



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 11:16 AM GMT

PDB ID : 3OE8  
Title : Crystal structure of the CXCR4 chemokine receptor in complex with a small molecule antagonist IT1t in P1 spacegroup  
Authors : Wu, B.; Mol, C.D.; Han, G.W.; Katritch, V.; Chien, E.Y.T.; Liu, W.; Cherezov, V.; Stevens, R.C.; Accelerated Technologies Center for Gene to 3D Structure (ATCG3D); GPCR Network (GPCR)  
Deposited on : 2010-08-12  
Resolution : 3.10 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](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

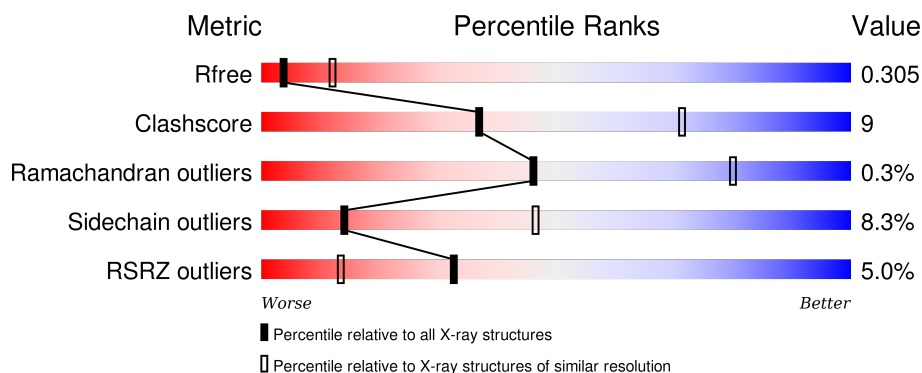
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.10 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))
$R_{free}$	91344	1114 (3.14-3.06)
Clashscore	102246	1222 (3.14-3.06)
Ramachandran outliers	100387	1174 (3.14-3.06)
Sidechain outliers	100360	1174 (3.14-3.06)
RSRZ outliers	91569	1119 (3.14-3.06)

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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. 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	502	<div> <div>3%</div> <div>66%</div> <div>15%</div> <div>• •</div> <div>14%</div> </div>
1	B	502	<div> <div>6%</div> <div>68%</div> <div>14%</div> <div>•</div> <div>15%</div> </div>
1	C	502	<div> <div>4%</div> <div>72%</div> <div>12%</div> <div>•</div> <div>15%</div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 10373 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 C-X-C chemokine receptor type 4, Lysozyme Chimera.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	433	Total	C	N	O	S	0	0	0
			3459	2269	577	597	16			
1	B	426	Total	C	N	O	S	0	0	0
			3404	2240	570	578	16			
1	C	428	Total	C	N	O	S	0	0	0
			3429	2263	567	583	16			

There are 54 discrepancies between the modelled and reference sequences:

