



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 21, 2020 – 02:17 PM BST

PDB ID : 4MGR  
Title : The crystal structure of Bacillus subtilis GabR, an autorepressor and PLP- and GABA-dependent transcriptional activator of gabT  
Authors : Wu, R.; Edayathumangalam, R.; Garcia, R.; Wang, Y.; Wang, W.; Kreinbring, C.A.; Bach, A.; Liao, J.; Stone, T.; Terwilliger, T.; Hoang, Q.Q.; Belitsky, B.R.; Petsko, G.A.; Ringe, D.; Liu, D.  
Deposited on : 2013-08-28  
Resolution : 2.55 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](mailto: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 ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) 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 : 2.13.1  
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) : 2.13.1

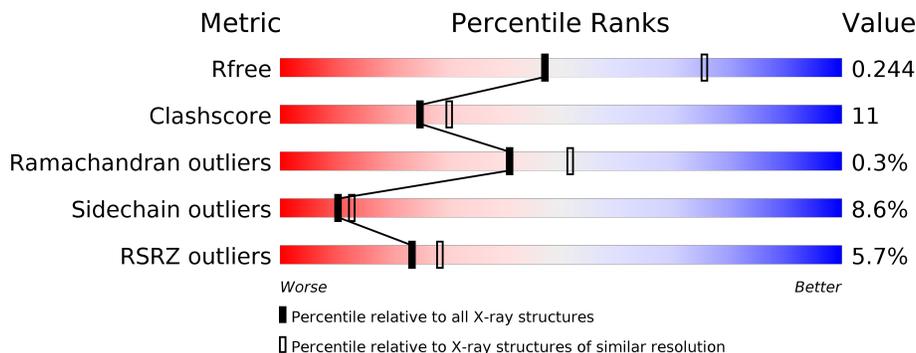
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.55 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	1284 (2.56-2.52)
Clashscore	141614	1332 (2.56-2.52)
Ramachandran outliers	138981	1315 (2.56-2.52)
Sidechain outliers	138945	1315 (2.56-2.52)
RSRZ outliers	127900	1272 (2.56-2.52)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	485	
1	B	485	
1	C	485	
1	D	485	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	ACT	B	503	-	-	X	-

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 15770 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 HTH-type transcriptional regulatory protein GabR.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	468	3800	2413	656	714	17	0	0	0
1	B	463	3757	2383	652	705	17	0	0	0
1	C	475	3855	2449	667	721	18	0	0	0
1	D	472	3831	2432	663	718	18	0	0	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	480	HIS	-	EXPRESSION TAG	UNP P94426
A	481	HIS	-	EXPRESSION TAG	UNP P94426
A	482	HIS	-	EXPRESSION TAG	UNP P94426
A	483	HIS	-	EXPRESSION TAG	UNP P94426
A	484	HIS	-	EXPRESSION TAG	UNP P94426
A	485	HIS	-	EXPRESSION TAG	UNP P94426
B	480	HIS	-	EXPRESSION TAG	UNP P94426
B	481	HIS	-	EXPRESSION TAG	UNP P94426
B	482	HIS	-	EXPRESSION TAG	UNP P94426
B	483	HIS	-	EXPRESSION TAG	UNP P94426
B	484	HIS	-	EXPRESSION TAG	UNP P94426
B	485	HIS	-	EXPRESSION TAG	UNP P94426
C	480	HIS	-	EXPRESSION TAG	UNP P94426
C	481	HIS	-	EXPRESSION TAG	UNP P94426
C	482	HIS	-	EXPRESSION TAG	UNP P94426
C	483	HIS	-	EXPRESSION TAG	UNP P94426
C	484	HIS	-	EXPRESSION TAG	UNP P94426
C	485	HIS	-	EXPRESSION TAG	UNP P94426
D	480	HIS	-	EXPRESSION TAG	UNP P94426
D	481	HIS	-	EXPRESSION TAG	UNP P94426
D	482	HIS	-	EXPRESSION TAG	UNP P94426

