



# Full wwPDB X-ray Structure Validation Report ⓘ

Jun 24, 2024 – 03:34 PM EDT

PDB ID : 7AID  
Title : HIV-1 REVERSE TRANSCRIPTASE COMPLEX WITH DNA AND D-ASPARTATE TENOFOVIR  
Authors : Gu, W.; Martinez, S.E.; Nguyen, H.; Xu, H.; Herdewijn, P.; de Jonghe, S.; Das, K.  
Deposited on : 2020-09-26  
Resolution : 3.15 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](mailto: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 : 2.37.1  
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) : 2.37.1

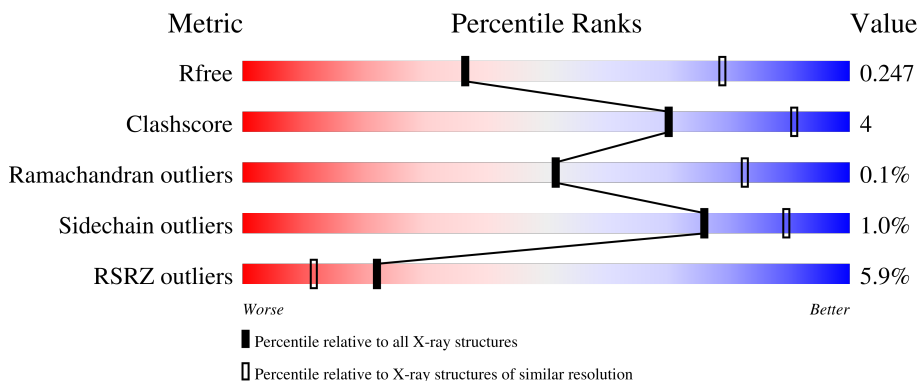
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.15 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.





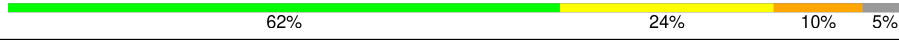

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	1665 (3.20-3.12)
Clashscore	141614	1804 (3.20-3.12)
Ramachandran outliers	138981	1770 (3.20-3.12)
Sidechain outliers	138945	1769 (3.20-3.12)
RSRZ outliers	127900	1616 (3.20-3.12)

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  $\geq 3$ , 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	556	 10% 90% 9%
1	C	556	 8% 89% 11%
2	B	428	 % 86% 9%
2	D	428	 3% 85% 11%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	E	27	
3	T	27	
4	F	21	
4	P	21	

## 2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 17683 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 Gag-Pol polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	555	Total	C	N	O	S	0	0	0
			4513	2922	751	832	8			
1	C	556	Total	C	N	O	S	0	0	0
			4521	2927	752	833	9			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MET	-	initiating methionine	UNP P03366
A	0	VAL	-	expression tag	UNP P03366
A	258	CYS	GLN	engineered mutation	UNP P03366
A	280	SER	CYS	engineered mutation	UNP P03366
A	498	ASN	ASP	engineered mutation	UNP P03366
C	-1	MET	-	initiating methionine	UNP P03366
C	0	VAL	-	expression tag	UNP P03366
C	258	CYS	GLN	engineered mutation	UNP P03366
C	280	SER	CYS	engineered mutation	UNP P03366
C	498	ASN	ASP	engineered mutation	UNP P03366

- Molecule 2 is a protein called Gag-Pol polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	411	Total	C	N	O	S	0	0	0
			3401	2215	563	616	7			
2	D	412	Total	C	N	O	S	0	0	0
			3400	2212	563	619	6			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	280	SER	CYS	engineered mutation	UNP P03366
D	280	SER	CYS	engineered mutation	UNP P03366

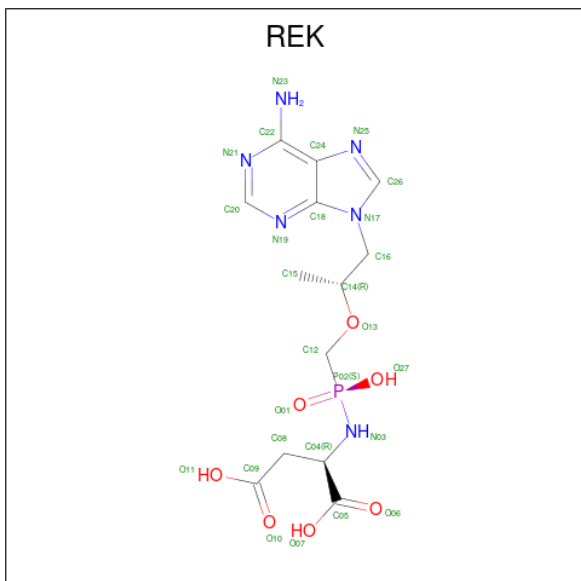
- Molecule 3 is a DNA chain called DNA (5'-D(P\*GP\*GP\*TP\*CP\*GP\*GP\*CP\*GP\*CP\*CP\*CP\*GP\*AP\*AP\*CP\*AP\*GP\*GP\*GP\*AP\*CP\*TP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	T	23	Total	C	N	O	P	0	0	0
			477	223	95	136	23			
3	E	23	Total	C	N	O	P	0	0	0
			477	223	95	136	23			

