



Full wwPDB X-ray Structure Validation Report ⓘ

May 18, 2020 – 01:37 am BST

PDB ID : 6P1A
Title : Transcription antitermination factor Q21 in complex with Q21-binding-element DNA
Authors : Yin, Z.; Ebright, R.H.
Deposited on : 2019-05-19
Resolution : 2.84 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

<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

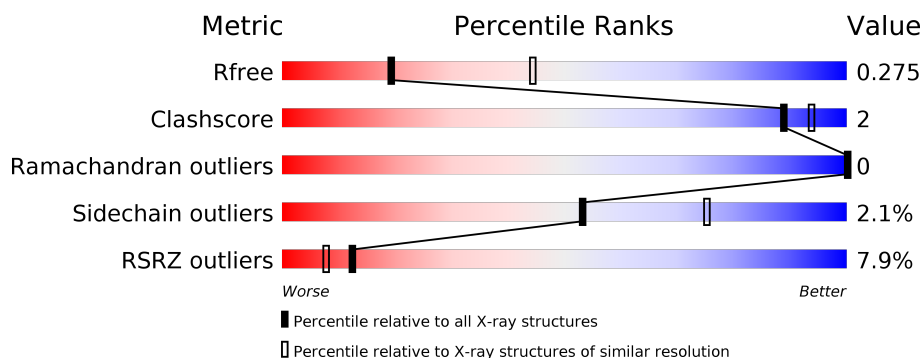
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.84 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	1031 (2.86-2.82)
Clashscore	141614	1078 (2.86-2.82)
Ramachandran outliers	138981	1050 (2.86-2.82)
Sidechain outliers	138945	1051 (2.86-2.82)
RSRZ outliers	127900	1019 (2.86-2.82)

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	162	<div> <div>2%</div> <div>90%</div> <div>6%</div> </div>
1	B	162	<div> <div>6%</div> <div>85%</div> <div>6%</div> <div>9%</div> </div>
1	C	162	<div> <div>10%</div> <div>87%</div> <div>6%</div> <div>7%</div> </div>
2	D	21	<div> <div>71%</div> <div>29%</div> </div>
2	F	21	<div> <div>29%</div> <div>90%</div> <div>10%</div> </div>
3	E	21	<div> <div>95%</div> <div>5%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	G	21	 A horizontal bar chart showing the quality of chain G. The bar is divided into two segments: a red segment on the left representing 33% and a green segment on the right representing 100%.

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 9963 atoms, of which 4595 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 Q protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	152	Total	C	H	N	O	S	0	5	0
			2501	794	1251	225	220	11			
1	B	147	Total	C	H	N	O	S	0	1	0
			2369	752	1185	212	209	11			
1	C	151	Total	C	H	N	O	S	0	0	0
			2389	763	1188	213	214	11			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	SER	-	expression tag	UNP Q9XJQ6
A	24	GLY	-	insertion	UNP Q9XJQ6
A	26	TRP	HIS	conflict	UNP Q9XJQ6
A	27	VAL	GLY	conflict	UNP Q9XJQ6
A	28	TYR	LEU	conflict	UNP Q9XJQ6
A	47	VAL	ILE	conflict	UNP Q9XJQ6
B	1	SER	-	expression tag	UNP Q9XJQ6
B	24	GLY	-	insertion	UNP Q9XJQ6
B	26	TRP	HIS	conflict	UNP Q9XJQ6
B	27	VAL	GLY	conflict	UNP Q9XJQ6
B	28	TYR	LEU	conflict	UNP Q9XJQ6
B	47	VAL	ILE	conflict	UNP Q9XJQ6
C	1	SER	-	expression tag	UNP Q9XJQ6
C	24	GLY	-	insertion	UNP Q9XJQ6
C	26	TRP	HIS	conflict	UNP Q9XJQ6
C	27	VAL	GLY	conflict	UNP Q9XJQ6
C	28	TYR	LEU	conflict	UNP Q9XJQ6
C	47	VAL	ILE	conflict	UNP Q9XJQ6

