



# Full wwPDB X-ray Structure Validation Report ⓘ

May 15, 2020 – 02:16 am BST

PDB ID : 4PJU  
Title : crystal structure of human Stromal Antigen 2 (SA2) in complex with Sister Chromatid Cohesion protein 1 (Scc1)  
Authors : Hara, K.; Chen, Z.; Tomchick, D.R.; Yu, H.  
Deposited on : 2014-05-12  
Resolution : 3.05 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.11  
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

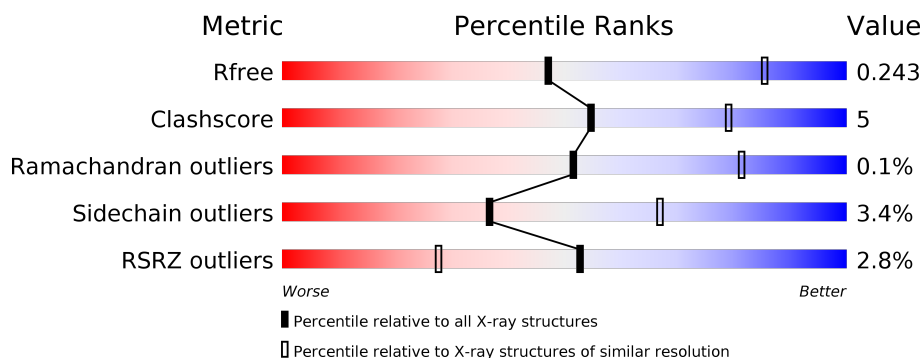
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.05 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))
$R_{free}$	130704	1754 (3.10-3.02)
Clashscore	141614	1864 (3.10-3.02)
Ramachandran outliers	138981	1794 (3.10-3.02)
Sidechain outliers	138945	1793 (3.10-3.02)
RSRZ outliers	127900	1713 (3.10-3.02)

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 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  $\leq 5\%$ . 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	981	<div> <div>2%</div> <div> <div></div> <div>77%</div> <div>14%</div> <div>9%</div> </div> </div>
2	B	140	<div> <div>2%</div> <div> <div></div> <div>47%</div> <div>5%</div> <div>47%</div> </div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 15923 atoms, of which 7992 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 Cohesin subunit SA-2.

Mol	Chain	Residues	Atoms								ZeroOcc	AltConf	Trace
1	A	896	Total	C	H	N	O	S	Se		0	0	0
			14666	4674	7353	1216	1369	20	34				

- Molecule 2 is a protein called Double-strand-break repair protein rad21 homolog.

Mol	Chain	Residues	Atoms								ZeroOcc	AltConf	Trace
2	B	74	Total	C	H	N	O	S	Se		0	0	0
			1235	385	639	101	107	1	2				

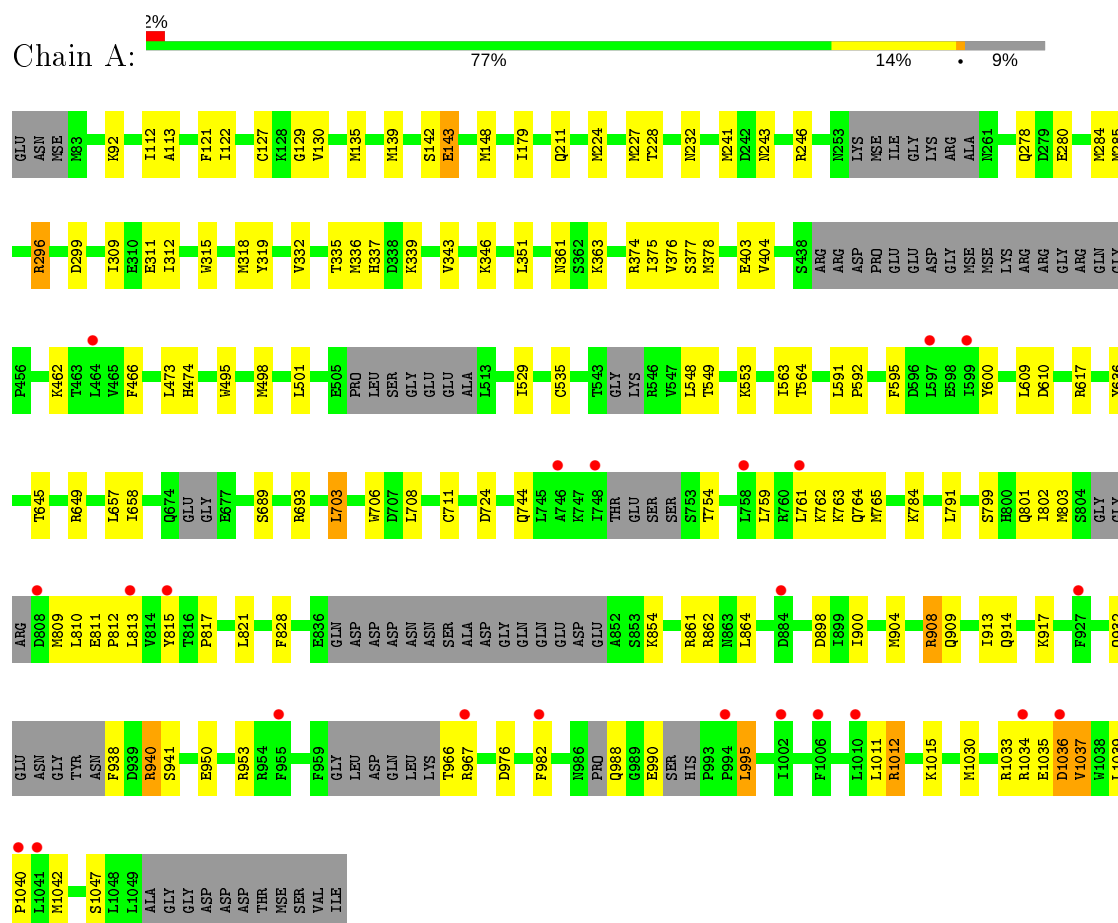
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	19	Total	O		
			19	19	0	0
3	B	3	Total	O		
			3	3	0	0

