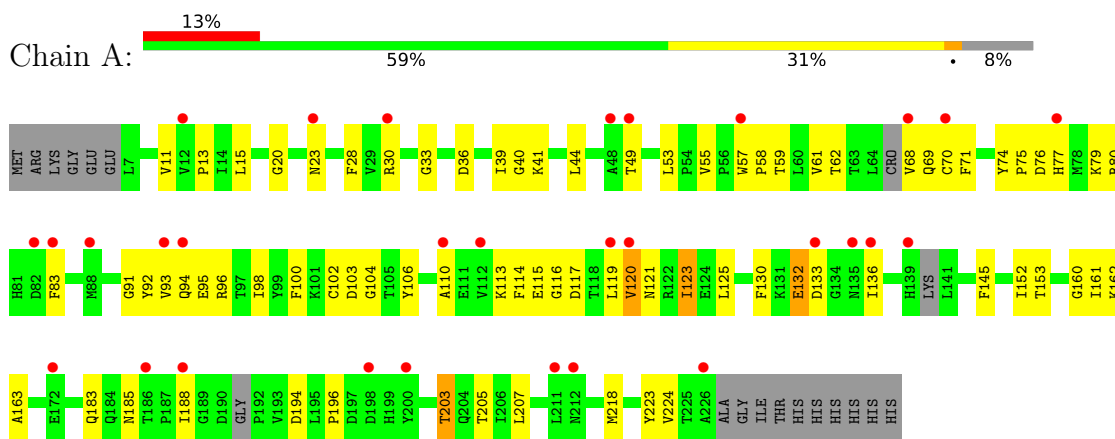
- Molecule 1 is a protein called fluorescent protein D102C.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	215	Total 1721	C 1098	N 291	O 327	S 5	0	0	0
1	B	205	Total 1649	C 1052	N 280	O 313	S 4	0	0	0
1	C	207	Total 1643	C 1052	N 277	O 309	S 5	0	0	0
1	D	194	Total 1552	C 993	N 261	O 293	S 5	0	0	0
1	E	208	Total 1661	C 1059	N 280	O 317	S 5	0	0	0
1	F	201	Total 1600	C 1020	N 270	O 307	S 3	0	0	0
1	G	188	Total 1497	C 949	N 262	O 282	S 4	0	0	0
1	H	191	Total 1537	C 980	N 264	O 289	S 4	0	0	0
1	I	209	Total 1671	C 1065	N 285	O 316	S 5	0	0	0
1	J	184	Total 1459	C 931	N 247	O 278	S 3	0	0	0
1	K	175	Total 1391	C 896	N 229	O 263	S 3	0	0	0
1	L	197	Total 1573	C 1002	N 269	O 298	S 4	0	0	0
1	M	168	Total 1352	C 864	N 233	O 251	S 4	0	0	0
1	N	174	Total 1387	C 876	N 235	O 273	S 3	0	0	0
1	O	213	Total 1677	C 1065	N 286	O 322	S 4	0	0	0
1	P	204	Total 1632	C 1038	N 281	O 308	S 5	0	0	0

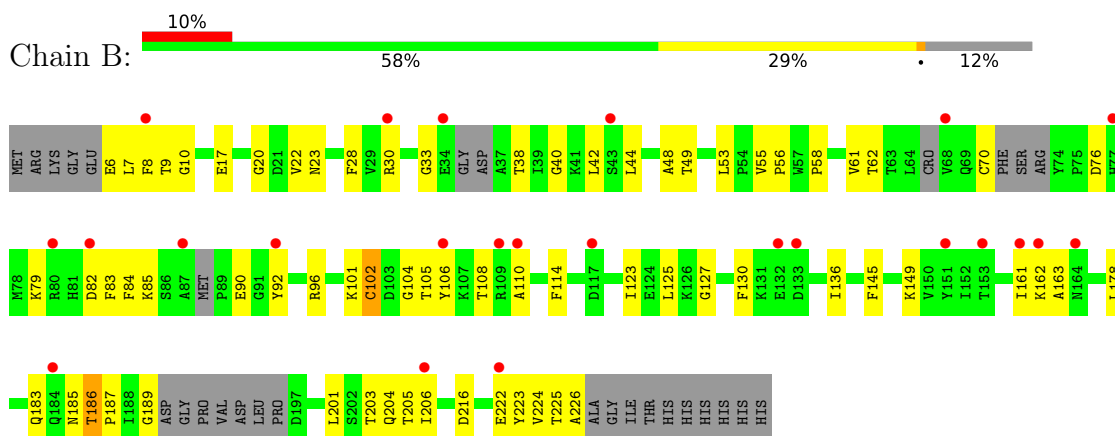
3 Residue-property plots [i](#)

