

wwPDB X-ray Structure Validation Summary Report (i)

May 15, 2020 – 12:25 am BST

PDB ID : 2YJD

Title : Stapled peptide bound to Estrogen Receptor Beta

Authors: Phillips, C.; Roberts, L.R.; Schade, M.; Bent, A.; Davies, N.L.; Moore, R.;

Pannifer, A.D.; Brown, D.G.; Pickford, A.R.; Irving, S.L.

Deposited on : 2011-05-19

Resolution : 1.93 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
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 (1)) 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)

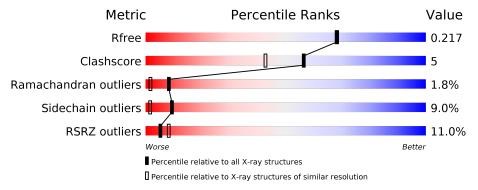
Validation Pipeline (wwPDB-VP) : 2.11

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 1.93 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	$\begin{array}{c} \text{Whole archive} \\ (\#\text{Entries}) \end{array}$	$\begin{array}{c} {\rm Similar \; resolution} \\ (\#{\rm Entries, \; resolution \; range(\AA)}) \end{array}$
R_{free}	130704	4310 (1.96-1.92)
Clashscore	141614	1023 (1.94-1.94)
Ramachandran outliers	138981	1007 (1.94-1.94)
Sidechain outliers	138945	1007 (1.94-1.94)
RSRZ outliers	127900	4250 (1.96-1.92)

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	240	8%	75%		18%	7%	
1	В	240	8%		15%	• 7%		
2	С	13	46% 31%		54%		15%	
2	D	13	23%	54%		23%		



2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 3960 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.

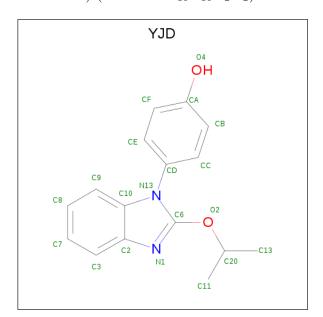
• Molecule 1 is a protein called ESTROGEN RECEPTOR BETA.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
1	A	223		C 1132			S 19	0	0	0
1	В	223		C 1132		O 312	S 19	0	0	0

• Molecule 2 is a protein called STAPLED PEPTIDE.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	С	13	Total	С	N	О	0	0	1
	C	10	97	64	17	16			
9	D	13	Total C N O	0	1				
2	D	Д 15	97	64	17	16	0	0	1

• Molecule 3 is 4-(2-PROPAN-2-YLOXYBENZIMIDAZOL-1-YL)PHENOL (three-letter code: YJD) (formula: C₁₆H₁₆N₂O₂).





Mol	Chain	Residues	${f Atoms}$			ZeroOcc	AltConf
3	A	1	Total 20			0	0
3	В	1	Total 20		N 2	0	0

• Molecule 4 is water.

Mol	Chain	Residues	${f Atoms}$	ZeroOcc	AltConf
4	A	103	Total O 103 103	0	0
4	В	98	Total O 98 98	0	0
4	С	5	Total O 5 5	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: ESTROGEN RECEPTOR BETA Chain A: 75% • Molecule 1: ESTROGEN RECEPTOR BETA Chain B: 15% • Molecule 2: STAPLED PEPTIDE 46% Chain C: 54% 15% • Molecule 2: STAPLED PEPTIDE 46% Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 4	Depositor
Cell constants	102.62Å 102.62Å 102.62Å	Depositor
a, b, c, α , β , γ	90.00° 90.00° 90.00°	Depositor
Resolution (Å)	13.03 - 1.93	Depositor
rtesolution (A)	12.83 - 1.93	EDS
% Data completeness	96.9 (13.03-1.93)	Depositor
(in resolution range)	96.9 (12.83-1.93)	EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) > 1$	$1.57 \; ({ m at} \; 1.93 { m \AA})$	Xtriage
Refinement program	BUSTER 2.9.6	Depositor
R, R_{free}	0.225 , 0.260	Depositor
It, It free	0.196 , 0.217	DCC
R_{free} test set	1944 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å ²)	29.1	Xtriage
Anisotropy	0.334	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$, $B_{sol}(Å^2)$	0.36 , 34.4	EDS
L-test for twinning ²	$< L >=0.43, < L^2>=0.26$	Xtriage
	0.059 for l,-k,h	
	0.067 for -l,-k,-h	
Estimated twinning fraction	0.057 for -h,-l,-k	Xtriage
	0.057 for -h,l,k	
	0.294 for -k,-h,-l	
F_o, F_c correlation	0.95	EDS
Total number of atoms	3960	wwPDB-VP
Average B, all atoms (Å ²)	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 3.93% of the height of the origin peak. No significant pseudotranslation is detected.

