



wwPDB X-ray Structure Validation Summary Report ⓘ

May 16, 2020 – 04:39 am BST

PDB ID : 4GFV
Title : PTPN18 in complex with HER2-pY1196 phosphor-peptides
Authors : Wang, H.M.; Yang, F.; Du, Y.J.; Yang, D.X.; Zhang, Y.; Yu, X.; Sun, J.P.
Deposited on : 2012-08-04
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

<https://www.wwpdb.org/validation/2017/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 : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.11
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) : 2.11

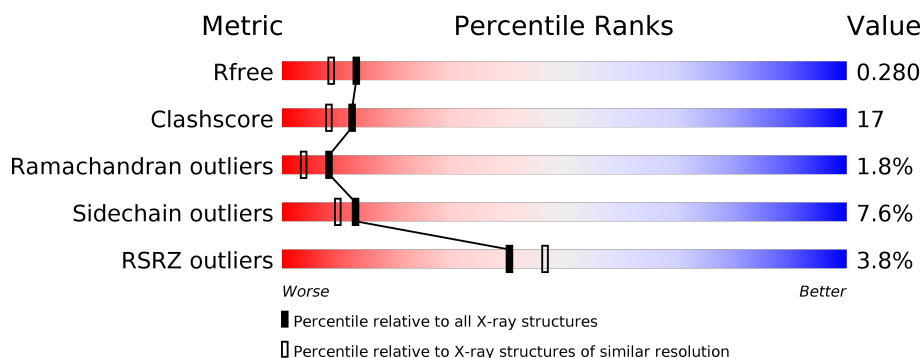
1 Overall quality at a glance ⓘ

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}	130704	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (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 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 $\leq 5\%$. 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	297	<div> <div>2%</div> <div> <div></div> <div>69%</div> <div>20%</div> <div>6%</div> <div>• •</div> </div> </div>
1	B	297	<div> <div>5%</div> <div> <div></div> <div>63%</div> <div>26%</div> <div>8%</div> <div>•</div> </div> </div>
2	E	6	<div> <div></div> <div> <div>33%</div> <div>33%</div> <div>17%</div> <div>17%</div> </div> </div>
2	F	6	<div> <div></div> <div> <div>33%</div> <div>33%</div> <div>17%</div> <div>17%</div> </div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	PTR	E	1196	-	-	X	-

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 4816 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 Tyrosine-protein phosphatase non-receptor type 18.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	286	Total	C	N	O	S	0	1	0
			2291	1449	410	415	17			
1	B	288	Total	C	N	O	S	0	1	0
			2305	1457	411	420	17			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	4	ALA	-	EXPRESSION TAG	UNP Q99952
A	5	ALA	-	EXPRESSION TAG	UNP Q99952
A	229	SER	CYS	ENGINEERED MUTATION	UNP Q99952
B	4	ALA	-	EXPRESSION TAG	UNP Q99952
B	5	ALA	-	EXPRESSION TAG	UNP Q99952
B	229	SER	CYS	ENGINEERED MUTATION	UNP Q99952

- Molecule 2 is a protein called HER2-pY1196 phosphor-peptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	6	Total	C	N	O	P	0	0	0
			54	34	6	13	1			
2	F	5	Total	C	N	O	P	0	0	0
			47	29	5	12	1			

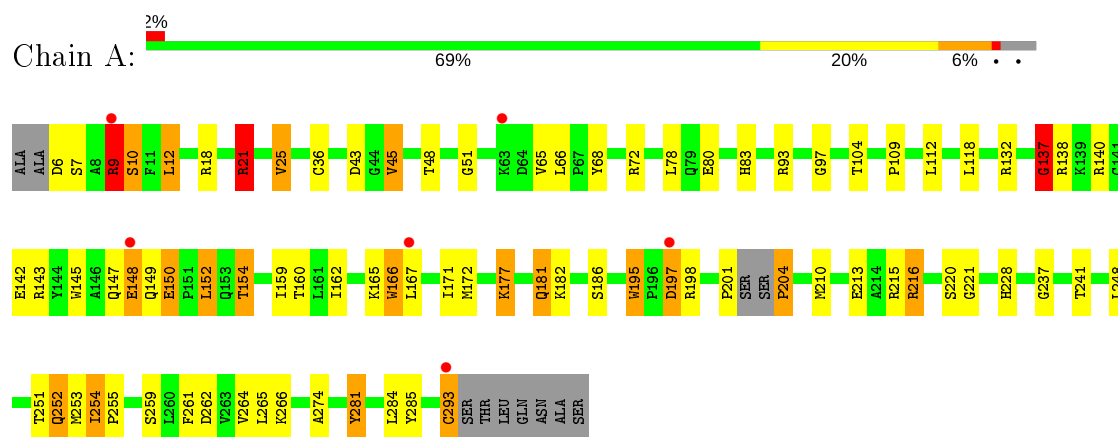
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	65	Total	O	0	0
			65	65		
3	B	52	Total	O	0	0
			52	52		
3	F	2	Total	O	0	0
			2	2		

