



wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 12, 2024 – 04:45 AM EDT

PDB ID : 1O3T
Title : PROTEIN-DNA RECOGNITION AND DNA DEFORMATION REVEALED
IN CRYSTAL STRUCTURES OF CAP-DNA COMPLEXES
Authors : Chen, S.; Vojtechovsky, J.; Parkinson, G.N.; Ebright, R.H.; Berman, H.M.
Deposited on : 2003-03-18
Resolution : 2.80 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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

MolProbity : 4.02b-467
Mogul : 2022.3.0, CSD as543be (2022)
Xtriage (Phenix) : 1.20.1
EDS : 2.36.2
buster-report : 1.1.7 (2018)
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) : 2.36.2

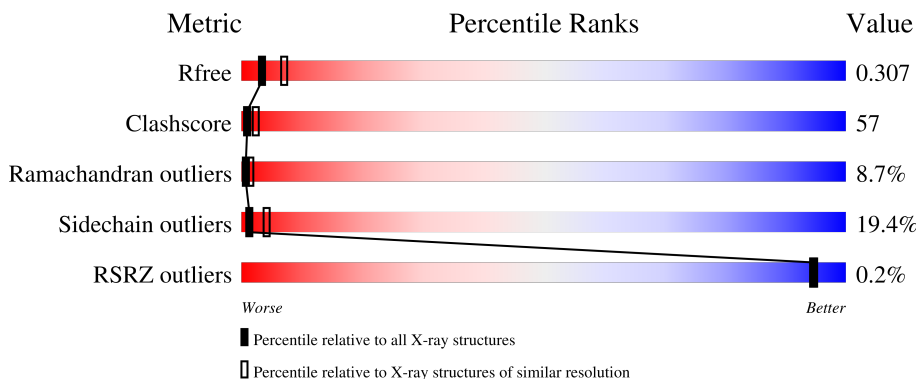
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.80 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	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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 $\leq 5\%$. 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	C	14	100%
1	F	14	14% 86%
2	D	17	18% 82%
2	E	17	18% 82%
3	A	200	30% 50% 19% .

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	B	200	

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
4	CMP	B	761	-	-	X	-

2 Entry composition i

There are 5 unique types of molecules in this entry. The entry contains 4491 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 DNA chain called 5'-D(*GP*CP*GP*AP*AP*AP*AP*TP*GP*CP*GP*AP*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	C	14	Total 289	C 138	N 60	O 78	P 13	0	0	0
1	F	14	Total 289	C 138	N 60	O 78	P 13	0	0	0

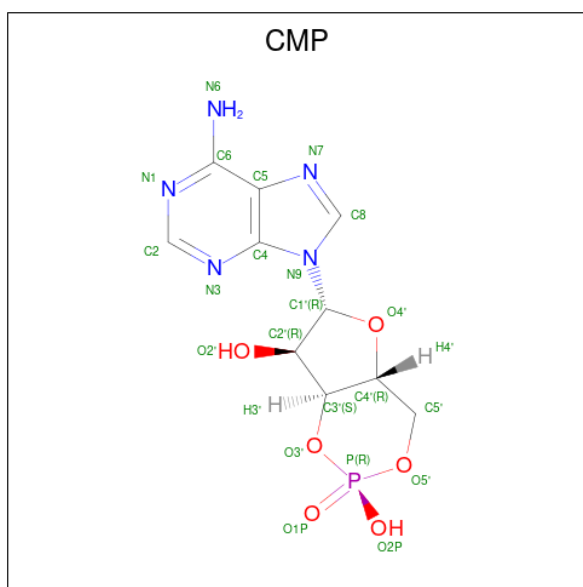
- Molecule 2 is a DNA chain called 5'-D(*CP*TP*AP*GP*AP*TP*CP*GP*CP*AP*TP*TP*TP*TP*TP*CP*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	D	17	Total 342	C 166	N 56	O 104	P 16	0	0	0
2	E	17	Total 342	C 166	N 56	O 104	P 16	0	0	0

