wwPDB X-ray Structure Validation Summary Report (i)

Dec 10, 2022 – 04:35 PM EST

PDB ID	:	1BF5
Title	:	TYROSINE PHOSPHORYLATED STAT-1/DNA COMPLEX
Authors	:	Kuriyan, J.; Zhao, Y.; Chen, X.; Vinkemeier, U.; Jeruzalmi, D.; Darnell Jr., J.E.
Deposited on	:	1998-05-27
Resolution	:	2.90 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 (i)) were used in the production of this report:

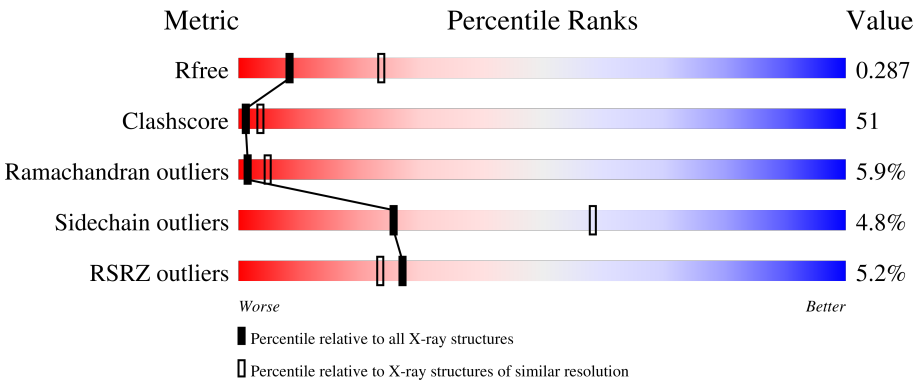
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.31.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.2

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 2.90 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	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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	B	18	11%; 6%, 83%, 11%
2	C	18	11%; 11%, 72%, 17%
3	A	575	5%; 38%, 47%, 8%, • 5%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 5117 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 DNA chain called DNA (5'-D(*AP*CP*AP*GP*TP*TP*TP*CP*CP*CP*GP*TP*AP*AP*AP*TP*G P*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	B	18	363	175	65	106	17	0	0	0

- Molecule 2 is a DNA chain called DNA (5'-D(*TP*GP*CP*AP*TP*TP*TP*AP*CP*GP*GP*GP*AP*AP*AP*CP*T P*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	C	18	369	177	69	106	17	0	0	0

- Molecule 3 is a protein called SIGNAL TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION 1-ALPHA/BETA.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	P	S			
3	A	545	4385	2808	751	807	1	18	0	0	0

There is a discrepancy between the modelled and reference sequences:

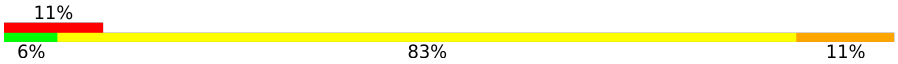
Chain	Residue	Modelled	Actual	Comment	Reference
A	701	PTR	TYR	modified residue	UNP P42224

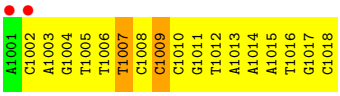
3 Residue-property plots (i)

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: DNA (5'-D(*AP*CP*AP*GP*TP*TP*TP*CP*CP*CP*GP*TP*AP*AP*AP*TP*G P*C)-3')

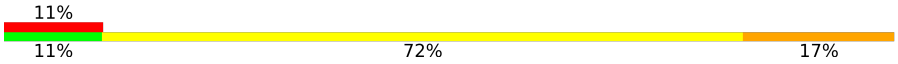
Chain B:





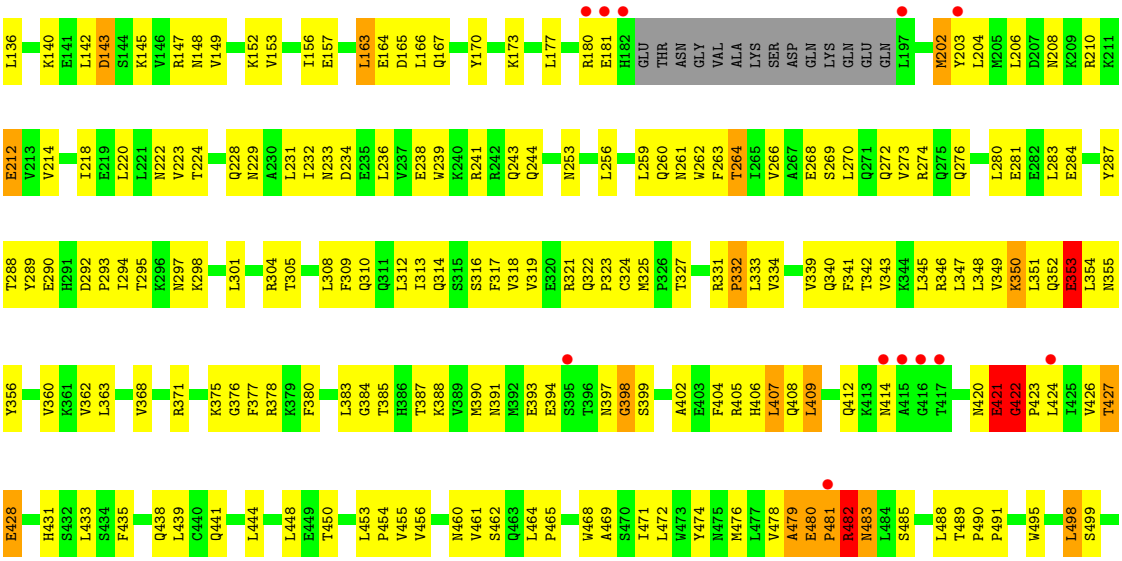
• Molecule 2: DNA (5'-D(*TP*GP*CP*AP*TP*TP*TP*AP*CP*GP*GP*GP*AP*AP*AP*CP*T P*G)-3')

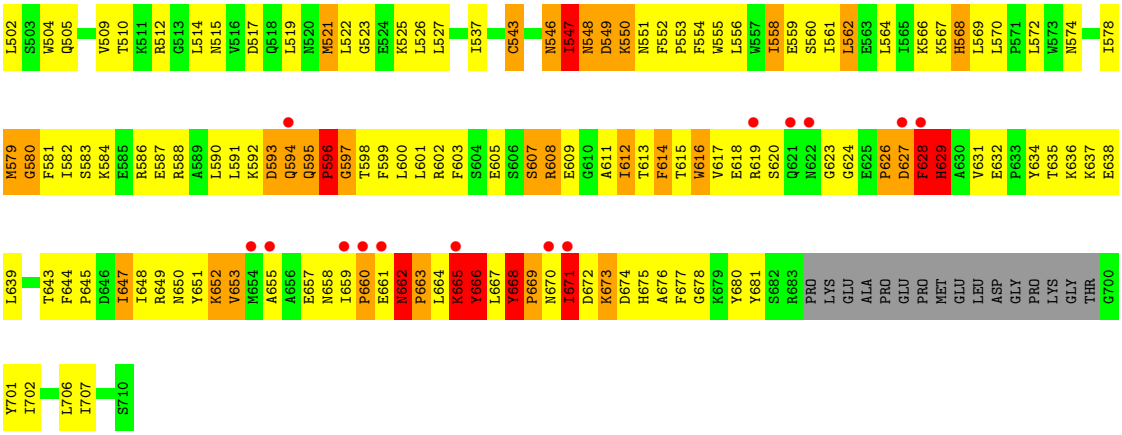
Chain C:



• Molecule 3: SIGNAL TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION 1-ALPHA/BETA

Chain A:




L502 S503 W504 Q505 V509 T510 K511 R512 G513 L514 N515 V516 D517 Q518 L519 N520 M521 L522 G523 E524 K525 L526 L527 I537 C543 N546 I547 N548 D549 K550 N551 F552 P553 F554 W555 L556 W557 I558 E559 S560 I561 L562 E563 L564 I565 K566 K567 H568 L569 L570 P571 L572 W573 N574 I578
M579 G580 F581 I582 S583 K584 E585 R586 E587 R588 A589 L590 L591 K592 D593 Q594 Q595 P596 G597 T598 F599 L600 L601 R602 F603 S604 E605 S606 S607 R608 E609 G610 A611 I612 T613 F614 T615 W616 V617 E618 R619 S620 Q621 N622 G623 G624 E625 P626 D627 F628 H629 A630 V631 E632 P633 Y634 T635 K636 K637 E638
L639 T643 F644 P645 D646 I647 I648 R649 N650 Y651 K652 V653 M654 A655 A656 E657 N658 I659 P660 E661 N662 P663 L664 K665 Y666 L667 Y668 P669 N670 I671 D672 K673 D674 H675 A676 F677 G678 K679 Y680 Y681 S682 R683 PRO LYS GLU ALA PRO GLU PRO MET GLU LEU ASP GLY PRO LYS GLY THR G700
Y701 I702 L706 I707 S710

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	76.60Å 148.20Å 181.10Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 2.90 29.24 – 2.79	Depositor EDS
% Data completeness (in resolution range)	89.8 (30.00-2.90) 84.3 (29.24-2.79)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	0.20	Depositor
$< I/\sigma(I) >$ [1]	4.32 (at 2.80Å)	Xtriage
Refinement program	CNS 3.0	Depositor
R, R_{free}	0.227 , 0.293 0.222 , 0.287	Depositor DCC
R_{free} test set	2177 reflections (9.91%)	wwPDB-VP
Wilson B-factor (Å^2)	45.5	Xtriage
Anisotropy	0.152	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 49.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.43$, $< L^2 > = 0.25$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.88	EDS
Total number of atoms	5117	wwPDB-VP
Average B, all atoms (Å^2)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.62% 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)

