



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 17, 2020 – 11:40 am BST

PDB ID : 2CV4  
Title : Crystal Structure of an Archaeal Peroxiredoxin from the Aerobic Hyperthermophilic Crenarchaeon Aeropyrum pernix K1  
Authors : Mizohata, E.; Sakai, H.; Fusatomi, E.; Terada, T.; Murayama, K.; Shirouzu, M.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)  
Deposited on : 2005-05-31  
Resolution : 2.30 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 : 1.8.5 (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

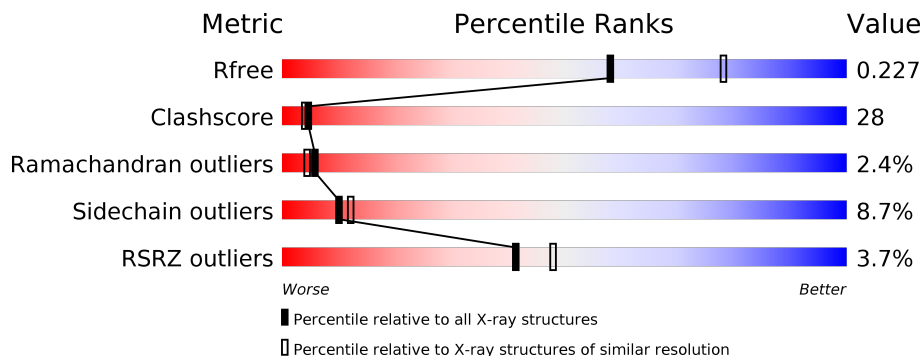
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.30 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	5042 (2.30-2.30)
Clashscore	141614	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)
RSRZ outliers	127900	4938 (2.30-2.30)

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	250	
1	B	250	
1	C	250	
1	D	250	
1	E	250	
1	F	250	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	G	250	
1	H	250	
1	I	250	
1	J	250	

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 criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
1	OCS	H	50	-	-	X	-
3	IPA	B	2005	-	-	X	-
3	IPA	D	2003	-	-	X	-
3	IPA	H	2009	-	-	X	-
3	IPA	J	2001	-	-	X	-

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 20374 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 peroxiredoxin.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	241	1953	1255	344	347	3	4	0	0	0
1	B	241	1953	1255	344	347	3	4	0	0	0
1	C	241	1953	1255	344	347	3	4	0	0	0
1	D	242	1962	1260	345	350	3	4	0	0	0
1	E	240	1941	1246	343	345	3	4	0	0	0
1	F	240	1941	1246	343	345	3	4	0	0	0
1	G	242	1962	1260	345	350	3	4	0	0	0
1	H	240	1941	1246	343	345	3	4	0	0	0
1	I	241	1953	1255	344	347	3	4	0	0	0
1	J	240	1941	1246	343	345	3	4	0	0	0

There are 60 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
A	15	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
A	50	OCS	CYS	MODIFIED RESIDUE	UNP Q9Y9L0
A	140	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
A	145	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
A	200	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
B	1	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
B	15	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
B	50	OCS	CYS	MODIFIED RESIDUE	UNP Q9Y9L0

