wwPDB X-ray Structure Validation Summary Report (i)

Oct 12, 2021 – 02:09 PM EDT

PDB ID	:	2AWI
Title	:	Structure of PrgX Y153C mutant
Authors	:	Shi, K.; Brown, C.K.; Gu, Z.Y.; Kozlowicz, B.k.; Dunny, G.M.; Ohlendorf, D.H.; Earhart, C.A.
Deposited on	:	2005-09-01
Resolution	:	2.25 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 following versions of software and data (see references (1)) were used in the production of this report:

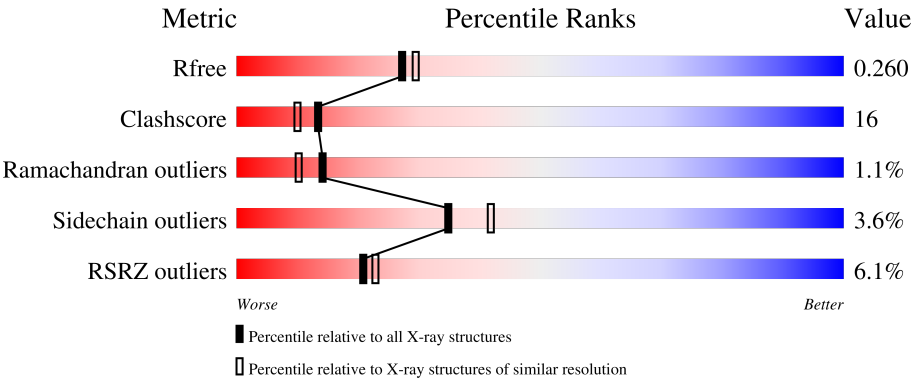
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 2.25 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	1377 (2.26-2.26)
Clashscore	141614	1487 (2.26-2.26)
Ramachandran outliers	138981	1449 (2.26-2.26)
Sidechain outliers	138945	1450 (2.26-2.26)
RSRZ outliers	127900	1356 (2.26-2.26)

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	317	3% 68% 24% · 6%
1	B	317	4% 63% 29% · 6%
1	C	317	5% 61% 30% · 6%
1	D	317	5% 62% 31% · 6%
1	E	317	4% 67% 24% · 6%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	317	8% 62% 30% • 6%
1	G	317	6% 65% 26% • 7%
1	H	317	6% 65% 26% • 7%
1	I	317	11% 57% 33% • 6%
1	J	317	7% 58% 33% • 7%
1	K	317	3% 66% 25% • 7%
1	L	317	5% 63% 28% • 7%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 30035 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 PrgX.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	298	Total	C	N	O	S	Se	0	0	0
			2461	1596	397	462	2	4			
1	B	298	Total	C	N	O	S	Se	0	0	0
			2461	1596	397	462	2	4			
1	C	297	Total	C	N	O	S	Se	0	0	0
			2450	1587	396	461	2	4			
1	D	297	Total	C	N	O	S	Se	0	0	0
			2450	1587	396	461	2	4			
1	E	297	Total	C	N	O	S	Se	0	0	0
			2450	1587	396	461	2	4			
1	F	297	Total	C	N	O	S	Se	0	0	0
			2449	1587	396	460	2	4			
1	G	296	Total	C	N	O	S	Se	0	0	0
			2441	1581	394	460	2	4			
1	H	296	Total	C	N	O	S	Se	0	0	0
			2441	1581	394	460	2	4			
1	I	298	Total	C	N	O	S	Se	0	0	0
			2461	1596	397	462	2	4			
1	J	296	Total	C	N	O	S	Se	0	0	0
			2441	1581	394	460	2	4			
1	K	296	Total	C	N	O	S	Se	0	0	0
			2441	1581	394	460	2	4			
1	L	296	Total	C	N	O	S	Se	0	0	0
			2441	1581	394	460	2	4			

There are 60 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	27	MSE	MET	modified residue	UNP Q04114
A	67	MSE	MET	modified residue	UNP Q04114
A	153	CYS	TYR	engineered mutation	UNP Q04114
A	175	MSE	MET	modified residue	UNP Q04114
A	203	MSE	MET	modified residue	UNP Q04114

