wwPDB X-ray Structure Validation Summary Report (i)

Jan 23, 2021 – 02:15 PM EST

PDB ID	:	1VI8
Title	:	Crystal structure of a putative thioesterase
Authors	:	Structural GenomiX; Burley, S.K.; New York SGX Research Center for Structural Genomics (NYSGXRC)
Deposited on	:	2003-12-01
Resolution	:	2.20 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:

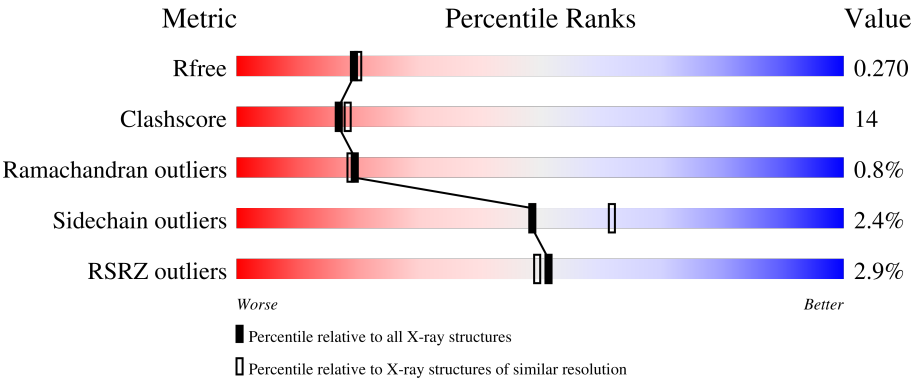
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.16
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.16

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 2.20 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	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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	148	72% 20% 7%
1	B	148	77% 21%
1	C	148	66% 27% 7%
1	D	148	11% 54% 36% 7%
1	E	148	75% 16% 7%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	148	76% 16% • 7%
1	G	148	% 68% 24% • 7%
1	H	148	5% 64% 28% • 7%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 9182 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 Hypothetical protein ydiI.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	138	Total	C	N	O	S	0	2	0
			1069	669	195	198	7			
1	B	146	Total	C	N	O	S	0	3	0
			1131	701	216	207	7			
1	C	138	Total	C	N	O	S	0	1	0
			1055	654	197	197	7			
1	D	138	Total	C	N	O	S	0	2	0
			1050	652	193	198	7			
1	E	138	Total	C	N	O	S	0	1	0
			1061	662	194	198	7			
1	F	138	Total	C	N	O	S	0	1	0
			1065	665	196	197	7			
1	G	137	Total	C	N	O	S	0	2	0
			1054	662	189	196	7			
1	H	138	Total	C	N	O	S	0	3	0
			1076	673	197	199	7			

There are 104 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MET	-	cloning artifact	UNP P77781
A	0	SER	-	cloning artifact	UNP P77781
A	1	LEU	-	cloning artifact	UNP P77781
A	137	GLU	-	cloning artifact	UNP P77781
A	138	GLY	-	cloning artifact	UNP P77781
A	139	GLY	-	cloning artifact	UNP P77781
A	140	SER	-	cloning artifact	UNP P77781
A	141	HIS	-	cloning artifact	UNP P77781
A	142	HIS	-	cloning artifact	UNP P77781
A	143	HIS	-	cloning artifact	UNP P77781
A	144	HIS	-	cloning artifact	UNP P77781
A	145	HIS	-	cloning artifact	UNP P77781
A	146	HIS	-	cloning artifact	UNP P77781

