

Full wwPDB X-ray Structure Validation Report (i)

Oct 10, 2023 – 12:13 AM EDT

PDB ID	:	7M4D
Title	:	DNA Polymerase Lambda, dCTP:At Ca2+ Ground State Ternary Complex
Authors	:	Jamsen, J.A.; Wilson, S.H.
Deposited on		
Resolution	:	1.82 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 (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 (1)) were used in the production of this report:

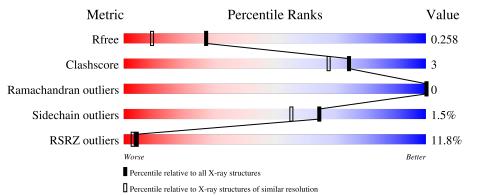
Xtriage (Phenix) EDS buster-report Percentile statistics Refmac CCP4 Ideal geometry (proteins) Ideal geometry (DNA, RNA)	: : : : :	20191225.v01 (using entries in the PDB archive December 25th 2019) 5.8.0158 7.0.044 (Gargrove) Engh & Huber (2001) Parkinson et al. (1996)
Ideal geometry (DNA, RNA) Validation Pipeline (wwPDB-VP)		Parkinson et al. (1996) 2.35.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure: $X\text{-}RAY \, DIFFRACTION$

The reported resolution of this entry is 1.82 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	$\begin{array}{c} \textbf{Whole archive} \\ (\#\textbf{Entries}) \end{array}$	${f Similar\ resolution}\ (\#{ m Entries,\ resolution\ range}({ m \AA}))$
R_{free}	130704	7484 (1.84-1.80)
Clashscore	141614	8401 (1.84-1.80)
Ramachandran outliers	138981	8290 (1.84-1.80)
Sidechain outliers	138945	8290 (1.84-1.80)
RSRZ outliers	127900	7371 (1.84-1.80)

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	А	329	90%		7% •
2	Т	11	73%	18%	9%
3	Р	6	83%		17%
4	D	4	100%		

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
8	EDO	А	606	-	-	-	Х



7M4D

2 Entry composition (i)

There are 9 unique types of molecules in this entry. The entry contains 3074 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 DNA polymerase lambda.

Mol	Chain	Residues		At	oms			ZeroOcc	AltConf	Trace
1	А	318	Total 2352	C 1483	N 425	0 432	S 12	8	6	0

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
А	463	LYS	SER	conflict	UNP Q9UGP5
А	464	GLY	GLN	conflict	UNP Q9UGP5
А	?	-	GLU	deletion	UNP Q9UGP5
А	?	-	ASN	deletion	UNP Q9UGP5
А	?	-	GLY	deletion	UNP Q9UGP5
А	?	-	GLN	deletion	UNP Q9UGP5
А	?	-	GLN	deletion	UNP Q9UGP5
А	471	THR	GLN	conflict	UNP Q9UGP5
А	543	ALA	CYS	engineered mutation	UNP Q9UGP5

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

Mol	Chain	Residues		Ate	oms			ZeroOcc	AltConf	Trace
2	Т	11	Total 264	C 126	N 51	O 75	Р 12	21	2	0

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

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
3	Р	6	Total	С	N	0	Р	0	0	0
	_		119	58	23	33	5	, i i i i i i i i i i i i i i i i i i i	Ū	, in the second s

• Molecule 4 is a DNA chain called DNA (5'-D(P*GP*CP*G)-3').



Mol	Chain	Residues		At	\mathbf{oms}			ZeroOcc	AltConf	Trace
4	П	4	Total	С	N	0	Р	0	0	0
4	D	4	83	38	16	25	4	0	0	0

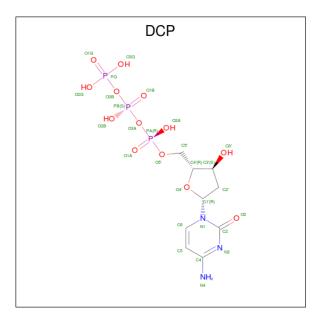
• Molecule 5 is CALCIUM ION (three-letter code: CA) (formula: Ca) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	А	1	Total Ca 1 1	0	0

• Molecule 6 is SODIUM ION (three-letter code: NA) (formula: Na) (labeled as "Ligand of Interest" by depositor).

