wwPDB X-ray Structure Validation Summary Report (i)

Mar 9, 2024 – 09:51 PM EST

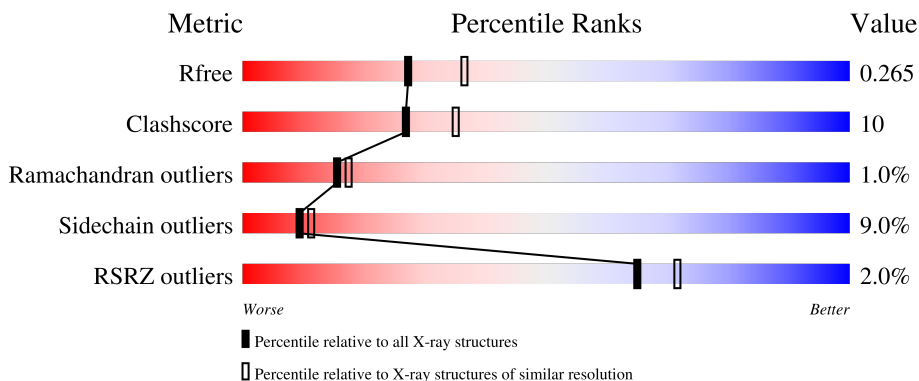|		
PDB ID	:	3IIP
Title	:	Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the KE07 series
Authors	:	Khersonsky, O.; Dym, O.; Tawfik, D.S.
Deposited on	:	2009-08-03
Resolution	:	2.30 Å(reported)

1   Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

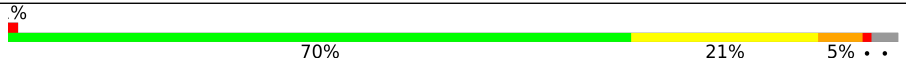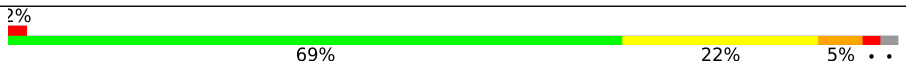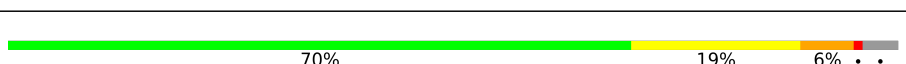The reported resolution of this entry is 2.30 Å.

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		0.265
Clashscore		10
Ramachandran outliers		1.0%
Sidechain outliers		9.0%
RSRZ outliers		2.0%

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$_{free}$	130704	5042 (2.30-2.30)
Clashscore	141614	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)
RSRZ outliers	127900	4938 (2.30-2.30)

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	256	73% 17% 7% · ·
1	B	256	.% 70% 21% 5% · ·
1	C	256	2% 69% 22% 5% · ·
1	D	256	3% 71% 20% 5% · ·
1	E	256	70% 19% 6% · ·

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	256	

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 11756 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 KE7 R6 3/7F.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	249	Total	C	N	O	S	0	0	0
			1935	1234	332	365	4			
1	B	248	Total	C	N	O	S	0	0	1
			1924	1226	331	363	4			
1	C	250	Total	C	N	O	S	0	0	0
			1939	1237	333	365	4			
1	D	247	Total	C	N	O	S	0	0	0
			1919	1224	330	361	4			
1	E	247	Total	C	N	O	S	0	0	0
			1921	1226	329	362	4			
1	F	247	Total	C	N	O	S	0	0	0
			1919	1225	329	361	4			

- Molecule 2 is POLYETHYLENE GLYCOL (N=34) (three-letter code: 15P) (formula: $C_{69}H_{140}O_{35}$).

