Full wwPDB X-ray Structure Validation Report ⓘ

Jul 14, 2016 – 07:29 PM EDT

PDB ID	:	5HAX
Title	:	Crystal structure of Chaetomium thermophilum Nup170 NTD-Nup53 complex
Authors	:	Lin, D.H.; Mobbs, G.; Hoelz, A.
Deposited on	:	2015-12-31
Resolution	:	2.10 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 ①) were used in the production of this report:

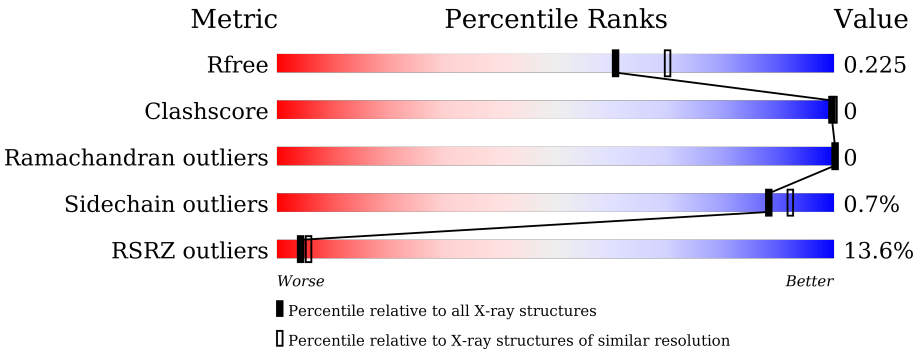
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027790
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027790

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 2.10 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 <=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	742	12% 93% · 6%
2	B	33	33% 39% · 58%

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 criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	PO4	A	903	-	-	-	X
3	PO4	A	904	-	-	-	X
4	EDO	A	905	-	-	-	X

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 11587 atoms, of which 5589 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 Nucleoporin NUP170.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	698	10917	3465	5447	953	1036	16	0	1	0

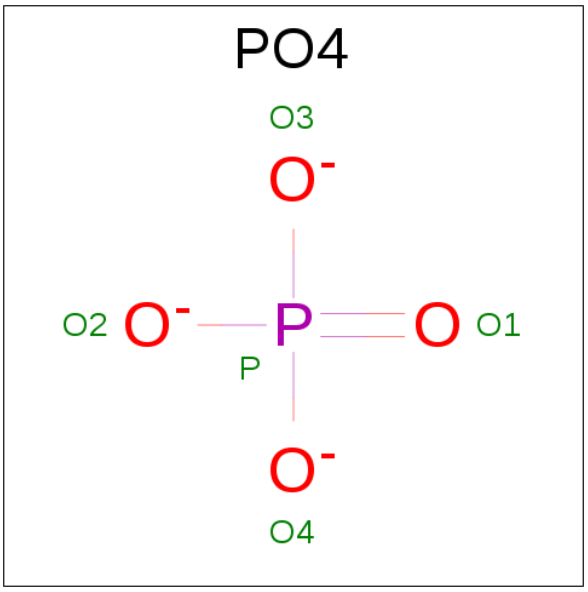
There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	73	SER	-	expression tag	UNP G0S7B6
A	?	-	PRO	deletion	UNP G0S7B6
A	?	-	GLY	deletion	UNP G0S7B6
A	?	-	TRP	deletion	UNP G0S7B6
A	?	-	SER	deletion	UNP G0S7B6
A	?	-	ALA	deletion	UNP G0S7B6
A	?	-	VAL	deletion	UNP G0S7B6
A	?	-	VAL	deletion	UNP G0S7B6
A	?	-	PRO	deletion	UNP G0S7B6
A	?	-	SER	deletion	UNP G0S7B6
A	?	-	LEU	deletion	UNP G0S7B6
A	?	-	ALA	deletion	UNP G0S7B6
A	?	-	GLY	deletion	UNP G0S7B6
A	?	-	LEU	deletion	UNP G0S7B6

- Molecule 2 is a protein called Nucleoporin NUP53.