Mo	l Chain	Residues	Atoms	ZeroOcc	AltConf
6	А	2	Total Na 2 2	0	0

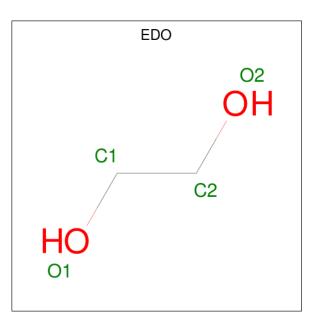
• Molecule 7 is 2'-DEOXYCYTIDINE-5'-TRIPHOSPHATE (three-letter code: DCP) (formula: C₉H₁₆N₃O₁₃P₃) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf		
7	А	1	Total 36	С 9	N 3	O 19	Р 5	0	1

• Molecule 8 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).





Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
8	А	1	$\begin{array}{ccc} \text{Total} & \text{C} & \text{O} \\ 4 & 2 & 2 \end{array}$	0	0
8	А	1	$\begin{array}{ccc} \text{Total} \text{C} \text{O} \\ 4 2 2 \end{array}$	0	0
8	А	1	$\begin{array}{ccc} \text{Total} & \text{C} & \text{O} \\ 4 & 2 & 2 \end{array}$	0	0
8	А	1	$\begin{array}{ccc} \text{Total} & \text{C} & \text{O} \\ 4 & 2 & 2 \end{array}$	0	0

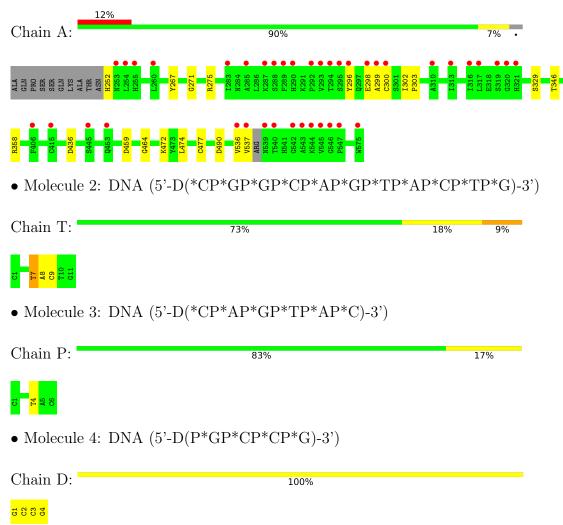
• Molecule 9 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
9	А	146	Total O 146 146	0	2
9	Т	25	TotalO2525	0	0
9	Р	22	TotalO2323	0	1
9	D	7	Total O 7 7	0	0



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 polymerase lambda



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants	55.75Å 59.92Å 141.30Å	Depositor
a, b, c, α , β , γ	90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.21 - 1.82	Depositor
Resolution (A)	39.21 - 1.82	EDS
% Data completeness	99.7 (39.21-1.82)	Depositor
(in resolution range)	99.7 (39.21-1.82)	EDS
R _{merge}	0.09	Depositor
R _{sym}	(Not available)	Depositor
$< I/\sigma(I) > 1$	$2.13 (at 1.82 \text{\AA})$	Xtriage
Refinement program	PHENIX v1.15.2-3472	Depositor
D D.	0.210 , 0.254	Depositor
R, R_{free}	0.213 , 0.258	DCC
R_{free} test set	2122 reflections (4.90%)	wwPDB-VP
Wilson B-factor $(Å^2)$	35.0	Xtriage
Anisotropy	0.326	Xtriage
Bulk solvent $k_{sol}(e/Å^3), B_{sol}(Å^2)$	0.31,63.0	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	0.95	EDS
Total number of atoms	3074	wwPDB-VP
Average B, all atoms $(Å^2)$	52.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 5.08% of the height of the origin peak. No significant pseudotranslation is detected.

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.



¹Intensities estimated from amplitudes.

