



# Full wwPDB X-ray Structure Validation Report

Feb 28, 2014 – 05:49 PM GMT

PDB ID : 3EN9  
Title : Structure of the Methanococcus jannaschii KAE1-BUD32 fusion protein  
Authors : Neculai, D.  
Deposited on : 2008-09-25  
Resolution : 2.67 Å(reported)

This is a full wwPDB validation 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 <http://wwpdb.org/ValidationPDFNotes.html>

---

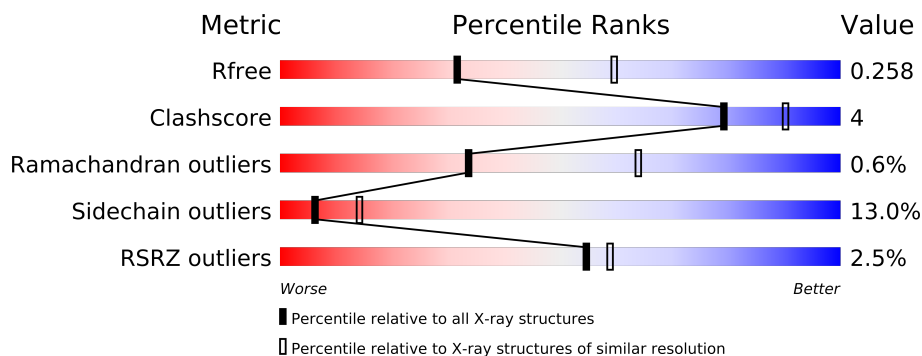
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.15 2013  
Xtriage (Phenix) : dev-1323  
EDS : stable22639  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable22683

# 1 Overall quality at a glance

The reported resolution of this entry is 2.67 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))
$R_{free}$	66092	2010 (2.70-2.66)
Clashscore	79885	2450 (2.70-2.66)
Ramachandran outliers	78287	2410 (2.70-2.66)
Sidechain outliers	78261	2410 (2.70-2.66)
RSRZ outliers	66119	2013 (2.70-2.66)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	540	
1	B	540	

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 8306 atoms, of which 0 are hydrogen and 0 are deuterium.

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 O-sialoglycoprotein endopeptidase/proteinkinase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	519	Total	C	N	O	S	0	0	0
			4130	2639	697	772	22			
1	B	511	Total	C	N	O	S	0	0	0
			4055	2591	681	762	21			

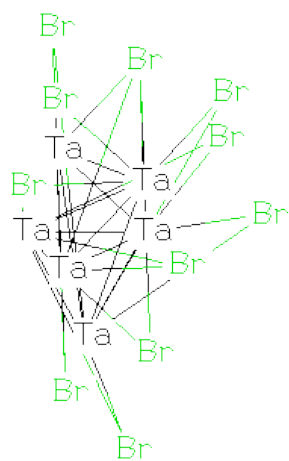
There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-4	GLY	-	EXPRESSION TAG	UNP Q58530
A	-3	ALA	-	EXPRESSION TAG	UNP Q58530
A	-2	MET	-	EXPRESSION TAG	UNP Q58530
A	-1	ASP	-	EXPRESSION TAG	UNP Q58530
A	0	PRO	-	EXPRESSION TAG	UNP Q58530
B	-4	GLY	-	EXPRESSION TAG	UNP Q58530
B	-3	ALA	-	EXPRESSION TAG	UNP Q58530
B	-2	MET	-	EXPRESSION TAG	UNP Q58530
B	-1	ASP	-	EXPRESSION TAG	UNP Q58530
B	0	PRO	-	EXPRESSION TAG	UNP Q58530

- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	1	Total	Mg	0	0
			1	1		
2	A	1	Total	Mg	0	0
			1	1		

- Molecule 3 is HEXATANTALUM DODECABROMIDE (three-letter code: TBR) (formula: Br<sub>12</sub>Ta<sub>6</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	B	1	Total	Br	Ta	0	0
			18	12	6		
3	A	1	Total	Br	Ta	0	0
			10	7	3		

- Molecule 4 is water.

