



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 10, 2018 – 03:10 pm GMT

PDB ID : 4J5X  
Title : Crystal Structure of the SR12813-bound PXR/RXRalpha LBD Heterotrimer Complex  
Authors : Wallace, B.D.; Betts, L.; Redinbo, M.R.  
Deposited on : 2013-02-10  
Resolution : 2.80 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

<https://www.wwpdb.org/validation/2017/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 : 1.7.3 (157068), CSD as539be (2018)  
Xtriage (Phenix) : 1.13  
EDS : trunk30967  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Refmac : 5.8.0158  
CCP4 : 7.0 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk30967

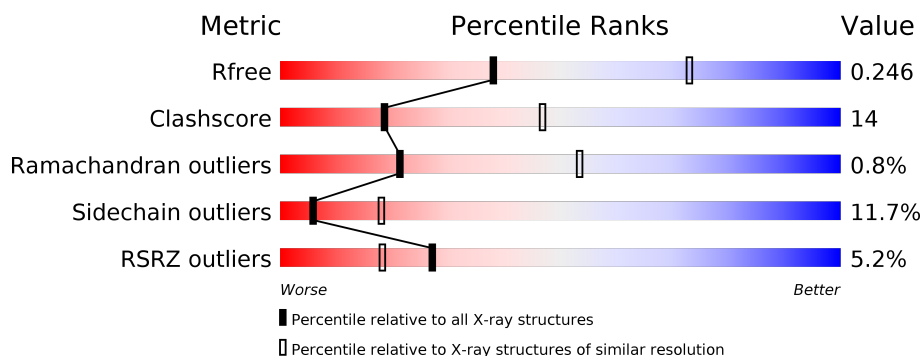
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.80 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))
$R_{free}$	111664	2792 (2.80-2.80)
Clashscore	122126	3209 (2.80-2.80)
Ramachandran outliers	120053	3158 (2.80-2.80)
Sidechain outliers	120020	3160 (2.80-2.80)
RSRZ outliers	108989	2726 (2.80-2.80)

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	C	264	<div> <div>2%</div> <div> <div></div> <div>70%</div> <div>17%</div> <div>•</div> <div>10%</div> </div> </div>
1	D	264	<div> <div>2%</div> <div> <div></div> <div>66%</div> <div>22%</div> <div>•</div> <div>9%</div> </div> </div>
2	A	336	<div> <div>5%</div> <div> <div></div> <div>53%</div> <div>29%</div> <div>• •</div> <div>13%</div> </div> </div>
2	B	336	<div> <div>7%</div> <div> <div></div> <div>51%</div> <div>31%</div> <div>5%</div> <div>13%</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 crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	SRL	B	501	-	-	-	X

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 8771 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 Retinoic acid receptor RXR-alpha, Nuclear receptor coactivator 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	D	239	Total	C	N	O	S	0	0	0
			1898	1215	329	343	11			
1	C	238	Total	C	N	O	S	0	0	0
			1887	1210	326	340	11			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	463	GLY	-	LINKER	UNP P19793
D	464	GLY	-	LINKER	UNP P19793
D	465	SER	-	LINKER	UNP P19793
D	466	GLY	-	LINKER	UNP Q15788
D	467	GLY	-	LINKER	UNP Q15788
C	463	GLY	-	LINKER	UNP P19793
C	464	GLY	-	LINKER	UNP P19793
C	465	SER	-	LINKER	UNP P19793
C	466	GLY	-	LINKER	UNP Q15788
C	467	GLY	-	LINKER	UNP Q15788

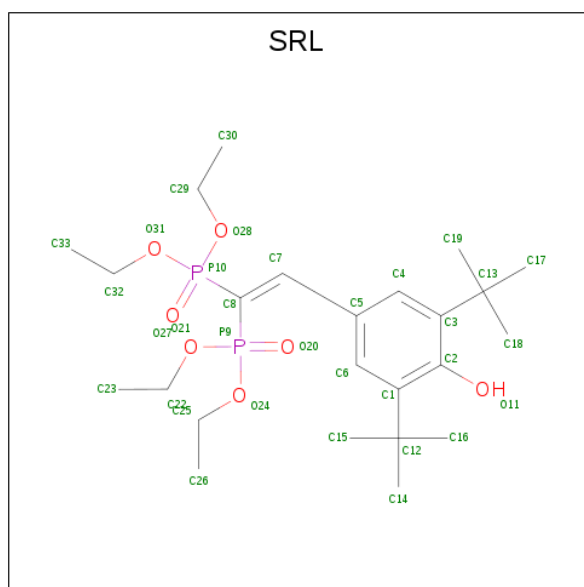
- Molecule 2 is a protein called Nuclear receptor subfamily 1 group I member 2, Nuclear receptor coactivator 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	A	293	Total	C	N	O	S	0	0	0
			2389	1532	413	426	18			
2	B	293	Total	C	N	O	S	0	0	0
			2371	1519	408	426	18			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	127	SER	-	EXPRESSION TAG	UNP O75469
A	128	ASN	-	EXPRESSION TAG	UNP O75469
A	129	ALA	-	EXPRESSION TAG	UNP O75469
A	435	GLY	-	LINKER	UNP Q15788
A	436	GLY	-	LINKER	UNP Q15788
A	437	SER	-	LINKER	UNP Q15788
A	438	GLY	-	LINKER	UNP Q15788
A	439	GLY	-	LINKER	UNP Q15788
B	127	SER	-	EXPRESSION TAG	UNP O75469
B	128	ASN	-	EXPRESSION TAG	UNP O75469
B	129	ALA	-	EXPRESSION TAG	UNP O75469
B	435	GLY	-	LINKER	UNP Q15788
B	436	GLY	-	LINKER	UNP Q15788
B	437	SER	-	LINKER	UNP Q15788
B	438	GLY	-	LINKER	UNP Q15788
B	439	GLY	-	LINKER	UNP Q15788

