wwPDB X-ray Structure Validation Summary Report (i)

Jun 15, 2020 – 08:09 am BST

PDB ID	:	4WKG
Title	:	The crystal structure of apo ArnA features an unexpected central binding pocket and provides an explanation for enzymatic coop-erativity
Authors	:	Grimm, C.
Deposited on	:	2014-10-02
Resolution	:	2.70 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:

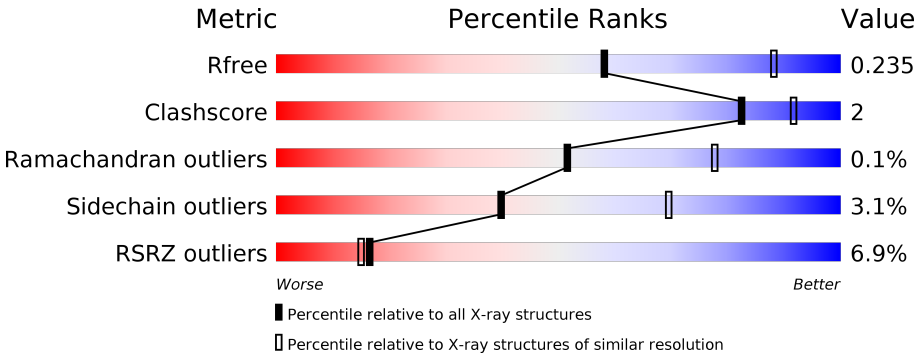
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 2.70 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	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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% 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	660	4% 90% 6%
1	B	660	6% 90% 7%
1	C	660	3% 87% 9%
1	D	660	18% 87% 10%
1	E	660	5% 88% 8%
1	F	660	5% 87% 9%

2 Entry composition

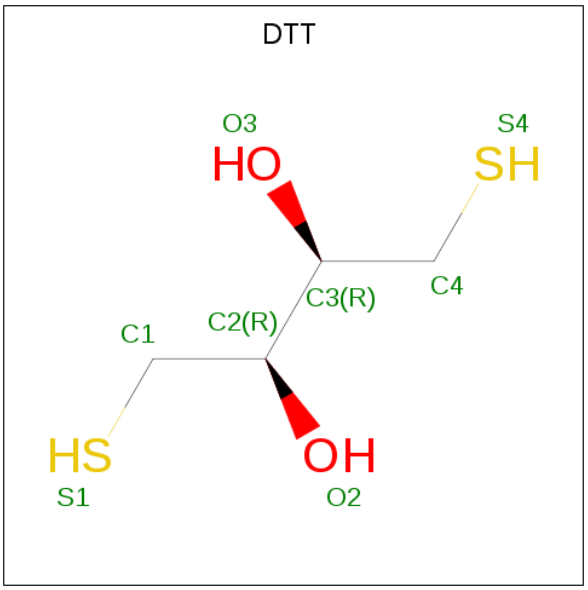
There are 4 unique types of molecules in this entry. The entry contains 60000 atoms, of which 29761 are hydrogens and 0 are deuteriums.

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 Bifunctional polymyxin resistance protein ArnA.

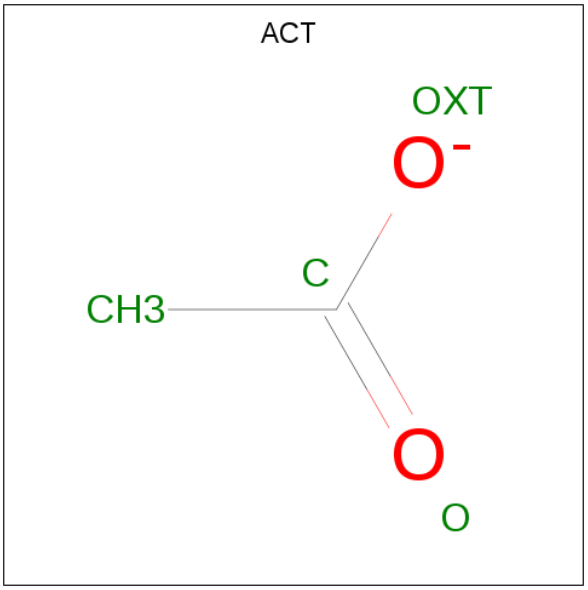
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	636	Total	C	H	N	O	S	0	0	0
			9951	3195	4945	889	903	19			
1	B	643	Total	C	H	N	O	S	0	0	0
			10036	3234	4972	897	914	19			
1	C	637	Total	C	H	N	O	S	0	0	0
			9963	3203	4947	888	906	19			
1	D	641	Total	C	H	N	O	S	0	0	0
			9973	3218	4936	889	911	19			
1	E	643	Total	C	H	N	O	S	0	0	0
			10032	3231	4974	894	914	19			
1	F	637	Total	C	H	N	O	S	0	0	0
			9950	3197	4941	888	905	19			