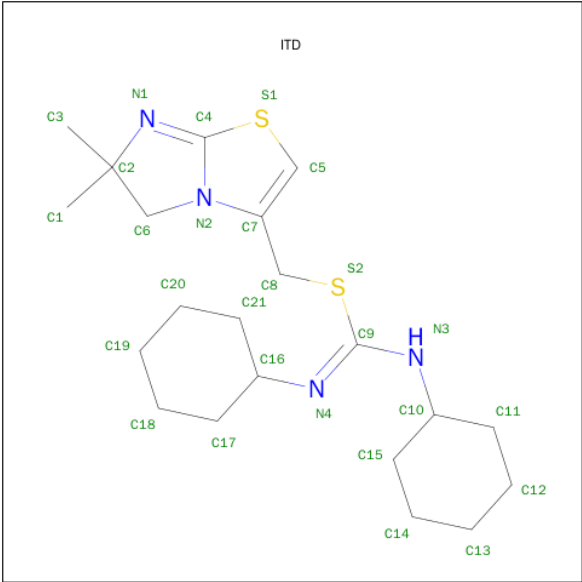
Chain	Residue	Modelled	Actual	Comment	Reference
A	-9	ASP	-	expression tag	UNP P61073
A	-8	TYR	-	expression tag	UNP P61073
A	-7	LYS	-	expression tag	UNP P61073
A	-6	ASP	-	expression tag	UNP P61073
A	-5	ASP	-	expression tag	UNP P61073
A	-4	ASP	-	expression tag	UNP P61073
A	-3	ASP	-	expression tag	UNP P61073
A	-2	ALA	-	expression tag	UNP P61073
A	-1	GLY	-	expression tag	UNP P61073
A	0	ALA	-	expression tag	UNP P61073
A	1	PRO	-	expression tag	UNP P61073
A	125	TRP	LEU	engineered	UNP P61073
A	900	GLY	-	linker	UNP P61073
A	901	SER	-	linker	UNP P61073
A	1200	GLY	-	linker	UNP P61073
A	1201	SER	-	linker	UNP P61073
A	1054	THR	CYS	engineered	UNP P00720
A	1097	ALA	CYS	engineered	UNP P00720
A	320	GLY	-	expression tag	UNP P61073
A	321	ARG	-	expression tag	UNP P61073
A	322	PRO	-	expression tag	UNP P61073
A	323	LEU	-	expression tag	UNP P61073
A	324	GLU	-	expression tag	UNP P61073

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	325	VAL	-	expression tag	UNP P61073
A	326	LEU	-	expression tag	UNP P61073
A	327	PHE	-	expression tag	UNP P61073
A	328	GLN	-	expression tag	UNP P61073
B	-9	ASP	-	expression tag	UNP P61073
B	-8	TYR	-	expression tag	UNP P61073
B	-7	LYS	-	expression tag	UNP P61073
B	-6	ASP	-	expression tag	UNP P61073
B	-5	ASP	-	expression tag	UNP P61073
B	-4	ASP	-	expression tag	UNP P61073
B	-3	ASP	-	expression tag	UNP P61073
B	-2	ALA	-	expression tag	UNP P61073
B	-1	GLY	-	expression tag	UNP P61073
B	0	ALA	-	expression tag	UNP P61073
B	1	PRO	-	expression tag	UNP P61073
B	125	TRP	LEU	engineered	UNP P61073
B	900	GLY	-	linker	UNP P61073
B	901	SER	-	linker	UNP P61073
B	1200	GLY	-	linker	UNP P61073
B	1201	SER	-	linker	UNP P61073
B	1054	THR	CYS	engineered	UNP P00720
B	1097	ALA	CYS	engineered	UNP P00720
B	320	GLY	-	expression tag	UNP P61073
B	321	ARG	-	expression tag	UNP P61073
B	322	PRO	-	expression tag	UNP P61073
B	323	LEU	-	expression tag	UNP P61073
B	324	GLU	-	expression tag	UNP P61073
B	325	VAL	-	expression tag	UNP P61073
B	326	LEU	-	expression tag	UNP P61073
B	327	PHE	-	expression tag	UNP P61073
B	328	GLN	-	expression tag	UNP P61073

- Molecule 2 is (6,6-DIMETHYL-5,6-DIHYDROIMIDAZO[2,1-B][1,3]THIAZOL-3-YL)METHYL N,N'-DICYCLOHEXYLIMIDOTHIOCARBAMATE (three-letter code: ITD) (formula: C<sub>21</sub>H<sub>34</sub>N<sub>4</sub>S<sub>2</sub>).

