



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 13, 2017 – 07:27 PM EDT

PDB ID : 2GBF  
Title : rat dpp-IV with alkynyl cyanopyrrolidine #1  
Authors : Longenecker, K.L.; Jakob, C.G.; Fry, E.H.; Wilk, S.  
Deposited on : unknown  
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 : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20030345  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
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) : rb-20030345

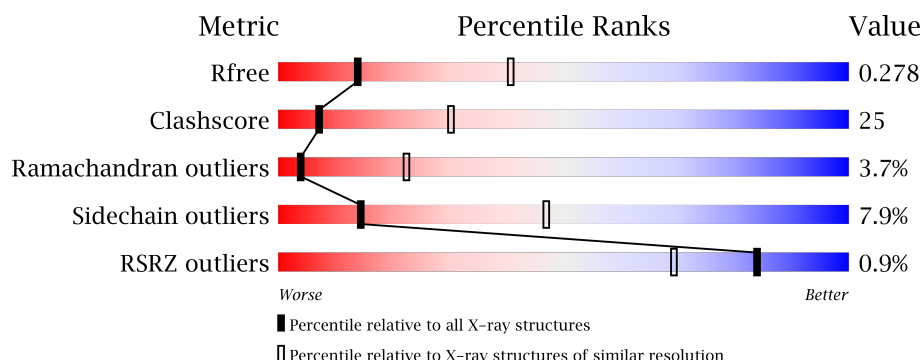
# 1 Overall quality at a glance

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))
$R_{free}$	100719	1001 (3.12-3.08)
Clashscore	112137	1099 (3.12-3.08)
Ramachandran outliers	110173	1057 (3.12-3.08)
Sidechain outliers	110143	1057 (3.12-3.08)
RSRZ outliers	101464	1006 (3.12-3.08)

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	730	<div> <div style="width: 100%; height: 10px; position: relative;"> <div style="position: absolute; top: -10px; left: 0; width: 100%; text-align: center;">%</div> <div style="position: absolute; top: 0; left: 0; width: 54%; height: 10px; background-color: green;"></div> <div style="position: absolute; top: 0; left: 54%; width: 40%; height: 10px; background-color: yellow;"></div> <div style="position: absolute; top: 0; left: 94%; width: 6%; height: 10px; background-color: orange;"></div> <div style="position: absolute; top: 0; left: 94%; width: 6%; height: 10px; background-color: red;"></div> <div style="position: absolute; top: 0; left: 0; width: 100%; text-align: center;">54% 40% 6% .</div> </div> </div>
1	B	730	<div> <div style="width: 100%; height: 10px; position: relative;"> <div style="position: absolute; top: -10px; left: 0; width: 100%; text-align: center;">%</div> <div style="position: absolute; top: 0; left: 0; width: 54%; height: 10px; background-color: green;"></div> <div style="position: absolute; top: 0; left: 54%; width: 39%; height: 10px; background-color: yellow;"></div> <div style="position: absolute; top: 0; left: 93%; width: 6%; height: 10px; background-color: orange;"></div> <div style="position: absolute; top: 0; left: 93%; width: 6%; height: 10px; background-color: red;"></div> <div style="position: absolute; top: 0; left: 0; width: 100%; text-align: center;">54% 39% 6% .</div> </div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	AIA	A	768	-	-	-	X