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	27	MSE	MET	modified residue	UNP Q04114
B	67	MSE	MET	modified residue	UNP Q04114
B	153	CYS	TYR	engineered mutation	UNP Q04114
B	175	MSE	MET	modified residue	UNP Q04114
B	203	MSE	MET	modified residue	UNP Q04114
C	27	MSE	MET	modified residue	UNP Q04114
C	67	MSE	MET	modified residue	UNP Q04114
C	153	CYS	TYR	engineered mutation	UNP Q04114
C	175	MSE	MET	modified residue	UNP Q04114
C	203	MSE	MET	modified residue	UNP Q04114
D	27	MSE	MET	modified residue	UNP Q04114
D	67	MSE	MET	modified residue	UNP Q04114
D	153	CYS	TYR	engineered mutation	UNP Q04114
D	175	MSE	MET	modified residue	UNP Q04114
D	203	MSE	MET	modified residue	UNP Q04114
E	27	MSE	MET	modified residue	UNP Q04114
E	67	MSE	MET	modified residue	UNP Q04114
E	153	CYS	TYR	engineered mutation	UNP Q04114
E	175	MSE	MET	modified residue	UNP Q04114
E	203	MSE	MET	modified residue	UNP Q04114
F	27	MSE	MET	modified residue	UNP Q04114
F	67	MSE	MET	modified residue	UNP Q04114
F	153	CYS	TYR	engineered mutation	UNP Q04114
F	175	MSE	MET	modified residue	UNP Q04114
F	203	MSE	MET	modified residue	UNP Q04114
G	27	MSE	MET	modified residue	UNP Q04114
G	67	MSE	MET	modified residue	UNP Q04114
G	153	CYS	TYR	engineered mutation	UNP Q04114
G	175	MSE	MET	modified residue	UNP Q04114
G	203	MSE	MET	modified residue	UNP Q04114
H	27	MSE	MET	modified residue	UNP Q04114
H	67	MSE	MET	modified residue	UNP Q04114
H	153	CYS	TYR	engineered mutation	UNP Q04114
H	175	MSE	MET	modified residue	UNP Q04114
H	203	MSE	MET	modified residue	UNP Q04114
I	27	MSE	MET	modified residue	UNP Q04114
I	67	MSE	MET	modified residue	UNP Q04114
I	153	CYS	TYR	engineered mutation	UNP Q04114
I	175	MSE	MET	modified residue	UNP Q04114
I	203	MSE	MET	modified residue	UNP Q04114
J	27	MSE	MET	modified residue	UNP Q04114
J	67	MSE	MET	modified residue	UNP Q04114

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
J	153	CYS	TYR	engineered mutation	UNP Q04114
J	175	MSE	MET	modified residue	UNP Q04114
J	203	MSE	MET	modified residue	UNP Q04114
K	27	MSE	MET	modified residue	UNP Q04114
K	67	MSE	MET	modified residue	UNP Q04114
K	153	CYS	TYR	engineered mutation	UNP Q04114
K	175	MSE	MET	modified residue	UNP Q04114
K	203	MSE	MET	modified residue	UNP Q04114
L	27	MSE	MET	modified residue	UNP Q04114
L	67	MSE	MET	modified residue	UNP Q04114
L	153	CYS	TYR	engineered mutation	UNP Q04114
L	175	MSE	MET	modified residue	UNP Q04114
L	203	MSE	MET	modified residue	UNP Q04114

- Molecule 2 is water.

