



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 07:48 PM GMT

PDB ID : 4PV7  
Title : Cocrystal structure of dipeptidyl-peptidase 4 with an indole scaffold inhibitor  
Authors : Xiao, P.; Guo, R.; Huang, S.; Cui, H.; Ye, S.; Zhang, Z.  
Deposited on : 2014-03-15  
Resolution : 3.24 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

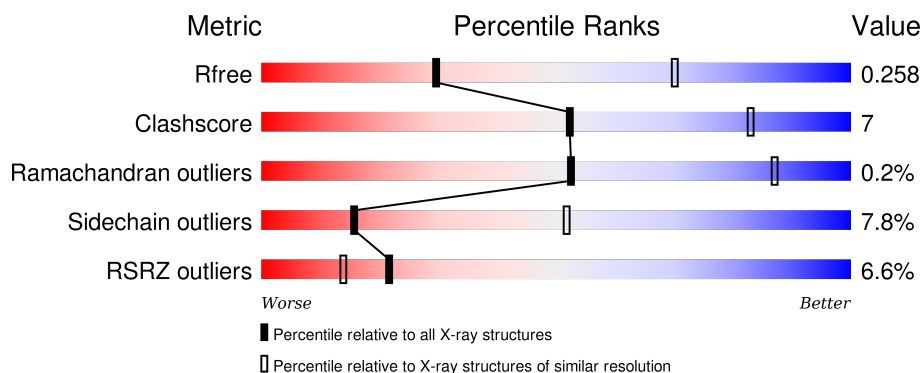
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.24 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	1092 (3.28-3.20)
Clashscore	102246	1227 (3.28-3.20)
Ramachandran outliers	100387	1204 (3.28-3.20)
Sidechain outliers	100360	1203 (3.28-3.20)
RSRZ outliers	91569	1097 (3.28-3.20)

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	752	<div> <div>6%</div> <div> <div></div> <div>74%</div> <div>20%</div> <div>• •</div> </div> </div>
1	B	752	<div> <div>7%</div> <div> <div></div> <div>76%</div> <div>19%</div> <div>• •</div> </div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 11948 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 soluble form.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	727	Total	C	N	O	S	0	0	0
			5951	3821	978	1126	26			
1	B	727	Total	C	N	O	S	0	0	0
			5951	3821	978	1126	26			

There are 48 discrepancies between the modelled and reference sequences:

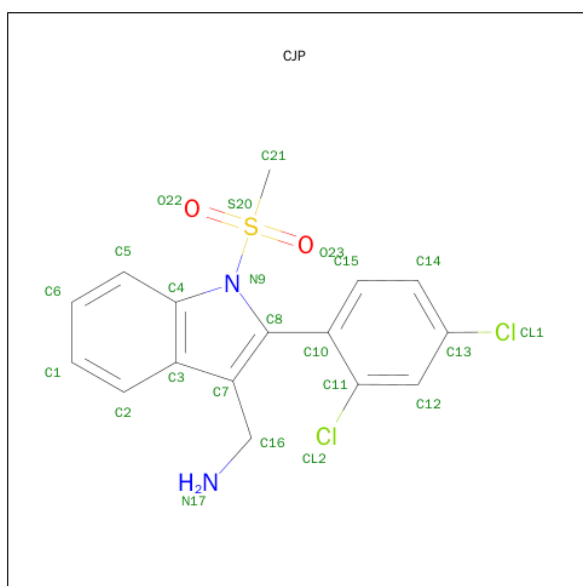
Chain	Residue	Modelled	Actual	Comment	Reference
A	15	HIS	-	EXPRESSION TAG	UNP P27487
A	16	HIS	-	EXPRESSION TAG	UNP P27487
A	17	HIS	-	EXPRESSION TAG	UNP P27487
A	18	HIS	-	EXPRESSION TAG	UNP P27487
A	19	HIS	-	EXPRESSION TAG	UNP P27487
A	20	HIS	-	EXPRESSION TAG	UNP P27487
A	21	ASP	-	EXPRESSION TAG	UNP P27487
A	22	TYR	-	EXPRESSION TAG	UNP P27487
A	23	ASP	-	EXPRESSION TAG	UNP P27487
A	24	ILE	-	EXPRESSION TAG	UNP P27487
A	25	PRO	-	EXPRESSION TAG	UNP P27487
A	26	THR	-	EXPRESSION TAG	UNP P27487
A	27	THR	-	EXPRESSION TAG	UNP P27487
A	28	GLU	-	EXPRESSION TAG	UNP P27487
A	29	ASN	-	EXPRESSION TAG	UNP P27487
A	30	LEU	-	EXPRESSION TAG	UNP P27487
A	31	TYR	-	EXPRESSION TAG	UNP P27487
A	32	PHE	-	EXPRESSION TAG	UNP P27487
A	33	GLN	-	EXPRESSION TAG	UNP P27487
A	34	GLY	-	EXPRESSION TAG	UNP P27487
A	35	ALA	-	EXPRESSION TAG	UNP P27487
A	36	MET	-	EXPRESSION TAG	UNP P27487
A	37	GLY	-	EXPRESSION TAG	UNP P27487
A	38	SER	-	EXPRESSION TAG	UNP P27487
B	15	HIS	-	EXPRESSION TAG	UNP P27487