- Molecule 3 is [2-(3,5-DI-TERT-BUTYL-4-HYDROXY-PHENYL)-1-(DIETHOXY-PHOSPHORYL)-VINYL]-PHOSPHONIC ACID DIETHYL ESTER (three-letter code: SRL) (formula: C<sub>24</sub>H<sub>42</sub>O<sub>7</sub>P<sub>2</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	O	P	0	0
			33	24	7	2		
3	B	1	Total	C	O	P	0	0
			33	24	7	2		

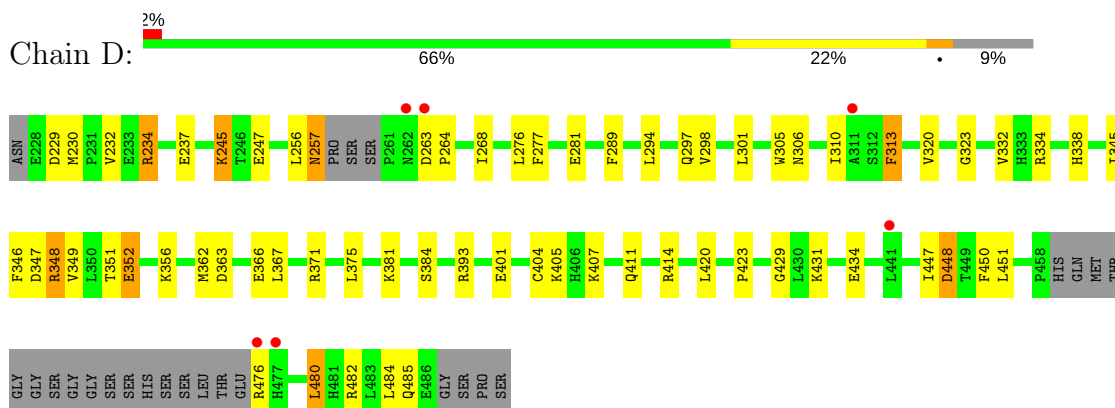
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	D	44	Total 44	O 44	0	0
4	C	34	Total 34	O 34	0	0
4	A	40	Total 40	O 40	0	0
4	B	42	Total 42	O 42	0	0

