W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Feb 19, 2016 – 10:09 PM GMT

PDB ID : 5BMU
Title : The crystal structure of the GST-like domains complex of AIMP3-EPRS mutant C92SC105SC123S
Authors : Cho, H.J.; Kang, B.S.
Deposited on : 2015-05-23
Resolution : 2.60 Å(reported)

The following versions of software and data (see references (1)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : unknown
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026982
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20026982

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 2.60 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.
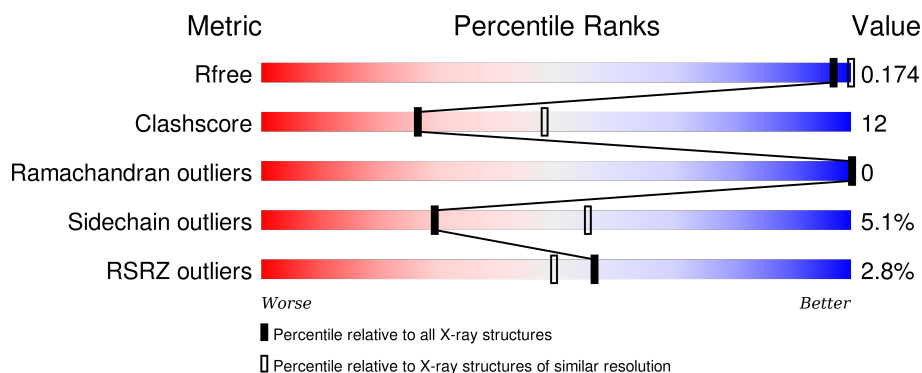
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	2328 (2.60-2.60)
Clashscore	102246	2679 (2.60-2.60)
Ramachandran outliers	100387	2635 (2.60-2.60)
Sidechain outliers	100360	2635 (2.60-2.60)
RSRZ outliers	91569	2334 (2.60-2.60)

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. 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	171	67% 27% · ·
1	C	171	80% 18% · ·
1	E	171	% 75% 22% · ·
1	G	171	2% 82% 15% · ·
2	B	175	5% 66% 26% · 5%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	D	175	4% / 64% / 25% / 6%
2	F	175	3% / 72% / 20% / 5%
2	H	175	6% / 72% / 22% / 6%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 10463 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 Eukaryotic translation elongation factor 1 epsilon-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	165	Total 1328	C 852	N 227	O 247	S 2	0	0	0
1	C	168	Total 1348	C 863	N 232	O 251	S 2	0	0	0
1	E	168	Total 1348	C 863	N 232	O 251	S 2	0	0	0
1	G	168	Total 1344	C 860	N 231	O 251	S 2	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	expression tag	UNP O43324
A	0	HIS	-	expression tag	UNP O43324
C	-1	GLY	-	expression tag	UNP O43324
C	0	HIS	-	expression tag	UNP O43324
E	-1	GLY	-	expression tag	UNP O43324
E	0	HIS	-	expression tag	UNP O43324
G	-1	GLY	-	expression tag	UNP O43324
G	0	HIS	-	expression tag	UNP O43324

- Molecule 2 is a protein called Glutamate–tRNA ligase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	166	Total 1278	C 812	N 214	O 251	S 1	0	0	0
2	D	164	Total 1266	C 804	N 212	O 249	S 1	0	0	0
2	F	166	Total 1284	C 814	N 216	O 253	S 1	0	0	0
2	H	165	Total 1261	C 798	N 212	O 250	S 1	0	0	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	92	SER	CYS	engineered mutation	UNP P07814
B	105	SER	CYS	engineered mutation	UNP P07814
B	123	SER	CYS	engineered mutation	UNP P07814
D	92	SER	CYS	engineered mutation	UNP P07814
D	105	SER	CYS	engineered mutation	UNP P07814
D	123	SER	CYS	engineered mutation	UNP P07814
F	92	SER	CYS	engineered mutation	UNP P07814
F	105	SER	CYS	engineered mutation	UNP P07814
F	123	SER	CYS	engineered mutation	UNP P07814
H	92	SER	CYS	engineered mutation	UNP P07814
H	105	SER	CYS	engineered mutation	UNP P07814
H	123	SER	CYS	engineered mutation	UNP P07814

- Molecule 3 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	4	Total 4	O 4	0	0
3	E	1	Total 1	O 1	0	0
3	F	1	Total 1	O 1	0	0