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: CA, NA, DCP, 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	Bo	nd lengths	Bo	ond angles
IVIOI	Chain	RMSZ	# Z > 5	RMSZ	# Z > 5
1	А	0.57	1/2404~(0.0%)	0.68	1/3259~(0.0%)
2	Т	1.37	1/296~(0.3%)	1.19	0/455
3	Р	1.48	1/133~(0.8%)	1.35	1/203~(0.5%)
4	D	1.70	2/92~(2.2%)	1.21	1/138~(0.7%)
All	All	0.80	5/2925~(0.2%)	0.82	3/4055~(0.1%)

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	D	1	DG	OP3-P	-10.85	1.48	1.61
2	Т	7	DT	C5-C7	-6.78	1.46	1.50
1	А	477	CYS	CB-SG	-5.91	1.72	1.81
3	Р	4	DT	C5-C7	-5.63	1.46	1.50
4	D	2	DC	C3'-O3'	-5.62	1.36	1.44

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Ζ	$Observed(^{o})$	$Ideal(^{o})$
1	А	358	ARG	NE-CZ-NH1	-5.89	117.36	120.30
3	Р	4	DT	O4'-C4'-C3'	-5.42	102.33	104.50
4	D	2	DC	O4'-C1'-C2'	5.26	110.11	105.90

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



Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	А	2352	0	2149	15	0
2	Т	264	0	147	2	0
3	Р	119	0	69	0	0
4	D	83	0	45	1	0
5	А	1	0	0	0	0
6	А	2	0	0	0	0
7	А	36	0	0	0	0
8	А	16	0	24	0	0
9	А	146	0	0	1	0
9	D	7	0	0	0	0
9	Р	23	0	0	0	0
9	Т	25	0	0	0	0
All	All	3074	0	2434	17	0

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.

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 3.

All (17) 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:302:ILE:HG22	1:A:303:PRO:HD2	1.55	0.89
1:A:296:TYR:O	1:A:299:ALA:N	2.32	0.61
1:A:536:VAL:O	1:A:537:VAL:HG23	2.04	0.58
1:A:474:LEU:HD22	1:A:490:ASP:CG	2.23	0.58
1:A:298:GLU:O	1:A:302:ILE:CG1	2.53	0.57
1:A:267:TYR:CZ	1:A:275[A]:ARG:HD3	2.40	0.56
1:A:472:LYS:NZ	9:A:706:HOH:O	2.39	0.56
1:A:302:ILE:CG2	1:A:303:PRO:HD2	2.35	0.54
1:A:302:ILE:HG22	1:A:303:PRO:CD	2.33	0.53
1:A:271:GLY:CA	1:A:346[B]:THR:HG21	2.41	0.50
1:A:296:TYR:O	1:A:300:CYS:N	2.42	0.50
1:A:299:ALA:O	1:A:302:ILE:CG1	2.60	0.48
1:A:464:GLY:HA3	2:T:9[A]:DC:O5'	2.14	0.48
4:D:3:DC:H2'	4:D:4:DG:C8	2.51	0.46
1:A:474:LEU:HD22	1:A:490:ASP:OD1	2.16	0.45
2:T:7:DT:H2"	2:T:8[A]:DA:H8	1.82	0.45
1:A:436:ASP:OD1	1:A:436:ASP:C	2.57	0.43

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	А	320/329~(97%)	310 (97%)	10 (3%)	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	А	209/274~(76%)	204~(98%)	5(2%)	49 35

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	А	252	HIS
1	А	329[A]	SER
1	А	329[B]	SER
1	А	459[A]	ASP
1	А	459[B]	ASP

Sometimes 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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 9 ligands modelled in this entry, 3 are monoatomic - leaving 6 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	ype Chain Res		Link	Bond lengths			Bond angles		
	Type	pe Chain			Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
8	EDO	А	608	-	$3,\!3,\!3$	0.50	0	$2,\!2,\!2$	0.31	0
8	EDO	А	605	-	$3,\!3,\!3$	0.41	0	2,2,2	0.47	0
8	EDO	А	607	-	$3,\!3,\!3$	0.52	0	2,2,2	0.36	0
8	EDO	А	606	-	$3,\!3,\!3$	0.48	0	2,2,2	0.30	0
7	DCP	А	604[B]	5	$25,\!29,\!29$	0.57	0	$37,\!45,\!45$	0.82	1 (2%)
7	DCP	А	604[A]	5	25,29,29	0.57	0	37,45,45	0.78	1 (2%)

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
8	EDO	А	608	-	-	0/1/1/1	-
8	EDO	А	605	-	-	1/1/1/1	-
8	EDO	А	607	-	-	0/1/1/1	-
8	EDO	А	606	-	-	0/1/1/1	-
7	DCP	А	604[B]	5	-	0/22/34/34	0/2/2/2
7	DCP	А	604[A]	5	-	1/22/34/34	0/2/2/2

There are no bond length outliers.