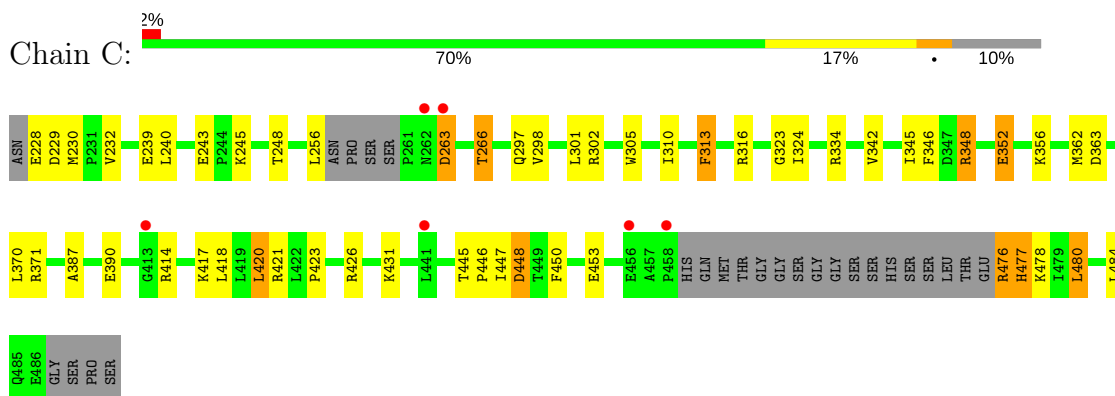
### 3 Residue-property plots

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

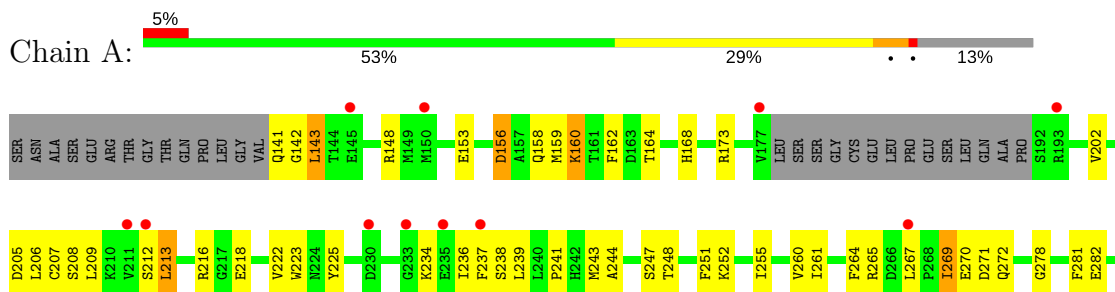
- Molecule 1: Retinoic acid receptor RXR-alpha, Nuclear receptor coactivator 1

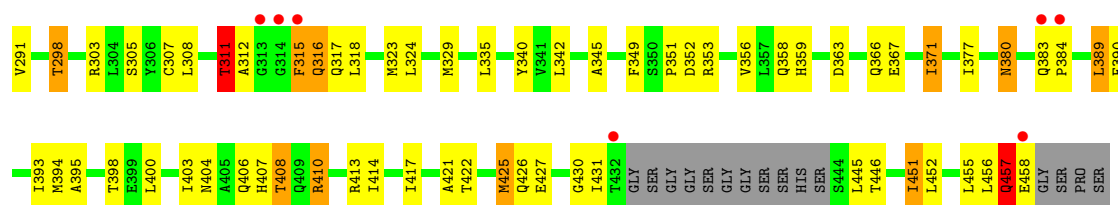


- Molecule 1: Retinoic acid receptor RXR-alpha, Nuclear receptor coactivator 1

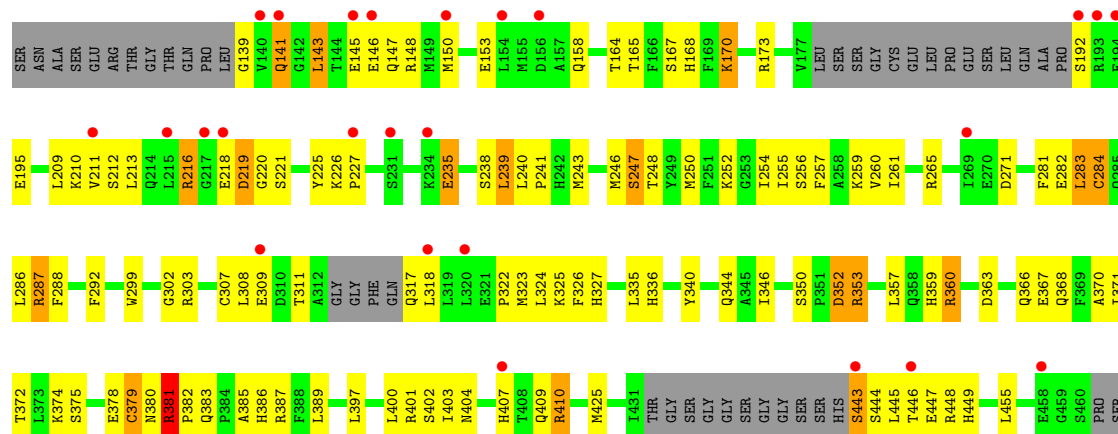


- Molecule 2: Nuclear receptor subfamily 1 group I member 2, Nuclear receptor coactivator 1





- Molecule 2: Nuclear receptor subfamily 1 group I member 2, Nuclear receptor coactivator 1





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	70.09Å 120.30Å 175.79Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.65 – 2.80 45.65 – 2.80	Depositor EDS
% Data completeness (in resolution range)	99.3 (45.65-2.80) 99.4 (45.65-2.80)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.25 (at 2.81Å)	Xtriage
Refinement program	PHENIX dev_1037	Depositor
R, $R_{free}$	0.245 , 0.298 0.249 , 0.246	Depositor DCC
$R_{free}$ test set	1852 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	37.4	Xtriage
Anisotropy	0.852	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 46.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.44$ , $\langle L^2 \rangle = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.89	EDS
Total number of atoms	8771	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 35.15 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 6.1035e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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 [i](#)

### 5.1 Standard geometry [i](#)

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

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	C	0.23	0/1924	0.42	0/2599
1	D	0.25	0/1935	0.44	0/2614
2	A	0.27	0/2440	0.52	0/3283
2	B	0.28	0/2420	0.50	0/3257
All	All	0.26	0/8719	0.48	0/11753

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 [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	C	1887	0	1926	34	1
1	D	1898	0	1936	38	1
2	A	2389	0	2402	82	0
2	B	2371	0	2374	98	0
3	A	33	0	42	3	0
3	B	33	0	42	18	0
4	A	40	0	0	14	0
4	B	42	0	0	11	0
4	C	34	0	0	1	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	D	44	0	0	9	0
All	All	8771	0	8722	246	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 14.