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.

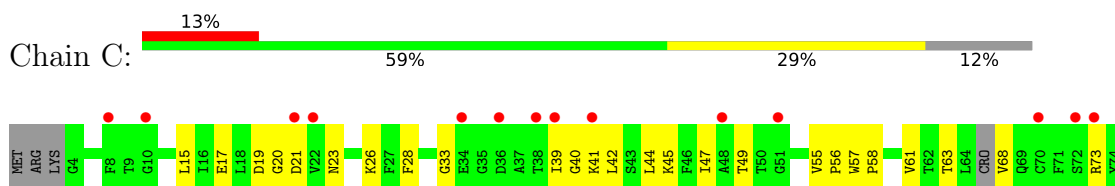
- Molecule 1: fluorescent protein D102C

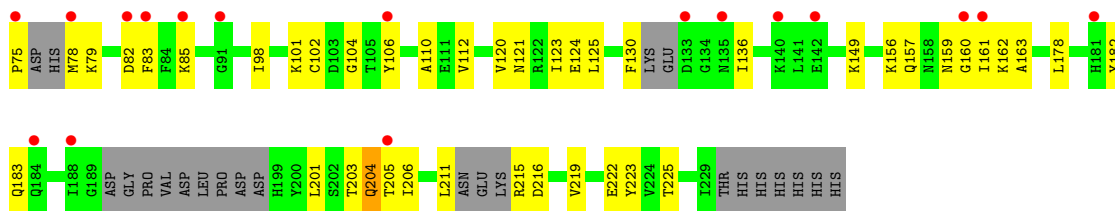


- Molecule 1: fluorescent protein D102C

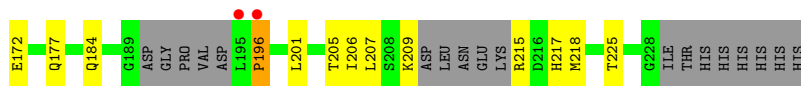
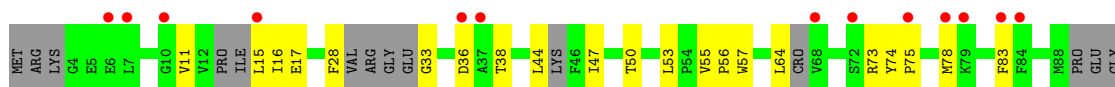


- Molecule 1: fluorescent protein D102C

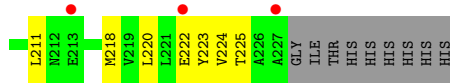




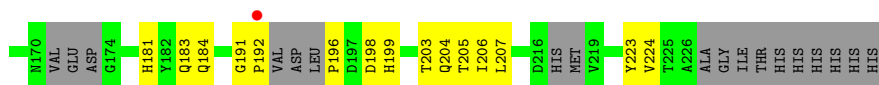
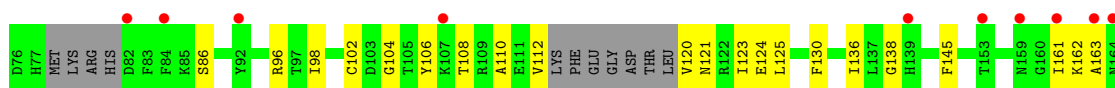
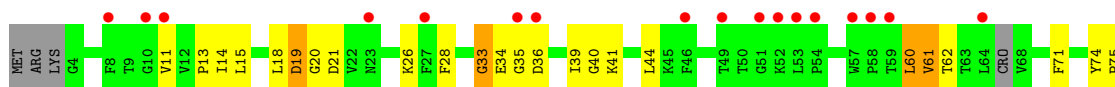
- Molecule 1: fluorescent protein D102C