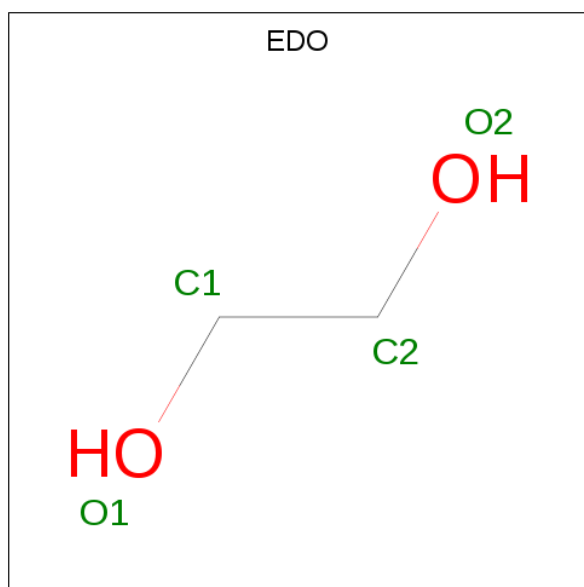
- Molecule 2 is a DNA chain called DNA (5'-D(P*CP*TP*CP*AP*TP*TP*GP*AP*GP*CP*AP*AP*AP*TP*GP*AP*GP*CP*AP*AP*G)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	D	21	Total	C	H	N	O	P	0	0	0
			670	206	236	85	122	21			
2	F	21	Total	C	H	N	O	P	0	0	0
			670	206	236	85	122	21			

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	E	21	Total	C	H	N	O	P	0	0	0
			665	205	241	71	128	20			
3	G	21	Total	C	H	N	O	P	0	0	0
			667	205	240	71	130	21			

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C₂H₆O₂).



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

- Molecule 5 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	1	Total 1	Cl 1	0	0
5	C	1	Total 1	Cl 1	0	0

- Molecule 1: Q protein



- | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| SER | GLY | ILE | ARG | GLU | LEU | N7 | Y28 | L32 | I40 | E48 | P49 | G50 | R51 | F61 | M10 | L11 | G12 | R13 | G14 | G15 | G16 | R17 | I18 | K153 | ARG | VAL | GLU | LYS | ILE | LYS | HIS | VAL | ALA |
|-----|-----|-----|-----|-----|-----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|

-
- Sequence logo for the 1000bp upstream region of the human POU3F1 gene. The y-axis represents information content in bits (0.00 to 0.15). The x-axis shows positions from -1000 to +1000 bp. The sequence is: SER, GLY, ILE, ARG, GLU, L6, K34, R35, M36, E48, Y52, M53, D61, F81, Y101, Y104, A105, R106, P107, R108, K109, M110, G115, G116, R117, I118, G119, T125, R154, E156, LYS, ILE, LYS, HIS, VAL, A. Red dots above the sequence indicate positions with significant enrichment.

- | | | | | | | |
|----|--|-----|--|-----|--|-----|
| C1 | | C10 | | C18 | | A20 |
| T2 | | A11 | | A19 | | G21 |

-
- Diagram illustrating a 16-bit bus system with 12 components (C1, T2, T5, T6, G7, A8, G9, C10, A11, G21). The components are connected to a central 16-bit bus. The bus is represented by a horizontal line with 16 segments. The components are connected to the bus via vertical lines. The components are color-coded: blue for C1, T2, T5, T6, G7, A8, and G9; yellow for C10 and A11; and green for G21. Red dots above the components indicate bus connections: C1, T2, T5, T6, G7, A8, and G9 have two dots; C10 and A11 have one dot; G21 has no dots.

