



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 04:42 AM GMT

PDB ID : 2NVW
Title : Crystal structure of transcriptional regulator Gal80p from *Kluyveromyces fragilis*
Authors : Thoden, J.B.; Sellick, C.A.; Reece, R.J.; Holden, H.M.
Deposited on : 2006-11-13
Resolution : 2.10 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
<http://wwpdb.org/validation/2016/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 : 1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
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) : trunk26865

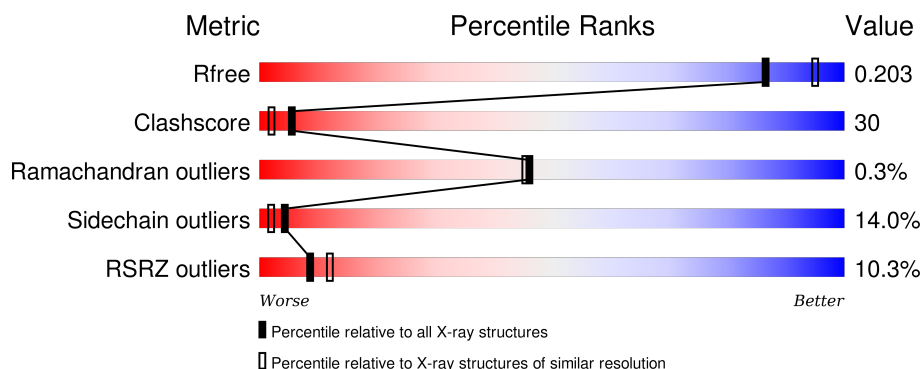
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.10 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	3939 (2.10-2.10)
Clashscore	102246	4460 (2.10-2.10)
Ramachandran outliers	100387	4413 (2.10-2.10)
Sidechain outliers	100360	4414 (2.10-2.10)
RSRZ outliers	91569	3948 (2.10-2.10)

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 $\leq 5\%$. 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	479	<div> <div>9%</div> <div>43%</div> <div>35%</div> <div>8%</div> <div>14%</div> </div>
1	B	479	<div> <div>9%</div> <div>39%</div> <div>35%</div> <div>10%</div> <div>15%</div> </div>

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 6895 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 Galactose/lactose metabolism regulatory protein GAL80.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	413	Total	C	N	O	S	0	1	0
			3310	2117	563	620	10			
1	B	405	Total	C	N	O	S	0	0	0
			3238	2072	547	609	10			

There are 46 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-21	MET	-	INITIATING METHIONINE	UNP Q06433
A	-20	GLY	-	CLONING ARTIFACT	UNP Q06433
A	-19	SER	-	CLONING ARTIFACT	UNP Q06433
A	-18	SER	-	CLONING ARTIFACT	UNP Q06433
A	-17	HIS	-	EXPRESSION TAG	UNP Q06433
A	-16	HIS	-	EXPRESSION TAG	UNP Q06433
A	-15	HIS	-	EXPRESSION TAG	UNP Q06433
A	-14	HIS	-	EXPRESSION TAG	UNP Q06433
A	-13	HIS	-	EXPRESSION TAG	UNP Q06433
A	-12	HIS	-	EXPRESSION TAG	UNP Q06433
A	-11	SER	-	CLONING ARTIFACT	UNP Q06433
A	-10	SER	-	CLONING ARTIFACT	UNP Q06433
A	-9	GLU	-	CLONING ARTIFACT	UNP Q06433
A	-8	ASN	-	CLONING ARTIFACT	UNP Q06433
A	-7	LEU	-	CLONING ARTIFACT	UNP Q06433
A	-6	TYR	-	CLONING ARTIFACT	UNP Q06433
A	-5	PHE	-	CLONING ARTIFACT	UNP Q06433
A	-4	GLN	-	CLONING ARTIFACT	UNP Q06433
A	-3	GLY	-	CLONING ARTIFACT	UNP Q06433
A	-2	HIS	-	CLONING ARTIFACT	UNP Q06433
A	-1	MET	-	CLONING ARTIFACT	UNP Q06433
A	0	LEU	-	CLONING ARTIFACT	UNP Q06433
A	1	ALA	-	CLONING ARTIFACT	UNP Q06433
B	-21	MET	-	INITIATING METHIONINE	UNP Q06433
B	-20	GLY	-	CLONING ARTIFACT	UNP Q06433