15P
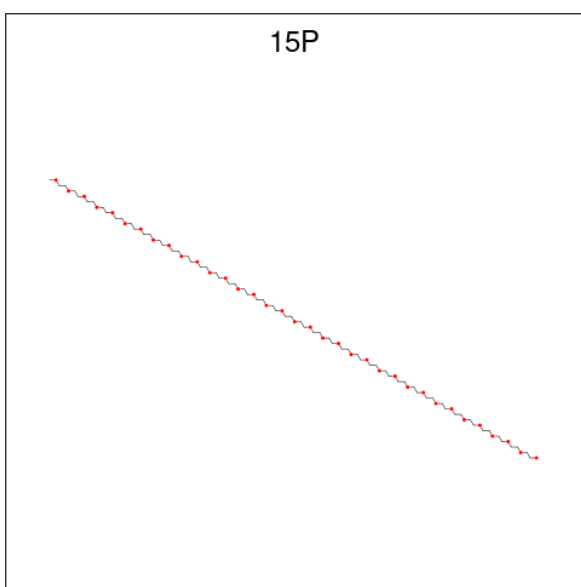
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total 10	C 7	O 3	0	0
2	B	1	Total 11	C 7	O 4	0	0
2	E	1	Total 11	C 7	O 4	0	0

• Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	35	Total 35	O 35	0	0
3	B	23	Total 23	O 23	0	0
3	C	34	Total 34	O 34	0	0
3	D	21	Total 21	O 21	0	0
3	E	34	Total 34	O 34	0	0
3	F	20	Total 20	O 20	0	0

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: KE7 R6 3/7F
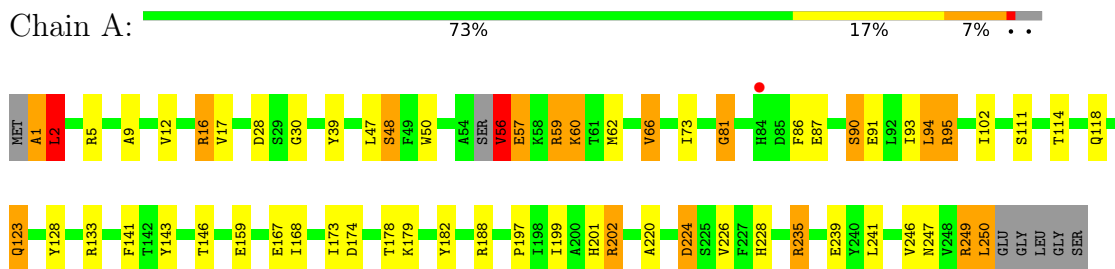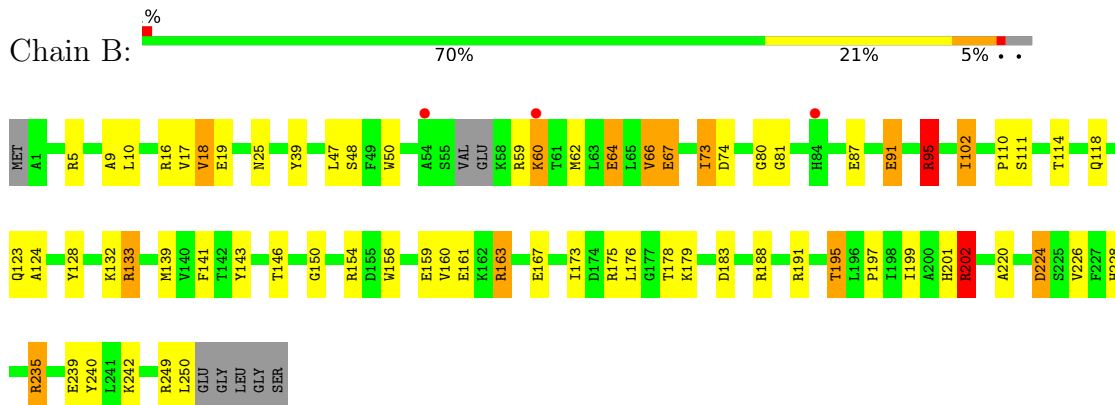
Chain A:

- Molecule 1: KE7 R6 3/7F

Chain B:

- Molecule 1: KE7 R6 3/7F

Chain C:

• Molecule 1: KE7 R6 3/7F

Chain D:

• Molecule 1: KE7 R6 3/7F

Chain E:

• Molecule 1: KE7 R6 3/7F

Chain F:
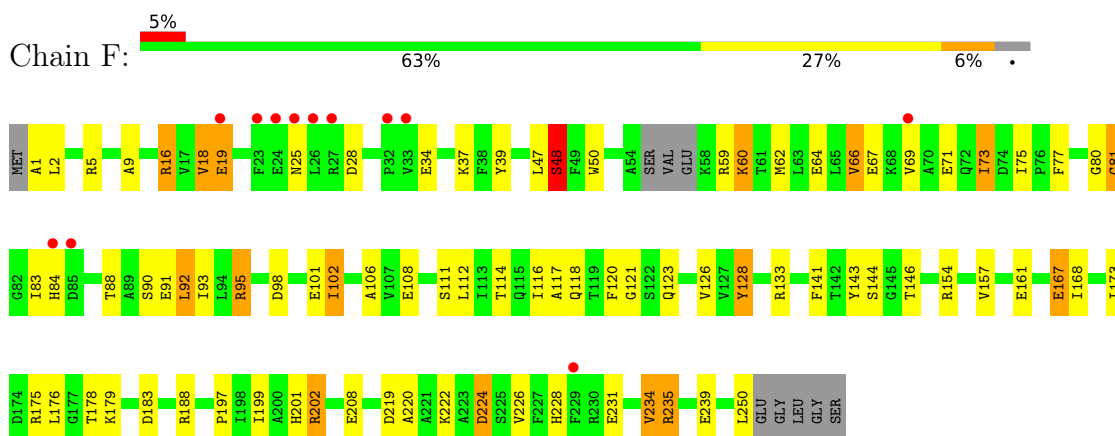
4 Data and refinement statistics ⓘ