*Continued on next page...*

*Continued from previous page...*

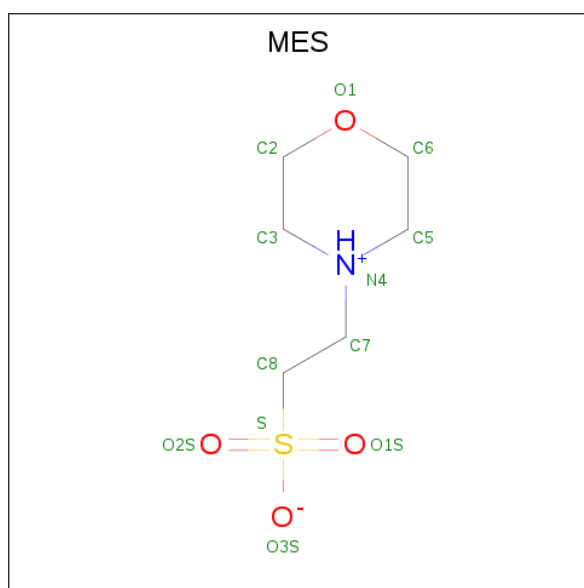
Chain	Residue	Modelled	Actual	Comment	Reference
B	140	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
B	145	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
B	200	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
C	1	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
C	15	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
C	50	OCS	CYS	MODIFIED RESIDUE	UNP Q9Y9L0
C	140	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
C	145	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
C	200	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
D	1	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
D	15	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
D	50	OCS	CYS	MODIFIED RESIDUE	UNP Q9Y9L0
D	140	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
D	145	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
D	200	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
E	1	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
E	15	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
E	50	OCS	CYS	MODIFIED RESIDUE	UNP Q9Y9L0
E	140	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
E	145	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
E	200	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
F	1	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
F	15	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
F	50	OCS	CYS	MODIFIED RESIDUE	UNP Q9Y9L0
F	140	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
F	145	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
F	200	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
G	1	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
G	15	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
G	50	OCS	CYS	MODIFIED RESIDUE	UNP Q9Y9L0
G	140	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
G	145	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
G	200	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
H	1	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
H	15	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
H	50	OCS	CYS	MODIFIED RESIDUE	UNP Q9Y9L0
H	140	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
H	145	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
H	200	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
I	1	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
I	15	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
I	50	OCS	CYS	MODIFIED RESIDUE	UNP Q9Y9L0

*Continued on next page...*

Continued from previous page...

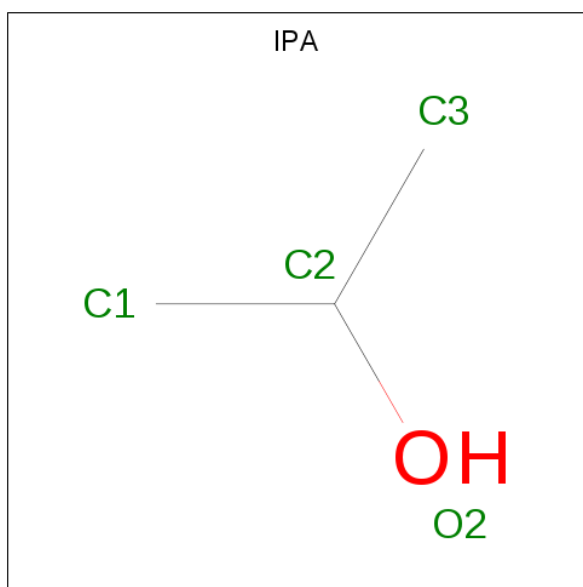
Chain	Residue	Modelled	Actual	Comment	Reference
I	140	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
I	145	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
I	200	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
J	1	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
J	15	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
J	50	OCS	CYS	MODIFIED RESIDUE	UNP Q9Y9L0
J	140	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
J	145	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0
J	200	MSE	MET	MODIFIED RESIDUE	UNP Q9Y9L0

- Molecule 2 is 2-(N-MORPHOLINO)-ETHANESULFONIC ACID (three-letter code: MES) (formula: C<sub>6</sub>H<sub>13</sub>NO<sub>4</sub>S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	N	O	S	0	0
			12	6	1	4	1		
2	F	1	Total	C	N	O	S	0	0
			12	6	1	4	1		

- Molecule 3 is ISOPROPYL ALCOHOL (three-letter code: IPA) (formula: C<sub>3</sub>H<sub>8</sub>O).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total C O 4 3 1	0	0
3	D	1	Total C O 4 3 1	0	0
3	E	1	Total C O 4 3 1	0	0
3	F	1	Total C O 4 3 1	0	0
3	G	1	Total C O 4 3 1	0	0
3	H	1	Total C O 4 3 1	0	0
3	H	1	Total C O 4 3 1	0	0
3	J	1	Total C O 4 3 1	0	0
3	J	1	Total C O 4 3 1	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	123	Total O 123 123	0	0
4	B	69	Total O 69 69	0	0
4	C	70	Total O 70 70	0	0