## 2 Entry composition [i](#)

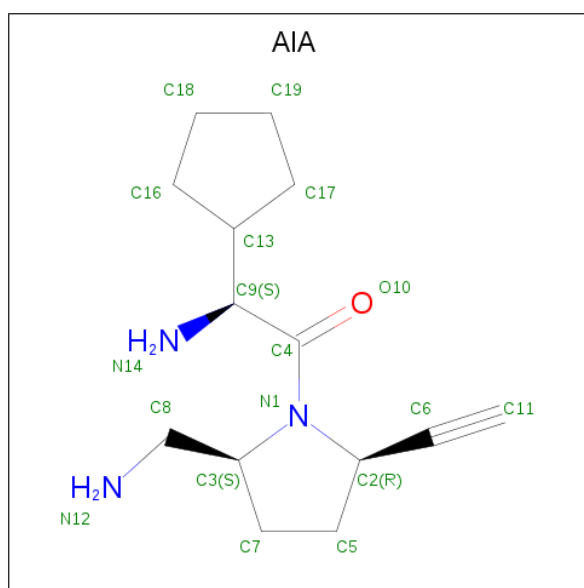
There are 2 unique types of molecules in this entry. The entry contains 11858 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Dipeptidyl peptidase 4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	730	Total	C	N	O	S	0	0	0
			5920	3789	981	1124	26			
1	B	730	Total	C	N	O	S	0	0	0
			5920	3789	981	1124	26			

- Molecule 2 is (1S)-2-[(2S,5R)-2-(AMINOMETHYL)-5-ETHYNYLPYRROLIDIN-1-YL]-1-CYCLOPENTYL-2-OXOETHANAMINE (three-letter code: AIA) (formula: C<sub>14</sub>H<sub>23</sub>N<sub>3</sub>O).

