wwPDB X-ray Structure Validation Summary Report (i)

Dec 6, 2023 – 05:21 am GMT

PDB ID	:	1HJB
Title	:	CRYSTAL STRUCTURE OF RUNX-1/AML1/CBFALPHA RUNT DOMAIN AND C/EBPBETA BZIP HOMODIMER BOUND TO A DNA FRAGMENT FROM THE CSF-1R PROMOTER
Authors	:	Tahirov, T.H.; Ogata, K.
Deposited on	:	2001-01-11
Resolution	:	3.00 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

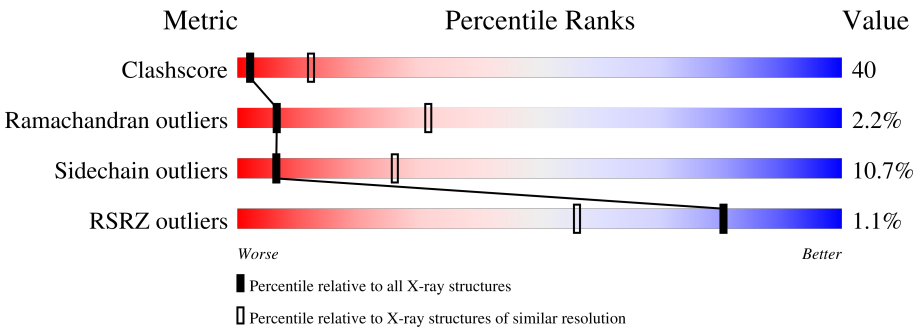
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 3.00 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))
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)

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 of 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	87	3%; 33%, 36%, 7%, 24%
1	B	87	2%; 22%, 47%, 8%, 23%
1	D	87	33%, 34%, 10%, 22%
1	E	87	39%, 29%, 10%, 22%
2	C	123	%; 42%, 51%, ·, ·
2	F	123	%; 38%, 51%, 8%, ·
3	G	26	8%, 88%, ·

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	I	26	12% 88%
4	H	26	8% 92%
4	J	26	96%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 6270 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 CCAAT/ENHANCER BINDING PROTEIN BETA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	66	560	341	114	104	1	0	0	0
1	B	67	568	347	116	104	1	0	0	0
1	D	68	579	354	117	107	1	0	0	0
1	E	68	575	352	117	105	1	0	0	0

- Molecule 2 is a protein called RUNT-RELATED TRANSCRIPTION FACTOR 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	C	120	934	586	175	169	4	0	0	0
2	F	120	934	586	175	169	4	0	0	0

- Molecule 3 is a DNA chain called DNA (5'-(*GP*AP*AP*GP*AP*TP*TP*TP*CP*CP* A P*AP*AP*CP*TP*CP*TP*GP*TP*GP*GP*TP*TP*GP*CP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	G	26	532	255	96	156	25	0	0	0
3	I	26	532	255	96	156	25	0	0	0

- Molecule 4 is a DNA chain called DNA (5'-(*CP*CP*GP*CP*AP*AP*CP*CP*AP*CP* A P*GP*AP*GP*TP*TP*TP*GP*GP*AP*AP*AP*TP*CP*TP*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
4	H	26	528	253	98	152	25	0	0	0

Continued on next page...

Continued from previous page...

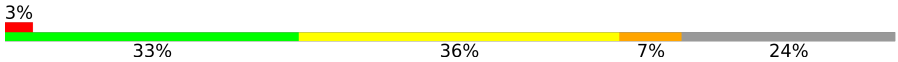
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
4	J	26	528	253	98	152	25	0	0	0

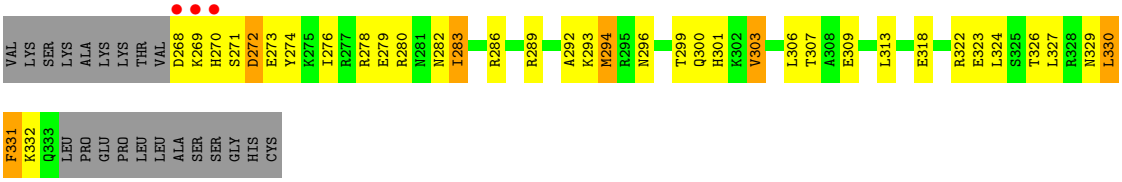
3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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: CCAAT/ENHANCER BINDING PROTEIN BETA

