



wwPDB X-ray Structure Validation Summary Report

Mar 1, 2014 – 12:46 AM GMT

PDB ID : 1X9N
Title : Crystal Structure of Human DNA Ligase I bound to 5'-adenylated, nicked DNA
Authors : Pascal, J.M.; O'Brien, P.J.; Tomkinson, A.E.; Ellenberger, T.
Deposited on : 2004-08-23
Resolution : 3.00 Å(reported)

This is a wwPDB 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/ValidationPDFNotes.html>

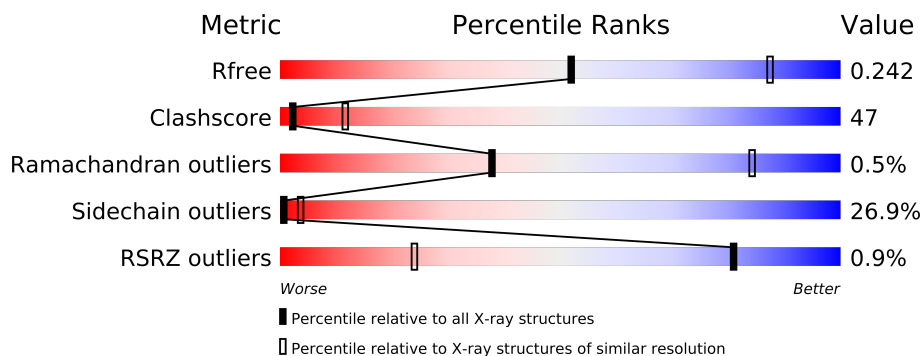
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable22639
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable22683

1 Overall quality at a glance

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))
R_{free}	66092	1216 (3.00-3.00)
Clashscore	79885	1594 (3.00-3.00)
Ramachandran outliers	78287	1537 (3.00-3.00)
Sidechain outliers	78261	1540 (3.00-3.00)
RSRZ outliers	66119	1217 (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 on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	B	13	
2	C	15	
3	D	28	
4	A	688	

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 5730 atoms, of which 0 are hydrogen and 0 are deuterium.

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 dideoxy terminated DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	11	Total	C	N	O	P	0	0	0
			222	107	40	65	10			

- Molecule 2 is a DNA chain called 5'-phosphorylated DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	9	Total	C	N	O	P	0	0	0
			187	88	35	55	9			

- Molecule 3 is a DNA chain called template DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	20	Total	C	N	O	P	0	0	0
			404	192	78	115	19			

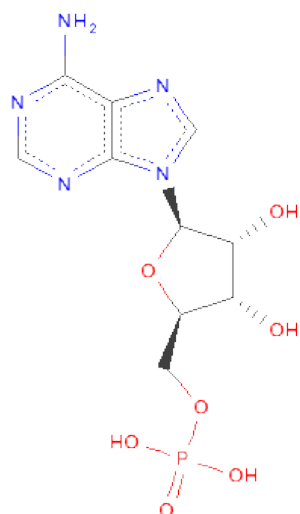
- Molecule 4 is a protein called DNA ligase I.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
4	A	632	Total	C	N	O	S	Se	0	0	0
			4894	3109	849	920	10	6			

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	232	MET	-	INITIATING METHIONINE	UNP P18858
A	308	MSE	MET	MODIFIED RESIDUE	UNP P18858
A	393	MSE	MET	MODIFIED RESIDUE	UNP P18858
A	480	MSE	MET	MODIFIED RESIDUE	UNP P18858
A	501	MSE	MET	MODIFIED RESIDUE	UNP P18858
A	543	MSE	MET	MODIFIED RESIDUE	UNP P18858
A	723	MSE	MET	MODIFIED RESIDUE	UNP P18858

- Molecule 5 is ADENOSINE MONOPHOSPHATE (three-letter code: AMP) (formula: $C_{10}H_{14}N_5O_7P$).

