wwPDB X-ray Structure Validation Summary Report (i)

Jun 17, 2024 – 05:44 AM EDT

PDB ID	:	3HM2
Title	:	Crystal Structure of Putative Precorrin-6Y C5,15-Methyltransferase Targeted Domain from Corynebacterium diphtheriae
Authors	:	Kim, Y.; Bigelow, L.; Keigher, L.; Joachimiak, A.; Midwest Center for Structural Genomics (MCSG)
Deposited on	:	2009-05-28
Resolution	:	2.21 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

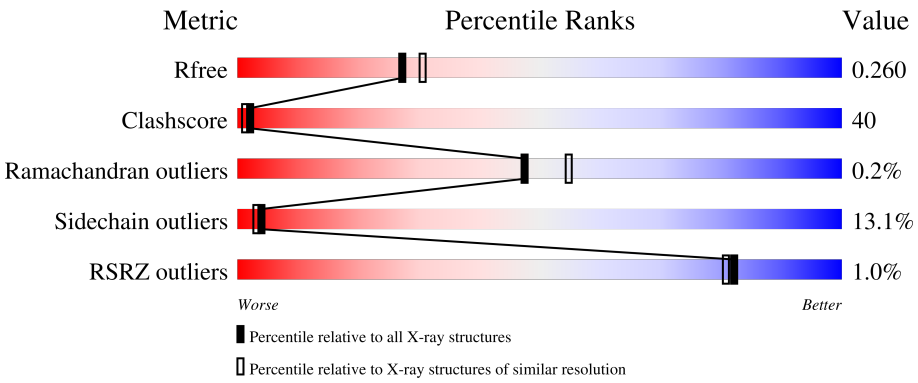
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.21 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	178	% 36% 56% · ·
1	B	178	2% 42% 49% · ·
1	C	178	4% 38% 51% 6% 5%
1	D	178	36% 51% 7% 6%
1	E	178	47% 45% · ·

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	178	43% 42% 9% 6%
1	G	178	% 44% 45% 7% •
1	H	178	% 41% 44% 8% • 6%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 10700 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 precorrin-6Y C5,15-methyltransferase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	171	1322	842	239	238	1	2	0	2	0
1	B	171	1304	831	234	236	1	2	0	0	0
1	C	169	1294	826	232	233	1	2	0	0	0
1	D	167	1279	815	230	231	1	2	0	0	0
1	E	170	1298	828	233	234	1	2	0	0	0
1	F	167	1294	826	231	234	1	2	0	1	0
1	G	171	1295	824	232	236	1	2	0	0	0
1	H	167	1282	816	232	231	1	2	0	1	0

There are 24 discrepancies between the modelled and reference sequences:

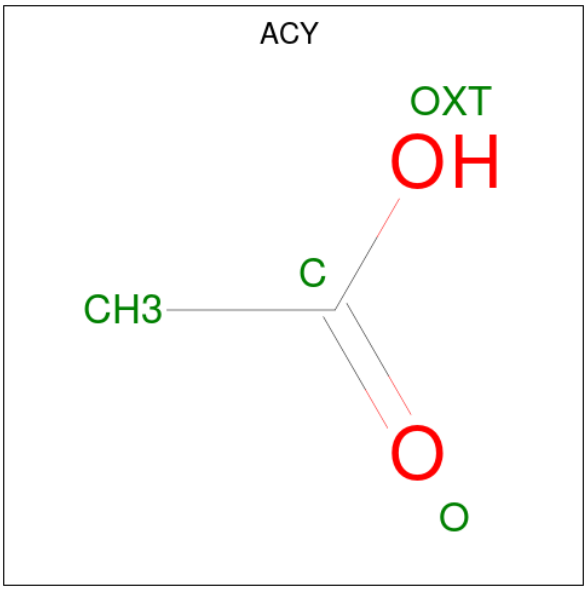
Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	SER	-	expression tag	UNP Q6NHA4
A	-1	ASN	-	expression tag	UNP Q6NHA4
A	0	ALA	-	expression tag	UNP Q6NHA4
B	-2	SER	-	expression tag	UNP Q6NHA4
B	-1	ASN	-	expression tag	UNP Q6NHA4
B	0	ALA	-	expression tag	UNP Q6NHA4
C	-2	SER	-	expression tag	UNP Q6NHA4
C	-1	ASN	-	expression tag	UNP Q6NHA4
C	0	ALA	-	expression tag	UNP Q6NHA4
D	-2	SER	-	expression tag	UNP Q6NHA4
D	-1	ASN	-	expression tag	UNP Q6NHA4
D	0	ALA	-	expression tag	UNP Q6NHA4
E	-2	SER	-	expression tag	UNP Q6NHA4