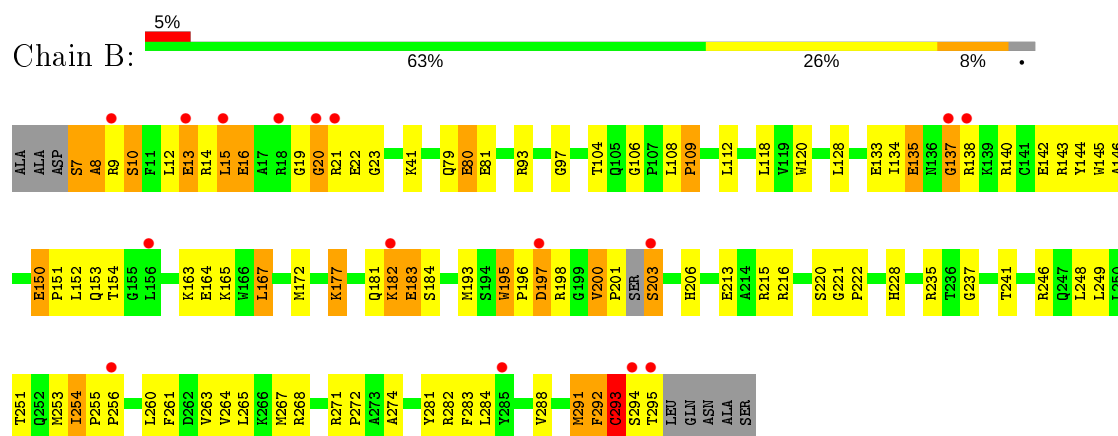
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA 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.

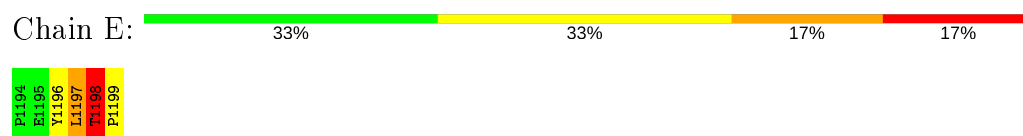
- Molecule 1: Tyrosine-protein phosphatase non-receptor type 18



- Molecule 1: Tyrosine-protein phosphatase non-receptor type 18

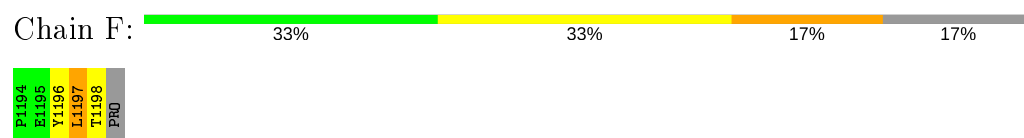


- Molecule 2: HER2-pY1196 phosphor-peptide



- Molecule 2: HER2-pY1196 phosphor-peptide

Chain F:



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	74.35Å 91.34Å 92.97Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.00 – 2.10 49.00 – 2.10	Depositor EDS
% Data completeness (in resolution range)	95.6 (49.00-2.10) 95.3 (49.00-2.10)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.99 (at 2.10Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.6.2_432)	Depositor
R, R_{free}	0.246 , 0.293 0.250 , 0.280	Depositor DCC
R_{free} test set	1810 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å ²)	19.6	Xtriage
Anisotropy	0.893	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 47.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.002 for -h,l,k	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	4816	wwPDB-VP
Average B, all atoms (Å ²)	28.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 53.31 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 4.2927e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹ Intensities estimated from amplitudes.