- Molecule 3 is a protein called CATABOLITE GENE ACTIVATOR PROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	A	199	Total 1573	C 997	N 275	O 292	S 9	0	0	0
3	B	197	Total 1556	C 986	N 273	O 288	S 9	0	0	0

- Molecule 4 is ADENOSINE-3',5'-CYCLIC-MONOPHOSPHATE (three-letter code: CMP) (formula: C₁₀H₁₂N₅O₆P).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	A	1	Total	C	N	O	P	0	0
			22	10	5	6	1		
4	B	1	Total	C	N	O	P	0	0
			22	10	5	6	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	C	4	Total	O	0	0
			4	4		
5	D	7	Total	O	0	0
			7	7		
5	E	5	Total	O	0	0
			5	5		
5	F	2	Total	O	0	0
			2	2		
5	A	19	Total	O	0	0
			19	19		
5	B	19	Total	O	0	0
			19	19		

3 Residue-property plots

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: 5'-D(*GP*CP*GP*AP*AP*AP*AP*AP*TP*GP*CP*GP*AP*T)-3'

Chain C:  100%

G-5
C-4
G-3
A-2
A-1
A1
A2
A3
T4
G5
C6
G7
A8
T9

- Molecule 1: 5'-D(*GP*CP*GP*AP*AP*AP*AP*AP*TP*GP*CP*GP*AP*T)-3'

Chain F:  14% 86%


G27
C26
G25
A24
A23
A22
A21
T19
G18
C17
G16
A15
T14

- Molecule 2: 5'-D(*CP*TP*AP*GP*AP*TP*CP*GP*CP*AP*TP*TP*TP*TP*TP*CP*G)-3',

Chain D:  18% 82%


G13
T12
A11
G10
A9
T8
C7
G6
C5
A4
T3
T2
T1
T-1
G-4

- Molecule 2: 5'-D(*CP*TP*AP*GP*AP*TP*CP*GP*CP*AP*TP*TP*TP*TP*TP*CP*G)-3',

Chain E:  18% 82%

G10
T11
A12
G13
A14
T15
C16
G17
C18
A19
T20
T21
T22
C25
G26

- Molecule 3: CATABOLITE GENE ACTIVATOR PROTEIN

Chain A:  30% 50% 19%

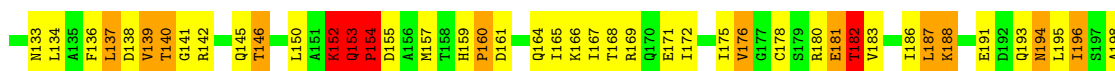
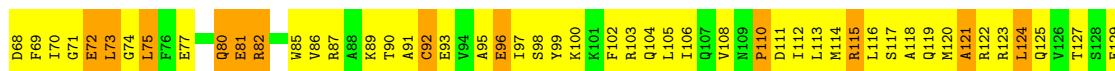
ASP
P9
T10
L11
E12
E77
E78
G79
Q80
E81
R82
S83
A84
W85
A88
E89
V94
V94
L97
S98
Y99
K100
K101
F102
R103
I106
M109
P110
D111
I112
L113
M114
R115
L116
S117
A118
Q119
M120
A121
R122
R123
L124
Q125
V126
T127
S128
M59
S62
Y63
L64
N133
N65
D68
F69
I70
G71
E72

L73
G74
L76
F76
E77
E78
G79
Q80
E81
R82
S83
A84
W85
A88
E89
V94
V94
L97
S98
Y99
K100
K101
F102
R103
I106
M109
P110
D111
I112
L113
M114
R115
L116
S117
A118
Q119
M120
A121
R122
R123
L124
Q125
V126
T127
S128
M59
S62
Y63
L64
N133
N65
D68
F69
I70
G71
E72