- Molecule 4 is a DNA chain called DNA (5'-D(\*CP\*AP\*GP\*TP\*CP\*CP\*CP\*TP\*GP\*TP\*TP\*CP\*GP\*GP\*(MRG)P\*CP\*GP\*CP\*CP\*(DDG))-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	P	20	Total	C	N	O	P	0	0	0
			403	192	72	120	19			
4	F	20	Total	C	N	O	P	0	0	0
			403	192	72	120	19			

- Molecule 5 is D-Aspartate Tenofovir (three-letter code: REK) (formula: C<sub>13</sub>H<sub>19</sub>N<sub>6</sub>O<sub>7</sub>P) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	C	1	Total	C	N	O	P	0	0
			27	13	6	7	1		

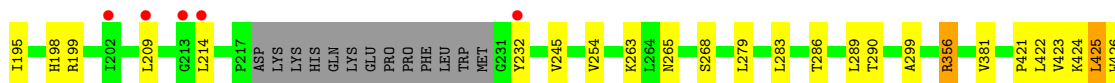
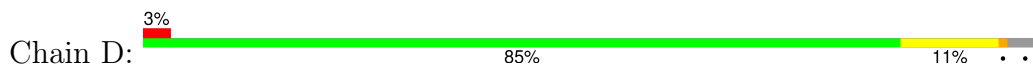
- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	14	Total O 14 14	0	0
6	B	17	Total O 17 17	0	0
6	C	14	Total O 14 14	0	0
6	D	13	Total O 13 13	0	0
6	E	3	Total O 3 3	0	0

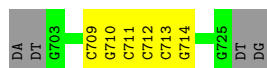




- Molecule 2: Gag-Pol polyprotein



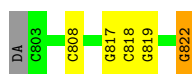
- Molecule 3: DNA (5'-D(P\*GP\*GP\*TP\*CP\*GP\*GP\*CP\*GP\*CP\*CP\*CP\*GP\*AP\*AP\*CP\*A P\*GP\*GP\*GP\*AP\*CP\*TP\*G)-3')



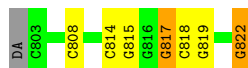
- Molecule 3: DNA (5'-D(P\*GP\*GP\*TP\*CP\*GP\*GP\*CP\*GP\*CP\*CP\*CP\*GP\*AP\*AP\*CP\*A P\*GP\*GP\*GP\*AP\*CP\*TP\*G)-3')



- Molecule 4: DNA (5'-D(\*CP\*AP\*GP\*TP\*CP\*CP\*CP\*TP\*GP\*TP\*TP\*CP\*GP\*GP\*(MRG) P\*CP\*GP\*CP\*CP\*(DDG))-3')



- Molecule 4: DNA (5'-D(\*CP\*AP\*GP\*TP\*CP\*CP\*CP\*TP\*GP\*TP\*TP\*CP\*GP\*GP\*(MRG) P\*CP\*GP\*CP\*CP\*(DDG))-3')





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	89.86Å 133.46Å 140.23Å 90.00° 98.35° 90.00°	Depositor
Resolution (Å)	53.37 – 3.15 53.37 – 3.15	Depositor EDS
% Data completeness (in resolution range)	98.9 (53.37-3.15) 98.9 (53.37-3.15)	Depositor EDS
$R_{merge}$	0.19	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.57 (at 3.13Å)	Xtrriage
Refinement program	PHENIX 1.18.2_3874	Depositor
R, $R_{free}$	0.214 , 0.250 0.212 , 0.247	Depositor DCC
$R_{free}$ test set	1630 reflections (2.90%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	87.7	Xtrriage
Anisotropy	0.262	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 61.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	17683	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	106.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.86% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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: REK, MRG, DDG

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	0.23	0/4631	0.40	0/6293
1	C	0.23	0/4639	0.40	0/6303
2	B	0.23	0/3499	0.40	0/4752
2	D	0.23	0/3497	0.41	0/4751
3	E	0.48	0/536	0.78	0/826
3	T	0.47	0/536	0.77	0/826
4	F	0.55	0/400	0.91	0/612
4	P	0.51	0/400	0.88	0/612
All	All	0.27	0/18138	0.47	0/24975

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	4513	0	4578	31	0
1	C	4521	0	4587	33	0
2	B	3401	0	3431	25	0
2	D	3400	0	3433	29	0
3	E	477	0	256	5	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	T	477	0	256	4	0
4	F	403	0	224	6	0
4	P	403	0	224	4	0
5	C	27	0	0	1	0
6	A	14	0	0	0	0
6	B	17	0	0	1	0
6	C	14	0	0	0	0
6	D	13	0	0	0	0
6	E	3	0	0	0	0
All	All	17683	0	16989	122	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 4.