*Continued on next page...*

*Continued from previous page...*

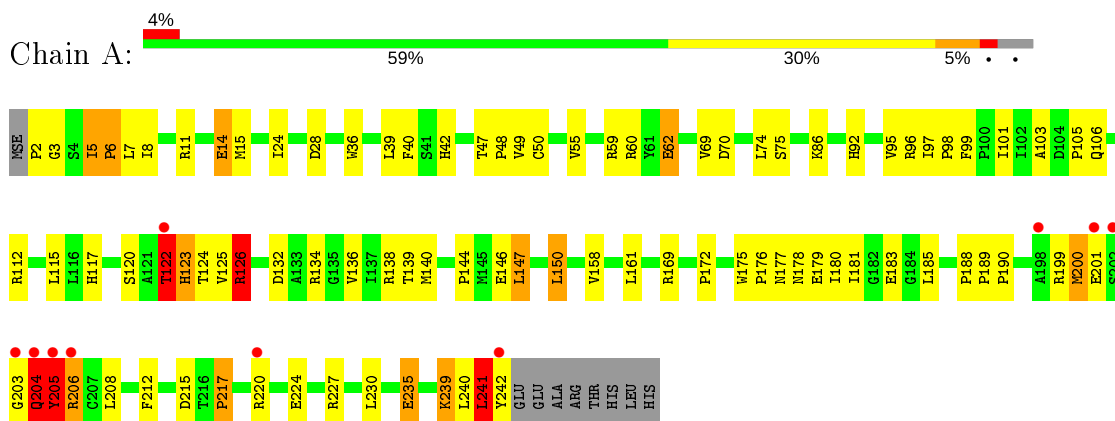
<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
4	D	66	Total 66	O 66	0	0
4	E	70	Total 70	O 70	0	0
4	F	75	Total 75	O 75	0	0
4	G	85	Total 85	O 85	0	0
4	H	82	Total 82	O 82	0	0
4	I	82	Total 82	O 82	0	0
4	J	92	Total 92	O 92	0	0



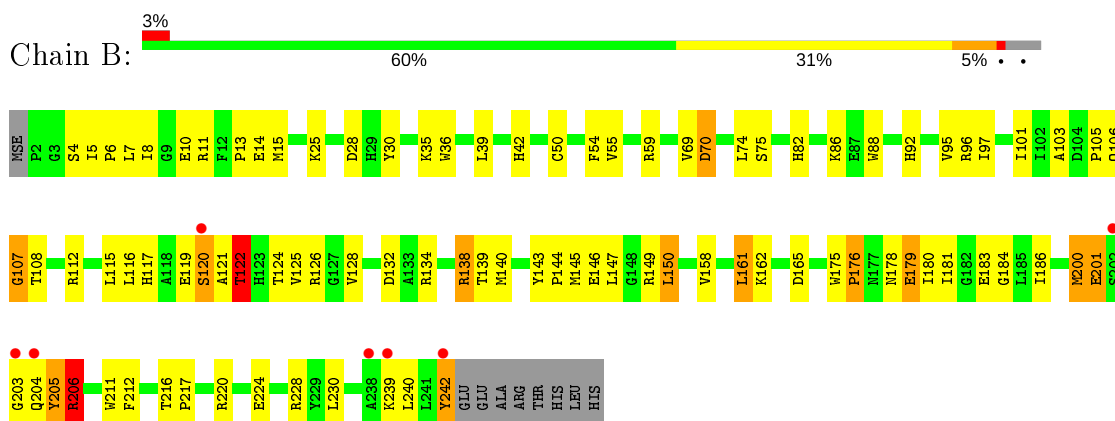
### 3 Residue-property plots [i](#)

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.