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.
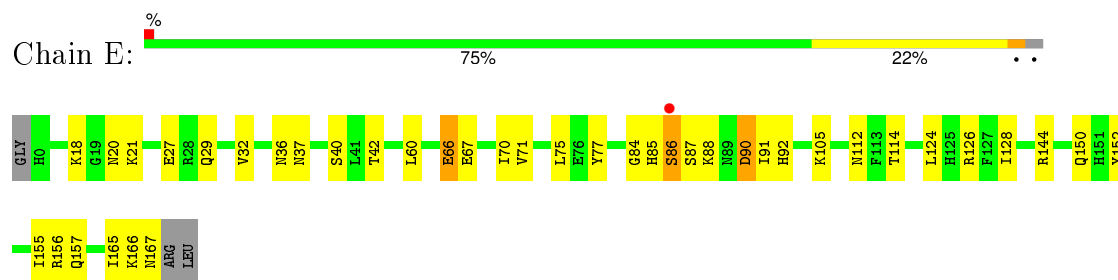
- Molecule 1: Eukaryotic translation elongation factor 1 epsilon-1
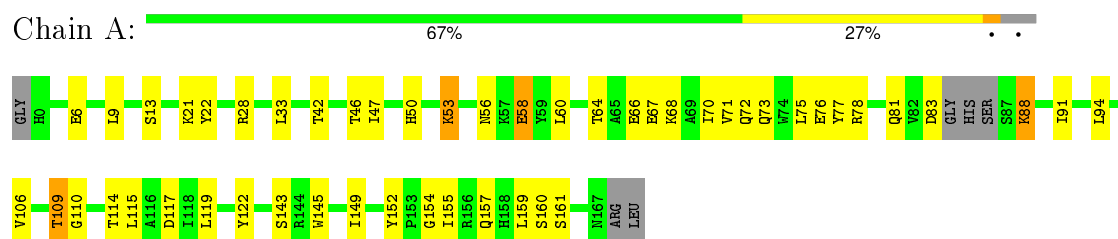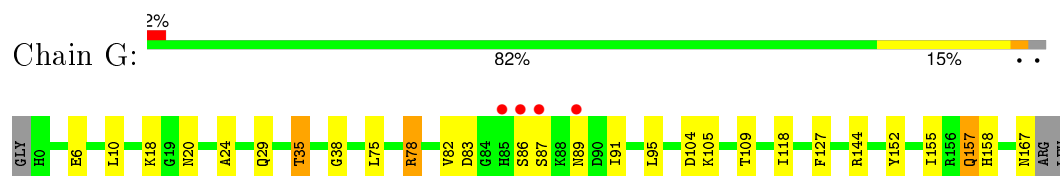

- Molecule 1: Eukaryotic translation elongation factor 1 epsilon-1
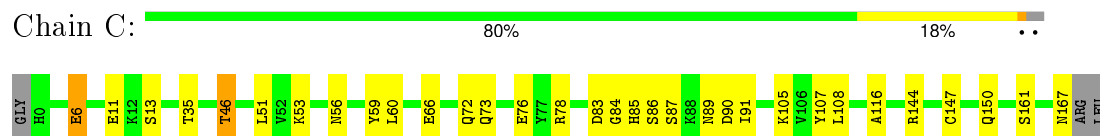
- Molecule 1: Eukaryotic translation elongation factor 1 epsilon-1

