



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 06:19 AM GMT

PDB ID : 2WPA  
Title : OPTIMISATION OF 6,6-DIMETHYL PYRROLO 3,4-C PYRAZOLES:  
IDENTIFICATION OF PHA-793887, A POTENT CDK INHIBITOR SUIT-  
ABLE FOR INTRAVENOUS DOSING  
Authors : Brasca, M.G.; Albanese, C.; Alzani, R.; Amici, R.; Avanzi, N.; Ballinari, D.;  
Bischoff, J.; Borghi, D.; Casale, E.; Croci, V.; Fiorentini, F.; Isacchi, A.;  
Mercurio, C.; Nesi, M.; Orsini, P.; Pastori, W.; Pesenti, E.; Pevarello, P.;  
Roussel, P.; Varasi, M.; Volpi, D.; Vulpetti, A.; Ciomei, M.  
Deposited on : 2009-08-03  
Resolution : 2.51 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

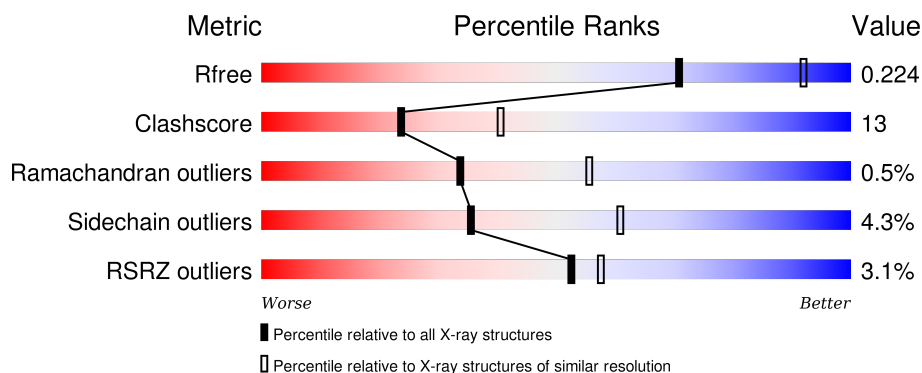
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.51 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	3553 (2.50-2.50)
Clashscore	102246	4242 (2.50-2.50)
Ramachandran outliers	100387	4156 (2.50-2.50)
Sidechain outliers	100360	4158 (2.50-2.50)
RSRZ outliers	91569	3562 (2.50-2.50)

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	309	<div> <div>2%</div> <div> <div></div> <div>75%</div> <div>20%</div> <div>• •</div> </div> </div>
1	C	309	<div> <div>5%</div> <div> <div></div> <div>65%</div> <div>28%</div> <div>• •</div> </div> </div>
2	B	265	<div> <div>3%</div> <div> <div></div> <div>75%</div> <div>21%</div> <div>• •</div> </div> </div>
2	D	265	<div> <div>2%</div> <div> <div></div> <div>73%</div> <div>20%</div> <div>• •</div> </div> </div>

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 9253 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 CELL DIVISION PROTEIN KINASE 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	302	Total	C	N	O	S	0	0	0
			2427	1579	412	428	8			
1	C	299	Total	C	N	O	S	0	0	0
			2405	1563	409	425	8			

- Molecule 2 is a protein called CYCLIN A2.

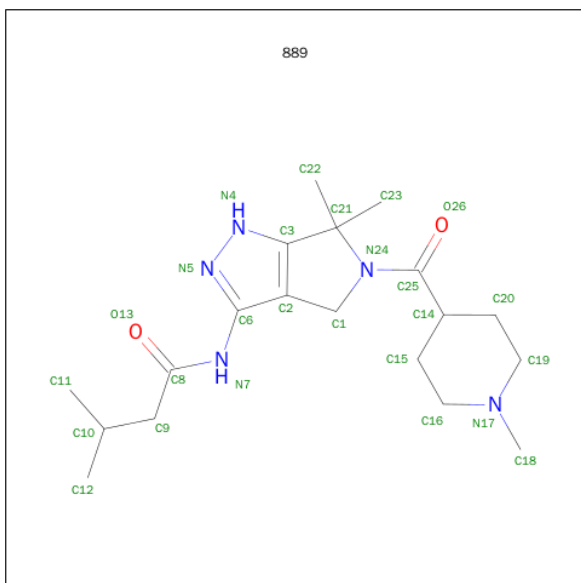
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	258	Total	C	N	O	S	0	0	0
			2084	1350	339	384	11			
2	D	254	Total	C	N	O	S	0	0	0
			2050	1327	335	377	11			