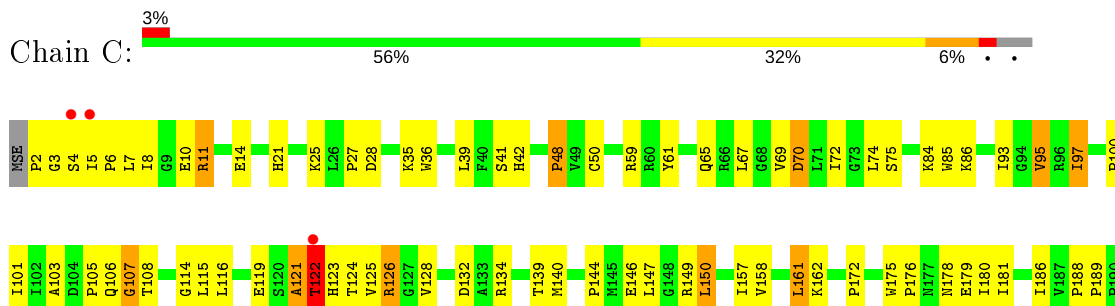
- Molecule 1: peroxiredoxin

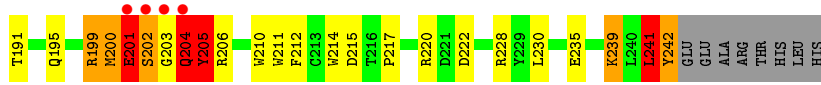


- Molecule 1: peroxiredoxin

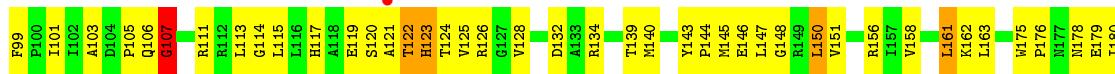
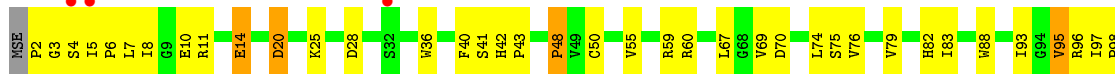


- Molecule 1: peroxiredoxin

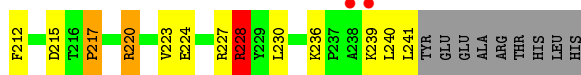
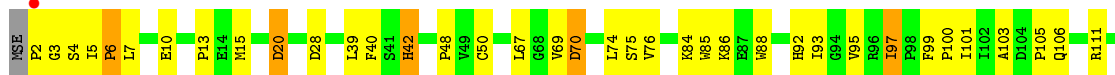




