



# Full wwPDB X-ray Structure Validation Report ⓘ

Aug 22, 2020 – 11:42 AM BST

PDB ID : 5W1A  
Title : The first X-ray crystal structure of an insect muscle myosin. *Drosophila melanogaster*, skeletal muscle myosin II, an embryonic isoform, subfragment-1  
Authors : Caldwell, J.T.; Bernstein, S.I.; Huxford, T.  
Deposited on : 2017-06-02  
Resolution : 2.23 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 : 2.13.1  
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) : 2.13.1

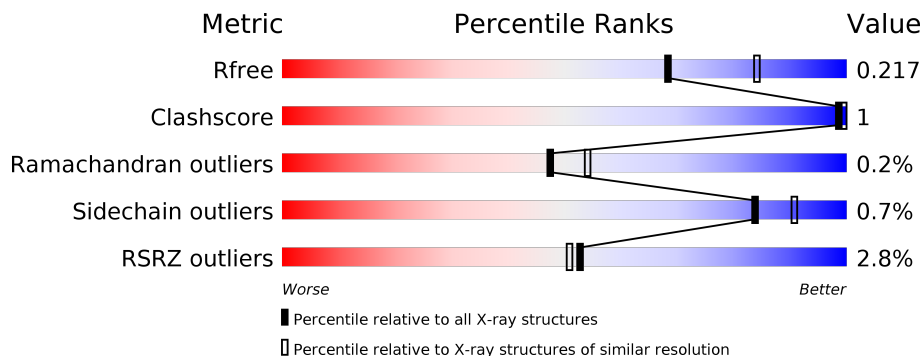
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.23 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	5912 (2.24-2.20)
Clashscore	141614	6646 (2.24-2.20)
Ramachandran outliers	138981	6543 (2.24-2.20)
Sidechain outliers	138945	6544 (2.24-2.20)
RSRZ outliers	127900	5797 (2.24-2.20)

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 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	810	
1	C	810	
2	B	155	
2	D	155	

## 2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 31091 atoms, of which 14967 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.

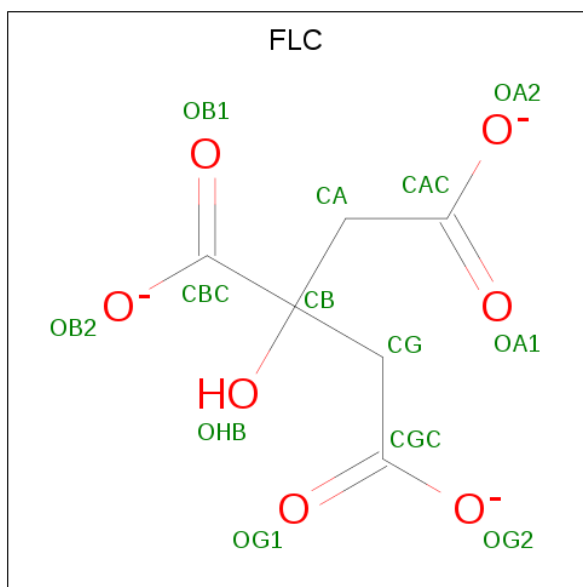
- Molecule 1 is a protein called Myosin heavy chain, muscle.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	H	N	O				S
1	A	786	12528	3987	6247	1081	1176	37	0	1	0
1	C	785	12579	3996	6281	1088	1176	38	0	2	0

- Molecule 2 is a protein called Myosin light chain alkali.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	H	N	O				S
2	B	154	2414	769	1197	193	246	9	0	0	0
2	D	151	2380	758	1181	191	242	8	0	0	0

- Molecule 3 is CITRATE ANION (three-letter code: FLC) (formula:  $C_6H_5O_7^-$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	H	O		
3	A	1	18	6	5	7	0	0
3	C	1	18	6	5	7	0	0

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	H	O		
4	A	1	12	3	6	3	0	0
4	A	1	12	3	6	3	0	0
4	C	1	14	3	8	3	0	0
4	C	1	13	3	7	3	0	0

- Molecule 5 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C<sub>2</sub>H<sub>6</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	1	Total	C	H	O	0	0
			10	2	6	2		
5	A	1	Total	C	H	O	0	0
			10	2	6	2		
5	A	1	Total	C	H	O	0	0
			10	2	6	2		
5	C	1	Total	C	H	O	0	0
			10	2	6	2		