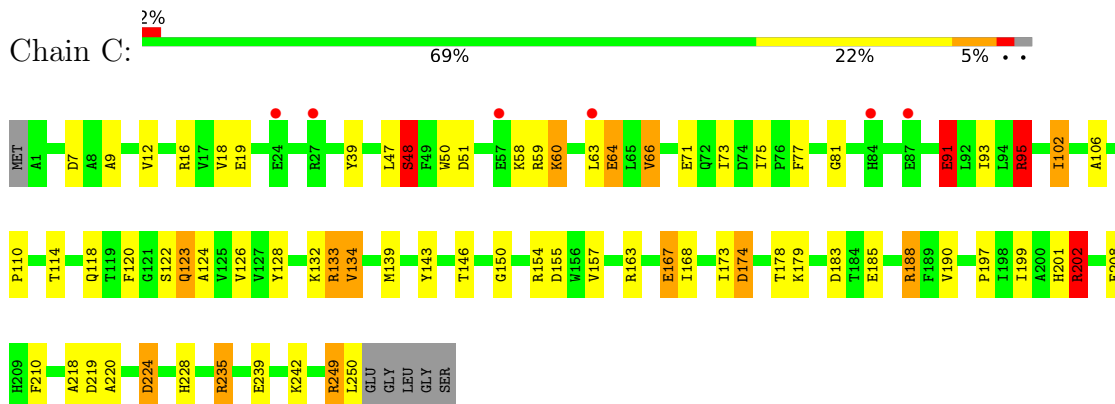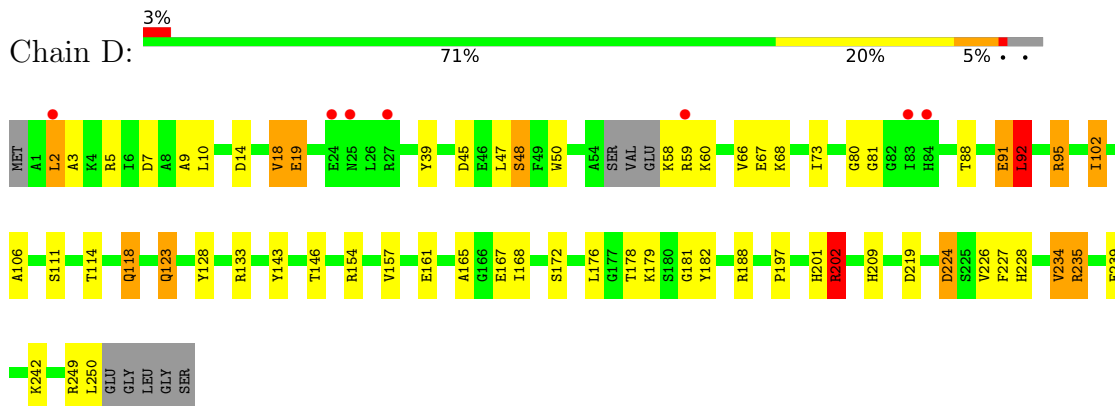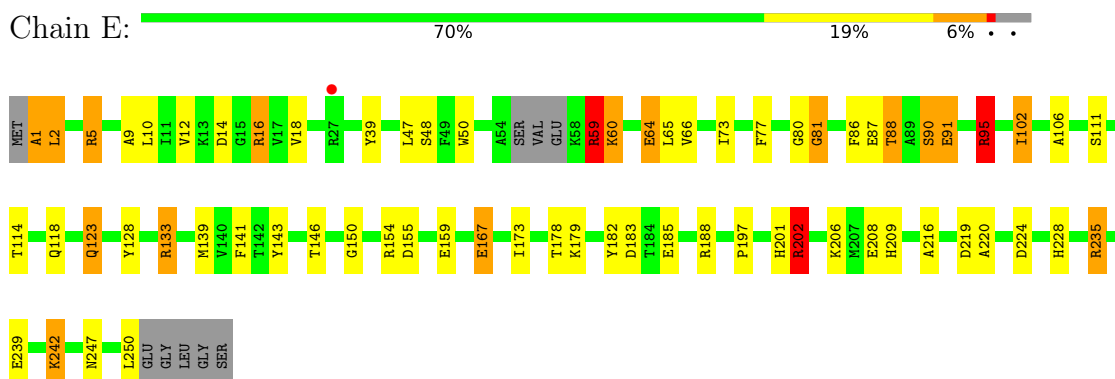| Property	Value	Source
Space group	P 31	Depositor
Cell constants a, b, c, α, β, γ	106.87Å 106.87Å 127.66Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	50.00 – 2.30 43.51 – 2.30	Depositor EDS
% Data completeness (in resolution range)	99.8 (50.00-2.30) 99.8 (43.51-2.30)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	3.27 (at 2.29Å)	Xtriage
Refinement program	REFMAC 5.4.0067	Depositor
R, R_{free}	0.218 , 0.266 0.217 , 0.265	Depositor DCC
R_{free} test set	3672 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å^2)	27.6	Xtriage
Anisotropy	0.045	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 34.4	EDS
L-test for twinning[2]	$<|L|> = 0.50$, $<L^2> = 0.33$	Xtriage
Estimated twinning fraction	0.000 for -h,-k,l 0.028 for h,-h-k,-l 0.000 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	11756	wwPDB-VP
Average B, all atoms (Å^2)	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 48.73 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 8.2518e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

---

[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 ⓘ

5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: 15P

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	1.50	11/1965 (0.6%)	1.38	20/2650 (0.8%)
1	B	1.56	14/1954 (0.7%)	1.33	18/2635 (0.7%)
1	C	1.52	17/1970 (0.9%)	1.45	16/2658 (0.6%)
1	D	1.47	13/1949 (0.7%)	1.33	22/2628 (0.8%)
1	E	1.48	15/1950 (0.8%)	1.39	22/2627 (0.8%)
1	F	1.59	15/1948 (0.8%)	1.37	18/2624 (0.7%)
All	All	1.52	85/11736 (0.7%)	1.38	116/15822 (0.7%)

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	2	4
1	B	0	1
1	C	1	0
1	E	0	1
1	F	0	2
All	All	3	8

The worst 5 of 85 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	91	GLU	CB-CG	-16.31	1.21	1.52
1	F	91	GLU	CB-CG	-13.27	1.26	1.52
1	A	91	GLU	CB-CG	-10.78	1.31	1.52
1	A	48	SER	CB-OG	-10.26	1.28	1.42
1	D	91	GLU	CB-CG	-9.97	1.33	1.52

The worst 5 of 116 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	174	ASP	CB-CG-OD1	-22.81	97.77	118.30
1	C	174	ASP	CB-CG-OD2	21.50	137.65	118.30
1	A	133	ARG	NE-CZ-NH2	-20.15	110.22	120.30
1	E	202	ARG	NE-CZ-NH1	18.32	129.46	120.30
1	A	133	ARG	NE-CZ-NH1	17.36	128.98	120.30

