



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 24, 2020 – 01:18 am BST

PDB ID : 6FEY  
Title : Crystal structure of Drosophila neural ectodermal development factor Imp-L2 with Drosophila DILP5 insulin  
Authors : Brzozowski, A.M.; Kulahin, N.; Kristensen, O.; Schluckebier, G.; Meyts, P.D.  
Deposited on : 2018-01-03  
Resolution : 3.48 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

<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  
Xtrriage (Phenix) : 1.13  
EDS : 2.11  
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) : 2.11

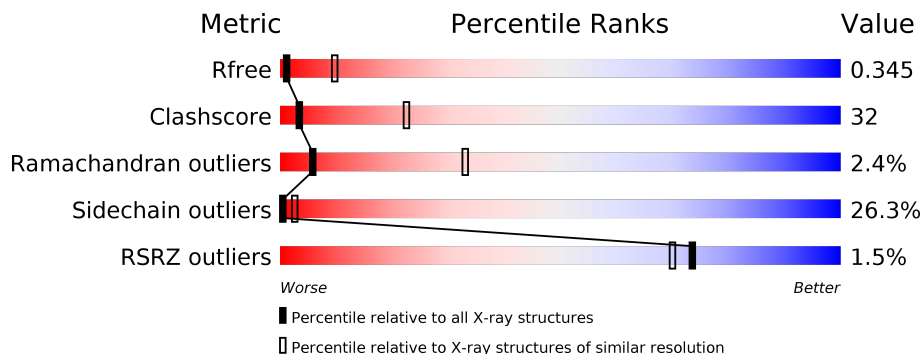
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## X-RAY DIFFRACTION

The reported resolution of this entry is 3.48 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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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	242	
1	B	242	
1	C	242	
1	D	242	
2	E	25	
2	G	25	

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	I	25	<p>52% 24% 20%</p>
2	K	25	<p>52% 32% 8%</p>
3	F	28	<p>43% 21% 32%</p>
3	H	28	<p>29% 25% 14% 32%</p>
3	J	28	<p>32% 7% 11% 50%</p>
3	L	28	<p>50% 7% 7% 36%</p>

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 5823 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 Neural/ectodermal development factor IMP-L2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	196	1477	932	262	275	8	0	0	0
1	B	207	1521	952	269	292	8	0	0	0
1	D	160	1084	675	198	203	8	0	0	0
1	C	124	821	508	149	156	8	0	0	0

- Molecule 2 is a protein called Probable insulin-like peptide 5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	E	21	134	78	24	28	4	0	0	0
2	G	20	116	70	20	22	4	0	0	0
2	I	20	110	63	23	20	4	0	0	0
2	K	23	147	85	32	26	4	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	12	ASN	LYS	engineered mutation	UNP Q7KUD5
G	12	ASN	LYS	engineered mutation	UNP Q7KUD5
I	12	ASN	LYS	engineered mutation	UNP Q7KUD5
K	12	ASN	LYS	engineered mutation	UNP Q7KUD5

- Molecule 3 is a protein called Probable insulin-like peptide 5.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace	
3	F	19	Total	C	N	O	S	0	0	0
			112	68	20	20	4			
3	H	19	Total	C	N	O	S	0	0	0
			121	77	20	20	4			
3	J	14	Total	C	N	O	S	0	0	0
			74	42	14	14	4			
3	L	18	Total	C	N	O	S	0	0	0
			102	61	19	18	4			

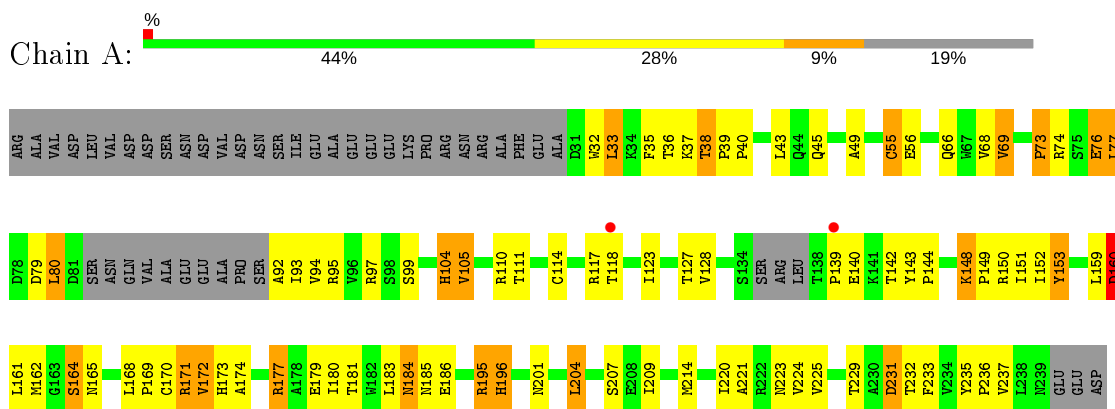
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	3	Total	O	0	0
			3	3		
4	C	1	Total	O	0	0
			1	1		