²Theoretical 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.



¹Intensities estimated from amplitudes.

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: NH2, YJD, ACE, MK8

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		
MIOI		RMSZ	# Z >5	RMSZ	# Z > 5	
1	A	0.48	0/1791	0.65	0/2417	
1	В	0.48	0/1791	0.63	0/2417	
2	С	1.17	0/75	1.70	1/98 (1.0%)	
2	D	1.20	0/75	1.75	1/98 (1.0%)	
All	All	0.53	0/3732	0.72	$2/5030 \ (0.0\%)$	

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 maintain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
2	С	0	1
2	D	0	3
All	All	0	4

There are no bond length outliers.

All (2) bond angle outliers are listed below:

\mathbf{Mol}	Chain	Res	Type	Atoms	Z	$\mathbf{Observed}(^o)$	$\mathbf{Ideal}(^o)$
2	С	3	ILE	CB-CA-C	7.33	126.27	111.60
2	D	5	HIS	N-CA-CB	5.38	120.27	110.60

There are no chirality outliers.

All (4) planarity outliers are listed below:

\mathbf{Mol}	Chain	Res	Type	Group
2	С	10	ASP	Peptide
2	D	10	ASP	Peptide

Continued on next page...



Continued from previous page...

Mol	Chain	Res	Type	Group
2	D	8	LEU	Peptide
2	D	9	GLN	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	$\mathbf{H}(\mathbf{model})$	H(added)	Clashes	Symm-Clashes
1	Α	1760	0	1842	20	0
1	В	1760	0	1842	18	0
2	С	97	0	98	2	0
2	D	97	0	98	5	0
3	A	20	0	16	1	0
3	В	20	0	15	0	0
4	A	103	0	0	0	0
4	В	98	0	0	0	0
4	С	5	0	0	0	0
All	All	3960	0	3911	38	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 5.

The worst 5 of 38 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	$egin{aligned} ext{Interatomic} \ ext{distance} \ (ext{Å}) \end{aligned}$	$egin{array}{c} ext{Clash} \ ext{overlap } (ext{Å}) \end{array}$
1:A:470:ASN:HD22	1:B:470:ASN:HD22	1.07	0.98
1:A:470:ASN:ND2	1:B:470:ASN:HD22	1.83	0.77
2:D:2:MK8:HG	2:D:6:MK8:HD	1.81	0.62
2:C:2:MK8:HG	2:C:6:MK8:HD	1.82	0.62
1:A:470:ASN:HD22	1:B:470:ASN:ND2	1.90	0.62

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	219/240 (91%)	210 (96%)	8 (4%)	1 (0%)	29 17
1	В	219/240 (91%)	206 (94%)	10 (5%)	3 (1%)	11 3
2	С	9/13 (69%)	7 (78%)	0	2 (22%)	0 0
2	D	9/13~(69%)	6 (67%)	1 (11%)	2 (22%)	0 0
All	All	$456/506 \ (90\%)$	429 (94%)	19 (4%)	8 (2%)	8 2

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	С	1	HIS
2	С	11	SER
2	D	1	HIS
2	D	11	SER
1	A	489	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	201/215~(94%)	184 (92%)	17 (8%)	10 2		
1	В	$201/215 \; (94\%)$	186 (92%)	15 (8%)	13 3		
2	С	9/9 (100%)	5 (56%)	4 (44%)	0 0		
2	D	9/9 (100%)	7 (78%)	2 (22%)	1 0		
All	All	420/448 (94%)	382 (91%)	38 (9%)	9 1		



5	of 38	residues	with a	non-rotamer	ic sidechair	are listed	below.