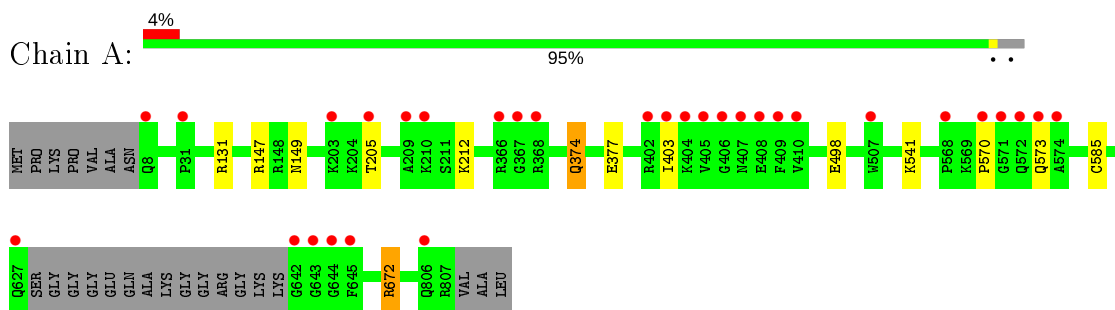
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	430	Total	O	0	0
			430	430		
6	B	52	Total	O	0	0
			52	52		
6	C	471	Total	O	0	0
			471	471		
6	D	110	Total	O	0	0
			110	110		

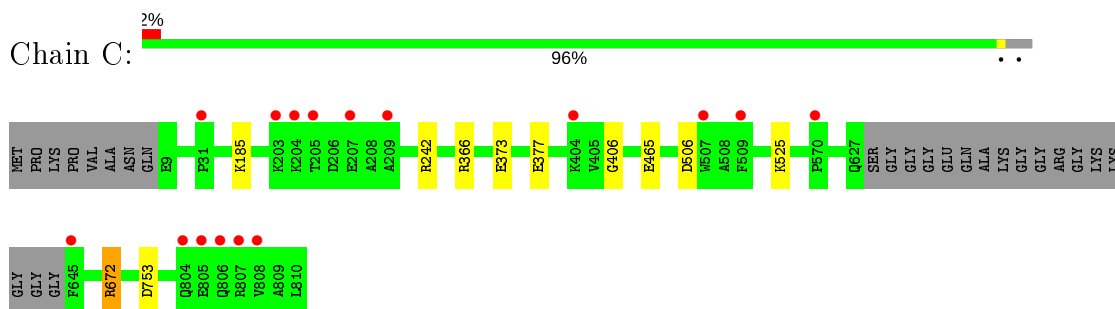
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

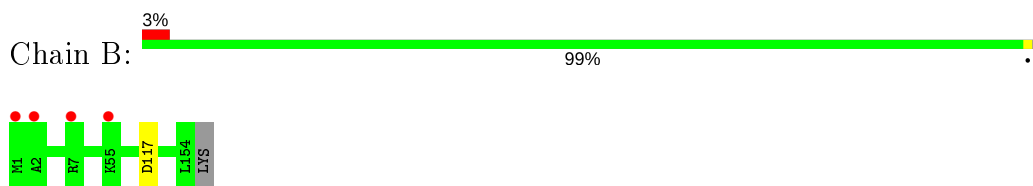
- Molecule 1: Myosin heavy chain, muscle



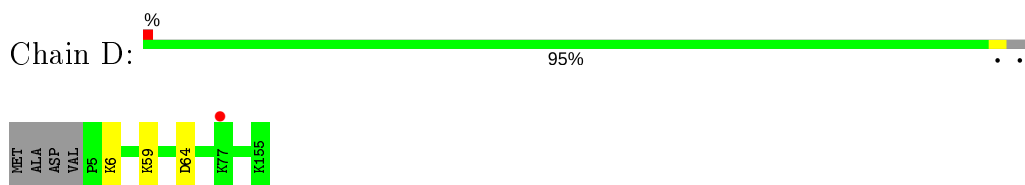
- Molecule 1: Myosin heavy chain, muscle



- Molecule 2: Myosin light chain alkali



- Molecule 2: Myosin light chain alkali



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	108.55Å 148.58Å 148.73Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	105.12 – 2.23 105.12 – 2.23	Depositor EDS
% Data completeness (in resolution range)	98.8 (105.12-2.23) 98.8 (105.12-2.23)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.15 (at 2.22Å)	Xtrriage
Refinement program	PHENIX (dev_2666)	Depositor
R, $R_{free}$	0.178 , 0.220 0.178 , 0.217	Depositor DCC
$R_{free}$ test set	5842 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	30.4	Xtrriage
Anisotropy	0.300	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.39 , 41.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.014 for -h,l,k	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	31091	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	41.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.71% 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 [i](#)

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

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, FLC, EDO

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	0.27	0/6412	0.42	0/8630
1	C	0.27	0/6425	0.43	0/8646
2	B	0.26	0/1236	0.41	0/1665
2	D	0.26	0/1218	0.43	0/1637
All	All	0.27	0/15291	0.42	0/20578