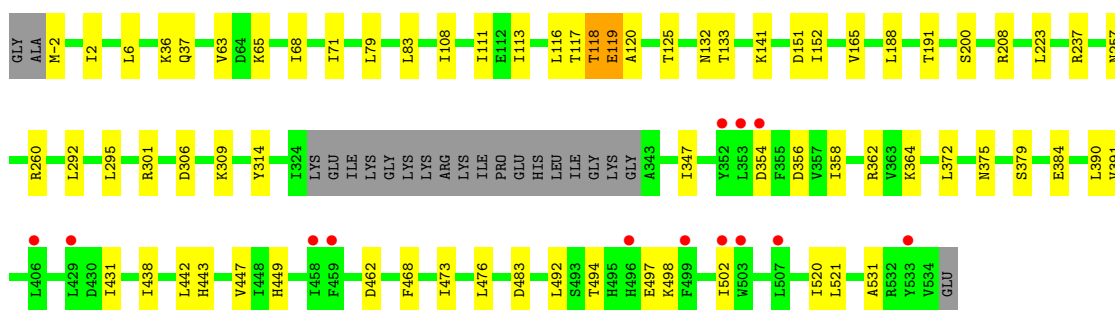
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	48	Total	O	0	0
			48	48		
4	B	43	Total	O	0	0
			43	43		

### 3 Residue-property plots

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

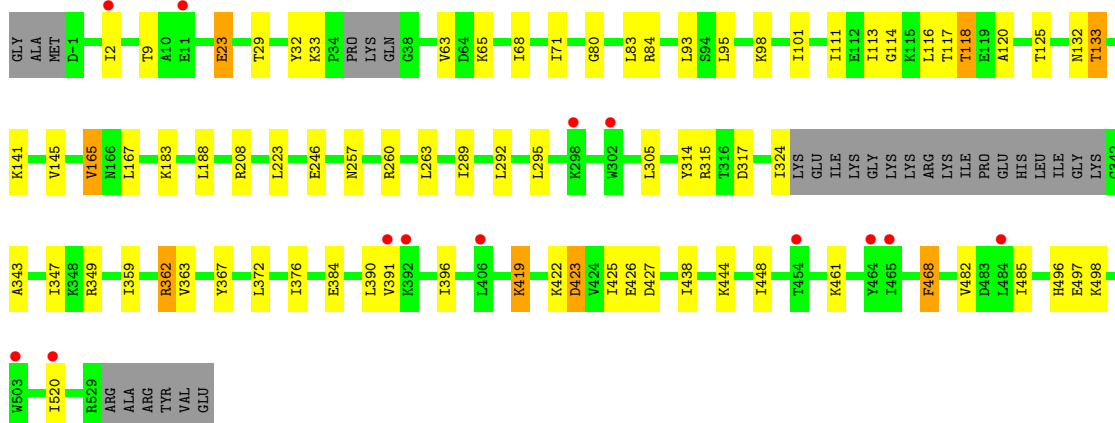
- Molecule 1: O-sialoglycoprotein endopeptidase/proteinkinase

Chain A: 



- Molecule 1: O-sialoglycoprotein endopeptidase/proteinkinase

Chain B: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	148.73Å 148.73Å 136.42Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	65.29 – 2.67 105.17 – 2.67	Depositor EDS
% Data completeness (in resolution range)	99.9 (65.29-2.67) 99.9 (105.17-2.67)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	55.01 (at 2.65Å)	Xtriage
Refinement program	PHENIX (phenix.refine)	Depositor
R, $R_{free}$	0.202 , 0.264 0.197 , 0.258	Depositor DCC
$R_{free}$ test set	2211 reflections (5.03%)	DCC
Wilson B-factor (Å <sup>2</sup> )	69.0	Xtriage
Anisotropy	0.056	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 51.6	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.44$ , $\langle L^2 \rangle = 0.27$	Xtriage
Outliers	0 of 43970 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	8306	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	71.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.77% of the height of the origin peak. No significant pseudotranslation is detected.*

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

## 5 Model quality

### 5.1 Standard geometry

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

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	0.44	0/4208	0.58	0/5681
1	B	0.41	0/4130	0.57	0/5575
All	All	0.43	0/8338	0.58	0/11256

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	4130	0	0	12	0
1	B	4055	0	0	17	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	10	0	0	0	0
3	B	18	0	0	2	0
4	A	48	0	0	1	0
4	B	43	0	0	2	0
All	All	8306	0	0	29	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 4.