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	-1	ASN	-	expression tag	UNP Q6NHA4
E	0	ALA	-	expression tag	UNP Q6NHA4
F	-2	SER	-	expression tag	UNP Q6NHA4
F	-1	ASN	-	expression tag	UNP Q6NHA4
F	0	ALA	-	expression tag	UNP Q6NHA4
G	-2	SER	-	expression tag	UNP Q6NHA4
G	-1	ASN	-	expression tag	UNP Q6NHA4
G	0	ALA	-	expression tag	UNP Q6NHA4
H	-2	SER	-	expression tag	UNP Q6NHA4
H	-1	ASN	-	expression tag	UNP Q6NHA4
H	0	ALA	-	expression tag	UNP Q6NHA4

- Molecule 2 is ACETIC ACID (three-letter code: ACY) (formula: $C_2H_4O_2$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
2	A	1	4	2	2	0	0
			Total	C	O		
2	D	1	4	2	2	0	0
			Total	C	O		
2	F	1	4	2	2	0	0

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total Mg 1 1	0	0
3	G	1	Total Mg 1 1	0	0

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	D	1	Total Ca 1 1	0	0
4	G	1	Total Ca 1 1	0	0

- Molecule 5 is water.

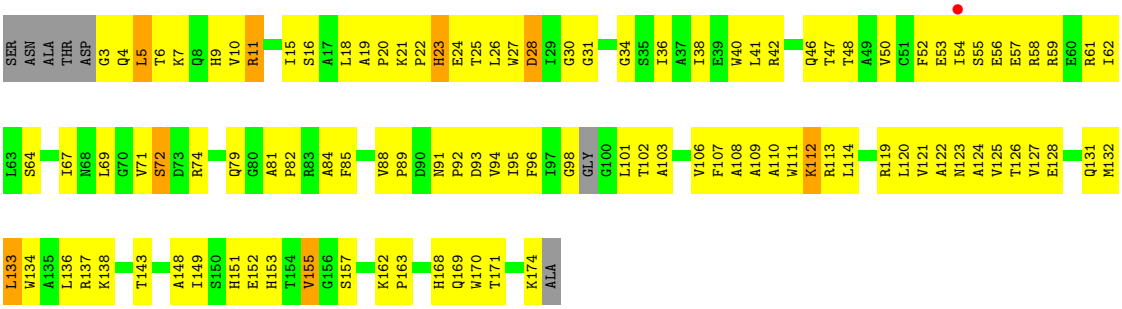
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	38	Total O 38 38	0	0
5	B	42	Total O 42 42	0	0
5	C	30	Total O 30 30	0	0
5	D	46	Total O 46 46	0	0
5	E	50	Total O 50 50	0	0
5	F	33	Total O 33 33	0	0
5	G	39	Total O 39 39	0	0
5	H	38	Total O 38 38	0	0

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.

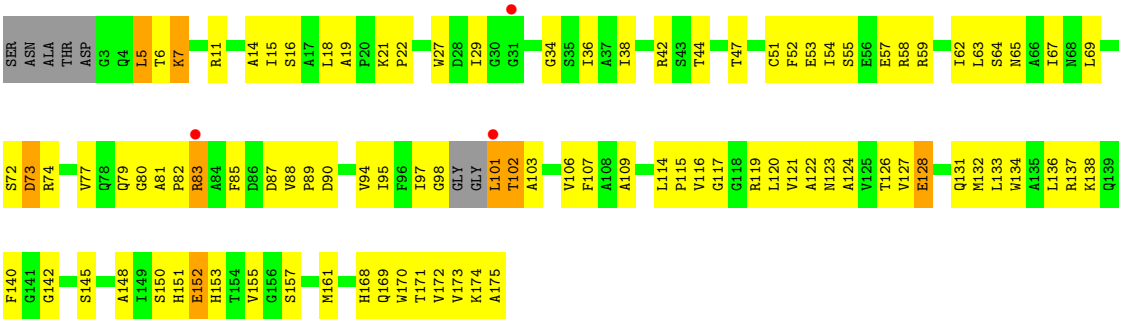
• Molecule 1: precorrin-6Y C5,15-methyltransferase

Chain A:



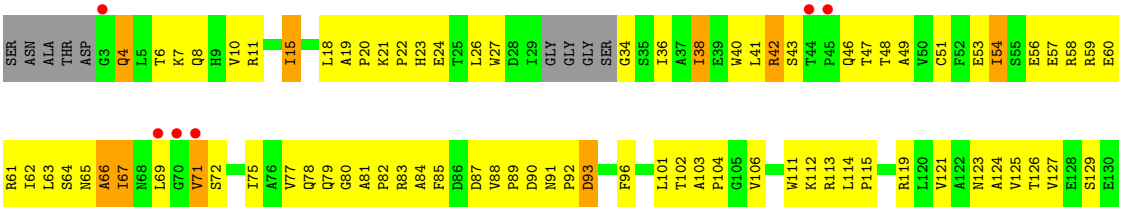
• Molecule 1: precorrin-6Y C5,15-methyltransferase

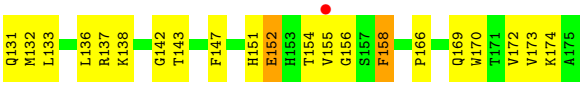
Chain B:



• Molecule 1: precorrin-6Y C5,15-methyltransferase

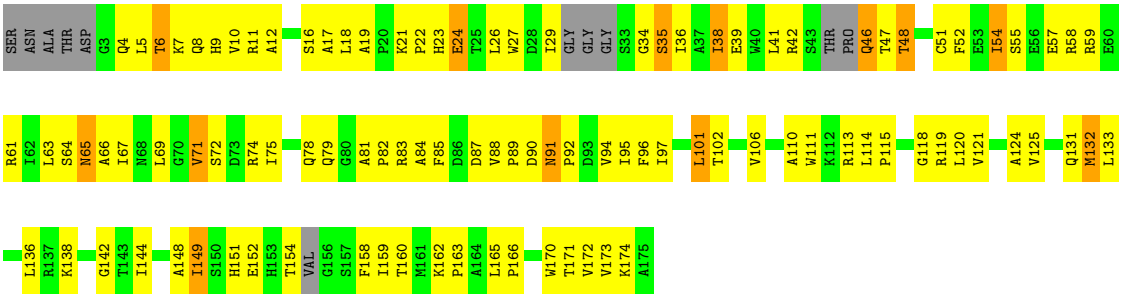
Chain C:





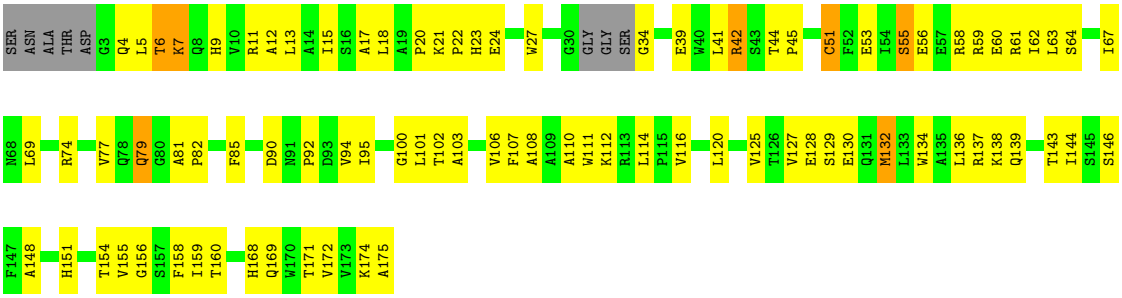
• Molecule 1: precorrin-6Y C5,15-methyltransferase

Chain D:



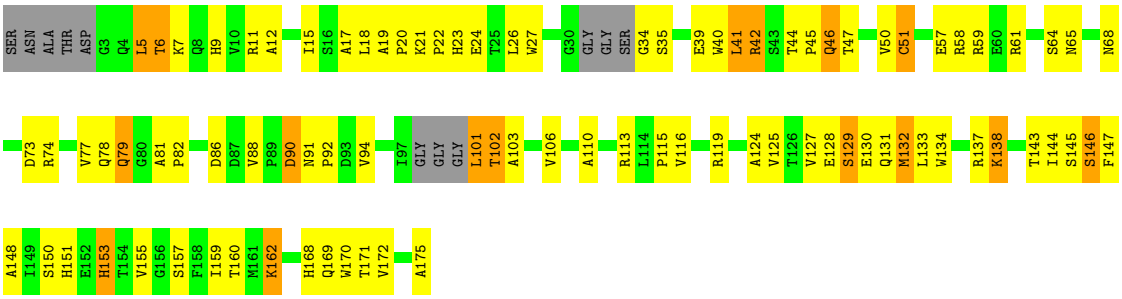
• Molecule 1: precorrin-6Y C5,15-methyltransferase

Chain E:



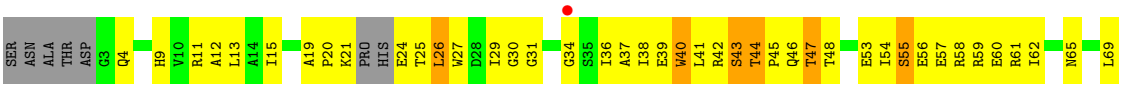
• Molecule 1: precorrin-6Y C5,15-methyltransferase

Chain F:



• Molecule 1: precorrin-6Y C5,15-methyltransferase

Chain G:



G70 V71 S72 D73 R74 V77 A81 P82 R83 A84 F85 D90 N91 V94 I95 F96 I97 G98 G99 G100 L101 T102 A103 V106 A109 K112 R113 L114 P115 R119 L120 V121 A122 N123 A124 V125 T126 Q131 M132 L133 W134 R137 K138 Q139 T143 F147 A148 I149

S150 H151 V155 F158 I159 T160 M161 A164 Q169 W170 T171 A175

• Molecule 1: precorrin-6Y C5,15-methyltransferase

Chain H:

SER ASN ALA THR ASP G3 Q4 L5 Q8 H9 V10 R11 A12 I15 S16 A17 L18 A19 P20 K21 PRO HIS E24 T25 L26 W27 G31 GLY SER G34 S35 I36 A37 I38 E39 W40 L41 R42 S43 T44 P45 Q46 T47 T48 A49 V50 C51 I54 S55 E56 E57 R58 R59 E60 S64

I67 N68 V71 R74 Q79 G80 A81 P82 R83 D86 D87 D90 N91 P92 D93 V94 I95 F96 I97 G98 G99 G100 L101 V106 R113 L114 P115 V116 G117 G118 R119 L120 V121 A122 N123 A124 V125 T126 V127 E128 S129 E130 Q131 M132 L133 W134 A135 L136 R137 K138 S145

A148 I149 S150 H151 E152 H153 THR VAL G156 I159 T160 M161 K162 P163 W170 K174 A175

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	77.78Å 113.85Å 84.65Å 90.00° 90.12° 90.00°	Depositor
Resolution (Å)	47.24 – 2.21 47.24 – 2.20	Depositor EDS
% Data completeness (in resolution range)	88.9 (47.24-2.21) 95.9 (47.24-2.20)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.17 (at 2.20Å)	Xtriage
Refinement program	PHENIX (phenix.refine)	Depositor
R, R_{free}	0.210 , 0.260 0.215 , 0.260	Depositor DCC
R_{free} test set	3621 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å^2)	36.2	Xtriage
Anisotropy	0.205	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 38.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.420 for h,-k,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	10700	wwPDB-VP
Average B, all atoms (Å^2)	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 57.44 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 2.3673e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: CA, MG, ACY

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	0.48	0/1350	0.64	1/1834 (0.1%)
1	B	0.52	0/1332	0.67	1/1810 (0.1%)
1	C	0.50	0/1322	0.63	0/1797
1	D	0.53	0/1304	0.69	0/1767
1	E	0.58	0/1326	0.69	0/1802
1	F	0.56	0/1321	0.71	1/1795 (0.1%)
1	G	0.57	0/1321	0.71	0/1793
1	H	0.53	1/1306 (0.1%)	0.66	0/1768
All	All	0.53	1/10582 (0.0%)	0.68	3/14366 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	H	57	GLU	C-N	5.86	1.47	1.34

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	F	119	ARG	NE-CZ-NH1	-6.47	117.06	120.30
1	B	133	LEU	CA-CB-CG	6.00	129.10	115.30
1	A	133	LEU	CA-CB-CG	5.28	127.45	115.30

There are no chirality outliers.

There are no planarity outliers.

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	1322	0	1334	132	0
1	B	1304	0	1313	103	0
1	C	1294	0	1305	106	0
1	D	1279	0	1285	116	0
1	E	1298	0	1308	90	0
1	F	1294	0	1301	100	0
1	G	1295	0	1305	114	0
1	H	1282	0	1291	124	0
2	A	4	0	3	0	0
2	D	4	0	3	0	0
2	F	4	0	3	1	0
3	B	1	0	0	0	0
3	G	1	0	0	0	0
4	D	1	0	0	0	0
4	G	1	0	0	0	0
5	A	38	0	0	10	0
5	B	42	0	0	1	0
5	C	30	0	0	4	0
5	D	46	0	0	8	0
5	E	50	0	0	2	0
5	F	33	0	0	2	0
5	G	39	0	0	4	0
5	H	38	0	0	2	0
All	All	10700	0	10451	838	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 40.