All (122) 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:181:TYR:HB2	1:A:188:TYR:HB3	1.77	0.67
1:C:181:TYR:HB2	1:C:188:TYR:HB3	1.75	0.67
1:C:184:MET:HG3	4:F:822:DDG:H1'	1.77	0.67
1:C:78:ARG:HD2	3:E:705:DT:H5''	1.77	0.67
1:A:543:GLY:N	2:B:283:LEU:O	2.30	0.65
1:A:131:THR:HG22	1:A:143:ARG:HG2	1.78	0.64
1:A:206:ARG:NH2	1:A:216:THR:O	2.30	0.64
1:C:65:LYS:NZ	5:C:601:REK:O06	2.29	0.64
2:B:358:ARG:NH2	2:B:405:TYR:O	2.31	0.64
2:D:263:LYS:HE3	2:D:425:LEU:HB2	1.85	0.58
2:B:199:ARG:HA	2:B:202:ILE:HD12	1.85	0.57
1:A:21:VAL:HB	1:A:59:PRO:HD3	1.87	0.57
2:B:365:VAL:HG11	2:B:401:TRP:HB2	1.85	0.57
1:C:125:ARG:HE	1:C:147:ASN:HA	1.70	0.57
1:A:288:ALA:HB3	1:A:291:GLU:HB2	1.87	0.56
1:A:184:MET:HG3	4:P:822:DDG:H1'	1.88	0.56
1:C:503:LEU:HD22	1:C:535:TRP:HB2	1.88	0.55
1:A:503:LEU:HD22	1:A:535:TRP:HB2	1.89	0.55
1:C:543:GLY:N	2:D:283:LEU:O	2.39	0.54
1:A:84:THR:HB	1:A:154:LYS:HE2	1.88	0.54
2:D:120:LEU:HD23	2:D:125:ARG:HG2	1.90	0.54
2:B:21:VAL:HB	2:B:59:PRO:HD3	1.90	0.53
1:C:84:THR:HB	1:C:154:LYS:HE2	1.91	0.53
2:B:390:LYS:NZ	2:B:415:GLU:OE2	2.37	0.53

Continued on next page...

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:D:185:ASP:N	2:D:185:ASP:OD1	2.42	0.53
1:C:63:ILE:HD11	1:C:72:ARG:HH12	1.75	0.52
1:A:442:VAL:HB	1:A:481:ALA:HB1	1.92	0.52
2:D:73:LYS:NZ	2:D:146:TYR:OH	2.40	0.50
1:A:185:ASP:N	1:A:185:ASP:OD1	2.44	0.50
3:E:711:DC:H2'	3:E:712:DC:C6	2.47	0.49
1:A:125:ARG:HD3	1:A:147:ASN:HA	1.95	0.48
2:D:191:SER:OG	2:D:198:HIS:ND1	2.36	0.48
1:C:34:LEU:HD21	1:C:62:ALA:HB2	1.95	0.48
2:D:28:GLU:HG2	2:D:32:LYS:HE2	1.95	0.48
4:F:818:DC:H2'	4:F:819:DG:C8	2.48	0.48
2:D:91:GLN:HG2	2:D:92:LEU:HG	1.96	0.48
3:E:718:DA:H1'	3:E:719:DG:C8	2.48	0.48
3:T:711:DC:H2'	3:T:712:DC:C6	2.49	0.48
2:B:120:LEU:HD23	2:B:125:ARG:HG2	1.96	0.48
1:C:41:MET:HB3	1:C:46:LYS:HB2	1.95	0.48
1:C:113:ASP:HB3	1:C:116:PHE:HD2	1.78	0.48
2:D:79:GLU:HG3	2:D:83:ARG:HE	1.78	0.48
1:A:458:VAL:HG12	2:B:286:THR:HG21	1.96	0.48
1:A:451:LYS:NZ	4:P:808:DC:OP1	2.48	0.47
3:T:709:DC:H2'	3:T:710:DG:C8	2.48	0.47
1:C:56:TYR:O	1:C:143:ARG:NH2	2.47	0.47
2:D:209:LEU:HB3	2:D:214:LEU:HB3	1.97	0.47
4:P:818:DC:H2'	4:P:819:DG:C8	2.50	0.47
4:F:818:DC:H2'	4:F:819:DG:H8	1.79	0.47
1:C:492:GLU:HG2	1:C:530:LYS:HB2	1.96	0.46
2:B:260:LEU:HD21	2:B:303:LEU:HD13	1.97	0.46
1:C:218:ASP:HA	1:C:221:HIS:HB2	1.96	0.46
1:C:494:ASN:HB3	2:D:289:LEU:HD12	1.98	0.46
2:B:282:LEU:HD21	2:B:295:LEU:HA	1.96	0.46
1:A:443:ASP:OD1	1:A:444:GLY:N	2.47	0.46
1:A:542:ILE:HG23	2:B:283:LEU:HD13	1.98	0.46
2:D:195:ILE:HD11	2:D:199:ARG:HE	1.81	0.46
2:D:265:ASN:O	2:D:268:SER:OG	2.27	0.46
1:C:458:VAL:HG12	2:D:286:THR:HG21	1.98	0.46
2:D:100:LEU:HG	2:D:381:VAL:HG13	1.98	0.45
4:P:818:DC:H2'	4:P:819:DG:H8	1.80	0.45
1:C:30:LYS:HG2	1:C:71:TRP:HZ3	1.81	0.45
1:A:494:ASN:HB3	2:B:289:LEU:HD12	1.99	0.45
2:B:60:VAL:HG22	2:B:75:VAL:HG22	1.98	0.45
1:C:317:VAL:HG11	1:C:347:LYS:HB3	1.98	0.45