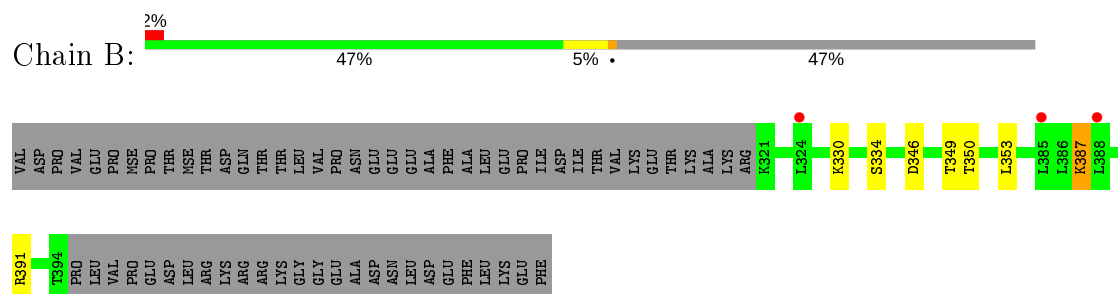
### 3 Residue-property plots

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: Cohesin subunit SA-2



#### • Molecule 2: Double-strand-break repair protein rad21 homolog



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	78.61Å 108.75Å 181.79Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.45 – 3.05 45.45 – 3.05	Depositor EDS
% Data completeness (in resolution range)	99.9 (45.45-3.05) 100.0 (45.45-3.05)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.15 (at 3.06Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.2_1309)	Depositor
R, $R_{free}$	0.195 , 0.238 0.201 , 0.243	Depositor DCC
$R_{free}$ test set	1521 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	92.4	Xtriage
Anisotropy	0.326	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 42.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	15923	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	87.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.45% 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

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	0.35	0/7402	0.54	0/9914
2	B	0.33	0/604	0.55	0/812
All	All	0.35	0/8006	0.54	0/10726

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	7313	7353	7331	80	0
2	B	596	639	639	5	0
3	A	19	0	0	1	0
3	B	3	0	0	0	0
All	All	7931	7992	7970	82	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.

All (82) 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:241:MSE:HE1	1:A:278:GLN:HB2	1.69	0.72