Mol	Chain	Res	Type
1	В	279	HIS
1	В	364	ARG
2	С	9	GLN
1	В	324	LEU
1	В	373	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	В	431	ASN
1	В	470	ASN
1	В	450	GLN
1	A	431	ASN
1	В	457	ASN

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	Mol Type	Chain	Res	Link	Bond lengths			Bond angles		
10101	туре	Chain	nes	LIIIK	Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
2	MK8	D	6	2	5,8,9	0.52	0	4,10,12	5.22	2 (50%)
2	MK8	С	2	2	5,8,9	1.61	1 (20%)	4,10,12	0.56	0
2	MK8	D	2	2	5,8,9	0.75	0	4,10,12	1.54	1 (25%)
2	MK8	С	6	2	5,8,9	0.49	0	4,10,12	4.95	3 (75%)

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	\mathbf{Type}	Chain	Res	Link	Chirals	Torsions	Rings
2	MK8	D	6	2	-	1/6/8/11	-
2	MK8	С	2	2	-	4/6/8/11	_
2	MK8	D	2	2	-	3/6/8/11	-
2	MK8	С	6	2	-	0/6/8/11	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	${f Observed(\AA)}$	$\mathbf{Ideal}(\mathbf{\AA})$
2	С	2	MK8	CB-CA	3.05	1.58	1.55

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	\mathbf{Z}	$Observed(^o)$	$\operatorname{Ideal}({}^o)$
2	D	6	MK8	CB-CG-CD	9.47	142.42	113.14
2	С	6	MK8	CB-CG-CD	8.83	140.42	113.14
2	D	6	MK8	CE-CD-CG	3.92	143.20	113.42
2	С	6	MK8	CE-CD-CG	3.68	141.39	113.42
2	D	2	MK8	CB1-CA-CB	-2.91	105.88	111.00

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	С	2	MK8	C-CA-CB-CG
2	С	2	MK8	N-CA-CB-CG
2	С	2	MK8	CB1-CA-CB-CG
2	D	2	MK8	N-CA-CB-CG
2	D	2	MK8	CB1-CA-CB-CG

There are no ring outliers.

4 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	D	6	MK8	2	0
2	С	2	MK8	2	0
2	D	2	MK8	3	0
2	С	6	MK8	1	0



5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

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	Chain	Res	$_{ m Res} \mid_{ m Link} \mid$	Во	Bond lengths			Bond angles		
MIOI	Mol Type Chain I	nes	LIIIK	Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2			
3	YJD	A	1498	-	21,22,22	1.66	4 (19%)	23,31,31	2.54	4 (17%)		
3	YJD	В	1498	-	21,22,22	1.62	5 (23%)	23,31,31	2.11	3 (13%)		

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	${ m Res}$	Link	Chirals	Torsions	Rings
3	YJD	A	1498	-	=	1/6/8/8	0/3/3/3
3	YJD	В	1498	-	-	1/6/8/8	0/3/3/3

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	\mathbf{Z}	$\operatorname{Observed}(\operatorname{\AA})$	$\operatorname{Ideal}(\operatorname{\AA})$
3	В	1498	YJD	CD-N13	-3.68	1.41	1.45
3	A	1498	YJD	CE-CD	3.32	1.44	1.38
3	A	1498	YJD	CC-CD	3.05	1.44	1.38
3	A	1498	YJD	CF-CA	3.05	1.44	1.38
3	В	1498	YJD	CE-CD	2.81	1.43	1.38

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	${f Atoms}$	\mathbf{Z}	$\mathbf{Observed}(^o)$	$\mathbf{Ideal}(^o)$
3	A	1498	YJD	C6-O2-C20	10.44	129.70	119.14
3	В	1498	YJD	C6-O2-C20	8.12	127.36	119.14

Continued on next page...



Continued from previous page...