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	A	6281	6247	6245	5	0
1	C	6298	6281	6277	6	0
2	B	1217	1197	1197	1	0
2	D	1199	1181	1181	2	0
3	A	13	5	5	1	0
3	C	13	5	5	1	0
4	A	12	12	16	0	0
4	C	12	15	16	0	0
5	A	12	18	18	0	0

*Continued on next page...*



Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	C	4	6	6	0	0
6	A	430	0	0	2	0
6	B	52	0	0	1	0
6	C	471	0	0	5	0
6	D	110	0	0	2	0
All	All	16124	14967	14966	16	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 1.

All (16) 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:C:753:ASP:OD1	6:C:2101:HOH:O	2.07	0.72
1:C:525:LYS:NZ	6:C:2105:HOH:O	2.24	0.70
3:C:2000:FLC:OG1	6:C:2102:HOH:O	2.11	0.68
3:A:2000:FLC:OHB	3:A:2000:FLC:OG1	2.12	0.65
1:A:498:GLU:OE1	6:A:2101:HOH:O	2.15	0.63
1:C:377:GLU:OE1	6:C:2103:HOH:O	2.16	0.60
2:D:64:ASP:OD1	6:D:201:HOH:O	2.18	0.56
2:D:59:LYS:NZ	6:D:203:HOH:O	2.37	0.49
2:B:117:ASP:OD2	6:B:201:HOH:O	2.19	0.47
1:A:374:GLN:NE2	1:A:377:GLU:OE2	2.48	0.47
1:C:366:ARG:NH2	1:C:373:GLU:OE2	2.49	0.44
1:A:147:ARG:NH1	1:A:149:ASN:OD1	2.52	0.43
1:A:672:ARG:HA	1:A:672:ARG:HD2	1.95	0.42
1:C:672:ARG:NH2	6:C:2142:HOH:O	2.53	0.42
1:A:672:ARG:NH2	6:A:2134:HOH:O	2.52	0.41
1:C:242:ARG:NH2	1:C:465:GLU:OE2	2.47	0.41

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	783/810 (97%)	768 (98%)	14 (2%)	1 (0%)	51	60
1	C	783/810 (97%)	768 (98%)	14 (2%)	1 (0%)	51	60
2	B	152/155 (98%)	151 (99%)	1 (1%)	0	100	100
2	D	149/155 (96%)	146 (98%)	2 (1%)	1 (1%)	22	21
All	All	1867/1930 (97%)	1833 (98%)	31 (2%)	3 (0%)	47	54

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	570	PRO
1	C	406	GLY
2	D	6	LYS

### 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	681/697 (98%)	672 (99%)	9 (1%)	69	80
1	C	684/697 (98%)	681 (100%)	3 (0%)	91	95
2	B	134/135 (99%)	134 (100%)	0	100	100
2	D	132/135 (98%)	132 (100%)	0	100	100
All	All	1631/1664 (98%)	1619 (99%)	12 (1%)	84	91

All (12) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	131	ARG
1	A	205	THR
1	A	212	LYS
1	A	374	GLN
1	A	403	ILE

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	541	LYS
1	A	573	GLN
1	A	585	CYS
1	A	672	ARG
1	C	185	LYS
1	C	506	ASP
1	C	672	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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 monosaccharides in this entry.

### 5.6 Ligand geometry [i](#)

10 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	$\# Z  > 2$
5	EDO	A	2003	-	3,3,3	0.52	0	2,2,2	0.21	0
3	FLC	C	2000	-	3,12,12	1.03	0	3,17,17	2.16	2 (66%)
5	EDO	A	2005	-	3,3,3	0.46	0	2,2,2	0.31	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
3	FLC	A	2000	-	3,12,12	1.15	0	3,17,17	2.57	2 (66%)
4	GOL	C	2002	-	5,5,5	0.92	0	5,5,5	0.93	0
4	GOL	A	2002	-	5,5,5	1.06	0	5,5,5	0.82	0
4	GOL	A	2001	-	5,5,5	0.93	0	5,5,5	0.99	0
4	GOL	C	2001	-	5,5,5	1.00	0	5,5,5	0.69	0
5	EDO	A	2004	-	3,3,3	0.46	0	2,2,2	0.32	0
5	EDO	C	2003	-	3,3,3	0.44	0	2,2,2	0.35	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
5	EDO	A	2003	-	-	0/1/1/1	-
3	FLC	C	2000	-	-	1/6/16/16	-
5	EDO	A	2005	-	-	0/1/1/1	-
3	FLC	A	2000	-	-	3/6/16/16	-
4	GOL	C	2002	-	-	2/4/4/4	-
4	GOL	A	2002	-	-	2/4/4/4	-
4	GOL	A	2001	-	-	0/4/4/4	-
4	GOL	C	2001	-	-	2/4/4/4	-
5	EDO	A	2004	-	-	0/1/1/1	-
5	EDO	C	2003	-	-	0/1/1/1	-