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:17:ASP:O	2:B:83:ARG:HD3	2.17	0.45
1:C:57:ASN:HD21	1:C:131:THR:HG22	1.81	0.45
2:D:17:ASP:O	2:D:83:ARG:HD3	2.15	0.45
3:T:713:DC:H2'	3:T:714:DG:C8	2.51	0.45
1:C:288:ALA:HB3	1:C:291:GLU:HB2	1.98	0.45
1:A:363:ASN:HA	1:A:511:ASP:OD1	2.17	0.44
2:D:421:PRO:HB2	2:D:423:VAL:HG22	1.99	0.44
1:A:41:MET:HB3	1:A:46:LYS:HB2	1.99	0.44
1:A:56:TYR:O	1:A:143:ARG:NH2	2.50	0.44
2:D:115:TYR:CD1	2:D:156:SER:HB3	2.52	0.44
2:D:114:ALA:HB2	2:D:214:LEU:HG	1.99	0.44
2:B:320:ASP:O	2:B:343:GLN:NE2	2.49	0.44
1:C:451:LYS:NZ	4:F:808:DC:OP1	2.50	0.44
1:C:376:THR:HG23	1:C:386:THR:HG22	2.00	0.43
1:C:114:ALA:HB1	1:C:160:PHE:CZ	2.53	0.43
2:B:154:LYS:HG2	2:B:184:MET:SD	2.58	0.43
2:D:245:VAL:O	2:D:263:LYS:NZ	2.42	0.43
1:A:376:THR:HG21	2:B:401:TRP:CH2	2.54	0.43
1:A:469:LEU:HD12	1:A:477:THR:HG22	1.99	0.43
2:B:359:GLY:HA3	2:B:366:LYS:HD2	2.00	0.43
2:D:356:ARG:H	2:D:356:ARG:HG3	1.63	0.43
3:T:709:DC:H2'	3:T:710:DG:H8	1.84	0.43
1:A:115:TYR:HD2	1:A:151:GLN:HG2	1.84	0.43
2:B:255:ASN:HB2	2:B:289:LEU:HB3	2.01	0.43
1:C:457:TYR:HE1	1:C:463:ARG:HG2	1.83	0.43
1:A:261:VAL:HG13	1:A:276:VAL:HG11	2.00	0.42
1:A:287:LYS:HD3	1:A:287:LYS:H	1.85	0.42
2:B:358:ARG:NH2	6:B:503:HOH:O	2.51	0.42
4:F:814:DC:H2''	4:F:815:DG:H8	1.85	0.42
1:A:24:TRP:HD1	1:A:59:PRO:HB3	1.83	0.42
1:A:541:GLY:HA2	1:A:546:GLU:HB2	2.01	0.42
2:B:191:SER:OG	2:B:198:HIS:ND1	2.47	0.42
2:D:423:VAL:HB	2:D:426:TRP:CD1	2.54	0.42
2:B:64:LYS:HE3	2:B:71:TRP:CE2	2.55	0.41
1:C:317:VAL:HG23	1:C:349:LEU:HD23	2.02	0.41
1:C:405:TYR:CE2	1:C:407:GLN:HB2	2.55	0.41
1:A:118:VAL:HG21	1:A:160:PHE:HD1	1.85	0.41
2:B:79:GLU:HG3	2:B:83:ARG:HE	1.85	0.41
1:C:80:LEU:O	1:C:84:THR:OG1	2.28	0.41
2:D:21:VAL:HB	2:D:59:PRO:HD3	2.02	0.41
3:E:706:DC:H2'	3:E:707:DG:C8	2.55	0.41