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-19	SER	-	CLONING ARTIFACT	UNP Q06433
B	-18	SER	-	CLONING ARTIFACT	UNP Q06433
B	-17	HIS	-	EXPRESSION TAG	UNP Q06433
B	-16	HIS	-	EXPRESSION TAG	UNP Q06433
B	-15	HIS	-	EXPRESSION TAG	UNP Q06433
B	-14	HIS	-	EXPRESSION TAG	UNP Q06433
B	-13	HIS	-	EXPRESSION TAG	UNP Q06433
B	-12	HIS	-	EXPRESSION TAG	UNP Q06433
B	-11	SER	-	CLONING ARTIFACT	UNP Q06433
B	-10	SER	-	CLONING ARTIFACT	UNP Q06433
B	-9	GLU	-	CLONING ARTIFACT	UNP Q06433
B	-8	ASN	-	CLONING ARTIFACT	UNP Q06433
B	-7	LEU	-	CLONING ARTIFACT	UNP Q06433
B	-6	TYR	-	CLONING ARTIFACT	UNP Q06433
B	-5	PHE	-	CLONING ARTIFACT	UNP Q06433
B	-4	GLN	-	CLONING ARTIFACT	UNP Q06433
B	-3	GLY	-	CLONING ARTIFACT	UNP Q06433
B	-2	HIS	-	CLONING ARTIFACT	UNP Q06433
B	-1	MET	-	CLONING ARTIFACT	UNP Q06433
B	0	LEU	-	CLONING ARTIFACT	UNP Q06433
B	1	ALA	-	CLONING ARTIFACT	UNP Q06433

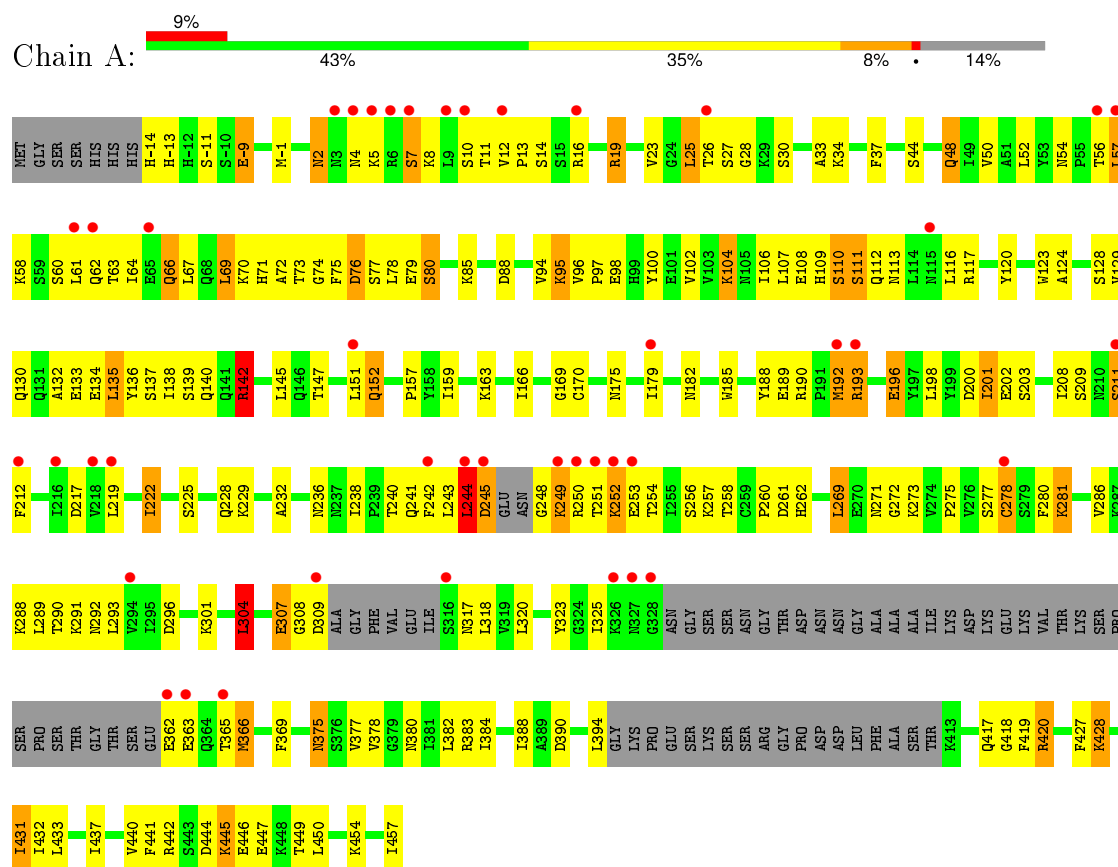
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	203	Total 203	O 203	0	0
2	B	144	Total 144	O 144	0	0