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	16	HIS	-	EXPRESSION TAG	UNP P27487
B	17	HIS	-	EXPRESSION TAG	UNP P27487
B	18	HIS	-	EXPRESSION TAG	UNP P27487
B	19	HIS	-	EXPRESSION TAG	UNP P27487
B	20	HIS	-	EXPRESSION TAG	UNP P27487
B	21	ASP	-	EXPRESSION TAG	UNP P27487
B	22	TYR	-	EXPRESSION TAG	UNP P27487
B	23	ASP	-	EXPRESSION TAG	UNP P27487
B	24	ILE	-	EXPRESSION TAG	UNP P27487
B	25	PRO	-	EXPRESSION TAG	UNP P27487
B	26	THR	-	EXPRESSION TAG	UNP P27487
B	27	THR	-	EXPRESSION TAG	UNP P27487
B	28	GLU	-	EXPRESSION TAG	UNP P27487
B	29	ASN	-	EXPRESSION TAG	UNP P27487
B	30	LEU	-	EXPRESSION TAG	UNP P27487
B	31	TYR	-	EXPRESSION TAG	UNP P27487
B	32	PHE	-	EXPRESSION TAG	UNP P27487
B	33	GLN	-	EXPRESSION TAG	UNP P27487
B	34	GLY	-	EXPRESSION TAG	UNP P27487
B	35	ALA	-	EXPRESSION TAG	UNP P27487
B	36	MET	-	EXPRESSION TAG	UNP P27487
B	37	GLY	-	EXPRESSION TAG	UNP P27487
B	38	SER	-	EXPRESSION TAG	UNP P27487

- Molecule 2 is 1-[2-(2,4-DICHLOROPHENYL)-1-(METHYLSULFONYL)-1H-INDOL-3-YL] METHANAMINE (three-letter code: CJP) (formula: C<sub>16</sub>H<sub>14</sub>Cl<sub>2</sub>N<sub>2</sub>O<sub>2</sub>S).

