



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 14, 2016 – 06:02 PM EDT

PDB ID : 4Z32  
Title : Crystal Structure of the FERM-SH2 Domains of Jak2  
Authors : McNally, R.; Eck, M.J.  
Deposited on : 2015-03-30  
Resolution : 3.04 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](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/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 : unknown  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20027790  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20027790

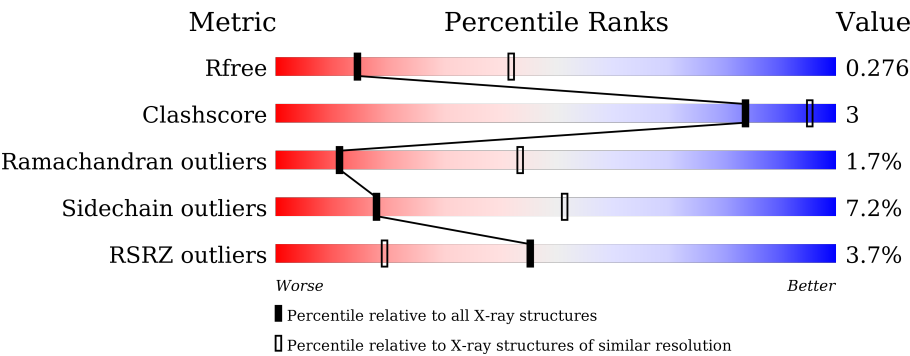
# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.04 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))
$R_{free}$	91344	1995 (3.08-3.00)
Clashscore	102246	2351 (3.08-3.00)
Ramachandran outliers	100387	2272 (3.08-3.00)
Sidechain outliers	100360	2275 (3.08-3.00)
RSRZ outliers	91569	2013 (3.08-3.00)

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  $\geq 3$ , 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	497	<div><div></div><div><div></div><div></div><div></div><div></div></div><div>75%13%11%</div></div>
1	B	497	<div><div></div><div><div></div><div></div><div></div><div></div></div><div>73%14%11%</div></div>
1	C	497	<div><div>6%</div><div><div></div><div></div><div></div><div></div></div><div>70%13%16%</div></div>
1	D	497	<div><div>3%</div><div><div></div><div></div><div></div><div></div></div><div>71%12%16%</div></div>
1	E	497	<div><div></div><div><div></div><div></div><div></div><div></div></div><div>73%12%14%</div></div>
1	F	497	<div><div></div><div><div></div><div></div><div></div><div></div></div><div>73%13%12%</div></div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	497	<div><div></div><div>4%</div><div>74%</div><div>8%</div><div>17%</div></div>
1	H	497	<div><div></div><div>7%</div><div>72%</div><div>9%</div><div>19%</div></div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 27929 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 Tyrosine-protein kinase JAK2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	441	Total	C	N	O	S	0	0	0
			3624	2334	622	646	22			
1	B	441	Total	C	N	O	S	0	0	0
			3624	2334	622	646	22			
1	C	418	Total	C	N	O	S	0	0	0
			3433	2211	588	612	22			
1	D	417	Total	C	N	O	S	0	0	0
			3423	2208	587	606	22			
1	E	427	Total	C	N	O	S	0	0	0
			3516	2271	602	622	21			
1	F	436	Total	C	N	O	S	0	0	0
			3586	2311	616	637	22			
1	G	411	Total	C	N	O	S	0	0	0
			3375	2179	578	597	21			
1	H	405	Total	C	N	O	S	0	0	0
			3336	2154	573	588	21			

There are 88 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	30	GLY	-	expression tag	UNP O60674
A	517	LEU	-	expression tag	UNP O60674
A	518	GLU	-	expression tag	UNP O60674
A	519	HIS	-	expression tag	UNP O60674
A	520	HIS	-	expression tag	UNP O60674
A	521	HIS	-	expression tag	UNP O60674
A	522	HIS	-	expression tag	UNP O60674
A	523	HIS	-	expression tag	UNP O60674
A	524	HIS	-	expression tag	UNP O60674
A	525	HIS	-	expression tag	UNP O60674
A	526	HIS	-	expression tag	UNP O60674
B	30	GLY	-	expression tag	UNP O60674
B	517	LEU	-	expression tag	UNP O60674