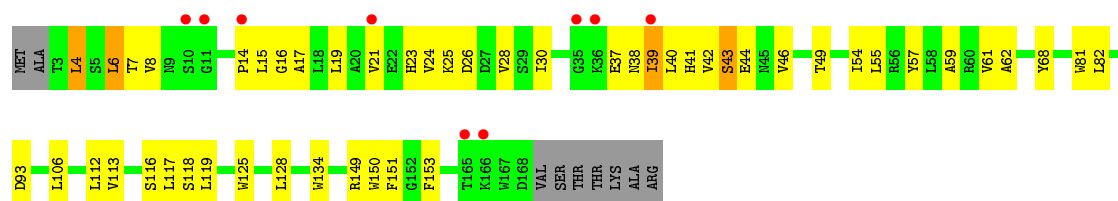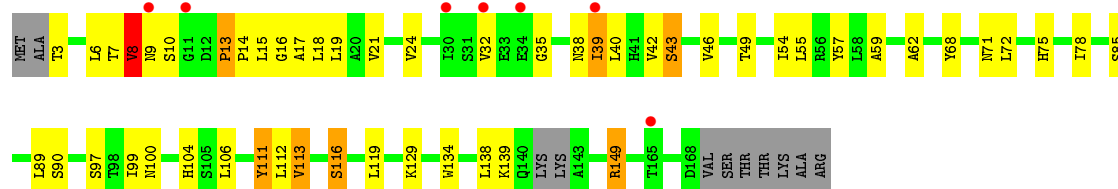- Molecule 1: Eukaryotic translation elongation factor 1 epsilon-1
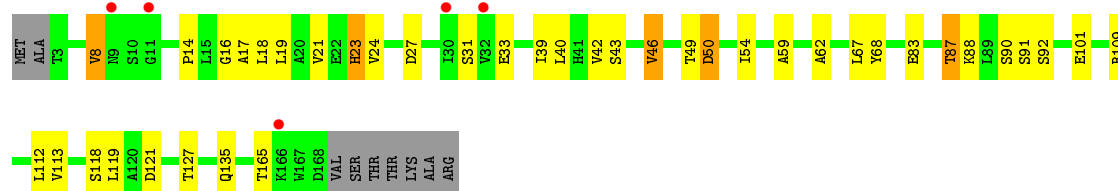

- Molecule 2: Glutamate–tRNA ligase
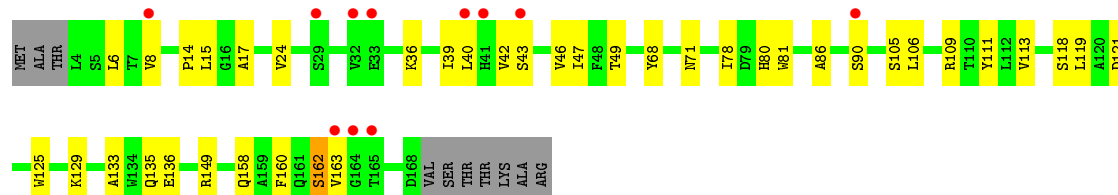
• Molecule 2: Glutamate–tRNA ligase

Chain D:

• Molecule 2: Glutamate–tRNA ligase

Chain F:

• Molecule 2: Glutamate–tRNA ligase

Chain H:

4   Data and refinement statistics (i)

Property	Value	Source
Space group	P 31	Depositor
Cell constants a, b, c, α, β, γ	92.06Å 92.06Å 185.95Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	48.93 – 2.60 48.93 – 2.60	Depositor EDS
% Data completeness (in resolution range)	99.8 (48.93-2.60) 99.8 (48.93-2.60)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.16	Depositor
$< I/\sigma(I) >$ [1]	14.26 (at 2.61Å)	Xtriage
Refinement program	REFMAC 5.8.0103	Depositor
R, R_{free}	0.138 , 0.170 0.146 , 0.174	Depositor DCC
R_{free} test set	1980 reflections (3.81%)	DCC
Wilson B-factor (Å^2)	25.7	Xtriage
Anisotropy	0.046	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 32.9	EDS
Estimated twinning fraction	0.037 for -h,-k,l 0.227 for h,-h-k,-l 0.041 for -k,-h,-l	Xtriage
L-test for twinning[2]	$< |L| > = 0.48, < L^2 > = 0.31$	Xtriage
Outliers	0 of 53895 reflections	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	10463	wwPDB-VP
Average B, all atoms (Å^2)	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.33% 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, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality (i)

5.1 Standard geometry (i)

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	1.04	3/1354 (0.2%)	0.88	2/1833 (0.1%)
1	C	1.10	4/1376 (0.3%)	0.89	0/1864
1	E	0.66	0/1376	0.82	1/1864 (0.1%)
1	G	0.67	0/1372	0.84	2/1860 (0.1%)
2	B	0.82	0/1305	0.86	1/1780 (0.1%)
2	D	0.93	1/1291 (0.1%)	0.90	2/1759 (0.1%)
2	F	0.88	0/1311	0.82	1/1788 (0.1%)
2	H	0.64	0/1286	0.73	0/1754
All	All	0.86	8/10671 (0.1%)	0.84	9/14502 (0.1%)

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	58	GLU	C-N	9.41	1.55	1.34
1	A	110	GLY	C-N	-6.28	1.19	1.34
1	A	143	SER	CB-OG	-5.76	1.34	1.42
1	C	107	TYR	CD2-CE2	-5.64	1.30	1.39
2	D	111	TYR	CE2-CZ	-5.61	1.31	1.38

