WORLDWIDE PDB
PROTEIN DATA BANK

Full wwPDB X-ray Structure Validation Report (i)

Sep 12, 2024 – 12:08 PM JST

PDB ID	:	9INK
Title	:	Crystal structure of beta-carotene-binding protein (BBP) from Schistocerca gregaria complexed with beta-carotene
Authors	:	Boyko, K.M.; Varfolomeeva, L.A.; Egorkin, N.A.; Popov, V.O.; Sluchanko, N.N.
Deposited on	:	2024-07-08
Resolution	:	2.70 Å(reported)

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

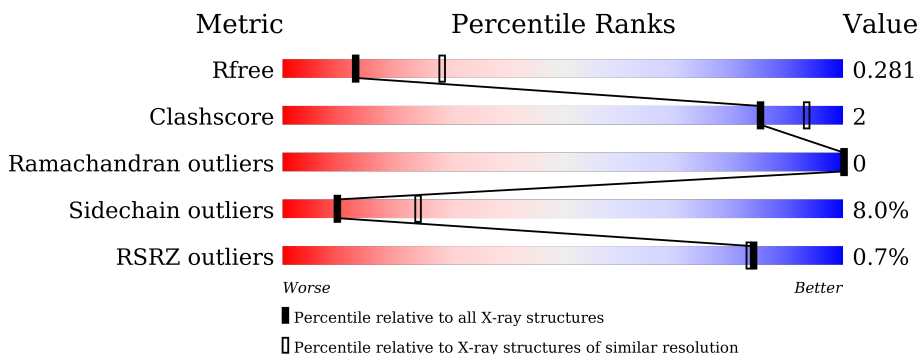|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	3.0
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.002 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.38.2

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 2.70 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	Percentile Ranks	Value
Rfree		0.281
Clashscore		2
Ramachandran outliers		0
Sidechain outliers		8.0%
RSRZ outliers		0.7%

Worse   Better

▌Percentile relative to all X-ray structures

▯Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	164625	3333 (2.70-2.70)
Clashscore	180529	3684 (2.70-2.70)
Ramachandran outliers	177936	3633 (2.70-2.70)
Sidechain outliers	177891	3633 (2.70-2.70)
RSRZ outliers	164620	3333 (2.70-2.70)

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	A	255	75% 8% • 13%
1	B	255	% 75% 9% • 14%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 3274 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 Yellow protein of the takeout family.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	221	Total	C	N	O	S	0	0	0
			1599	1014	267	315	3			
1	B	220	Total	C	N	O	S	0	0	0
			1590	1009	265	313	3			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	19	GLY	-	expression tag	UNP A0A6N3ISN1
A	20	PRO	-	expression tag	UNP A0A6N3ISN1
A	21	HIS	-	expression tag	UNP A0A6N3ISN1
A	22	MET	-	expression tag	UNP A0A6N3ISN1
B	19	GLY	-	expression tag	UNP A0A6N3ISN1
B	20	PRO	-	expression tag	UNP A0A6N3ISN1
B	21	HIS	-	expression tag	UNP A0A6N3ISN1
B	22	MET	-	expression tag	UNP A0A6N3ISN1

- Molecule 2 is BETA-CAROTENE (three-letter code: BCR) (formula: $C_{40}H_{56}$) (labeled as "Ligand of Interest" by depositor).