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:58:THR:HG21	1:A:77:PHE:CE1	2.56	0.41
1:C:500:GLN:CD	2:D:422:LEU:HD21	2.40	0.41
1:C:206:ARG:NH2	1:C:216:THR:O	2.48	0.41
1:C:198:HIS:O	1:C:202:ILE:HG12	2.21	0.41
2:D:154:LYS:HG2	2:D:184:MET:SD	2.61	0.41
2:D:254:VAL:HG13	2:D:283:LEU:HD22	2.03	0.41
2:D:279:LEU:HD23	2:D:299:ALA:HB1	2.02	0.41
1:C:435:VAL:HG22	2:D:290:THR:HG21	2.03	0.41
2:D:149:LEU:HA	2:D:150:PRO:HD3	1.95	0.41
3:E:709:DC:H2'	3:E:710:DG:C8	2.56	0.41
4:F:817:MRG:H2'	4:F:818:DC:C6	2.56	0.41
2:B:266:TRP:CE2	2:B:422:LEU:HD13	2.56	0.40
1:A:246:LEU:HD22	1:A:260:LEU:HD12	2.03	0.40
2:B:38:CYS:SG	2:B:132:ILE:HD11	2.60	0.40
1:A:12:LEU:HD22	1:A:84:THR:HA	2.02	0.40
1:C:281:LYS:HG2	1:C:284:ARG:NH2	2.37	0.40

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	553/556 (100%)	534 (97%)	19 (3%)	0	100	100
1	C	554/556 (100%)	536 (97%)	17 (3%)	1 (0%)	47	78
2	B	407/428 (95%)	388 (95%)	19 (5%)	0	100	100
2	D	408/428 (95%)	389 (95%)	19 (5%)	0	100	100
All	All	1922/1968 (98%)	1847 (96%)	74 (4%)	1 (0%)	51	83

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	332	GLN

### 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	496/497 (100%)	489 (99%)	7 (1%)	67 85
1	C	497/497 (100%)	494 (99%)	3 (1%)	86 94
2	B	373/390 (96%)	370 (99%)	3 (1%)	81 92
2	D	374/390 (96%)	369 (99%)	5 (1%)	69 86
All	All	1740/1774 (98%)	1722 (99%)	18 (1%)	76 89

All (18) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	12	LEU
1	A	16	MET
1	A	17	ASP
1	A	185	ASP
1	A	287	LYS
1	A	324	ASP
1	A	388	LYS
2	B	92	LEU
2	B	414	TRP
2	B	422	LEU
1	C	276	VAL
1	C	301	LEU
1	C	498	ASN
2	D	185	ASP
2	D	232	TYR
2	D	356	ARG
2	D	424	LYS
2	D	425	LEU

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](#)

4 non-standard protein/DNA/RNA residues 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	DDG	P	822	3,4	17,23,24	1.25	2 (11%)	16,33,36	1.55	3 (18%)
4	DDG	F	822	3,4	17,23,24	1.25	2 (11%)	16,33,36	1.66	3 (18%)
4	MRG	P	817	3,4	18,24,29	1.11	3 (16%)	19,35,42	0.70	1 (5%)
4	MRG	F	817	3,4	18,24,29	1.12	3 (16%)	19,35,42	0.71	1 (5%)

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	DDG	P	822	3,4	-	0/3/18/19	0/3/3/3
4	DDG	F	822	3,4	-	0/3/18/19	0/3/3/3
4	MRG	P	817	3,4	-	0/3/21/27	0/3/3/3
4	MRG	F	817	3,4	-	0/3/21/27	0/3/3/3

All (10) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	P	822	DDG	C5-C6	-2.78	1.41	1.47
4	F	822	DDG	C5-C6	-2.78	1.41	1.47
4	F	817	MRG	C5-C6	-2.75	1.42	1.47
4	P	817	MRG	C5-C6	-2.75	1.42	1.47
4	P	822	DDG	C8-N7	-2.73	1.30	1.34
4	F	822	DDG	C8-N7	-2.72	1.30	1.34

*Continued on next page...*



Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	F	817	MRG	C8-N7	-2.51	1.30	1.34
4	P	817	MRG	C8-N7	-2.50	1.30	1.34
4	F	817	MRG	C5-C4	-2.03	1.38	1.43
4	P	817	MRG	C5-C4	-2.02	1.38	1.43