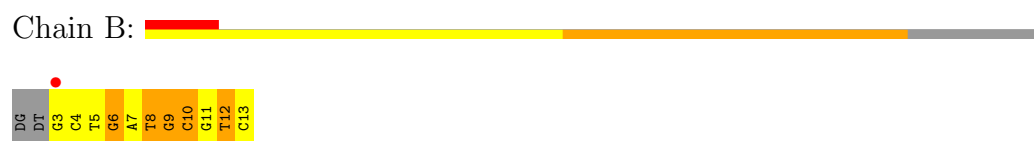


Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
5	C	1	23	10	5	7	1	0	0

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.

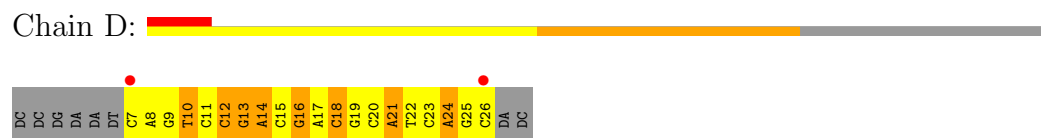
- Molecule 1: dideoxy terminated DNA



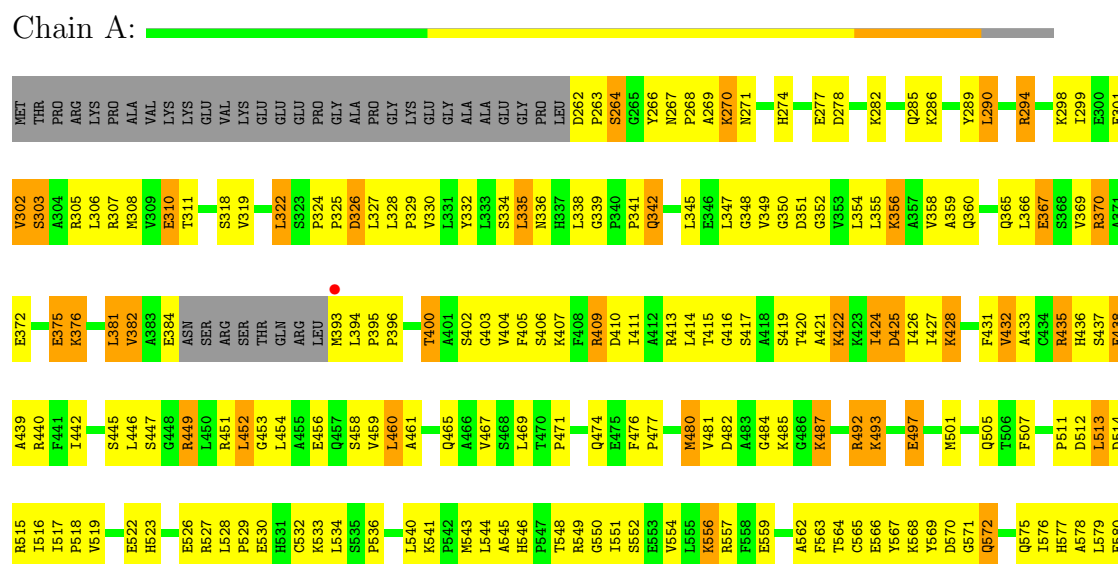
- Molecule 2: 5'-phosphorylated DNA



- Molecule 3: template DNA



- Molecule 4: DNA ligase I



GLY GLU ASP SER GLY SER ASP PRO GLU ASP THR TYR	V844	K845	C846	A847	D848	L849	S850	L851	S852	P853	I854	Y855	P856	A857											V862	D863	S864	D865	K866	G867	I868	S869	L870	R871	F872	P873	R874	F875	I876	R877	V878											D881	K882	Q883	P884	E885	Q886	A887	T888	S889	S890	A891	Q892	V893	A894	C895	L896	Y897	R898	K899	Q900	S901	GLN	ILE	GLN	ASN	GLN	GLN																			
	S782	Y783	D784	E785	D786	S787	E788	E789	L790	Q791	A792	I793	C794	K795	L796	G797	T798	G799	R800	S801	D802			L805	H808	H809	Q810	S811	L812	K813	A814	L815	V816	L817	P818	S819	D820	E821	P822	Y823	V824	I825	R826	D827	Q828	A829	V830	I831	P832	D833	H834	W835	L836	D837	P838	S839	A840	W841	L842	E843																																					
	Q713	S714	K715	L716	D717	S718	C719	E720	E721	L722	M723	V724	K725	T726	L727											T732	Y733	E734	I735	K737	S739	H740	N741	W742	L743	K744	L745	K746	K747	D748	Y749	L750	D751	G752	M753	G754	D755	T756	L757	D758	V759	W760	V761	L762	Y765											R768	G769	K770	R771											R774	L779	L780	A781														
	D647	A648	S649	E650	I651	R652	V653	Q654	V655	C656	L657	Y658											D661	L662	I663	Y664	L665	N666											L670	V671	K672	E673	P674	L675	S676	R677	R678	R679	Q680	L681	L682	R683	E684	N685	F686	V687	E688	T689											F693	V694	F695	A696	S697	L698	P699	D700	T701	K702	D703	I704	E705	T706	Q706	I707											F710	L711	E712
	V684	K685	I686	R689	U690	Q691	E692	D693	L694	K695	Y697	Y698	P699	D600	I601	Y602	S603	R604	I605	P606	K607	L608	K609	L610	P611	S612	V613	T614	S615	F616	L617	L618	D619	T620	E621	A622	V623	A624	W625	D626											K629	K630	Q631	Q632	P633	P634	F635	Q636	V637	L638	T639	T640	R641	K642	R643	K644	E645	V646																													