Mol	Chain	Res	Type	Atoms	Z	$Observed(^{o})$	$Ideal(^{o})$
7	А	604[B]	DCP	PB-O3B-PG	-2.85	123.05	132.83
7	А	604[A]	DCP	PB-O3B-PG	-2.24	125.14	132.83

All (2) bond angle outliers are listed below:

There are no chirality outliers.

All (2) torsion outliers are listed below:

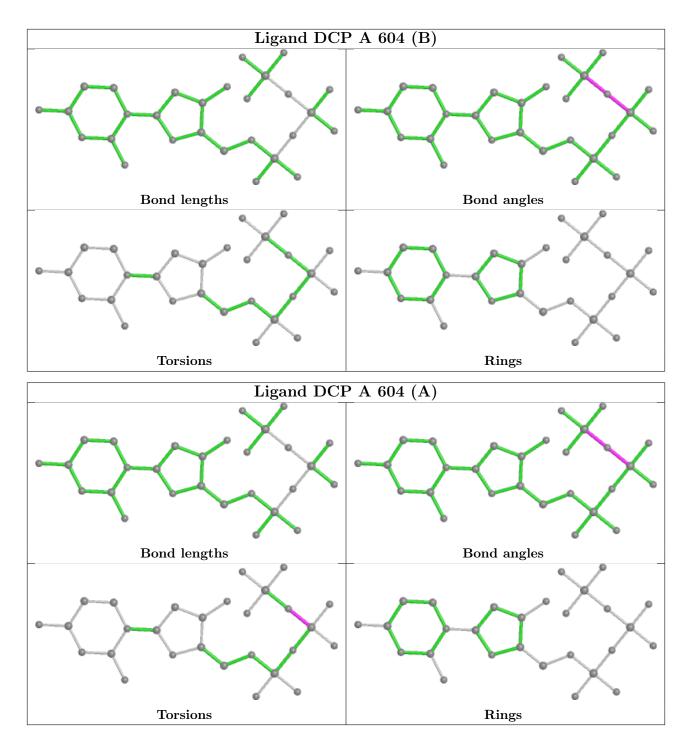
Mol	Chain	\mathbf{Res}	Type	Atoms
8	А	605	EDO	O1-C1-C2-O2
7	А	604[A]	DCP	PG-O3B-PB-O1B

There are no ring outliers.

No monomer is involved in short contacts.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and sufficient the outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.





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, 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	А	318/329~(96%)	0.71	40 (12%) 3 2	26, 45, 111, 235	0
2	Т	11/11 (100%)	-0.29	0 100 100	40, 44, 53, 54	0
3	Р	6/6~(100%)	0.15	0 100 100	29, 31, 35, 40	0
4	D	4/4 (100%)	-0.46	0 100 100	50, 51, 51, 53	0
All	All	339/350~(96%)	0.65	40 (11%) 4 3	26, 45, 108, 235	0

All (40) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	А	542	GLY	10.6
1	А	290	HIS	7.8
1	А	540	THR	7.5
1	А	317	LEU	6.2
1	А	543	ALA	6.2
1	А	287	LYS	6.2
1	А	537	VAL	6.1
1	А	293	VAL	5.9
1	А	320	GLY	5.7
1	А	547	PRO	4.9
1	А	316	ILE	4.7
1	А	254	LEU	4.7
1	А	300	CYS	4.7
1	А	253	ASN	4.6
1	А	285	ALA	4.5
1	А	539	ASN	4.4
1	А	546	GLY	4.2
1	А	295	SER	3.9
1	А	294	THR	3.8
1	А	319	SER	3.8
1	А	575	TRP	3.7

Continued on next page...



Mol	Chain	Res	Type	RSRZ
1	А	255	HIS	3.7
1	А	321	HIS	3.5
1	А	260	LEU	3.5
1	А	545	VAL	3.5
1	А	310	ALA	3.4
1	А	313	ILE	3.4
1	А	299	ALA	3.0
1	А	283	ILE	2.9
1	А	288	SER	2.8
1	А	292	PRO	2.7
1	А	406	PHE	2.7
1	А	544	LYS	2.6
1	А	445	SER	2.3
1	А	296	TYR	2.3
1	А	289	PHE	2.2
1	А	536	VAL	2.2
1	А	453	GLN	2.1
1	А	415	CYS	2.1
1	А	298	GLU	2.0

Continued from previous page...

