Full wwPDB X-ray Structure Validation Report (i)

Nov 25, 2024 – 06:25 PM EST

PDB ID	:	3KG7
Title	:	Dehydratase domain from CurH module of Curacin polyketide synthase
Authors	:	Akey, D.L.; Smith, J.L.
Deposited on	:	2009-10-28
Resolution	:	2.77 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (1)) were used in the production of this report:

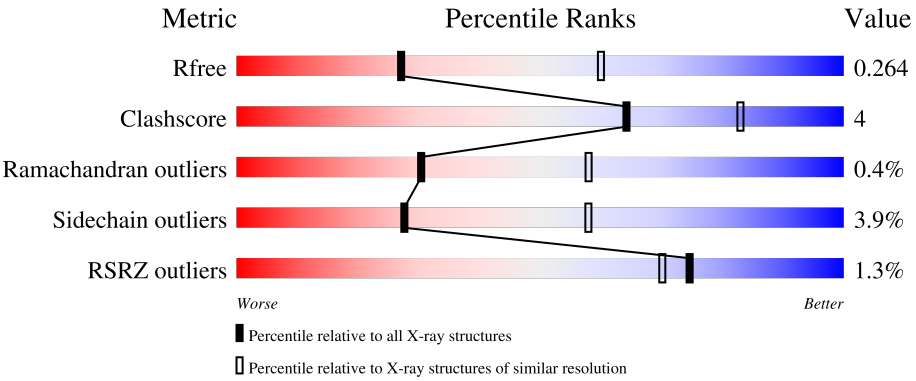
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.21
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.004 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.40

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 2.77 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}	164625	4924 (2.80-2.76)
Clashscore	180529	5458 (2.80-2.76)
Ramachandran outliers	177936	5386 (2.80-2.76)
Sidechain outliers	177891	5388 (2.80-2.76)
RSRZ outliers	164620	4926 (2.80-2.76)

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%

Mol	Chain	Length	Quality of chain
1	A	293	87% 10%
1	B	293	84% 12%
1	C	293	83% 14%
1	D	293	83% 14%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 9076 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 CurH.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	286	Total	C	N	O	S	Se	0	0	0
			2259	1440	381	431	2	5			
1	B	285	Total	C	N	O	S	Se	0	0	0
			2252	1435	380	430	2	5			
1	C	287	Total	C	N	O	S	Se	0	0	0
			2268	1445	383	433	2	5			
1	D	285	Total	C	N	O	S	Se	0	0	0
			2252	1435	380	430	2	5			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	931	SER	-	expression tag	UNP Q6DNE5
A	932	ASN	-	expression tag	UNP Q6DNE5
A	933	ALA	-	expression tag	UNP Q6DNE5
B	931	SER	-	expression tag	UNP Q6DNE5
B	932	ASN	-	expression tag	UNP Q6DNE5
B	933	ALA	-	expression tag	UNP Q6DNE5
C	931	SER	-	expression tag	UNP Q6DNE5
C	932	ASN	-	expression tag	UNP Q6DNE5
C	933	ALA	-	expression tag	UNP Q6DNE5
D	931	SER	-	expression tag	UNP Q6DNE5
D	932	ASN	-	expression tag	UNP Q6DNE5
D	933	ALA	-	expression tag	UNP Q6DNE5

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	17	Total	O	0	0
			17	17		
2	B	9	Total	O	0	0
			9	9		

Continued on next page...

Continued from previous page...

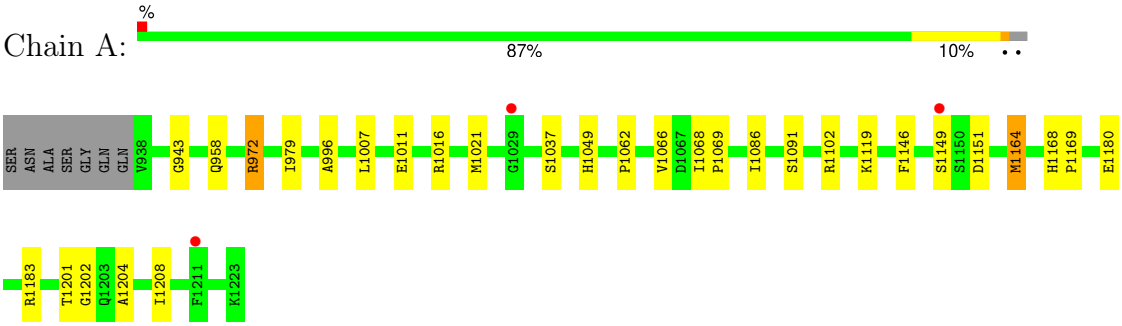
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	C	9	Total O 9 9	0	0
2	D	10	Total O 10 10	0	0