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:224:MSE:HB3	1:A:311:GLU:HG3	1.74	0.69
1:A:337:HIS:NE2	3:A:1117:HOH:O	2.26	0.68
2:B:349:THR:OG1	2:B:350:THR:N	2.33	0.61
1:A:296:ARG:NH1	1:A:299:ASP:OD1	2.34	0.60
1:A:228:THR:HG23	1:A:318:MSE:HE1	1.84	0.59
1:A:803:MSE:N	1:A:803:MSE:SE	2.85	0.58
1:A:862:ARG:NH2	1:A:898:ASP:OD1	2.38	0.56
2:B:387:LYS:O	2:B:391:ARG:HB2	2.06	0.56
1:A:122:ILE:HG21	1:A:148:MSE:CE	2.35	0.56
1:A:361:ASN:OD1	1:A:363:LYS:N	2.39	0.56
1:A:950:GLU:OE2	1:A:953:ARG:NH2	2.37	0.55
1:A:122:ILE:HG21	1:A:148:MSE:HE2	1.89	0.54
1:A:243:ASN:OD1	1:A:246:ARG:NH2	2.37	0.54
1:A:564:THR:OG1	1:A:595:PHE:HB2	2.07	0.54
1:A:139:MSE:HE2	1:A:143:GLU:HB2	1.89	0.53
1:A:374:ARG:HG3	1:A:378:MSE:HE2	1.90	0.53
1:A:900:ILE:HG22	1:A:904:MSE:HE3	1.92	0.52
1:A:932:GLN:OE1	1:A:932:GLN:N	2.43	0.52
1:A:744:GLN:NE2	1:A:764:GLN:OE1	2.44	0.51
1:A:799:SER:O	1:A:801:GLN:N	2.43	0.50
1:A:1035:GLU:HG3	1:A:1036:ASP:HA	1.92	0.50
1:A:966:THR:OG1	1:A:967:ARG:N	2.46	0.48
1:A:129:GLY:C	1:A:148:MSE:HE1	2.33	0.48
1:A:861:ARG:HA	1:A:864:LEU:HD12	1.95	0.48
1:A:940:ARG:HD3	1:A:941:SER:N	2.29	0.48
1:A:228:THR:CG2	1:A:318:MSE:HE1	2.43	0.48
1:A:375:ILE:HA	1:A:378:MSE:HE3	1.95	0.48
1:A:280:GLU:HG3	1:A:284:MSE:HE3	1.95	0.47
1:A:917:LYS:NZ	1:A:976:ASP:OD2	2.42	0.47
1:A:241:MSE:SE	1:A:278:GLN:NE2	2.98	0.47
1:A:473:LEU:HD13	2:B:353:LEU:HD21	1.97	0.47
1:A:549:THR:O	1:A:553:LYS:HG2	2.15	0.46
1:A:913:ILE:HG13	1:A:914:GLN:N	2.30	0.46
1:A:988:GLN:NE2	1:A:990:GLU:OE2	2.46	0.46
1:A:724:ASP:N	1:A:724:ASP:OD1	2.48	0.46
1:A:600:TYR:CE2	1:A:609:LEU:HB2	2.51	0.46
1:A:982:PHE:CE1	1:A:1030:MSE:HE1	2.50	0.46
1:A:529:ILE:HG23	1:A:563:ILE:HD13	1.98	0.46
1:A:591:LEU:HB2	1:A:592:PRO:HD3	1.98	0.46
1:A:759:LEU:O	1:A:763:LYS:HG3	2.16	0.46
1:A:765:MSE:HE1	1:A:815:TYR:HB2	1.98	0.45

