Full wwPDB X-ray Structure Validation Report (i)

May 25, 2020 – 05:39 am BST

PDB ID	:	4O9V
Title	:	Crystal structure of matriptase in complex with inhibitor
Authors	:	Rao, K.N.; Chandra, B.R.; Ashok, K.N.; Chakshusmathi, G.; Ramesh, K.S.; Subramanya, H.S.
Deposited on	:	2014-01-03
Resolution	:	1.90 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

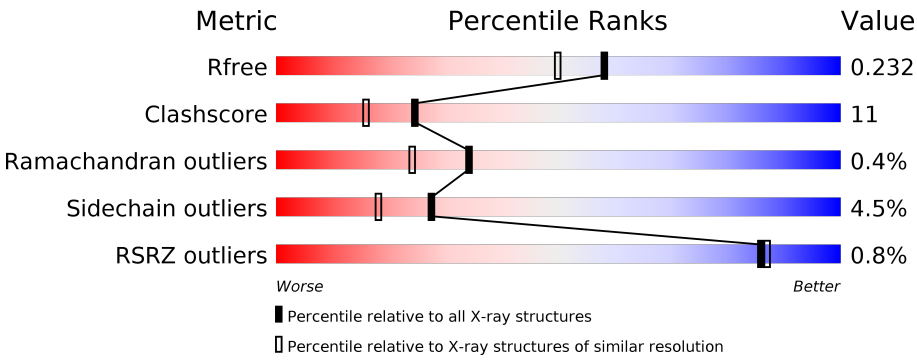
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 1.90 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	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)

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 >=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	241	79% 15% 5%
2	B	4	50% 50% 50%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 2206 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.

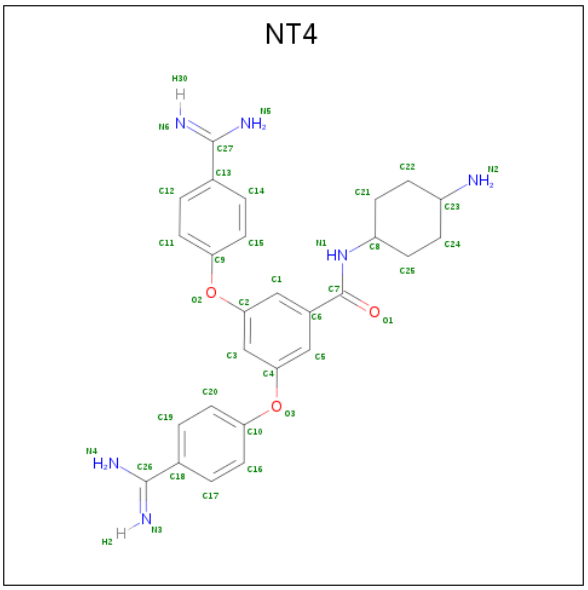
- Molecule 1 is a protein called Suppressor of tumorigenicity 14 protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	241	1857	1176	325	346	10	0	0	0

- Molecule 2 is a protein called Peptide CGLR.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	4	29	17	7	4	1	0	0	0

- Molecule 3 is N-(trans-4-aminocyclohexyl)-3,5-bis(4-carbamimidoylphenoxy)benzamide (three-letter code: NT4) (formula: $C_{27}H_{30}N_6O_3$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
3	A	1	36	27	6	3	0	0

- Molecule 4 is water.

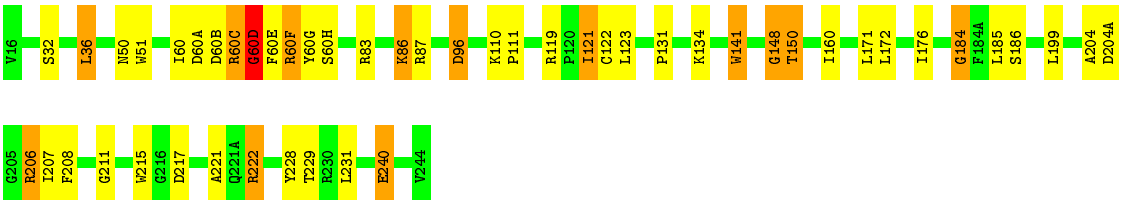
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	284	284	284	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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: Suppressor of tumorigenicity 14 protein