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total O S 5 4 1	0	0
3	D	1	Total O S 5 4 1	0	0

- Molecule 4 is N-{6,6-DIMETHYL-5-[(1-METHYLPIPERIDIN-4-YL)CARBONYL]-1,4,5,6-TETRAHYDROPYRROLO[3,4-C]PYRAZOL-3-YL}-3-METHYLBUTANAMIDE (three-letter code: 889) (formula: C<sub>19</sub>H<sub>31</sub>N<sub>5</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C N O 26 19 5 2	0	0
4	C	1	Total C N O 26 19 5 2	0	0

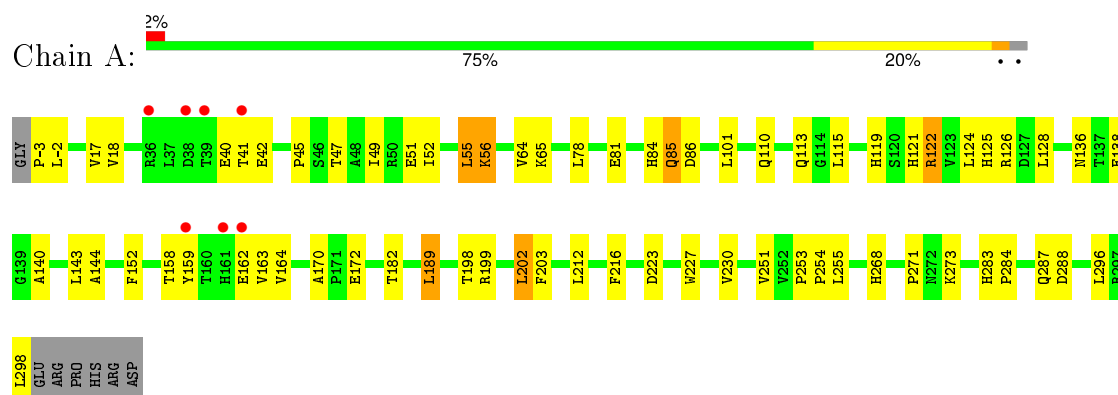
- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	53	Total O 53 53	0	0
5	B	47	Total O 47 47	0	0
5	C	29	Total O 29 29	0	0
5	D	96	Total O 96 96	0	0

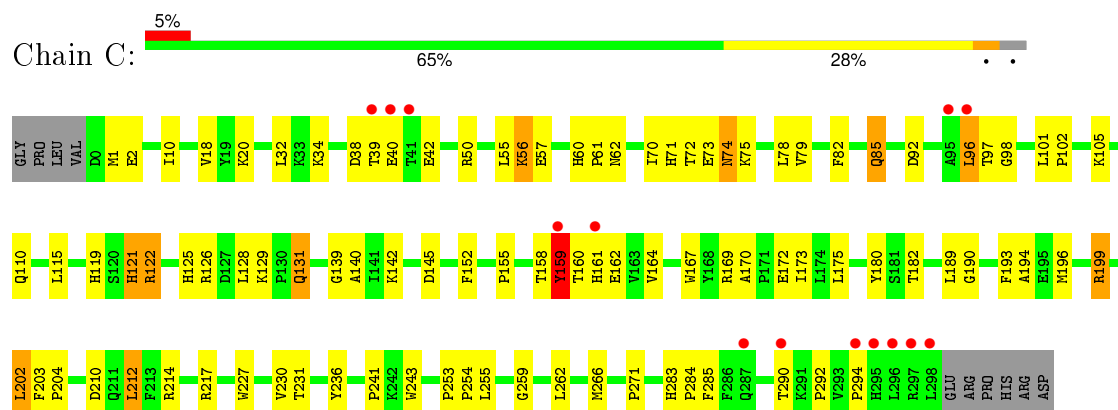
### 3 Residue-property plots [i](#)