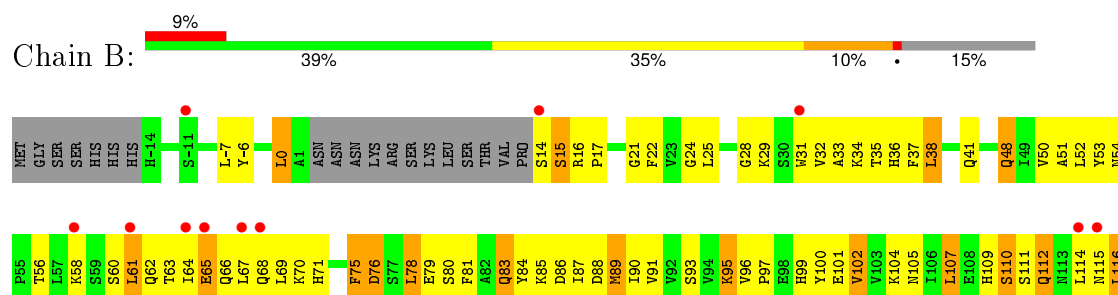
3 Residue-property plots

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: Galactose/lactose metabolism regulatory protein GAL80



- Molecule 1: Galactose/lactose metabolism regulatory protein GAL80



R420	F421	E422	K428	D429	A430	I431	I432	L433	H434	R435	I436	I437	D438	F441	R442	K445	E446	E447	K448	T449	L450	D451	V452	S453	K454	I455	M456	I457	F393	L394	G395	LYS	PRO	GLU	SER	LYS	SER	SER	ARG	GLY	PRO	ASP	ASP	LEU	PHE	ALA	SER	THR	K413	F414	D415	K416	Q417	G418	F419	
GLU	LYS	VAL	THR	LYS	SER	PRO	SER	PRO	SER	THR	GLY	THR	SER	E361	E362	T365	M366	F369	R372	N375	N380	R383	I384	S387	F393	L394	G395	LYS	PRO	GLU	SER	LYS	SER	SER	ARG	GLY	SER	SER	PRO	ASP	ASP	THR	ASP	ASN	ASN	GLY	ALA	ALA	ALA	ILE	LYS	ASP	LYS			
F280	K281	T284	P285	V286	K287	K288	L289	N292	L293	D294	V294	I295	D296	I297	H298	G299	T300	K301	L304	D309	A310	G311	PHE	VAL	GLU	ILE	S316	V319	L320	L321	F322	Y323	G324	I325	K326	N327	G328	ASN	GLY	SER	SER	SER	ASP	GLY	THR	ASP	ASN	ASN	GLY	ALA	ALA	ALA	ILE	LYS	ASP	LYS
E189	R190	P191	M192	R193	S194	P195	E196	Y197	L198	Y199	D200	L201	E202	S203	L207	S211	D217	V218	L219	Q220	Y221	I222	F227	Q228	K229	A232	N236	F242	L243	L244	D245	E246	N247	GLY	LYS	R250	T251	K252	E253	T254	I255	S256	K257	T258	C259	P260	D261	L264	C278	S279						
R117	Y120	Y121	E122	H123	A124	L125	A126	A127	S128	V129	Q130	Q131	A132	E133	E134	L135	Y136	S137	I138	S139	Q140	Q141	R142	L145	I149	G150	L151	Q152	G153	R154	K155	S156	P157	Y158	I159	V160	K163	E164	L165	I166	S167	D173	S176	I177	E178	I179	G183	G184	W185	Y186	G187	Y188				