3 Residue-property plots (i)

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. 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.

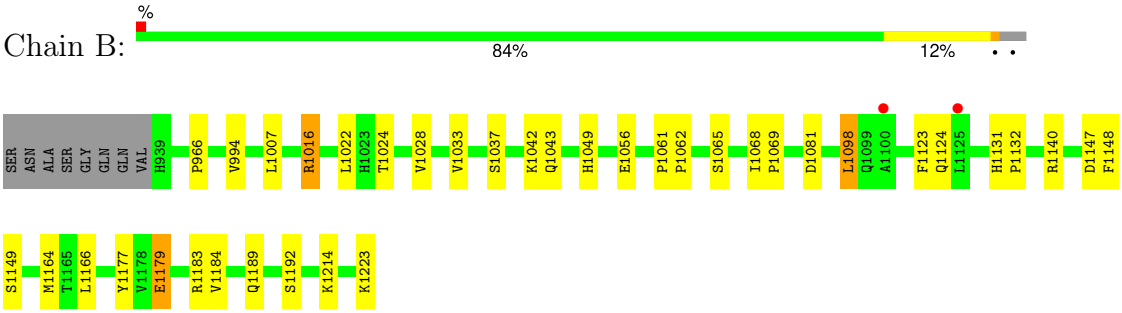
• Molecule 1: CurH

Chain A:



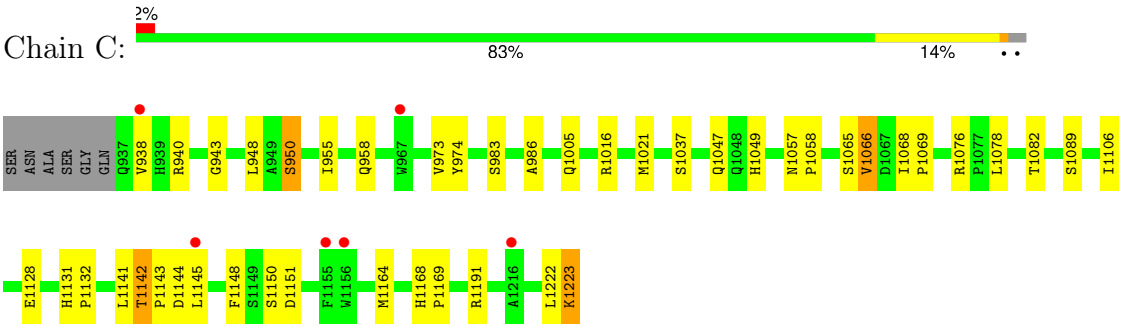
• Molecule 1: CurH

Chain B:



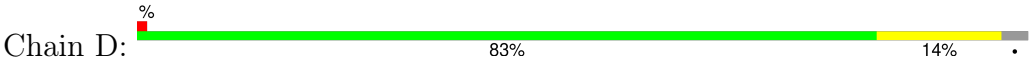
• Molecule 1: CurH

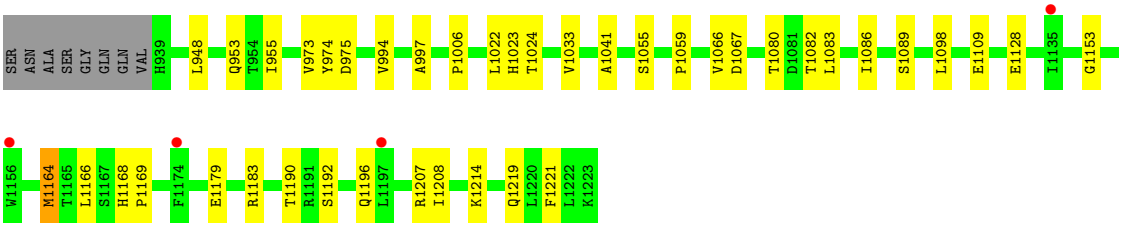
Chain C:



• Molecule 1: CurH

Chain D:




SER
ASN
ALA
SER
GLY
GLN
GLN
VAL
H939
L948
Q953
T954
T955
V973
Y974
D975
V994
A997
P1006
L1022
H1023
T1024
V1033
A1041
S1055
P1059
V1066
D1067
T1080
D1081
T1082
L1083
I1086
S1089
L1098
E1109
E1128
I1135
G1153
W1156
M1164
T1165
L1166
S1167
H1168
P1169
F1174
E1179
R1183
T1190
R1191
S1192
Q1196
L1197
R1207
I1208
K1214
Q1219
L1220
F1221
L1222
K1223

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	82.16Å 74.45Å 103.48Å 90.00° 108.84° 90.00°	Depositor
Resolution (Å)	48.97 – 2.77 48.97 – 2.77	Depositor EDS
% Data completeness (in resolution range)	98.8 (48.97-2.77) 98.8 (48.97-2.77)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	0.11	Depositor
$< I/\sigma(I) >$ [1]	3.26 (at 2.77Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.203 , 0.269 0.200 , 0.264	Depositor DCC
R_{free} test set	1996 reflections (6.69%)	wwPDB-VP
Wilson B-factor (Å^2)	73.1	Xtriage
Anisotropy	0.469	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 75.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	9076	wwPDB-VP
Average B, all atoms (Å^2)	85.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.18% 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	0.43	0/2313	0.56	0/3154
1	B	0.40	0/2306	0.55	0/3144
1	C	0.48	2/2322 (0.1%)	0.58	2/3166 (0.1%)
1	D	0.39	0/2306	0.54	0/3144
All	All	0.43	2/9247 (0.0%)	0.56	2/12608 (0.0%)

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	C	0	1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	1076	ARG	CZ-NH1	6.76	1.41	1.33
1	C	1076	ARG	NE-CZ	5.77	1.40	1.33

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	1076	ARG	NE-CZ-NH1	5.42	123.01	120.30
1	C	1076	ARG	NE-CZ-NH2	-5.26	117.67	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	1222	LEU	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	2259	0	2209	14	0
1	B	2252	0	2200	17	0
1	C	2268	0	2217	20	0
1	D	2252	0	2200	18	0
2	A	17	0	0	2	0
2	B	9	0	0	0	0
2	C	9	0	0	0	0
2	D	10	0	0	1	0
All	All	9076	0	8826	68	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 4.

All (68) 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:1016:ARG:NH1	1:B:1037:SER:OG	2.23	0.70
1:B:1028:VAL:HG12	1:B:1056:GLU:OE1	1.93	0.68
1:D:1168:HIS:HB2	1:D:1169:PRO:HD2	1.79	0.63
1:D:994:VAL:HG11	1:D:1166:LEU:HD23	1.81	0.63
1:A:979:ILE:HD12	1:A:1007:LEU:HD22	1.80	0.62
1:C:1047:GLN:OE1	1:C:1223:LYS:HD2	1.99	0.61
1:B:1183:ARG:HG3	1:B:1192:SER:HB3	1.83	0.59
1:D:953:GLN:HG2	1:D:1023:HIS:HA	1.85	0.58
1:D:997:ALA:HB2	1:D:1059:PRO:HB3	1.88	0.55
1:C:1142:THR:HG22	1:C:1143:PRO:HD2	1.88	0.55
1:C:1037:SER:OG	1:C:1049:HIS:NE2	2.36	0.54
1:C:1089:SER:HB3	1:C:1150:SER:HB2	1.88	0.53
1:B:1022:LEU:HD12	1:B:1033:VAL:HG22	1.90	0.53
1:C:1068:ILE:HG22	1:C:1069:PRO:HD3	1.91	0.52
1:B:1131:HIS:CG	1:B:1132:PRO:HD2	2.45	0.52
1:B:1098:LEU:O	1:B:1140:ARG:NH2	2.39	0.51
1:C:1106:ILE:HD11	1:C:1143:PRO:HA	1.93	0.51
1:D:1022:LEU:HD12	1:D:1033:VAL:HG22	1.93	0.51
1:C:1144:ASP:OD2	1:C:1191:ARG:NH2	2.44	0.51