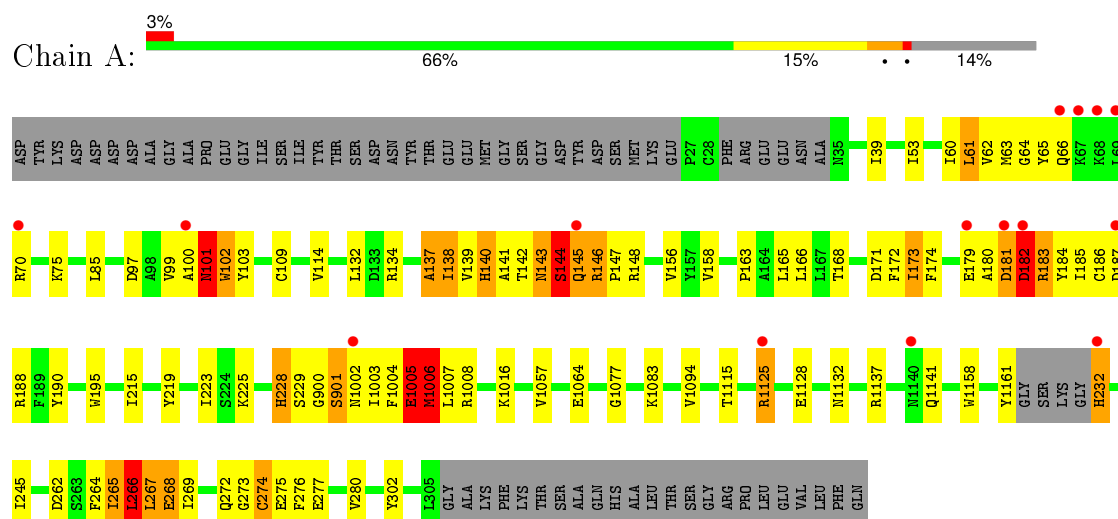


Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	S	0	0
			27	21	4	2		
2	B	1	Total	C	N	S	0	0
			27	21	4	2		
2	C	1	Total	C	N	S	0	0
			27	21	4	2		

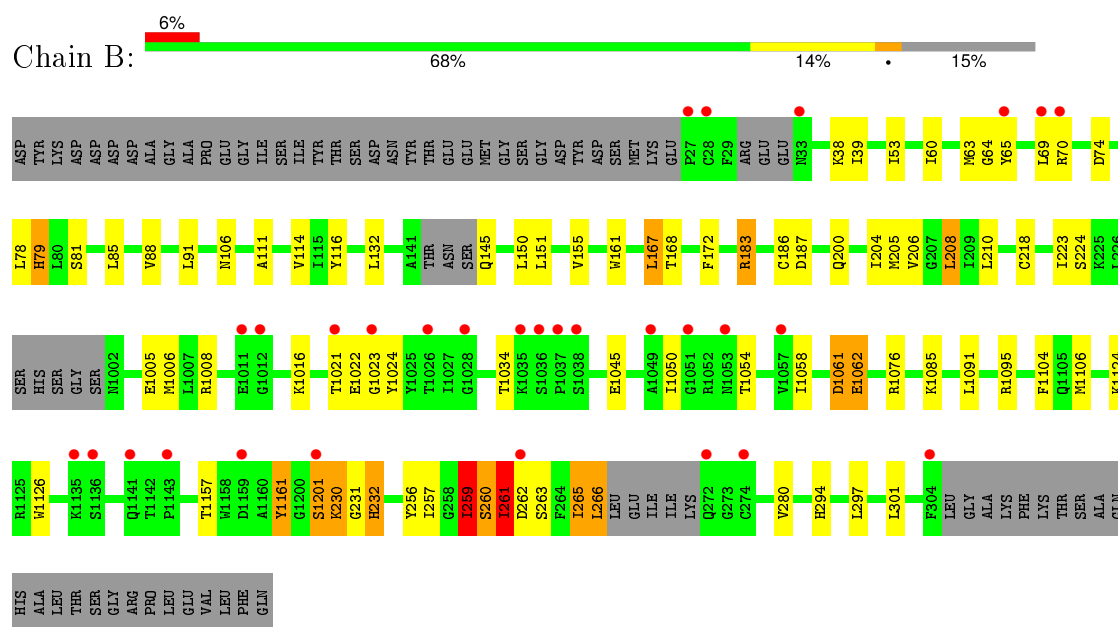
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