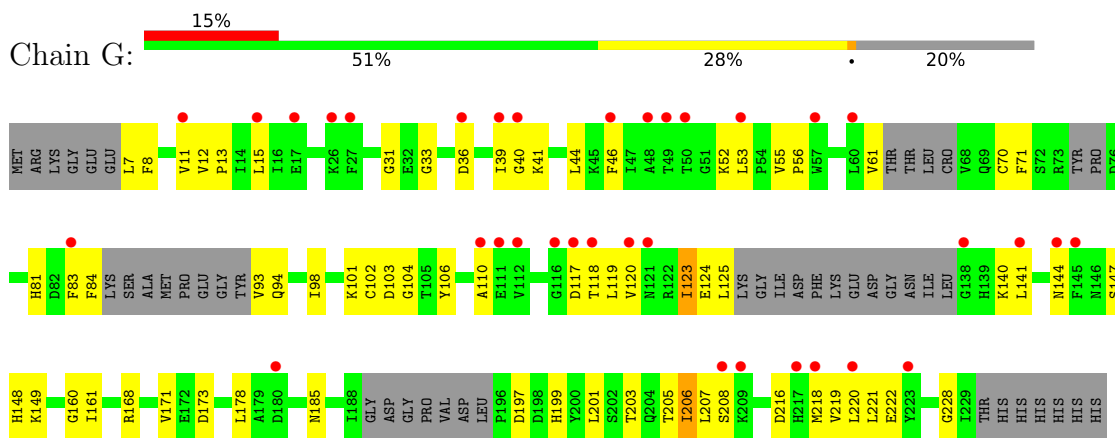
- Molecule 1: fluorescent protein D102C



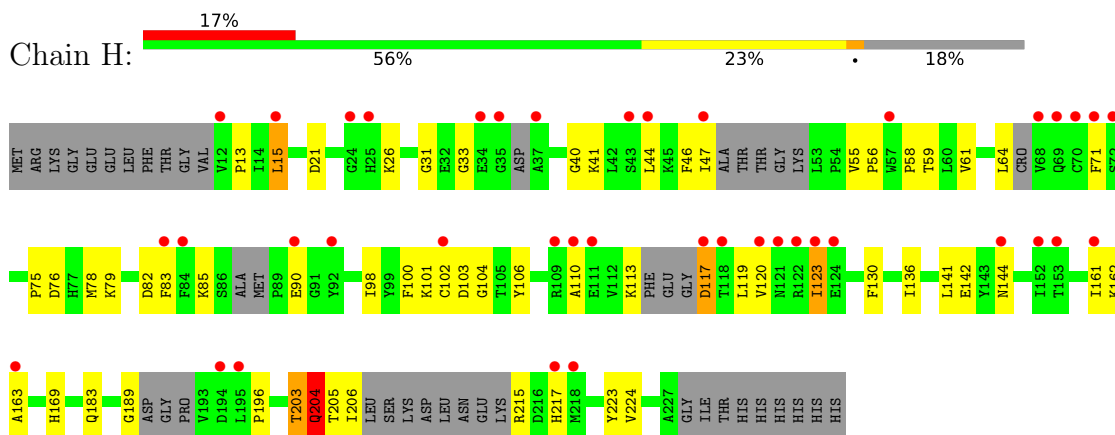
- Molecule 1: fluorescent protein D102C



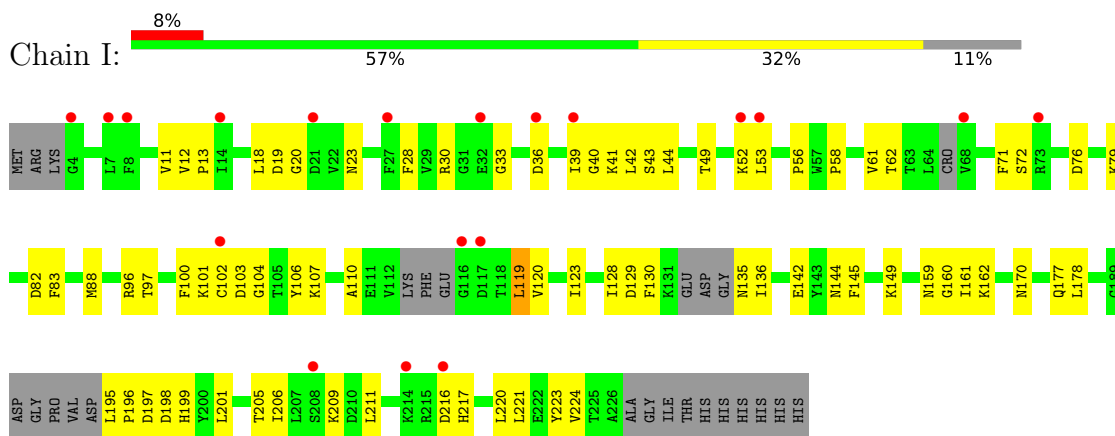
- Molecule 1: fluorescent protein D102C