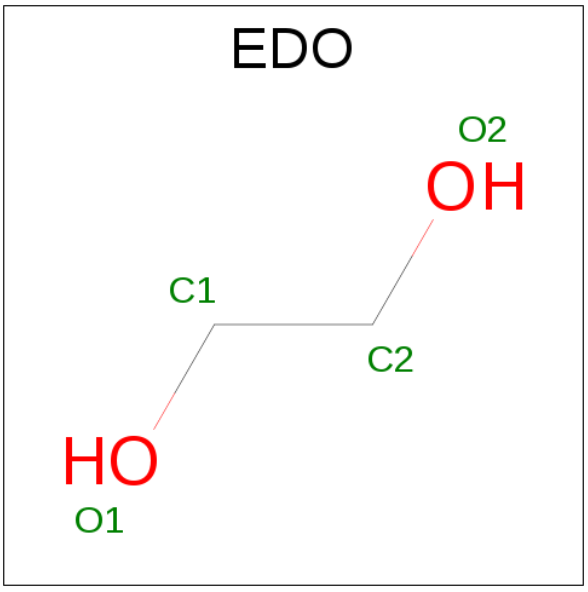
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
2	B	14	239	73	128	19	18	1	0	0	0

- Molecule 3 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).



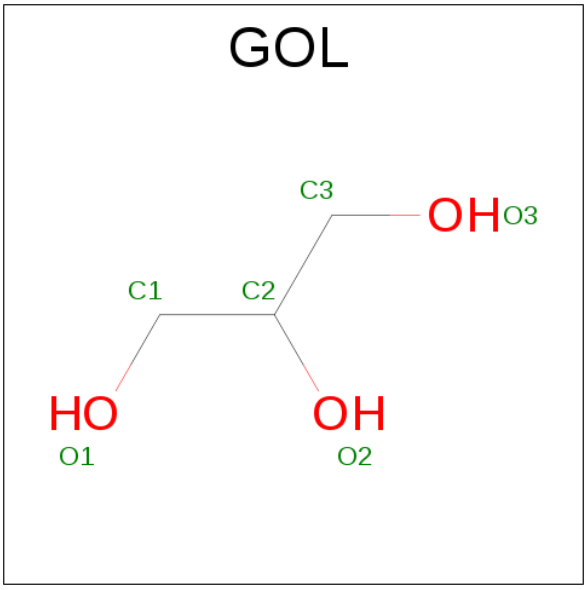
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total O P 5 4 1	0	0
3	A	1	Total O P 5 4 1	0	0
3	A	1	Total O P 5 4 1	0	0
3	A	1	Total O P 5 4 1	0	0

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total	C	H	O	0	0
			10	2	6	2		

- Molecule 5 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	1	Total	C	H	O	0	0
			14	3	8	3		

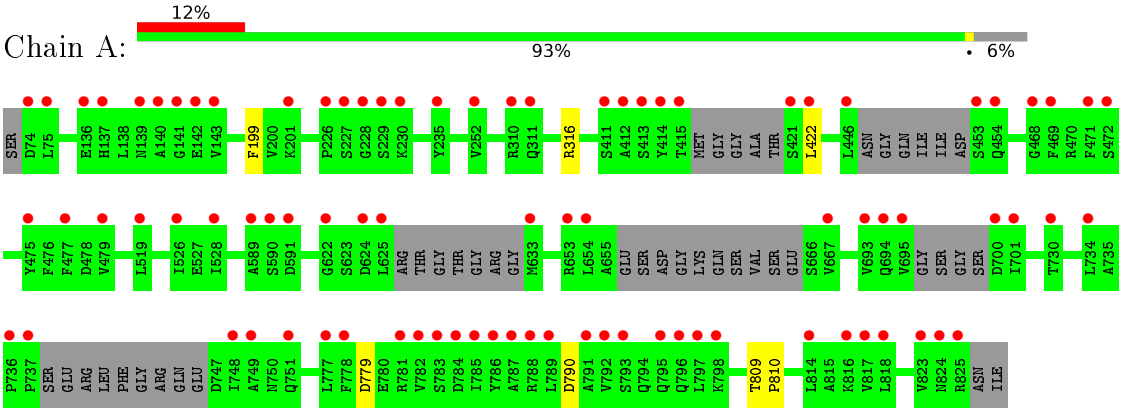
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	385	Total	O	0	2
			387	387		