• Molecule 1: peroxiredoxin



• Molecule 1: peroxiredoxin

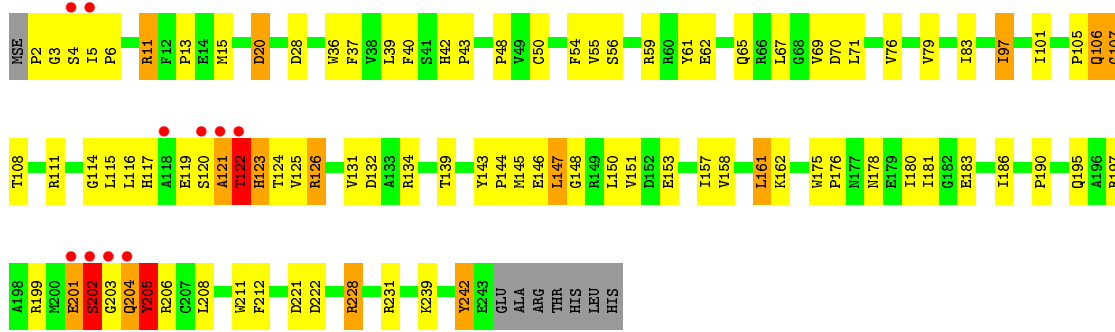


• Molecule 1: peroxiredoxin



• Molecule 1: peroxiredoxin

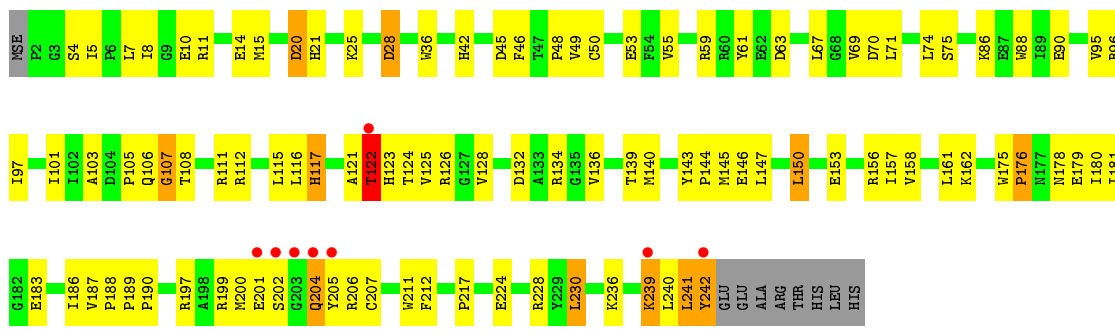




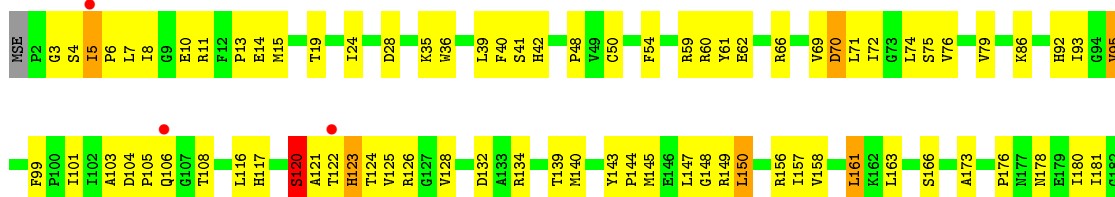
• Molecule 1: peroxiredoxin



• Molecule 1: peroxiredoxin



• Molecule 1: peroxiredoxin





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	76.74Å 103.43Å 102.81Å 105.18° 92.91° 105.39°	Depositor
Resolution (Å)	49.20 – 2.30 49.20 – 2.31	Depositor EDS
% Data completeness (in resolution range)	93.4 (49.20-2.30) 93.6 (49.20-2.31)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.45 (at 2.32Å)	Xtrriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.178 , 0.230 0.176 , 0.227	Depositor DCC
$R_{free}$ test set	6058 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	28.4	Xtrriage
Anisotropy	0.306	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 61.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	20374	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	36.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.18% 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: OCS, IPA, MES

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	0.79	1/1994 (0.1%)	1.00	8/2703 (0.3%)
1	B	0.76	0/1994	0.91	3/2703 (0.1%)
1	C	0.77	0/1994	0.98	4/2703 (0.1%)
1	D	0.76	0/2003	1.02	8/2715 (0.3%)
1	E	0.73	0/1981	0.95	9/2685 (0.3%)
1	F	0.75	0/1981	1.01	11/2685 (0.4%)
1	G	0.75	0/2003	0.97	9/2715 (0.3%)
1	H	0.74	0/1981	0.92	3/2685 (0.1%)
1	I	0.72	0/1994	0.89	1/2703 (0.0%)
1	J	0.71	0/1981	0.96	4/2685 (0.1%)
All	All	0.75	1/19906 (0.0%)	0.96	60/26982 (0.2%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1
1	B	0	1
1	C	0	1
1	E	0	1
1	F	0	1
All	All	0	5

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	126	ARG	CB-CG	-5.33	1.38	1.52

The worst 5 of 60 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	J	203	GLY	N-CA-C	-9.93	88.28	113.10
1	E	202	SER	N-CA-C	8.54	134.05	111.00
1	J	204	GLN	N-CA-C	7.68	131.74	111.00
1	A	241	LEU	CA-CB-CG	7.68	132.97	115.30
1	F	202	SER	CB-CA-C	7.67	124.68	110.10

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	205	TYR	Sidechain
1	B	30	TYR	Sidechain
1	C	205	TYR	Sidechain
1	E	205	TYR	Sidechain
1	F	50	OCS	Mainchain