BCR
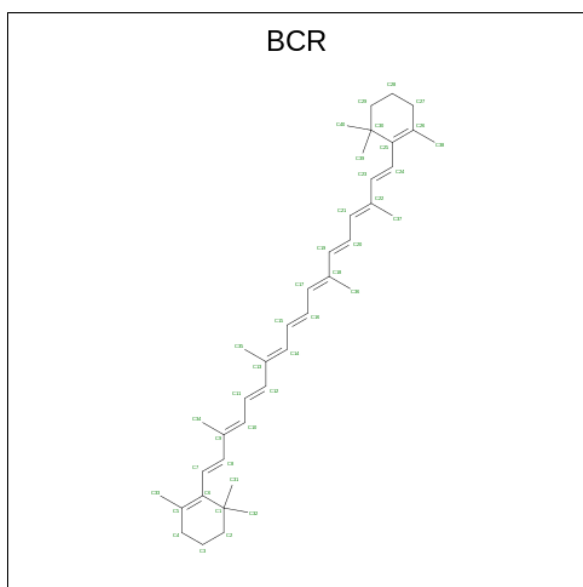
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total 40	C 40	0	0
2	B	1	Total 40	C 40	0	0

• Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	3	Total 3	O 3	0	0
3	B	2	Total 2	O 2	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.
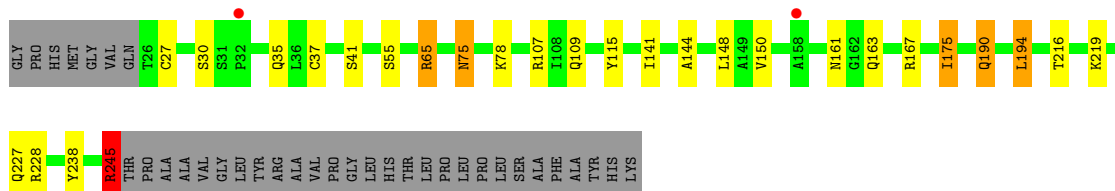
• Molecule 1: Yellow protein of the takeout family

Chain A:
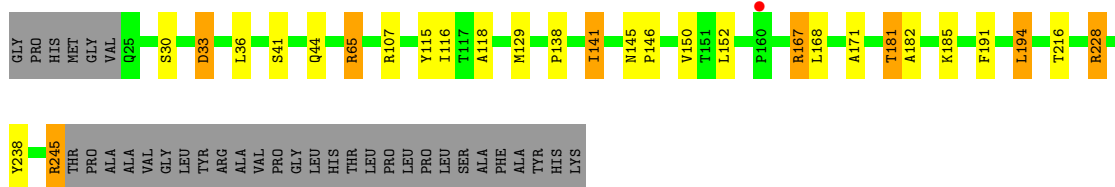
• Molecule 1: Yellow protein of the takeout family

Chain B:

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	55.39Å 75.51Å 224.22Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	28.88 – 2.70 28.88 – 2.70	Depositor EDS
% Data completeness (in resolution range)	99.5 (28.88-2.70) 99.5 (28.88-2.70)	Depositor EDS
R_{merge}	0.41	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.25 (at 2.72Å)	Xtriage
Refinement program	REFMAC 5.8.0425	Depositor
R, R_{free}	0.221 , 0.278 0.224 , 0.281	Depositor DCC
R_{free} test set	653 reflections (4.91%)	wwPDB-VP
Wilson B-factor (Å^2)	47.4	Xtriage
Anisotropy	0.264	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 29.7	EDS
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	3274	wwPDB-VP
Average B, all atoms (Å^2)	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.82% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ 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: BCR

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	0.63	0/1633	1.28	11/2240 (0.5%)
1	B	0.58	0/1624	1.24	11/2228 (0.5%)
All	All	0.61	0/3257	1.26	22/4468 (0.5%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	3
1	B	0	3
All	All	0	6

There are no bond length outliers.

All (22) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	245	ARG	CG-CD-NE	10.05	132.90	111.80
1	A	245	ARG	NE-CZ-NH2	9.77	125.18	120.30
1	A	181	THR	CA-CB-OG1	-9.59	88.86	109.00
1	A	107	ARG	NE-CZ-NH1	8.87	124.73	120.30
1	B	228	ARG	NE-CZ-NH2	-7.86	116.37	120.30
1	B	245	ARG	CD-NE-CZ	7.64	134.29	123.60
1	A	167	ARG	NE-CZ-NH1	7.10	123.85	120.30
1	A	194	LEU	CB-CG-CD1	6.72	122.43	111.00
1	B	163	GLN	CB-CA-C	6.57	123.54	110.40
1	B	107	ARG	NE-CZ-NH2	-6.15	117.22	120.30
1	A	168	LEU	CB-CG-CD1	-6.11	100.62	111.00
1	B	219	LYS	N-CA-CB	5.97	121.36	110.60