*Continued on next page...*

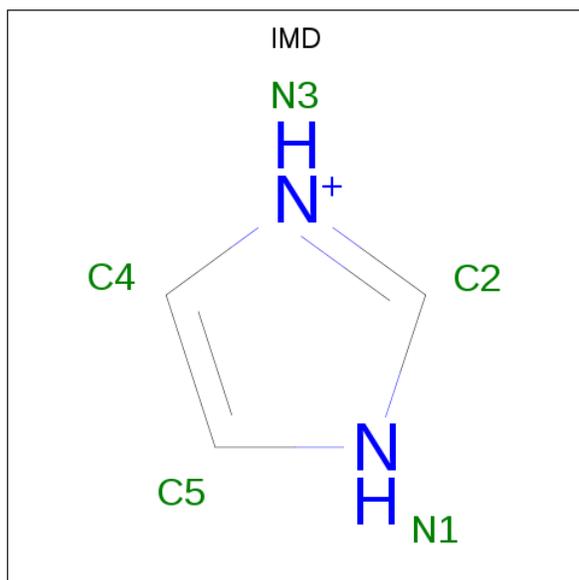
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	483	HIS	-	EXPRESSION TAG	UNP P94426
D	484	HIS	-	EXPRESSION TAG	UNP P94426
D	485	HIS	-	EXPRESSION TAG	UNP P94426

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

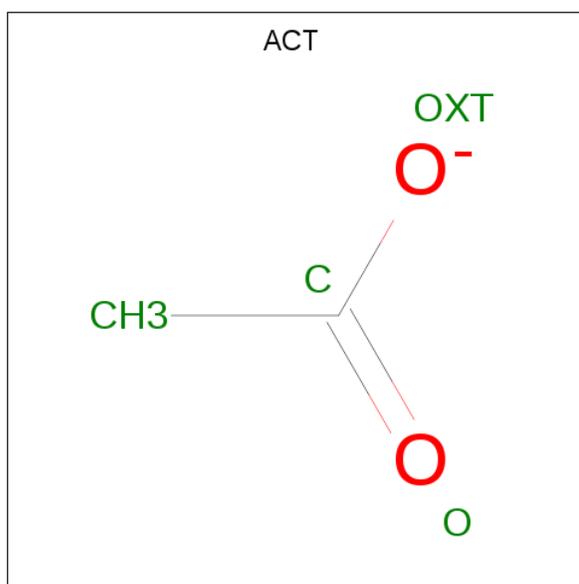
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	1	Total Zn 1 1	0	0
2	A	1	Total Zn 1 1	0	0
2	D	1	Total Zn 1 1	0	0
2	C	1	Total Zn 1 1	0	0

- Molecule 3 is IMIDAZOLE (three-letter code: IMD) (formula: C<sub>3</sub>H<sub>5</sub>N<sub>2</sub>).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C N 5 3 2	0	0
3	B	1	Total C N 5 3 2	0	0
3	C	1	Total C N 5 3 2	0	0
3	D	1	Total C N 5 3 2	0	0

- Molecule 4 is ACETATE ION (three-letter code: ACT) (formula: C<sub>2</sub>H<sub>3</sub>O<sub>2</sub>).



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

- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	101	Total O 101 101	0	0
5	B	102	Total O 102 102	0	0