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

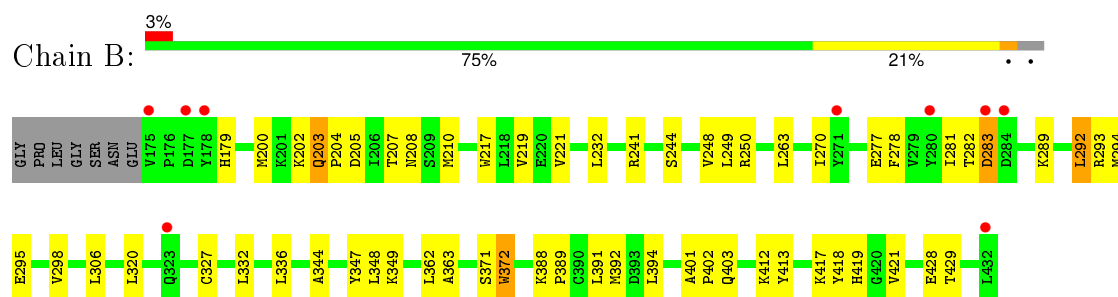
#### • Molecule 1: CELL DIVISION PROTEIN KINASE 2



#### • Molecule 1: CELL DIVISION PROTEIN KINASE 2



#### • Molecule 2: CYCLIN A2



#### • Molecule 2: CYCLIN A2

L282	R293	A308	P309	T316	Q323	P324	M334	E338	A344	R345	P346	Y347	L348	K349	H361	S371	W372	P373	E374	R378	K379	K388	P389	C390	L391	Q396	A401	P402	I409	K414	Y418	H419	G420	V421	L432	L287	K288	K289	Q290	W291								
GLY	PRO	LEU	GLY	SER	ASN	GLU	VAL	PRO	ASP	TYR	H179	E180	I182	H183	T184	Y185	E188	K194	P195	K196	V197	Q203	P204	I205	I206	M208	S209	M210	R241	S244	S247	Y248	L249	R250	P273	E274	V275	A276	E277	F278	V279	Y280	D283	T287	K288	K289	Q290	W291

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 6 <sub>2</sub> 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	186.53Å 186.53Å 215.13Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	29.87 – 2.51 48.15 – 2.51	Depositor EDS
% Data completeness (in resolution range)	99.0 (29.87-2.51) 99.1 (48.15-2.51)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.12 (at 2.51Å)	Xtriage
Refinement program	CNX	Depositor
R, $R_{free}$	0.221 , 0.261 0.218 , 0.224	Depositor DCC
$R_{free}$ test set	3771 reflections (5.29%)	DCC
Wilson B-factor (Å <sup>2</sup> )	49.1	Xtriage
Anisotropy	0.250	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 52.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 75161 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	9253	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	53.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.96% 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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: 889, SO4

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	0.44	0/2490	0.55	0/3381
1	C	0.35	0/2467	0.50	0/3349
2	B	0.37	0/2134	0.48	0/2897
2	D	0.43	0/2098	0.52	0/2846
All	All	0.40	0/9189	0.52	0/12473

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
2	D	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	D	206	ILE	Peptide

### 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	2427	0	2481	60	0
1	C	2405	0	2454	82	0
2	B	2084	0	2107	40	0
2	D	2050	0	2078	67	0
3	A	5	0	0	0	0
3	D	5	0	0	0	0
4	A	26	0	31	2	0
4	C	26	0	31	3	0
5	A	53	0	0	1	0
5	B	47	0	0	3	0
5	C	29	0	0	1	0
5	D	96	0	0	6	1
All	All	9253	0	9182	240	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 13.