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	228	ARG	NE-CZ-NH2	-5.64	117.48	120.30
1	A	181	THR	CA-CB-CG2	5.55	120.17	112.40
1	B	109	GLN	CB-CA-C	5.47	121.33	110.40
1	B	194	LEU	CB-CG-CD1	5.42	120.21	111.00
1	B	65	ARG	NE-CZ-NH1	5.40	123.00	120.30
1	A	115	TYR	CB-CA-C	5.39	121.18	110.40
1	A	167	ARG	CG-CD-NE	5.36	123.06	111.80
1	A	33	ASP	CB-CA-C	5.34	121.08	110.40
1	B	75	ASN	CB-CA-C	5.09	120.59	110.40
1	B	78	LYS	CB-CG-CD	5.07	124.79	111.60

There are no chirality outliers.

All (6) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	167	ARG	Sidechain
1	A	228	ARG	Sidechain
1	A	65	ARG	Sidechain
1	B	161	ASN	Peptide
1	B	167	ARG	Sidechain
1	B	245	ARG	Sidechain

5.2 Too-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 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	1599	0	1609	8	0
1	B	1590	0	1601	5	0
2	A	40	0	56	0	0
2	B	40	0	56	0	0
3	A	3	0	0	0	0
3	B	2	0	0	0	0
All	All	3274	0	3322	12	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 (12) 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:B:27:CYS:HG	1:B:37:CYS:HG	1.32	0.69
1:A:118:ALA:HB3	1:A:141:ILE:HD12	1.88	0.54
1:A:152:LEU:HD23	1:A:171:ALA:HB2	1.91	0.51
1:A:118:ALA:HB3	1:A:141:ILE:CD1	2.41	0.50
1:B:115:TYR:CE2	1:B:144:ALA:HB2	2.47	0.49
1:A:141:ILE:HG23	1:A:182:ALA:HB2	1.98	0.45
1:A:116:ILE:HG12	1:A:146:PRO:HG3	1.98	0.44
1:A:138:PRO:HD2	1:A:185:LYS:HB2	2.00	0.43
1:B:27:CYS:CB	1:B:37:CYS:HG	2.31	0.43
1:A:129:MET:O	1:B:190:GLN:HG3	2.18	0.43
1:A:191:PHE:HD2	1:A:194:LEU:HB2	1.85	0.40
1:B:148:LEU:HD13	1:B:175:ILE:HG22	2.02	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	219/255 (86%)	215 (98%)	4 (2%)	0	100	100
1	B	218/255 (86%)	215 (99%)	3 (1%)	0	100	100
All	All	437/510 (86%)	430 (98%)	7 (2%)	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	176/201 (88%)	163 (93%)	13 (7%)	11	28
1	B	175/201 (87%)	160 (91%)	15 (9%)	8	21
All	All	351/402 (87%)	323 (92%)	28 (8%)	10	24

All (28) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	30	SER
1	A	33	ASP
1	A	36	LEU
1	A	41	SER
1	A	44	GLN
1	A	65	ARG
1	A	141	ILE
1	A	145	ASN
1	A	150	VAL
1	A	181	THR
1	A	216	THR
1	A	238	TYR
1	A	245	ARG
1	B	30	SER
1	B	35	GLN
1	B	41	SER
1	B	55	SER
1	B	65	ARG
1	B	75	ASN
1	B	141	ILE
1	B	150	VAL
1	B	175	ILE
1	B	190	GLN
1	B	194	LEU
1	B	216	THR
1	B	227	GLN
1	B	238	TYR
1	B	245	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	25	GLN
1	A	44	GLN

5.3.3　RNA ⓘ

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

5.5　Carbohydrates ⓘ

There are no oligosaccharides in this entry.

