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)

Sep 25, 2023 – 01:04 pm BST

PDB ID	:	8P26
Title	:	Crystal structure of Arabidopsis thaliana PAXX
Authors	:	Ochi, T.
Deposited on	:	2023-05-15
Resolution	:	3.60 Å(reported)

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

|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
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)	:	2.35.1

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 3.60 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.
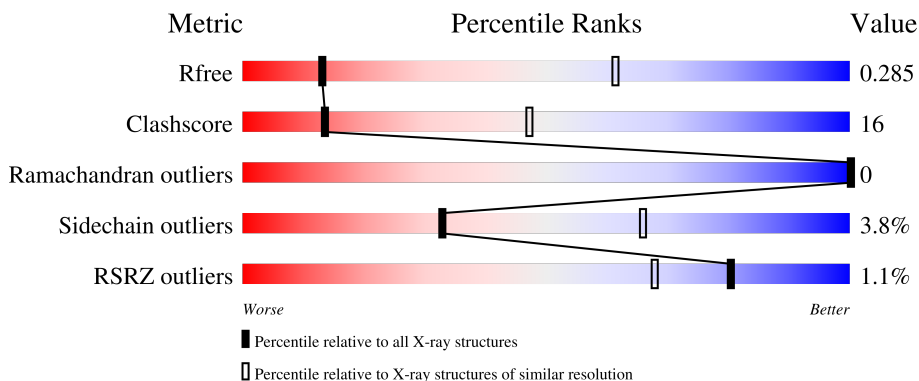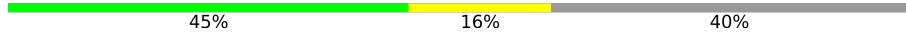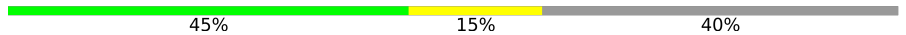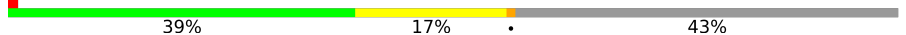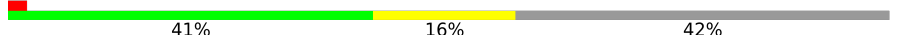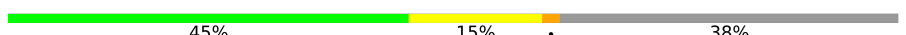

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1257 (3.70-3.50)
Clashscore	141614	1353 (3.70-3.50)
Ramachandran outliers	138981	1307 (3.70-3.50)
Sidechain outliers	138945	1307 (3.70-3.50)
RSRZ outliers	127900	1161 (3.70-3.50)

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 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	238	40% 17% • 41%
1	B	238	43% 15% • 41%
1	C	238	43% 14% 42%
1	D	238	42% 16% • 41%
1	E	238	% 42% 15% 42%

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	238	45% 16% 40%
1	G	238	45% 15% 40%
1	H	238	.% 39% 17% · 43%
1	I	238	2% 41% 16% 42%
1	J	238	45% 15% · 38%

2 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 10142 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 U2 small nuclear ribonucleoprotein auxiliary factor-like protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	140	Total	C	N	O	S	0	0	0
			1025	651	165	203	6			
1	B	140	Total	C	N	O	S	0	0	0
			1036	660	170	200	6			
1	C	137	Total	C	N	O	S	0	0	0
			970	625	155	184	6			
1	D	140	Total	C	N	O	S	0	0	0
			1005	643	161	196	5			
1	E	137	Total	C	N	O	S	0	0	0
			984	632	159	188	5			
1	F	143	Total	C	N	O	S	0	0	0
			1047	665	174	202	6			
1	G	143	Total	C	N	O	S	0	0	0
			1038	660	169	204	5			
1	H	135	Total	C	N	O	S	0	0	0
			987	630	161	190	6			
1	I	137	Total	C	N	O	S	0	0	0
			984	631	156	191	6			
1	J	147	Total	C	N	O	S	0	0	0
			1044	660	173	207	4			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	expression tag	UNP F4KC77
A	0	SER	-	expression tag	UNP F4KC77
B	-1	GLY	-	expression tag	UNP F4KC77
B	0	SER	-	expression tag	UNP F4KC77
C	-1	GLY	-	expression tag	UNP F4KC77
C	0	SER	-	expression tag	UNP F4KC77
D	-1	GLY	-	expression tag	UNP F4KC77
D	0	SER	-	expression tag	UNP F4KC77
E	-1	GLY	-	expression tag	UNP F4KC77