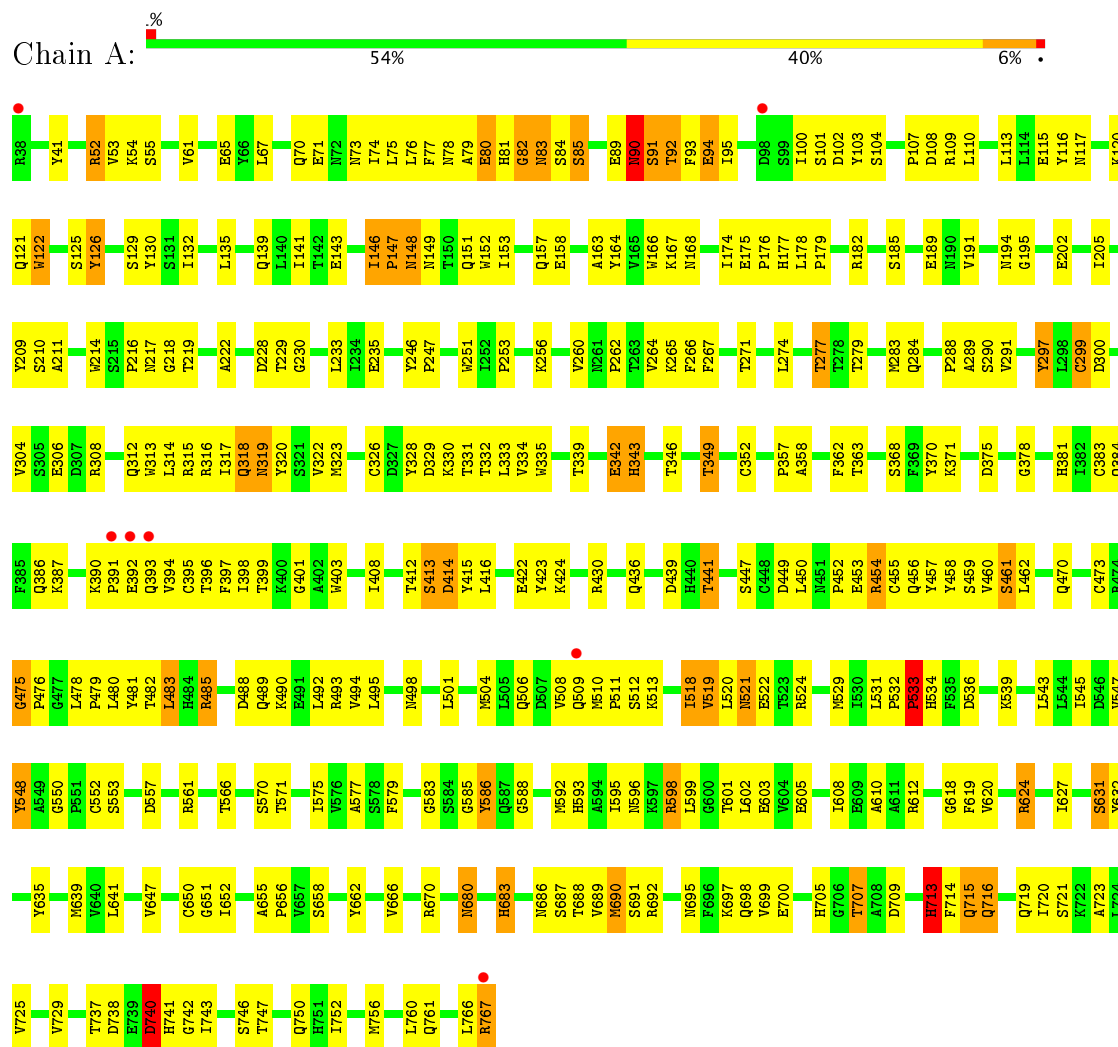


Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	0	0
			18	14	3	1		

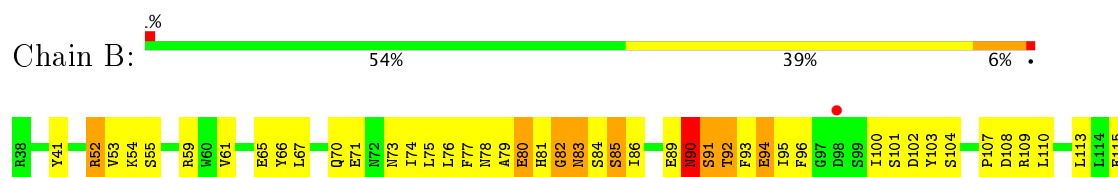
### 3 Residue-property plots [i](#)

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: Dipeptidyl peptidase 4



#### • Molecule 1: Dipeptidyl peptidase 4



Q719	K623	I544	Q384	L298	G195	Y116
I720	R624	I545	F385	C299	E202	N117
S721	D546	D546	K386	D300	K120	K120
K722	I627	V547	G475	K387	K121	W121
A723	I627	Y548	P476	K390	I205	W122
D726	Y632	A549	K390	E306	Y209	S125
V729	Y635	G550	P391	D307	S210	Y126
D738	Y639	F551	E392	R308	A211	S129
E739	M639	S553	T482	Q112	W214	S215
D740	V640	D557	L483	C395	S215	P216
H741	L641	R561	H484	T396	P216	S131
G742	V647	T487	R485	F397	N217	I132
I743	G650	R566	S486	I398	L135	L135
I752	I652	T566	K400	T399	T219	L135
M756	G651	S570	G401	K400	A222	Q139
Q761	I652	T571	A401	N319	T228	T142
L766	A655	I575	W403	Y320	T229	E143
R767	S658	V576	L408	S321	G230	E144
	R659	A577	T412	K323	K145	K145
	Y662	S578	D414	C326	L233	I146
	D664	F579	Y415	D327	I234	P147
	S665	G583	L416	Y328	E235	N149
	V666	S584	E422	D329	W251	T150
	R670	G585	Y423	T331	I251	Q151
		Y586	Q506	T332	P253	W152
		G587	K424	L333	K256	I153
		G588	R430	V334		Q157
		M592	V508	W335	V260	E158
		H593	Q509	T339	N261	K161
		A594	Q436	E342	P262	L162
		I595	D439	H243	T263	A163
		I596	H440		K265	Y164
		R597	T441	T346	F266	V165
		R598	K444	T349	I267	W166
		L599			I268	K167
		G600	S447	P357		N168
		M601	C448	A358	T271	I174
		L602	D449	L450	L274	E175
		V604	N451	T363	T277	P176
		E605	P452	S368	L178	H177
		L608	E453	F369	T278	L178
		R609	R454	V370	T279	P179
		A610	C455	K371	M283	R182
		G611	Q456	Y457	Q284	E189
		R612	Y458	D375	I285	H190
		L615	F535	G378	A289	S290
		K616	S459	V460	S290	V291
		H617	S461	L462		Y191
		G618	L462	C383		N194
		F619	Q470			
		V620				