5.6　Ligand geometry ⓘ

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	#$|Z| > 2$
2	BCR	A	301	-	41,41,41	1.98	10 (24%)	56,56,56	1.63	11 (19%)
2	BCR	B	301	-	41,41,41	1.83	9 (21%)	56,56,56	1.64	11 (19%)

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
2	BCR	A	301	-	-	1/29/63/63	0/2/2/2
2	BCR	B	301	-	-	5/29/63/63	0/2/2/2

All (19) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	301	BCR	C14-C13	6.46	1.44	1.35
2	B	301	BCR	C17-C18	6.34	1.44	1.35
2	A	301	BCR	C21-C22	5.54	1.43	1.35
2	B	301	BCR	C14-C13	4.07	1.41	1.35
2	A	301	BCR	C10-C9	3.76	1.40	1.35
2	B	301	BCR	C10-C9	3.63	1.40	1.35
2	B	301	BCR	C8-C9	-3.34	1.38	1.45
2	A	301	BCR	C17-C18	3.22	1.40	1.35
2	A	301	BCR	C19-C18	-3.15	1.39	1.45
2	A	301	BCR	C15-C14	-2.67	1.35	1.43
2	A	301	BCR	C23-C22	-2.67	1.40	1.45
2	B	301	BCR	C12-C13	-2.45	1.40	1.45
2	B	301	BCR	C21-C22	2.37	1.38	1.35
2	A	301	BCR	C8-C7	2.28	1.40	1.33
2	B	301	BCR	C23-C22	-2.19	1.41	1.45
2	A	301	BCR	C29-C28	-2.10	1.47	1.52
2	A	301	BCR	C2-C3	-2.09	1.47	1.52
2	B	301	BCR	C16-C15	2.07	1.41	1.36
2	B	301	BCR	C34-C9	2.01	1.55	1.50

All (22) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	301	BCR	C37-C22-C21	-5.62	115.05	122.92
2	A	301	BCR	C34-C9-C10	-4.42	116.72	122.92
2	A	301	BCR	C34-C9-C8	4.23	124.74	118.08
2	A	301	BCR	C16-C15-C14	3.59	130.82	123.47
2	A	301	BCR	C16-C17-C18	3.40	132.17	127.31
2	B	301	BCR	C15-C14-C13	3.34	132.07	127.31
2	B	301	BCR	C37-C22-C23	3.29	123.26	118.08
2	B	301	BCR	C15-C16-C17	3.13	129.89	123.47
2	B	301	BCR	C29-C30-C25	2.86	114.88	110.48
2	B	301	BCR	C4-C5-C6	-2.85	118.59	122.73
2	B	301	BCR	C34-C9-C8	2.76	122.43	118.08
2	B	301	BCR	C33-C5-C6	2.67	127.53	124.53
2	B	301	BCR	C34-C9-C10	-2.63	119.24	122.92
2	B	301	BCR	C16-C17-C18	2.62	131.06	127.31
2	A	301	BCR	C29-C30-C25	2.44	114.24	110.48
2	A	301	BCR	C21-C20-C19	2.42	130.76	123.22
2	A	301	BCR	C37-C22-C23	2.28	121.67	118.08
2	A	301	BCR	C35-C13-C14	-2.16	119.90	122.92
2	A	301	BCR	C37-C22-C21	-2.14	119.92	122.92

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	301	BCR	C36-C18-C19	2.11	121.39	118.08
2	B	301	BCR	C35-C13-C12	2.09	121.38	118.08
2	A	301	BCR	C11-C12-C13	2.07	132.23	126.42