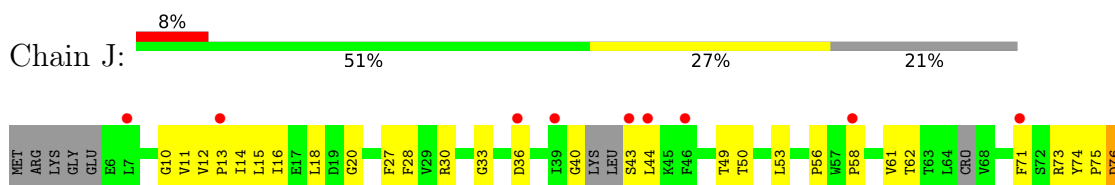
- Molecule 1: fluorescent protein D102C

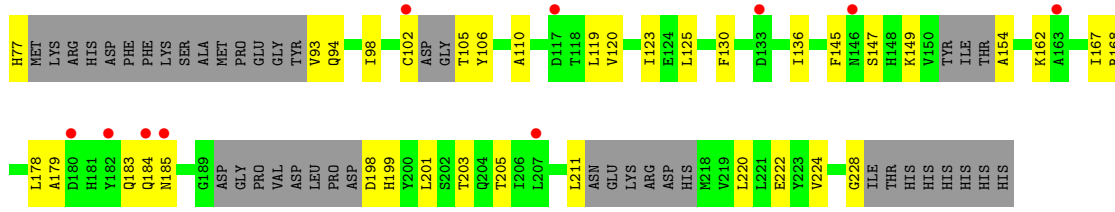


- Molecule 1: fluorescent protein D102C

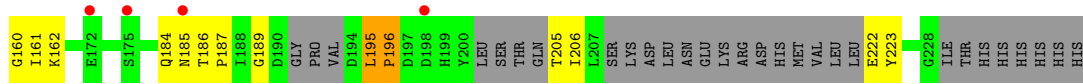
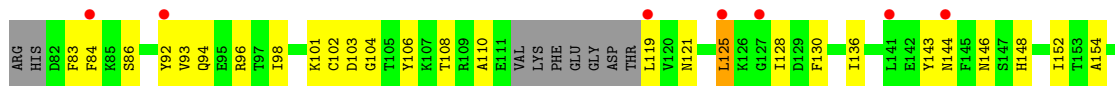
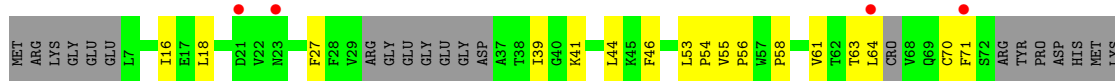