- Molecule 2 is 2,3-DIHYDROXY-1,4-DITHIOBUTANE (three-letter code: DTT) (formula: $C_4H_{10}O_2S_2$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	H	O	S	0	1
			36	8	20	4	4		
2	D	1	Total	C	H	O	S	0	1
			36	8	20	4	4		

- Molecule 3 is ACETATE ION (three-letter code: ACT) (formula: $C_2H_3O_2$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	H	O	0	0
			7	2	3	2		
3	F	1	Total	C	H	O	0	0
			7	2	3	2		

- Molecule 4 is water.

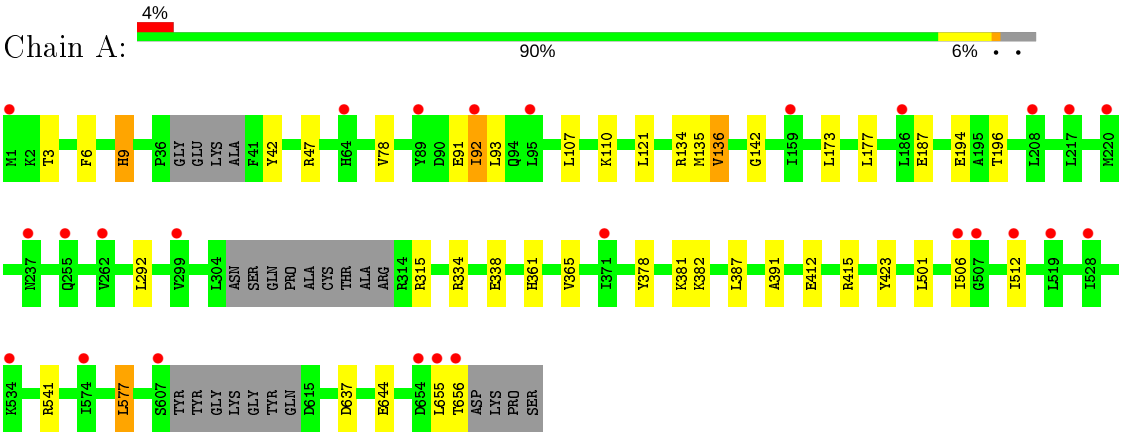
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	O	0	0
			1	1		
4	B	2	Total	O	0	0
			2	2		
4	C	2	Total	O	0	0
			2	2		
4	D	1	Total	O	0	0
			1	1		
4	E	2	Total	O	0	0
			2	2		
4	F	1	Total	O	0	0
			1	1		

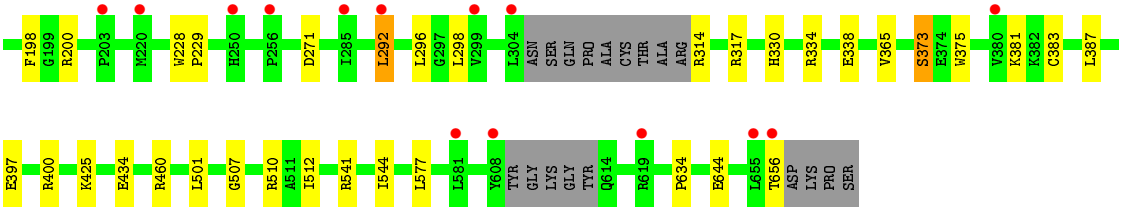
3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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.

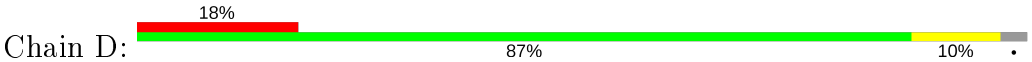
- Molecule 1: Bifunctional polymyxin resistance protein ArnA