The worst 5 of 240 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)
2:D:287:THR:HG22	2:D:290:GLN:H	1.16	1.09
2:D:334:MET:HG2	5:D:2047:HOH:O	1.61	0.97
1:A:227:TRP:O	1:A:230:VAL:HG23	1.66	0.95
1:A:172:GLU:HG2	1:A:271:PRO:HG3	1.53	0.90
1:A:85:GLN:HG3	1:A:86:ASP:N	1.87	0.89

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)
5:D:2006:HOH:O	5:D:2006:HOH:O[10_775]	2.09	0.11

## 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	300/309 (97%)	295 (98%)	5 (2%)	0	100	100
1	C	297/309 (96%)	285 (96%)	8 (3%)	4 (1%)	15	26
2	B	256/265 (97%)	253 (99%)	2 (1%)	1 (0%)	39	61
2	D	252/265 (95%)	249 (99%)	3 (1%)	0	100	100
All	All	1105/1148 (96%)	1082 (98%)	18 (2%)	5 (0%)	34	55

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	159	TYR
1	C	160	THR
1	C	162	GLU
1	C	145	ASP
2	B	372	TRP

### 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	267/273 (98%)	255 (96%)	12 (4%)	34	59
1	C	264/273 (97%)	249 (94%)	15 (6%)	25	46
2	B	232/237 (98%)	225 (97%)	7 (3%)	48	76
2	D	228/237 (96%)	219 (96%)	9 (4%)	39	66
All	All	991/1020 (97%)	948 (96%)	43 (4%)	35	61

5 of 43 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	50	ARG
1	C	85	GLN
2	D	287	THR
1	C	55	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	C	56	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	296	HIS
2	B	312	ASN
2	D	208	ASN
2	B	208	ASN
2	D	396	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](#)

4 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	$\# Z  > 2$
3	SO4	A	1300	-	4,4,4	0.28	0	6,6,6	0.17	0
4	889	A	1301	-	27,28,28	2.39	7 (25%)	28,42,42	1.61	6 (21%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
4	889	C	1300	-	27,28,28	2.09	7 (25%)	28,42,42	1.89	7 (25%)
3	SO4	D	1433	-	4,4,4	0.24	0	6,6,6	0.14	0

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	SO4	A	1300	-	-	0/0/0/0	0/0/0/0
4	889	A	1301	-	-	0/14/42/42	0/3/3/3
4	889	C	1300	-	-	0/14/42/42	0/3/3/3
3	SO4	D	1433	-	-	0/0/0/0	0/0/0/0

The worst 5 of 14 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	1301	889	C1-C2	-3.95	1.45	1.50
4	C	1300	889	C21-C3	-2.90	1.48	1.52
4	A	1301	889	C6-N5	-2.88	1.30	1.34
4	C	1300	889	C2-C3	-2.54	1.34	1.39
4	C	1300	889	C1-C2	-2.48	1.47	1.50

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	C	1300	889	C6-C2-C3	-4.39	102.68	106.38
4	A	1301	889	C6-C2-C3	-4.37	102.69	106.38
4	C	1300	889	C14-C25-N24	-3.16	116.06	119.01
4	C	1300	889	C9-C8-N7	-3.14	110.82	114.48
4	A	1301	889	C15-C14-C25	2.04	113.49	109.69

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	1301	889	2	0
4	C	1300	889	3	0

## 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, 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	302/309 (97%)	-0.14	7 (2%) 64 67	30, 42, 77, 106	0
1	C	299/309 (96%)	0.20	14 (4%) 35 40	36, 62, 92, 128	0
2	B	258/265 (97%)	0.03	9 (3%) 48 53	38, 55, 77, 105	0
2	D	254/265 (95%)	-0.16	4 (1%) 74 78	29, 41, 67, 93	0
All	All	1113/1148 (96%)	-0.01	34 (3%) 52 57	29, 51, 84, 128	0

The worst 5 of 34 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	298	LEU	6.0
1	A	161	HIS	5.6
1	C	159	TYR	5.6
2	B	175	VAL	5.5
1	C	40	GLU	4.9

### 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
4	889	C	1300	26/26	0.92	0.19	0.78	49,54,71,72	0
4	889	A	1301	26/26	0.96	0.14	0.21	30,35,53,59	0
3	SO4	A	1300	5/5	0.86	0.19	-	124,124,125,125	0
3	SO4	D	1433	5/5	0.84	0.22	-	123,123,124,124	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.