



# Full wwPDB X-ray Structure Validation Report ⓘ

Oct 16, 2017 – 09:35 PM EDT

PDB ID : 3D1M  
Title : Crystal Structure of Sonic Hedgehog Bound to the third FNIII domain of CDO  
Authors : McLellan, J.S.; Leahy, D.J.  
Deposited on : unknown  
Resolution : 1.70 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

<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  
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

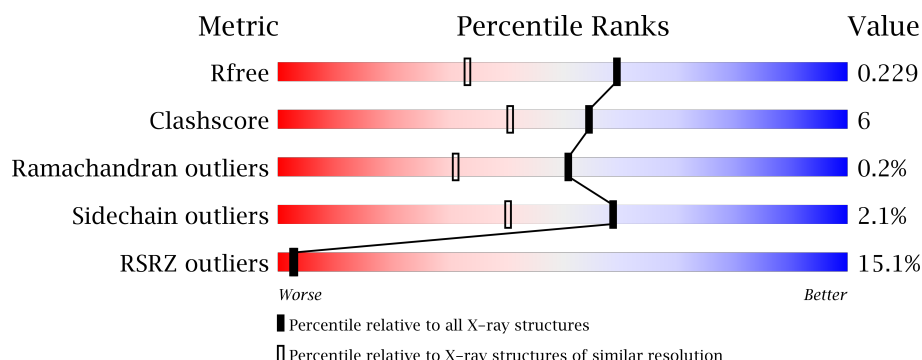
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.70 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	3453 (1.70-1.70)
Clashscore	112137	3876 (1.70-1.70)
Ramachandran outliers	110173	3815 (1.70-1.70)
Sidechain outliers	110143	3815 (1.70-1.70)
RSRZ outliers	101464	3491 (1.70-1.70)

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	168	<div> <div>10%</div> <div> <div></div> <div>80%</div> <div>8%</div> <div>12%</div> </div> </div>
1	B	168	<div> <div>8%</div> <div> <div></div> <div>79%</div> <div>9%</div> <div>11%</div> </div> </div>
2	C	102	<div> <div>24%</div> <div> <div></div> <div>75%</div> <div>11%</div> <div>15%</div> </div> </div>
2	D	102	<div> <div>20%</div> <div> <div></div> <div>85%</div> <div>10%</div> <div>5%</div> </div> </div>

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 4327 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 Sonic hedgehog protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	148	Total	C	N	O	S	0	0	0
			1189	743	212	229	5			
1	B	150	Total	C	N	O	S	0	0	0
			1207	754	214	234	5			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	22	GLY	-	EXPRESSION TAG	UNP Q62226
A	23	PRO	-	EXPRESSION TAG	UNP Q62226
A	24	GLY	-	EXPRESSION TAG	UNP Q62226
A	25	SER	-	EXPRESSION TAG	UNP Q62226
B	22	GLY	-	EXPRESSION TAG	UNP Q62226
B	23	PRO	-	EXPRESSION TAG	UNP Q62226
B	24	GLY	-	EXPRESSION TAG	UNP Q62226
B	25	SER	-	EXPRESSION TAG	UNP Q62226

- Molecule 2 is a protein called Cell adhesion molecule.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	87	Total	C	N	O	S	0	0	0
			714	450	114	144	6			
2	D	99	Total	C	N	O	S	0	0	0
			801	503	129	163	6			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	823	GLY	-	EXPRESSION TAG	UNP Q4KMG0
C	824	SER	-	EXPRESSION TAG	UNP Q4KMG0
C	825	THR	-	EXPRESSION TAG	UNP Q4KMG0
D	823	GLY	-	EXPRESSION TAG	UNP Q4KMG0

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
D	824	SER	-	EXPRESSION TAG	UNP Q4KMG0
D	825	THR	-	EXPRESSION TAG	UNP Q4KMG0

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	1	Total 1	Zn 1	0	0
3	A	1	Total 1	Zn 1	0	0

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	2	Total 2	Ca 2	0	0
4	A	2	Total 2	Ca 2	0	0

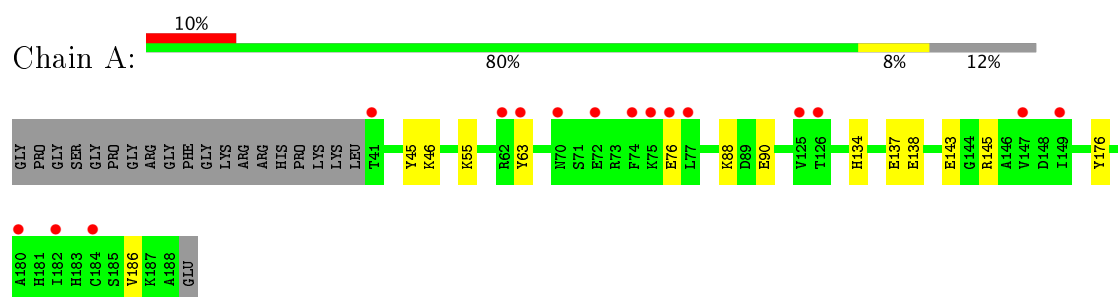
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	136	Total 136	O 136	0	0
5	B	136	Total 136	O 136	0	0
5	C	62	Total 62	O 62	0	0
5	D	76	Total 76	O 76	0	0