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	518	GLU	-	expression tag	UNP O60674
B	519	HIS	-	expression tag	UNP O60674
B	520	HIS	-	expression tag	UNP O60674
B	521	HIS	-	expression tag	UNP O60674
B	522	HIS	-	expression tag	UNP O60674
B	523	HIS	-	expression tag	UNP O60674
B	524	HIS	-	expression tag	UNP O60674
B	525	HIS	-	expression tag	UNP O60674
B	526	HIS	-	expression tag	UNP O60674
C	30	GLY	-	expression tag	UNP O60674
C	517	LEU	-	expression tag	UNP O60674
C	518	GLU	-	expression tag	UNP O60674
C	519	HIS	-	expression tag	UNP O60674
C	520	HIS	-	expression tag	UNP O60674
C	521	HIS	-	expression tag	UNP O60674
C	522	HIS	-	expression tag	UNP O60674
C	523	HIS	-	expression tag	UNP O60674
C	524	HIS	-	expression tag	UNP O60674
C	525	HIS	-	expression tag	UNP O60674
C	526	HIS	-	expression tag	UNP O60674
D	30	GLY	-	expression tag	UNP O60674
D	517	LEU	-	expression tag	UNP O60674
D	518	GLU	-	expression tag	UNP O60674
D	519	HIS	-	expression tag	UNP O60674
D	520	HIS	-	expression tag	UNP O60674
D	521	HIS	-	expression tag	UNP O60674
D	522	HIS	-	expression tag	UNP O60674
D	523	HIS	-	expression tag	UNP O60674
D	524	HIS	-	expression tag	UNP O60674
D	525	HIS	-	expression tag	UNP O60674
D	526	HIS	-	expression tag	UNP O60674
E	30	GLY	-	expression tag	UNP O60674
E	517	LEU	-	expression tag	UNP O60674
E	518	GLU	-	expression tag	UNP O60674
E	519	HIS	-	expression tag	UNP O60674
E	520	HIS	-	expression tag	UNP O60674
E	521	HIS	-	expression tag	UNP O60674
E	522	HIS	-	expression tag	UNP O60674
E	523	HIS	-	expression tag	UNP O60674
E	524	HIS	-	expression tag	UNP O60674
E	525	HIS	-	expression tag	UNP O60674
E	526	HIS	-	expression tag	UNP O60674

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
F	30	GLY	-	expression tag	UNP O60674
F	517	LEU	-	expression tag	UNP O60674
F	518	GLU	-	expression tag	UNP O60674
F	519	HIS	-	expression tag	UNP O60674
F	520	HIS	-	expression tag	UNP O60674
F	521	HIS	-	expression tag	UNP O60674
F	522	HIS	-	expression tag	UNP O60674
F	523	HIS	-	expression tag	UNP O60674
F	524	HIS	-	expression tag	UNP O60674
F	525	HIS	-	expression tag	UNP O60674
F	526	HIS	-	expression tag	UNP O60674
G	30	GLY	-	expression tag	UNP O60674
G	517	LEU	-	expression tag	UNP O60674
G	518	GLU	-	expression tag	UNP O60674
G	519	HIS	-	expression tag	UNP O60674
G	520	HIS	-	expression tag	UNP O60674
G	521	HIS	-	expression tag	UNP O60674
G	522	HIS	-	expression tag	UNP O60674
G	523	HIS	-	expression tag	UNP O60674
G	524	HIS	-	expression tag	UNP O60674
G	525	HIS	-	expression tag	UNP O60674
G	526	HIS	-	expression tag	UNP O60674
H	30	GLY	-	expression tag	UNP O60674
H	517	LEU	-	expression tag	UNP O60674
H	518	GLU	-	expression tag	UNP O60674
H	519	HIS	-	expression tag	UNP O60674
H	520	HIS	-	expression tag	UNP O60674
H	521	HIS	-	expression tag	UNP O60674
H	522	HIS	-	expression tag	UNP O60674
H	523	HIS	-	expression tag	UNP O60674
H	524	HIS	-	expression tag	UNP O60674
H	525	HIS	-	expression tag	UNP O60674
H	526	HIS	-	expression tag	UNP O60674

- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	5	Total O 5 5	0	0
2	B	4	Total O 4 4	0	0
2	D	1	Total O 1 1	0	0