*Continued on next page...*

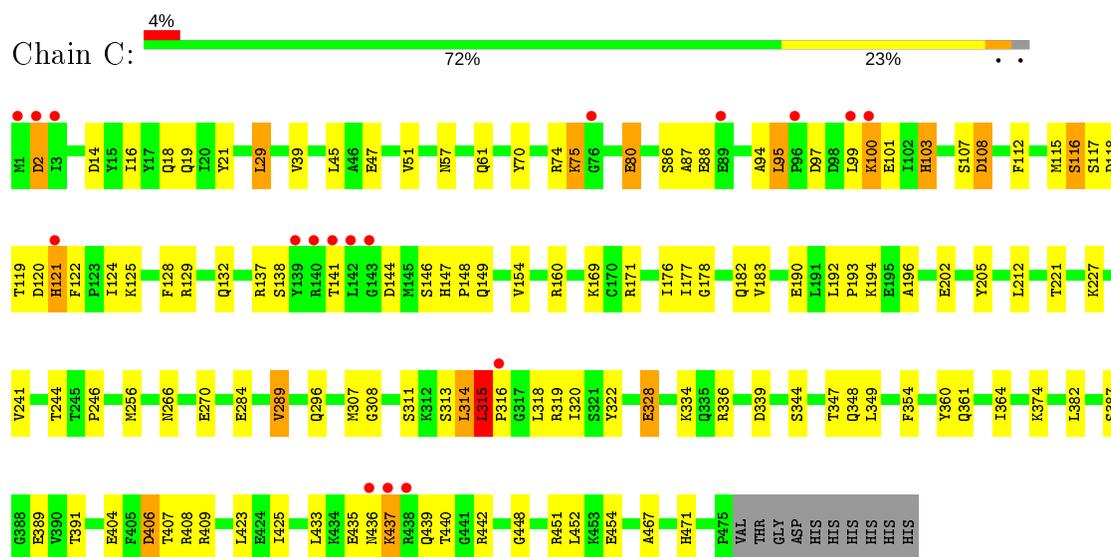
*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
5	C	134	Total 134	O 134	0	0
5	D	130	Total 130	O 130	0	0

