wwPDB X-ray Structure Validation Summary Report (i)

May 24, 2020 – 12:33 am BST

PDB ID	:	6GCU
Title	:	MET receptor in complex with InlB internalin domain and DARPin A3A
Authors	:	Meyer, T.; Andres, F.; Iamele, L.; Gherardi, E.; Pluckthun, A.; Niemann, H.H.
Deposited on	:	2018-04-19
Resolution	:	6.00 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
https://www.wwpdb.org/validation/2017/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:

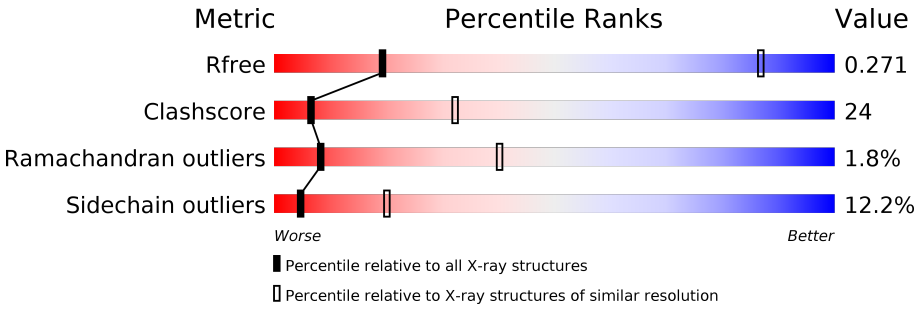
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 6.00 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	1000 (8.00-3.88)
Clashscore	141614	1049 (8.00-3.90)
Ramachandran outliers	138981	1016 (8.00-3.86)
Sidechain outliers	138945	1017 (8.00-3.82)

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 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%

Mol	Chain	Length	Quality of chain
1	A	727	50% 34% 8% • 7%
1	D	727	52% 32% 9% • 7%
2	B	289	51% 45% • •
2	E	289	53% 42% • •
3	C	173	80% 8% • 10%
3	F	173	81% 8% • 10%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 17448 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 Hepatocyte growth factor receptor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	675	Total	C	N	O	S	0	0	0
			5314	3370	904	1002	38			
1	D	675	Total	C	N	O	S	0	0	0
			5314	3370	904	1002	38			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	22	GLU	-	expression tag	UNP P08581
A	23	THR	-	expression tag	UNP P08581
A	24	ARG	-	expression tag	UNP P08581
A	742	LEU	-	expression tag	UNP P08581
A	743	HIS	-	expression tag	UNP P08581
A	744	HIS	-	expression tag	UNP P08581
A	745	HIS	-	expression tag	UNP P08581
A	746	HIS	-	expression tag	UNP P08581
A	747	HIS	-	expression tag	UNP P08581
A	748	HIS	-	expression tag	UNP P08581
D	22	GLU	-	expression tag	UNP P08581
D	23	THR	-	expression tag	UNP P08581
D	24	ARG	-	expression tag	UNP P08581
D	742	LEU	-	expression tag	UNP P08581
D	743	HIS	-	expression tag	UNP P08581
D	744	HIS	-	expression tag	UNP P08581
D	745	HIS	-	expression tag	UNP P08581
D	746	HIS	-	expression tag	UNP P08581
D	747	HIS	-	expression tag	UNP P08581
D	748	HIS	-	expression tag	UNP P08581

- Molecule 2 is a protein called Internalin B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	286	Total	C	N	O	S	0	0	0
			2252	1435	378	437	2			
2	E	286	Total	C	N	O	S	0	0	0
			2252	1435	378	437	2			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	33	GLY	-	expression tag	UNP P25147
B	34	ALA	-	expression tag	UNP P25147
B	35	MET	-	expression tag	UNP P25147
E	33	GLY	-	expression tag	UNP P25147
E	34	ALA	-	expression tag	UNP P25147
E	35	MET	-	expression tag	UNP P25147

- Molecule 3 is a protein called DARPin A3A.