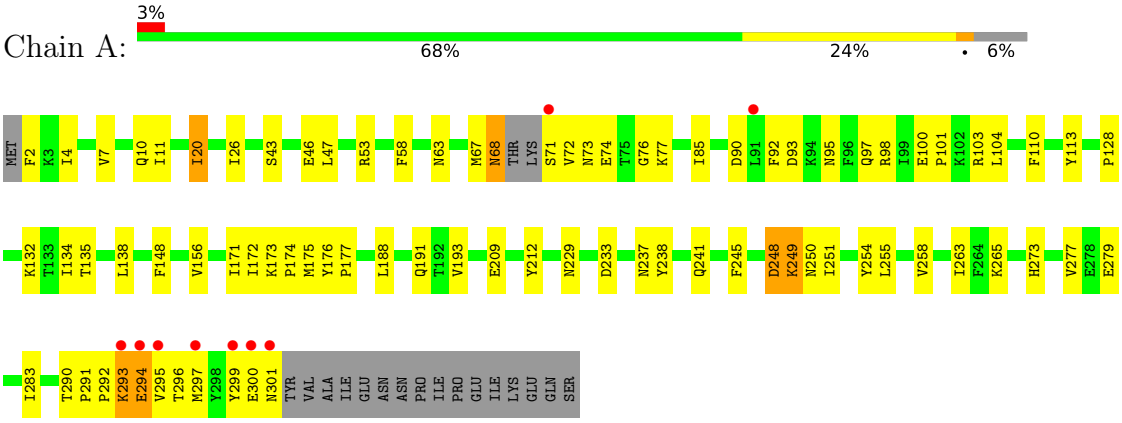
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	67	Total O 67 67	0	0
2	B	96	Total O 96 96	0	0
2	C	43	Total O 43 43	0	0
2	D	57	Total O 57 57	0	0
2	E	67	Total O 67 67	0	0
2	F	59	Total O 59 59	0	0
2	G	42	Total O 42 42	0	0
2	H	37	Total O 37 37	0	0
2	I	23	Total O 23 23	0	0
2	J	47	Total O 47 47	0	0
2	K	77	Total O 77 77	0	0
2	L	33	Total O 33 33	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.

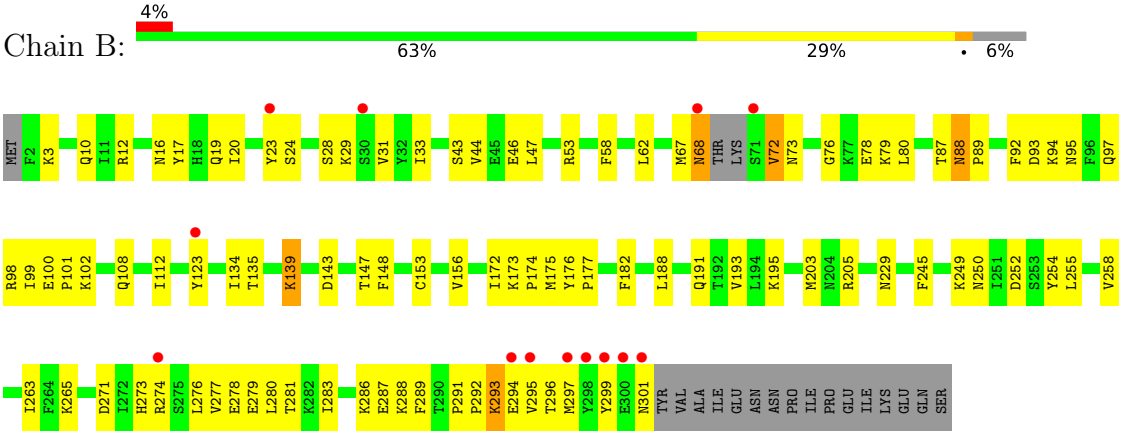
• Molecule 1: PrgX

Chain A:



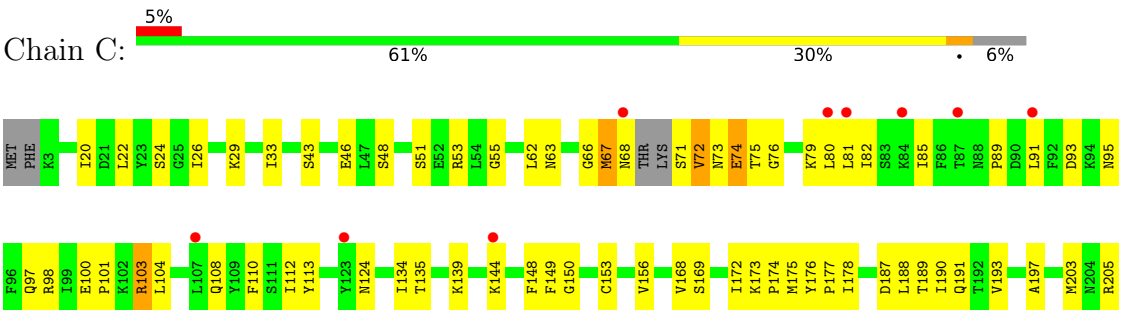
• Molecule 1: PrgX

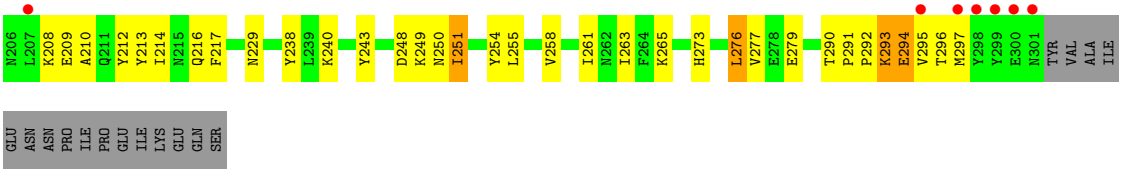
Chain B:



• Molecule 1: PrgX

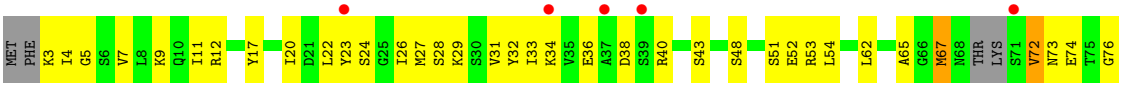
Chain C:

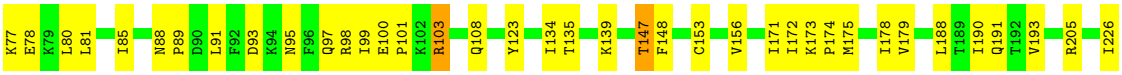


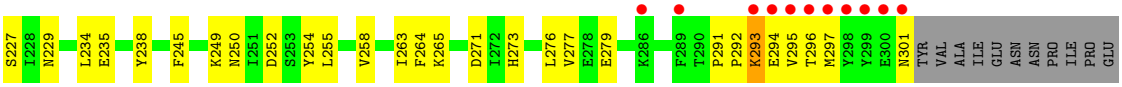


• Molecule 1: PrgX

Chain D:

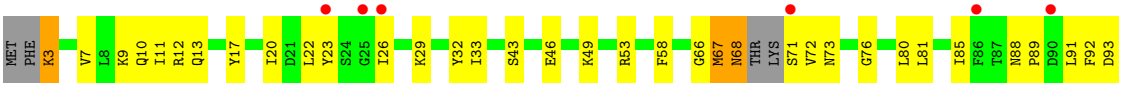


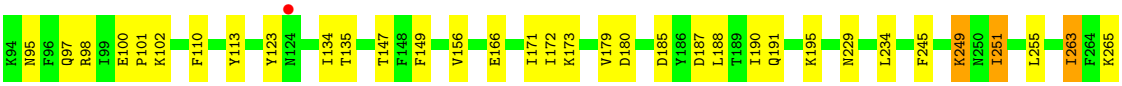


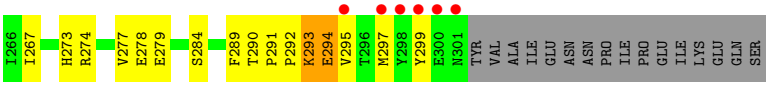


• Molecule 1: PrgX

Chain E:

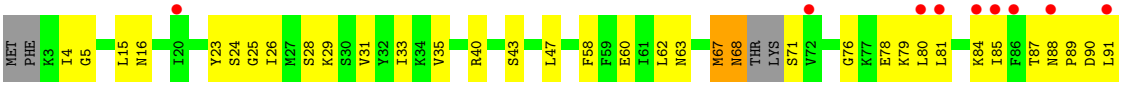


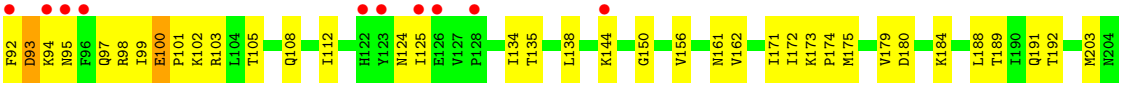


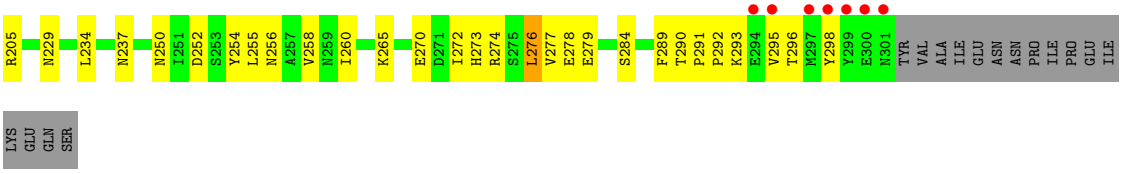


• Molecule 1: PrgX

Chain F:

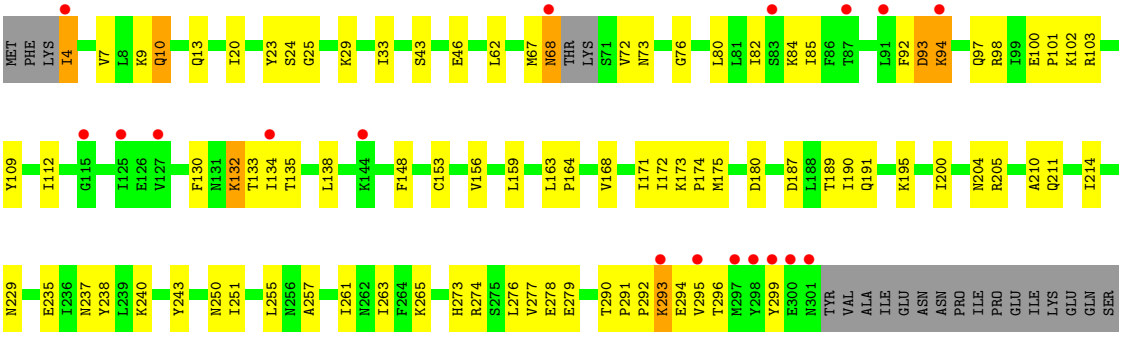






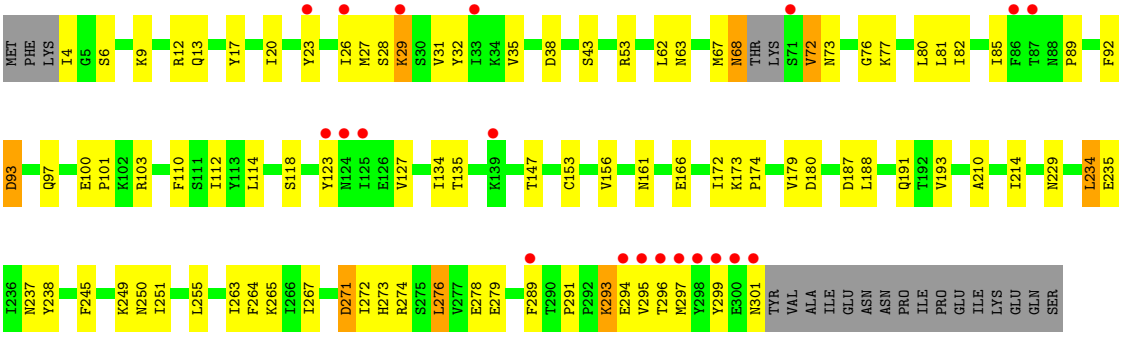
• Molecule 1: PrgX

Chain G:



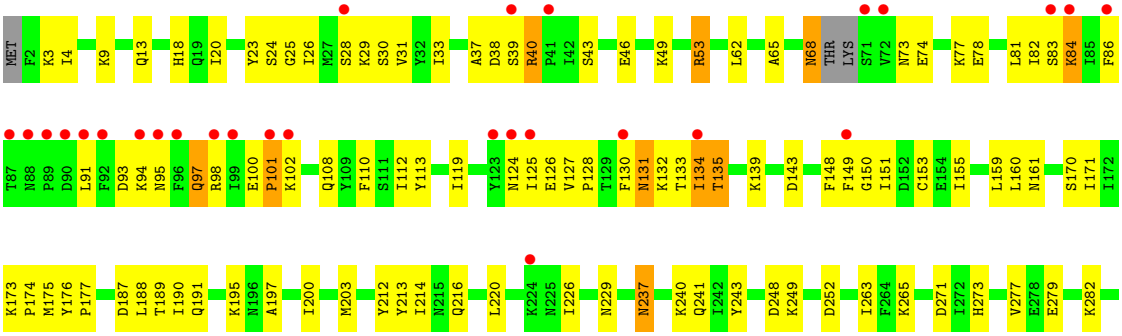
• Molecule 1: PrgX

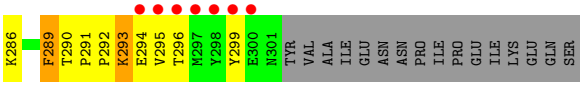
Chain H:



• Molecule 1: PrgX

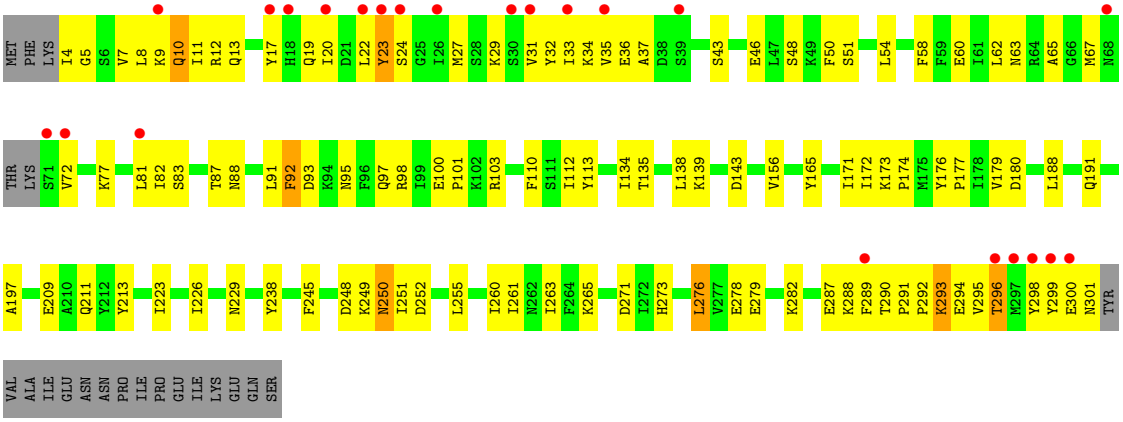
Chain I:





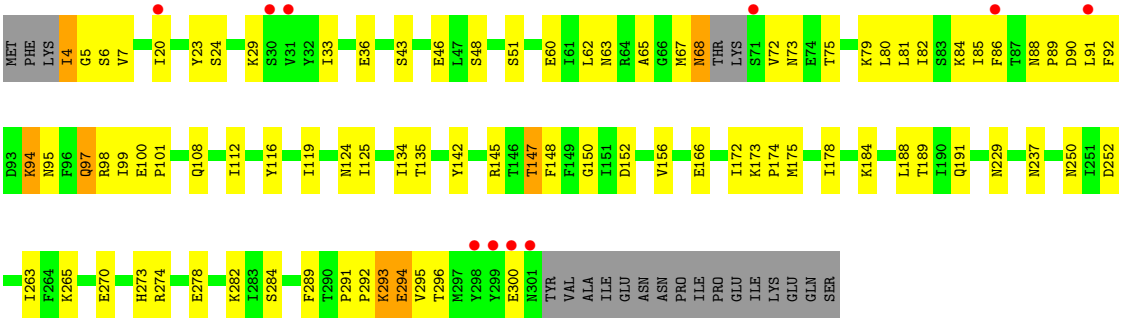
• Molecule 1: PrgX

Chain J:



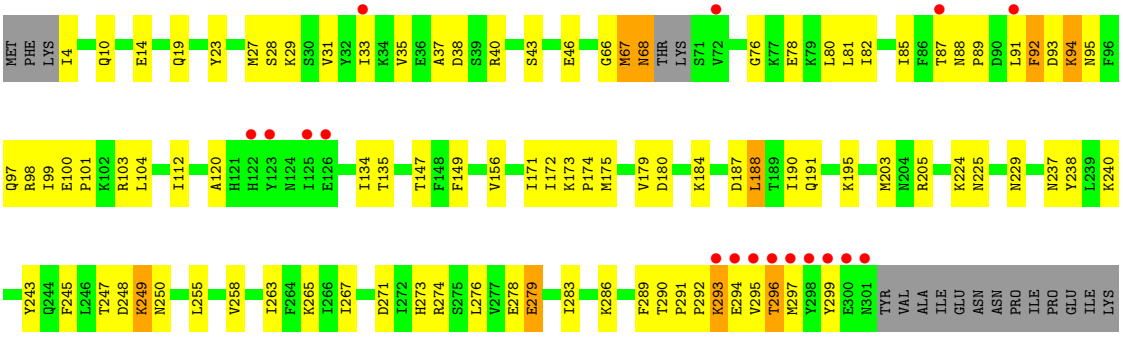
• Molecule 1: PrgX

Chain K:



• Molecule 1: PrgX

Chain L:




GLU
GLN
SER

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	92.53Å 134.37Å 195.20Å 90.00° 100.28° 90.00°	Depositor
Resolution (Å)	29.90 – 2.25 29.85 – 2.25	Depositor EDS
% Data completeness (in resolution range)	98.6 (29.90-2.25) 95.4 (29.85-2.25)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.83 (at 2.24Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.218 , 0.271 0.234 , 0.260	Depositor DCC
R_{free} test set	10751 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å^2)	42.1	Xtriage
Anisotropy	0.383	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 52.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	30035	wwPDB-VP
Average B, all atoms (Å^2)	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.89% 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

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	0.33	0/2506	0.53	0/3376
1	B	0.33	0/2506	0.54	0/3376
1	C	0.31	0/2494	0.51	0/3360
1	D	0.32	0/2494	0.52	0/3360
1	E	0.32	0/2494	0.52	0/3360
1	F	0.31	0/2493	0.52	0/3359
1	G	0.30	0/2485	0.52	0/3349
1	H	0.31	0/2485	0.52	0/3349
1	I	0.30	0/2506	0.49	0/3376
1	J	0.31	0/2485	0.51	0/3349
1	K	0.32	0/2485	0.51	0/3349
1	L	0.31	0/2485	0.52	0/3349
All	All	0.31	0/29918	0.52	0/40312

There are no bond length outliers.

There are no bond angle outliers.

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	2461	0	2493	76	0
1	B	2461	0	2493	81	0
1	C	2450	0	2484	83	0
1	D	2450	0	2484	76	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	E	2450	0	2484	75	0
1	F	2449	0	2484	90	0
1	G	2441	0	2471	80	0
1	H	2441	0	2471	88	0
1	I	2461	0	2493	119	0
1	J	2441	0	2471	90	0
1	K	2441	0	2471	73	0
1	L	2441	0	2471	81	0
2	A	67	0	0	10	0
2	B	96	0	0	8	0
2	C	43	0	0	2	0
2	D	57	0	0	4	0
2	E	67	0	0	5	0
2	F	59	0	0	8	0
2	G	42	0	0	2	0
2	H	37	0	0	2	0
2	I	23	0	0	6	0
2	J	47	0	0	8	0
2	K	77	0	0	9	0
2	L	33	0	0	3	0
All	All	30035	0	29770	952	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 16.