The worst 5 of 838 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:81:ALA:HB2	1:A:85:PHE:CE1	1.70	1.26
1:C:27:TRP:CE2	1:C:92:PRO:HG3	1.80	1.17
1:C:54:ILE:HD12	1:C:80:GLY:HA3	1.29	1.15
1:C:41:LEU:HB3	1:C:71:VAL:HG11	1.26	1.12
1:G:56:GLU:O	1:G:60:GLU:HG3	1.48	1.12

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	169/178 (95%)	149 (88%)	20 (12%)	0	100	100
1	B	167/178 (94%)	148 (89%)	19 (11%)	0	100	100
1	C	165/178 (93%)	151 (92%)	12 (7%)	2 (1%)	13	10
1	D	159/178 (89%)	147 (92%)	12 (8%)	0	100	100
1	E	166/178 (93%)	152 (92%)	14 (8%)	0	100	100
1	F	162/178 (91%)	156 (96%)	6 (4%)	0	100	100
1	G	167/178 (94%)	151 (90%)	16 (10%)	0	100	100
1	H	160/178 (90%)	143 (89%)	17 (11%)	0	100	100
All	All	1315/1424 (92%)	1197 (91%)	116 (9%)	2 (0%)	47	55

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	66	ALA
1	C	156	GLY

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	139/139 (100%)	123 (88%)	16 (12%)	5	5
1	B	137/139 (99%)	122 (89%)	15 (11%)	6	5
1	C	136/139 (98%)	116 (85%)	20 (15%)	3	2

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	D	134/139 (96%)	119 (89%)	15 (11%)	6	5
1	E	136/139 (98%)	125 (92%)	11 (8%)	11	12
1	F	137/139 (99%)	114 (83%)	23 (17%)	2	1
1	G	135/139 (97%)	117 (87%)	18 (13%)	4	3
1	H	133/139 (96%)	107 (80%)	26 (20%)	1	1
All	All	1087/1112 (98%)	943 (87%)	144 (13%)	4	3

5 of 144 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	137	ARG
1	H	159	ILE
1	H	5	LEU
1	H	64	SER
1	C	152	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 30 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	91	ASN
1	H	91	ASN
1	F	23	HIS
1	H	151	HIS
1	G	91	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 7 ligands modelled in this entry, 4 are monoatomic - leaving 3 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	ACY	D	177	-	3,3,3	0.77	0	3,3,3	1.19	0
2	ACY	F	176	-	3,3,3	0.82	0	3,3,3	0.73	0
2	ACY	A	176	-	3,3,3	0.79	0	3,3,3	1.19	0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	F	176	ACY	1	0

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 (i)

6.1 Protein, DNA and RNA chains (i)

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(Å^2)	Q<0.9
1	A	169/178 (94%)	-0.16	1 (0%)	89	88	25, 42, 67, 90	0
1	B	169/178 (94%)	-0.21	3 (1%)	68	66	25, 43, 62, 90	0
1	C	167/178 (93%)	0.11	7 (4%)	36	34	29, 47, 81, 105	0
1	D	165/178 (92%)	-0.17	0	100	100	24, 44, 68, 97	0
1	E	168/178 (94%)	-0.30	0	100	100	22, 39, 62, 94	0
1	F	165/178 (92%)	-0.34	0	100	100	23, 38, 58, 74	0
1	G	169/178 (94%)	-0.14	1 (0%)	89	88	27, 44, 66, 83	0
1	H	165/178 (92%)	-0.15	2 (1%)	79	77	28, 46, 69, 89	0
All	All	1337/1424 (93%)	-0.17	14 (1%)	82	81	22, 43, 69, 105	0

The worst 5 of 14 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	70	GLY	7.9
1	C	44	THR	7.9
1	C	69	LEU	6.0
1	C	155	VAL	5.0
1	G	34	GLY	4.4

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)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	ACY	F	176	4/4	0.82	0.23	42,54,57,71	0
2	ACY	D	177	4/4	0.93	0.18	35,45,47,49	0
2	ACY	A	176	4/4	0.93	0.11	38,43,49,55	0
3	MG	G	176	1/1	0.96	0.06	43,43,43,43	0
4	CA	D	176	1/1	0.96	0.09	64,64,64,64	0
4	CA	G	177	1/1	0.97	0.14	56,56,56,56	0
3	MG	B	176	1/1	0.98	0.12	39,39,39,39	0

6.5 Other polymers (i)

There are no such residues in this entry.