All (3) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	A	94	LEU	CA
1	A	249	ARG	CA
1	C	174	ASP	CA

5 of 8 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	1	ALA	Peptide
1	A	56	VAL	Peptide
1	A	81	GLY	Peptide
1	A	93	ILE	Peptide
1	B	80	GLY	Peptide

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	1935	0	1959	33	0
1	B	1924	0	1948	35	1
1	C	1939	0	1965	47	1
1	D	1919	0	1944	33	0
1	E	1921	0	1946	42	0
1	F	1919	0	1944	46	0
2	A	10	0	10	0	0
2	B	11	0	15	4	0
2	E	11	0	12	0	0
3	A	35	0	0	2	0
3	B	23	0	0	1	0
3	C	34	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	D	21	0	0	0	0
3	E	34	0	0	0	0
3	F	20	0	0	0	0
All	All	11756	0	11743	237	1

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 10.

The worst 5 of 237 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:60:LYS:O	1:B:64:GLU:HG3	1.48	1.12
1:F:143:TYR:O	1:F:146:THR:HG23	1.52	1.05
1:E:143:TYR:O	1:E:146:THR:HG23	1.57	1.03
1:F:83:ILE:C	1:F:84:HIS:CA	2.28	1.01
1:C:39:TYR:OH	1:C:228:HIS:HD2	1.45	0.99

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:95:ARG:NH2	1:C:95:ARG:NH1[3_664]	2.19	0.01

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	245/256 (96%)	232 (95%)	10 (4%)	3 (1%)	13	14
1	B	244/256 (95%)	230 (94%)	11 (4%)	3 (1%)	13	14
1	C	248/256 (97%)	240 (97%)	7 (3%)	1 (0%)	34	42

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	D	243/256 (95%)	234 (96%)	6 (2%)	3 (1%)	13	14
1	E	241/256 (94%)	231 (96%)	8 (3%)	2 (1%)	19	23
1	F	241/256 (94%)	228 (95%)	10 (4%)	3 (1%)	13	14
All	All	1462/1536 (95%)	1395 (95%)	52 (4%)	15 (1%)	15	17

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	94	LEU
1	D	19	GLU
1	D	202	ARG
1	F	18	VAL
1	F	19	GLU

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	203/209 (97%)	187 (92%)	16 (8%)	12	15
1	B	202/209 (97%)	184 (91%)	18 (9%)	9	11
1	C	203/209 (97%)	185 (91%)	18 (9%)	9	11
1	D	201/209 (96%)	180 (90%)	21 (10%)	7	8
1	E	201/209 (96%)	184 (92%)	17 (8%)	10	13
1	F	200/209 (96%)	181 (90%)	19 (10%)	8	10
All	All	1210/1254 (96%)	1101 (91%)	109 (9%)	9	11

5 of 109 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	91	GLU
1	D	242	LYS
1	F	111	SER
1	D	95	ARG

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	D	179	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 22 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	201	HIS
1	E	228	HIS
1	E	123	GLN
1	E	247	ASN
1	B	201	HIS

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 monosaccharides in this entry.

5.6   Ligand geometry ⓘ

3 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	15P	A	256	-	9,9,103	0.90	0	8,8,102	1.13	0
2	15P	E	256	-	10,10,103	0.77	0	9,9,102	0.96	1 (11%)
2	15P	B	256	-	10,10,103	0.83	0	9,9,102	1.62	2 (22%)

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	15P	A	256	-	-	4/7/7/101	-
2	15P	E	256	-	-	5/8/8/101	-
2	15P	B	256	-	-	3/8/8/101	-

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	256	15P	O32-C64-C63	-3.57	94.39	110.07
2	B	256	15P	O32-C65-C66	2.47	121.53	110.39
2	E	256	15P	O33-C67-C68	-2.23	100.27	110.07

There are no chirality outliers.