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:592:PRO:HA	1:A:595:PHE:CZ	2.51	0.45
1:A:332:VAL:HG11	1:A:351:LEU:CD1	2.47	0.45
1:A:744:GLN:HB2	1:A:761:LEU:HD13	1.97	0.45
1:A:811:GLU:N	1:A:812:PRO:HD2	2.31	0.45
1:A:127:CYS:SG	1:A:148:MSE:HE3	2.56	0.45
1:A:908:ARG:HG2	1:A:909:GLN:N	2.31	0.45
1:A:802:ILE:HB	1:A:803:MSE:SE	2.66	0.45
1:A:112:ILE:HG13	1:A:113:ALA:N	2.32	0.45
1:A:689:SER:O	1:A:693:ARG:HG3	2.17	0.45
1:A:810:LEU:HB3	1:A:813:LEU:HD12	1.99	0.44
1:A:462:LYS:O	1:A:466:PHE:HD1	2.01	0.44
1:A:312:ILE:HA	1:A:315:TRP:CE3	2.53	0.43
1:A:335:THR:HG22	1:A:343:VAL:HG12	2.00	0.43
1:A:784:LYS:HB3	1:A:828:PHE:HE1	1.83	0.43
1:A:309:ILE:HG13	1:A:335:THR:HG21	2.00	0.43
1:A:1011:LEU:HD12	1:A:1012:ARG:H	1.84	0.43
1:A:336:MSE:HE3	1:A:374:ARG:CG	2.48	0.43
1:A:762:LYS:HG3	1:A:813:LEU:HD23	2.00	0.42
1:A:799:SER:C	1:A:801:GLN:H	2.21	0.42
1:A:706:TRP:O	1:A:708:LEU:N	2.52	0.42
1:A:610:ASP:OD2	1:A:649:ARG:NH2	2.52	0.42
1:A:1034:ARG:HG2	1:A:1042:MSE:SE	2.69	0.42
1:A:332:VAL:HG11	1:A:351:LEU:HD11	2.02	0.42
1:A:617:ARG:HD2	1:A:657:LEU:HB2	2.02	0.42
1:A:335:THR:CG2	1:A:343:VAL:HG12	2.49	0.42
1:A:121:PHE:HE1	1:A:179:ILE:HD11	1.83	0.42
1:A:1039:LEU:N	1:A:1040:PRO:CD	2.83	0.41
1:A:658:ILE:HD13	1:A:703:LEU:CD2	2.50	0.41
1:A:339:LYS:HE3	2:B:346:ASP:OD2	2.20	0.41
1:A:1037:VAL:HG12	1:A:1037:VAL:O	2.20	0.41
1:A:817:PRO:HB3	1:A:821:LEU:HD23	2.02	0.41
1:A:809:MSE:HE3	1:A:811:GLU:HB2	2.03	0.41
1:A:336:MSE:SE	1:A:378:MSE:HE1	2.70	0.41
1:A:495:TRP:HA	1:A:498:MSE:HE3	2.02	0.41
1:A:211:GLN:HG2	2:B:330:LYS:HD3	2.01	0.40
1:A:535:CYS:HB3	1:A:548:LEU:HD11	2.02	0.40
1:A:1012:ARG:HA	1:A:1015:LYS:HB2	2.04	0.40
1:A:135:MSE:HE3	1:A:139:MSE:SE	2.72	0.40
1:A:227:MSE:HE3	1:A:285:MSE:HB3	2.03	0.40
1:A:938:PHE:CE1	1:A:995:LEU:HG	2.56	0.40

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	870/981 (89%)	836 (96%)	33 (4%)	1 (0%)	51	81
2	B	72/140 (51%)	70 (97%)	2 (3%)	0	100	100
All	All	942/1121 (84%)	906 (96%)	35 (4%)	1 (0%)	51	81

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	1037	VAL

### 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	815/843 (97%)	787 (97%)	28 (3%)	37	67
2	B	69/125 (55%)	67 (97%)	2 (3%)	42	70
All	All	884/968 (91%)	854 (97%)	30 (3%)	37	67

All (30) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	92	LYS
1	A	130	VAL
1	A	142	SER
1	A	143	GLU
1	A	232	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	296	ARG
1	A	319	TYR
1	A	346	LYS
1	A	376	VAL
1	A	377	SER
1	A	403	GLU
1	A	404	VAL
1	A	474	HIS
1	A	501	LEU
1	A	636	TYR
1	A	645	THR
1	A	703	LEU
1	A	711	CYS
1	A	754	THR
1	A	791	LEU
1	A	854	LYS
1	A	908	ARG
1	A	940	ARG
1	A	995	LEU
1	A	1012	ARG
1	A	1033	ARG
1	A	1036	ASP
1	A	1047	SER
2	B	334	SER
2	B	387	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	247	GLN
1	A	261	ASN
1	A	275	GLN
1	A	277	ASN
1	A	278	GLN

### 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](#)

There are no ligands in this entry.

## 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

### 6.1 Protein, DNA and RNA chains

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	862/981 (87%)	0.34	23 (2%) 54 28	32, 78, 128, 174	0
2	B	72/140 (51%)	0.31	3 (4%) 36 17	43, 67, 99, 131	0
All	All	934/1121 (83%)	0.33	26 (2%) 53 28	32, 78, 127, 174	0

All (26) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	813	LEU	3.9
1	A	758	LEU	3.8
1	A	1040	PRO	3.7
1	A	597	LEU	3.5
1	A	599	ILE	3.5
1	A	927	PHE	3.4
1	A	994	PRO	3.2
1	A	815	TYR	3.1
1	A	1041	LEU	3.0
1	A	761	LEU	2.7
1	A	982	PHE	2.7
1	A	1006	PHE	2.6
1	A	1010	LEU	2.6
2	B	385	LEU	2.6
1	A	746	ALA	2.4
2	B	388	LEU	2.4
1	A	808	ASP	2.4
2	B	324	LEU	2.3
1	A	955	PHE	2.2
1	A	1034	ARG	2.2
1	A	884	ASP	2.1
1	A	1036	ASP	2.1
1	A	748	ILE	2.0
1	A	967	ARG	2.0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	464	LEU	2.0
1	A	1002	ILE	2.0

## 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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.