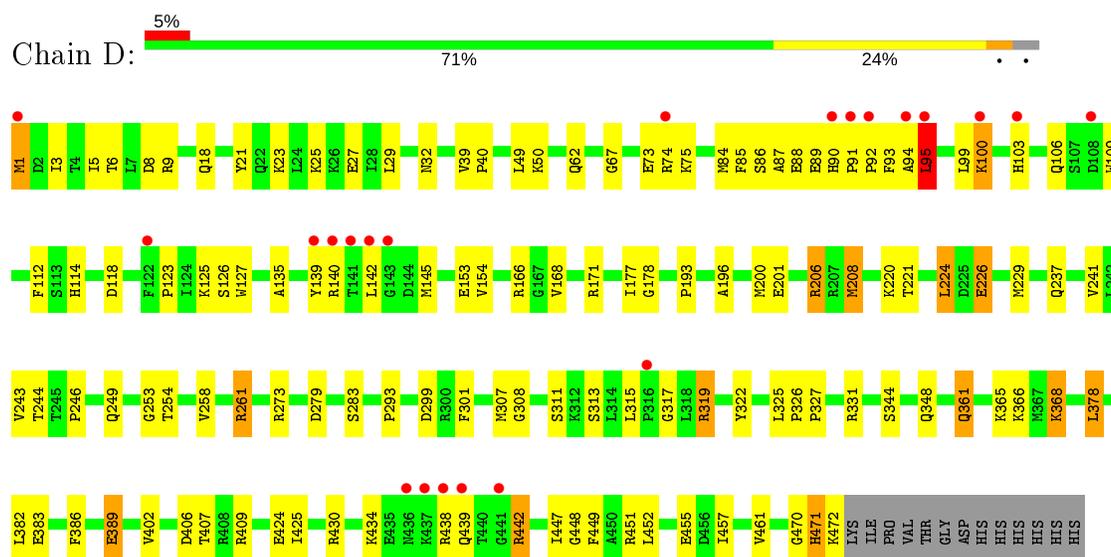


K472  
 LYS  
 ILE  
 PRO  
 VAL  
 THR  
 GLY  
 ASP  
 HIS  
 HIS  
 HIS  
 HIS  
 HIS  
 HIS

• Molecule 1: HTH-type transcriptional regulatory protein GabR



• Molecule 1: HTH-type transcriptional regulatory protein GabR



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	99.93Å 101.36Å 213.20Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.30 – 2.55 49.31 – 2.35	Depositor EDS
% Data completeness (in resolution range)	95.6 (49.30-2.55) 87.7 (49.31-2.35)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.15 (at 2.34Å)	Xtrriage
Refinement program	PHENIX (phenix.refine: 1.8.2_1309)	Depositor
R, $R_{free}$	0.194 , 0.244 0.194 , 0.244	Depositor DCC
$R_{free}$ test set	4245 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	39.3	Xtrriage
Anisotropy	0.470	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 53.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	0.025 for k,h,-l	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	15770	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	46.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.76% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  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: ZN, IMD, ACT

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	0.27	0/3876	0.55	2/5227 (0.0%)
1	B	0.26	0/3829	0.52	0/5157
1	C	0.27	0/3933	0.54	1/5302 (0.0%)
1	D	0.26	0/3908	0.52	1/5268 (0.0%)
All	All	0.26	0/15546	0.53	4/20954 (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	A	0	2
1	B	0	3
1	C	0	6
All	All	0	11

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	74	ARG	N-CA-C	-5.88	95.14	111.00
1	A	435	GLU	N-CA-C	5.71	126.40	111.00
1	D	95	LEU	CA-CB-CG	5.66	128.32	115.30
1	C	315	LEU	C-N-CD	5.03	138.96	128.40

There are no chirality outliers.

5 of 11 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	435	GLU	Peptide
1	A	436	ASN	Peptide
1	B	314	LEU	Peptide
1	B	97	ASP	Peptide
1	B	99	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	3800	0	3810	96	0
1	B	3757	0	3774	88	0
1	C	3855	0	3876	89	0
1	D	3831	0	3845	79	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
3	A	5	0	5	0	0
3	B	5	0	5	0	0
3	C	5	0	5	1	0
3	D	5	0	5	0	0
4	A	4	0	3	0	0
4	B	16	0	12	3	0
4	C	8	0	6	1	0
4	D	8	0	6	0	0
5	A	101	0	0	12	0
5	B	102	0	0	11	0
5	C	134	0	0	19	0
5	D	130	0	0	18	0
All	All	15770	0	15352	331	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 11.

The worst 5 of 331 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:86:SER:O	5:A:664:HOH:O	1.90	0.91
1:C:171:ARG:NH1	5:C:730:HOH:O	2.03	0.90
4:B:503:ACT:O	5:B:663:HOH:O	1.91	0.88
1:B:75:LYS:HB3	1:B:76:GLY:HA2	1.54	0.88
1:C:284:GLU:O	5:C:606:HOH:O	1.91	0.88

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	466/485 (96%)	434 (93%)	30 (6%)	2 (0%)	34	46
1	B	459/485 (95%)	425 (93%)	33 (7%)	1 (0%)	47	60
1	C	473/485 (98%)	443 (94%)	29 (6%)	1 (0%)	47	60
1	D	470/485 (97%)	435 (93%)	34 (7%)	1 (0%)	47	60
All	All	1868/1940 (96%)	1737 (93%)	126 (7%)	5 (0%)	41	51

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	98	ASP
1	A	435	GLU
1	A	74	ARG
1	C	88	GLU
1	D	95	LEU

### 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	416/431 (96%)	378 (91%)	38 (9%)	9	11
1	B	411/431 (95%)	380 (92%)	31 (8%)	13	17
1	C	422/431 (98%)	386 (92%)	36 (8%)	10	13
1	D	419/431 (97%)	380 (91%)	39 (9%)	9	10
All	All	1668/1724 (97%)	1524 (91%)	144 (9%)	10	13

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

Mol	Chain	Res	Type
1	B	437	LYS
1	C	116	SER
1	D	366	LYS
1	B	462	GLN
1	C	47	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	12	GLN
1	B	352	GLN
1	C	182	GLN
1	C	369	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 monosaccharides in this entry.