Chain A:



- Molecule 2: Peptide CGLR

Chain B:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 2	Depositor
Cell constants a, b, c, α, β, γ	66.64Å 140.68Å 51.86Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	28.03 – 1.90 28.03 – 1.90	Depositor EDS
% Data completeness (in resolution range)	94.7 (28.03-1.90) 94.8 (28.03-1.90)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.79 (at 1.91Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, R_{free}	0.170 , 0.239 0.184 , 0.232	Depositor DCC
R_{free} test set	954 reflections (5.12%)	wwPDB-VP
Wilson B-factor (Å^2)	12.4	Xtriage
Anisotropy	0.039	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 42.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	2206	wwPDB-VP
Average B, all atoms (Å^2)	13.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.94% 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: NT4

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	1.40	16/1909 (0.8%)	1.23	15/2599 (0.6%)
2	B	0.99	0/28	1.08	0/35
All	All	1.40	16/1937 (0.8%)	1.23	15/2634 (0.6%)

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	1

All (16) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	60(C)	ARG	C-N	20.72	1.70	1.33
1	A	148	GLY	C-N	15.82	1.70	1.34
1	A	184	GLY	C-N	11.19	1.59	1.34
1	A	60(E)	PHE	C-N	10.34	1.57	1.34
1	A	60(G)	TYR	C-N	10.23	1.57	1.34
1	A	186	SER	C-N	10.12	1.51	1.33
1	A	217	ASP	C-N	10.11	1.51	1.33
1	A	60(D)	GLY	C-N	9.87	1.56	1.34
1	A	222	ARG	CD-NE	8.75	1.61	1.46
1	A	60(B)	ASP	C-N	8.50	1.53	1.34
1	A	221	ALA	C-N	7.92	1.52	1.34
1	A	51	TRP	CD2-CE2	7.36	1.50	1.41
1	A	60(F)	ARG	C-N	6.80	1.49	1.34
1	A	141	TRP	CD2-CE2	5.80	1.48	1.41
1	A	240	GLU	CG-CD	5.77	1.60	1.51
1	A	204	ALA	C-N	5.12	1.45	1.34

All (15) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	148	GLY	O-C-N	-17.82	94.19	122.70
1	A	206	ARG	NE-CZ-NH2	-13.50	113.55	120.30
1	A	206	ARG	NE-CZ-NH1	12.95	126.78	120.30
1	A	148	GLY	CA-C-N	11.49	142.48	117.20
1	A	60(G)	TYR	O-C-N	-10.94	105.19	122.70
1	A	148	GLY	C-N-CA	7.35	140.07	121.70
1	A	217	ASP	O-C-N	-7.10	111.12	123.20
1	A	60(G)	TYR	CA-C-N	7.07	132.75	117.20
1	A	60(A)	ASP	C-N-CA	-6.06	106.54	121.70
1	A	87	ARG	NE-CZ-NH1	-5.85	117.38	120.30
1	A	186	SER	O-C-N	-5.53	113.81	123.20
1	A	217	ASP	CA-C-N	5.42	127.05	116.20
1	A	123	LEU	CB-CG-CD1	-5.31	101.98	111.00
1	A	60(B)	ASP	O-C-N	-5.27	114.27	122.70
1	A	60(H)	SER	O-C-N	-5.26	114.28	122.70

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	60(D)	GLY	Mainchain

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	1857	0	1769	35	3
2	B	29	0	30	11	0
3	A	36	0	29	1	0
4	A	284	0	0	8	3
All	All	2206	0	1828	41	3

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 11.