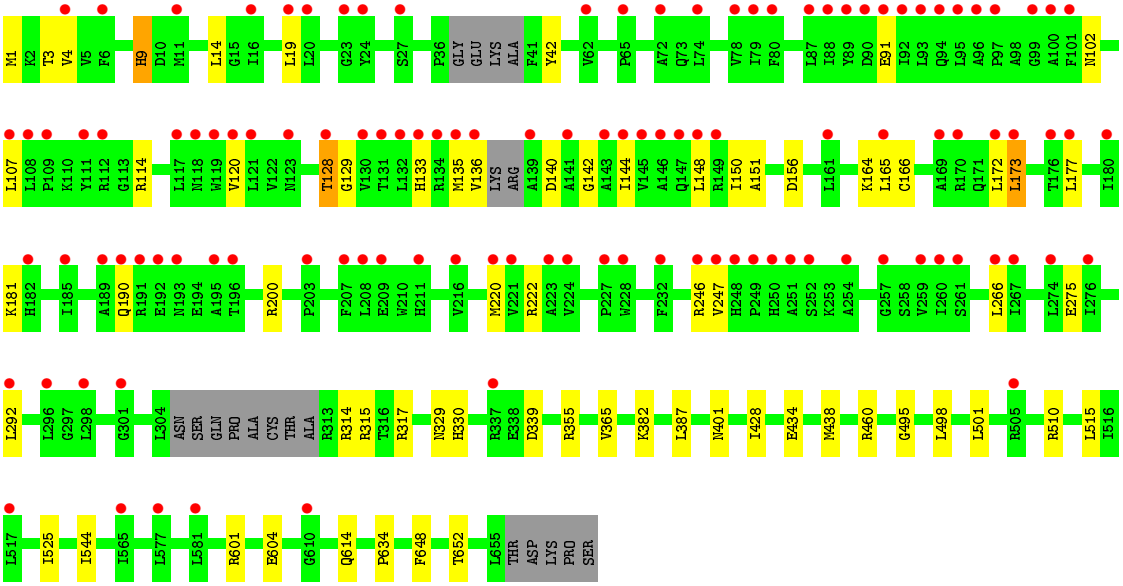
Chain A:



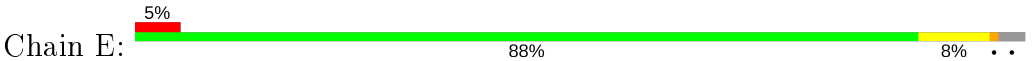


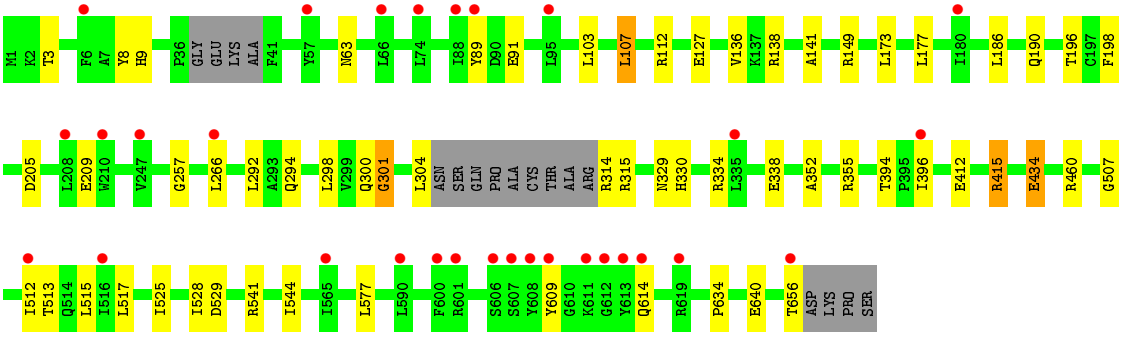
• Molecule 1: Bifunctional polymyxin resistance protein ArnA



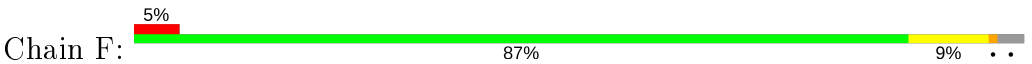


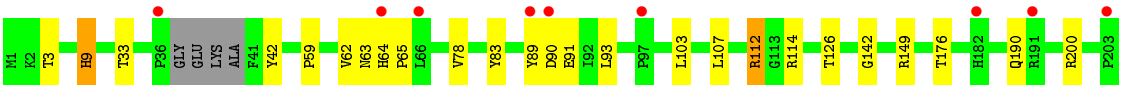
• Molecule 1: Bifunctional polymyxin resistance protein ArnA