## 5.6 Ligand geometry

Of 17 ligands modelled in this entry, 4 are monoatomic - leaving 13 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
4	ACT	C	503	-	1,3,3	0.98	0	0,3,3	0.00	-
4	ACT	D	503	-	1,3,3	0.96	0	0,3,3	0.00	-
4	ACT	B	506	-	1,3,3	1.13	0	0,3,3	0.00	-
4	ACT	D	504	-	1,3,3	1.24	0	0,3,3	0.00	-
3	IMD	C	502	-	3,5,5	0.41	0	4,5,5	0.60	0
4	ACT	B	504	-	1,3,3	1.21	0	0,3,3	0.00	-
4	ACT	B	505	-	1,3,3	1.22	0	0,3,3	0.00	-
3	IMD	A	502	-	3,5,5	0.40	0	4,5,5	0.58	0
3	IMD	B	502	-	3,5,5	0.40	0	4,5,5	0.59	0
4	ACT	C	504	-	1,3,3	1.24	0	0,3,3	0.00	-
3	IMD	D	502	-	3,5,5	0.40	0	4,5,5	0.60	0
4	ACT	A	503	-	1,3,3	0.98	0	0,3,3	0.00	-
4	ACT	B	503	-	1,3,3	1.23	0	0,3,3	0.00	-

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
3	IMD	C	502	-	-	-	0/1/1/1
3	IMD	A	502	-	-	-	0/1/1/1
3	IMD	D	502	-	-	-	0/1/1/1
3	IMD	B	502	-	-	-	0/1/1/1

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.

4 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	C	503	ACT	1	0
3	C	502	IMD	1	0
4	B	505	ACT	1	0
4	B	503	ACT	2	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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	468/485 (96%)	0.28	38 (8%) 12 15	20, 49, 102, 126	0
1	B	463/485 (95%)	0.21	29 (6%) 20 23	15, 42, 98, 143	0
1	C	475/485 (97%)	0.02	18 (3%) 40 47	16, 36, 93, 137	0
1	D	472/485 (97%)	0.04	22 (4%) 31 38	13, 36, 91, 123	0
All	All	1878/1940 (96%)	0.14	107 (5%) 23 28	13, 40, 97, 143	0

The worst 5 of 107 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	142	LEU	8.6
1	A	387	SER	8.0
1	D	141	THR	7.8
1	A	99	LEU	7.0
1	B	139	TYR	5.4

### 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, 95<sup>th</sup> 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( $\text{\AA}^2$ )	Q<0.9
4	ACT	D	504	4/4	0.53	0.26	55,62,62,65	0
2	ZN	C	501	1/1	0.78	0.14	78,78,78,78	0
4	ACT	B	505	4/4	0.85	0.20	45,57,58,63	0
2	ZN	B	501	1/1	0.86	0.18	88,88,88,88	0
4	ACT	B	504	4/4	0.88	0.13	24,42,43,50	0
4	ACT	C	504	4/4	0.88	0.14	51,58,58,61	0
3	IMD	A	502	5/5	0.89	0.15	41,41,43,43	0
4	ACT	A	503	4/4	0.91	0.15	28,29,33,44	0
4	ACT	B	503	4/4	0.91	0.30	27,29,37,49	0
3	IMD	C	502	5/5	0.93	0.13	27,31,37,40	0
4	ACT	C	503	4/4	0.93	0.15	26,38,42,44	0
2	ZN	D	501	1/1	0.94	0.11	81,81,81,81	0
4	ACT	D	503	4/4	0.95	0.13	38,42,43,44	0
4	ACT	B	506	4/4	0.96	0.14	60,61,62,64	0
3	IMD	B	502	5/5	0.96	0.11	17,20,22,26	0
3	IMD	D	502	5/5	0.98	0.13	28,29,37,38	0
2	ZN	A	501	1/1	0.99	0.09	63,63,63,63	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.