5 of 12 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	256	15P	O31-C63-C64-O32
2	E	256	15P	C64-C63-O31-C62
2	B	256	15P	O32-C65-C66-O33
2	A	256	15P	C68-C67-O33-C66
2	B	256	15P	C65-C66-O33-C67

There are no ring outliers.

1 monomer is involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	256	15P	4	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 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.
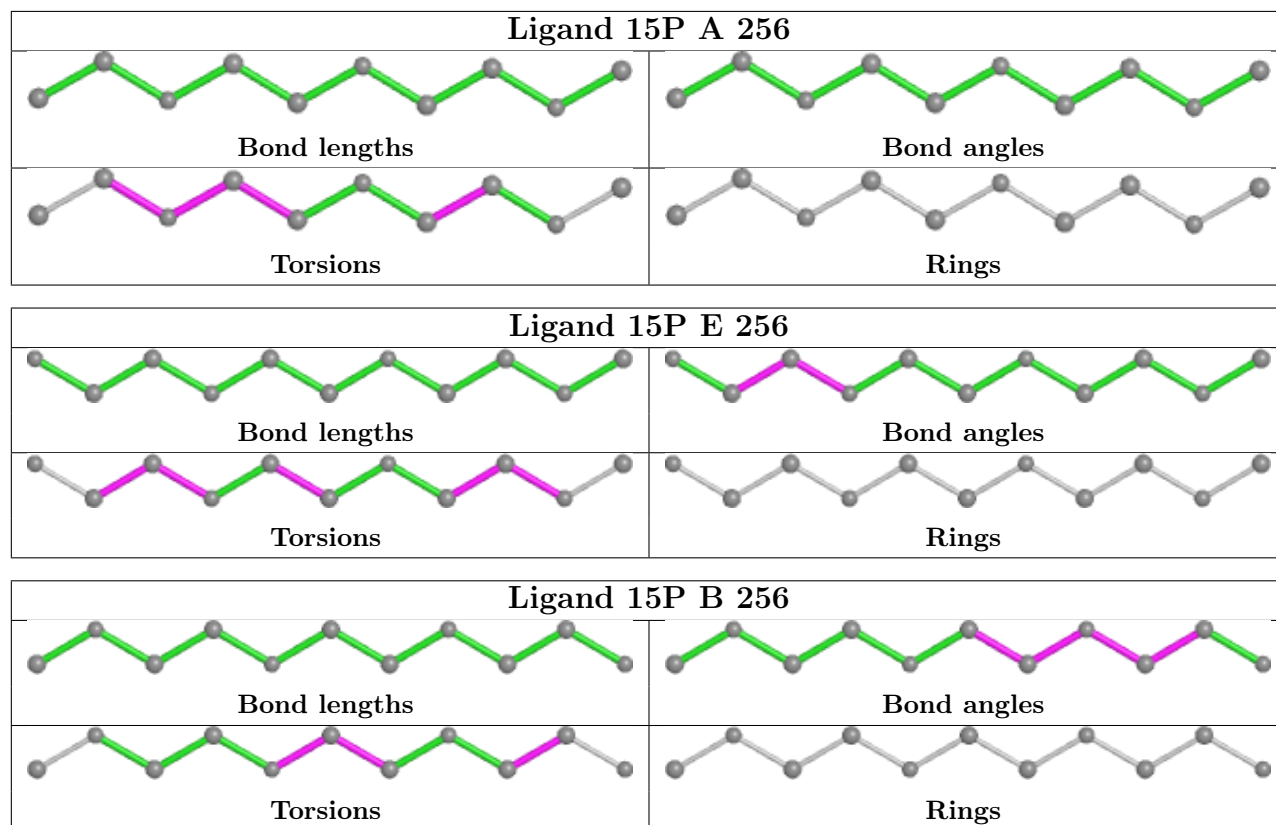
5.7 Other polymers ⓘ

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(Å²)	Q<0.9
1	A	249/256 (97%)	-0.07	1 (0%)	92	95	11, 26, 42, 59	0
1	B	248/256 (96%)	-0.04	3 (1%)	79	83	11, 26, 42, 59	0
1	C	250/256 (97%)	-0.07	6 (2%)	59	66	11, 26, 43, 59	0
1	D	247/256 (96%)	-0.12	7 (2%)	53	60	11, 26, 42, 59	0
1	E	247/256 (96%)	-0.10	1 (0%)	92	95	10, 26, 42, 59	0
1	F	247/256 (96%)	0.14	12 (4%)	29	36	11, 26, 42, 59	0
All	All	1488/1536 (96%)	-0.04	30 (2%)	65	71	10, 26, 43, 59	0