● Molecule 3: CATABOLITE GENE ACTIVATOR PROTEIN

Chain B: 24% 55% 18%



4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α , β , γ	136.99Å 152.80Å 76.06Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.80 28.91 – 2.49	Depositor EDS
% Data completeness (in resolution range)	72.2 (20.00-2.80) 63.5 (28.91-2.49)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	5.47 (at 2.48Å)	Xtrriage
Refinement program	CNS 0.5	Depositor
R, R_{free}	0.230 , 0.303 0.228 , 0.307	Depositor DCC
R_{free} test set	919 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å ²)	25.9	Xtrriage
Anisotropy	0.396	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.25 , 81.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.40$, $\langle L^2 \rangle = 0.22$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.86	EDS
Total number of atoms	4491	wwPDB-VP
Average B, all atoms (Å ²)	30.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.65% 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: CMP

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	C	0.49	0/326	0.75	0/502
1	F	0.54	0/326	0.77	0/502
2	D	0.52	0/381	0.84	0/586
2	E	0.56	0/381	0.84	0/586
3	A	0.55	1/1598 (0.1%)	0.83	1/2150 (0.0%)
3	B	0.56	0/1580	0.83	1/2127 (0.0%)
All	All	0.55	1/4592 (0.0%)	0.82	2/6453 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	178	CYS	CB-SG	-5.38	1.73	1.81

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	182	THR	N-CA-C	-5.54	96.03	111.00
3	A	93	GLU	N-CA-C	-5.27	96.76	111.00

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	C	289	0	158	34	0
1	F	289	0	158	24	0
2	D	342	0	196	31	0
2	E	342	0	196	27	0
3	A	1573	0	1612	167	0
3	B	1556	0	1600	216	0
4	A	22	0	11	3	0
4	B	22	0	11	7	0
5	A	19	0	0	2	0
5	B	19	0	0	4	0
5	C	4	0	0	1	0
5	D	7	0	0	1	0
5	E	5	0	0	0	0
5	F	2	0	0	0	0
All	All	4491	0	3942	478	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 57.

The worst 5 of 478 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)
4:A:762:CMP:C2	4:A:762:CMP:H2	0.97	1.49
4:B:761:CMP:C2	4:B:761:CMP:H2	0.97	1.48
1:F:26:DC:H2''	1:F:25:DG:H5''	1.24	1.17
3:B:196:ILE:HG22	3:B:204:VAL:HA	1.25	1.15
3:B:115:ARG:HB2	3:B:115:ARG:HH11	1.04	1.06

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	
3	A	197/200 (98%)	152 (77%)	33 (17%)	12 (6%)	1	4
3	B	195/200 (98%)	141 (72%)	32 (16%)	22 (11%)	0	1
All	All	392/400 (98%)	293 (75%)	65 (17%)	34 (9%)	1	1

5 of 34 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	98	SER
3	A	155	ASP
3	A	182	THR
3	A	206	TYR
3	B	26	LYS

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	
3	A	171/172 (99%)	133 (78%)	38 (22%)	1	2
3	B	170/172 (99%)	142 (84%)	28 (16%)	2	7
All	All	341/344 (99%)	275 (81%)	66 (19%)	1	4

5 of 66 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	B	154	PRO
3	B	160	PRO
3	B	202	THR
3	A	122	ARG
3	A	109	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
3	B	119	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	B	125	GLN
3	B	199	HIS
3	B	164	GLN
3	B	174	GLN

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](#)

2 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$
4	CMP	B	761	-	21,25,25	1.18	2 (9%)	24,39,39	1.67	8 (33%)
4	CMP	A	762	-	21,25,25	1.28	3 (14%)	24,39,39	2.14	5 (20%)

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	CMP	B	761	-	-	0/0/31/31	0/4/4/4
4	CMP	A	762	-	-	0/0/31/31	0/4/4/4

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	762	CMP	O3'-C3'	-3.16	1.39	1.44
4	A	762	CMP	P-O5'	2.59	1.60	1.57
4	B	761	CMP	P-O5'	2.46	1.60	1.57
4	B	761	CMP	O3'-C3'	-2.08	1.41	1.44
4	A	762	CMP	C8-N7	-2.04	1.31	1.34

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	762	CMP	O3'-C3'-C4'	-7.13	105.33	110.71
4	A	762	CMP	O2P-P-O1P	3.73	120.02	108.56
4	B	761	CMP	O2P-P-O1P	3.56	119.52	108.56
4	B	761	CMP	O3'-C3'-C2'	3.32	118.86	115.61
4	A	762	CMP	O3'-C3'-C2'	3.03	118.57	115.61

There are no chirality outliers.