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-1	MET	-	cloning artifact	UNP P77781
B	0	SER	-	cloning artifact	UNP P77781
B	1	LEU	-	cloning artifact	UNP P77781
B	137	GLU	-	cloning artifact	UNP P77781
B	138	GLY	-	cloning artifact	UNP P77781
B	139	GLY	-	cloning artifact	UNP P77781
B	140	SER	-	cloning artifact	UNP P77781
B	141	HIS	-	cloning artifact	UNP P77781
B	142	HIS	-	cloning artifact	UNP P77781
B	143	HIS	-	cloning artifact	UNP P77781
B	144	HIS	-	cloning artifact	UNP P77781
B	145	HIS	-	cloning artifact	UNP P77781
B	146	HIS	-	cloning artifact	UNP P77781
C	-1	MET	-	cloning artifact	UNP P77781
C	0	SER	-	cloning artifact	UNP P77781
C	1	LEU	-	cloning artifact	UNP P77781
C	137	GLU	-	cloning artifact	UNP P77781
C	138	GLY	-	cloning artifact	UNP P77781
C	139	GLY	-	cloning artifact	UNP P77781
C	140	SER	-	cloning artifact	UNP P77781
C	141	HIS	-	cloning artifact	UNP P77781
C	142	HIS	-	cloning artifact	UNP P77781
C	143	HIS	-	cloning artifact	UNP P77781
C	144	HIS	-	cloning artifact	UNP P77781
C	145	HIS	-	cloning artifact	UNP P77781
C	146	HIS	-	cloning artifact	UNP P77781
D	-1	MET	-	cloning artifact	UNP P77781
D	0	SER	-	cloning artifact	UNP P77781
D	1	LEU	-	cloning artifact	UNP P77781
D	137	GLU	-	cloning artifact	UNP P77781
D	138	GLY	-	cloning artifact	UNP P77781
D	139	GLY	-	cloning artifact	UNP P77781
D	140	SER	-	cloning artifact	UNP P77781
D	141	HIS	-	cloning artifact	UNP P77781
D	142	HIS	-	cloning artifact	UNP P77781
D	143	HIS	-	cloning artifact	UNP P77781
D	144	HIS	-	cloning artifact	UNP P77781
D	145	HIS	-	cloning artifact	UNP P77781
D	146	HIS	-	cloning artifact	UNP P77781
E	-1	MET	-	cloning artifact	UNP P77781
E	0	SER	-	cloning artifact	UNP P77781
E	1	LEU	-	cloning artifact	UNP P77781

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	137	GLU	-	cloning artifact	UNP P77781
E	138	GLY	-	cloning artifact	UNP P77781
E	139	GLY	-	cloning artifact	UNP P77781
E	140	SER	-	cloning artifact	UNP P77781
E	141	HIS	-	cloning artifact	UNP P77781
E	142	HIS	-	cloning artifact	UNP P77781
E	143	HIS	-	cloning artifact	UNP P77781
E	144	HIS	-	cloning artifact	UNP P77781
E	145	HIS	-	cloning artifact	UNP P77781
E	146	HIS	-	cloning artifact	UNP P77781
F	-1	MET	-	cloning artifact	UNP P77781
F	0	SER	-	cloning artifact	UNP P77781
F	1	LEU	-	cloning artifact	UNP P77781
F	137	GLU	-	cloning artifact	UNP P77781
F	138	GLY	-	cloning artifact	UNP P77781
F	139	GLY	-	cloning artifact	UNP P77781
F	140	SER	-	cloning artifact	UNP P77781
F	141	HIS	-	cloning artifact	UNP P77781
F	142	HIS	-	cloning artifact	UNP P77781
F	143	HIS	-	cloning artifact	UNP P77781
F	144	HIS	-	cloning artifact	UNP P77781
F	145	HIS	-	cloning artifact	UNP P77781
F	146	HIS	-	cloning artifact	UNP P77781
G	-1	MET	-	cloning artifact	UNP P77781
G	0	SER	-	cloning artifact	UNP P77781
G	1	LEU	-	cloning artifact	UNP P77781
G	137	GLU	-	cloning artifact	UNP P77781
G	138	GLY	-	cloning artifact	UNP P77781
G	139	GLY	-	cloning artifact	UNP P77781
G	140	SER	-	cloning artifact	UNP P77781
G	141	HIS	-	cloning artifact	UNP P77781
G	142	HIS	-	cloning artifact	UNP P77781
G	143	HIS	-	cloning artifact	UNP P77781
G	144	HIS	-	cloning artifact	UNP P77781
G	145	HIS	-	cloning artifact	UNP P77781
G	146	HIS	-	cloning artifact	UNP P77781
H	-1	MET	-	cloning artifact	UNP P77781
H	0	SER	-	cloning artifact	UNP P77781
H	1	LEU	-	cloning artifact	UNP P77781
H	137	GLU	-	cloning artifact	UNP P77781
H	138	GLY	-	cloning artifact	UNP P77781
H	139	GLY	-	cloning artifact	UNP P77781

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
H	140	SER	-	cloning artifact	UNP P77781
H	141	HIS	-	cloning artifact	UNP P77781
H	142	HIS	-	cloning artifact	UNP P77781
H	143	HIS	-	cloning artifact	UNP P77781
H	144	HIS	-	cloning artifact	UNP P77781
H	145	HIS	-	cloning artifact	UNP P77781
H	146	HIS	-	cloning artifact	UNP P77781

- Molecule 2 is water.