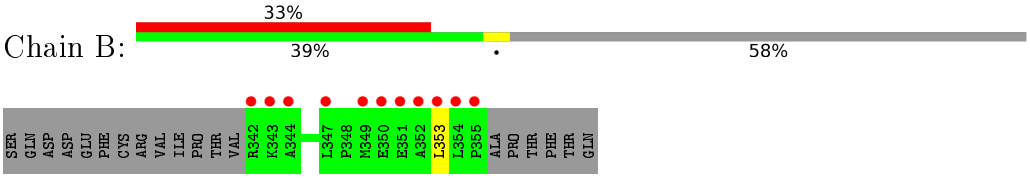
3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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: Nucleoporin NUP170



- Molecule 2: Nucleoporin NUP53



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	69.23Å 106.16Å 120.08Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.75 – 2.10 39.75 – 2.10	Depositor EDS
% Data completeness (in resolution range)	100.0 (39.75-2.10) 100.0 (39.75-2.10)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.79 (at 2.10Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, R_{free}	0.191 , 0.227 0.188 , 0.225	Depositor DCC
R_{free} test set	2617 reflections (5.00%)	DCC
Wilson B-factor (Å^2)	48.2	Xtriage
Anisotropy	0.271	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 59.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	11587	wwPDB-VP
Average B, all atoms (Å^2)	68.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.17% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, PO4, EDO

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	0.24	0/5583	0.43	0/7573
2	B	0.25	0/112	0.48	0/149
All	All	0.24	0/5695	0.43	0/7722

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	5470	5447	5439	2	0
2	B	111	128	127	1	0
3	A	20	0	0	0	0
4	A	4	6	6	0	0
5	A	6	8	8	0	0
6	A	387	0	0	0	0
All	All	5998	5589	5580	2	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 0.

All (2) 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:199:PHE:HA	2:B:353:LEU:HA	1.99	0.44
1:A:809:THR:HB	1:A:810:PRO:HD3	1.98	0.44

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	685/742 (92%)	669 (98%)	16 (2%)	0	100	100
2	B	12/33 (36%)	11 (92%)	1 (8%)	0	100	100
All	All	697/775 (90%)	680 (98%)	17 (2%)	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	600/634 (95%)	596 (99%)	4 (1%)	88	92
2	B	12/30 (40%)	12 (100%)	0	100	100
All	All	612/664 (92%)	608 (99%)	4 (1%)	88	92

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	316	ARG
1	A	422	LEU
1	A	779	ASP
1	A	790	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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)

6 ligands 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	$\#\|Z\| > 2$
3	PO4	A	901	-	4,4,4	0.69	0	6,6,6	0.23	0
3	PO4	A	902	-	4,4,4	0.67	0	6,6,6	0.23	0
3	PO4	A	903	-	4,4,4	0.64	0	6,6,6	0.23	0
3	PO4	A	904	-	4,4,4	0.55	0	6,6,6	0.24	0
4	EDO	A	905	-	3,3,3	0.45	0	2,2,2	0.39	0
5	GOL	A	906	-	5,5,5	0.46	0	5,5,5	0.17	0

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	PO4	A	901	-	-	0/0/0/0	0/0/0/0
3	PO4	A	902	-	-	0/0/0/0	0/0/0/0
3	PO4	A	903	-	-	0/0/0/0	0/0/0/0
3	PO4	A	904	-	-	0/0/0/0	0/0/0/0
4	EDO	A	905	-	-	0/1/1/1	0/0/0/0
5	GOL	A	906	-	-	0/4/4/4	0/0/0/0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

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, 95^{th} 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	698/742 (94%)	0.58	86 (12%)	5	7	35, 55, 105, 134	0
2	B	14/33 (42%)	3.50	11 (78%)	0	0	93, 110, 125, 128	0
All	All	712/775 (91%)	0.63	97 (13%)	4	6	35, 55, 109, 134	0