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	109	THR	N-CA-C	-6.87	92.46	111.00
1	G	78	ARG	NE-CZ-NH1	6.26	123.43	120.30
2	D	113	VAL	CB-CA-C	-6.19	99.63	111.40
2	D	8	VAL	CB-CA-C	-5.66	100.65	111.40
1	A	110	GLY	O-C-N	-5.56	113.81	122.70

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	1328	0	1338	49	0
1	C	1348	0	1355	20	0
1	E	1348	0	1355	25	0
1	G	1344	0	1344	20	0
2	B	1278	0	1235	41	0
2	D	1266	0	1231	48	0
2	F	1284	0	1243	39	0
2	H	1261	0	1213	26	0
3	A	4	0	0	0	0
3	E	1	0	0	0	0
3	F	1	0	0	0	0
All	All	10463	0	10314	253	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 12.

The worst 5 of 253 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)
2:D:8:VAL:HG11	2:D:14:PRO:HD3	1.15	1.07
1:A:46:THR:HG23	1:C:76:GLU:OE2	1.53	1.05
2:D:8:VAL:HG11	2:D:14:PRO:CD	1.87	1.04
1:G:87:SER:O	1:G:91:ILE:HG23	1.66	0.94
2:D:113:VAL:HG21	2:D:116:SER:O	1.69	0.93

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	161/171 (94%)	147 (91%)	14 (9%)	0	100	100
1	C	166/171 (97%)	149 (90%)	17 (10%)	0	100	100
1	E	166/171 (97%)	158 (95%)	8 (5%)	0	100	100
1	G	166/171 (97%)	157 (95%)	9 (5%)	0	100	100
2	B	164/175 (94%)	154 (94%)	10 (6%)	0	100	100
2	D	160/175 (91%)	147 (92%)	13 (8%)	0	100	100
2	F	164/175 (94%)	153 (93%)	11 (7%)	0	100	100
2	H	163/175 (93%)	151 (93%)	12 (7%)	0	100	100
All	All	1310/1384 (95%)	1216 (93%)	94 (7%)	0	100	100

There are no Ramachandran outliers to report.

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	145/149 (97%)	141 (97%)	4 (3%)	51	78
1	C	147/149 (99%)	144 (98%)	3 (2%)	63	85
1	E	147/149 (99%)	138 (94%)	9 (6%)	23	46
1	G	146/149 (98%)	143 (98%)	3 (2%)	61	85
2	B	139/151 (92%)	129 (93%)	10 (7%)	18	35
2	D	139/151 (92%)	127 (91%)	12 (9%)	13	25
2	F	141/151 (93%)	130 (92%)	11 (8%)	16	30
2	H	137/151 (91%)	131 (96%)	6 (4%)	35	63
All	All	1141/1200 (95%)	1083 (95%)	58 (5%)	29	55

5 of 58 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	129	LYS
1	E	40	SER
2	H	68	TYR
2	D	139	LYS
1	E	27	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	81	GLN
1	C	167	ASN
1	E	167	ASN
1	C	36	ASN
1	E	85	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 carbohydrates 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)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	A	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	110:GLY	C	111:TYR	N	1.19

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	165/171 (96%)	-0.64	0	100	100	10, 19, 47, 68	0
1	C	168/171 (98%)	-0.70	0	100	100	9, 20, 45, 67	0
1	E	168/171 (98%)	-0.64	1 (0%)	90	88	11, 21, 40, 59	0
1	G	168/171 (98%)	-0.59	4 (2%)	62	56	8, 22, 43, 78	0
2	B	166/175 (94%)	-0.25	9 (5%)	29	22	17, 35, 67, 80	0
2	D	164/175 (93%)	-0.21	7 (4%)	39	31	17, 36, 68, 80	0
2	F	166/175 (94%)	-0.19	5 (3%)	54	47	22, 43, 79, 96	0
2	H	165/175 (94%)	-0.07	11 (6%)	21	15	17, 45, 78, 95	0
All	All	1330/1384 (96%)	-0.41	37 (2%)	56	49	8, 28, 68, 96	0

The worst 5 of 37 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	32	VAL	6.1
2	H	32	VAL	5.0
2	D	165	THR	4.6
2	D	11	GLY	4.4
2	B	35	GLY	4.1

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 carbohydrates in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.