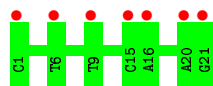
- Molecule 3: DNA (5'-D(*CP*TP*TP*GP*CP*TP*CP*AP*TP*TP*TP*GP*CP*TP*CP*AP*AP*TP*GP*AP*G)-3')

Chain E:  95% 5%



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

Chain G:  33% 100%



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α , β , γ	69.57Å 263.22Å 56.77Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.81 – 2.84 47.81 – 2.84	Depositor EDS
% Data completeness (in resolution range)	89.4 (47.81-2.84) 89.4 (47.81-2.84)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.90 (at 2.86Å)	Xtriage
Refinement program	PHENIX 1.14 _3260	Depositor
R, R_{free}	0.231 , 0.275 0.231 , 0.275	Depositor DCC
R_{free} test set	1999 reflections (8.73%)	wwPDB-VP
Wilson B-factor (Å ²)	49.1	Xtriage
Anisotropy	0.060	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 52.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.89	EDS
Total number of atoms	9963	wwPDB-VP
Average B, all atoms (Å ²)	62.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.61% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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 [i](#)

5.1 Standard geometry [i](#)

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

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	0.26	0/1296	0.43	0/1744
1	B	0.27	0/1214	0.45	0/1634
1	C	0.26	0/1228	0.44	0/1655
2	D	0.56	0/488	0.85	0/751
2	F	0.53	0/488	0.86	0/751
3	E	0.55	0/473	0.98	0/728
3	G	0.57	0/476	0.97	0/732
All	All	0.39	0/5663	0.66	0/7995

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 [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	1250	1251	1239	5	0
1	B	1184	1185	1185	3	0
1	C	1201	1188	1188	6	0
2	D	434	236	236	4	0
2	F	434	236	236	1	0
3	E	424	241	241	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	G	427	240	240	0	0
4	A	8	12	12	1	0
4	B	4	6	6	0	0
5	B	1	0	0	0	0
5	C	1	0	0	0	0
All	All	5368	4595	4583	19	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 2.

All (19) 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)
2:D:1:DC:H2'	2:D:2:DT:H72	1.68	0.74
1:C:104:VAL:HG12	1:C:104:VAL:O	1.95	0.66
1:A:109:LYS:NZ	4:A:201:EDO:O2	2.31	0.63
1:C:101:TYR:O	1:C:105:ALA:HB2	1.99	0.62
1:B:48:GLU:OE1	1:B:50:GLY:O	2.27	0.53
1:C:48:GLU:OE1	1:C:48:GLU:N	2.32	0.52
1:B:28:TYR:O	1:B:51:ARG:NH2	2.43	0.52
1:A:127:ARG:NH2	3:E:4:DG:N7	2.62	0.48
1:A:52:VAL:HG23	1:A:52:VAL:O	2.14	0.47
2:D:10:DC:H2''	2:D:11:DA:C8	2.51	0.45
2:D:1:DC:C2'	2:D:2:DT:H72	2.44	0.45
2:F:10:DC:H2''	2:F:11:DA:C8	2.52	0.45
1:C:110:MET:CE	1:C:110:MET:HA	2.48	0.43
1:C:108:ARG:CG	1:C:109:LYS:H	2.32	0.43
1:A:108:ARG:NH2	1:A:129:GLU:OE2	2.52	0.42
2:D:18:DC:H2''	2:D:19:DA:C8	2.54	0.42
1:A:53:MET:HA	1:A:53:MET:CE	2.51	0.40
1:B:32:LEU:HD23	1:B:40:ILE:HG21	2.02	0.40
1:C:108:ARG:CG	1:C:109:LYS:N	2.84	0.40

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	
1	A	155/162 (96%)	151 (97%)	4 (3%)	0	100	100
1	B	146/162 (90%)	139 (95%)	7 (5%)	0	100	100
1	C	149/162 (92%)	141 (95%)	8 (5%)	0	100	100
All	All	450/486 (93%)	431 (96%)	19 (4%)	0	100	100

There are no Ramachandran outliers to report.