All (29) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:B:23:GLU:OE1	3:B:601:TBR:BR2	2.71	0.64
1:B:23:GLU:OE1	3:B:601:TBR:BR1	2.77	0.58
1:B:362:ARG:NH2	1:B:384:GLU:OE2	2.40	0.54
1:B:165:VAL:CG1	1:B:165:VAL:O	2.57	0.53
1:A:257:ASN:ND2	1:A:260:ARG:NH2	2.56	0.53
1:B:315:ARG:NH2	1:B:317:ASP:OD2	2.43	0.51
1:B:257:ASN:ND2	1:B:260:ARG:NH2	2.58	0.51
1:A:118:THR:CG2	1:A:119:GLU:N	2.77	0.48
1:B:84:ARG:NE	4:B:625:HOH:O	2.48	0.47
1:B:419:LYS:N	1:B:419:LYS:CD	2.79	0.46
1:B:114:GLY:O	1:B:118:THR:CB	2.64	0.46
1:B:384:GLU:OE1	1:B:468:PHE:N	2.49	0.45
1:B:98:LYS:O	1:B:98:LYS:CD	2.65	0.45
1:B:132:ASN:O	1:B:133:THR:CG2	2.65	0.45
1:B:98:LYS:O	1:B:98:LYS:CG	2.64	0.45
1:A:165:VAL:O	1:A:165:VAL:CG1	2.65	0.45
1:A:443:HIS:NE2	1:A:483:ASP:OD2	2.50	0.44
1:A:132:ASN:O	1:A:133:THR:CG2	2.65	0.44
1:B:317:ASP:C	1:B:317:ASP:OD1	2.55	0.44
1:A:118:THR:CG2	1:A:120:ALA:N	2.81	0.43
1:A:449:HIS:O	1:A:483:ASP:OD1	2.37	0.42
1:A:473:ILE:O	1:A:473:ILE:CD1	2.68	0.42
1:A:531:ALA:CB	4:A:640:HOH:O	2.67	0.41
1:B:118:THR:CG2	1:B:120:ALA:N	2.84	0.41
1:B:367:TYR:CD1	1:B:367:TYR:C	2.94	0.41
1:A:362:ARG:NH2	1:A:384:GLU:CD	2.73	0.41
1:A:151:ASP:OD1	1:A:152:ILE:N	2.53	0.40
1:B:444:LYS:NZ	4:B:624:HOH:O	2.54	0.40
1:A:375:ASN:O	1:A:379:SER:OG	2.39	0.40

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone ⓘ

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	515/540 (95%)	480 (93%)	34 (7%)	1 (0%)	56	85
1	B	505/540 (94%)	467 (92%)	33 (6%)	5 (1%)	22	49
All	All	1020/1080 (94%)	947 (93%)	67 (7%)	6 (1%)	33	64

All (6) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	343	ALA
1	B	423	ASP
1	B	497	GLU
1	B	422	LYS
1	A	354	ASP
1	B	80	GLY

### 5.3.2 Protein sidechains ⓘ

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	444/460 (96%)	391 (88%)	53 (12%)	8	16
1	B	436/460 (95%)	375 (86%)	61 (14%)	5	12
All	All	880/920 (96%)	766 (87%)	114 (13%)	6	14

All (114) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	-2	MET
1	A	2	ILE
1	A	6	LEU
1	A	36	LYS
1	A	37	GLN
1	A	63	VAL
1	A	65	LYS
1	A	68	ILE
1	A	71	ILE

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	79	LEU
1	A	83	LEU
1	A	108	ILE
1	A	111	ILE
1	A	113	ILE
1	A	116	LEU
1	A	117	THR
1	A	118	THR
1	A	119	GLU
1	A	125	THR
1	A	141	LYS
1	A	188	LEU
1	A	191	THR
1	A	200	SER
1	A	208	ARG
1	A	223	LEU
1	A	237	ARG
1	A	292	LEU
1	A	295	LEU
1	A	301	ARG
1	A	306	ASP
1	A	309	LYS
1	A	314	TYR
1	A	347	ILE
1	A	356	ASP
1	A	358	ILE
1	A	364	LYS
1	A	372	LEU
1	A	390	LEU
1	A	391	VAL
1	A	431	ILE
1	A	438	ILE
1	A	442	LEU
1	A	447	VAL
1	A	462	ASP
1	A	468	PHE
1	A	476	LEU
1	A	492	LEU
1	A	494	THR
1	A	497	GLU
1	A	498	LYS
1	A	502	ILE