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	2000	FLC	CB-CG-CGC	-3.32	109.67	114.98
3	A	2000	FLC	CB-CA-CAC	-2.96	110.24	114.98
3	C	2000	FLC	CB-CA-CAC	-2.92	110.31	114.98
3	C	2000	FLC	CB-CG-CGC	-2.05	111.70	114.98

There are no chirality outliers.

All (10) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	2000	FLC	CBC-CB-CG-CGC
4	C	2002	GOL	O1-C1-C2-O2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms
4	C	2002	GOL	O1-C1-C2-C3
4	A	2002	GOL	O1-C1-C2-O2
4	A	2002	GOL	O1-C1-C2-C3
4	C	2001	GOL	O1-C1-C2-C3
4	C	2001	GOL	O1-C1-C2-O2
3	A	2000	FLC	OHB-CB-CG-CGC
3	A	2000	FLC	CA-CB-CG-CGC
3	C	2000	FLC	OHB-CB-CG-CGC

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	C	2000	FLC	1	0
3	A	2000	FLC	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	786/810 (97%)	0.05	31 (3%) 39 37	20, 34, 76, 141	0
1	C	785/810 (96%)	-0.10	16 (2%) 65 63	18, 29, 64, 100	0
2	B	154/155 (99%)	0.15	4 (2%) 56 54	29, 44, 75, 115	0
2	D	151/155 (97%)	-0.15	1 (0%) 87 86	20, 30, 53, 73	0
All	All	1876/1930 (97%)	-0.02	52 (2%) 53 51	18, 33, 71, 141	0

All (52) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	642	GLY	16.1
2	B	1	MET	12.8
1	A	571	GLY	7.4
1	A	210	LYS	6.9
1	A	407	ASN	6.8
1	A	367	GLY	6.7
1	A	572	GLN	6.3
1	A	643	GLY	6.0
1	A	645	PHE	5.8
2	B	2	ALA	5.7
1	A	570	PRO	5.2
1	A	368	ARG	4.9
1	A	209	ALA	4.8
1	C	807	ARG	4.6
1	C	509	PHE	4.5
1	A	644	GLY	4.3
1	C	645	PHE	4.2
1	A	405	VAL	4.2
1	A	410	VAL	4.0
1	A	403	ILE	3.7
1	C	507	TRP	3.6

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	409	PHE	3.6
1	A	406	GLY	3.2
1	A	203	LYS	3.0
1	A	205	THR	3.0
1	C	806	GLN	3.0
1	A	507	TRP	2.8
1	C	203	LYS	2.8
1	C	207	GLU	2.8
1	A	402	ARG	2.8
1	A	8	GLN	2.8
1	C	31	PRO	2.7
1	A	573	GLN	2.6
1	C	209	ALA	2.6
1	C	804	GLN	2.5
1	A	408	GLU	2.5
1	A	366	ARG	2.5
1	C	205	THR	2.4
2	B	7	ARG	2.3
1	A	31	PRO	2.3
1	C	805	GLU	2.3
1	A	404	LYS	2.3
1	C	570	PRO	2.2
1	C	808	VAL	2.2
1	C	404	LYS	2.2
2	B	55	LYS	2.2
1	C	204	LYS	2.2
1	A	627	GLN	2.2
1	A	806	GLN	2.1
1	A	568	PRO	2.1
2	D	77	LYS	2.1
1	A	574	ALA	2.1

## 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 monosaccharides 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(Å <sup>2</sup> )	Q<0.9
5	EDO	A	2005	4/4	0.79	0.17	52,62,70,70	0
4	GOL	C	2001	6/6	0.80	0.24	29,36,54,54	0
5	EDO	A	2003	4/4	0.81	0.16	33,47,57,57	0
3	FLC	C	2000	13/13	0.86	0.17	32,52,65,68	0
4	GOL	C	2002	6/6	0.88	0.19	48,57,63,74	0
3	FLC	A	2000	13/13	0.89	0.16	35,50,60,63	0
4	GOL	A	2002	6/6	0.90	0.23	33,45,54,55	0
5	EDO	C	2003	4/4	0.91	0.11	50,60,61,67	0
5	EDO	A	2004	4/4	0.94	0.34	40,50,60,67	0
4	GOL	A	2001	6/6	0.95	0.13	35,42,48,51	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.