## 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	1953	0	1939	115	0
1	B	1953	0	1940	108	0
1	C	1953	0	1940	131	0
1	D	1962	0	1946	151	0
1	E	1941	0	1931	115	0
1	F	1941	0	1931	112	0
1	G	1962	0	1946	118	1
1	H	1941	0	1931	131	0
1	I	1953	0	1940	121	0
1	J	1941	0	1931	145	0
2	A	12	0	13	1	0
2	F	12	0	13	0	0
3	B	4	0	8	6	0
3	D	4	0	8	8	0
3	E	4	0	8	3	0
3	F	4	0	8	1	0
3	G	4	0	8	0	0
3	H	8	0	16	7	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	J	8	0	16	7	0
4	A	123	0	0	7	0
4	B	69	0	0	4	1
4	C	70	0	0	7	0
4	D	66	0	0	5	0
4	E	70	0	0	6	0
4	F	75	0	0	8	0
4	G	85	0	0	8	0
4	H	82	0	0	9	0
4	I	82	0	0	6	0
4	J	92	0	0	11	0
All	All	20374	0	19473	1101	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 28.

The worst 5 of 1101 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:H:199:ARG:O	1:H:200:MSE:HE3	1.41	1.20
1:D:202:SER:C	1:D:204:GLN:H	1.49	1.11
1:G:121:ALA:O	1:G:122:THR:HG23	1.48	1.11
1:I:117:HIS:HB2	1:I:125:VAL:CG2	1.84	1.08
1:J:200:MSE:HE3	1:J:203:GLY:HA3	1.30	1.07

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:G:231:ARG:NH2	4:B:2033:HOH:O[1_544]	2.09	0.11

## 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	238/250 (95%)	221 (93%)	13 (6%)	4 (2%)	9	8
1	B	238/250 (95%)	222 (93%)	12 (5%)	4 (2%)	9	8
1	C	238/250 (95%)	223 (94%)	7 (3%)	8 (3%)	3	2
1	D	239/250 (96%)	222 (93%)	8 (3%)	9 (4%)	3	1
1	E	237/250 (95%)	222 (94%)	8 (3%)	7 (3%)	4	2
1	F	237/250 (95%)	224 (94%)	8 (3%)	5 (2%)	7	5
1	G	239/250 (96%)	222 (93%)	11 (5%)	6 (2%)	5	4
1	H	237/250 (95%)	223 (94%)	9 (4%)	5 (2%)	7	5
1	I	238/250 (95%)	223 (94%)	11 (5%)	4 (2%)	9	8
1	J	237/250 (95%)	214 (90%)	18 (8%)	5 (2%)	7	5
All	All	2378/2500 (95%)	2216 (93%)	105 (4%)	57 (2%)	6	4

5 of 57 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	204	GLN
1	C	107	GLY
1	C	201	GLU
1	C	202	SER
1	C	205	TYR

### 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	207/210 (99%)	184 (89%)	23 (11%)	6	7
1	B	207/210 (99%)	187 (90%)	20 (10%)	8	9
1	C	207/210 (99%)	188 (91%)	19 (9%)	9	11
1	D	208/210 (99%)	191 (92%)	17 (8%)	11	14
1	E	206/210 (98%)	189 (92%)	17 (8%)	11	14

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	F	206/210 (98%)	189 (92%)	17 (8%)	11	14
1	G	208/210 (99%)	193 (93%)	15 (7%)	14	18
1	H	206/210 (98%)	190 (92%)	16 (8%)	12	16
1	I	207/210 (99%)	188 (91%)	19 (9%)	9	11
1	J	206/210 (98%)	190 (92%)	16 (8%)	12	16
All	All	2068/2100 (98%)	1889 (91%)	179 (9%)	10	12

5 of 179 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	28	ASP
1	F	59	ARG
1	J	28	ASP
1	E	97	ILE
1	E	212	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 54 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	178	ASN
1	F	195	GLN
1	J	92	HIS
1	E	195	GLN
1	F	42	HIS

