wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 02:19 PM GMT

PDB ID	:	3WTW
Title	:	Crystal structure of the complex comprised of ETS1(K167A), RUNX1, CBF-BETA, and the tcralpha gene enhancer DNA
Authors	:	Shiina, M.; Hamada, K.; Ogata, K.
Deposited on	:	2014-04-21
Resolution	:	2.90 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
http://wwpdb.org/validation/2016/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:

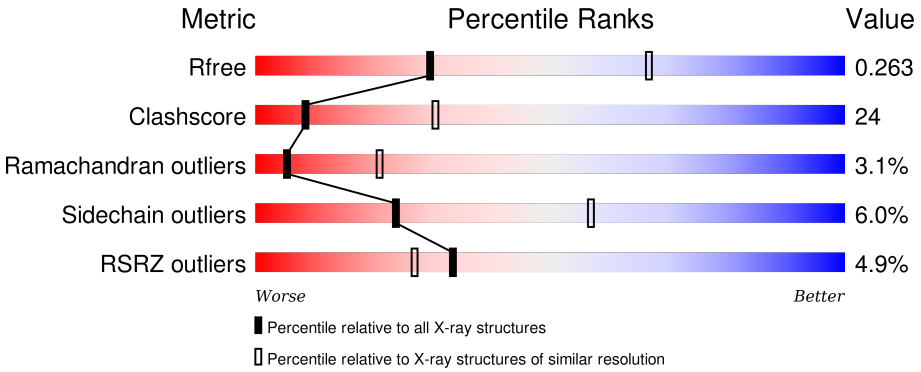
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 2.90 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.



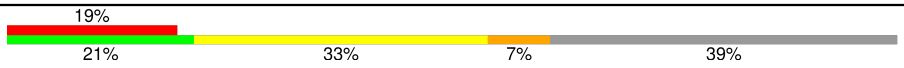
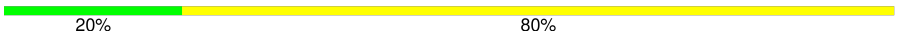
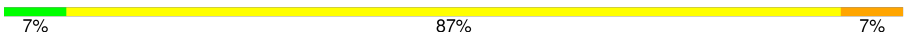
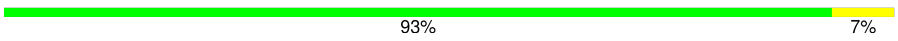
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	1451 (2.90-2.90)
Clashscore	102246	1668 (2.90-2.90)
Ramachandran outliers	100387	1630 (2.90-2.90)
Sidechain outliers	100360	1632 (2.90-2.90)
RSRZ outliers	91569	1456 (2.90-2.90)

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 >=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 <=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	204	37% 18% • 42%
1	F	204	39% 18% • 42%
2	B	142	% 51% 38% • 8%
2	G	142	% 49% 36% 6% 9%
3	C	166	% 45% 23% • 29%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	H	166	
4	D	15	
4	I	15	
5	E	15	
5	J	15	

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 7014 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 Runt-related transcription factor 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	118	Total	C	N	O	S	0	0	0
			910	571	168	167	4			
1	F	119	Total	C	N	O	S	0	0	0
			921	577	172	168	4			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	94	LYS	LEU	ENGINEERED MUTATION	UNP Q03347
A	167	ALA	LYS	ENGINEERED MUTATION	UNP Q03347
F	94	LYS	LEU	ENGINEERED MUTATION	UNP Q03347
F	167	ALA	LYS	ENGINEERED MUTATION	UNP Q03347

- Molecule 2 is a protein called Core-binding factor subunit beta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	130	Total	C	N	O	S	0	0	0
			1071	671	195	199	6			
2	G	129	Total	C	N	O	S	0	0	0
			1062	666	193	197	6			

- Molecule 3 is a protein called Protein C-ets-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	118	Total	C	N	O	S	0	0	0
			967	627	165	171	4			
3	H	101	Total	C	N	O	S	0	0	0
			853	553	147	149	4			