Mol	Chain	Res	Type	$\Gamma ext{ype} A ext{toms} $		$\mathbf{Observed}(^o)$	$\operatorname{Ideal}({}^o)$
3	В	1498	YJD	O2-C20-C13	4.02	122.01	107.93
3	A	1498	YJD	O2-C20-C13	3.37	119.73	107.93
3	A	1498	YJD	CE-CD-N13	3.09	122.11	119.24

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	1498	YJD	C13-C20-O2-C6
3	В	1498	YJD	C13-C20-O2-C6

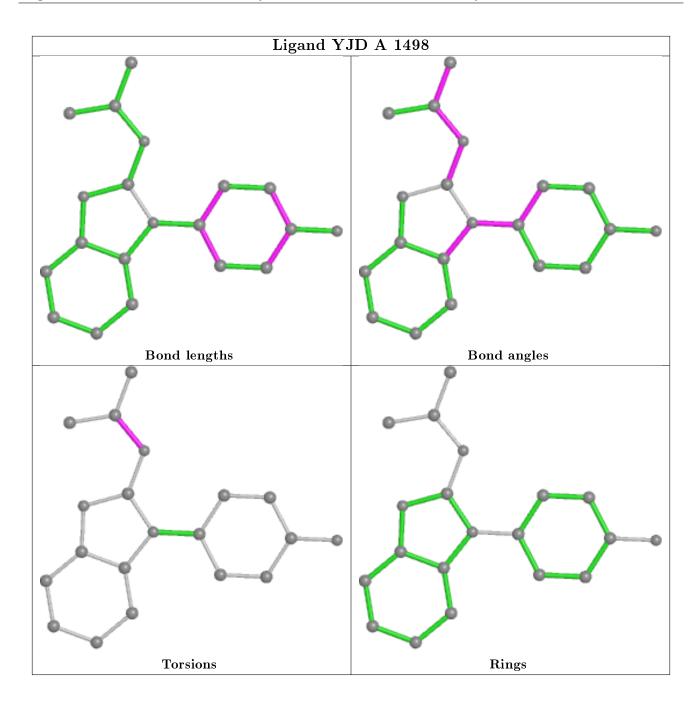
There are no ring outliers.

1 monomer is involved in 1 short contact:

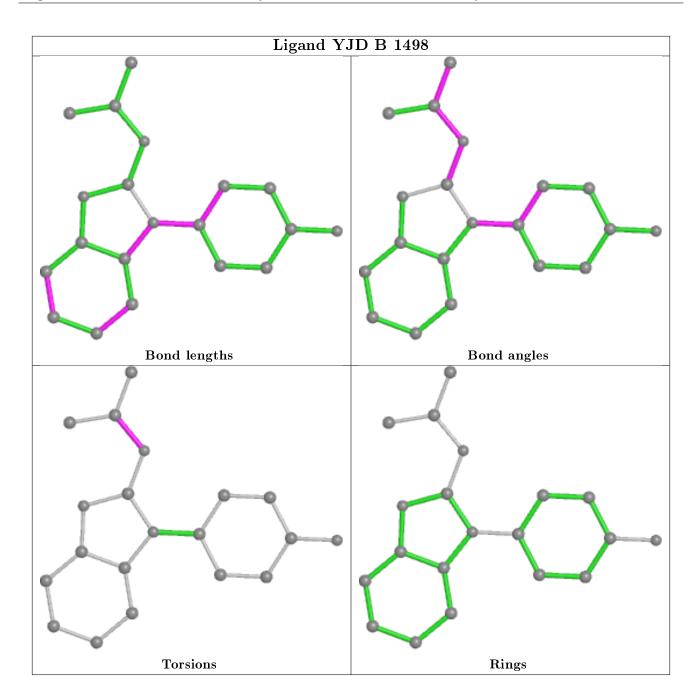
Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	1498	YJD	1	0

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.