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	0	SER	-	expression tag	UNP F4KC77
F	-1	GLY	-	expression tag	UNP F4KC77
F	0	SER	-	expression tag	UNP F4KC77
G	-1	GLY	-	expression tag	UNP F4KC77
G	0	SER	-	expression tag	UNP F4KC77
H	-1	GLY	-	expression tag	UNP F4KC77
H	0	SER	-	expression tag	UNP F4KC77
I	-1	GLY	-	expression tag	UNP F4KC77
I	0	SER	-	expression tag	UNP F4KC77
J	-1	GLY	-	expression tag	UNP F4KC77
J	0	SER	-	expression tag	UNP F4KC77

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	3	Total 3	O 3	0	0
2	B	2	Total 2	O 2	0	0
2	C	2	Total 2	O 2	0	0
2	D	2	Total 2	O 2	0	0
2	E	3	Total 3	O 3	0	0
2	F	1	Total 1	O 1	0	0
2	G	4	Total 4	O 4	0	0
2	H	1	Total 1	O 1	0	0
2	I	3	Total 3	O 3	0	0
2	J	1	Total 1	O 1	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: U2 small nuclear ribonucleoprotein auxiliary factor-like protein

Chain A:

• Molecule 1: U2 small nuclear ribonucleoprotein auxiliary factor-like protein

Chain B:

• Molecule 1: U2 small nuclear ribonucleoprotein auxiliary factor-like protein

Chain C:

• Molecule 1: U2 small nuclear ribonucleoprotein auxiliary factor-like protein

Chain D:

• Molecule 1: U2 small nuclear ribonucleoprotein auxiliary factor-like protein

Chain E:
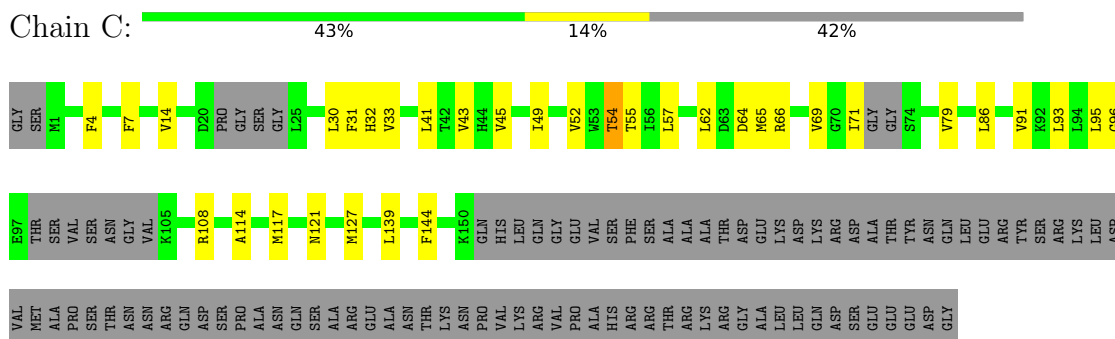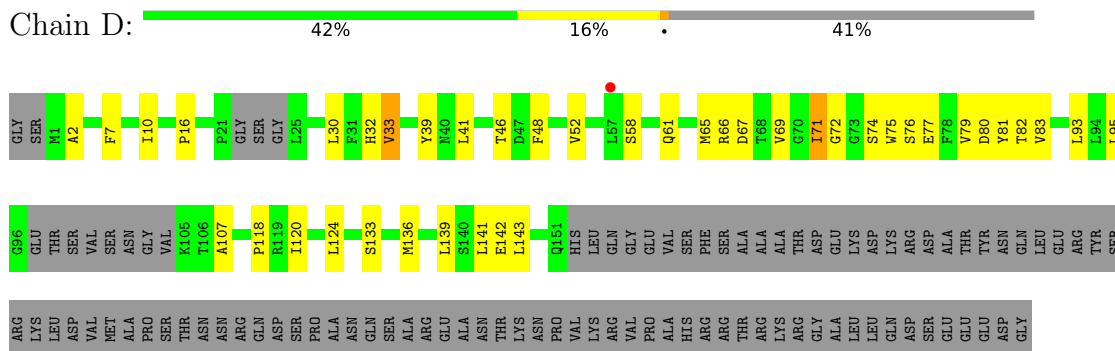

• Molecule 1: U2 small nuclear ribonucleoprotein auxiliary factor-like protein

Chain F:

• Molecule 1: U2 small nuclear ribonucleoprotein auxiliary factor-like protein
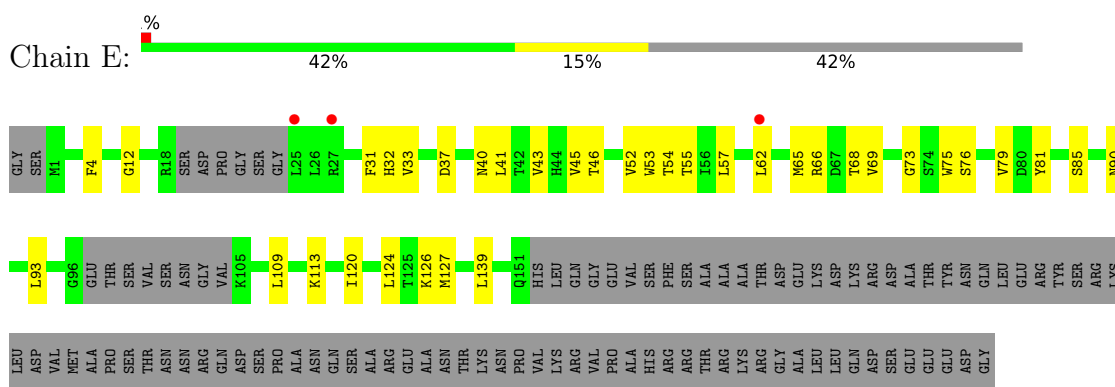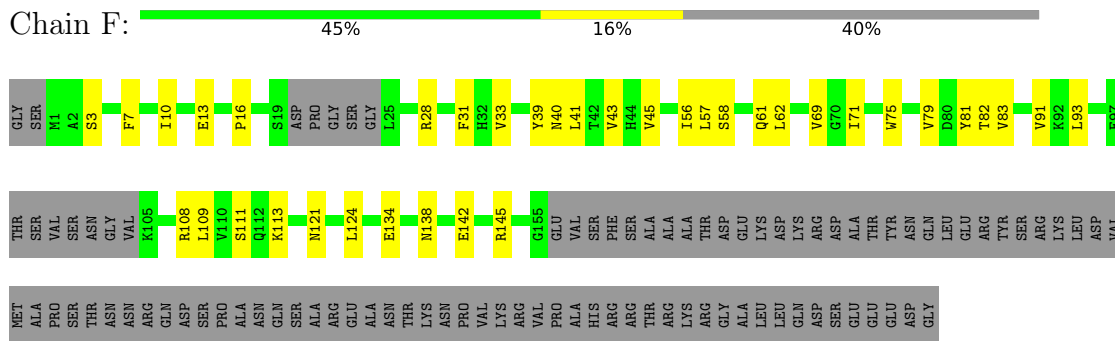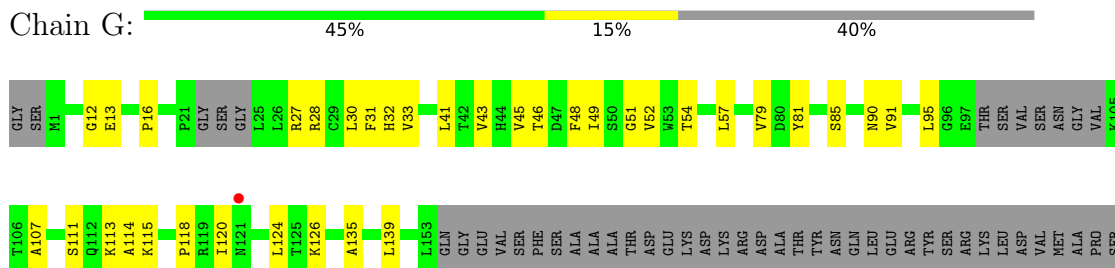
Chain G:

THR ASN ARG GLN ASP PRO ALA ASN GLN SER ALA ARG GLU ALA ASN THR LYS ASN PRO VAL LYS ARG VAL PRO ALA HIS ARG ARG THR ARG LYS ARG GLY ALA LEU LEU GLN ASP SER GLU GLU ASP GLY

• Molecule 1: U2 small nuclear ribonucleoprotein auxiliary factor-like protein

Chain H:

%

39% 17% • 43%

GLY SER M1 F4 F7 F11 V14 V15 P16 GLU ARG SER ASP PRO GLY SER GLY L25 L26 R27 R28 F31 H32 Y39 N40 L41 V45 T46 D47 F48 I49 V52 W53 T54 L57 Q61 M65 I71 F78 V79 V83 A84 S85 L86 K87

L93 L94 E97 THR SER VAL SER ASN GLY VAL K105 R108 L109 V110 S111 Q112 K113 A114 N121 V122 K126 S130 E134 A135 L139 S140 L141 F144 K150 GLN HIS LEU GLN GLY GLU VAL SER PHE SER ALA ALA THR ASP LYS ASP LYS ARG ASP ALA