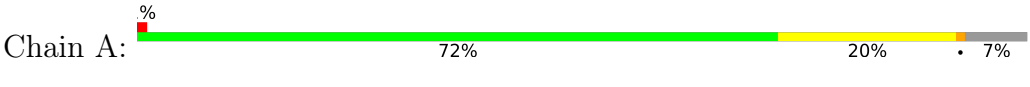
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	81	Total O 81 81	0	0
2	B	107	Total O 107 107	0	0
2	C	58	Total O 58 58	0	0
2	D	62	Total O 62 62	0	0
2	E	95	Total O 95 95	0	0
2	F	75	Total O 75 75	0	0
2	G	63	Total O 63 63	0	0
2	H	80	Total O 80 80	0	0

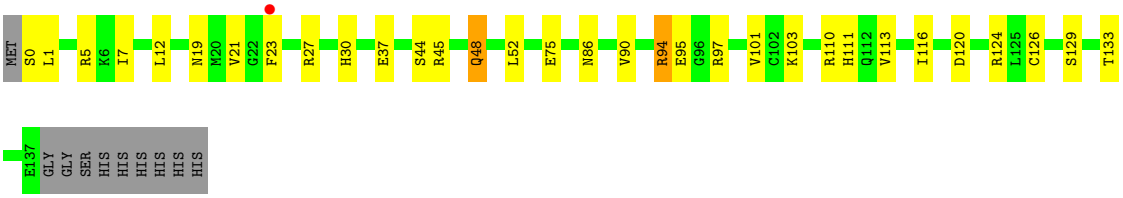
3 Residue-property plots (i)

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 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: Hypothetical protein ydiI

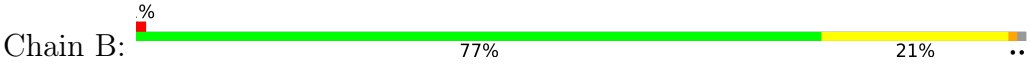
Chain A:

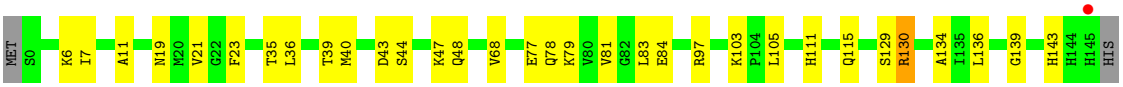




• Molecule 1: Hypothetical protein ydiI

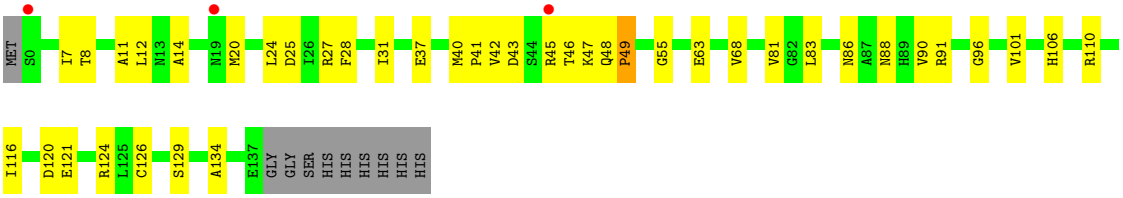
Chain B:





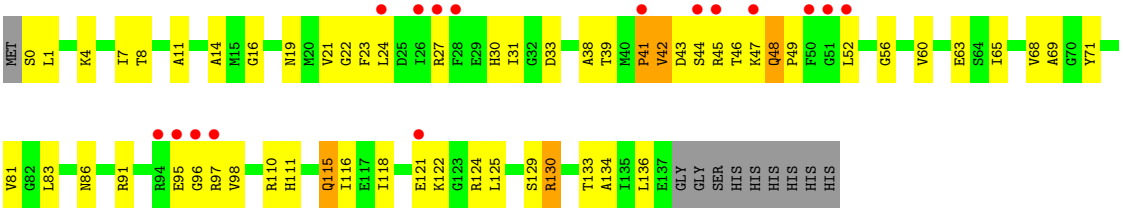
• Molecule 1: Hypothetical protein ydiI

Chain C:



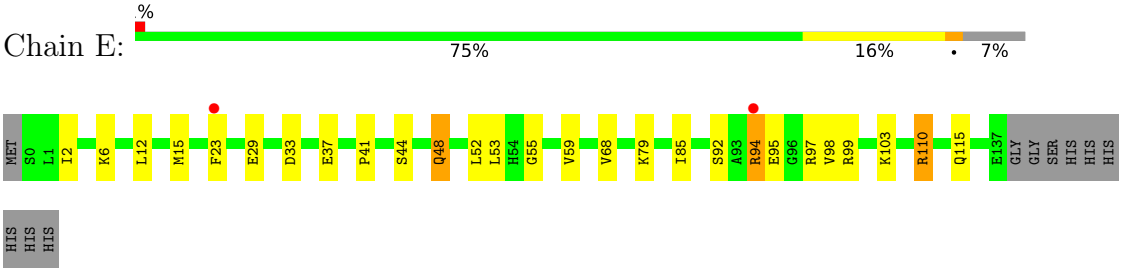
• Molecule 1: Hypothetical protein ydiI

Chain D:



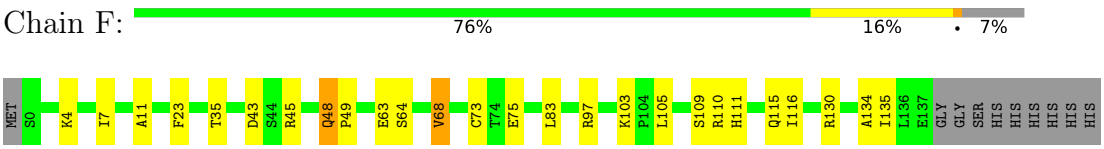
• Molecule 1: Hypothetical protein ydiI

Chain E:



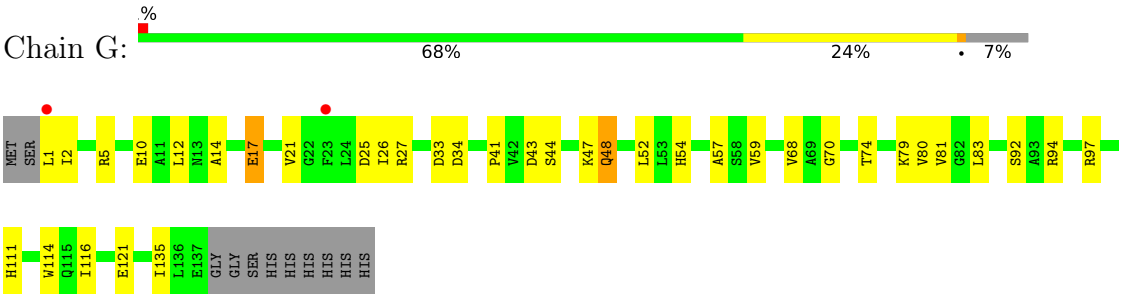
• Molecule 1: Hypothetical protein ydiI

Chain F:



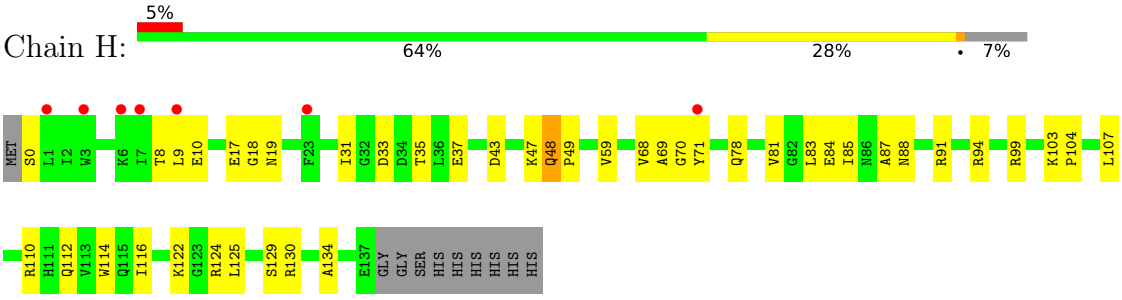
• Molecule 1: Hypothetical protein ydiI

Chain G:



• Molecule 1: Hypothetical protein ydiI

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	78.66Å 67.46Å 105.19Å 90.00° 92.33° 90.00°	Depositor
Resolution (Å)	46.51 – 2.20 46.51 – 2.20	Depositor EDS
% Data completeness (in resolution range)	(Not available) (46.51-2.20) 95.5 (46.51-2.20)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.73 (at 2.20Å)	Xtriage
Refinement program	REFMAC 4.0	Depositor
R, R_{free}	0.241 , 0.302 0.214 , 0.270	Depositor DCC
R_{free} test set	2733 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å^2)	39.4	Xtriage
Anisotropy	0.292	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 53.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.028 for h,-k,-l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	9182	wwPDB-VP
Average B, all atoms (Å^2)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.47% 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	0.36	0/1096	0.62	0/1476
1	B	0.37	0/1167	0.61	0/1570
1	C	0.32	0/1077	0.59	0/1450
1	D	0.32	0/1074	0.56	0/1446
1	E	0.38	0/1082	0.64	0/1457
1	F	0.34	0/1087	0.61	0/1464
1	G	0.32	0/1081	0.59	0/1458
1	H	0.34	0/1108	0.59	0/1491
All	All	0.35	0/8772	0.60	0/11812

There are no bond length outliers.