There are no chirality outliers.
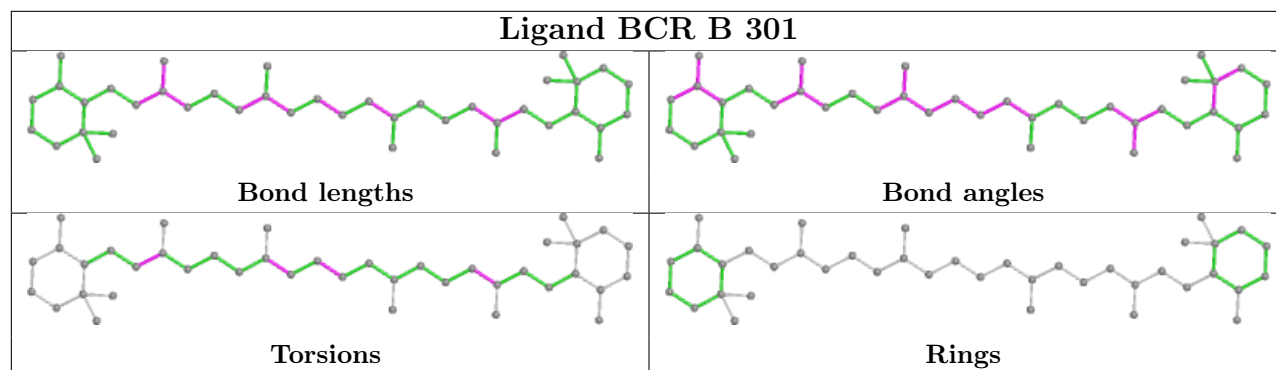
All (6) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	301	BCR	C7-C8-C9-C34
2	A	301	BCR	C10-C11-C12-C13
2	B	301	BCR	C20-C21-C22-C37
2	B	301	BCR	C35-C13-C14-C15
2	B	301	BCR	C14-C15-C16-C17
2	B	301	BCR	C7-C8-C9-C10

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 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.

Ligand BCR A 301

Bond lengths	Bond angles

Torsions	Rings

Ligand BCR B 301	
Bond lengths	Bond angles
Torsions	Rings
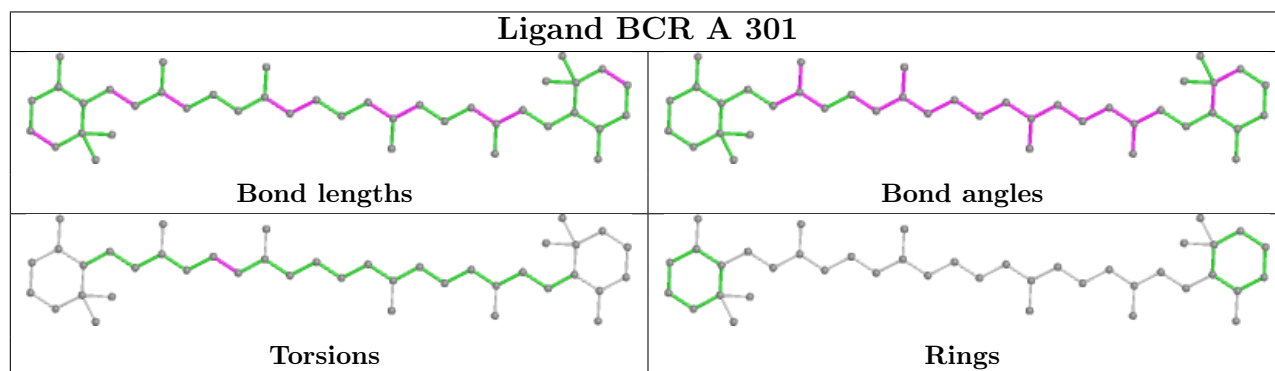
5.7 Other polymers ⓘ

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	A	221/255 (86%)	0.04	1 (0%)	87	86	37, 52, 77, 112	0
1	B	220/255 (86%)	-0.16	2 (0%)	81	80	30, 49, 86, 113	0
All	All	441/510 (86%)	-0.06	3 (0%)	84	83	30, 50, 79, 113	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	160	PRO	3.2
1	B	32	PRO	2.5
1	B	158	ALA	2.3