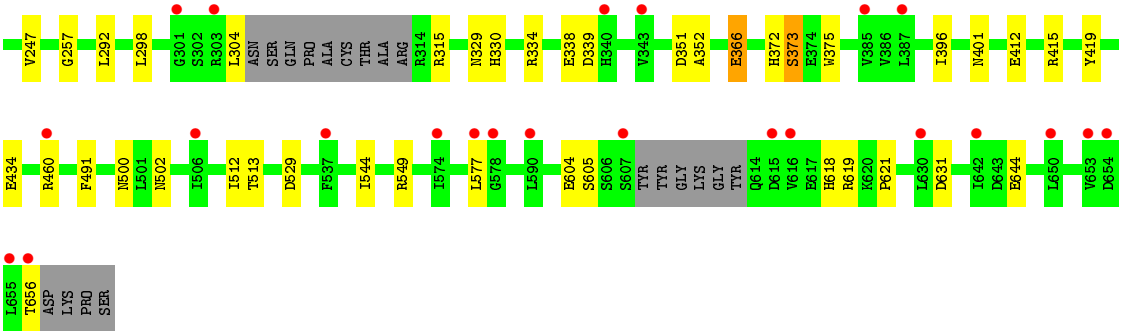




• Molecule 1: Bifunctional polymyxin resistance protein ArnA






V247
G257
L292
L298
G301
S302
R303
L304
ASN
SER
GLN
PRO
ALA
CYS
THR
ALA
ARG
R314
R315
N329
H330
R334
E338
D339
H340
V343
D351
A352
E366
H372
S373
E374
W375
V385
V386
L387
I396
N401
E412
R415
Y419
E434
R460
F491
N500
L501
N502
I506
I512
T513
D529
F537
I544
R549
I574
L577
G578
L590
E604
S605
S606
S607
TYR
TYR
GLY
LYS
GLY
TYR
Q614
D615
V616
E617
H618
R619
K620
P621
L630
D631
I642
D643
E644
L650
V653
D654
L655
T656
ASP
LYS
PRO
SER

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	106.72Å 112.82Å 113.61Å 81.92° 82.96° 83.80°	Depositor
Resolution (Å)	48.16 – 2.70 48.16 – 2.70	Depositor EDS
% Data completeness (in resolution range)	97.4 (48.16-2.70) 90.0 (48.16-2.70)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.08 (at 2.69Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.2_1309)	Depositor
R, R_{free}	0.205 , 0.232 0.208 , 0.235	Depositor DCC
R_{free} test set	6967 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	69.0	Xtriage
Anisotropy	0.218	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 37.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.014 for -h,-l,-k	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	60000	wwPDB-VP
Average B, all atoms (Å^2)	63.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.50% 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

Bond lengths and bond angles in the following residue types are not validated in this section: ACT, DTT

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	0.29	0/5126	0.52	1/6965 (0.0%)
1	B	0.30	0/5188	0.54	1/7051 (0.0%)
1	C	0.27	0/5137	0.51	1/6981 (0.0%)
1	D	0.29	0/5160	0.53	0/7013
1	E	0.29	0/5182	0.53	2/7044 (0.0%)
1	F	0.29	0/5129	0.52	0/6970
All	All	0.29	0/30922	0.53	5/42024 (0.0%)

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	E	301	GLY	C-N-CA	6.51	137.98	121.70
1	B	507	GLY	N-CA-C	5.87	127.77	113.10
1	E	507	GLY	N-CA-C	5.35	126.47	113.10
1	A	577	LEU	CB-CG-CD2	-5.18	102.19	111.00
1	C	507	GLY	N-CA-C	5.10	125.85	113.10

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	5006	4945	4933	20	0
1	B	5064	4972	4977	23	0
1	C	5016	4947	4937	26	0
1	D	5037	4936	4936	33	0
1	E	5058	4974	4966	25	0
1	F	5009	4941	4930	29	0
2	A	16	20	20	0	0
2	D	16	20	20	1	0
3	A	4	3	3	0	0
3	F	4	3	3	1	0
4	A	1	0	0	0	0
4	B	2	0	0	0	0
4	C	2	0	0	0	0
4	D	1	0	0	0	0
4	E	2	0	0	0	0
4	F	1	0	0	0	0
All	All	30239	29761	29725	150	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 2.

