

Full wwPDB X-ray Structure Validation Report (i)

May 22, 2020 – 02:56 pm BST

PDB ID : 4IDI

Title : Crystal Structure of Rurm1-related protein from Plasmodium Yoelii, PY06420 Authors : Wernimont, A.K.; Tempel, W.; Lew, J.; Walker, J.; Arrowsmith, C.H.; Ed-

wards, A.M.; Schapira, M.; Bountra, C.; Hui, R.; Artz, J.D.; Structural Ge-

nomics Consortium (SGC)

Deposited on : 2012-12-12

Resolution : 1.90 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
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) 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) roteins) : Engh & Huber (2001)

Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)

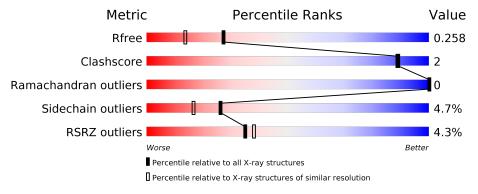
Validation Pipeline (wwPDB-VP) : 2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure: X-RAY DIFFRACTION

The reported resolution of this entry is 1.90 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	$\begin{array}{c} \text{Whole archive} \\ (\#\text{Entries}) \end{array}$	$\begin{array}{c} {\rm Similar\ resolution} \\ (\#{\rm Entries,\ resolution\ range(\AA)}) \end{array}$
R_{free}	130704	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)

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 >=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 <=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	144	74%	8% •	17%
1	В	144	74%	6% •	19%
1	С	144	77%	5% •	17%
1	D	144	74%	7%	19%



2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 4190 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 Oryza sativa Rurm1-related.

Mol	Chain	Residues		Atoms				ZeroOcc	AltConf	Trace	
1	A 110	119	Total	С	N	О	S	Se	0	6	0
1	A	119	1014	658	154	198	3	1	0	0	0
1	В	117	Total	С	N	О	S	Se	0	2	0
1		117	968	632	148	185	2	1	U		
1	1 C 119	110	Total	С	N	О	S	Se	0	4	0
1		119	991	646	150	191	3	1	0	4	0
1	1 D	116	Total	С	N	О	S	Se	0	5	0
	116	973	636	145	189	2	1	U	Э	0	

There are 80 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-18	MSE	-	EXPRESSION TAG	UNP Q7RAS9
A	-17	GLY	-	EXPRESSION TAG	UNP Q7RAS9
A	-16	SER	-	EXPRESSION TAG	UNP Q7RAS9
A	-15	SER	-	EXPRESSION TAG	UNP Q7RAS9
A	-14	HIS	-	EXPRESSION TAG	UNP Q7RAS9
A	-13	HIS	-	EXPRESSION TAG	UNP Q7RAS9
A	-12	HIS	-	EXPRESSION TAG	UNP Q7RAS9
A	-11	HIS	-	EXPRESSION TAG	UNP Q7RAS9
A	-10	HIS	_	EXPRESSION TAG	UNP Q7RAS9
A	-9	HIS	-	EXPRESSION TAG	UNP Q7RAS9
A	-8	SER	_	EXPRESSION TAG	UNP Q7RAS9
A	-7	SER	_	EXPRESSION TAG	UNP Q7RAS9
A	-6	GLY	_	EXPRESSION TAG	UNP Q7RAS9
A	-5	LEU	_	EXPRESSION TAG	UNP Q7RAS9
A	-4	VAL	-	EXPRESSION TAG	UNP Q7RAS9
A	-3	PRO	_	EXPRESSION TAG	UNP Q7RAS9
A	-2	ARG	=	EXPRESSION TAG	UNP Q7RAS9
A	-1	GLY	-	EXPRESSION TAG	UNP Q7RAS9
A	0	SER	-	EXPRESSION TAG	UNP Q7RAS9
A	70	MSE	LEU	ENGINEERED MUTATION	UNP Q7RAS9
В	-18	MSE	_	EXPRESSION TAG	UNP Q7RAS9

Continued on next page...



Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
В	-17	GLY	-	EXPRESSION TAG	UNP Q7RAS9
В	-16	SER	-	EXPRESSION TAG	UNP Q7RAS9
В	-15	SER	_	EXPRESSION TAG	UNP Q7RAS9
В	-14	HIS	-	EXPRESSION TAG	UNP Q7RAS9
В	-13	HIS	-	EXPRESSION TAG	UNP Q7RAS9
В	-12	HIS	_	EXPRESSION TAG	UNP Q7RAS9
В	-11	HIS	_	EXPRESSION TAG	UNP Q7RAS9
В	-10	HIS	-	EXPRESSION TAG	UNP Q7RAS9
В	-9	HIS	-	EXPRESSION TAG	UNP Q7RAS9
В	-8	SER	_	EXPRESSION TAG	UNP Q7RAS9
В	-7	SER	-	EXPRESSION TAG	UNP Q7RAS9
В	-6	GLY	-	EXPRESSION TAG	UNP Q7RAS9
В	-5	LEU	_	EXPRESSION TAG	UNP Q7RAS9
В	-4	VAL	-	EXPRESSION TAG	UNP Q7RAS9
В	-3	PRO	-	EXPRESSION TAG	UNP Q7RAS9
В	-2	ARG	_	EXPRESSION TAG	UNP Q7RAS9
В	-1	GLY	-	EXPRESSION TAG	UNP Q7RAS9
В	0	SER	_	EXPRESSION TAG	UNP Q7RAS9
В	70	MSE	LEU	ENGINEERED MUTATION	UNP Q7RAS9
С	-18	MSE	-	EXPRESSION TAG	UNP Q7RAS9
С	-17	GLY	-	EXPRESSION TAG	UNP Q7RAS9
С	-16	SER	-	EXPRESSION TAG	UNP Q7RAS9
С	-15	SER	-	EXPRESSION TAG	UNP Q7RAS9
С	-14	HIS	-	EXPRESSION TAG	UNP Q7RAS9
С	-13	HIS	_	EXPRESSION TAG	UNP Q7RAS9
С	-12	HIS	-	EXPRESSION TAG	UNP Q7RAS9
С	-11	HIS	-	EXPRESSION TAG	UNP Q7RAS9
С	-10	HIS	-	EXPRESSION TAG	UNP Q7RAS9
С	-9	HIS	_	EXPRESSION TAG	UNP Q7RAS9
С	-8	SER	_	EXPRESSION TAG	UNP Q7RAS9
С	-7	SER	_	EXPRESSION TAG	UNP Q7RAS9
С	-6	GLY	_	EXPRESSION TAG	UNP Q7RAS9
С	-5	LEU	_	EXPRESSION TAG	UNP Q7RAS9
С	-4	VAL	_	EXPRESSION TAG	UNP Q7RAS9
С	-3	PRO	-	EXPRESSION TAG	UNP Q7RAS9
С	-2	ARG	-	EXPRESSION TAG	UNP Q7RAS9
С	-1	GLY		EXPRESSION TAG	UNP Q7RAS9
С	0	SER	-	EXPRESSION TAG	UNP Q7RAS9
С	70	MSE	LEU	ENGINEERED MUTATION	UNP Q7RAS9
D	-18	MSE	-	EXPRESSION TAG	UNP Q7RAS9
D	-17	GLY	-	EXPRESSION TAG	UNP Q7RAS9
D	-16	SER	-	EXPRESSION TAG	UNP Q7RAS9

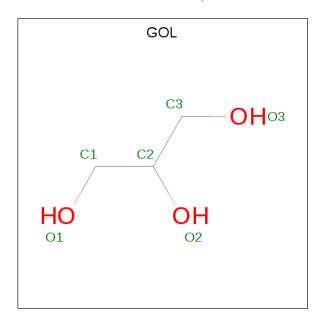
Continued on next page...



Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	-15	SER	_	EXPRESSION TAG	UNP Q7RAS9
D	-14	HIS	-	EXPRESSION TAG	UNP Q7RAS9
D	-13	HIS	_	EXPRESSION TAG	UNP Q7RAS9
D	-12	HIS	_	EXPRESSION TAG	UNP Q7RAS9
D	-11	HIS	_	EXPRESSION TAG	UNP Q7RAS9
D	-10	HIS	_	EXPRESSION TAG	UNP Q7RAS9
D	-9	HIS	_	EXPRESSION TAG	UNP Q7RAS9
D	-8	SER	_	EXPRESSION TAG	UNP Q7RAS9
D	-7	SER	_	EXPRESSION TAG	UNP Q7RAS9
D	-6	GLY	_	EXPRESSION TAG	UNP Q7RAS9
D	-5	LEU	_	EXPRESSION TAG	UNP Q7RAS9
D	-4	VAL	_	EXPRESSION TAG	UNP Q7RAS9
D	-3	PRO	_	EXPRESSION TAG	UNP Q7RAS9
D	-2	ARG	-	EXPRESSION TAG	UNP Q7RAS9
D	-1	GLY	-	EXPRESSION TAG	UNP Q7RAS9
D	0	SER	-	EXPRESSION TAG	UNP Q7RAS9
D	70	MSE	LEU	ENGINEERED MUTATION	UNP Q7RAS9

 \bullet Molecule 2 is GLYCEROL (three-letter code: GOL) (formula: $\mathrm{C_3H_8O_3}).$



ľ	Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
	2	В	1	Total C O 6 3 3	0	0

• Molecule 3 is water.