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

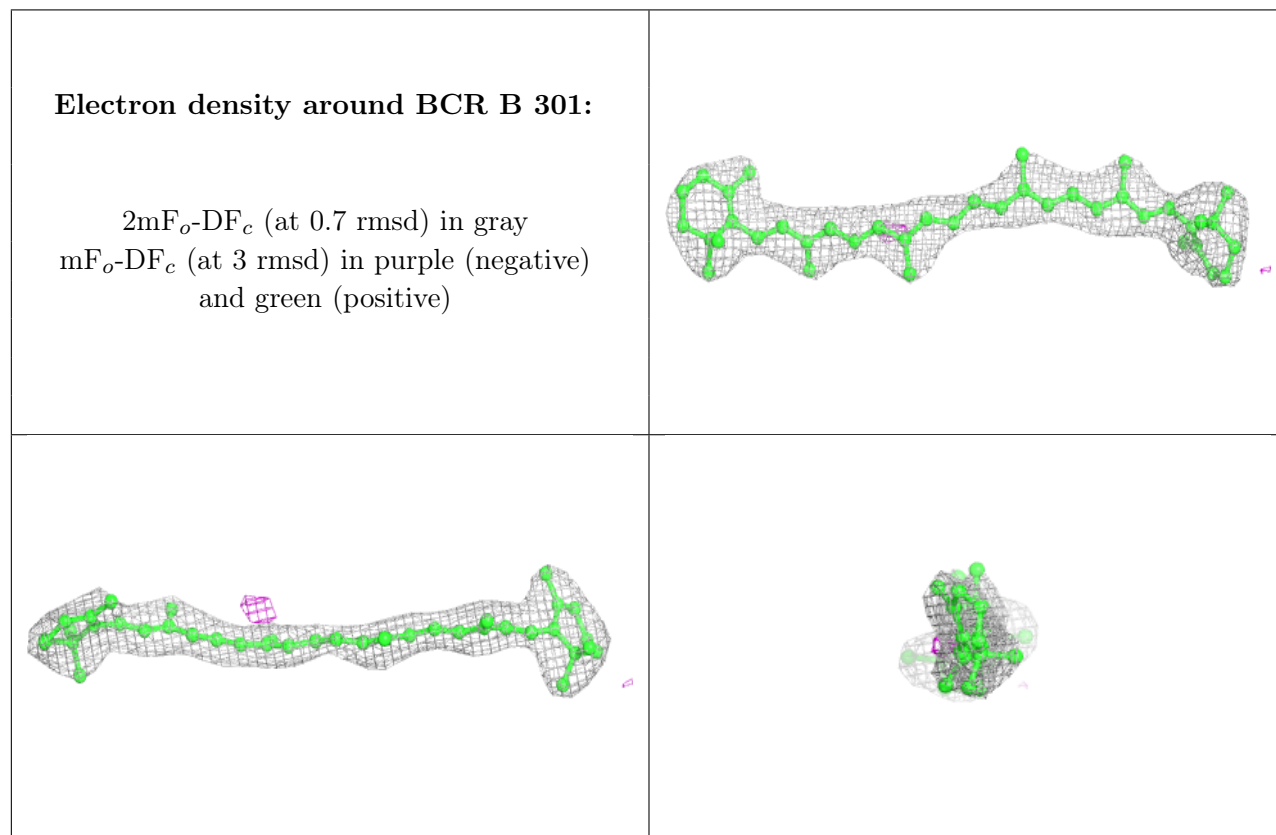There are no non-standard protein/DNA/RNA residues in this entry.