The worst 5 of 30 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	27	ARG	5.0
1	F	24	GLU	4.0
1	C	27	ARG	3.8
1	D	24	GLU	3.5
1	F	23	PHE	3.4

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

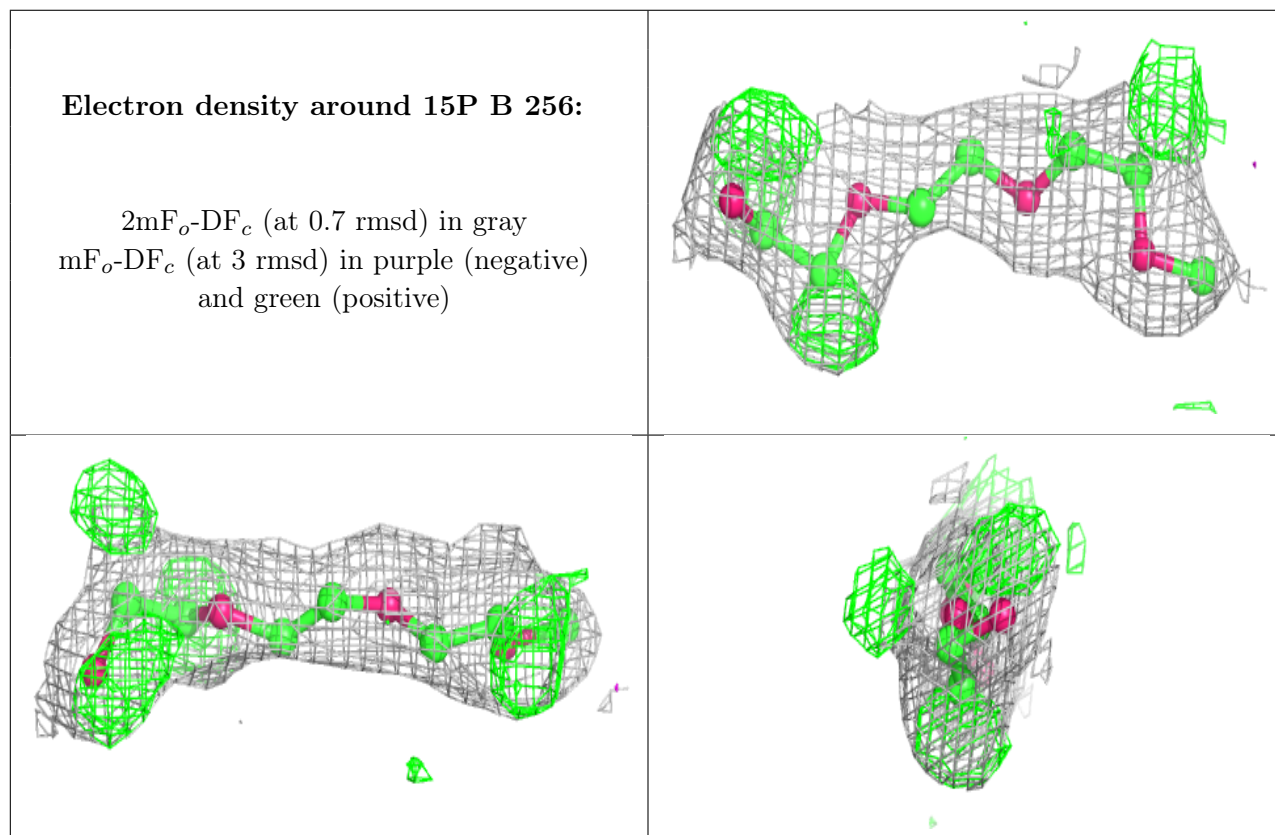There are no non-standard protein/DNA/RNA residues in this entry.