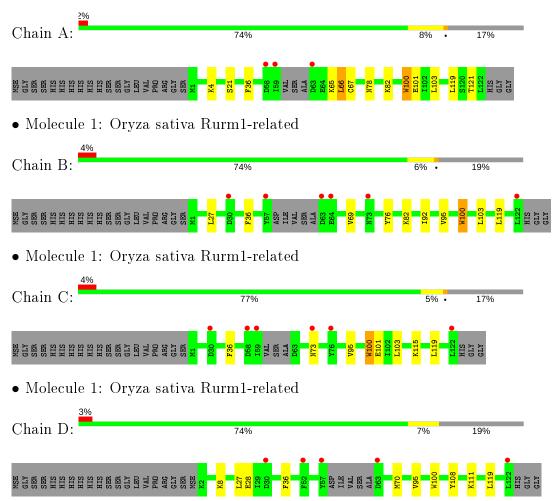
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	62	Total O 62 62	0	0
3	В	57	Total O 57 57	0	0
3	С	60	Total O 60 60	0	0
3	D	59	Total O 59 59	0	0



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: Oryza sativa Rurm1-related





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants	$32.59 ext{Å}$ $72.72 ext{Å}$ $107.51 ext{Å}$	Depositor
a, b, c, α , β , γ	90.00° 92.64° 90.00°	Depositor
Resolution (Å)	43.20 - 1.90	Depositor
Resolution (A)	43.20 - 1.90	EDS
% Data completeness	98.8 (43.20-1.90)	Depositor
(in resolution range)	98.9 (43.20-1.90)	EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) > 1$	$1.29 \; ({\rm at} \; 1.89 {\rm \AA})$	Xtriage
Refinement program	BUSTER-TNT, BUSTER 2.10.0	Depositor
D.D.	0.227 , 0.252	Depositor
R, R_{free}	0.240 , 0.258	DCC
R_{free} test set	1971 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å ²)	22.4	Xtriage
Anisotropy	0.539	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$, $B_{sol}(Å^2)$	0.36, 47.3	EDS
L-test for twinning ²	$< L >=0.50, < L^2>=0.34$	Xtriage
Estimated twinning fraction	0.117 for h,-k,-l	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	4190	wwPDB-VP
Average B, all atoms (Å ²)	30.0	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 43.91 % 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 1.6461e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.

²Theoretical values of <|L|>, $<L^2>$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.



¹Intensities estimated from amplitudes.

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

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		
MIOI	Moi Chain		$\mid \text{RMSZ} \mid \# Z > 5$		# Z > 5	
1	A	0.51	0/1037	0.64	0/1388	
1	В	0.50	0/985	0.65	0/1322	
1	С	0.50	0/1014	0.64	0/1360	
1	D	0.50	0/996	0.62	0/1338	
All	All	0.50	0/4032	0.64	0/5408	

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)	$\mathbf{H}(\mathbf{added})$	Clashes	Symm-Clashes
1	A	1014	0	1017	7	0
1	В	968	0	974	4	0
1	С	991	0	993	5	0
1	D	973	0	965	3	0
2	В	6	0	8	0	0
3	A	62	0	0	0	0
3	В	57	0	0	0	0
3	С	60	0	0	0	0
3	D	59	0	0	1	0
All	All	4190	0	3957	15	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 2.

