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)

Apr 21, 2024 – 08:52 am BST

PDB ID	:	4UF6
Title	:	UCH-L5 in complex with ubiquitin-propargyl bound to an activating fragment of INO80G
Authors	:	Sahtoe, D.D.; Van Dijk, W.J.; El Oualid, F.; Ekkebus, R.; Ovaa, H.; Sixma, T.K.
Deposited on	:	2014-12-23
Resolution	:	3.69 Å(reported)

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

|		
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36.2
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)	:	2.36.2

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 3.69 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.
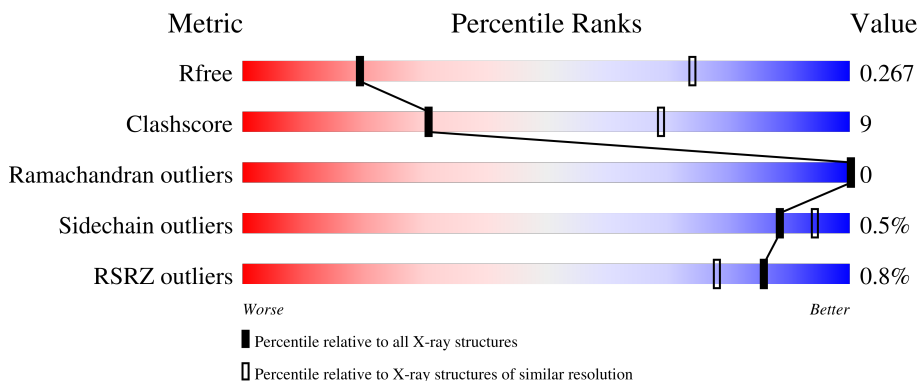
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1049 (3.88-3.52)
Clashscore	141614	1027 (3.86-3.54)
Ramachandran outliers	138981	1069 (3.88-3.52)
Sidechain outliers	138945	1065 (3.88-3.52)
RSRZ outliers	127900	1578 (3.90-3.50)

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	331	66% 22% 12%
1	D	331	69% 20% 11%
1	G	331	66% 21% 13%
1	J	331	68% 21% 11%
2	B	76	80% 18% ·

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	E	76	87% 13%
2	H	76	91% 9%
2	K	76	83% 17%
3	C	65	2% 52% 25% 23%
3	F	65	3% 58% 18% 23%
3	I	65	2% 55% 20% 25%
3	L	65	2% 52% 23% 25%

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 13486 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 UBIQUITIN CARBOXYL-TERMINAL HYDROLASE ISOZYME L5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	292	Total	C	N	O	S	0	0	0
			2360	1503	405	440	12			
1	D	295	Total	C	N	O	S	0	0	0
			2382	1515	409	446	12			
1	G	287	Total	C	N	O	S	0	0	0
			2313	1475	397	430	11			
1	J	294	Total	C	N	O	S	0	0	0
			2377	1514	407	444	12			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	expression tag	UNP Q9Y5K5
A	-1	PRO	-	expression tag	UNP Q9Y5K5
A	0	GLY	-	expression tag	UNP Q9Y5K5
D	-2	GLY	-	expression tag	UNP Q9Y5K5
D	-1	PRO	-	expression tag	UNP Q9Y5K5
D	0	GLY	-	expression tag	UNP Q9Y5K5
G	-2	GLY	-	expression tag	UNP Q9Y5K5
G	-1	PRO	-	expression tag	UNP Q9Y5K5
G	0	GLY	-	expression tag	UNP Q9Y5K5
J	-2	GLY	-	expression tag	UNP Q9Y5K5
J	-1	PRO	-	expression tag	UNP Q9Y5K5
J	0	GLY	-	expression tag	UNP Q9Y5K5

- Molecule 2 is a protein called POLYUBIQUITIN-B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	76	Total	C	N	O	S	0	0	0
			601	379	105	116	1			
2	E	76	Total	C	N	O	S	0	0	0
			601	379	105	116	1			

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	H	76	Total 601	C 379	N 105	O 116	S 1	0	0	0
2	K	76	Total 601	C 379	N 105	O 116	S 1	0	0	0

- Molecule 3 is a protein called NUCLEAR FACTOR RELATED TO KAPPA-B-BINDING PROTEIN.

| Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	50	Total 417	C 267	N 64	O 86	0	0	0
3	F	50	Total 417	C 267	N 64	O 86	0	0	0
3	I	49	Total 408	C 262	N 63	O 83	0	0	0
3	L	49	Total 408	C 262	N 63	O 83	0	0	0

There are 12 discrepancies between the modelled and reference sequences:

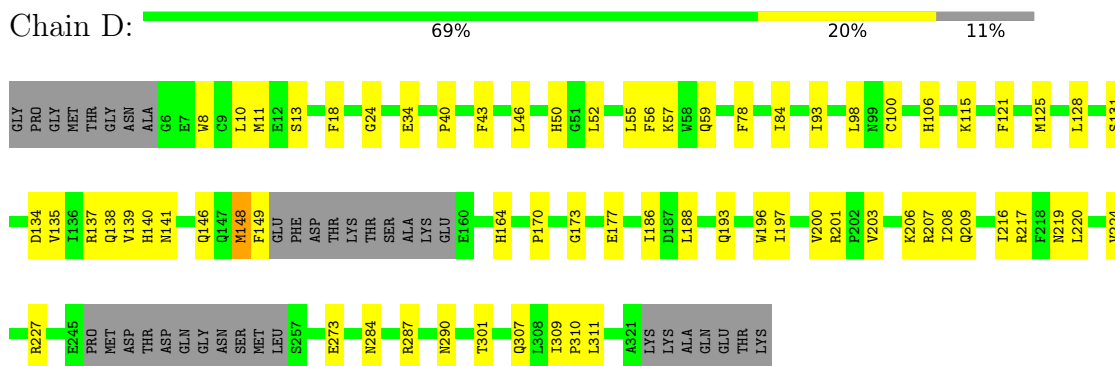| Chain	Residue	Modelled	Actual	Comment	Reference
C	37	GLY	-	expression tag	UNP Q6P4R8
C	38	PRO	-	expression tag	UNP Q6P4R8
C	39	GLY	-	expression tag	UNP Q6P4R8
F	37	GLY	-	expression tag	UNP Q6P4R8
F	38	PRO	-	expression tag	UNP Q6P4R8
F	39	GLY	-	expression tag	UNP Q6P4R8
I	37	GLY	-	expression tag	UNP Q6P4R8
I	38	PRO	-	expression tag	UNP Q6P4R8
I	39	GLY	-	expression tag	UNP Q6P4R8
L	37	GLY	-	expression tag	UNP Q6P4R8
L	38	PRO	-	expression tag	UNP Q6P4R8
L	39	GLY	-	expression tag	UNP Q6P4R8

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.
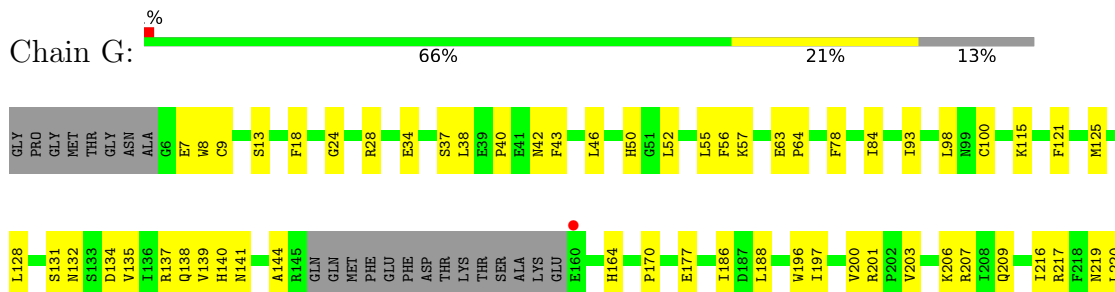
• Molecule 1: UBIQUITIN CARBOXYL-TERMINAL HYDROLASE ISOZYME L5
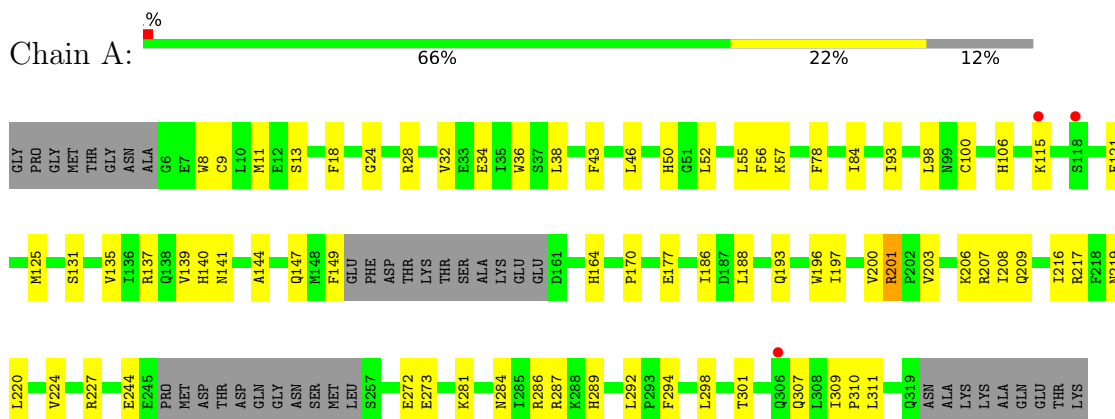
Chain A:

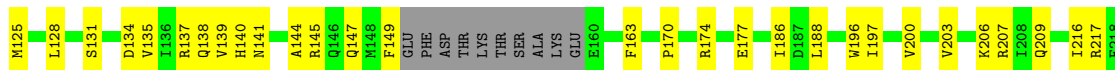• Molecule 1: UBIQUITIN CARBOXYL-TERMINAL HYDROLASE ISOZYME L5

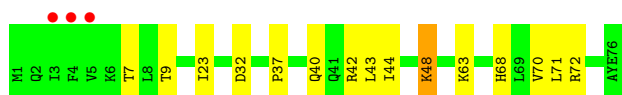Chain D:

• Molecule 1: UBIQUITIN CARBOXYL-TERMINAL HYDROLASE ISOZYME L5

Chain G:

V224 R227 A243 GLU GLU PRO MET ASP THR ASP GLN GLY ASN SER MET L256 E272 E273 N284 R287 K288 H289 M296 I309 P310 L311 V312 K318 GLN ASN ALA LYS LYS ALA GLN THR LYS

• **Molecule 1: UBIQUITIN CARBOXYL-TERMINAL HYDROLASE ISOZYME L5**

Chain J:

| 68% | 21% | 11% |

GLY PRO GLY MET THR GLY ASN ALA G6 E7 W8 C9 L10 M11 E12 S13 F18 G24 R28 E34 S37 N42 F43 L46 H50 G51 L52 F56 K57 F78 I84 I93 L98 N99 C100 Q103 D104 V105 H106 E113 F114 K115 F121

M125 L128 S131 D134 V135 I136 R137 Q138 V139 H140 N141 A144 R145 Q146 Q147 M148 F149 GLU PHE ASP THR LYS THR SER ALA LYS GLU E160 F163 P170 R174 E177 I186 D187 L188 W196 I197 V200 V203 K206 R207 I208 Q209 I216 R217 F218

N219 L220 V224 R227 E245 PRO MET ASP THR ASP GLN GLY ASN SER MET L256 E273 N284 R287 E297 L298 T301 I309 P310 L311 K318 Q319 ASN ALA LYS LYS ALA GLN GLU THR LYS

• **Molecule 2: POLYUBIQUITIN-B**

Chain B:

4%
| 80% | 18% | . |

M1 Q2 I3 F4 V5 K6 T7 L8 T9 I23 D32 P37 Q40 Q41 R42 L43 I44 K48 K63 H68 L69 V70 L71 R72 AYE76

• **Molecule 2: POLYUBIQUITIN-B**

Chain E:

| 87% | 13% |

M1 T7 L8 T9 I23 L43 I44 K63 H68 L69 V70 L71 R74 G75 AYE76

• **Molecule 2: POLYUBIQUITIN-B**

Chain H:

| 91% | 9% |

M1 I23 L43 I44 K48 H68 L69 V70 L71 AYE76

• **Molecule 2: POLYUBIQUITIN-B**

Chain K:

| 83% | 17% |

M1 T7 L8 T9 I23 P37 Q40 Q41 R42 L43 K48 N60 V70 L71 R72 L73 AYE76

• **Molecule 3: NUCLEAR FACTOR RELATED TO KAPPA-B-BINDING PROTEIN**

Chain C:

2%
| 52% | 25% | 23% |

GLY PRO GLY GLU ASP LEU LEU E44 L55 S56 T57 W58 Q59 E60 V61 L62 S63 D64 S65 Q66 H69 L70 Q71 P75 Q76 S81 A82 Q85 L88 F93 SER GLY GLU ASN PHE ARG PHE GLY

• Molecule 3: NUCLEAR FACTOR RELATED TO KAPPA-B-BINDING PROTEIN

Chain F:

3%
58%   18%   23%

GLY PRO GLY GLU ASP LEU LEU E44 L55 S56 T57 W58 Q59 E60 V61 L62 S63 D64 S65 Q66 Q71 Q72 S81 Q85 F93 SER GLY GLU ASN PHE ARG PHE GLY

• Molecule 3: NUCLEAR FACTOR RELATED TO KAPPA-B-BINDING PROTEIN

Chain I:

2%
55%   20%   25%

GLY PRO GLY GLU ASP LEU LEU GLU D45 P46 E47 I48 L55 S56 T57 W58 Q59 S63 D64 S65 Q66 P75 D80 S81 A82 E83 Q84 Q85 F93 SER GLY GLU ASN PHE ARG PHE GLY

• Molecule 3: NUCLEAR FACTOR RELATED TO KAPPA-B-BINDING PROTEIN

Chain L:

2%
52%   23%   25%

GLY PRO GLY GLU ASP LEU LEU GLU D45 P46 V52 L55 S56 T57 W58 Q59 S63 D64 S65 Q66 H69 L70 Q71 Q72 F73 E79 D80 S81 Q84 Q85 F93 SER GLY GLU ASN PHE ARG PHE GLY

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	152.08Å 137.79Å 98.92Å 90.00° 102.54° 90.00°	Depositor
Resolution (Å)	100.99 – 3.69 47.74 – 3.69	Depositor EDS
% Data completeness (in resolution range)	97.9 (100.99-3.69) 98.0 (47.74-3.69)	Depositor EDS
R_{merge}	0.20	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.43 (at 3.67Å)	Xtriage
Refinement program	REFMAC 5.8.0073	Depositor
R, R_{free}	0.238 , 0.269 0.240 , 0.267	Depositor DCC
R_{free} test set	1059 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	82.8	Xtriage
Anisotropy	0.568	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 29.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.46$, $< L^2 > = 0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	13486	wwPDB-VP
Average B, all atoms (Å^2)	87.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.54% 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 ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: AYE

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	0.44	0/2404	0.64	2/3236 (0.1%)
1	D	0.44	0/2426	0.62	0/3266
1	G	0.41	0/2356	0.59	0/3173
1	J	0.46	0/2421	0.62	0/3259
2	B	0.34	0/603	0.57	0/811
2	E	0.39	0/603	0.61	0/811
2	H	0.35	0/603	0.57	0/811
2	K	0.35	0/603	0.57	0/811
3	C	0.43	0/427	0.63	0/581
3	F	0.45	0/427	0.68	0/581
3	I	0.43	0/418	0.63	0/569
3	L	0.46	0/418	0.63	0/569
All	All	0.43	0/13709	0.61	2/18478 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	201	ARG	NE-CZ-NH1	7.37	123.99	120.30
1	A	201	ARG	NE-CZ-NH2	-6.35	117.13	120.30

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	2360	0	2348	62	1
1	D	2382	0	2365	54	0
1	G	2313	0	2311	53	0
1	J	2377	0	2365	50	1
2	B	601	0	630	12	2
2	E	601	0	630	10	1
2	H	601	0	630	6	0
2	K	601	0	630	8	1
3	C	417	0	383	17	0
3	F	417	0	383	11	0
3	I	408	0	377	7	1
3	L	408	0	377	11	1
All	All	13486	0	13429	247	4

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 9.