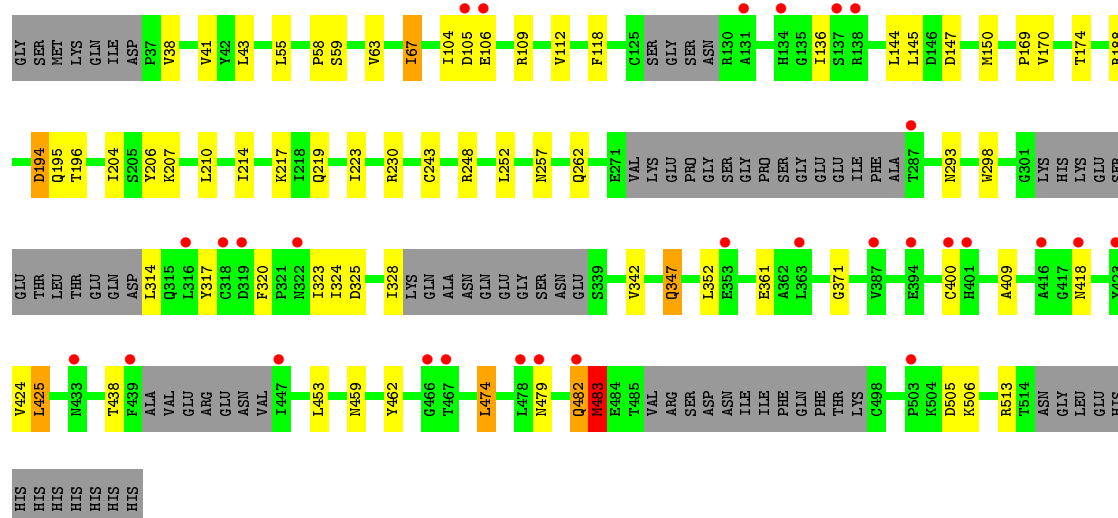
*Continued on next page...*

*Continued from previous page...*

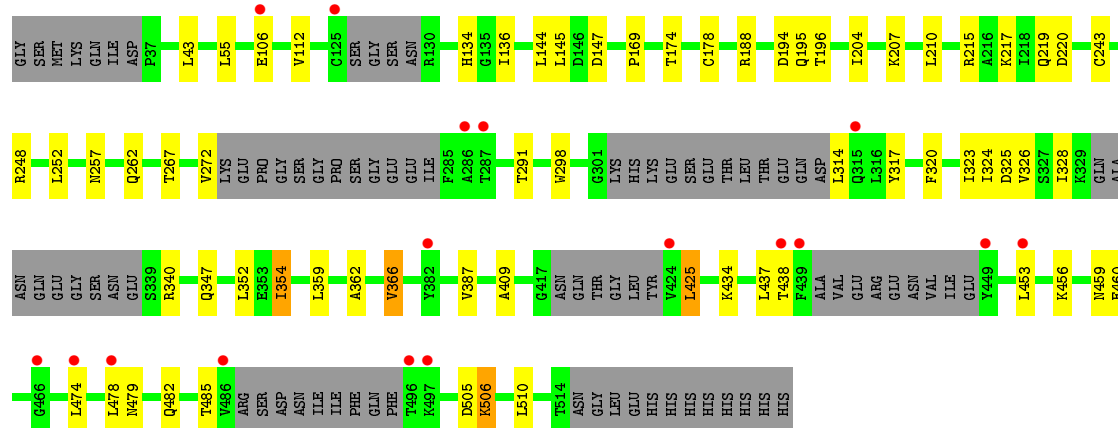
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	E	1	Total	O	0	0
			1	1		
2	F	1	Total	O	0	0
			1	1		