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

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	B	0.71	0/405	0.82	0/620
2	C	3.93	1/414 (0.2%)	1.07	3/638 (0.5%)
3	A	0.94	11/4460 (0.2%)	1.08	30/6035 (0.5%)
All	All	1.41	12/5279 (0.2%)	1.06	33/7293 (0.5%)

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	B	0	2
2	C	0	2
3	A	1	7
All	All	1	11

The worst 5 of 12 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	2010	DG	O3'-P	-78.87	0.66	1.61
3	A	623	GLY	C-N	-23.42	0.90	1.33
3	A	480	GLU	N-CA	21.34	1.89	1.46
3	A	543	CYS	CB-SG	15.16	2.08	1.82
3	A	422	GLY	N-CA	14.87	1.68	1.46

The worst 5 of 33 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	628	PHE	N-CA-CB	14.14	136.05	110.60
3	A	480	GLU	N-CA-C	-12.88	76.22	111.00

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	629	HIS	N-CA-C	11.21	141.26	111.00
2	C	2010	DG	P-O3'-C3'	11.13	133.06	119.70
3	A	595	GLN	N-CA-C	-10.61	82.36	111.00

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
3	A	628	PHE	CA

5 of 11 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	A	421	GLU	Peptide
1	B	1007	DT	Sidechain
1	B	1009	DC	Sidechain
2	C	2007	DT	Sidechain
2	C	2016	DC	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	B	363	0	206	42	0
2	C	369	0	206	21	0
3	A	4385	0	4357	451	2
All	All	5117	0	4769	504	2

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 51.

The worst 5 of 504 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:A:617:VAL:HG22	3:A:628:PHE:CD1	1.07	1.57
3:A:422:GLY:N	3:A:422:GLY:CA	1.68	1.53
3:A:617:VAL:CG2	3:A:628:PHE:CD1	1.89	1.52

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:A:422:GLY:H	3:A:423:PRO:CD	1.21	1.51
3:A:543:CYS:CB	3:A:543:CYS:SG	2.08	1.41

All (2) 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)
3:A:165:ASP:OD2	3:A:165:ASP:OD2[3_655]	1.82	0.38
3:A:438:GLN:OE1	3:A:670:ASN:ND2[1_455]	2.15	0.05

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
3	A	538/575 (94%)	428 (80%)	78 (14%)	32 (6%)	1 5

5 of 32 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	202	MET
3	A	548	ASN
3	A	580	GLY
3	A	594	GLN
3	A	596	PRO

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
3	A	480/528 (91%)	457 (95%)	23 (5%)	25 58

5 of 23 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	596	PRO
3	A	616	TRP
3	A	614	PHE
3	A	628	PHE
3	A	409	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 21 such sidechains are listed below:

Mol	Chain	Res	Type
3	A	408	GLN
3	A	483	ASN
3	A	629	HIS
3	A	505	GLN
3	A	438	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)

1 non-standard protein/DNA/RNA residue is modelled in this entry.

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	PTR	A	701	3	15,16,17	1.23	0	19,22,24	1.11	1 (5%)

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	PTR	A	701	3	-	0/10/11/13	0/1/1/1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	701	PTR	CD2-CE2-CZ	-2.20	117.05	119.73

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
3	A	2
1	B	1
2	C	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	B	1010:DC	O3'	1011:DG	P	3.61
1	A	608:ARG	C	609:GLU	N	1.12
1	A	623:GLY	C	624:GLY	N	0.90
1	C	2010:DG	O3'	2011:DG	P	0.66

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	B	18/18 (100%)	0.05	2 (11%)	5	4	14, 32, 56, 64	18 (100%)
2	C	18/18 (100%)	0.14	2 (11%)	5	4	15, 31, 59, 66	18 (100%)
3	A	544/575 (94%)	-0.16	26 (4%)	30	27	5, 37, 103, 119	0
All	All	580/611 (94%)	-0.14	30 (5%)	27	23	5, 37, 103, 119	36 (6%)

The worst 5 of 30 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	416	GLY	12.3
3	A	660	PRO	6.9
3	A	627	ASP	6.3
3	A	654	MET	6.0
1	B	1001	DA	4.9

6.2 Non-standard residues in protein, DNA, RNA chains (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, 95th 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(Å^2)	Q<0.9
3	PTR	A	701	16/17	0.96	0.15	20,30,54,57	0

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.