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 monosaccharides 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, 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	B-factors(Å ²)	Q < 0.9
8	EDO	А	607	4/4	0.63	0.19	64,69,69,72	0
8	EDO	А	606	4/4	0.74	0.57	125,125,128,128	0
7	DCP	А	604[A]	28/28	0.85	0.24	47,68,89,94	14
7	DCP	А	604[B]	28/28	0.85	0.24	47,68,91,94	14

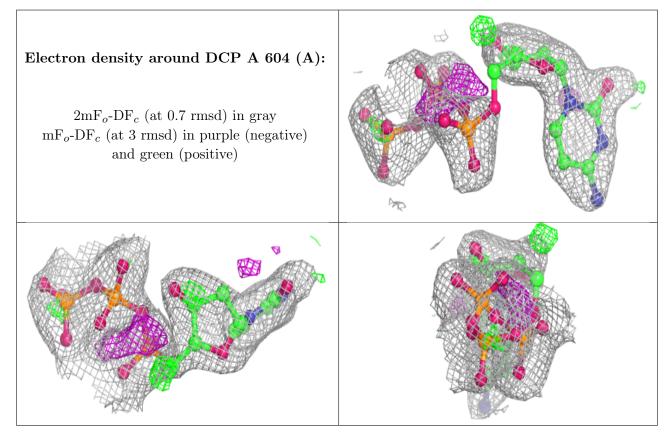
Continued on next page...



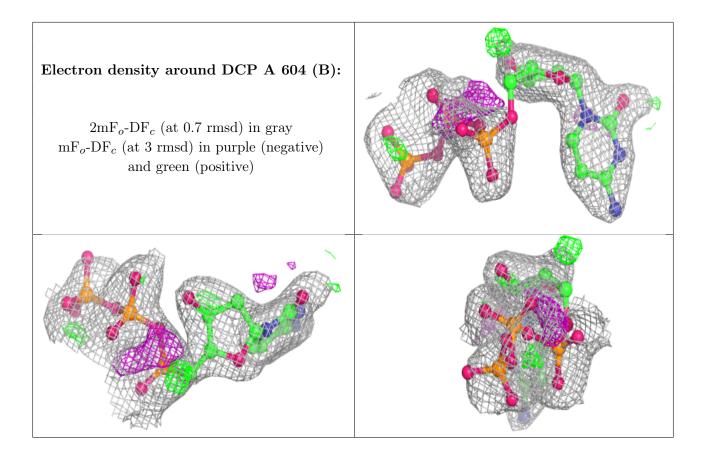
001000	continuou front protious page								
Mol	Type	Chain	\mathbf{Res}	Atoms	RSCC	RSR	$B-factors(Å^2)$	Q<0.9	
6	NA	А	602	1/1	0.86	0.17	59, 59, 59, 59, 59	0	
8	EDO	А	608	4/4	0.89	0.12	61,73,74,77	0	
8	EDO	А	605	4/4	0.92	0.17	49,54,55,58	0	
5	CA	А	601	1/1	0.98	0.08	63,63,63,63	0	
6	NA	А	603	1/1	0.99	0.07	25,25,25,25	0	

Continued from previous page..