All (8) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	P	822	DDG	C8-N7-C5	3.93	109.24	102.55
4	F	822	DDG	C8-N7-C5	3.92	109.22	102.55
4	F	822	DDG	C2-N1-C6	-2.68	120.21	125.11
4	P	822	DDG	C2-N1-C6	-2.67	120.23	125.11
4	P	822	DDG	C5-C6-N1	2.28	118.42	114.07
4	F	822	DDG	C5-C6-N1	2.28	118.41	114.07
4	F	817	MRG	O6-C6-C5	2.01	128.31	124.32
4	P	817	MRG	O6-C6-C5	2.01	128.30	124.32

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	P	822	DDG	1	0
4	F	822	DDG	1	0
4	F	817	MRG	1	0

## 5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

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	REK	C	601	-	22,28,28	1.59	3 (13%)	18,40,40	1.28	2 (11%)

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	REK	C	601	-	-	9/18/23/23	0/2/2/2

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	C	601	REK	C16-N17	2.87	1.51	1.48
5	C	601	REK	C20-N21	2.80	1.38	1.33
5	C	601	REK	P02-O27	-2.53	1.50	1.56

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	C	601	REK	C24-C22-N23	2.31	123.83	120.31
5	C	601	REK	O27-P02-O01	2.04	117.64	112.36

There are no chirality outliers.

All (9) torsion outliers are listed below:

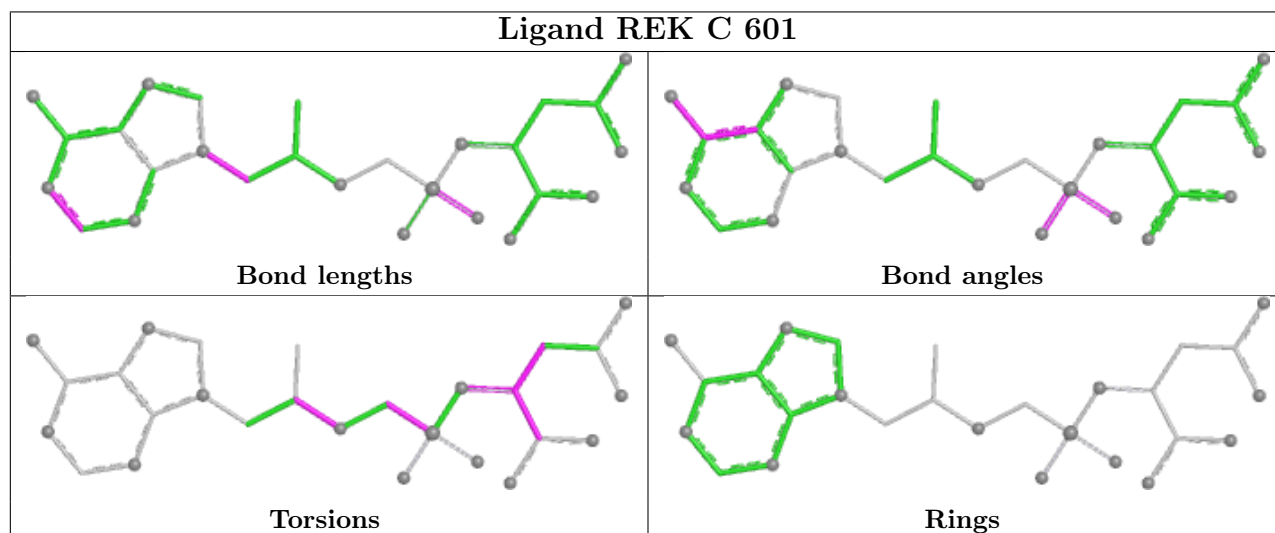
Mol	Chain	Res	Type	Atoms
5	C	601	REK	O13-C12-P02-O01
5	C	601	REK	O13-C12-P02-O27
5	C	601	REK	C15-C14-O13-C12
5	C	601	REK	N03-C04-C05-O06
5	C	601	REK	N03-C04-C05-O07
5	C	601	REK	C08-C04-N03-P02
5	C	601	REK	C16-C14-O13-C12
5	C	601	REK	C05-C04-N03-P02
5	C	601	REK	C05-C04-C08-C09

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	C	601	REK	1	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