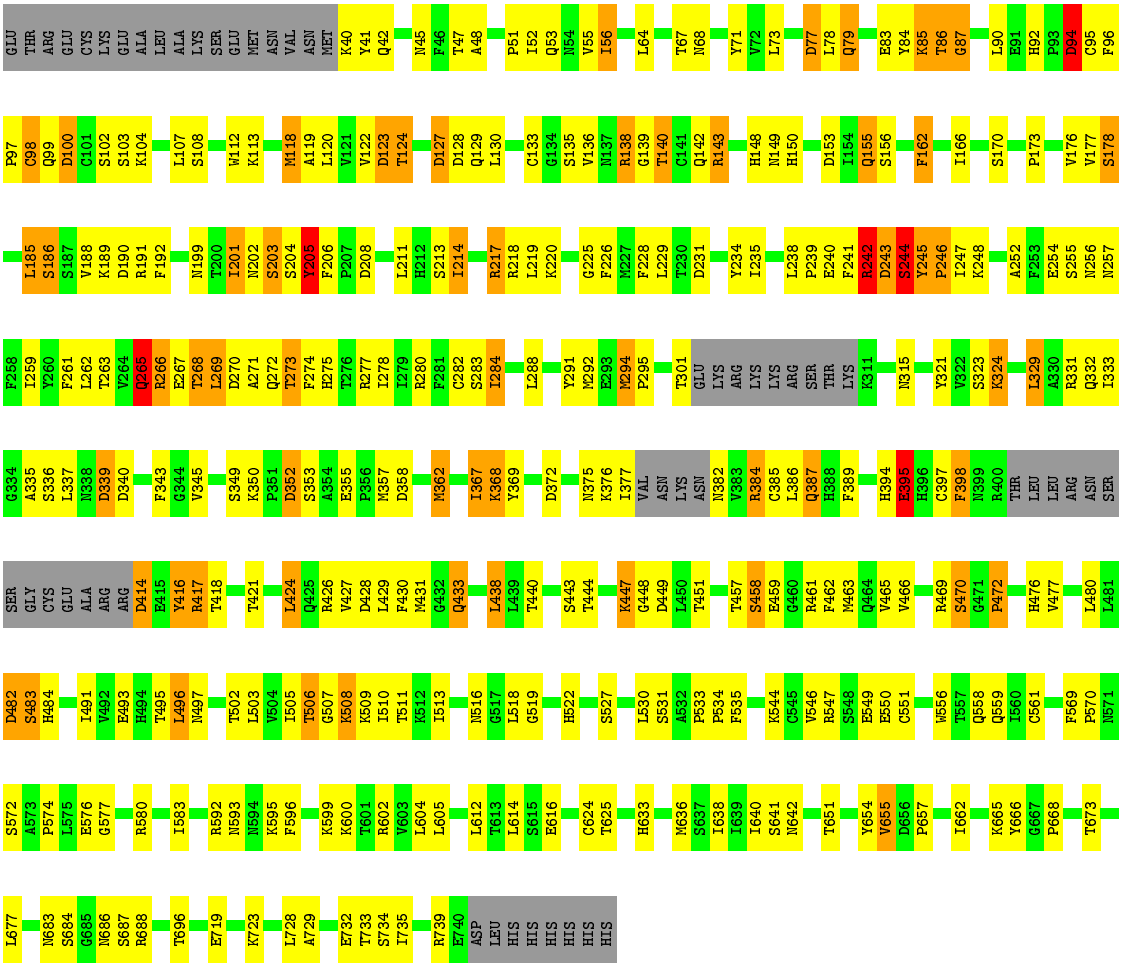
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	155	Total	C	N	O	S	0	0	0
			1158	717	208	231	2			
3	F	155	Total	C	N	O	S	0	0	0
			1158	717	208	231	2			

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. 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. 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.