THR TYR ASN GLN LEU GLU GLN ARG SER TYR SER ARG LYS LEU ASP VAL MET ALA PRO SER THR ASN ASN ARG GLN ASP SER PRO ALA ASN GLN SER ALA ARG GLU ALA ASN THR LYS ASN PRO VAL LYS ARG VAL PRO ALA HIS ARG ARG THR ARG LYS ARG GLY ALA LEU LEU GLN ASP SER GLU

GLU GLU ASP GLY

• Molecule 1: U2 small nuclear ribonucleoprotein auxiliary factor-like protein

Chain I:

2%

41% 16% 42%

GLY SER M1 I10 P16 S19 ASP PRO GLY SER GLY L25 L26 R27 F31 H32 V33 Y39 N40 L41 T42 V43 H44 V45 T46 D47 F48 I49 V52 W53 T54 T55 I56 L57 S58 V59 R66 W75 V79 D80 Y81 T82 V83 N90 V91 K92 L93 L94 L95 G96

GLU THR SER VAL SER ASN GLY VAL K105 T106 A107 R108 L109 V110 S111 A114 M117 I120 P123 L124 T125 K126 S130 L139 S140 L141 K150 GLN HIS LEU GLN GLY GLU VAL SER PHE SER ALA ALA ALA THR ASP GLU LYS ASP LYS ARG ASP ALA THR TYR ASN GLN

LEU GLU ARG TYR SER ARG LYS LEU ASP VAL MET ALA PRO SER THR ASN ASN ARG GLN ASP SER PRO ALA ASN GLN SER ALA ARG GLU ALA ASN THR LYS ASN PRO VAL LYS ARG VAL PRO ALA HIS ARG ARG THR ARG LYS ARG GLY ALA LEU LEU GLN ASP SER GLU GLU GLU ASP GLY

• Molecule 1: U2 small nuclear ribonucleoprotein auxiliary factor-like protein

Chain J:

45% 15% • 38%

GLY SER M1 F7 E8 P9 I10 V14 V15 P16 D20 PRO GLY SER G24 R28 C29 L30 F31 H32 V33 Y34 A35 Y39 N40 L41 H44 V45 W53 S58 Q61 D64 M65 R66 G73 S74 W75 V79 T82 V83 V91 K92 L93 V100 S101

S111 I120 N121 V122 P123 L124 T125 L139 S140 L141 K150 GLN HIS LEU GLN GLY GLU VAL PRO ALA HIS ARG ARG THR ARG LYS ARG GLY ALA LEU LEU GLN ASP SER GLU GLU GLU ASP GLY

ASP SER PRO ALA ASN GLN SER ALA ARG GLU GLU ALA ASN THR LYS ASN PRO VAL LYS ARG VAL PRO ALA HIS ARG ARG THR ARG LYS ARG GLY ALA LEU LEU GLN ASP SER GLU GLU GLU ASP GLY

4   Data and refinement statistics   ⓘ

Property	Value	Source
Space group	P 21 2 21	Depositor
Cell constants a, b, c, α, β, γ	107.03Å 152.42Å 187.09Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	152.42 – 3.60 152.42 – 3.60	Depositor EDS
% Data completeness (in resolution range)	99.6 (152.42-3.60) 99.6 (152.42-3.60)	Depositor EDS
R_{merge}	0.24	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.96 (at 3.58Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.248 , 0.286 0.246 , 0.285	Depositor DCC
R_{free} test set	1794 reflections (4.97%)	wwPDB-VP
Wilson B-factor (Å²)	107.3	Xtriage
Anisotropy	0.632	Xtriage
Bulk solvent k_{sol}(e/Å³), B_{sol}(Å²)	0.30 , 110.7	EDS
L-test for twinning [2]	$< \lvert L \rvert > = 0.47, < L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	10142	wwPDB-VP
Average B, all atoms (Å²)	119.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.46% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $< \lvert L \rvert >$, $< L^2 >$ 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	0.27	0/1043	0.53	0/1422
1	B	0.31	0/1055	0.52	0/1436
1	C	0.35	0/987	0.54	0/1349
1	D	0.47	0/1024	0.61	0/1400
1	E	0.34	0/1002	0.57	0/1369
1	F	0.31	0/1066	0.53	0/1451
1	G	0.35	0/1057	0.56	0/1442
1	H	0.34	0/1005	0.53	0/1368
1	I	0.36	0/1002	0.56	0/1368
1	J	0.38	0/1063	0.59	0/1454
All	All	0.35	0/10304	0.56	0/14059