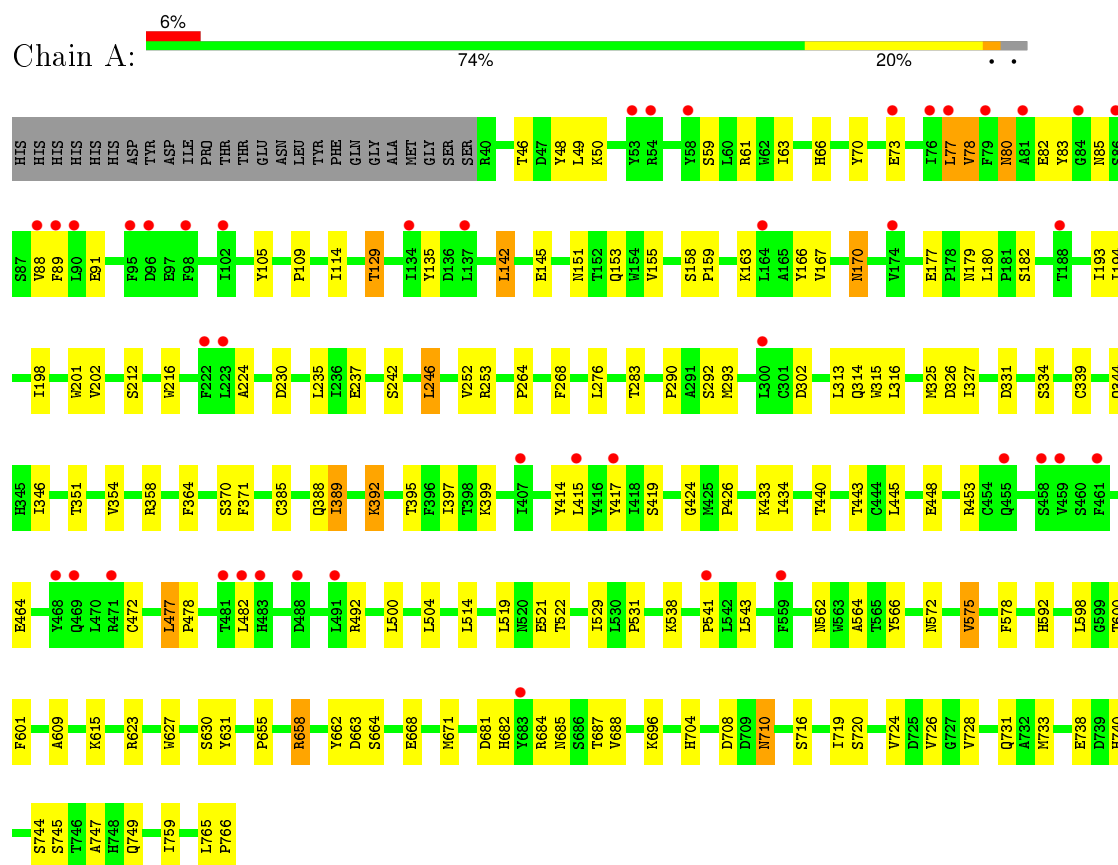


Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
2	A	1	Total	C	Cl	N	O	S	0	0
			23	16	2	2	2	1		
2	B	1	Total	C	Cl	N	O	S	0	0
			23	16	2	2	2	1		

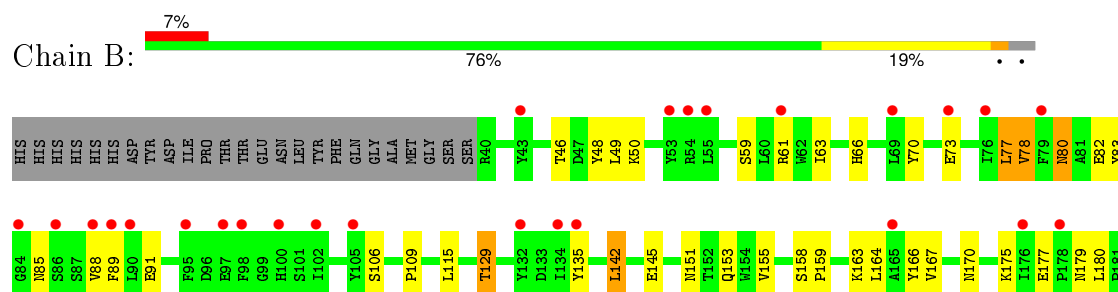
### 3 Residue-property plots

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: Dipeptidyl peptidase 4 soluble form



- Molecule 1: Dipeptidyl peptidase 4 soluble form



P766	I626	A465	I327	S182
	W627	K466	Y330	T188
	S630	Y467	D831	
	Y631	C472	S334	T198
	T636	L477		S212
		P478	W337	A213
	G650	L482	N338	L214
	I651		C339	W215
	P655	S485	Q344	W216
	R658	D488	R345	F222
			I346	L223
	D663	L491	T351	A224
	S664	R492	V354	Y225
	E668	V493		A226
		L494	R358	Q227
	M671	L504	F364	D230
	D681	V507	S370	L235
	N685	P510	F371	S242
	S686	L514	C385	L246
	T687		Q388	P249
	F695	L519	I389	R253
	K696	N520	K392	P264
	L701	T522	T395	T265
	H704	K523	F396	W266
		F524	I397	K267
	N710	I529	T398	F268
	S716	K538	K399	T273
	S720	N562	Y414	L276
		W563	L415	
	V724	A564	Y416	V279
	D725	T565	Y417	
	W726	Y566	I418	T283
	G727		S419	
	V728	S569	K433	P290
	Q731	V575	I434	A291
	A732	F578	T440	S292
	H740	H592	L445	M293
	S744	L598	E448	D302
	S745	G599	R453	S312
	T746	T600	F461	L313
	A747	F601	E464	Q314
	H748			W315
	Q749			L316
	L765			M325
				D326