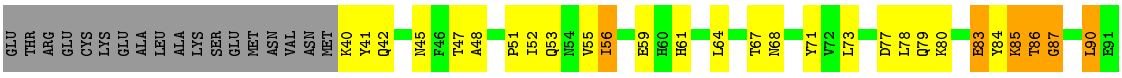
- Molecule 1: Hepatocyte growth factor receptor

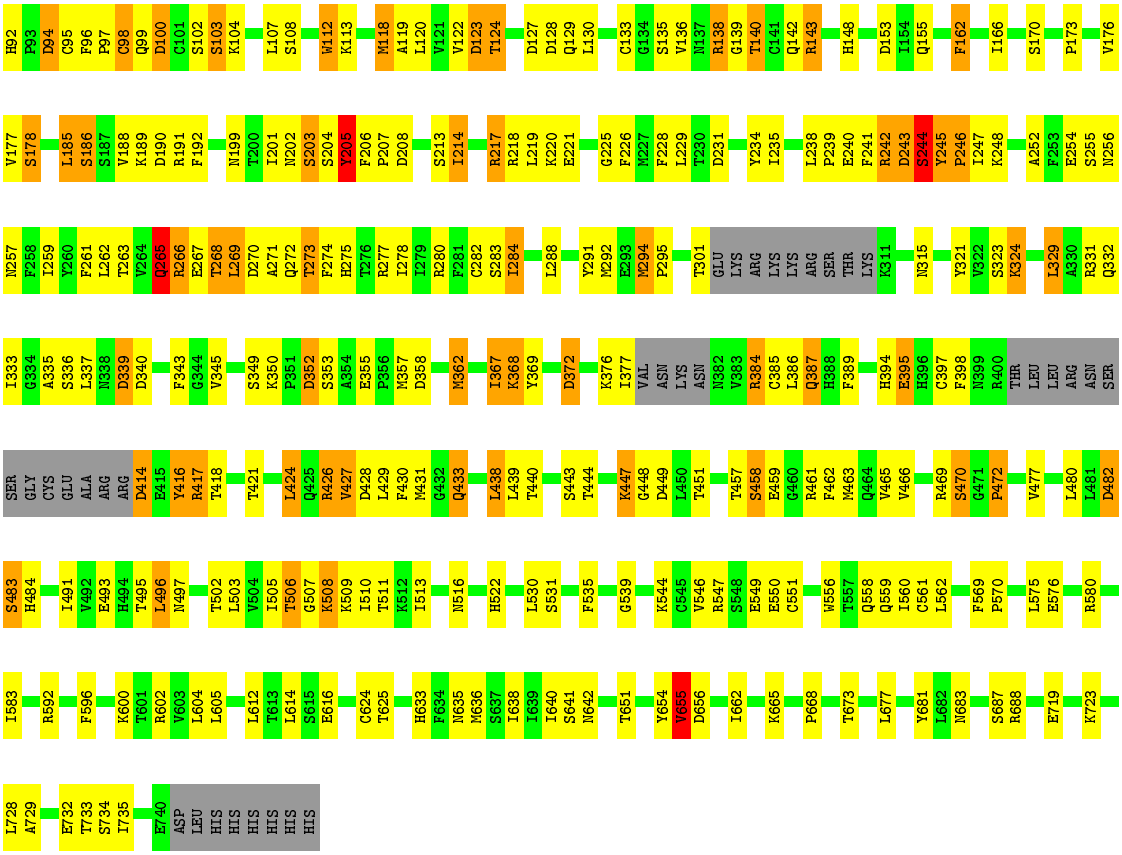
Chain A:



- Molecule 1: Hepatocyte growth factor receptor

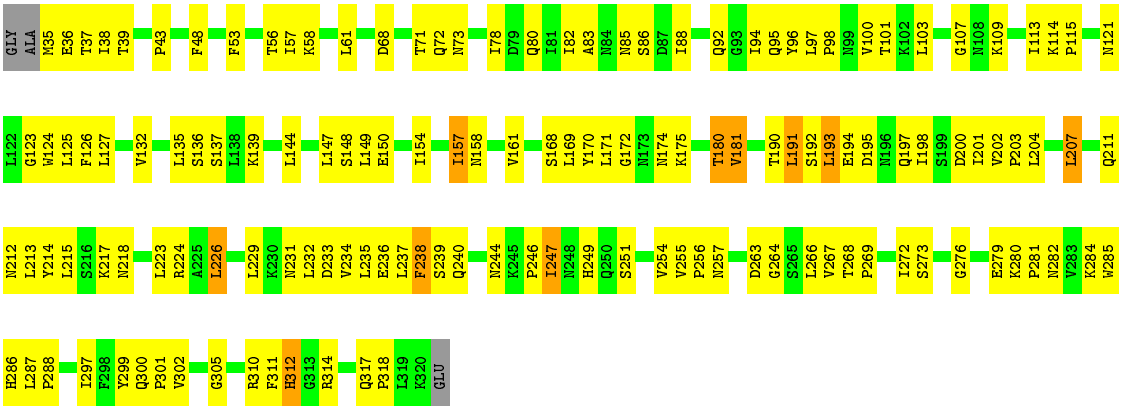
Chain D:





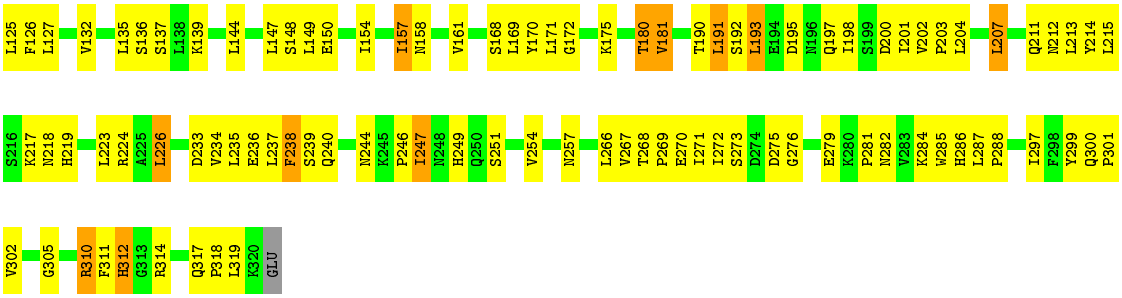
- Molecule 2: Internalin B

Chain B:



- Molecule 2: Internalin B

Chain E:

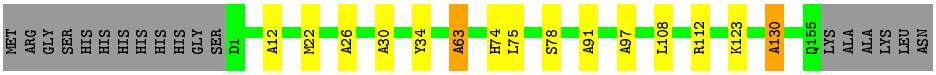


- Molecule 3: DARPin A3A

Chain C:

- Molecule 3: DARPin A3A

Chain F:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 31	Depositor
Cell constants a, b, c, α, β, γ	144.87Å 144.87Å 128.90Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	48.15 – 6.00 48.15 – 6.00	Depositor EDS
% Data completeness (in resolution range)	99.8 (48.15-6.00) 99.8 (48.15-6.00)	Depositor EDS
R_{merge}	0.28	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.47 (at 6.15Å)	Xtriage
Refinement program	PHENIX (dev_3071: ???)	Depositor
R, R_{free}	0.263 , 0.271 0.263 , 0.271	Depositor DCC
R_{free} test set	339 reflections (4.49%)	wwPDB-VP
Wilson B-factor (Å^2)	248.6	Xtriage
Anisotropy	0.111	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 296.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.46$, $< L^2 > = 0.28$	Xtriage
Estimated twinning fraction	0.017 for -h,-k,l 0.042 for h,-h-k,-l 0.043 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.82	EDS
Total number of atoms	17448	wwPDB-VP
Average B, all atoms (Å^2)	317.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.19% 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

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	0.51	2/5439 (0.0%)	0.98	23/7377 (0.3%)
1	D	0.50	1/5439 (0.0%)	0.77	17/7377 (0.2%)
2	B	0.25	0/2288	0.46	1/3105 (0.0%)
2	E	0.25	0/2288	0.52	3/3105 (0.1%)
3	C	0.93	3/1175 (0.3%)	0.82	5/1594 (0.3%)
3	F	0.94	3/1175 (0.3%)	0.83	5/1594 (0.3%)
All	All	0.54	9/17804 (0.1%)	0.79	54/24152 (0.2%)

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
1	A	0	4
1	D	0	5
3	C	0	2
3	F	0	2
All	All	0	13