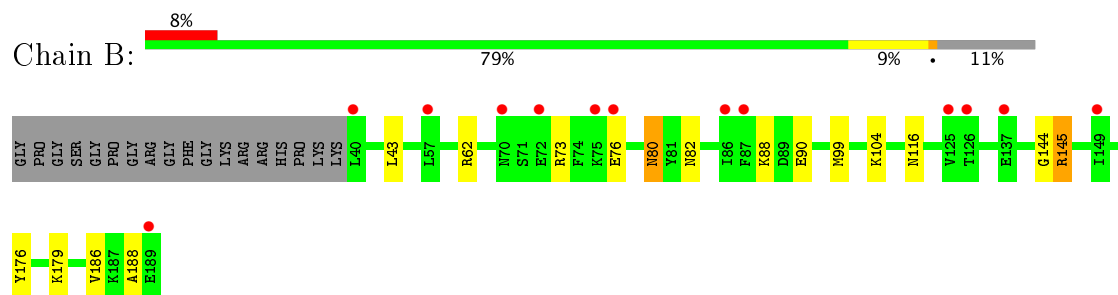
### 3 Residue-property plots [i](#)

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

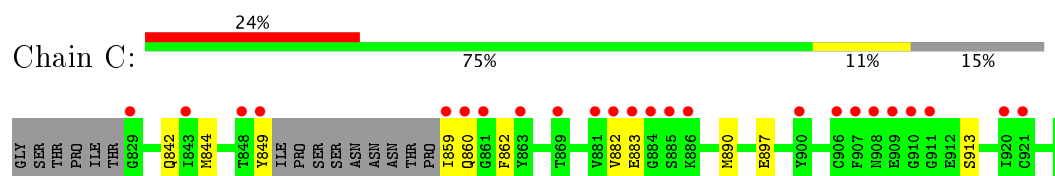
- Molecule 1: Sonic hedgehog protein



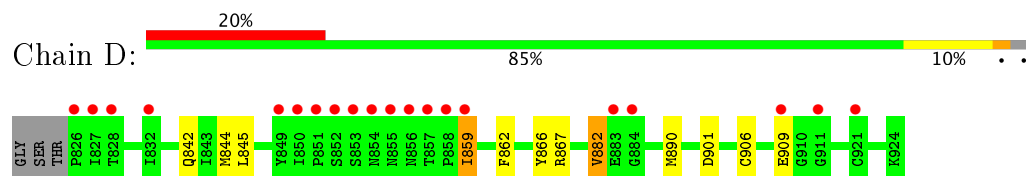
- Molecule 1: Sonic hedgehog protein



- Molecule 2: Cell adhesion molecule



- Molecule 2: Cell adhesion molecule



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	71.18Å 98.58Å 144.05Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.56 – 1.70 30.55 – 1.70	Depositor EDS
% Data completeness (in resolution range)	99.0 (30.56-1.70) 99.0 (30.55-1.70)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.53 (at 1.70Å)	Xtriage
Refinement program	REFMAC	Depositor
R, $R_{free}$	0.186 , 0.229 0.186 , 0.229	Depositor DCC
$R_{free}$ test set	2693 reflections (5.11%)	DCC
Wilson B-factor (Å <sup>2</sup> )	24.6	Xtriage
Anisotropy	0.041	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 44.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	4327	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	33.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 11.33% 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: ZN, CA

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	0.64	0/1215	0.72	0/1640
1	B	0.67	0/1233	0.80	4/1663 (0.2%)
2	C	0.85	2/732 (0.3%)	0.77	0/987
2	D	0.66	0/823	0.75	1/1116 (0.1%)
All	All	0.70	2/4003 (0.0%)	0.76	5/5406 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	860	GLN	CD-NE2	7.14	1.50	1.32
2	C	860	GLN	CD-OE1	5.37	1.35	1.24

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	145	ARG	NE-CZ-NH1	8.29	124.45	120.30
1	B	145	ARG	NE-CZ-NH2	-8.25	116.17	120.30
1	B	43	LEU	CA-CB-CG	5.84	128.72	115.30
2	D	867	ARG	CG-CD-NE	-5.73	99.77	111.80
1	B	43	LEU	CB-CG-CD2	5.05	119.59	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	1189	0	1149	10	0
1	B	1207	0	1166	13	0
2	C	714	0	655	18	0
2	D	801	0	742	16	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	2	0	0	0	0
4	B	2	0	0	0	0
5	A	136	0	0	2	0
5	B	136	0	0	2	0
5	C	62	0	0	0	0
5	D	76	0	0	1	0
All	All	4327	0	3712	46	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 6.