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	15P	B	256	11/104	0.79	0.20	33,38,41,42	0
2	15P	E	256	11/104	0.84	0.21	33,42,48,49	0
2	15P	A	256	10/104	0.92	0.22	33,39,40,42	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.
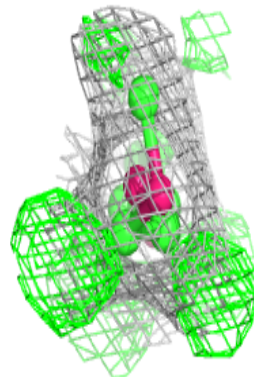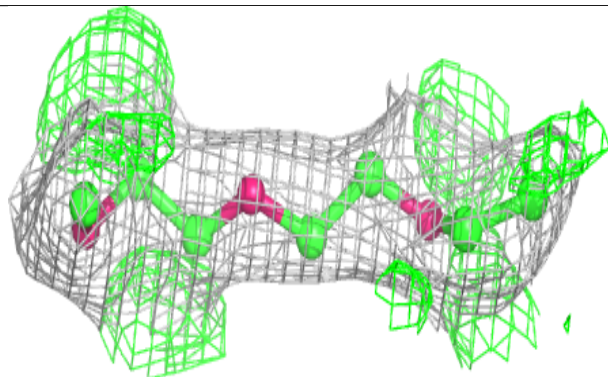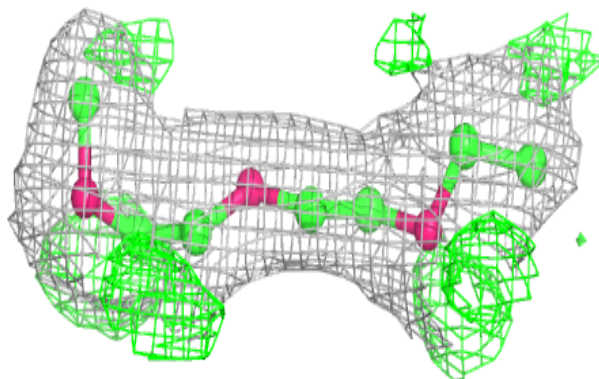

Electron density around 15P B 256:

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

Electron density around 15P E 256:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)
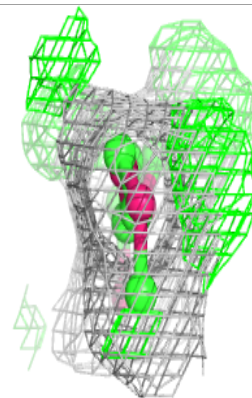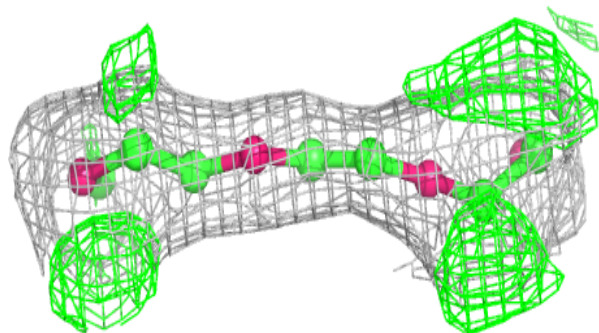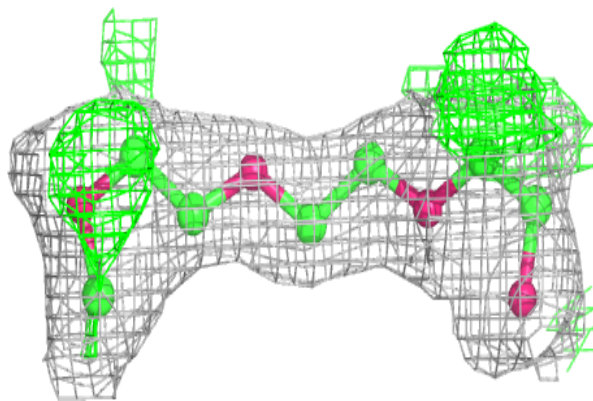
Electron density around 15P A 256:

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

6.5 Other polymers ⓘ

There are no such residues in this entry.