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:972:ARG:HD2	2:A:5:HOH:O	2.12	0.50
1:C:1168:HIS:HB2	1:C:1169:PRO:CD	2.43	0.49
1:B:966:PRO:HD2	1:B:1124:GLN:OE1	2.12	0.49
1:A:1016:ARG:NH1	1:A:1037:SER:OG	2.45	0.49
1:A:1037:SER:OG	1:A:1049:HIS:NE2	2.39	0.48
1:D:1089:SER:O	1:D:1153:GLY:HA2	2.13	0.48
1:D:1168:HIS:HB2	1:D:1169:PRO:CD	2.42	0.48
1:B:1068:ILE:HB	1:B:1069:PRO:HD3	1.96	0.48
1:D:997:ALA:HB3	1:D:1055:SER:HB3	1.97	0.47
1:C:1168:HIS:HB2	1:C:1169:PRO:HD2	1.97	0.47
1:D:1006:PRO:HG3	1:D:1221:PHE:CE1	2.49	0.47
1:A:1168:HIS:HB2	1:A:1169:PRO:HD2	1.97	0.47
1:B:1148:PHE:HZ	1:B:1184:VAL:HG21	1.81	0.46
1:D:1080:THR:HA	1:D:1083:LEU:HD12	1.97	0.46
1:D:1082:THR:O	1:D:1086:ILE:HG12	2.15	0.46
1:C:1065:SER:O	1:C:1066:VAL:HG23	2.16	0.46
1:D:975:ASP:HA	2:D:8:HOH:O	2.15	0.46
1:C:943:GLY:HA3	1:C:958:GLN:HB3	1.96	0.46
1:A:943:GLY:HA3	1:A:958:GLN:HB3	1.98	0.45
1:C:973:VAL:HG12	1:C:974:TYR:CD2	2.52	0.45
1:A:1062:PRO:HB3	1:A:1204:ALA:O	2.17	0.45
1:A:1068:ILE:N	1:A:1069:PRO:HD2	2.31	0.45
1:C:1131:HIS:CG	1:C:1132:PRO:HD2	2.52	0.44
1:A:1164:MSE:HG2	1:A:1208:ILE:HG12	1.99	0.44
1:D:1196:GLN:OE1	1:D:1207:ARG:HG2	2.17	0.44
1:B:1223:LYS:HE3	1:B:1223:LYS:HB2	1.80	0.44
1:D:973:VAL:HG12	1:D:974:TYR:CD2	2.53	0.44
1:B:994:VAL:HG11	1:B:1166:LEU:HD23	2.00	0.43
1:C:1057:ASN:HA	1:C:1058:PRO:HD2	1.80	0.43
1:B:1177:TYR:CE2	1:B:1179:GLU:HG3	2.53	0.43
1:D:948:LEU:HD21	1:D:955:ILE:HD12	2.00	0.43
1:B:1007:LEU:HD12	1:B:1049:HIS:CD2	2.54	0.43
1:B:1148:PHE:HA	1:B:1149:SER:HA	1.83	0.43
1:B:1189:GLN:O	1:B:1214:LYS:HG3	2.19	0.43
1:A:1021:MSE:HE3	1:C:950:SER:HB3	2.00	0.43
1:D:1067:ASP:OD1	1:D:1067:ASP:N	2.49	0.43
1:B:1061:PRO:HA	1:B:1062:PRO:HD3	1.89	0.43
1:A:996:ALA:HA	2:A:17:HOH:O	2.17	0.42
1:A:1011:GLU:HA	1:A:1011:GLU:OE1	2.19	0.42
1:C:983:SER:O	1:C:986:ALA:HB3	2.20	0.41
1:C:1141:LEU:HA	1:C:1145:LEU:HD13	2.02	0.41