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 3	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	208.17Å 208.17Å 208.17Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 3.10 19.94 – 3.11	Depositor EDS
% Data completeness (in resolution range)	98.8 (20.00-3.10) 99.3 (19.94-3.11)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.57 (at 3.09Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.250 , 0.289 0.243 , 0.278	Depositor DCC
$R_{free}$ test set	2743 reflections (5.10%)	DCC
Wilson B-factor (Å <sup>2</sup> )	73.1	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 59.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.54$ , $\langle L^2 \rangle = 0.38$	Xtriage
Estimated twinning fraction	0.000 for l,-k,h	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	11858	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	65.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.78% 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, 0.333 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: AIA

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	0.46	2/6088 (0.0%)	0.71	2/8278 (0.0%)
1	B	0.43	0/6088	0.70	2/8278 (0.0%)
All	All	0.45	2/12176 (0.0%)	0.71	4/16556 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	631	SER	C-O	8.88	1.40	1.23
1	A	352	CYS	CB-SG	-5.14	1.73	1.81

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	713	HIS	N-CA-C	7.29	130.67	111.00
1	A	713	HIS	N-CA-C	7.22	130.50	111.00
1	A	90	ASN	N-CA-C	-5.15	97.10	111.00
1	B	90	ASN	N-CA-C	-5.12	97.18	111.00

There are no chirality outliers.

There are no planarity outliers.

### 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	5920	0	5631	294	0
1	B	5920	0	5632	292	0
2	A	18	0	21	0	0
All	All	11858	0	11284	584	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 25.

The worst 5 of 584 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:761:GLN:HB3	1:A:767:ARG:HB3	1.39	1.03
1:B:349:THR:HB	1:B:593:HIS:HD2	1.26	0.99
1:A:460:VAL:HG22	1:A:461:SER:H	1.28	0.99
1:A:349:THR:HB	1:A:593:HIS:HD2	1.25	0.98
1:B:761:GLN:HB3	1:B:767:ARG:HB3	1.42	0.98

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	728/730 (100%)	614 (84%)	87 (12%)	27 (4%)	4	22
1	B	728/730 (100%)	615 (84%)	86 (12%)	27 (4%)	4	22
All	All	1456/1460 (100%)	1229 (84%)	173 (12%)	54 (4%)	4	22

5 of 54 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	80	GLU
1	A	143	GLU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	450	LEU
1	A	521	ASN
1	A	533	PRO

### 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	643/651 (99%)	594 (92%)	49 (8%)	15	49
1	B	643/651 (99%)	591 (92%)	52 (8%)	14	45
All	All	1286/1302 (99%)	1185 (92%)	101 (8%)	14	47

5 of 101 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	715	GLN
1	B	149	ASN
1	B	686	ASN
1	A	740	ASP
1	B	94	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 41 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	719	GLN
1	B	284	GLN
1	B	715	GLN
1	A	749	HIS
1	B	148	ASN

### 5.3.3 RNA ⓘ

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	$\# Z  > 2$
2	AIA	A	768	1	17,19,19	1.39	1 (5%)	14,26,26	1.33	2 (14%)

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	AIA	A	768	1	-	0/14/36/36	0/2/2/2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	768	AIA	C8-N12	-4.81	1.35	1.47

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	768	AIA	C7-C3-N1	2.36	104.32	101.72
2	A	768	AIA	C13-C9-C4	2.63	116.56	111.06

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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	730/730 (100%)	-0.37	7 (0%) 82 67	32, 58, 91, 165	0
1	B	730/730 (100%)	-0.25	6 (0%) 86 71	35, 65, 104, 161	0
All	All	1460/1460 (100%)	-0.31	13 (0%) 84 69	32, 61, 99, 165	0

The worst 5 of 13 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	767	ARG	6.9
1	A	393	GLN	3.0
1	B	392	GLU	2.9
1	B	393	GLN	2.8
1	A	98	ASP	2.5

### 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. 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	AIA	A	768	18/18	0.92	0.29	3.15	65,65,65,65	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.