4 Data and refinement statistics

Property	Value	Source
Space group	P 2 ₁ 2 ₁ 2	Depositor
Cell constants a, b, c, α , β , γ	112.20 Å 137.10 Å 72.90 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 2.10 37.30 – 2.09	Depositor EDS
% Data completeness (in resolution range)	97.7 (30.00-2.10) 97.4 (37.30-2.09)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.07	Depositor
$\langle I/\sigma(I) \rangle$ ¹	0.61 (at 2.08 Å)	Xtriage
Refinement program	TNT	Depositor
R, R_{free}	0.195 , 0.245 0.199 , 0.203	Depositor DCC
R_{free} test set	6516 reflections (11.16%)	DCC
Wilson B-factor (Å ²)	28.2	Xtriage
Anisotropy	0.149	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.28 , 112.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	1 of 65500 reflections (0.002%)	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	6895	wwPDB-VP
Average B, all atoms (Å ²)	46.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.80% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

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	0.77	0/3381	1.50	34/4560 (0.7%)
1	B	0.79	1/3302 (0.0%)	1.52	41/4453 (0.9%)
All	All	0.78	1/6683 (0.0%)	1.51	75/9013 (0.8%)

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	1	1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	422	GLU	CB-CG	5.55	1.62	1.52

The worst 5 of 75 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	304	LEU	CB-CG-CD1	-9.85	94.25	111.00
1	A	145	LEU	CB-CG-CD2	-9.68	94.54	111.00
1	B	219	LEU	CA-CB-CG	-9.24	94.06	115.30
1	B	435	ARG	NE-CZ-NH1	-8.64	115.98	120.30
1	A	296	ASP	CB-CG-OD2	-8.54	110.61	118.30

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	A	112	GLN	CA

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	2	ASN	Mainchain

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	3310	0	3309	172	1
1	B	3238	0	3218	217	0
2	A	203	0	0	15	0
2	B	144	0	0	7	1
All	All	6895	0	6527	387	1

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 30.

The worst 5 of 387 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:251:THR:HG22	1:A:253:GLU:H	1.06	1.15
1:B:25:LEU:HD23	1:B:52:LEU:HD21	1.36	1.07
1:B:159:ILE:HG23	1:B:222:ILE:HD11	1.38	1.04
1:A:70:LYS:HG3	1:A:71:HIS:CD2	1.98	0.99
1:A:250:ARG:HG2	1:A:251:THR:H	1.26	0.98

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:76:ASP:OD2	2:B:601:HOH:O[3_556]	2.16	0.04

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	404/479 (84%)	386 (96%)	18 (4%)	0	100	100
1	B	393/479 (82%)	361 (92%)	30 (8%)	2 (0%)	34	30
All	All	797/958 (83%)	747 (94%)	48 (6%)	2 (0%)	46	45

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	15	SER
1	B	102	VAL

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	370/422 (88%)	318 (86%)	52 (14%)	4	2
1	B	359/422 (85%)	309 (86%)	50 (14%)	4	2
All	All	729/844 (86%)	627 (86%)	102 (14%)	4	2

5 of 102 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	366	MET
1	B	75	PHE
1	B	372	ARG
1	A	428	LYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	454	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 26 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	152	GLN
1	A	271	ASN
1	B	228	GLN
1	A	228	GLN
1	A	236	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 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](#)

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	413/479 (86%)	0.57	43 (10%) 8 12	19, 39, 81, 100	0
1	B	405/479 (84%)	0.67	41 (10%) 9 12	22, 41, 79, 100	0
All	All	818/958 (85%)	0.62	84 (10%) 9 12	19, 40, 81, 100	0

The worst 5 of 84 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	361	GLU	7.4
1	B	327	ASN	7.0
1	A	316	SER	6.3
1	B	328	GLY	6.2
1	B	245	ASP	5.8

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.