All (15) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	$egin{array}{ll} ext{Interatomic} \ ext{distance} \ (ext{Å}) \end{array}$	Clash overlap (Å)
1:D:8[A]:LYS:NZ	3:D:252:HOH:O	2.18	0.77
1:A:4:LYS:CE	1:C:115:LYS:HE2	2.36	0.56
1:A:21:SER:HB2	1:C:73:ASN:CG	2.26	0.56
1:A:121:THR:HG22	1:B:92[B]:ILE:HD11	1.94	0.49
1:A:65:LYS:HB2	1:A:82:LYS:HD2	1.93	0.49
1:C:100:TRP:HA	1:C:103:LEU:HD12	1.95	0.49
1:A:100:TRP:HA	1:A:103:LEU:HD12	1.95	0.49
1:B:100:TRP:HA	1:B:103:LEU:HD12	1.97	0.46
1:A:4:LYS:HE2	1:C:115:LYS:HE2	1.96	0.46
1:A:66:LEU:HG	1:A:78:ASN:HA	2.00	0.44
1:D:95:VAL:HG11	1:D:108:TYR:CE2	2.54	0.43
1:B:95:VAL:HB	1:B:103:LEU:CD1	2.50	0.42
1:C:95:VAL:HB	1:C:103:LEU:CD1	2.50	0.41
1:B:69:VAL:HG23	1:B:76:TYR:HB2	2.03	0.40
1:D:28:GLU:HB2	1:D:70:MSE:HB2	2.04	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	Perce	\mathbf{ntiles}
1	A	121/144 (84%)	118 (98%)	3 (2%)	0	100	100
1	В	115/144 (80%)	111 (96%)	4 (4%)	0	100	100
1	С	119/144 (83%)	115 (97%)	4 (3%)	0	100	100
1	D	117/144 (81%)	113 (97%)	4 (3%)	0	100	100
All	All	472/576 (82%)	457 (97%)	15 (3%)	0	100	100



There are no Ramachandran outliers to report.

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	115/129~(89%)	107 (93%)	8 (7%)	15 7
1	В	108/129~(84%)	103 (95%)	5 (5%)	27 17
1	С	111/129 (86%)	107 (96%)	4 (4%)	35 26
1	D	108/129 (84%)	103 (95%)	5 (5%)	27 17
All	All	442/516~(86%)	420 (95%)	22 (5%)	26 15

All (22) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	36	PHE
1	A	66	LEU
1	A	67[A]	CYS
1	A	67[B]	CYS
1	A	100	TRP
1	A	101[A]	GLU
1	A	101[B]	GLU
1	A	119	LEU
1	В	27	LEU
1	В	36	PHE
1	В	82	LYS
1	В	100	TRP
1	В	119	LEU
1	C C C	36	PHE
1	С	100	TRP
1		101	GLU
1	С	119	LEU
1	D	27	LEU
1	D	36	PHE
1	D	100	TRP
1	D	111	LYS
1	D	119	LEU



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 carbohydrates in this entry.

5.6 Ligand geometry (i)

1 ligand is 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	Tuno	Chain	Chain	Chain	Chain	Chain	\mathbf{Res}	$_{ m Res} \mid_{ m Link} \mid$	Bond lengths			Bond angles		
	туре		nes		Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2				
2	2	GOL	В	201	_	5,5,5	0.13	0	5,5,5	0.17	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
2	GOL	В	201	_	-	0/4/4/4	-

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 (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, 95th 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 $>$	$\#\mathrm{RSRZ}{>}2$	$OWAB(A^2)$	Q < 0.9
1	A	117/144 (81%)	0.35	3 (2%) 56 58	16, 27, 56, 65	0
1	В	115/144 (79%)	0.41	6 (5%) 27 30	15, 28, 54, 69	0
1	С	117/144 (81%)	0.41	6 (5%) 28 31	15, 27, 57, 68	0
1	D	115/144 (79%)	0.29	5 (4%) 35 38	16, 27, 47, 59	0
All	All	464/576 (80%)	0.36	20 (4%) 35 38	15, 27, 54, 69	0

All (20) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	С	73	ASN	4.3
1	В	122	LEU	3.8
1	D	122	LEU	3.6
1	В	63	ASP	3.3
1	В	30	ASP	3.1
1	В	73	ASN	3.0
1	В	64	GLU	2.9
1	С	122	LEU	2.9
1	A	63	ASP	2.8
1	С	59	ILE	2.8
1	D	63	ASP	2.6
1	С	30	ASP	2.5
1	В	57	TYR	2.5
1	A	58	ASP	2.4
1	D	30	ASP	2.3
1	С	76	TYR	2.2
1	D	57	TYR	2.1
1	A	59	ILE	2.1
1	С	58	ASP	2.1
1	D	52	PHE	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. The B-factors column lists the minimum, median, 95^{th} 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	$\mathbf{B} ext{-}\mathbf{factors}(\mathbf{\mathring{A}}^2)$	Q < 0.9
2	GOL	В	201	6/6	0.74	0.18	37,42,45,48	0

6.5 Other polymers (i)

There are no such residues in this entry.