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	520	ILE
1	A	521	LEU
1	B	2	ILE
1	B	9	THR
1	B	23	GLU
1	B	29	THR
1	B	32	TYR
1	B	33	LYS
1	B	63	VAL
1	B	65	LYS
1	B	68	ILE
1	B	71	ILE
1	B	83	LEU
1	B	93	LEU
1	B	95	LEU
1	B	101	ILE
1	B	111	ILE
1	B	113	ILE
1	B	116	LEU
1	B	117	THR
1	B	118	THR
1	B	125	THR
1	B	133	THR
1	B	141	LYS
1	B	145	VAL
1	B	165	VAL
1	B	167	LEU
1	B	183	LYS
1	B	188	LEU
1	B	208	ARG
1	B	223	LEU
1	B	246	GLU
1	B	263	LEU
1	B	289	ILE
1	B	292	LEU
1	B	295	LEU
1	B	305	LEU
1	B	314	TYR
1	B	324	ILE
1	B	347	ILE
1	B	349	ARG
1	B	359	ILE

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	362	ARG
1	B	363	VAL
1	B	372	LEU
1	B	376	ILE
1	B	390	LEU
1	B	391	VAL
1	B	396	ILE
1	B	419	LYS
1	B	423	ASP
1	B	425	ILE
1	B	426	GLU
1	B	427	ASP
1	B	438	ILE
1	B	448	ILE
1	B	461	LYS
1	B	468	PHE
1	B	482	VAL
1	B	485	ILE
1	B	496	HIS
1	B	498	LYS
1	B	520	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

There are no RNA chains 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 ⓘ

Of 4 ligands modelled in this entry, 2 are monoatomic - leaving 2 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	$\# Z  > 2$
3	TBR	A	601	-	0,11,36	0.00	-	0,28,180	0.00	-
3	TBR	B	601	-	0,36,36	0.00	-	0,180,180	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	TBR	A	601	-	-	0/0/34/696	0/0/2/19
3	TBR	B	601	-	-	0/0/696/696	0/0/19/19

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.

## 5.7 Other polymers ⓘ

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<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	519/540 (96%)	0.46	13 (2%) 54 59	43, 64, 95, 147	0
1	B	511/540 (94%)	0.40	13 (2%) 54 59	43, 66, 127, 167	0
All	All	1030/1080 (95%)	0.43	26 (2%) 54 59	43, 65, 114, 167	0

All (26) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	352	TYR	4.5
1	A	533	TYR	3.6
1	A	496	HIS	3.1
1	A	354	ASP	3.0
1	A	429	LEU	3.0
1	B	392	LYS	3.0
1	B	302	TRP	3.0
1	A	502	ILE	2.7
1	A	406	LEU	2.5
1	B	503	TRP	2.5
1	A	353	LEU	2.5
1	A	503	TRP	2.5
1	B	298	LYS	2.4
1	A	499	PHE	2.4
1	B	464	TYR	2.4
1	B	391	VAL	2.3
1	B	2	ILE	2.3
1	A	459	PHE	2.2
1	B	11	GLU	2.2
1	B	465	ILE	2.1
1	B	454	THR	2.1
1	A	458	ILE	2.1
1	B	520	ILE	2.1
1	B	406	LEU	2.1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	B	484	LEU	2.1
1	A	507	LEU	2.1

## 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
2	MG	A	600	1/1	0.20	0.40	76,76,76,76	0
3	TBR	B	601	18/18	0.19	-0.87	57,65,71,75	18
2	MG	B	600	1/1	0.08	-2.98	77,77,77,77	0
3	TBR	A	601	10/18	0.14	-6.62	91,113,178,192	10

## 6.5 Other polymers ⓘ

There are no such residues in this entry.