The worst 5 of 247 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:G:40:PRO:HB3	1:G:201:ARG:HH22	1.20	1.02
1:D:135:VAL:O	1:D:139:VAL:HG23	1.63	0.99
1:A:135:VAL:O	1:A:139:VAL:HG23	1.62	0.98
1:J:135:VAL:O	1:J:139:VAL:HG23	1.62	0.97
1:G:135:VAL:O	1:G:139:VAL:HG23	1.62	0.97

All (4) 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:A:244:GLU:O	1:J:174:ARG:NH1[2_755]	1.99	0.21
2:B:32:ASP:O	2:K:60:ASN:ND2[4_655]	2.05	0.15
2:E:63:LYS:NZ	3:I:82:ALA:O[4_656]	2.14	0.06
2:B:63:LYS:NZ	3:L:79:GLU:OE1[4_655]	2.18	0.02

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	286/331 (86%)	273 (96%)	13 (4%)	0	100	100
1	D	289/331 (87%)	276 (96%)	13 (4%)	0	100	100
1	G	281/331 (85%)	269 (96%)	12 (4%)	0	100	100
1	J	288/331 (87%)	276 (96%)	12 (4%)	0	100	100
2	B	73/76 (96%)	71 (97%)	2 (3%)	0	100	100
2	E	73/76 (96%)	70 (96%)	3 (4%)	0	100	100
2	H	73/76 (96%)	71 (97%)	2 (3%)	0	100	100
2	K	73/76 (96%)	71 (97%)	2 (3%)	0	100	100
3	C	48/65 (74%)	41 (85%)	7 (15%)	0	100	100
3	F	48/65 (74%)	41 (85%)	7 (15%)	0	100	100
3	I	47/65 (72%)	39 (83%)	8 (17%)	0	100	100
3	L	47/65 (72%)	39 (83%)	8 (17%)	0	100	100
All	All	1626/1888 (86%)	1537 (94%)	89 (6%)	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	257/288 (89%)	257 (100%)	0	100	100
1	D	259/288 (90%)	257 (99%)	2 (1%)	81	89

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	G	252/288 (88%)	252 (100%)	0	100	100
1	J	259/288 (90%)	258 (100%)	1 (0%)	91	95
2	B	68/68 (100%)	67 (98%)	1 (2%)	65	81
2	E	68/68 (100%)	68 (100%)	0	100	100
2	H	68/68 (100%)	68 (100%)	0	100	100
2	K	68/68 (100%)	67 (98%)	1 (2%)	65	81
3	C	48/59 (81%)	48 (100%)	0	100	100
3	F	48/59 (81%)	48 (100%)	0	100	100
3	I	47/59 (80%)	46 (98%)	1 (2%)	53	74
3	L	47/59 (80%)	46 (98%)	1 (2%)	53	74
All	All	1489/1660 (90%)	1482 (100%)	7 (0%)	88	94

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	I	46	PRO
1	J	145	ARG
3	L	46	PRO
2	K	48	LYS
1	D	148	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	106	HIS
1	D	42	ASN
1	D	233	GLN

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 ⓘ

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(Å²)	Q<0.9
1	A	292/331 (88%)	-0.09	3 (1%)	82	73	53, 79, 143, 181	0
1	D	295/331 (89%)	-0.13	0	100	100	48, 66, 125, 169	0
1	G	287/331 (86%)	-0.12	3 (1%)	82	73	50, 77, 146, 185	0
1	J	294/331 (88%)	-0.17	0	100	100	42, 63, 138, 170	0
2	B	75/76 (98%)	0.28	3 (4%)	38	28	107, 140, 187, 199	0
2	E	75/76 (98%)	-0.03	0	100	100	69, 86, 103, 106	0
2	H	75/76 (98%)	0.07	0	100	100	76, 90, 100, 102	0
2	K	75/76 (98%)	0.03	0	100	100	58, 76, 92, 97	0
3	C	50/65 (76%)	0.08	1 (2%)	65	53	102, 114, 126, 132	0
3	F	50/65 (76%)	0.20	2 (4%)	38	28	99, 113, 127, 133	0
3	I	49/65 (75%)	0.23	1 (2%)	65	53	106, 124, 144, 154	0
3	L	49/65 (75%)	0.33	1 (2%)	65	53	91, 115, 149, 158	0
All	All	1666/1888 (88%)	-0.05	14 (0%)	86	78	42, 80, 146, 199	0

The worst 5 of 14 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	F	71	GLN	3.9
3	F	81	SER	3.8
1	G	243	ALA	3.8
3	C	71	GLN	3.0
3	L	71	GLN	2.7

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 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.