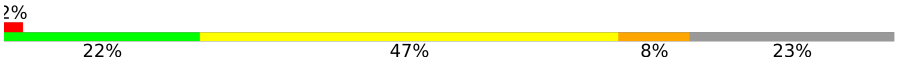
Chain A:

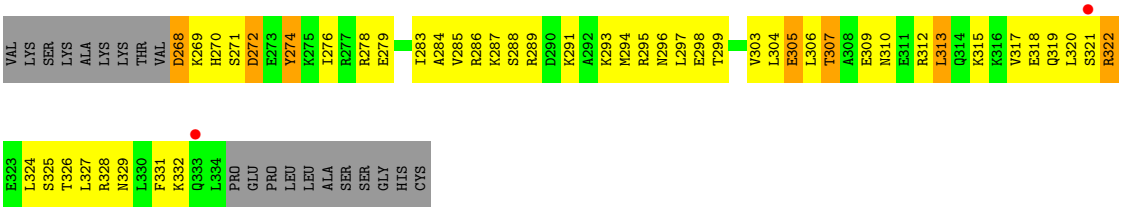




• Molecule 1: CCAAT/ENHANCER BINDING PROTEIN BETA

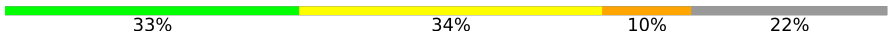
Chain B:

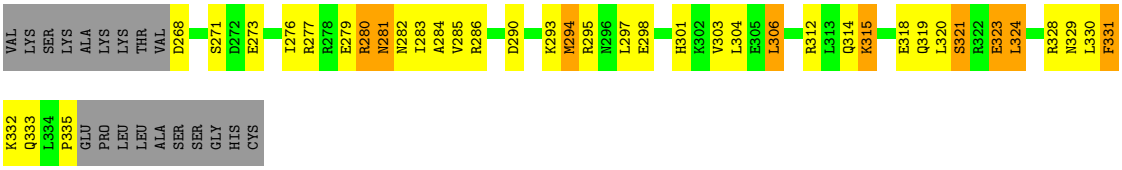




• Molecule 1: CCAAT/ENHANCER BINDING PROTEIN BETA

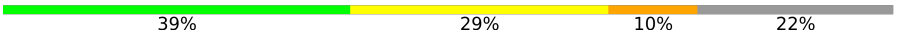
Chain D:

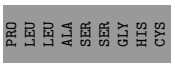




• Molecule 1: CCAAT/ENHANCER BINDING PROTEIN BETA

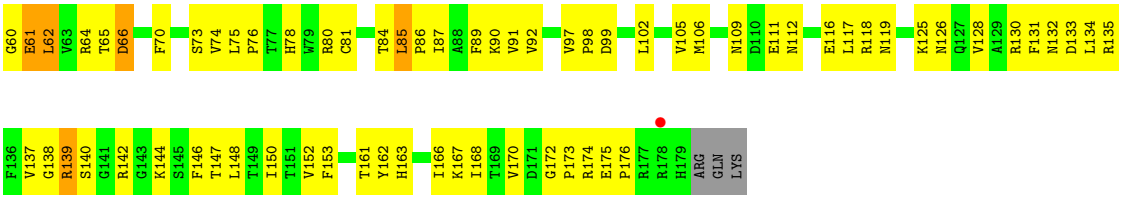
Chain E:





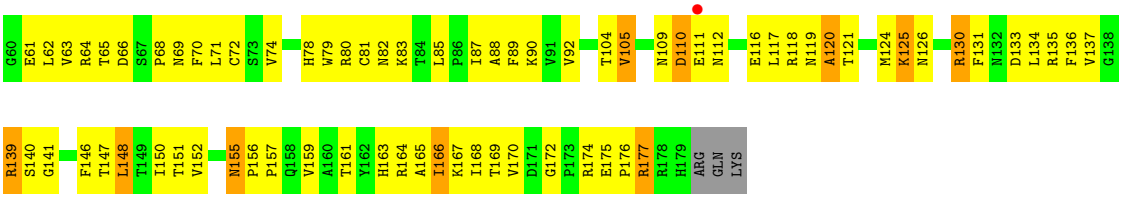
• Molecule 2: RUNT-RELATED TRANSCRIPTION FACTOR 1

Chain C:



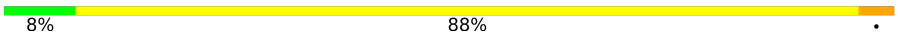
• Molecule 2: RUNT-RELATED TRANSCRIPTION FACTOR 1

Chain F:



• Molecule 3: DNA (5'-(*GP*AP*AP*GP*AP*TP*TP*TP*CP*CP* AP*AP*AP*CP*TP*CP* TP*GP*TP*GP*GP*TP*TP*GP*CP*G)-3')

Chain G:

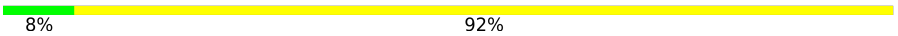


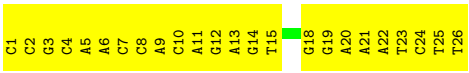
• Molecule 3: DNA (5'-(*GP*AP*AP*GP*AP*TP*TP*TP*CP*CP* AP*AP*AP*CP*TP*CP* TP*GP*TP*GP*GP*TP*TP*GP*CP*G)-3')

Chain I:

• Molecule 4: DNA (5'-(*CP*CP*GP*CP*AP*AP*CP*CP*AP*CP* AP*GP*AP*GP*TP*TP* TP*GP*GP*AP*AP*AP*TP*CP*TP*T)-3')

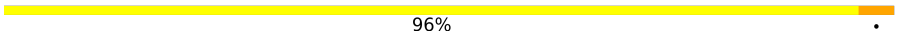
Chain H:

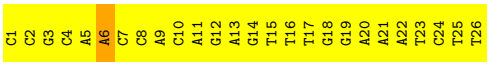




• Molecule 4: DNA (5'-(*CP*CP*GP*CP*AP*AP*CP*CP*AP*CP* AP*GP*AP*GP*TP*TP* TP*GP*GP*AP*AP*AP*TP*CP*TP*T)-3')

Chain J:


96%


C1
C2
G3
C4
A5
A6
C7
C8
A9
C10
A11
G12
A13
G14
T15
T16
T17
G18
G19
A20
A21
A22
T23
C24
T25
T26

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	102.17Å 109.27Å 127.41Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 3.00 30.03 – 3.00	Depositor EDS
% Data completeness (in resolution range)	89.6 (30.00-3.00) 89.7 (30.03-3.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.07	Depositor
$< I/\sigma(I) >$ [1]	2.58 (at 3.00Å)	Xtriage
Refinement program	CNS 0.9	Depositor
R, R_{free}	0.244 , 0.313 0.258 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	60.0	Xtriage
Anisotropy	0.378	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.15 , 33.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.41$, $< L^2 > = 0.23$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	6270	wwPDB-VP
Average B, all atoms (Å^2)	72.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.92% 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 (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	0.37	0/563	0.59	0/746
1	B	0.36	0/571	0.61	0/757
1	D	0.35	0/583	0.54	0/774
1	E	0.38	0/579	0.61	0/769
2	C	0.39	0/954	0.71	0/1297
2	F	0.42	0/954	0.73	0/1297
3	G	0.50	0/596	0.77	0/919
3	I	0.48	0/596	0.79	0/919
4	H	0.48	0/592	0.77	0/911
4	J	0.51	0/592	0.82	0/911
All	All	0.43	0/6580	0.71	0/9300