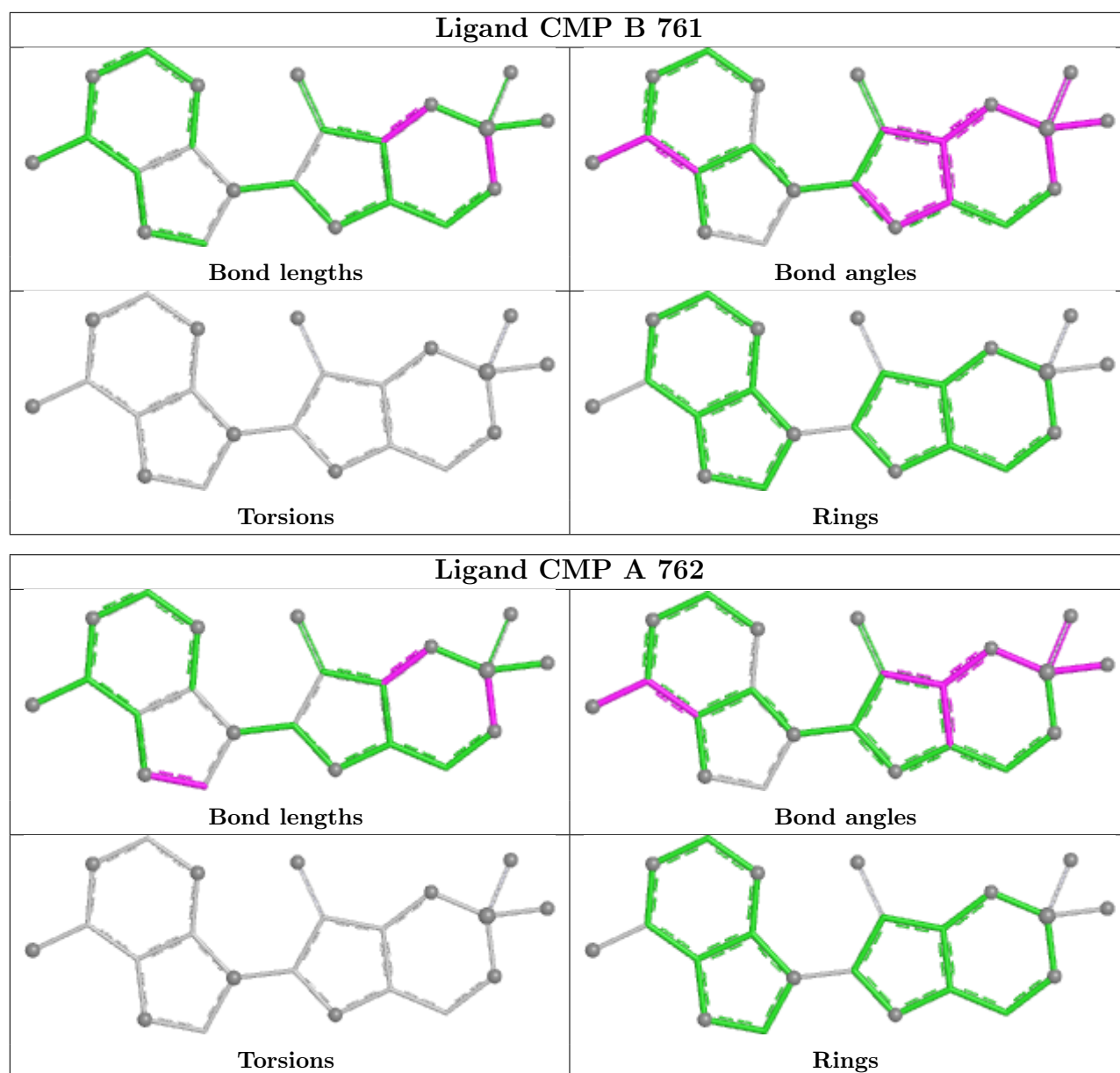
There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 10 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	B	761	CMP	7	0
4	A	762	CMP	3	0