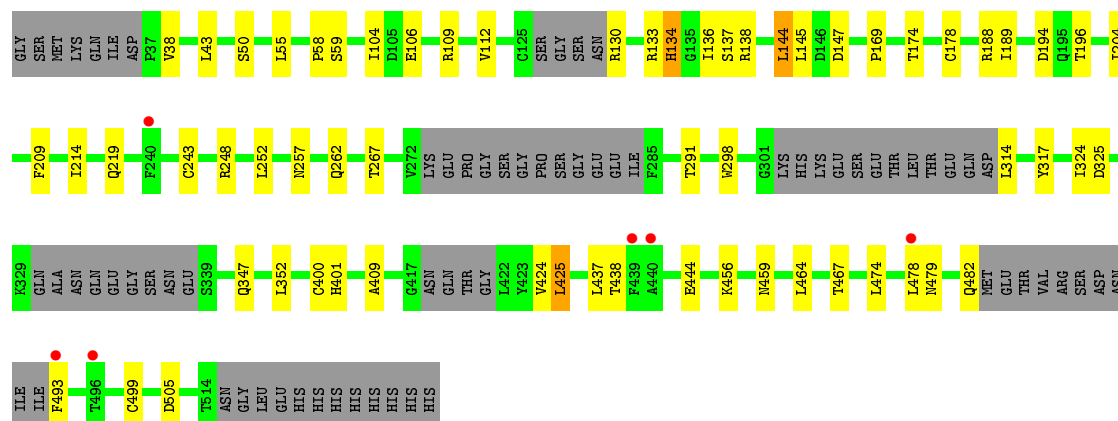
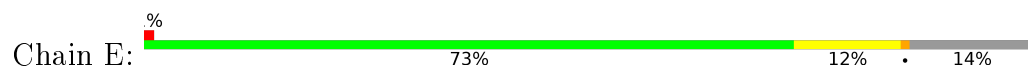




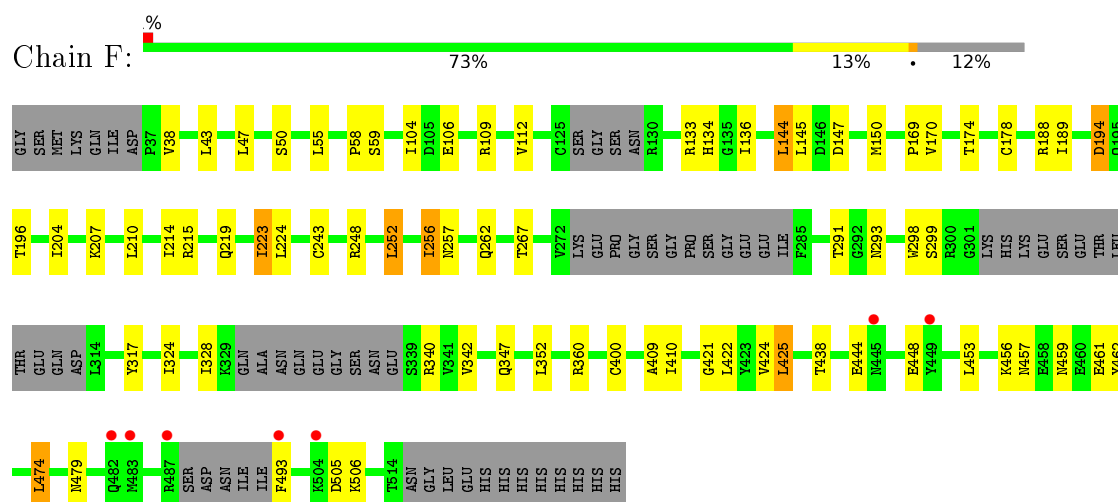
• Molecule 1: Tyrosine-protein kinase JAK2



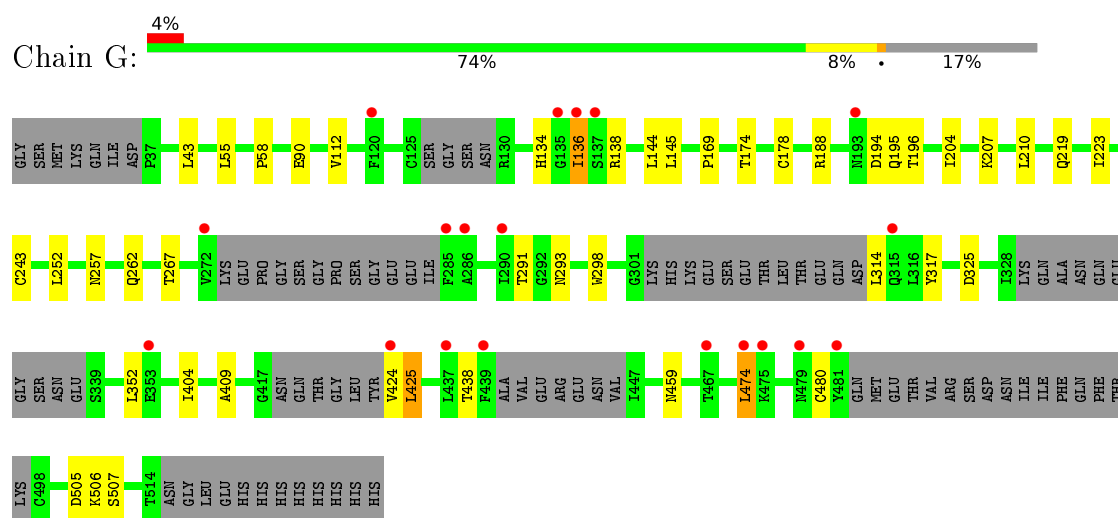
• Molecule 1: Tyrosine-protein kinase JAK2