All (97) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	342	ARG	8.0
1	A	701	ILE	7.0
1	A	414	TYR	6.9
1	A	695	VAL	6.0
1	A	413	SER	5.8
1	A	143	VAL	5.7
2	B	355	PRO	5.7
1	A	782	VAL	5.7
1	A	625	LEU	5.6
1	A	415	THR	5.6
1	A	141	GLY	5.4
1	A	624	ASP	4.8
2	B	353	LEU	4.7
1	A	787	ALA	4.7
1	A	421	SER	4.6
1	A	788	ARG	4.5
1	A	700	ASP	4.5
1	A	140	ALA	4.2
1	A	783	SER	4.2
1	A	825	ARG	4.2
1	A	412	ALA	4.1
1	A	693	VAL	4.0
1	A	789	LEU	4.0
1	A	784	ASP	4.0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	653	ARG	4.0
1	A	798	LYS	3.9
1	A	818	LEU	3.8
1	A	748	ILE	3.8
2	B	349	MET	3.7
2	B	344	ALA	3.7
1	A	654	LEU	3.5
2	B	351	GLU	3.5
1	A	142	GLU	3.5
1	A	201	LYS	3.5
1	A	781	ARG	3.5
1	A	136	GLU	3.5
1	A	734	LEU	3.4
2	B	347	LEU	3.4
2	B	354	LEU	3.4
1	A	139	ASN	3.3
1	A	411	SER	3.3
1	A	422	LEU	3.3
1	A	778	PHE	3.2
1	A	785	ILE	3.2
1	A	446	LEU	3.2
1	A	137	HIS	3.1
1	A	74	ASP	3.1
1	A	227	SER	3.0
1	A	591	ASP	3.0
1	A	791	ALA	3.0
1	A	75	LEU	2.9
1	A	823	VAL	2.8
2	B	352	ALA	2.8
1	A	786	TYR	2.8
1	A	519	LEU	2.8
1	A	777	LEU	2.8
1	A	694	GLN	2.8
1	A	454	GLN	2.8
1	A	749	ALA	2.7
1	A	528	ILE	2.7
1	A	792	VAL	2.6
2	B	350	GLU	2.6
1	A	226	PRO	2.6
1	A	796	GLN	2.6
1	A	453	SER	2.6
2	B	343	LYS	2.6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	471	PHE	2.6
1	A	793	SER	2.5
1	A	817	VAL	2.5
1	A	667	VAL	2.4
1	A	824	ASN	2.4
1	A	235	TYR	2.4
1	A	751	GLN	2.4
1	A	230	LYS	2.4
1	A	736	PRO	2.4
1	A	633	MET	2.4
1	A	310	ARG	2.4
1	A	622	GLY	2.4
1	A	472	SER	2.3
1	A	590	SER	2.3
1	A	795	GLN	2.3
1	A	228	GLY	2.3
1	A	475	TYR	2.2
1	A	814	LEU	2.2
1	A	526	ILE	2.2
1	A	229	SER	2.2
1	A	477	PHE	2.2
1	A	468	GLY	2.1
1	A	252	VAL	2.1
1	A	479	VAL	2.1
1	A	816	LYS	2.1
1	A	737	PRO	2.0
1	A	469	PHE	2.0
1	A	797	LEU	2.0
1	A	589	ALA	2.0
1	A	730	THR	2.0
1	A	311	GLN	2.0

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

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95^{th} 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	LLDF	B-factors(Å^2)	Q<0.9
3	PO4	A	904	5/5	0.90	0.52	29.38	57,73,80,136	0
3	PO4	A	903	5/5	0.82	0.46	8.14	53,78,88,106	0
4	EDO	A	905	4/4	0.91	0.24	3.40	51,61,71,74	0
5	GOL	A	906	6/6	0.80	0.15	0.73	74,89,96,96	0
3	PO4	A	901	5/5	0.84	0.16	-0.04	74,80,83,92	0
3	PO4	A	902	5/5	0.96	0.22	-	69,74,92,93	0

6.5 Other polymers

There are no such residues in this entry.