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

10 non-standard protein/DNA/RNA residues 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
1	OCS	H	50	1	7,8,9	2.65	5 (71%)	6,11,13	1.93	2 (33%)
1	OCS	E	50	1	7,8,9	3.37	3 (42%)	6,11,13	3.43	3 (50%)
1	OCS	J	50	1	7,8,9	2.68	5 (71%)	6,11,13	1.25	0
1	OCS	I	50	1	7,8,9	3.36	5 (71%)	6,11,13	3.45	3 (50%)
1	OCS	C	50	1	7,8,9	3.65	4 (57%)	6,11,13	2.07	3 (50%)
1	OCS	B	50	1	7,8,9	2.30	2 (28%)	6,11,13	2.66	2 (33%)
1	OCS	G	50	1	7,8,9	1.52	1 (14%)	6,11,13	2.08	4 (66%)
1	OCS	D	50	1	7,8,9	1.99	2 (28%)	6,11,13	2.10	3 (50%)
1	OCS	A	50	1	7,8,9	1.94	2 (28%)	6,11,13	1.75	1 (16%)
1	OCS	F	50	1	7,8,9	1.93	3 (42%)	6,11,13	3.16	3 (50%)

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
1	OCS	H	50	1	-	0/4/7/9	-
1	OCS	E	50	1	-	0/4/7/9	-
1	OCS	J	50	1	-	0/4/7/9	-
1	OCS	I	50	1	-	0/4/7/9	-
1	OCS	C	50	1	-	0/4/7/9	-
1	OCS	B	50	1	-	0/4/7/9	-
1	OCS	G	50	1	-	0/4/7/9	-
1	OCS	D	50	1	-	0/4/7/9	-
1	OCS	A	50	1	-	0/4/7/9	-
1	OCS	F	50	1	-	0/4/7/9	-

The worst 5 of 32 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	50	OCS	CB-CA	-6.98	1.47	1.53
1	I	50	OCS	OD1-SG	6.16	1.63	1.45
1	C	50	OCS	OD1-SG	5.92	1.62	1.45
1	J	50	OCS	OD1-SG	4.74	1.59	1.45
1	C	50	OCS	OD3-SG	4.71	1.58	1.45

The worst 5 of 24 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	I	50	OCS	OD2-SG-CB	-7.13	94.38	105.74
1	B	50	OCS	OD2-SG-CB	-5.86	96.40	105.74
1	E	50	OCS	OD2-SG-CB	-5.70	96.65	105.74
1	E	50	OCS	OD3-SG-CB	5.06	112.95	106.94
1	F	50	OCS	OD3-SG-CB	5.01	112.89	106.94

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

10 monomers are involved in 25 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	H	50	OCS	6	0
1	E	50	OCS	2	0
1	J	50	OCS	2	0
1	I	50	OCS	2	0
1	C	50	OCS	1	0
1	B	50	OCS	1	0
1	G	50	OCS	3	0
1	D	50	OCS	3	0
1	A	50	OCS	2	0
1	F	50	OCS	3	0

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

11 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	IPA	D	2003	-	3,3,3	0.33	0	3,3,3	1.82	1 (33%)
2	MES	A	1001	-	12,12,12	0.90	0	14,16,16	0.74	0
3	IPA	G	2008	-	3,3,3	0.34	0	3,3,3	1.82	1 (33%)
3	IPA	J	2002	-	3,3,3	0.39	0	3,3,3	1.48	1 (33%)
3	IPA	F	2007	-	3,3,3	0.77	0	3,3,3	1.47	1 (33%)
3	IPA	J	2001	-	3,3,3	0.47	0	3,3,3	1.68	1 (33%)
3	IPA	E	2006	-	3,3,3	0.61	0	3,3,3	1.73	1 (33%)
3	IPA	B	2005	-	3,3,3	0.60	0	3,3,3	1.53	1 (33%)
3	IPA	H	2004	-	3,3,3	0.50	0	3,3,3	1.83	1 (33%)
2	MES	F	1002	-	12,12,12	1.12	0	14,16,16	1.14	1 (7%)
3	IPA	H	2009	-	3,3,3	0.62	0	3,3,3	1.71	1 (33%)

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	MES	F	1002	-	-	1/6/14/14	0/1/1/1
2	MES	A	1001	-	-	2/6/14/14	0/1/1/1

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	G	2008	IPA	C3-C2-C1	-3.10	89.98	113.47
2	F	1002	MES	O1S-S-C8	-3.04	103.25	106.92
3	D	2003	IPA	C3-C2-C1	-3.02	90.62	113.47
3	H	2004	IPA	C3-C2-C1	-2.99	90.83	113.47
3	E	2006	IPA	C3-C2-C1	-2.93	91.27	113.47

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	1001	MES	C8-C7-N4-C5
2	F	1002	MES	C8-C7-N4-C5
2	A	1001	MES	C7-C8-S-O1S

There are no ring outliers.