### 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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	555/556 (99%)	0.40	58 (10%) <b>6</b> <b>3</b>	38, 108, 211, 237	0
1	C	556/556 (100%)	0.36	42 (7%) <b>13</b> <b>7</b>	38, 111, 204, 221	0
2	B	411/428 (96%)	0.01	4 (0%) <b>82</b> <b>73</b>	36, 79, 132, 156	0
2	D	412/428 (96%)	0.21	13 (3%) <b>47</b> <b>30</b>	45, 96, 152, 168	0
3	E	23/27 (85%)	0.13	2 (8%) <b>10</b> <b>5</b>	87, 137, 201, 237	0
3	T	23/27 (85%)	-0.27	0 <b>100</b> <b>100</b>	103, 157, 215, 241	0
4	F	18/21 (85%)	-0.18	0 <b>100</b> <b>100</b>	83, 124, 201, 202	0
4	P	18/21 (85%)	-0.50	0 <b>100</b> <b>100</b>	99, 133, 193, 197	0
All	All	2016/2064 (97%)	0.25	119 (5%) <b>22</b> <b>12</b>	36, 100, 188, 241	0

All (119) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	67	ASP	6.8
2	D	214	LEU	6.7
1	C	26	LEU	6.1
1	C	133	PRO	6.1
1	A	73	LYS	6.1
1	C	30	LYS	5.9
1	A	132	ILE	5.6
1	A	69	THR	5.4
1	C	63	ILE	5.4
1	A	66	LYS	5.4
1	C	142	ILE	5.3
1	A	135	ILE	5.3
1	A	134	SER	5.2
1	C	74	LEU	5.2
2	D	4	PRO	5.2
1	A	26	LEU	5.1

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Res</b>	<b>Type</b>	<b>RSRZ</b>
1	A	133	PRO	5.1
2	B	214	LEU	5.0
1	C	136	ASN	4.7
1	C	139	THR	4.6
1	A	30	LYS	4.5
1	C	67	ASP	4.4
1	A	70	LYS	4.4
1	C	293	ILE	4.4
1	C	73	LYS	4.3
1	C	61	PHE	4.3
1	A	249	LYS	4.2
1	C	138	GLU	4.2
1	A	63	ILE	4.2
1	C	34	LEU	4.1
1	A	252	TRP	4.1
1	A	62	ALA	3.9
1	A	142	ILE	3.9
1	A	554	ALA	3.8
1	C	32	LYS	3.8
1	C	294	PRO	3.8
1	A	72	ARG	3.7
1	A	286	THR	3.6
1	A	71	TRP	3.6
1	C	135	ILE	3.6
1	C	140	PRO	3.5
1	C	132	ILE	3.4
1	C	295	LEU	3.4
3	E	703	DG	3.4
1	A	29	GLU	3.3
1	C	252	TRP	3.3
1	A	138	GLU	3.3
1	A	140	PRO	3.3
1	A	64	LYS	3.3
2	D	174	GLN	3.3
1	A	75	VAL	3.2
1	C	33	ALA	3.1
1	A	61	PHE	3.1
1	A	137	ASN	3.1
1	A	287	LYS	3.1
1	C	134	SER	3.0
1	A	68	SER	3.0
1	C	49	LYS	3.0

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Res</b>	<b>Type</b>	<b>RSRZ</b>
1	C	144	TYR	3.0
1	A	284	ARG	3.0
1	A	28	GLU	3.0
1	A	34	LEU	3.0
1	A	74	LEU	3.0
1	C	27	THR	2.9
1	A	289	LEU	2.9
1	C	544	GLY	2.9
2	D	85	GLN	2.9
1	C	137	ASN	2.7
2	D	209	LEU	2.7
1	A	295	LEU	2.7
1	C	24	TRP	2.7
1	A	27	THR	2.7
1	A	193	LEU	2.7
1	C	545	ASN	2.6
1	A	544	GLY	2.6
2	D	213	GLY	2.6
1	C	17	ASP	2.6
1	A	296	THR	2.6
1	A	139	THR	2.6
2	B	213	GLY	2.6
2	D	232	TYR	2.5
1	C	247	PRO	2.5
1	A	545	ASN	2.5
1	C	-1	MET	2.5
2	D	168	LEU	2.5
1	A	141	GLY	2.4
2	D	66	LYS	2.4
1	A	136	ASN	2.4
1	A	254	VAL	2.4
1	A	41	MET	2.4
1	A	247	PRO	2.4
1	C	37	ILE	2.4
1	C	62	ALA	2.4
2	B	295	LEU	2.4
1	C	248	GLU	2.4
1	A	251	SER	2.3
1	A	292	VAL	2.3
1	A	282	LEU	2.3
1	A	549	ASP	2.3
3	E	704	DG	2.3

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	C	25	PRO	2.2
1	C	299	ALA	2.2
1	A	32	LYS	2.2
2	D	202	ILE	2.2
1	A	546	GLU	2.2
1	A	294	PRO	2.1
1	A	65	LYS	2.1
1	C	52	PRO	2.1
1	C	106	VAL	2.1
1	A	303	LEU	2.1
1	C	296	THR	2.1
2	D	5	ILE	2.1
1	A	131	THR	2.1
2	D	38	CYS	2.0
1	C	286	THR	2.0
2	B	4	PRO	2.0
1	A	285	GLY	2.0
1	A	550	LYS	2.0
2	D	178	ILE	2.0

## 6.2 Non-standard residues in protein, DNA, RNA chains [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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
4	MRG	P	817	22/27	0.86	0.11	139,156,164,165	0
4	MRG	F	817	22/27	0.87	0.15	119,136,148,165	0
4	DDG	P	822	21/22	0.91	0.17	89,107,126,130	0
4	DDG	F	822	21/22	0.92	0.21	94,113,122,123	0

## 6.3 Carbohydrates [i](#)

There are no monosaccharides in this entry.