All (46) 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:C:842:GLN:HE22	2:D:842:GLN:HE22	1.14	0.93
1:A:134:HIS:HD2	5:D:130:HOH:O	1.67	0.76
2:C:890:MET:CE	2:D:890:MET:CE	2.69	0.70
2:C:890:MET:CE	2:D:890:MET:HE1	2.21	0.70
1:A:55:LYS:CD	5:A:251:HOH:O	2.40	0.69
1:A:55:LYS:HD2	5:A:251:HOH:O	2.00	0.62
1:B:80:ASN:HD21	1:B:82:ASN:HB2	1.64	0.60
1:A:63:TYR:CE1	1:A:138:GLU:HA	2.37	0.60
1:B:80:ASN:HD22	1:B:82:ASN:H	1.52	0.57
1:B:116:ASN:HB2	5:B:274:HOH:O	2.04	0.57
1:B:80:ASN:C	1:B:80:ASN:HD22	2.08	0.57
2:C:844:MET:CE	2:D:844:MET:HE2	2.36	0.56
1:B:80:ASN:ND2	1:B:82:ASN:H	2.04	0.56
1:A:63:TYR:CZ	1:A:138:GLU:HA	2.42	0.54
2:C:890:MET:HE2	2:D:890:MET:HE1	1.89	0.54
2:C:890:MET:CE	2:D:890:MET:HE3	2.36	0.54
2:C:862:PHE:HB2	2:C:882:VAL:HG22	1.90	0.53
2:D:862:PHE:HB2	2:D:882:VAL:HG22	1.89	0.53
2:C:844:MET:CE	2:C:890:MET:HE2	2.39	0.52
1:B:73:ARG:O	1:B:76:GLU:HG2	2.10	0.51

*Continued on next page...*



*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:63:TYR:CD1	1:A:143:GLU:OE2	2.64	0.50
2:C:844:MET:CE	2:C:890:MET:CE	2.89	0.50
2:C:844:MET:HE3	2:C:890:MET:CE	2.41	0.50
2:C:890:MET:HE1	2:D:890:MET:HE3	1.94	0.50
1:B:145:ARG:HD2	1:B:186:VAL:O	2.13	0.49
1:B:80:ASN:HD21	1:B:82:ASN:CB	2.25	0.49
2:C:849:TYR:CE2	2:C:859:ILE:HD11	2.49	0.48
2:C:849:TYR:HE2	2:C:859:ILE:HD11	1.78	0.47
1:A:137:GLU:H	1:A:137:GLU:CD	2.18	0.47
1:A:45:TYR:CZ	1:A:46:LYS:HE2	2.50	0.47
1:A:88:LYS:HE3	1:A:90:GLU:HB3	1.97	0.47
1:B:145:ARG:HD3	1:B:188:ALA:HA	1.96	0.46
1:A:145:ARG:HD2	1:A:186:VAL:O	2.17	0.45
1:B:88:LYS:HE3	1:B:90:GLU:HB3	1.98	0.45
2:D:859:ILE:HD12	2:D:906:CYS:HB2	1.98	0.45
1:B:179:LYS:HD2	2:C:897:GLU:OE1	2.16	0.45
2:C:890:MET:HE1	2:D:890:MET:CE	2.47	0.44
1:B:99:MET:HG3	1:B:104:LYS:HG3	1.99	0.44
2:C:844:MET:HE3	2:C:890:MET:HE1	2.00	0.44
2:D:859:ILE:HD11	2:D:862:PHE:CZ	2.53	0.44
2:D:909:GLU:CD	2:D:909:GLU:H	2.21	0.44
1:B:144:GLY:HA2	5:B:308:HOH:O	2.17	0.44
2:C:844:MET:CE	2:D:844:MET:CE	2.97	0.43
2:D:866:TYR:HA	2:D:901:ASP:O	2.20	0.41
2:C:844:MET:HE3	2:D:844:MET:CE	2.50	0.40
2:D:844:MET:CE	2:D:890:MET:HE2	2.51	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	146/168 (87%)	144 (99%)	2 (1%)	0	100	100
1	B	148/168 (88%)	146 (99%)	2 (1%)	0	100	100
2	C	83/102 (81%)	80 (96%)	2 (2%)	1 (1%)	15	3
2	D	97/102 (95%)	93 (96%)	4 (4%)	0	100	100
All	All	474/540 (88%)	463 (98%)	10 (2%)	1 (0%)	51	31