- Molecule 1: fluorescent protein D102C

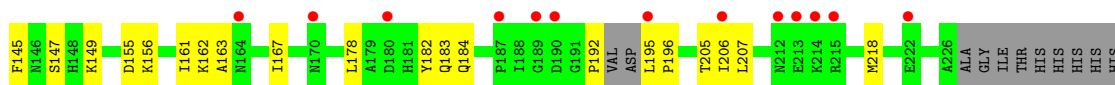
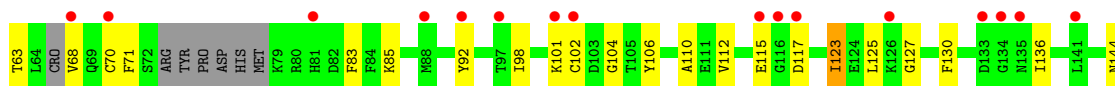
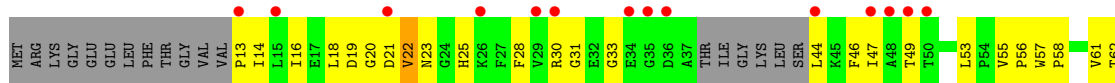




• Molecule 1: fluorescent protein D102C

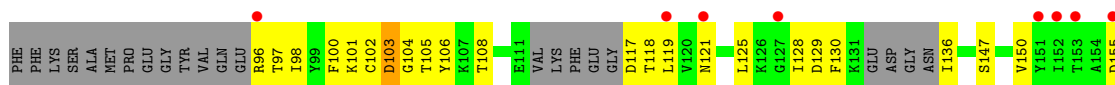
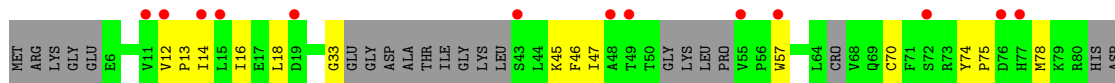