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	1025	0	944	57	0
1	B	1036	0	972	31	0
1	C	970	0	888	22	0
1	D	1005	0	919	38	0
1	E	984	0	913	29	0
1	F	1047	0	972	30	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	G	1038	0	950	21	0
1	H	987	0	921	43	0
1	I	984	0	909	40	0
1	J	1044	0	945	32	0
2	A	3	0	0	0	0
2	B	2	0	0	0	0
2	C	2	0	0	0	0
2	D	2	0	0	0	0
2	E	3	0	0	0	0
2	F	1	0	0	0	0
2	G	4	0	0	0	0
2	H	1	0	0	0	0
2	I	3	0	0	0	0
2	J	1	0	0	0	0
All	All	10142	0	9333	312	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 312 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:A:48:PHE:HA	1:A:136:MET:SD	1.62	1.36
1:F:3:SER:O	1:J:100:VAL:HG22	1.39	1.19
1:A:62:LEU:HD23	1:A:75:TRP:CZ2	1.81	1.15
1:I:25:LEU:HG	1:I:26:LEU:H	1.03	1.10
1:D:10:ILE:CD1	1:D:83:VAL:HG13	1.81	1.09

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	134/238 (56%)	129 (96%)	5 (4%)	0	100	100
1	B	134/238 (56%)	130 (97%)	4 (3%)	0	100	100
1	C	129/238 (54%)	121 (94%)	8 (6%)	0	100	100
1	D	134/238 (56%)	122 (91%)	12 (9%)	0	100	100
1	E	131/238 (55%)	126 (96%)	5 (4%)	0	100	100
1	F	137/238 (58%)	132 (96%)	5 (4%)	0	100	100
1	G	137/238 (58%)	130 (95%)	7 (5%)	0	100	100
1	H	129/238 (54%)	125 (97%)	4 (3%)	0	100	100
1	I	131/238 (55%)	128 (98%)	3 (2%)	0	100	100
1	J	143/238 (60%)	134 (94%)	9 (6%)	0	100	100
All	All	1339/2380 (56%)	1277 (95%)	62 (5%)	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	107/205 (52%)	103 (96%)	4 (4%)	34	66
1	B	109/205 (53%)	106 (97%)	3 (3%)	43	72
1	C	97/205 (47%)	93 (96%)	4 (4%)	30	64
1	D	102/205 (50%)	98 (96%)	4 (4%)	32	65
1	E	100/205 (49%)	98 (98%)	2 (2%)	55	79
1	F	108/205 (53%)	106 (98%)	2 (2%)	57	80
1	G	106/205 (52%)	101 (95%)	5 (5%)	26	61
1	H	102/205 (50%)	98 (96%)	4 (4%)	32	65
1	I	101/205 (49%)	97 (96%)	4 (4%)	31	65
1	J	105/205 (51%)	98 (93%)	7 (7%)	16	50
All	All	1037/2050 (51%)	998 (96%)	39 (4%)	33	66

5 of 39 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	I	42	THR
1	J	66	ARG
1	I	79	VAL
1	J	33	VAL
1	J	125	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	F	61	GLN
1	G	112	GLN
1	J	32	HIS
1	H	112	GLN
1	E	90	ASN

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 monosaccharides 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 (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, 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	140/238 (58%)	0.16	0	100	100	72, 110, 181, 211	0
1	B	140/238 (58%)	0.11	1 (0%)	87	78	73, 102, 156, 218	0
1	C	137/238 (57%)	-0.01	0	100	100	79, 114, 169, 201	0
1	D	140/238 (58%)	0.07	1 (0%)	87	78	87, 125, 192, 259	0
1	E	137/238 (57%)	0.13	3 (2%)	62	45	77, 112, 172, 202	0
1	F	143/238 (60%)	0.10	0	100	100	78, 118, 186, 248	0
1	G	143/238 (60%)	0.17	1 (0%)	87	78	77, 109, 167, 256	0
1	H	135/238 (56%)	0.15	3 (2%)	62	45	85, 123, 180, 235	0
1	I	137/238 (57%)	0.15	5 (3%)	42	28	82, 120, 189, 217	0
1	J	147/238 (61%)	0.09	1 (0%)	87	78	83, 118, 204, 255	0
All	All	1399/2380 (58%)	0.11	15 (1%)	80	68	72, 116, 184, 259	0

The worst 5 of 15 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	25	LEU	3.0
1	H	27	ARG	2.9
1	I	109	LEU	2.7
1	I	16	PRO	2.7
1	D	57	LEU	2.6

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

There are no non-standard protein/DNA/RNA residues in this entry.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

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.