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 than 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 any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. 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, 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	C	14/14 (100%)	-0.43	0 100 100	32, 50, 59, 63	0
1	F	14/14 (100%)	-0.63	0 100 100	26, 40, 48, 56	0
2	D	17/17 (100%)	-0.74	0 100 100	35, 44, 59, 63	0
2	E	17/17 (100%)	-0.74	0 100 100	25, 36, 51, 53	0
3	A	199/200 (99%)	-0.88	1 (0%) 91 88	4, 24, 45, 55	0
3	B	197/200 (98%)	-0.91	0 100 100	2, 23, 45, 55	0
All	All	458/462 (99%)	-0.86	1 (0%) 95 94	2, 26, 50, 63	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	207	GLY	2.7

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](#)

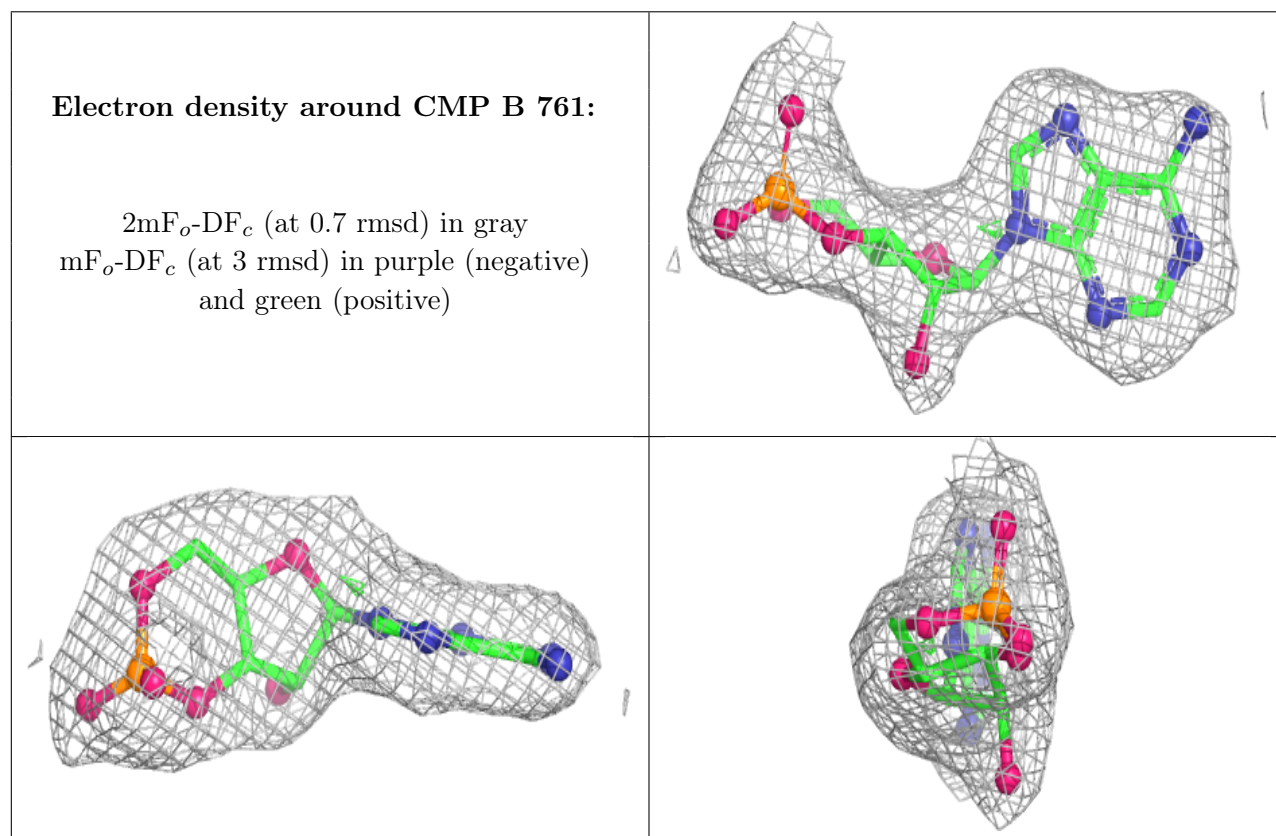
There are no monosaccharides in this entry.