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:938:VAL:HG23	1:C:943:GLY:HA2	2.03	0.41
1:C:1005:GLN:H	1:C:1005:GLN:HG2	1.73	0.41
1:A:1086:ILE:HD11	1:A:1146:PHE:HD1	1.85	0.41
1:C:948:LEU:HD11	1:C:955:ILE:HD11	2.02	0.41
1:B:1042:LYS:O	1:B:1043:GLN:HB2	2.20	0.41
1:D:1190:THR:HG22	1:D:1214:LYS:HB2	2.03	0.40
1:A:1066:VAL:O	1:A:1202:GLY:HA3	2.21	0.40
1:D:1164:MSE:HG2	1:D:1208:ILE:HG12	2.02	0.40

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	284/293 (97%)	274 (96%)	9 (3%)	1 (0%)	30	58
1	B	283/293 (97%)	272 (96%)	11 (4%)	0	100	100
1	C	285/293 (97%)	265 (93%)	18 (6%)	2 (1%)	19	45
1	D	283/293 (97%)	274 (97%)	8 (3%)	1 (0%)	30	58
All	All	1135/1172 (97%)	1085 (96%)	46 (4%)	4 (0%)	30	58

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	1148	PHE
1	A	1149	SER
1	C	1151	ASP
1	D	1041	ALA

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	250/250 (100%)	241 (96%)	9 (4%)	30	61
1	B	249/250 (100%)	240 (96%)	9 (4%)	30	61
1	C	251/250 (100%)	240 (96%)	11 (4%)	24	53
1	D	249/250 (100%)	239 (96%)	10 (4%)	27	57
All	All	999/1000 (100%)	960 (96%)	39 (4%)	27	58

All (39) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	972	ARG
1	A	1091	SER
1	A	1102	ARG
1	A	1119	LYS
1	A	1151	ASP
1	A	1164	MSE
1	A	1180	GLU
1	A	1183	ARG
1	A	1201	THR
1	B	1016	ARG
1	B	1024	THR
1	B	1065	SER
1	B	1081	ASP
1	B	1098	LEU
1	B	1123	PHE
1	B	1147	ASP
1	B	1164	MSE
1	B	1179	GLU
1	C	940	ARG
1	C	950	SER
1	C	1016	ARG
1	C	1021	MSE
1	C	1066	VAL
1	C	1078	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	1082	THR
1	C	1128	GLU
1	C	1142	THR
1	C	1164	MSE
1	C	1223	LYS
1	D	1024	THR
1	D	1066	VAL
1	D	1098	LEU
1	D	1109	GLU
1	D	1128	GLU
1	D	1164	MSE
1	D	1179	GLU
1	D	1183	ARG
1	D	1192	SER
1	D	1219	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	1043	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 oligosaccharides 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

Warning: The R factor obtained from EDS is 0.2568, which does not match the depositor's R factor of 0.203. Please interpret the results in this section carefully.

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	281/293 (95%)	0.03	3 (1%)	77	73	68, 83, 98, 114	0
1	B	280/293 (95%)	0.02	2 (0%)	84	80	74, 83, 98, 124	0
1	C	282/293 (96%)	0.39	6 (2%)	63	57	46, 83, 98, 113	0
1	D	280/293 (95%)	0.42	4 (1%)	73	68	74, 84, 104, 113	0
All	All	1123/1172 (95%)	0.22	15 (1%)	74	70	46, 83, 100, 124	0

All (15) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	1174	PHE	3.7
1	B	1100	ALA	3.4
1	C	1155	PHE	3.4
1	D	1156	TRP	2.7
1	C	1156	TRP	2.6
1	C	967	TRP	2.5
1	A	1029	GLY	2.5
1	C	1145	LEU	2.4
1	C	1216	ALA	2.4
1	A	1149	SER	2.3
1	C	938	VAL	2.2
1	A	1211	PHE	2.1
1	B	1125	LEU	2.1
1	D	1135	ILE	2.1
1	D	1197	LEU	2.0

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 (i)

There are no monosaccharides 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.