- Molecule 4 is a DNA chain called DNA (5'-D(*GP*AP*AP*GP*CP*CP*AP*CP*AP*TP*CP*CP*TP*CP*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	15	Total	C	N	O	P	0	0	0
			299	144	54	87	14			
4	I	15	Total	C	N	O	P	0	0	0
			299	144	54	87	14			

- Molecule 5 is a DNA chain called DNA (5'-D(*AP*GP*AP*GP*GP*AP*TP*GP*TP*GP*GP*CP*TP*TP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	E	15	Total	C	N	O	P	0	0	0
			310	148	59	89	14			
5	J	15	Total	C	N	O	P	0	0	0
			310	148	59	89	14			

- Molecule 6 is water.

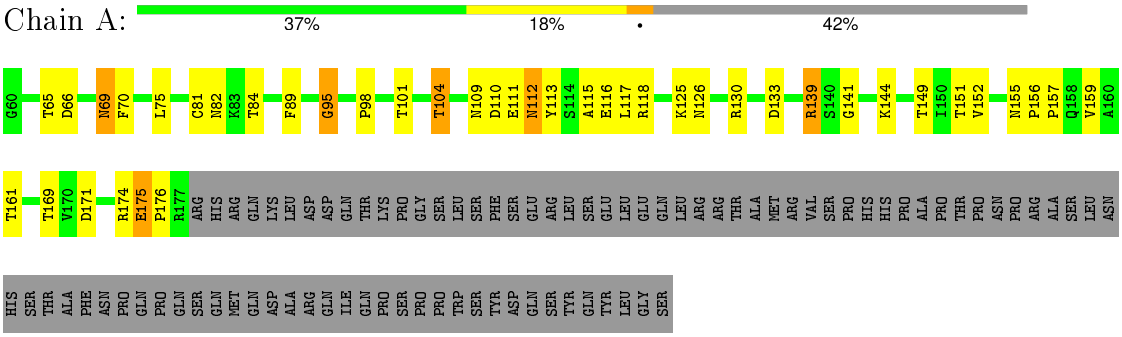
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	1	Total	O	0	0
			1	1		
6	B	1	Total	O	0	0
			1	1		
6	C	3	Total	O	0	0
			3	3		
6	F	5	Total	O	0	0
			5	5		
6	G	1	Total	O	0	0
			1	1		
6	E	1	Total	O	0	0
			1	1		

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 errors 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.

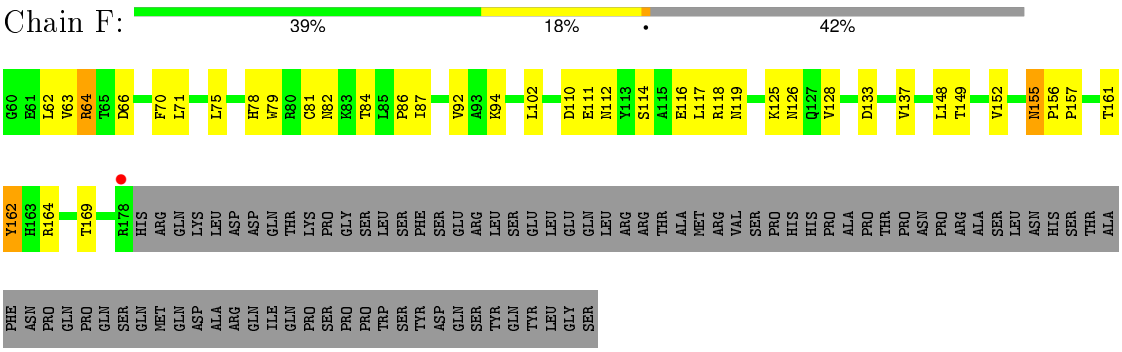
- Molecule 1: Runt-related transcription factor 1

Chain A:



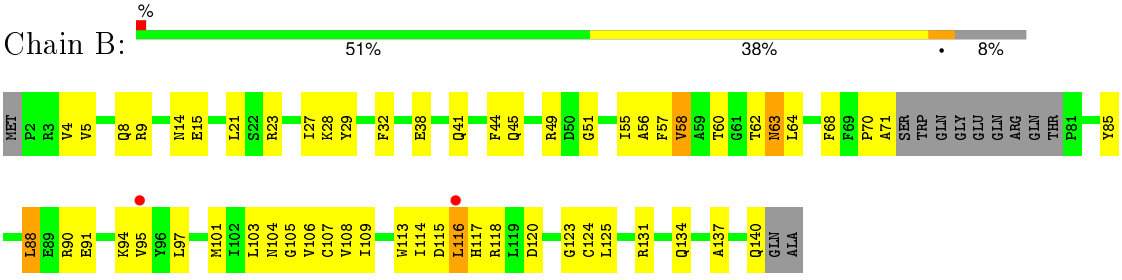
- Molecule 1: Runt-related transcription factor 1

Chain F:



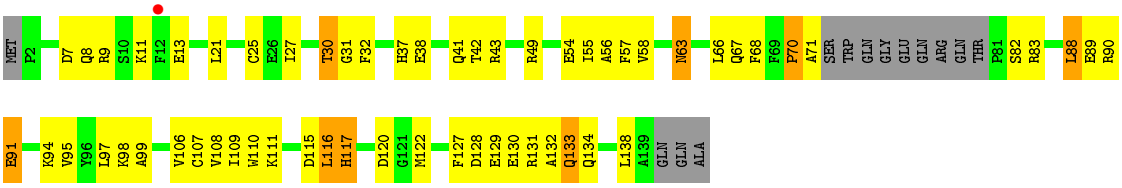
- Molecule 2: Core-binding factor subunit beta

Chain B:

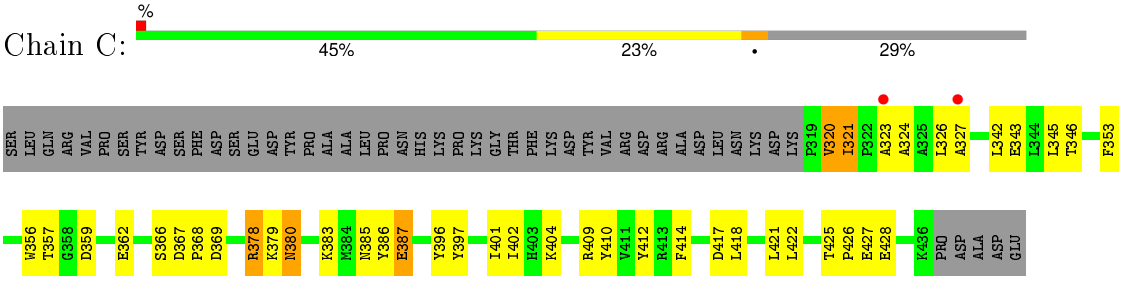


- Molecule 2: Core-binding factor subunit beta

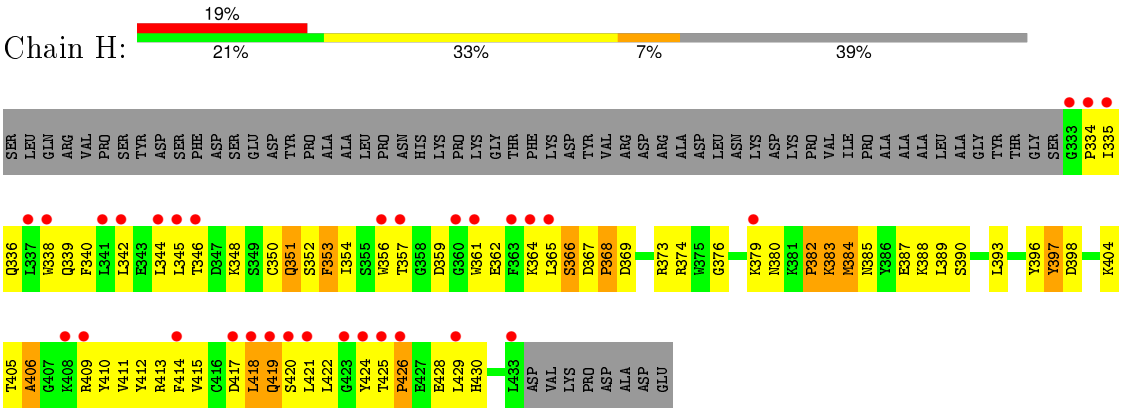
Chain G:



• Molecule 3: Protein C-ets-1

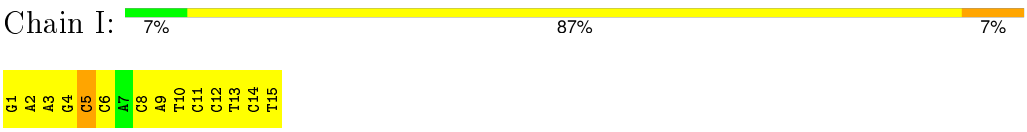


• Molecule 3: Protein C-ets-1



• Molecule 4: DNA (5'-D(*GP*AP*AP*GP*CP*CP*AP*CP*AP*TP*CP*CP*TP*CP*T)-3')

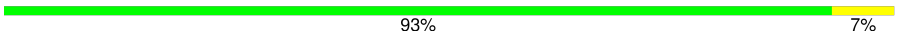
• Molecule 4: DNA (5'-D(*GP*AP*AP*GP*CP*CP*AP*CP*AP*TP*CP*CP*TP*CP*T)-3')



• Molecule 5: DNA (5'-D(*AP*GP*AP*GP*GP*AP*TP*GP*TP*GP*GP*CP*TP*TP*C)-3')

• Molecule 5: DNA (5'-D(*AP*GP*AP*GP*GP*AP*TP*GP*TP*GP*GP*CP*TP*TP*C)-3')

Chain J:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	78.50Å 100.61Å 194.40Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	42.36 – 2.90 42.36 – 2.90	Depositor EDS
% Data completeness (in resolution range)	97.9 (42.36-2.90) 98.0 (42.36-2.90)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.60 (at 2.90Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, R_{free}	0.219 , 0.264 0.218 , 0.263	Depositor DCC
R_{free} test set	3414 reflections (9.99%)	DCC
Wilson B-factor (Å^2)	77.0	Xtriage
Anisotropy	0.301	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 56.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Outliers	3 of 34189 reflections (0.009%)	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	7014	wwPDB-VP
Average B, all atoms (Å^2)	88.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.25% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.58	1/929 (0.1%)	0.71	0/1264
1	F	0.62	0/940	0.79	0/1278
2	B	0.44	0/1093	0.57	0/1466
2	G	0.50	0/1084	0.62	0/1454
3	C	0.57	0/994	0.67	0/1342
3	H	0.42	0/877	0.52	0/1181
4	D	0.69	0/334	0.90	0/512
4	I	0.73	0/334	0.92	0/512
5	E	0.71	0/348	0.81	0/537
5	J	0.72	0/348	0.84	0/537
All	All	0.57	1/7281 (0.0%)	0.70	0/10083

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
4	D	0	3
4	I	0	2
5	E	0	1
All	All	0	6

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	81	CYS	CB-SG	-5.17	1.73	1.81

There are no bond angle outliers.