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

5 Model quality

5.1 Standard geometry

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	A	1.26	10/2343 (0.4%)	0.90	13/3165 (0.4%)
1	B	1.38	11/2354 (0.5%)	0.95	14/3182 (0.4%)
2	E	0.53	0/38	0.94	0/50
2	F	0.49	0/30	0.82	0/38
All	All	1.32	21/4765 (0.4%)	0.93	27/6435 (0.4%)

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	0	3
1	B	0	1
All	All	0	4

The worst 5 of 21 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	36	CYS	CB-SG	-8.36	1.68	1.82
1	A	197	ASP	CB-CG	-7.88	1.35	1.51
1	B	133	GLU	CB-CG	-7.84	1.37	1.52
1	A	45	VAL	CB-CG1	-7.54	1.37	1.52
1	A	285	TYR	CD1-CE1	-7.02	1.28	1.39

The worst 5 of 27 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	197	ASP	CB-CG-OD1	8.15	125.63	118.30
1	A	43	ASP	CB-CG-OD1	7.49	125.04	118.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	216[A]	ARG	CA-C-O	-6.87	105.67	120.10
1	A	216[B]	ARG	CA-C-O	-6.87	105.67	120.10
1	B	197	ASP	CB-CG-OD2	-6.83	112.15	118.30

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	137	GLY	Peptide
1	A	216[A]	ARG	Mainchain
1	A	216[B]	ARG	Mainchain
1	B	293	CYS	Peptide

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	2291	0	2276	62	0
1	B	2305	0	2285	89	0
2	E	54	0	45	9	0
2	F	47	0	38	7	0
3	A	65	0	0	2	0
3	B	52	0	0	3	0
3	F	2	0	0	0	0
All	All	4816	0	4644	158	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 17.

The worst 5 of 158 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:149:GLN:N	1:A:150:GLU:OE1	1.87	1.07
1:B:197:ASP:OD2	2:E:1196:PTR:CZ	2.03	1.05
1:B:197:ASP:OD2	2:E:1196:PTR:CE1	2.04	1.05

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:137:GLY:HA3	1:A:138:ARG:HG2	1.39	1.01
1:B:197:ASP:OD1	3:B:411:HOH:O	1.80	1.00

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	283/297 (95%)	267 (94%)	13 (5%)	3 (1%)	14	9
1	B	285/297 (96%)	268 (94%)	11 (4%)	6 (2%)	7	3
2	E	3/6 (50%)	2 (67%)	0	1 (33%)	0	0
2	F	2/6 (33%)	2 (100%)	0	0	100	100
All	All	573/606 (95%)	539 (94%)	24 (4%)	10 (2%)	8	4

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	80	GLU
1	A	148	GLU
1	B	21	ARG
1	B	8	ALA
1	B	80	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	246/255 (96%)	228 (93%)	18 (7%)	14	11
1	B	248/255 (97%)	231 (93%)	17 (7%)	15	12
2	E	5/5 (100%)	3 (60%)	2 (40%)	0	0
2	F	4/5 (80%)	3 (75%)	1 (25%)	0	0
All	All	503/520 (97%)	465 (92%)	38 (8%)	13	10

5 of 38 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	266	LYS
1	B	15	LEU
2	E	1197	LEU
1	B	10	SER
1	B	16	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	252	GLN
1	B	83	HIS
1	B	147	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

2 non-standard protein/DNA/RNA residues are 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
2	PTR	F	1196	2	15,16,17	2.20	2 (13%)	19,22,24	0.83	1 (5%)
2	PTR	E	1196	2	15,16,17	1.97	1 (6%)	19,22,24	0.77	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
2	PTR	F	1196	2	-	1/10/11/13	0/1/1/1
2	PTR	E	1196	2	-	0/10/11/13	0/1/1/1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	F	1196	PTR	OH-CZ	-7.79	1.23	1.40
2	E	1196	PTR	OH-CZ	-7.36	1.23	1.40
2	F	1196	PTR	P-OH	2.48	1.63	1.59

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	F	1196	PTR	O2P-P-OH	2.54	113.20	105.24
2	E	1196	PTR	O2P-P-OH	2.36	112.61	105.24

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	F	1196	PTR	O-C-CA-CB

There are no ring outliers.

2 monomers are involved in 11 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	F	1196	PTR	4	0
2	E	1196	PTR	7	0

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	286/297 (96%)	0.09	6 (2%) 63 68	15, 26, 44, 61	0
1	B	288/297 (96%)	0.17	16 (5%) 24 29	14, 26, 50, 62	0
2	E	5/6 (83%)	1.10	0 100 100	30, 49, 57, 73	0
2	F	4/6 (66%)	1.12	0 100 100	37, 46, 55, 60	0
All	All	583/606 (96%)	0.14	22 (3%) 40 46	14, 26, 49, 73	0

The worst 5 of 22 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	293	CYS	5.2
1	B	197	ASP	5.1
1	B	294	SER	4.0
1	B	203	SER	3.8
1	A	197	ASP	3.5

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
2	PTR	F	1196	16/17	0.95	0.13	13,19,29,38	0
2	PTR	E	1196	16/17	0.96	0.12	15,20,33,34	0

6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.