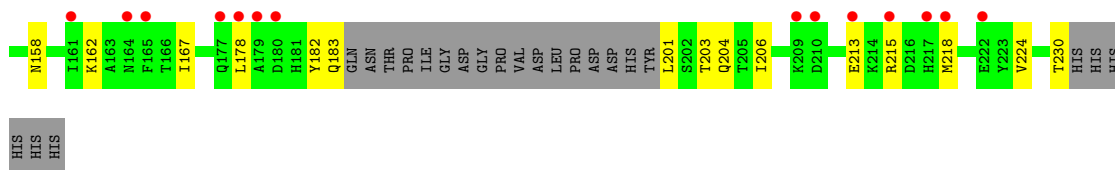
• Molecule 1: fluorescent protein D102C



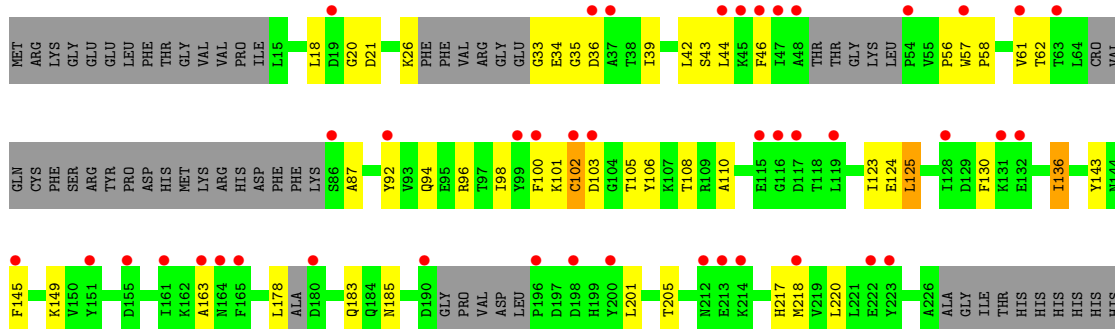
HIS

• Molecule 1: fluorescent protein D102C

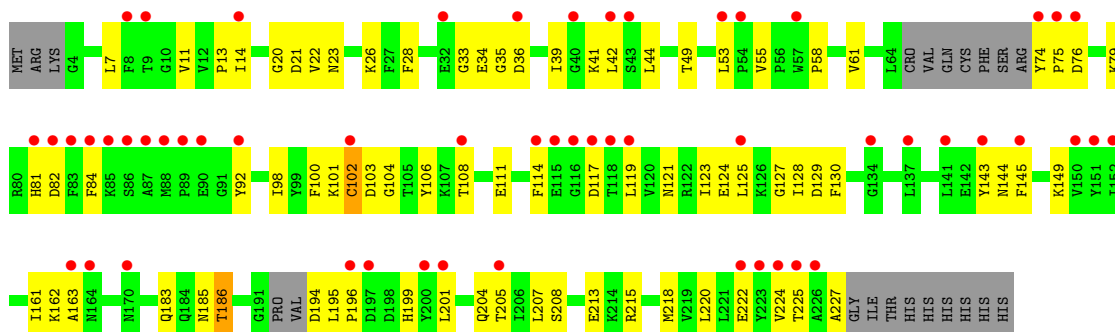




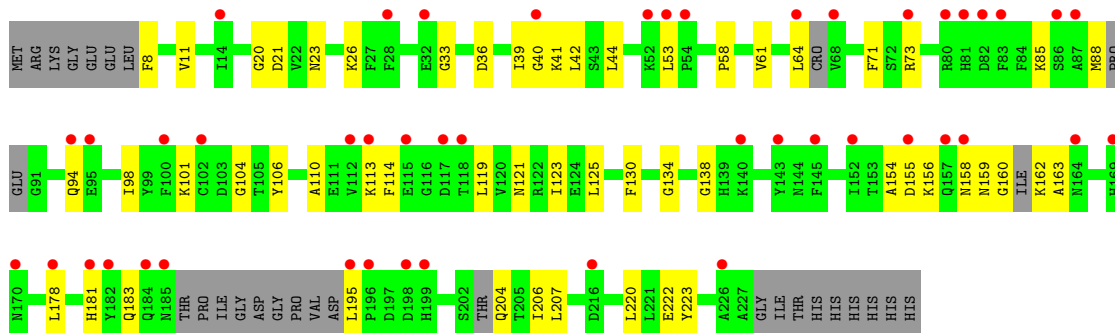
• Molecule 1: fluorescent protein D102C



• Molecule 1: fluorescent protein D102C



• Molecule 1: fluorescent protein D102C



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	92.42Å 92.56Å 124.53Å 94.94° 96.17° 102.25°	Depositor
Resolution (Å)	89.88 – 3.47 89.88 – 3.47	Depositor EDS
% Data completeness (in resolution range)	89.4 (89.88-3.47) 86.7 (89.88-3.47)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.09	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.70 (at 3.49Å)	Xtrriage
Refinement program	PHENIX (PHENIX.REFINE: DEV_1555)	Depositor
R, R_{free}	0.307 , 0.357 0.326 , 0.343	Depositor DCC
R_{free} test set	4609 reflections (10.00%)	wwPDB-VP
Wilson B-factor (Å ²)	101.2	Xtrriage
Anisotropy	0.575	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.28 , 40.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.28$	Xtrriage
Estimated twinning fraction	0.025 for -k,-h,-l	Xtrriage
F_o, F_c correlation	0.85	EDS
Total number of atoms	25002	wwPDB-VP
Average B, all atoms (Å ²)	113.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.95% 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 [i](#)