All (41) 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:60(C):ARG:C	1:A:60(D):GLY:N	1.70	1.42
1:A:148:GLY:C	1:A:150:THR:N	1.70	1.40
1:A:172:LEU:HB2	1:A:176:ILE:HD11	1.37	1.05
2:B:8:ARG:N	2:B:8:ARG:HD3	1.88	0.87
1:A:119:ARG:CD	2:B:7:LEU:O	2.23	0.87
1:A:172:LEU:HB2	1:A:176:ILE:CD1	2.07	0.84
1:A:119:ARG:HD3	2:B:7:LEU:O	1.80	0.81
1:A:119:ARG:HE	2:B:7:LEU:H	1.30	0.78
1:A:148:GLY:O	1:A:150:THR:N	2.17	0.78
1:A:60(F):ARG:CZ	4:A:1168:HOH:O	2.39	0.71
2:B:6:GLY:O	2:B:7:LEU:HD13	1.95	0.67
1:A:110:LYS:HG2	4:A:1163:HOH:O	1.94	0.66
1:A:121:ILE:HD13	1:A:122:CYS:H	1.61	0.66
2:B:8:ARG:N	2:B:8:ARG:CD	2.61	0.64
1:A:60(F):ARG:NH1	4:A:1168:HOH:O	2.31	0.64
1:A:121:ILE:HD13	1:A:122:CYS:N	2.13	0.63
1:A:119:ARG:NE	2:B:7:LEU:O	2.33	0.62
1:A:110:LYS:CG	4:A:1163:HOH:O	2.52	0.56
1:A:160:ILE:HB	1:A:184:GLY:HA2	1.91	0.53
1:A:119:ARG:HD2	4:A:1125:HOH:O	2.08	0.52
4:A:1173:HOH:O	2:B:7:LEU:CD2	2.60	0.49
1:A:121:ILE:HD13	1:A:207:ILE:O	2.12	0.49
1:A:199:LEU:HD21	1:A:228:TYR:CE2	2.47	0.49
1:A:148:GLY:HA2	4:A:1196:HOH:O	2.12	0.48
1:A:60:ILE:HD13	1:A:60:ILE:N	2.28	0.47
1:A:121:ILE:CD1	1:A:122:CYS:N	2.75	0.47
2:B:6:GLY:O	2:B:7:LEU:CD1	2.62	0.47
1:A:131:PRO:HG2	1:A:134:LYS:HB2	1.98	0.46
1:A:86:LYS:O	1:A:86:LYS:HD3	2.15	0.46
1:A:119:ARG:NE	2:B:7:LEU:H	2.08	0.45
1:A:50:ASN:HD21	1:A:111:PRO:HG3	1.82	0.44
1:A:215:TRP:CD2	3:A:900:NT4:H14	2.53	0.43
1:A:121:ILE:HD11	1:A:208:PHE:C	2.39	0.43
1:A:204(A):ASP:OD1	1:A:206:ARG:HD3	2.20	0.42
1:A:36:LEU:HA	1:A:36:LEU:HD23	1.76	0.42
1:A:32:SER:HB2	1:A:141:TRP:CZ3	2.55	0.42
1:A:83:ARG:CZ	1:A:110:LYS:HE3	2.51	0.41
1:A:204(A):ASP:OD1	1:A:206:ARG:CD	2.69	0.41
1:A:211:GLY:HA2	1:A:229:THR:O	2.20	0.41
4:A:1162:HOH:O	2:B:5:CYS:HB3	2.21	0.41
1:A:171:LEU:HD11	1:A:185:LEU:HD22	2.04	0.40

All (3) symmetry-related close contacts are listed below. The label for Atom-2 includes the sym-

metry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:222:ARG:CZ	4:A:1071:HOH:O[3_556]	1.82	0.38
1:A:222:ARG:NE	4:A:1071:HOH:O[3_556]	1.92	0.28
1:A:222:ARG:NH1	4:A:1042:HOH:O[3_556]	2.13	0.07

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	239/241 (99%)	230 (96%)	8 (3%)	1 (0%)	34	24
2	B	2/4 (50%)	2 (100%)	0	0	100	100
All	All	241/245 (98%)	232 (96%)	8 (3%)	1 (0%)	34	24

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	96	ASP

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	196/197 (100%)	189 (96%)	7 (4%)	35	26
2	B	3/3 (100%)	1 (33%)	2 (67%)	0	0
All	All	199/200 (100%)	190 (96%)	9 (4%)	27	18