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	63	ALA	C-N	-17.98	0.92	1.34
3	F	63	ALA	C-N	-17.95	0.92	1.34
1	D	655	VAL	C-N	12.36	1.62	1.34
1	A	382	ASN	N-CA	-11.45	1.23	1.46
3	C	97	ALA	C-N	-10.00	1.11	1.34

The worst 5 of 54 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	655	VAL	O-C-N	30.93	172.18	122.70

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	655	VAL	CA-C-N	-22.94	66.74	117.20
1	A	382	ASN	N-CA-CB	21.14	148.66	110.60
1	A	382	ASN	N-CA-C	-19.68	57.87	111.00
1	A	655	VAL	C-N-CA	-18.14	76.34	121.70

There are no chirality outliers.

5 of 13 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	135	SER	Peptide
1	A	244	SER	Peptide
1	A	265	GLN	Peptide
1	A	274	PHE	Peptide
3	C	63	ALA	Mainchain

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	5314	0	5162	311	23
1	D	5314	0	5159	322	17
2	B	2252	0	2322	190	14
2	E	2252	0	2322	180	5
3	C	1158	0	1132	31	11
3	F	1158	0	1130	32	9
All	All	17448	0	17227	823	44

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 823 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:732:GLU:OE2	1:D:723:LYS:NZ	1.60	1.35
2:B:264:GLY:HA3	2:E:180:THR:CG2	1.57	1.31

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:602:ARG:NH2	2:E:170:TYR:OH	1.66	1.28
1:D:372:ASP:HB3	2:E:275:ASP:OD1	1.14	1.27
1:D:41:TYR:HE1	1:D:522:HIS:CD2	1.52	1.27

The worst 5 of 44 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:D:459:GLU:OE2	1:D:635:ASN:ND2[2_455]	0.86	1.34
1:D:459:GLU:OE2	1:D:635:ASN:CG[2_455]	1.00	1.20
1:A:155:GLN:OE1	2:B:71:THR:CG2[2_554]	1.06	1.14
1:A:155:GLN:CD	2:B:71:THR:CG2[2_554]	1.06	1.14
3:C:123:LYS:NZ	2:E:175:LYS:CB[3_544]	1.09	1.11

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	663/727 (91%)	589 (89%)	58 (9%)	16 (2%)	6 33
1	D	663/727 (91%)	588 (89%)	60 (9%)	15 (2%)	6 34
2	B	284/289 (98%)	228 (80%)	52 (18%)	4 (1%)	11 46
2	E	284/289 (98%)	227 (80%)	53 (19%)	4 (1%)	11 46
3	C	153/173 (88%)	149 (97%)	4 (3%)	0	100 100
3	F	153/173 (88%)	149 (97%)	4 (3%)	0	100 100
All	All	2200/2378 (92%)	1930 (88%)	231 (10%)	39 (2%)	8 40

5 of 39 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	87	GLY

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	270	ASP
1	A	272	GLN
1	A	483	SER
2	B	193	LEU

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	605/654 (92%)	496 (82%)	109 (18%)	1 10
1	D	605/654 (92%)	497 (82%)	108 (18%)	2 10
2	B	264/265 (100%)	252 (96%)	12 (4%)	27 52
2	E	264/265 (100%)	252 (96%)	12 (4%)	27 52
3	C	118/132 (89%)	118 (100%)	0	100 100
3	F	118/132 (89%)	118 (100%)	0	100 100
All	All	1974/2102 (94%)	1733 (88%)	241 (12%)	5 20

5 of 241 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	37	THR
1	D	100	ASP
1	D	495	THR
2	B	157	ILE
1	D	42	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 38 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	231	ASN
1	D	129	GLN
2	E	211	GLN
1	D	53	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	D	256	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 carbohydrates 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

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	D	3
1	A	2
3	C	2
3	F	2

The worst 5 of 9 chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	D	518:LEU	C	519:GLY	N	6.49
1	A	518:LEU	C	519:GLY	N	5.09
1	D	560:ILE	C	561:CYS	N	4.19

Continued on next page...

Continued from previous page...

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	560:ILE	C	561:CYS	N	3.62
1	D	655:VAL	C	656:ASP	N	1.62

6 Fit of model and data

6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.