5.1 Standard geometry [i](#)

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	0.38	0/1758	0.82	3/2375 (0.1%)
1	B	0.34	0/1681	0.77	1/2265 (0.0%)
1	C	0.33	0/1675	0.80	1/2258 (0.0%)
1	D	0.38	0/1579	0.89	3/2121 (0.1%)
1	E	0.33	0/1694	0.78	4/2287 (0.2%)
1	F	0.36	0/1631	0.81	2/2199 (0.1%)
1	G	0.35	0/1524	0.82	0/2053
1	H	0.34	0/1567	0.82	4/2112 (0.2%)
1	I	0.35	0/1705	0.84	4/2302 (0.2%)
1	J	0.35	0/1483	0.82	0/1998
1	K	0.41	0/1417	0.81	2/1912 (0.1%)
1	L	0.33	0/1605	0.76	1/2164 (0.0%)
1	M	0.31	0/1373	0.84	0/1847
1	N	0.33	0/1410	0.86	3/1899 (0.2%)
1	O	0.40	0/1712	0.81	1/2314 (0.0%)
1	P	0.33	0/1663	0.79	1/2239 (0.0%)
All	All	0.35	0/25477	0.81	30/34345 (0.1%)

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	D	0	1

There are no bond length outliers.

The worst 5 of 30 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	H	15	LEU	CA-CB-CG	7.89	133.45	115.30
1	D	125	LEU	CA-CB-CG	7.63	132.85	115.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	N	220	LEU	CA-CB-CG	7.57	132.72	115.30
1	A	203	THR	N-CA-C	7.12	130.21	111.00
1	H	203	THR	N-CA-C	6.86	129.52	111.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	111	GLU	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	1721	0	1680	100	0
1	B	1649	0	1615	84	0
1	C	1643	0	1611	59	1
1	D	1552	0	1511	61	1
1	E	1661	0	1623	38	0
1	F	1600	0	1557	47	1
1	G	1497	0	1460	86	0
1	H	1537	0	1487	56	0
1	I	1671	0	1633	95	0
1	J	1459	0	1431	58	1
1	K	1391	0	1359	51	0
1	L	1573	0	1532	60	0
1	M	1352	0	1339	46	0
1	N	1387	0	1351	40	0
1	O	1677	0	1616	102	0
1	P	1632	0	1586	42	0
All	All	25002	0	24391	986	2

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 20.

The worst 5 of 986 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:I:28:PHE:CE2	1:I:30:ARG:HG3	1.44	1.51
1:A:152:ILE:O	1:A:153:THR:HG22	1.19	1.34
1:D:15:LEU:O	1:D:15:LEU:HD12	1.31	1.30
1:B:106:TYR:CE1	1:B:130:PHE:CE1	2.24	1.24
1:G:44:LEU:O	1:G:219:VAL:HG13	1.36	1.24

All (2) 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:D:157:GLN:OE1	1:F:184:GLN:NE2[1_445]	2.14	0.06
1:C:17:GLU:OE2	1:J:50:THR:OG1[1_455]	2.17	0.03

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	207/234 (88%)	195 (94%)	11 (5%)	1 (0%)	29	66
1	B	193/234 (82%)	184 (95%)	8 (4%)	1 (0%)	29	66
1	C	195/234 (83%)	188 (96%)	6 (3%)	1 (0%)	29	66
1	D	174/234 (74%)	161 (92%)	12 (7%)	1 (1%)	25	63
1	E	198/234 (85%)	191 (96%)	6 (3%)	1 (0%)	29	66
1	F	187/234 (80%)	180 (96%)	6 (3%)	1 (0%)	29	66
1	G	176/234 (75%)	167 (95%)	8 (4%)	1 (1%)	25	63
1	H	175/234 (75%)	169 (97%)	6 (3%)	0	100	100
1	I	199/234 (85%)	192 (96%)	6 (3%)	1 (0%)	29	66
1	J	168/234 (72%)	163 (97%)	5 (3%)	0	100	100
1	K	159/234 (68%)	154 (97%)	4 (2%)	1 (1%)	25	63
1	L	187/234 (80%)	181 (97%)	5 (3%)	1 (0%)	29	66