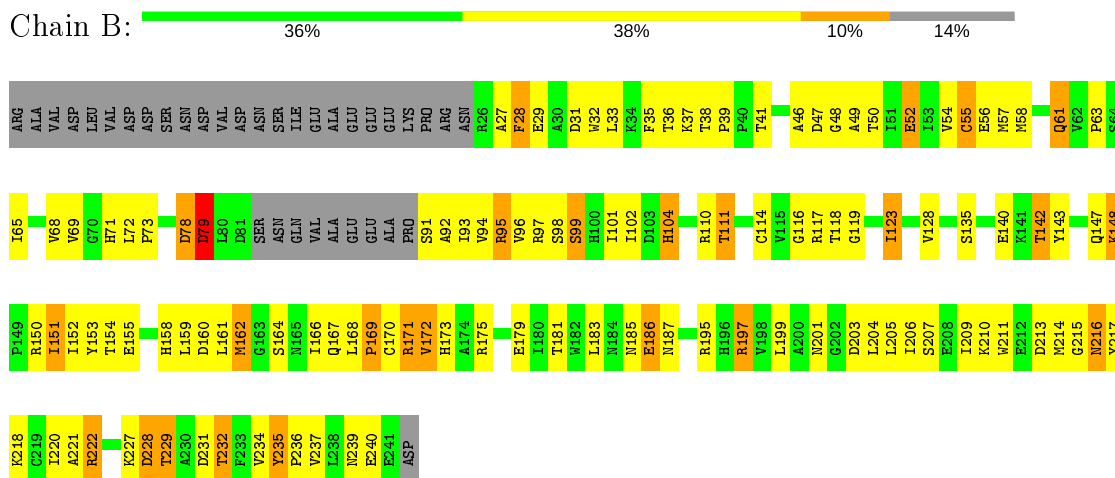
### 3 Residue-property plots [i](#)

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.

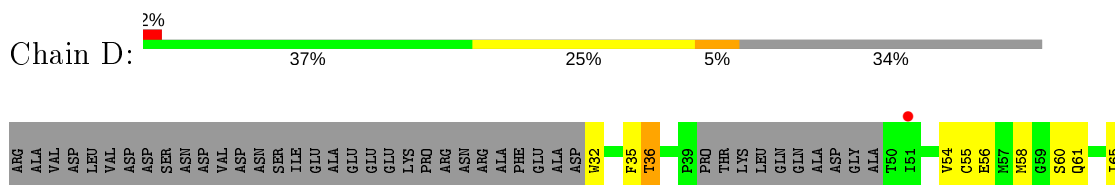
- Molecule 1: Neural/ectodermal development factor IMP-L2