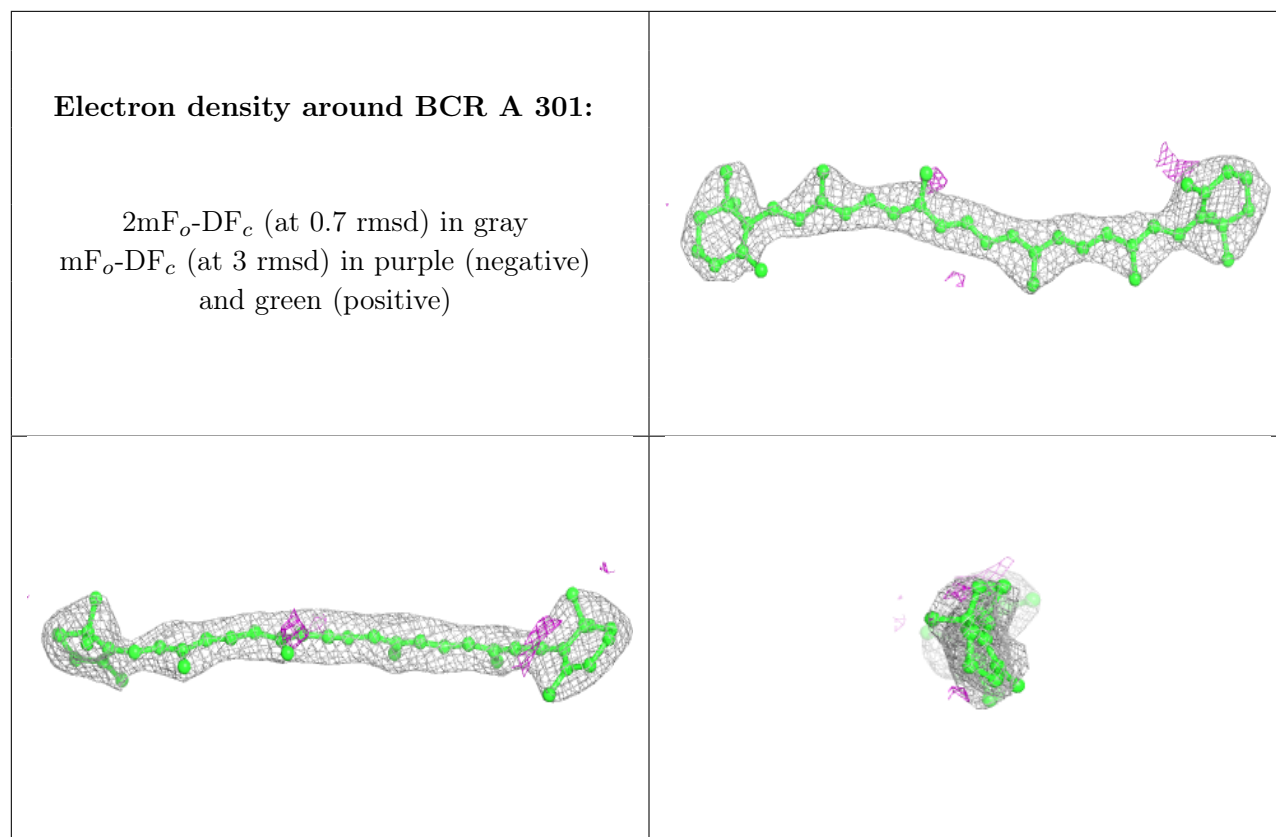## 6.3 Carbohydrates ⓘ

There are no monosaccharides 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. 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(Å^2)	Q<0.9
2	BCR	B	301	40/40	0.92	0.09	26,36,48,51	0
2	BCR	A	301	40/40	0.94	0.09	28,34,43,48	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.

Electron density around BCR B 301:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

Electron density around BCR A 301:

$2mF_o\text{-}DF_c$ (at 0.7 rmsd) in gray
$mF_o\text{-}DF_c$ (at 3 rmsd) in purple (negative)
and green (positive)

6.5 Other polymers ⓘ

There are no such residues in this entry.