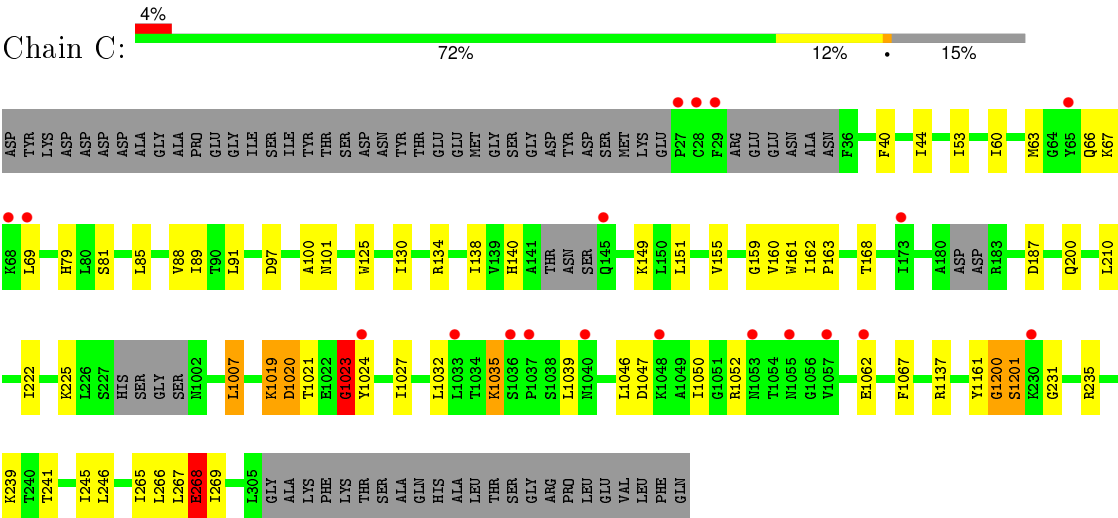
- Molecule 1: C-X-C chemokine receptor type 4, Lysozyme Chimera



- Molecule 1: C-X-C chemokine receptor type 4, Lysozyme Chimera



- Molecule 1: C-X-C chemokine receptor type 4, Lysozyme Chimera



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	69.37Å 76.60Å 91.72Å 96.00° 97.78° 97.38°	Depositor
Resolution (Å)	19.97 – 3.10 19.97 – 3.10	Depositor EDS
% Data completeness (in resolution range)	(Not available) (19.97-3.10) 83.8 (19.97-3.10)	Depositor EDS
$R_{merge}$	0.15	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.31 (at 3.09Å)	Xtriage
Refinement program	BUSTER 2.8.0	Depositor
R, $R_{free}$	0.250 , 0.295 0.251 , 0.305	Depositor DCC
$R_{free}$ test set	1441 reflections (5.30%)	DCC
Wilson B-factor (Å <sup>2</sup> )	97.5	Xtriage
Anisotropy	0.090	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 79.5	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 28647 reflections	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	10373	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	103.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.49% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: ITD

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	0.50	0/3543	0.92	30/4814 (0.6%)
1	B	0.48	0/3485	0.77	12/4731 (0.3%)
1	C	0.50	3/3511 (0.1%)	0.85	16/4767 (0.3%)
All	All	0.49	3/10539 (0.0%)	0.85	58/14312 (0.4%)

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	C	0	3

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	1201	SER	C-O	10.02	1.42	1.23
1	C	1201	SER	CA-C	-8.01	1.32	1.52
1	C	1200	GLY	CA-C	-6.40	1.41	1.51

The worst 5 of 58 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	1201	SER	N-CA-C	13.97	148.71	111.00
1	C	1201	SER	CB-CA-C	-13.45	84.54	110.10
1	C	1201	SER	N-CA-C	13.16	146.55	111.00
1	A	232	HIS	N-CA-CB	-12.66	87.82	110.60
1	A	229	SER	O-C-N	-12.10	102.64	123.20

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	1200	GLY	Mainchain,Peptide
1	C	1201	SER	Mainchain

## 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	3459	0	3494	105	0
1	B	3404	0	3454	66	0
1	C	3429	0	3495	27	0
2	A	27	0	34	2	0
2	B	27	0	34	4	0
2	C	27	0	34	4	0
All	All	10373	0	10545	196	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 9.