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 (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 $>$	$\#\mathrm{RSRZ}{>}2$	$OWAB(Å^2)$	Q < 0.9
1	A	223/240 (92%)	0.68	20 (8%) 9 14	18, 30, 59, 79	0
1	В	223/240 (92%)	0.63	19 (8%) 10 16	18, 30, 59, 83	0
2	С	9/13 (69%)	3.83	6 (66%) 0 0	36, 49, 85, 89	0
2	D	9/13 (69%)	3.33	6 (66%) 0 0	38, 46, 86, 92	0
All	All	464/506 (91%)	0.77	51 (10%) 5 8	18, 30, 62, 92	0

The worst 5 of 51 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	С	11	SER	11.6
1	В	487	VAL	11.0
1	В	285	PRO	10.3
1	A	488	TYR	9.8
1	В	497	ALA	9.0

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^{th} 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	$\mathbf{B} ext{-}\mathbf{factors}(\mathbf{\mathring{A}}^2)$	Q < 0.9
2	MK8	D	6	9/10	0.86	0.15	42,43,45,45	0
2	MK8	С	2	9/10	0.88	0.16	34,36,37,38	0
2	MK8	D	2	9/10	0.89	0.17	44,49,50,52	0
2	MK8	С	6	9/10	0.92	0.15	33,38,39,40	0



6.3 Carbohydrates (i)

There are no carbohydrates in this entry.

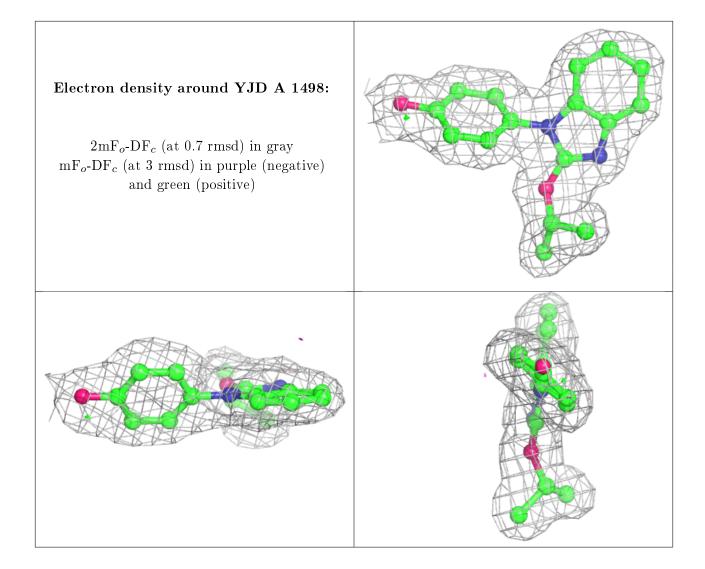
6.4 Ligands (i)

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, 95^{th} 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.

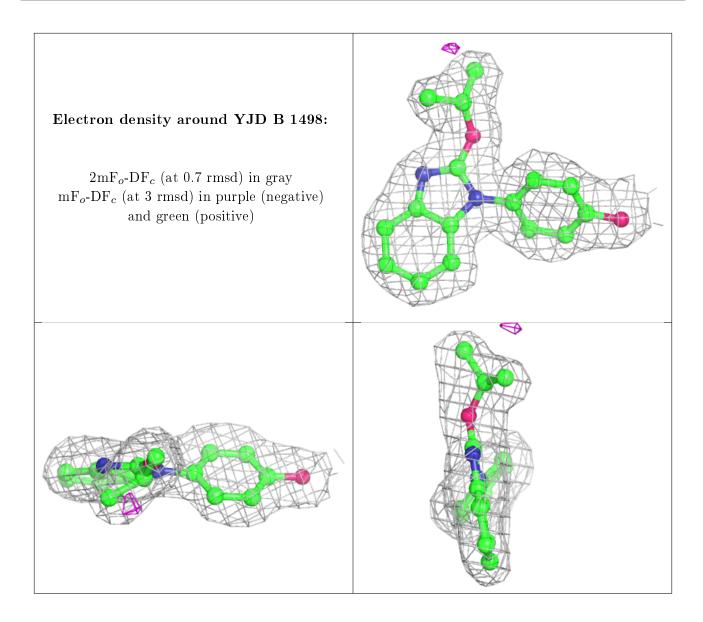
Mol	Type	Chain	Res	Atoms	RSCC	RSR	${f B-factors}({f \AA}^2)$	Q<0.9
3	YJD	A	1498	20/20	0.95	0.11	19,26,28,28	0
3	YJD	В	1498	20/20	0.95	0.12	22,25,28,28	0

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.