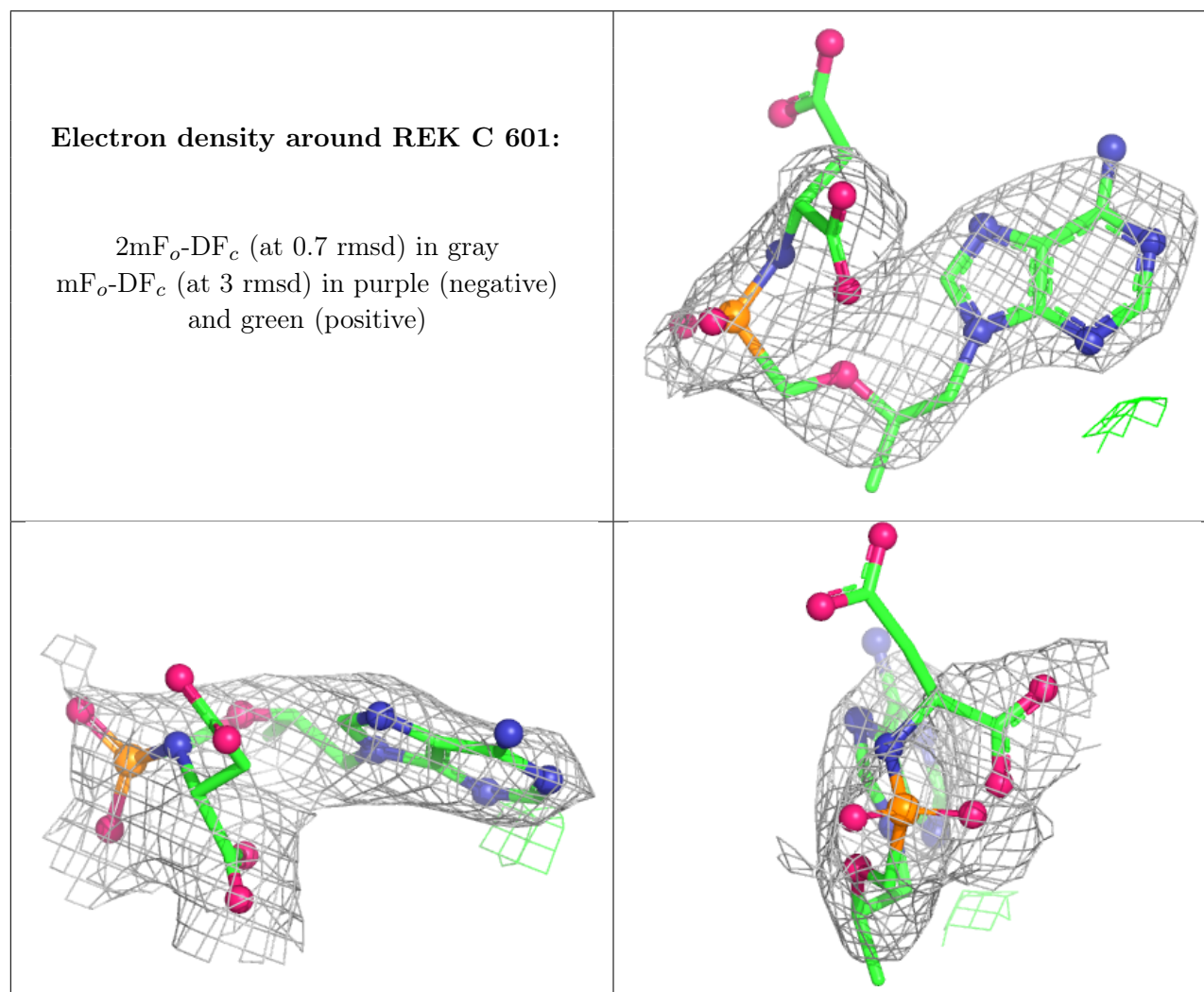
## 6.4 Ligands [i](#)

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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
5	REK	C	601	27/27	0.80	0.27	140,145,164,174	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.