The worst 5 of 196 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:265:ILE:HG22	1:A:266:LEU:N	1.55	1.12
1:C:1024:TYR:HE2	1:C:1035:LYS:HB3	1.19	1.06
1:A:143:ASN:ND2	1:A:143:ASN:H	1.52	1.06
1:C:267:LEU:O	1:C:268:GLU:HG3	1.59	1.01
1:C:1024:TYR:CE2	1:C:1035:LYS:HB3	1.96	1.01

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	427/502 (85%)	395 (92%)	29 (7%)	3 (1%)	26	65
1	B	416/502 (83%)	394 (95%)	22 (5%)	0	100	100
1	C	418/502 (83%)	402 (96%)	15 (4%)	1 (0%)	52	84
All	All	1261/1506 (84%)	1191 (94%)	66 (5%)	4 (0%)	46	80

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	1023	GLY
1	A	182	ASP
1	A	274	CYS
1	A	1077	GLY

### 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	368/432 (85%)	336 (91%)	32 (9%)	13	44
1	B	359/432 (83%)	328 (91%)	31 (9%)	13	45
1	C	365/432 (84%)	337 (92%)	28 (8%)	16	50
All	All	1092/1296 (84%)	1001 (92%)	91 (8%)	14	46

5 of 91 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	167	LEU
1	B	1062	GLU
1	C	1137	ARG
1	B	168	THR
1	B	205	MET

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

Mol	Chain	Res	Type
1	B	79	HIS
1	B	119	ASN
1	C	203	HIS
1	A	272	GLN
1	C	66	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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

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	ITD	A	1500	-	25,30,30	2.03	5 (20%)	27,42,42	2.34	7 (25%)
2	ITD	B	1500	-	25,30,30	2.09	5 (20%)	27,42,42	2.06	10 (37%)
2	ITD	C	1500	-	25,30,30	2.08	5 (20%)	27,42,42	1.74	7 (25%)

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	ITD	A	1500	-	-	0/11/39/39	0/3/4/4
2	ITD	B	1500	-	-	0/11/39/39	0/3/4/4
2	ITD	C	1500	-	-	0/11/39/39	0/3/4/4

The worst 5 of 15 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	1500	ITD	C9-S2	-8.00	1.66	1.75
2	B	1500	ITD	C9-S2	-7.89	1.66	1.75
2	A	1500	ITD	C9-S2	-7.52	1.67	1.75
2	A	1500	ITD	C2-N1	-3.20	1.46	1.49
2	B	1500	ITD	C2-N1	-2.93	1.46	1.49

The worst 5 of 24 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	1500	ITD	C7-C8-S2	-3.34	106.58	112.84
2	C	1500	ITD	C7-C8-S2	-3.28	106.69	112.84
2	B	1500	ITD	C7-C8-S2	-2.73	107.73	112.84
2	B	1500	ITD	C3-C2-C6	-2.27	108.68	112.13
2	B	1500	ITD	C12-C11-C10	-2.07	107.93	111.13

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 10 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1500	ITD	2	0
2	B	1500	ITD	4	0
2	C	1500	ITD	4	0

## 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 [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<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	433/502 (86%)	0.02	15 (3%)	48 23	68, 89, 125, 161	0
1	B	426/502 (84%)	0.21	30 (7%)	19 7	63, 100, 181, 198	0
1	C	428/502 (85%)	0.13	19 (4%)	38 17	72, 98, 169, 207	0
All	All	1287/1506 (85%)	0.12	64 (4%)	32 13	63, 94, 169, 207	0

The worst 5 of 64 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	232	HIS	6.4
1	A	66	GLN	5.0
1	A	69	LEU	4.8
1	C	1037	PRO	4.7
1	C	29	PHE	4.7

### 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 carbohydrates 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	ITD	A	1500	27/27	0.88	0.23	-0.20	37,105,187,191	0
2	ITD	C	1500	27/27	0.91	0.21	-0.25	35,105,168,218	0
2	ITD	B	1500	27/27	0.95	0.19	-0.28	24,89,203,224	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.