4 Data and refinement statistics

Property	Value	Source
Space group	P 63	Depositor
Cell constants a, b, c, α , β , γ	161.89Å 161.89Å 88.45Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	20.00 – 3.00 19.99 – 3.00	Depositor EDS
% Data completeness (in resolution range)	50.0 (20.00-3.00) 99.0 (19.99-3.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.08	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.60 (at 2.98Å)	Xtriage
Refinement program	REFMAC 5.1.80	Depositor
R, R_{free}	0.235 , 0.268 0.241 , 0.242	Depositor DCC
R_{free} test set	1287 reflections (5.16%)	DCC
Wilson B-factor (Å ²)	78.1	Xtriage
Anisotropy	0.568	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 25.2	EDS
Estimated twinning fraction	0.026 for h,-h-k,-l	Xtriage
L-test for twinning	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.34$	Xtriage
Outliers	0 of 26244 reflections	Xtriage
F_o, F_c correlation	0.90	EDS
Total number of atoms	5730	wwPDB-VP
Average B, all atoms (Å ²)	41.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.83% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

5 Model quality ⓘ

5.1 Standard geometry ⓘ

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

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	1.01	0/228	2.16	12/351 (3.4%)
2	C	1.40	3/209 (1.4%)	2.04	9/321 (2.8%)
3	D	1.15	0/453	1.95	19/696 (2.7%)
4	A	0.58	0/4987	0.82	15/6753 (0.2%)
All	All	0.71	3/5877 (0.1%)	1.12	55/8121 (0.7%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	1	DG	C3'-O3'	-7.01	1.34	1.44
2	C	3	DC	C3'-O3'	-5.51	1.36	1.44
2	C	4	DG	C3'-O3'	-5.22	1.37	1.44

The worst 5 of 55 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	8	DT	O4'-C1'-N1	18.87	121.21	108.00
3	D	16	DG	O4'-C4'-C3'	-9.17	100.50	106.00
3	D	15	DC	O4'-C4'-C3'	-8.89	100.67	106.00
3	D	11	DC	O4'-C1'-N1	-7.93	102.45	108.00
3	D	21	DA	O4'-C1'-N9	-7.78	102.56	108.00

There are no chirality outliers.

There are no planarity outliers.