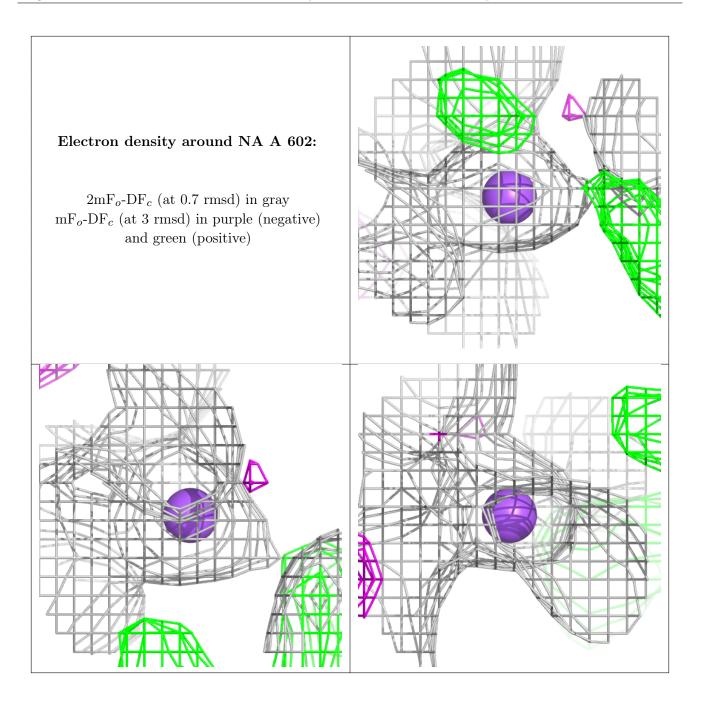
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.



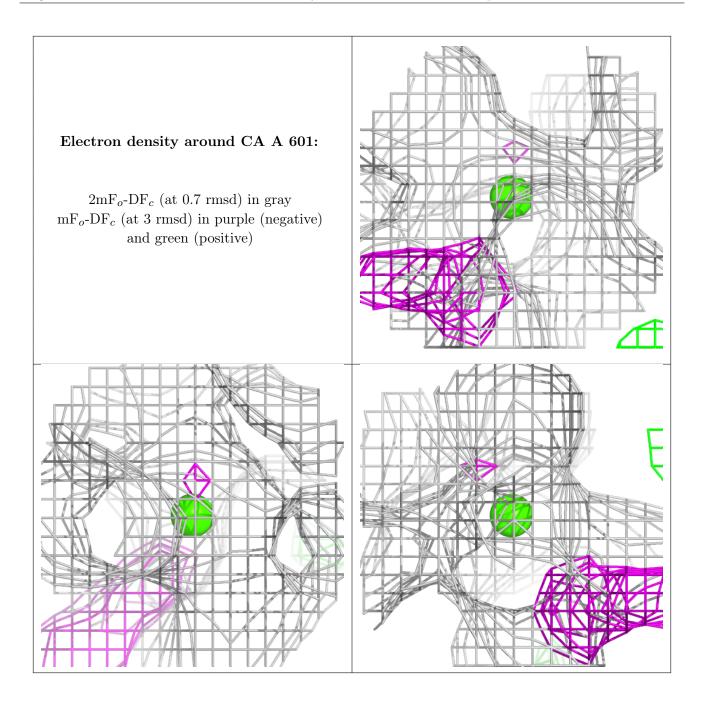




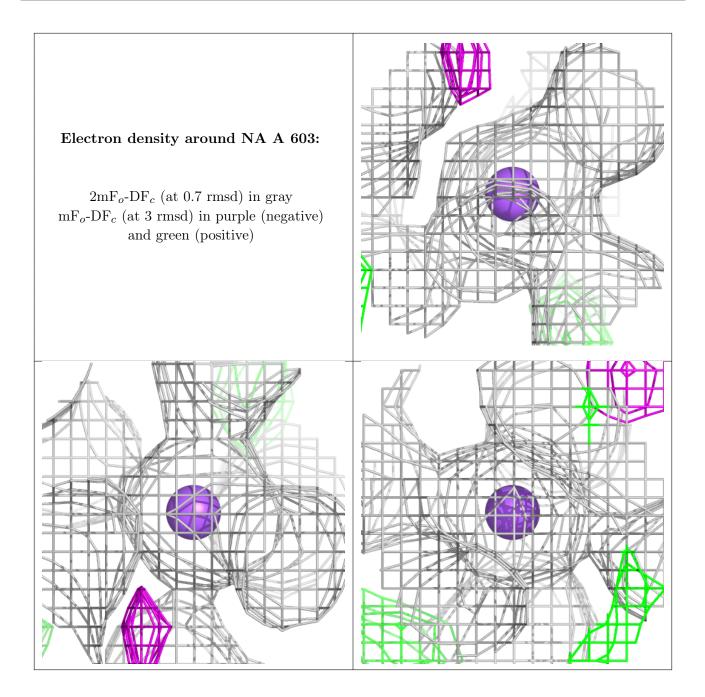












6.5 Other polymers (i)

There are no such residues in this entry.