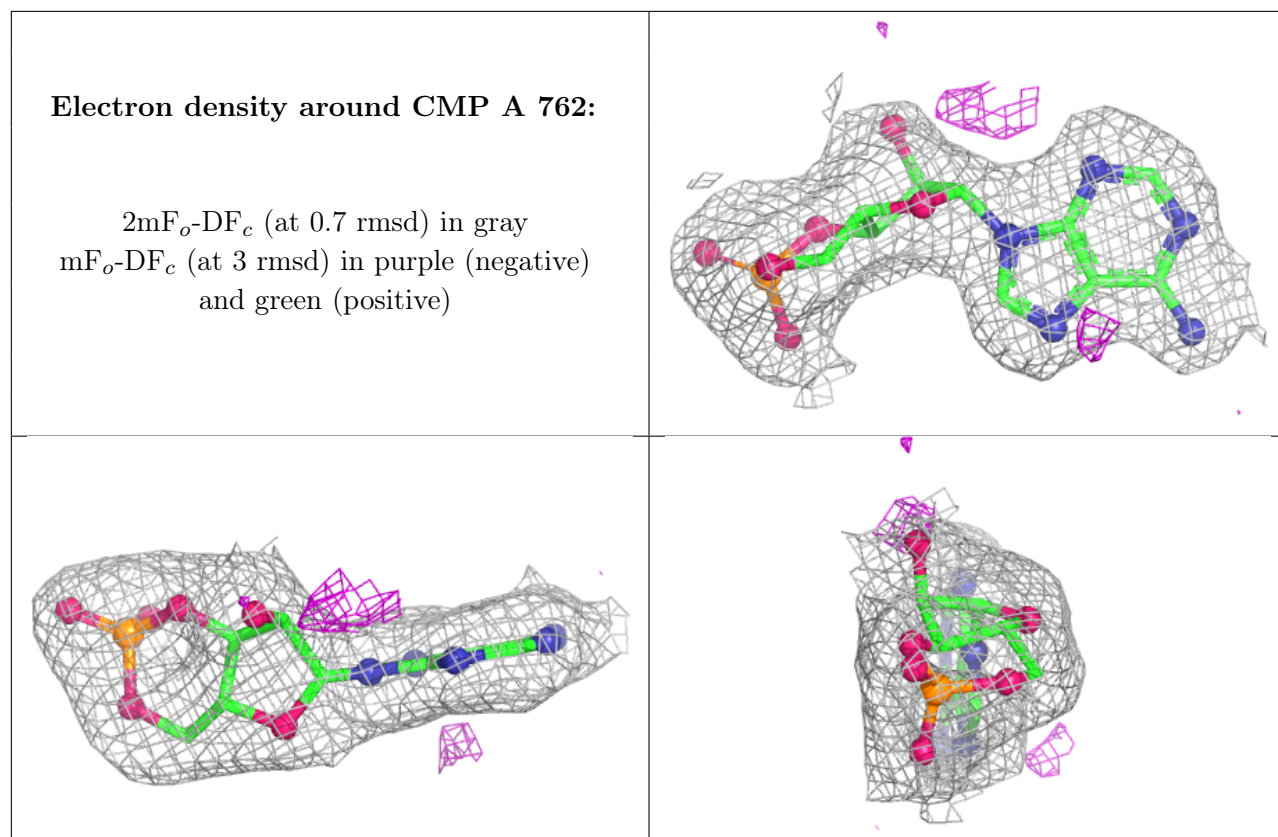
6.4 Ligands [i](#)

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, 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	CMP	B	761	22/22	0.97	0.13	9,14,20,21	0
4	CMP	A	762	22/22	0.98	0.13	0,7,12,15	0

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.