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	36	LEU
1	A	86	LYS
1	A	96	ASP
1	A	121	ILE
1	A	150	THR
1	A	231	LEU
1	A	240	GLU
2	B	7	LEU
2	B	8	ARG

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

Mol	Chain	Res	Type
1	A	50	ASN
1	A	192	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)

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	$\#\|Z\| > 2$
3	NT4	A	900	-	39,39,39	1.20	3 (7%)	46,54,54	1.86	10 (21%)

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
3	NT4	A	900	-	-	10/24/34/34	0/4/4/4

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	900	NT4	C13-C27	3.03	1.52	1.47
3	A	900	NT4	O2-C9	-2.61	1.34	1.39
3	A	900	NT4	C18-C26	2.36	1.51	1.47

All (10) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	900	NT4	C21-C22-C23	-5.37	105.38	111.53
3	A	900	NT4	C8-N1-C7	4.57	128.88	122.55
3	A	900	NT4	C9-O2-C2	4.15	128.51	118.80
3	A	900	NT4	C24-C23-C22	-3.54	106.61	110.28
3	A	900	NT4	O2-C2-C1	3.46	129.97	119.10
3	A	900	NT4	O2-C2-C3	-3.19	109.10	119.10
3	A	900	NT4	C25-C8-N1	2.99	116.76	110.56
3	A	900	NT4	C24-C25-C8	-2.34	108.89	111.48
3	A	900	NT4	C21-C8-N1	-2.29	105.83	110.56
3	A	900	NT4	O3-C4-C5	2.17	125.92	119.10

There are no chirality outliers.

All (10) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	900	NT4	C25-C8-N1-C7
3	A	900	NT4	C5-C6-C7-N1
3	A	900	NT4	C5-C6-C7-O1
3	A	900	NT4	C1-C6-C7-N1
3	A	900	NT4	C1-C6-C7-O1
3	A	900	NT4	C17-C18-C26-N4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
3	A	900	NT4	C19-C18-C26-N4
3	A	900	NT4	C14-C13-C27-N6
3	A	900	NT4	C17-C18-C26-N3
3	A	900	NT4	C19-C18-C26-N3

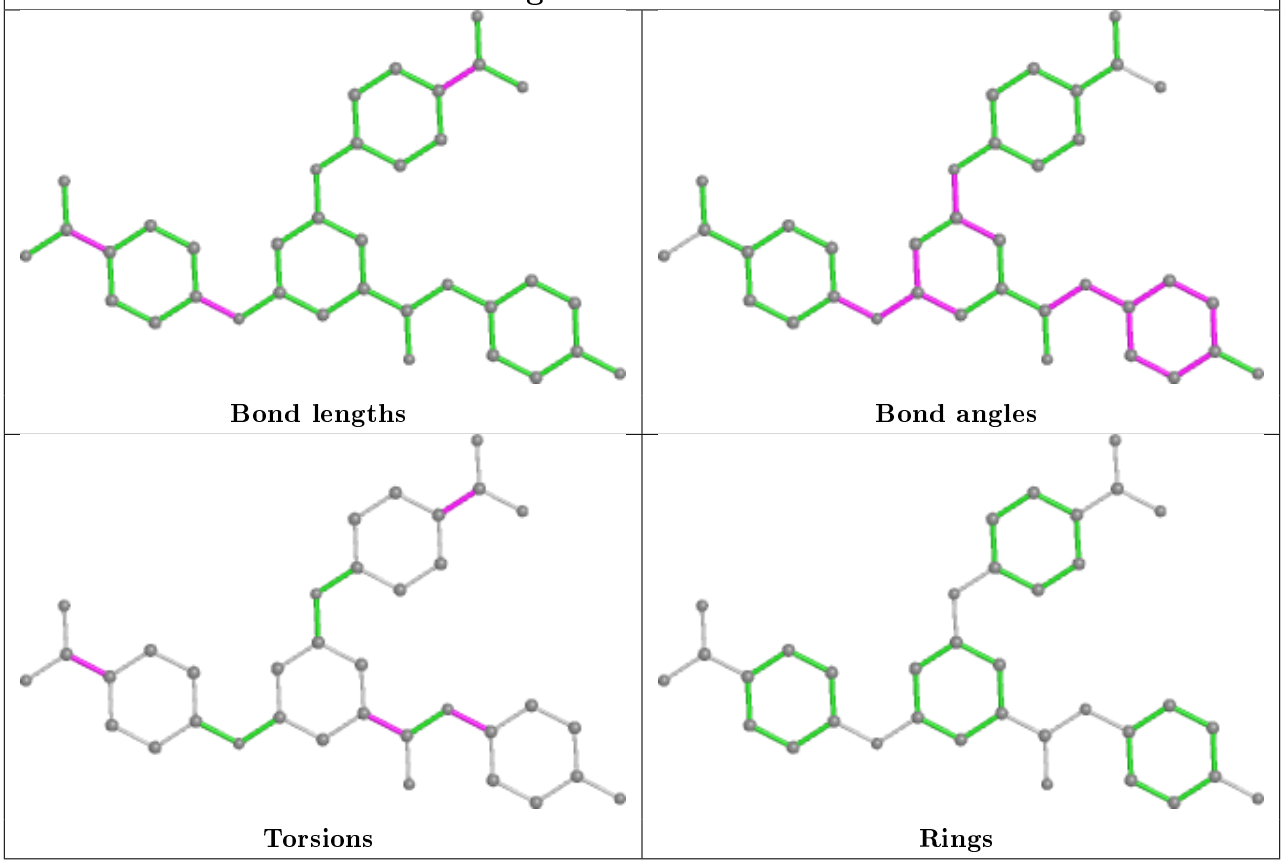
There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	900	NT4	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 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 NT4 A 900