The worst 5 of 150 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:D:434:GLU:OE1	1:D:460:ARG:NH2	2.16	0.77
1:A:382:LYS:NZ	1:B:366:GLU:OE2	2.19	0.76
1:B:434:GLU:HG2	1:B:619:ARG:HH22	1.50	0.75
1:F:500:ASN:ND2	1:F:502:ASN:OD1	2.21	0.74
1:F:415:ARG:NH2	1:F:419:TYR:OH	2.21	0.73

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	628/660 (95%)	613 (98%)	14 (2%)	1 (0%)	47 73
1	B	637/660 (96%)	623 (98%)	13 (2%)	1 (0%)	47 73
1	C	629/660 (95%)	617 (98%)	12 (2%)	0	100 100
1	D	633/660 (96%)	619 (98%)	14 (2%)	0	100 100
1	E	637/660 (96%)	619 (97%)	17 (3%)	1 (0%)	47 73
1	F	629/660 (95%)	618 (98%)	11 (2%)	0	100 100
All	All	3793/3960 (96%)	3709 (98%)	81 (2%)	3 (0%)	51 78

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	136	VAL
1	E	136	VAL
1	B	136	VAL

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	528/559 (94%)	515 (98%)	13 (2%)	47 76
1	B	532/559 (95%)	517 (97%)	15 (3%)	43 73
1	C	529/559 (95%)	513 (97%)	16 (3%)	41 70
1	D	528/559 (94%)	514 (97%)	14 (3%)	44 74
1	E	531/559 (95%)	513 (97%)	18 (3%)	37 66
1	F	528/559 (94%)	507 (96%)	21 (4%)	31 60
All	All	3176/3354 (95%)	3079 (97%)	97 (3%)	40 69

5 of 97 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	656	THR
1	D	365	VAL

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	F	373	SER
1	D	4	VAL
1	D	173	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	118	ASN
1	E	190	GLN
1	F	64	HIS
1	D	104	HIS
1	F	9	HIS

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

6 ligands are modelled in this entry.

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	DTT	D	701[B]	-	7,7,7	0.57	0	4,8,8	0.29	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\|>2$	Counts	RMSZ	$\#\|Z\|>2$
3	ACT	F	701	-	1,3,3	0.48	0	0,3,3	0.00	-
3	ACT	A	702	-	1,3,3	1.51	0	0,3,3	0.00	-
2	DTT	A	701[A]	-	7,7,7	0.72	0	4,8,8	0.53	0
2	DTT	A	701[B]	-	7,7,7	0.61	0	4,8,8	0.35	0
2	DTT	D	701[A]	-	7,7,7	0.60	0	4,8,8	0.32	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	DTT	D	701[B]	-	-	1/8/8/8	-
2	DTT	A	701[A]	-	-	2/8/8/8	-
2	DTT	A	701[B]	-	-	4/8/8/8	-
2	DTT	D	701[A]	-	-	1/8/8/8	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	D	701[B]	DTT	S1-C1-C2-O2
2	A	701[A]	DTT	S1-C1-C2-O2
2	A	701[A]	DTT	S1-C1-C2-C3
2	A	701[B]	DTT	S1-C1-C2-O2
2	A	701[B]	DTT	S1-C1-C2-C3

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	D	701[B]	DTT	1	0
3	F	701	ACT	1	0

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues

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	636/660 (96%)	0.43	26 (4%)	37	36	19, 51, 96, 152	0
1	B	643/660 (97%)	0.50	39 (6%)	21	20	20, 56, 105, 170	0
1	C	637/660 (96%)	0.43	23 (3%)	42	42	27, 55, 103, 158	0
1	D	641/660 (97%)	0.98	116 (18%)	1	1	20, 71, 157, 204	0
1	E	643/660 (97%)	0.41	30 (4%)	31	30	16, 49, 98, 147	0
1	F	637/660 (96%)	0.47	32 (5%)	28	27	30, 60, 100, 146	0
All	All	3837/3960 (96%)	0.54	266 (6%)	16	15	16, 55, 121, 204	0

The worst 5 of 266 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	79	ILE	12.9
1	D	100	ALA	12.2
1	D	101	PHE	8.7
1	D	136	VAL	8.0
1	E	608	TYR	7.8

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

There are no carbohydrates in this entry.

6.4 Ligands

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, 95^{th} 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	DTT	A	701[A]	8/8	0.80	0.35	32,55,55,55	18
2	DTT	A	701[B]	8/8	0.80	0.35	55,55,55,55	18
2	DTT	D	701[B]	8/8	0.81	0.28	46,64,64,64	18
2	DTT	D	701[A]	8/8	0.81	0.28	35,64,64,64	18
3	ACT	F	701	4/4	0.86	0.28	20,20,58,58	0
3	ACT	A	702	4/4	0.95	0.42	20,20,44,44	0

6.5 Other polymers

There are no such residues in this entry.