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	913	SER

### 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	126/140 (90%)	124 (98%)	2 (2%)	68	53
1	B	128/140 (91%)	125 (98%)	3 (2%)	56	36
2	C	79/93 (85%)	78 (99%)	1 (1%)	73	60
2	D	91/93 (98%)	88 (97%)	3 (3%)	43	21
All	All	424/466 (91%)	415 (98%)	9 (2%)	59	40

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	76	GLU
1	A	176	TYR
1	B	62	ARG
1	B	80	ASN
1	B	176	TYR
2	C	883	GLU
2	D	845	LEU
2	D	859	ILE
2	D	882	VAL

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

Mol	Chain	Res	Type
1	B	80	ASN
2	D	842	GLN
2	D	855	ASN
2	D	860	GLN
2	D	905	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](#)

Of 6 ligands modelled in this entry, 6 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

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 [i](#)

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 ⓘ

### 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	148/168 (88%)	0.62	16 (10%) 6 7	29, 32, 42, 47	0
1	B	150/168 (89%)	0.40	13 (8%) 11 13	27, 30, 39, 46	0
2	C	87/102 (85%)	1.32	24 (27%) 1 0	29, 33, 44, 52	0
2	D	99/102 (97%)	1.07	20 (20%) 1 1	28, 32, 44, 49	0
All	All	484/540 (89%)	0.77	73 (15%) 3 3	27, 32, 43, 52	0

All (73) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	827	ILE	8.5
2	C	849	TYR	8.2
2	D	857	THR	6.7
2	C	911	GLY	6.5
1	A	70	ASN	6.3
2	D	856	ASN	6.0
2	C	907	PHE	5.6
2	C	859	ILE	5.3
2	D	855	ASN	5.3
2	D	854	ASN	5.3
2	D	853	SER	5.2
2	D	826	PRO	4.7
2	D	858	PRO	4.7
2	C	910	GLY	4.4
1	B	70	ASN	4.2
2	C	920	ILE	4.2
1	B	40	LEU	4.0
2	C	909	GLU	4.0
1	A	75	LYS	3.8
2	C	885	SER	3.8
1	A	125	VAL	3.6

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
2	D	852	SER	3.5
2	C	908	ASN	3.4
2	D	849	TYR	3.3
1	B	125	VAL	3.3
2	C	883	GLU	3.1
2	D	859	ILE	3.1
2	C	860	GLN	3.0
2	C	921	CYS	3.0
2	D	909	GLU	2.9
1	A	72	GLU	2.9
1	B	86	ILE	2.9
2	D	850	ILE	2.9
1	A	63	TYR	2.9
1	A	126	THR	2.8
2	C	829	GLY	2.8
2	C	863	TYR	2.7
1	B	76	GLU	2.7
1	B	189	GLU	2.7
2	D	921	CYS	2.6
1	B	126	THR	2.6
2	C	906	CYS	2.6
2	C	861	GLY	2.6
1	B	87	PHE	2.6
1	A	147	VAL	2.6
2	D	884	GLY	2.5
1	A	184	CYS	2.5
1	A	76	GLU	2.4
1	A	182	ILE	2.4
1	B	149	ILE	2.4
2	C	882	VAL	2.4
1	B	72	GLU	2.4
2	D	828	THR	2.4
2	C	881	VAL	2.4
2	D	832	ILE	2.4
2	C	900	TYR	2.4
2	D	911	GLY	2.4
1	A	62	ARG	2.3
2	C	886	LYS	2.3
1	A	149	ILE	2.3
1	B	57	LEU	2.2
1	A	74	PHE	2.2
2	D	883	GLU	2.2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	B	137	GLU	2.2
1	A	180	ALA	2.2
2	D	851	PRO	2.1
1	A	41	THR	2.1
2	C	843	ILE	2.1
2	C	869	THR	2.0
1	A	77	LEU	2.0
1	B	75	LYS	2.0
2	C	848	THR	2.0
2	C	884	GLY	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](#)

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(Å <sup>2</sup> )	Q<0.9
4	CA	B	4	1/1	1.00	0.04	-2.13	22,22,22,22	0
4	CA	B	3	1/1	1.00	0.03	-3.17	19,19,19,19	0
4	CA	A	2	1/1	0.99	0.05	-4.26	21,21,21,21	0
4	CA	A	190	1/1	0.99	0.03	-4.45	19,19,19,19	0
3	ZN	B	2	1/1	0.98	0.05	-4.58	26,26,26,26	1
3	ZN	A	1	1/1	0.97	0.06	-6.54	25,25,25,25	1

## 6.5 Other polymers [i](#)

There are no such residues in this entry.