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 32	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	79.81Å 79.81Å 286.76Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	49.75 – 3.24 49.76 – 3.24	Depositor EDS
% Data completeness (in resolution range)	100.0 (49.75-3.24) 99.9 (49.76-3.24)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	0.11	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.15 (at 3.25Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.223 , 0.264 0.217 , 0.258	Depositor DCC
$R_{free}$ test set	1329 reflections (4.27%)	DCC
Wilson B-factor (Å <sup>2</sup> )	114.4	Xtriage
Anisotropy	0.417	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 86.1	EDS
Estimated twinning fraction	0.019 for -h,-k,l 0.065 for h,-h-k,-l 0.043 for -k,-h,-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtriage
Outliers	0 of 32574 reflections	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	11948	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	146.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.65% 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: CJP

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	0.43	0/6123	0.57	0/8329
1	B	0.43	0/6123	0.57	0/8329
All	All	0.43	0/12246	0.57	0/16658

There are no bond length outliers.

There are no bond angle outliers.

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	5951	0	5669	82	0
1	B	5951	0	5669	77	0
2	A	23	0	14	0	0
2	B	23	0	14	1	0
All	All	11948	0	11366	156	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 7.

The worst 5 of 156 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:253:ARG:HH21	1:B:253:ARG:HH21	1.11	0.92
1:B:153:GLN:HE22	1:B:170:ASN:H	1.26	0.83
1:A:153:GLN:HE22	1:A:170:ASN:H	1.27	0.78
1:A:696:LYS:HG3	1:A:728:VAL:HG22	1.67	0.75
1:A:80:ASN:HD22	1:A:82:GLU:H	1.36	0.73

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	725/752 (96%)	672 (93%)	51 (7%)	2 (0%)	46	83
1	B	725/752 (96%)	672 (93%)	52 (7%)	1 (0%)	56	90
All	All	1450/1504 (96%)	1344 (93%)	103 (7%)	3 (0%)	52	87

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	73	GLU
1	B	73	GLU
1	A	389	ILE

### 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	651/674 (97%)	600 (92%)	51 (8%)	16	52
1	B	651/674 (97%)	601 (92%)	50 (8%)	16	52
All	All	1302/1348 (97%)	1201 (92%)	101 (8%)	16	52

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

Mol	Chain	Res	Type
1	A	627	TRP
1	B	80	ASN
1	B	566	TYR
1	A	663	ASP
1	B	61	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 26 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	592	HIS
1	B	153	GLN
1	B	592	HIS
1	A	710	ASN
1	B	80	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

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	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
2	CJP	A	1000	-	21,25,25	1.94	5 (23%)	27,38,38	2.86	8 (29%)
2	CJP	B	1000	-	21,25,25	2.02	5 (23%)	27,38,38	2.82	10 (37%)

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	CJP	A	1000	-	-	0/2/12/12	0/3/3/3
2	CJP	B	1000	-	-	0/2/12/12	0/3/3/3

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1000	CJP	C21-S20	-4.00	1.68	1.75
2	B	1000	CJP	C21-S20	-3.90	1.68	1.75
2	A	1000	CJP	C11-CL2	2.40	1.79	1.73
2	B	1000	CJP	C11-CL2	2.99	1.80	1.73
2	A	1000	CJP	C8-C7	3.31	1.47	1.40

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	1000	CJP	O22-S20-O23	-8.75	106.48	118.67
2	B	1000	CJP	O22-S20-O23	-8.32	107.08	118.67
2	A	1000	CJP	C10-C8-C7	-5.77	119.54	129.57
2	B	1000	CJP	C10-C8-C7	-5.65	119.75	129.57
2	B	1000	CJP	C12-C11-C10	-4.15	118.74	121.76

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	1000	CJP	1	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	727/752 (96%)	0.32	43 (5%) 26 16	92, 140, 193, 226	0
1	B	727/752 (96%)	0.46	53 (7%) 18 11	96, 150, 199, 235	0
All	All	1454/1504 (96%)	0.39	96 (6%) 22 14	92, 145, 198, 235	0

The worst 5 of 96 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	469	GLN	6.2
1	A	468	TYR	6.1
1	B	223	LEU	5.5
1	B	88	VAL	5.1
1	B	267	LYS	5.0

### 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	CJP	B	1000	23/23	0.83	0.31	0.60	139,139,142,142	0
2	CJP	A	1000	23/23	0.84	0.34	0.55	139,139,142,142	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.