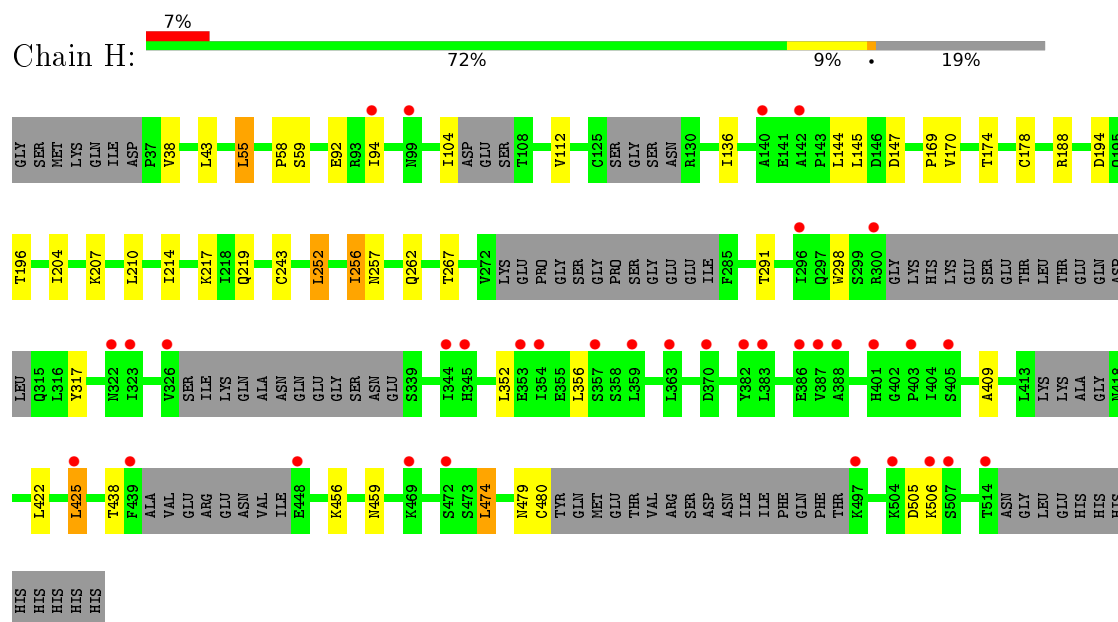
• Molecule 1: Tyrosine-protein kinase JAK2



• Molecule 1: Tyrosine-protein kinase JAK2



• Molecule 1: Tyrosine-protein kinase JAK2



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	118.19Å 188.74Å 118.58Å 90.00° 113.87° 90.00°	Depositor
Resolution (Å)	42.95 – 3.04 47.19 – 3.00	Depositor EDS
% Data completeness (in resolution range)	96.2 (42.95-3.04) 95.3 (47.19-3.00)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.01 (at 3.01Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, $R_{free}$	0.257 , 0.276 0.257 , 0.276	Depositor DCC
$R_{free}$ test set	4306 reflections (4.91%)	DCC
Wilson B-factor (Å <sup>2</sup> )	67.8	Xtriage
Anisotropy	0.315	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 43.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.116 for l,-k,h	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	27929	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	88.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 34.10 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 7.1980e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<sup>1</sup> Intensities estimated from amplitudes.