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	
1	A	134/140 (96%)	132 (98%)	2 (2%)	65	82
1	B	127/140 (91%)	124 (98%)	3 (2%)	49	72
1	C	127/140 (91%)	124 (98%)	3 (2%)	49	72
All	All	388/420 (92%)	380 (98%)	8 (2%)	53	75

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	53	MET
1	A	81	PHE
1	B	81	PHE
1	B	111	LEU
1	B	117	ARG

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	53	MET
1	C	61	ASP
1	C	81	PHE

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](#)

Of 5 ligands modelled in this entry, 2 are monoatomic - leaving 3 for Mogul analysis.

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	$\# Z > 2$
4	EDO	B	201	-	3,3,3	0.50	0	2,2,2	0.18	0
4	EDO	A	202	-	3,3,3	0.48	0	2,2,2	0.26	0
4	EDO	A	201	-	3,3,3	0.48	0	2,2,2	0.27	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
4	EDO	B	201	-	-	0/1/1/1	-
4	EDO	A	202	-	-	1/1/1/1	-
4	EDO	A	201	-	-	1/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	202	EDO	O1-C1-C2-O2
4	A	201	EDO	O1-C1-C2-O2

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	201	EDO	1	0

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 ⓘ

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	A	152/162 (93%)	0.05	4 (2%) 56 51	13, 30, 91, 118	0
1	B	147/162 (90%)	0.22	9 (6%) 21 15	14, 35, 96, 118	0
1	C	151/162 (93%)	0.59	16 (10%) 6 3	20, 58, 114, 133	0
2	D	21/21 (100%)	0.03	0 100 100	21, 42, 63, 65	0
2	F	21/21 (100%)	1.55	6 (28%) 0 0	91, 119, 147, 159	0
3	E	21/21 (100%)	-0.07	0 100 100	18, 36, 70, 72	0
3	G	21/21 (100%)	1.52	7 (33%) 0 0	79, 124, 146, 156	0
All	All	534/570 (93%)	0.36	42 (7%) 12 8	13, 41, 118, 159	0

All (42) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	35	ARG	5.5
1	C	118	ILE	4.7
1	B	115	GLY	4.1
2	F	9	DG	3.9
1	B	49	PRO	3.7
1	A	155	VAL	3.6
2	F	1	DC	3.5
1	A	156	GLU	3.5
1	C	108	ARG	3.4
2	F	8	DA	3.2
1	C	155	VAL	3.2
1	C	119	GLN	3.2
1	C	117	ARG	3.1
1	C	53	MET	3.0
1	A	154	ARG	3.0
3	G	15	DC	2.9
3	G	21	DG	2.7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
3	G	1	DC	2.7
1	B	111	LEU	2.7
1	A	50	GLY	2.7
3	G	9	DT	2.7
1	B	113	ARG	2.7
1	B	118	ILE	2.5
1	C	116	GLY	2.4
1	B	112	CYS	2.3
1	C	154	ARG	2.3
1	C	36	MET	2.3
1	B	110	MET	2.2
1	C	125	THR	2.2
2	F	5	DT	2.2
2	F	6	DT	2.2
3	G	20	DA	2.2
1	B	51	ARG	2.2
2	F	2	DT	2.1
3	G	16	DA	2.1
1	C	107	PRO	2.1
1	C	106	ARG	2.1
1	B	114	GLY	2.1
1	C	34	LYS	2.0
1	C	115	GLY	2.0
1	C	52	VAL	2.0
3	G	6	DT	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 [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(\AA^2)	Q<0.9
4	EDO	A	201	4/4	0.70	0.39	49,71,96,96	0
5	CL	B	202	1/1	0.86	0.15	59,59,59,59	0
4	EDO	B	201	4/4	0.87	0.20	25,46,55,55	0
4	EDO	A	202	4/4	0.90	0.38	25,43,55,55	0
5	CL	C	201	1/1	0.97	0.21	61,61,61,61	0

6.5 Other polymers [i](#)

There are no such residues in this entry.