5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	A	2

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	60(C):ARG	C	60(D):GLY	N	1.70
1	A	148:GLY	C	150:THR	N	1.70

6 Fit of model and data (i)

6.1 Protein, DNA and RNA chains (i)

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å^2)	Q<0.9
1	A	241/241 (100%)	-0.19	0 100 100	6, 10, 18, 24	0
2	B	4/4 (100%)	1.89	2 (50%) 0 0	13, 15, 31, 48	0
All	All	245/245 (100%)	-0.15	2 (0%) 86 87	6, 10, 18, 48	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	7	LEU	5.2
2	B	8	ARG	3.5

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.

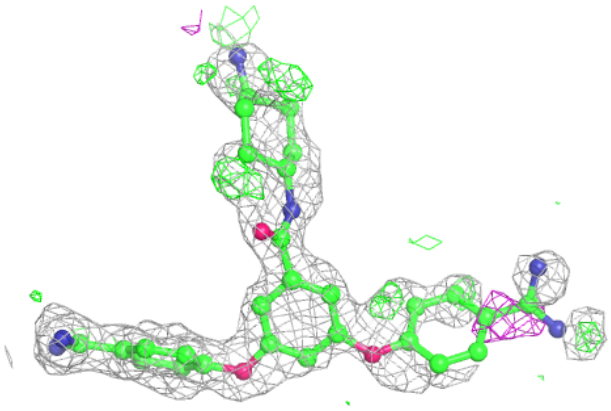
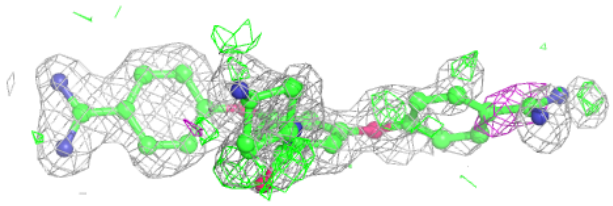
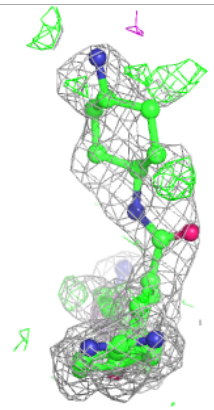
6.4 Ligands (i)

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. 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
3	NT4	A	900	36/36	0.79	0.22	11,31,59,61	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 NT4 A 900: $2mF_o$-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 (i)

There are no such residues in this entry.