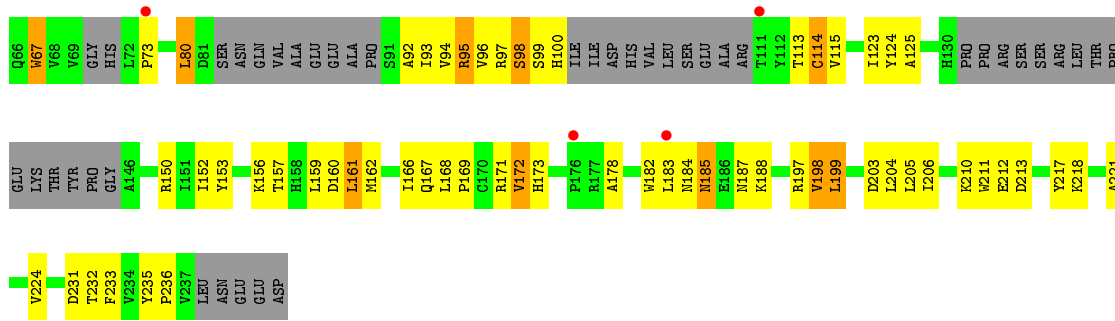


- Molecule 1: Neural/ectodermal development factor IMP-L2

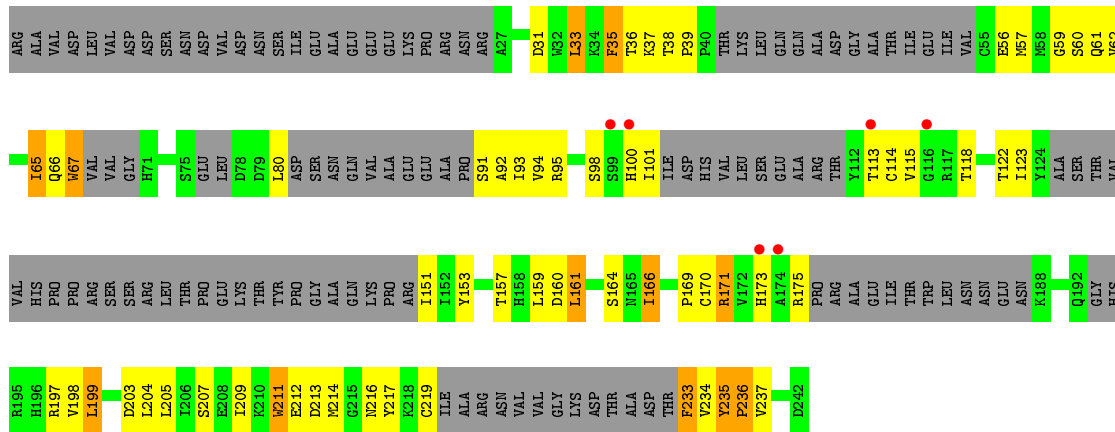


- Molecule 1: Neural/ectodermal development factor IMP-L2

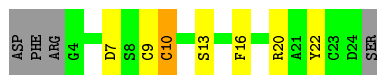




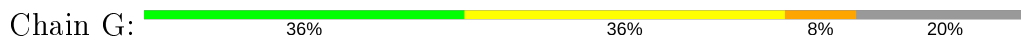
• Molecule 1: Neural/ectodermal development factor IMP-L2



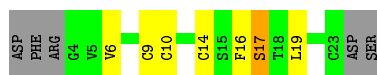
• Molecule 2: Probable insulin-like peptide 5



• Molecule 2: Probable insulin-like peptide 5

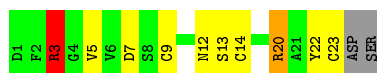


• Molecule 2: Probable insulin-like peptide 5



• Molecule 2: Probable insulin-like peptide 5

Chain K:  52% 32% 8%




- Molecule 3: Probable insulin-like peptide 5

Chain F:  43% 21% 32%



- Molecule 3: Probable insulin-like peptide 5

Chain H:  29% 25% 14% 32%



- Molecule 3: Probable insulin-like peptide 5

Chain J:  32% 7% 11% 50%



- Molecule 3: Probable insulin-like peptide 5

Chain L:  50% 7% 7% 36%





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	150.83Å 150.83Å 125.32Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.60 – 3.48 29.58 – 3.48	Depositor EDS
% Data completeness (in resolution range)	98.8 (29.60-3.48) 99.0 (29.58-3.48)	Depositor EDS
$R_{merge}$	0.17	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.46 (at 3.47Å)	Xtrriage
Refinement program	REFMAC 5.8.0189	Depositor
R, $R_{free}$	0.261 , 0.345 0.261 , 0.345	Depositor DCC
$R_{free}$ test set	971 reflections (5.16%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	115.5	Xtrriage
Anisotropy	0.145	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 120.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.41$ , $\langle L^2 \rangle = 0.24$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.88	EDS
Total number of atoms	5823	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	104.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.37% of the height of the origin peak. No significant pseudotranslation is detected.*

<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	0.57	0/1509	0.84	2/2059 (0.1%)
1	B	0.62	0/1555	0.93	2/2118 (0.1%)
1	C	0.51	0/832	0.75	0/1116
1	D	0.57	0/1103	0.83	0/1499
2	E	0.66	0/135	0.80	0/179
2	G	0.59	0/117	0.60	0/155
2	I	0.71	0/110	0.85	0/142
2	K	0.53	0/147	1.06	1/193 (0.5%)
3	F	0.56	0/113	0.75	0/148
3	H	0.66	0/123	1.05	0/166
3	J	0.61	0/73	0.80	0/93
3	L	0.53	0/102	0.84	0/133
All	All	0.58	0/5919	0.86	5/8001 (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
2	K	0	1
3	F	0	1
All	All	0	2

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	160	ASP	CB-CG-OD2	7.47	125.02	118.30
1	B	228	ASP	CB-CG-OD1	-7.44	111.60	118.30
1	B	228	ASP	CB-CG-OD2	5.61	123.35	118.30
2	K	3	ARG	NE-CZ-NH1	5.14	122.87	120.30
1	A	160	ASP	CB-CG-OD1	-5.00	113.80	118.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	F	5	ALA	Peptide
2	K	13	SER	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	1477	0	1391	79	0
1	B	1521	0	1408	106	0
1	C	821	0	634	60	0
1	D	1084	0	897	64	0
2	E	134	0	94	3	0
2	G	116	0	74	16	0
2	I	110	0	71	5	0
2	K	147	0	120	7	0
3	F	112	0	90	8	0
3	H	121	0	108	21	0
3	J	74	0	51	6	0
3	L	102	0	81	1	0
4	A	3	0	0	0	0
4	C	1	0	0	0	0
All	All	5823	0	5019	351	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 32.