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
3	G	0	1
4	J	0	3
All	All	0	4

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	G	18	DG	Sidechain
4	J	1	DC	Sidechain
4	J	2	DC	Sidechain
4	J	6	DA	Sidechain

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	560	0	583	49	0
1	B	568	0	596	62	0
1	D	579	0	607	34	0
1	E	575	0	603	60	0
2	C	934	0	940	61	0
2	F	934	0	940	72	1
3	G	532	0	296	34	0
3	I	532	0	296	48	0
4	H	528	0	294	58	0
4	J	528	0	294	47	0
All	All	6270	0	5449	471	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 40.

The worst 5 of 471 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)
3:G:9:DC:H2"	3:G:10:DC:H5"	1.25	1.14
1:E:295:ARG:HH11	1:E:295:ARG:HB2	1.07	1.12
4:H:20:DA:H2"	4:H:21:DA:H5"	1.32	1.10
4:J:20:DA:H2"	4:J:21:DA:H5"	1.33	1.06
3:I:24:DG:H2"	3:I:25:DC:H5'	1.34	1.05

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)
2:F:111:GLU:CB	2:F:111:GLU:CB[2_575]	2.13	0.07

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	64/87 (74%)	50 (78%)	12 (19%)	2 (3%)	4 23
1	B	65/87 (75%)	57 (88%)	7 (11%)	1 (2%)	10 42
1	D	66/87 (76%)	55 (83%)	10 (15%)	1 (2%)	10 42
1	E	66/87 (76%)	56 (85%)	8 (12%)	2 (3%)	4 24
2	C	118/123 (96%)	107 (91%)	9 (8%)	2 (2%)	9 39
2	F	118/123 (96%)	105 (89%)	10 (8%)	3 (2%)	5 28
All	All	497/594 (84%)	430 (86%)	56 (11%)	11 (2%)	6 31

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	272	ASP
2	F	82	ASN
2	C	61	GLU
2	F	120	ALA
2	C	111	GLU

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	62/81 (76%)	57 (92%)	5 (8%)	11 40
1	B	63/81 (78%)	55 (87%)	8 (13%)	4 19
1	D	65/81 (80%)	53 (82%)	12 (18%)	1 8

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	E	64/81 (79%)	57 (89%)	7 (11%)	6 25
2	C	102/105 (97%)	95 (93%)	7 (7%)	15 48
2	F	102/105 (97%)	92 (90%)	10 (10%)	8 30
All	All	458/534 (86%)	409 (89%)	49 (11%)	6 26

5 of 49 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	324	LEU
1	E	307	THR
1	D	329	ASN
1	E	274	TYR
1	E	319	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	319	GLN
2	F	69	ASN
2	F	155	ASN
1	E	310	ASN
1	B	333	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 (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	66/87 (75%)	-0.55	3 (4%)	33	12	27, 81, 163, 183	0
1	B	67/87 (77%)	-0.49	2 (2%)	50	22	29, 78, 147, 166	0
1	D	68/87 (78%)	-0.63	0	100	100	35, 87, 161, 193	0
1	E	68/87 (78%)	-0.81	0	100	100	27, 73, 121, 154	0
2	C	120/123 (97%)	-0.68	1 (0%)	86	65	35, 73, 122, 170	0
2	F	120/123 (97%)	-0.70	1 (0%)	86	65	28, 65, 116, 173	0
3	G	26/26 (100%)	-1.06	0	100	100	33, 55, 77, 89	0
3	I	26/26 (100%)	-1.10	0	100	100	32, 53, 83, 102	0
4	H	26/26 (100%)	-1.10	0	100	100	35, 57, 69, 77	0
4	J	26/26 (100%)	-1.10	0	100	100	38, 53, 78, 95	0
All	All	613/698 (87%)	-0.73	7 (1%)	80	56	27, 69, 132, 193	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	270	HIS	5.5
2	F	111	GLU	4.9
1	A	269	LYS	3.7
1	A	268	ASP	2.8
1	B	321	SER	2.4

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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.