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
4	D	4	DG	Sidechain
4	D	5	DC	Sidechain
4	D	6	DC	Sidechain
5	E	13	DT	Sidechain
4	I	4	DG	Sidechain

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	910	0	914	34	0
1	F	921	0	927	37	0
2	B	1071	0	1034	65	0
2	G	1062	0	1026	51	0
3	C	967	0	966	34	0
3	H	853	0	847	81	0
4	D	299	0	170	14	0
4	I	299	0	170	17	0
5	E	310	0	171	1	0
5	J	310	0	171	1	0
6	A	1	0	0	0	0
6	B	1	0	0	1	0
6	C	3	0	0	0	0
6	E	1	0	0	0	0
6	F	5	0	0	1	0
6	G	1	0	0	0	0
All	All	7014	0	6396	321	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 24.

The worst 5 of 321 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)
3:C:385:ASN:HD21	3:C:387:GLU:HB2	1.22	1.01
1:A:82:ASN:HD21	1:A:118:ARG:HH11	1.04	0.89
2:B:8:GLN:HE21	2:B:107:CYS:H	1.16	0.88

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:161:THR:H	2:B:104:ASN:HD21	1.20	0.87
1:A:109:ASN:HD21	1:A:111:GLU:HB2	1.39	0.86

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	116/204 (57%)	108 (93%)	6 (5%)	2 (2%)	11 38
1	F	117/204 (57%)	109 (93%)	7 (6%)	1 (1%)	21 57
2	B	126/142 (89%)	99 (79%)	24 (19%)	3 (2%)	7 29
2	G	125/142 (88%)	107 (86%)	13 (10%)	5 (4%)	4 15
3	C	116/166 (70%)	109 (94%)	6 (5%)	1 (1%)	21 57
3	H	99/166 (60%)	68 (69%)	21 (21%)	10 (10%)	1 2
All	All	699/1024 (68%)	600 (86%)	77 (11%)	22 (3%)	5 21

5 of 22 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	C	320	VAL
3	H	366	SER
3	H	382	PRO
2	B	15	GLU
2	B	116	LEU

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	99/178 (56%)	92 (93%)	7 (7%)	18	47
1	F	100/178 (56%)	94 (94%)	6 (6%)	24	57
2	B	113/123 (92%)	109 (96%)	4 (4%)	43	78
2	G	112/123 (91%)	104 (93%)	8 (7%)	18	47
3	C	102/145 (70%)	96 (94%)	6 (6%)	24	58
3	H	91/145 (63%)	85 (93%)	6 (7%)	21	51
All	All	617/892 (69%)	580 (94%)	37 (6%)	24	57

5 of 37 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	C	428	GLU
1	F	155	ASN
3	H	384	MET
1	F	64	ARG
1	F	112	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 24 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	134	GLN
1	F	82	ASN
3	H	339	GLN
3	C	380	ASN
3	C	385	ASN

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^{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	118/204 (57%)	0.05	0 100 100	57, 77, 94, 111	0
1	F	119/204 (58%)	0.01	1 (0%) 87 86	51, 66, 86, 132	0
2	B	130/142 (91%)	0.41	2 (1%) 76 74	80, 105, 131, 136	0
2	G	129/142 (90%)	0.22	1 (0%) 87 86	68, 89, 117, 122	0
3	C	118/166 (71%)	0.17	2 (1%) 73 70	52, 72, 104, 115	0
3	H	101/166 (60%)	1.52	32 (31%) 1 0	77, 131, 172, 180	0
4	D	15/15 (100%)	-0.23	0 100 100	59, 67, 79, 80	0
4	I	15/15 (100%)	-0.24	0 100 100	61, 72, 90, 92	0
5	E	15/15 (100%)	-0.33	0 100 100	55, 67, 81, 85	0
5	J	15/15 (100%)	-0.25	0 100 100	59, 69, 92, 98	0
All	All	775/1084 (71%)	0.32	38 (4%) 33 27	51, 82, 143, 180	0

The worst 5 of 38 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	H	360	GLY	7.8
3	H	345	LEU	6.1
3	H	424	TYR	5.1
1	F	178	ARG	5.0
3	H	333	GLY	4.8

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 carbohydrates 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.