9 monomers are involved in 33 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	D	2003	IPA	8	0
2	A	1001	MES	1	0
3	J	2002	IPA	3	0
3	F	2007	IPA	1	0
3	J	2001	IPA	4	0
3	E	2006	IPA	3	0
3	B	2005	IPA	6	0
3	H	2004	IPA	1	0
3	H	2009	IPA	6	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	236/250 (94%)	-0.41	10 (4%) 36 43	15, 27, 63, 101	0
1	B	236/250 (94%)	-0.39	7 (2%) 50 57	14, 27, 65, 98	0
1	C	236/250 (94%)	-0.35	7 (2%) 50 57	17, 28, 68, 111	0
1	D	237/250 (94%)	-0.23	11 (4%) 32 39	18, 31, 70, 102	0
1	E	235/250 (94%)	-0.27	8 (3%) 45 52	18, 33, 69, 107	0
1	F	235/250 (94%)	-0.38	9 (3%) 40 47	17, 29, 74, 100	0
1	G	237/250 (94%)	-0.28	10 (4%) 36 43	18, 31, 76, 106	0
1	H	235/250 (94%)	-0.32	9 (3%) 40 47	18, 29, 76, 102	0
1	I	236/250 (94%)	-0.23	8 (3%) 45 52	17, 32, 70, 112	0
1	J	235/250 (94%)	-0.25	9 (3%) 40 47	17, 33, 70, 109	0
All	All	2358/2500 (94%)	-0.31	88 (3%) 41 48	14, 30, 71, 112	0

The worst 5 of 88 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	202	SER	11.0
1	B	204	GLN	7.0
1	A	204	GLN	6.9
1	G	4	SER	6.3
1	I	204	GLN	6.2

### 6.2 Non-standard residues in protein, DNA, RNA chains [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
1	OCS	I	50	9/10	0.97	0.14	27,29,37,37	0
1	OCS	J	50	9/10	0.98	0.11	23,28,36,38	0
1	OCS	H	50	9/10	0.98	0.09	24,27,30,40	0
1	OCS	C	50	9/10	0.98	0.12	24,25,34,36	0
1	OCS	B	50	9/10	0.98	0.09	23,25,33,34	0
1	OCS	G	50	9/10	0.98	0.09	23,27,37,47	0
1	OCS	D	50	9/10	0.98	0.10	16,26,38,43	0
1	OCS	A	50	9/10	0.98	0.09	20,24,36,41	0
1	OCS	F	50	9/10	0.98	0.08	20,23,35,36	0
1	OCS	E	50	9/10	0.99	0.09	22,27,37,42	0

### 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	IPA	F	2007	4/4	0.84	0.21	39,40,43,45	0
3	IPA	H	2004	4/4	0.85	0.18	55,61,61,63	0
3	IPA	E	2006	4/4	0.86	0.20	52,57,61,62	0
2	MES	F	1002	12/12	0.88	0.18	70,75,82,84	0
3	IPA	B	2005	4/4	0.94	0.10	40,42,46,47	0
2	MES	A	1001	12/12	0.94	0.15	41,51,65,66	0
3	IPA	J	2002	4/4	0.94	0.16	38,39,39,41	0
3	IPA	D	2003	4/4	0.95	0.11	33,38,40,44	0
3	IPA	J	2001	4/4	0.96	0.10	32,38,40,42	0
3	IPA	G	2008	4/4	0.96	0.13	31,34,42,46	0
3	IPA	H	2009	4/4	0.96	0.10	38,46,47,53	0

### 6.5 Other polymers [i](#)

There are no such residues in this entry.