There are no bond angle outliers.

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	1069	0	1080	31	0
1	B	1131	0	1129	31	0
1	C	1055	0	1058	36	0
1	D	1050	0	1049	49	0
1	E	1061	0	1074	24	0
1	F	1065	0	1076	23	0
1	G	1054	0	1052	36	0
1	H	1076	0	1089	35	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	81	0	0	6	0
2	B	107	0	0	10	0
2	C	58	0	0	3	0
2	D	62	0	0	5	0
2	E	95	0	0	4	0
2	F	75	0	0	7	0
2	G	63	0	0	10	0
2	H	80	0	0	10	0
All	All	9182	0	8607	242	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 14.

The worst 5 of 242 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:E:110:ARG:HB2	1:E:110:ARG:HH21	1.24	0.99
1:A:111:HIS:HE1	2:A:211:HOH:O	1.58	0.86
1:C:12:LEU:HD11	1:C:68:VAL:HG21	1.60	0.84
1:G:14:ALA:O	1:G:17:GLU:HB2	1.79	0.81
1:D:115:GLN:NE2	1:D:130:ARG:HD2	1.97	0.80

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	138/148 (93%)	133 (96%)	5 (4%)	0	100 100
1	B	147/148 (99%)	142 (97%)	5 (3%)	0	100 100
1	C	137/148 (93%)	131 (96%)	4 (3%)	2 (2%)	10 8

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	D	137/148 (93%)	119 (87%)	12 (9%)	6 (4%)	2 1
1	E	137/148 (93%)	135 (98%)	2 (2%)	0	100 100
1	F	137/148 (93%)	130 (95%)	7 (5%)	0	100 100
1	G	137/148 (93%)	127 (93%)	10 (7%)	0	100 100
1	H	139/148 (94%)	127 (91%)	11 (8%)	1 (1%)	22 22
All	All	1109/1184 (94%)	1044 (94%)	56 (5%)	9 (1%)	19 19

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	23	PHE
1	D	44	SER
1	D	49	PRO
1	D	42	VAL
1	H	49	PRO

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	116/122 (95%)	114 (98%)	2 (2%)	60 74
1	B	122/122 (100%)	119 (98%)	3 (2%)	47 60
1	C	113/122 (93%)	112 (99%)	1 (1%)	78 88
1	D	112/122 (92%)	108 (96%)	4 (4%)	35 45
1	E	115/122 (94%)	112 (97%)	3 (3%)	46 58
1	F	115/122 (94%)	112 (97%)	3 (3%)	46 58
1	G	112/122 (92%)	109 (97%)	3 (3%)	44 57
1	H	117/122 (96%)	114 (97%)	3 (3%)	46 58
All	All	922/976 (94%)	900 (98%)	22 (2%)	49 62

5 of 22 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	130	ARG
1	E	110	ARG
1	H	48	GLN
1	E	48	GLN
1	E	94	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	115	GLN
1	H	30	HIS
1	F	115	GLN
1	D	30	HIS
1	F	30	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 monosaccharides 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	138/148 (93%)	-0.08	1 (0%) 87 86	21, 35, 62, 72	0
1	B	146/148 (98%)	-0.31	1 (0%) 87 86	21, 32, 51, 63	0
1	C	138/148 (93%)	0.27	3 (2%) 62 59	25, 47, 68, 79	0
1	D	138/148 (93%)	0.46	16 (11%) 4 4	28, 51, 71, 79	0
1	E	138/148 (93%)	-0.35	2 (1%) 75 73	19, 33, 50, 64	0
1	F	138/148 (93%)	-0.21	0 100 100	23, 38, 56, 67	0
1	G	137/148 (92%)	0.02	2 (1%) 73 72	28, 48, 64, 69	0
1	H	138/148 (93%)	0.13	7 (5%) 28 26	26, 44, 61, 67	0
All	All	1111/1184 (93%)	-0.01	32 (2%) 51 49	19, 40, 65, 79	0

The worst 5 of 32 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	44	SER	4.5
1	G	23[A]	PHE	4.1
1	B	145	HIS	3.9
1	C	19	ASN	3.6
1	H	1	LEU	3.4

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.