The worst 5 of 952 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:K:82:ILE:HD11	1:K:112:ILE:HG23	1.35	1.08
1:K:263:ILE:HG13	1:L:263:ILE:HG13	1.42	1.01
1:A:263:ILE:HG13	1:B:263:ILE:HG13	1.42	0.99
1:J:62:LEU:HB3	1:J:67:MSE:HG3	1.44	0.97
1:I:20:ILE:HD12	1:I:20:ILE:H	1.25	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	294/317 (93%)	281 (96%)	10 (3%)	3 (1%)	15	13
1	B	294/317 (93%)	281 (96%)	11 (4%)	2 (1%)	22	21
1	C	293/317 (92%)	276 (94%)	10 (3%)	7 (2%)	6	3
1	D	293/317 (92%)	274 (94%)	17 (6%)	2 (1%)	22	21
1	E	293/317 (92%)	277 (94%)	13 (4%)	3 (1%)	15	13
1	F	293/317 (92%)	269 (92%)	21 (7%)	3 (1%)	15	13
1	G	292/317 (92%)	273 (94%)	16 (6%)	3 (1%)	15	13
1	H	292/317 (92%)	278 (95%)	11 (4%)	3 (1%)	15	13
1	I	294/317 (93%)	262 (89%)	27 (9%)	5 (2%)	9	4
1	J	292/317 (92%)	271 (93%)	18 (6%)	3 (1%)	15	13
1	K	292/317 (92%)	267 (91%)	22 (8%)	3 (1%)	15	13
1	L	292/317 (92%)	272 (93%)	17 (6%)	3 (1%)	15	13
All	All	3514/3804 (92%)	3281 (93%)	193 (6%)	40 (1%)	14	10

5 of 40 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	249	LYS
1	C	72	VAL
1	C	249	LYS
1	E	92	PHE
1	F	24	SER

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	281/295 (95%)	273 (97%)	8 (3%)	43	52
1	B	281/295 (95%)	267 (95%)	14 (5%)	24	26
1	C	280/295 (95%)	272 (97%)	8 (3%)	42	51
1	D	280/295 (95%)	273 (98%)	7 (2%)	47	56
1	E	280/295 (95%)	269 (96%)	11 (4%)	32	38
1	F	280/295 (95%)	272 (97%)	8 (3%)	42	51
1	G	279/295 (95%)	270 (97%)	9 (3%)	39	47
1	H	279/295 (95%)	270 (97%)	9 (3%)	39	47
1	I	281/295 (95%)	268 (95%)	13 (5%)	27	30
1	J	279/295 (95%)	268 (96%)	11 (4%)	32	38
1	K	279/295 (95%)	269 (96%)	10 (4%)	35	42
1	L	279/295 (95%)	265 (95%)	14 (5%)	24	26
All	All	3358/3540 (95%)	3236 (96%)	122 (4%)	35	42

5 of 122 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	10	GLN
1	L	68	ASN
1	H	271	ASP
1	L	67	MSE
1	L	279	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 168 such sidechains are listed below:

Mol	Chain	Res	Type
1	I	68	ASN
1	K	68	ASN
1	I	121	HIS
1	J	19	GLN
1	K	215	ASN

5.3.3 RNA

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

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, 95^{th} 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	294/317 (92%)	0.11	9 (3%) 49 52	17, 40, 79, 114	0
1	B	294/317 (92%)	0.15	13 (4%) 34 37	12, 39, 75, 117	0
1	C	293/317 (92%)	0.41	16 (5%) 25 27	25, 52, 93, 113	0
1	D	293/317 (92%)	0.32	16 (5%) 25 27	20, 46, 88, 119	0
1	E	293/317 (92%)	0.22	13 (4%) 34 37	22, 45, 86, 114	0
1	F	293/317 (92%)	0.45	26 (8%) 9 10	24, 52, 98, 118	0
1	G	292/317 (92%)	0.42	18 (6%) 20 22	27, 56, 95, 112	0
1	H	292/317 (92%)	0.45	20 (6%) 17 18	25, 52, 95, 130	0
1	I	294/317 (92%)	0.61	35 (11%) 4 3	29, 64, 110, 124	0
1	J	292/317 (92%)	0.30	23 (7%) 12 14	23, 49, 98, 114	0
1	K	292/317 (92%)	0.36	10 (3%) 45 47	22, 47, 95, 112	0
1	L	292/317 (92%)	0.33	17 (5%) 23 25	23, 49, 89, 126	0
All	All	3514/3804 (92%)	0.34	216 (6%) 21 23	12, 49, 96, 130	0

The worst 5 of 216 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	H	299	TYR	9.1
1	H	301	ASN	8.3
1	I	91	LEU	8.3
1	H	300	GLU	7.6
1	H	297	MET	7.1

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

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.