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	M	152/234 (65%)	149 (98%)	2 (1%)	1 (1%)	22	60
1	N	162/234 (69%)	159 (98%)	3 (2%)	0	100	100
1	O	207/234 (88%)	198 (96%)	7 (3%)	2 (1%)	15	52
1	P	192/234 (82%)	186 (97%)	6 (3%)	0	100	100
All	All	2931/3744 (78%)	2817 (96%)	101 (3%)	13 (0%)	34	70

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	120	VAL
1	F	61	VAL
1	G	206	ILE
1	D	196	PRO
1	C	204	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	189/204 (93%)	188 (100%)	1 (0%)	88	95
1	B	181/204 (89%)	180 (99%)	1 (1%)	86	94
1	C	178/204 (87%)	177 (99%)	1 (1%)	86	94
1	D	170/204 (83%)	170 (100%)	0	100	100
1	E	182/204 (89%)	181 (100%)	1 (0%)	88	95
1	F	175/204 (86%)	174 (99%)	1 (1%)	86	94
1	G	163/204 (80%)	162 (99%)	1 (1%)	86	94
1	H	168/204 (82%)	166 (99%)	2 (1%)	71	87
1	I	183/204 (90%)	181 (99%)	2 (1%)	73	88
1	J	160/204 (78%)	159 (99%)	1 (1%)	86	94
1	K	152/204 (74%)	151 (99%)	1 (1%)	84	93
1	L	171/204 (84%)	170 (99%)	1 (1%)	86	94

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	M	149/204 (73%)	148 (99%)	1 (1%)	84	93
1	N	153/204 (75%)	152 (99%)	1 (1%)	84	93
1	O	179/204 (88%)	178 (99%)	1 (1%)	86	94
1	P	177/204 (87%)	177 (100%)	0	100	100
All	All	2730/3264 (84%)	2714 (99%)	16 (1%)	86	94

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

Mol	Chain	Res	Type
1	N	102	CYS
1	M	121	ASN
1	I	119	LEU
1	L	123	ILE
1	H	204	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 18 such sidechains are listed below:

Mol	Chain	Res	Type
1	N	164	ASN
1	P	181	HIS
1	P	94	GLN
1	J	183	GLN
1	L	184	GLN

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](#)

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, 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	215/234 (91%)	0.97	30 (13%) 2 4	50, 80, 148, 324	0
1	B	205/234 (87%)	0.74	24 (11%) 4 6	51, 87, 175, 274	0
1	C	207/234 (88%)	0.89	31 (14%) 2 3	55, 91, 164, 338	0
1	D	194/234 (82%)	0.58	18 (9%) 8 10	57, 92, 169, 237	0
1	E	208/234 (88%)	0.67	20 (9%) 8 10	64, 96, 157, 204	0
1	F	201/234 (85%)	0.89	28 (13%) 2 4	69, 106, 174, 357	0
1	G	188/234 (80%)	0.93	35 (18%) 1 2	52, 103, 181, 284	0
1	H	191/234 (81%)	1.08	40 (20%) 1 1	69, 102, 172, 382	0
1	I	209/234 (89%)	0.64	19 (9%) 9 11	64, 98, 165, 249	0
1	J	184/234 (78%)	0.60	19 (10%) 6 8	51, 95, 166, 270	0
1	K	175/234 (74%)	0.64	15 (8%) 10 12	73, 102, 181, 323	0
1	L	197/234 (84%)	1.14	43 (21%) 0 1	67, 114, 206, 293	0
1	M	168/234 (71%)	1.25	35 (20%) 1 1	75, 119, 203, 375	0
1	N	174/234 (74%)	1.22	43 (24%) 0 0	72, 130, 219, 388	0
1	O	213/234 (91%)	1.47	55 (25%) 0 0	54, 126, 205, 302	0
1	P	204/234 (87%)	1.20	46 (22%) 0 0	73, 134, 217, 464	0
All	All	3133/3744 (83%)	0.93	501 (15%) 1 2	50, 104, 188, 464	0

The worst 5 of 501 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	N	164	ASN	11.3
1	M	164	ASN	9.7
1	O	86	SER	9.2
1	M	76	ASP	8.9
1	O	223	TYR	8.6

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](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.