The worst 5 of 351 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:B:151:ILE:HD13	1:B:170:CYS:SG	1.82	1.19
2:G:9:CYS:O	2:G:12:ASN:O	1.59	1.19
1:A:153:TYR:HB3	1:A:171:ARG:HG3	1.18	1.15
1:B:78:ASP:HA	1:B:79:ASP:HB2	1.21	1.10
1:B:168:LEU:HD22	1:B:232:THR:HG21	1.27	1.09

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	190/242 (78%)	177 (93%)	9 (5%)	4 (2%)	7	35
1	B	203/242 (84%)	186 (92%)	11 (5%)	6 (3%)	4	28
1	C	104/242 (43%)	83 (80%)	17 (16%)	4 (4%)	3	24
1	D	148/242 (61%)	134 (90%)	10 (7%)	4 (3%)	5	31
2	E	19/25 (76%)	17 (90%)	2 (10%)	0	100	100
2	G	18/25 (72%)	18 (100%)	0	0	100	100
2	I	18/25 (72%)	14 (78%)	3 (17%)	1 (6%)	2	15
2	K	21/25 (84%)	18 (86%)	3 (14%)	0	100	100
3	F	17/28 (61%)	16 (94%)	1 (6%)	0	100	100
3	H	17/28 (61%)	15 (88%)	2 (12%)	0	100	100
3	J	12/28 (43%)	12 (100%)	0	0	100	100
3	L	16/28 (57%)	15 (94%)	1 (6%)	0	100	100
All	All	783/1180 (66%)	705 (90%)	59 (8%)	19 (2%)	6	33

5 of 19 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	27	ALA
1	B	29	GLU
1	B	79	ASP
1	D	73	PRO
1	C	212	GLU

### 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	150/214 (70%)	115 (77%)	35 (23%)	1	3
1	B	156/214 (73%)	114 (73%)	42 (27%)	0	2
1	C	68/214 (32%)	44 (65%)	24 (35%)	0	1
1	D	92/214 (43%)	77 (84%)	15 (16%)	2	12
2	E	12/23 (52%)	8 (67%)	4 (33%)	0	1
2	G	9/23 (39%)	7 (78%)	2 (22%)	1	4
2	I	8/23 (35%)	7 (88%)	1 (12%)	4	21
2	K	13/23 (56%)	9 (69%)	4 (31%)	0	2
3	F	9/22 (41%)	7 (78%)	2 (22%)	1	4
3	H	11/22 (50%)	6 (54%)	5 (46%)	0	0
3	J	5/22 (23%)	1 (20%)	4 (80%)	0	0
3	L	7/22 (32%)	3 (43%)	4 (57%)	0	0
All	All	540/1036 (52%)	398 (74%)	142 (26%)	0	2

5 of 142 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	172	VAL
1	D	58	MET
3	J	10	LEU
1	B	181	THR
1	B	216	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	216	ASN
3	H	20	ASN
1	D	185	ASN
1	B	187	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	C	100	HIS

### 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 [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, 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	196/242 (80%)	-0.23	2 (1%) 82 78	62, 97, 129, 143	0
1	B	207/242 (85%)	-0.30	0 100 100	51, 85, 114, 132	0
1	C	124/242 (51%)	-0.08	6 (4%) 30 29	82, 116, 145, 165	0
1	D	160/242 (66%)	-0.03	5 (3%) 49 45	75, 114, 148, 176	0
2	E	21/25 (84%)	-0.19	0 100 100	87, 114, 177, 195	0
2	G	20/25 (80%)	-0.74	0 100 100	94, 116, 147, 158	0
2	I	20/25 (80%)	-0.71	0 100 100	122, 140, 157, 163	0
2	K	23/25 (92%)	-0.69	0 100 100	92, 115, 130, 138	0
3	F	19/28 (67%)	-0.65	0 100 100	103, 123, 145, 149	0
3	H	19/28 (67%)	-0.26	0 100 100	81, 118, 138, 152	0
3	J	14/28 (50%)	-0.69	0 100 100	114, 125, 153, 163	0
3	L	18/28 (64%)	-0.58	0 100 100	83, 108, 128, 130	0
All	All	841/1180 (71%)	-0.25	13 (1%) 73 70	51, 105, 145, 195	0

The worst 5 of 13 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	99	SER	4.8
1	C	173	HIS	3.1
1	D	183	LEU	3.0
1	C	174	ALA	3.0
1	A	139	PRO	2.6

### 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

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.