5.2 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 hydrogens

added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	222	0	126	18	0
2	C	187	0	102	15	0
3	D	404	0	224	30	0
4	A	4894	0	4925	468	0
5	C	23	0	12	3	0
All	All	5730	0	5389	520	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 47.

The worst 5 of 520 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
4:A:480:MSE:CE	4:A:480:MSE:SE	2.15	1.44
4:A:548:THR:CG2	4:A:745:LEU:HD13	1.72	1.19
2:C:7:DC:H2"	2:C:8:DT:H5"	1.24	1.15
4:A:400:THR:HG22	4:A:403:GLY:H	1.08	1.14
4:A:551:ILE:CD1	4:A:707:ILE:HG22	1.79	1.13

There are no symmetry-related clashes.

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
4	A	628/688 (91%)	559 (89%)	66 (10%)	3 (0%)	38 84

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	A	748	ASP
4	A	753	VAL

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
4	A	644	LYS

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
4	A	525/576 (91%)	384 (73%)	141 (27%)	1 4

5 of 141 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	A	597	LYS
4	A	652	GLN
4	A	862	VAL
4	A	607	LYS
4	A	633	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
4	A	474	GLN
4	A	594	ASN
4	A	713	GLN
4	A	360	GLN
4	A	706	GLN

5.3.3 RNA ⓘ

There are no RNA chains in this entry.

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

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	$\# Z > 2$
1	DOC	B	13	1,3	17,19,20	0.99	1 (5%)	20,26,29	2.43	5 (25%)

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
1	DOC	B	13	1,3	-	0/5/18/19	0/2/2/2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	13	DOC	C6-C5	-2.26	1.32	1.38

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	13	DOC	C4'-O4'-C1'	-8.25	106.61	110.05
1	B	13	DOC	C2'-C1'-N1	-3.81	105.04	112.66
1	B	13	DOC	C3'-C2'-C1'	-3.19	99.26	102.80
1	B	13	DOC	C2-N3-C4	2.32	118.92	115.57
1	B	13	DOC	O4'-C1'-C2'	2.01	108.77	106.37

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

5.6 Ligand geometry ⓘ

1 ligand 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$
5	AMP	C	100	2	25,25,25	1.02	2 (8%)	38,38,38	2.03	5 (13%)

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
5	AMP	C	100	2	-	0/10/26/26	0/1/3/3

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	C	100	AMP	C2-N3	3.09	1.38	1.32
5	C	100	AMP	C2-N1	2.20	1.38	1.33

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	C	100	AMP	N3-C2-N1	-9.38	120.87	128.71
5	C	100	AMP	N3-C4-N9	4.09	132.81	125.43
5	C	100	AMP	P-O5'-C5'	2.77	126.19	118.19
5	C	100	AMP	C5-C4-N3	-2.69	119.85	125.70
5	C	100	AMP	C2-N3-C4	2.09	119.95	114.01

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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, 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	11/13 (84%)	0.45	1 (9%) 9 2	39, 49, 65, 66	0
2	C	9/15 (60%)	0.37	1 (11%) 6 2	40, 46, 55, 61	0
3	D	20/28 (71%)	0.12	2 (10%) 8 2	38, 46, 58, 70	0
4	A	632/688 (91%)	-0.05	2 (0%) 91 48	26, 41, 50, 61	0
All	All	672/744 (90%)	-0.03	6 (0%) 81 24	26, 41, 52, 70	0

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	3	DG	3.5
4	A	393	MSE	2.9
4	A	866	LYS	2.6
3	D	26	DC	2.5
3	D	7	DC	2.3

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

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, 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	RSR	LLDF	B-factors(Å ²)	Q<0.9
1	DOC	B	13	18/19	0.18	-0.50	35,37,42,42	0

6.3 Carbohydrates

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, 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	RSR	LLDF	B-factors(Å ²)	Q<0.9
5	AMP	C	100	23/23	0.14	-1.59	36,38,45,48	0

6.5 Other polymers

There are no such residues in this entry.