The worst 5 of 246 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:B:209:LEU:HB3	3:B:501:SRL:H301	1.52	0.89
3:A:501:SRL:H233	3:A:501:SRL:H173	1.55	0.87
2:B:282:GLU:HG3	2:B:404:ASN:HD22	1.44	0.83
1:D:352:GLU:HG2	2:A:352:ASP:HB3	1.64	0.79
2:A:237:PHE:O	2:A:241:PRO:HD3	1.84	0.78

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)
1:D:247:GLU:OE2	1:C:248:THR:OG1[1_665]	2.13	0.07

## 5.3 Torsion angles

### 5.3.1 Protein backbone

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	C	232/264 (88%)	220 (95%)	10 (4%)	2 (1%)	19	50
1	D	233/264 (88%)	222 (95%)	10 (4%)	1 (0%)	36	70
2	A	287/336 (85%)	263 (92%)	22 (8%)	2 (1%)	24	57
2	B	285/336 (85%)	253 (89%)	29 (10%)	3 (1%)	16	45
All	All	1037/1200 (86%)	958 (92%)	71 (7%)	8 (1%)	21	53

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	A	457	GLN
2	B	219	ASP
2	B	381	ARG
1	C	263	ASP
2	A	311	THR

### 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	C	205/227 (90%)	184 (90%)	21 (10%)	8	24
1	D	207/227 (91%)	184 (89%)	23 (11%)	7	20
2	A	262/295 (89%)	234 (89%)	28 (11%)	7	21
2	B	260/295 (88%)	223 (86%)	37 (14%)	3	11
All	All	934/1044 (90%)	825 (88%)	109 (12%)	6	18

5 of 109 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	A	168	HIS
2	A	371	ILE
2	B	381	ARG
2	A	213	LEU
2	A	311	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
2	A	141	GLN
2	B	404	ASN
2	B	141	GLN
1	C	477	HIS
2	B	201	GLN

### 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
3	SRL	A	501	-	30,33,33	1.20	4 (13%)	45,50,50	1.97	12 (26%)
3	SRL	B	501	-	30,33,33	1.68	6 (20%)	45,50,50	2.07	16 (35%)

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	SRL	A	501	-	-	0/30/44/44	0/1/1/1
3	SRL	B	501	-	-	0/30/44/44	0/1/1/1

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	501	SRL	C13-C3	-4.26	1.47	1.54

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	501	SRL	C4-C5	-2.99	1.34	1.39
3	B	501	SRL	C12-C1	-2.33	1.50	1.54
3	A	501	SRL	C13-C3	-2.21	1.50	1.54
3	A	501	SRL	C5-C7	2.12	1.51	1.46

The worst 5 of 28 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	501	SRL	C5-C7-C8	-5.81	120.35	131.09
3	B	501	SRL	C3-C2-C1	-5.29	116.89	122.62
3	A	501	SRL	O31-P10-O27	-3.54	105.37	115.07
3	A	501	SRL	C3-C2-C1	-3.05	119.32	122.62
3	B	501	SRL	O28-P10-C8	-2.79	97.38	105.18

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 21 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	501	SRL	3	0
3	B	501	SRL	18	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	C	238/264 (90%)	0.32	6 (2%) 57 47	28, 47, 73, 123	0
1	D	239/264 (90%)	0.27	6 (2%) 57 47	30, 46, 79, 101	0
2	A	293/336 (87%)	0.51	18 (6%) 21 13	28, 54, 90, 128	0
2	B	293/336 (87%)	0.54	25 (8%) 11 5	33, 55, 93, 111	0
All	All	1063/1200 (88%)	0.42	55 (5%) 27 18	28, 51, 87, 128	0

The worst 5 of 55 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	477	HIS	4.9
2	A	314	GLY	4.6
2	A	313	GLY	4.6
2	A	383	GLN	4.0
1	C	262	ASN	3.9

### 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. 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	B-factors( $\text{\AA}^2$ )	Q<0.9
3	SRL	B	501	33/33	0.73	0.42	56,88,109,114	0
3	SRL	A	501	33/33	0.77	0.40	57,80,95,114	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.