<sup>2</sup> Theoretical 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	0.39	0/3713	0.59	0/5015
1	B	0.39	0/3713	0.61	0/5015
1	C	0.39	0/3517	0.59	0/4748
1	D	0.38	0/3506	0.58	0/4730
1	E	0.39	0/3603	0.58	0/4863
1	F	0.39	0/3674	0.59	0/4960
1	G	0.39	0/3458	0.59	0/4667
1	H	0.38	0/3418	0.58	0/4613
All	All	0.39	0/28602	0.59	0/38611

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 [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	3624	0	3603	24	0
1	B	3624	0	3603	28	0
1	C	3433	0	3406	29	0
1	D	3423	0	3409	23	0
1	E	3516	0	3496	21	0
1	F	3586	0	3565	28	0
1	G	3375	0	3354	12	0
1	H	3336	0	3311	13	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	5	0	0	0	0
2	B	4	0	0	0	0
2	D	1	0	0	0	0
2	E	1	0	0	0	0
2	F	1	0	0	0	0
All	All	27929	0	27747	165	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 3.

The worst 5 of 165 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:C:41:VAL:HG21	1:C:67:ILE:HD11	1.42	1.01
1:H:252:LEU:O	1:H:256:ILE:HD13	1.75	0.87
1:F:252:LEU:O	1:F:256:ILE:HD13	1.76	0.86
1:B:147:ASP:HB3	1:B:248:ARG:HH12	1.42	0.84
1:F:207:LYS:HA	1:F:210:LEU:HD12	1.61	0.83

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	431/497 (87%)	399 (93%)	23 (5%)	9 (2%)	9	37
1	B	431/497 (87%)	400 (93%)	20 (5%)	11 (3%)	7	30
1	C	404/497 (81%)	375 (93%)	21 (5%)	8 (2%)	9	38
1	D	401/497 (81%)	378 (94%)	20 (5%)	3 (1%)	26	67
1	E	413/497 (83%)	388 (94%)	18 (4%)	7 (2%)	11	43

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	F	424/497 (85%)	393 (93%)	22 (5%)	9 (2%)	9	37
1	G	395/497 (80%)	372 (94%)	19 (5%)	4 (1%)	19	58
1	H	387/497 (78%)	362 (94%)	21 (5%)	4 (1%)	19	58
All	All	3286/3976 (83%)	3067 (93%)	164 (5%)	55 (2%)	11	43

5 of 55 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	136	ILE
1	B	106	GLU
1	B	136	ILE
1	C	106	GLU
1	C	136	ILE

### 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	398/446 (89%)	371 (93%)	27 (7%)	20	54
1	B	398/446 (89%)	368 (92%)	30 (8%)	17	49
1	C	377/446 (84%)	349 (93%)	28 (7%)	17	50
1	D	376/446 (84%)	345 (92%)	31 (8%)	14	45
1	E	385/446 (86%)	358 (93%)	27 (7%)	19	53
1	F	393/446 (88%)	363 (92%)	30 (8%)	16	49
1	G	370/446 (83%)	347 (94%)	23 (6%)	23	59
1	H	366/446 (82%)	341 (93%)	25 (7%)	20	54
All	All	3063/3568 (86%)	2842 (93%)	221 (7%)	18	51

5 of 221 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	262	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	E	188	ARG
1	H	219	GLN
1	D	325	ASP
1	D	482	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 23 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	238	GLN
1	D	433	ASN
1	G	451	HIS
1	D	322	ASN
1	E	238	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 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 [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<sup>th</sup> 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 <sup>2</sup> )	Q < 0.9
1	A	441/497 (88%)	0.09	6 (1%)	78 51	33, 64, 120, 158	0
1	B	441/497 (88%)	0.05	5 (1%)	82 58	32, 64, 116, 134	0
1	C	418/497 (84%)	0.43	29 (6%)	20 7	52, 104, 152, 167	0
1	D	417/497 (83%)	0.26	17 (4%)	41 17	44, 86, 149, 165	0
1	E	427/497 (85%)	0.12	6 (1%)	78 51	40, 77, 138, 183	0
1	F	436/497 (87%)	0.14	7 (1%)	74 47	34, 73, 127, 167	0
1	G	411/497 (82%)	0.34	19 (4%)	36 15	48, 99, 156, 181	0
1	H	405/497 (81%)	0.56	35 (8%)	13 4	56, 114, 160, 176	0
All	All	3396/3976 (85%)	0.24	124 (3%)	45 19	32, 82, 149, 183	0

The worst 5 of 124 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	496	THR	5.4
1	H	363	LEU	5.0
1	E	493	PHE	4.9
1	C	482	GLN	4.6
1	C	416	ALA	4.6

### 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 ⓘ

There are no carbohydrates 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.