



Full wwPDB NMR Structure Validation Report ⓘ

Apr 27, 2016 – 03:22 AM BST

PDB ID : 2MU2
Title : NMR structure of the cap domain of NP_346487.1, a putative phosphoglycolate phosphatase from *Streptococcus pneumoniae* TIGR4
Authors : Jaudzems, K.; Serrano, P.; Pedrini, B.; Geralt, M.; Wuthrich, K.; Joint Center for Structural Genomics (JCSG)
Deposited on : 2014-09-03

This is a Full wwPDB NMR Structure Validation Report for a publicly released PDB entry.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
<http://wwpdb.org/validation/2016/NMRValidationReportHelp>
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:

Cyrange : Kirchner and Güntert (2011)
NmrClust : Kelley et al. (1996)
MolProbity : 4.02b-467
Mogul : unknown
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI : v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV : Wang et al. (2010)
ShiftChecker : rb-20027457
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20027457

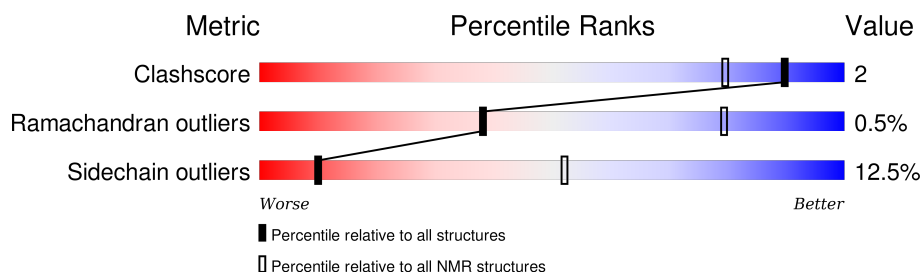
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

SOLUTION NMR

The overall completeness of chemical shifts assignment is 76%.

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)	NMR archive (#Entries)
Clashscore	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

The table below summarises the geometric issues observed across the polymeric chains and their fit to the experimental data. The red, orange, yellow and green segments indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. A cyan segment indicates the fraction of residues that are not part of the well-defined cores, and 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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	208	

2 Ensemble composition and analysis

This entry contains 20 models. Model 7 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average*.

The following residues are included in the computation of the global validation metrics.

Well-defined (core) protein residues			
Well-defined core	Residue range (total)	Backbone RMSD (Å)	Medoid model
1	A:21-A:74 (54)	0.30	7

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

The models can be grouped into 3 clusters and 2 single-model clusters were found.

Cluster number	Models
1	1, 2, 3, 5, 6, 7, 9, 10, 11, 14, 15, 16, 20
2	13, 17, 19
3	4, 12
Single-model clusters	8; 18

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1156 atoms, of which 578 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Hydrolase, haloacid dehalogenase-like family.

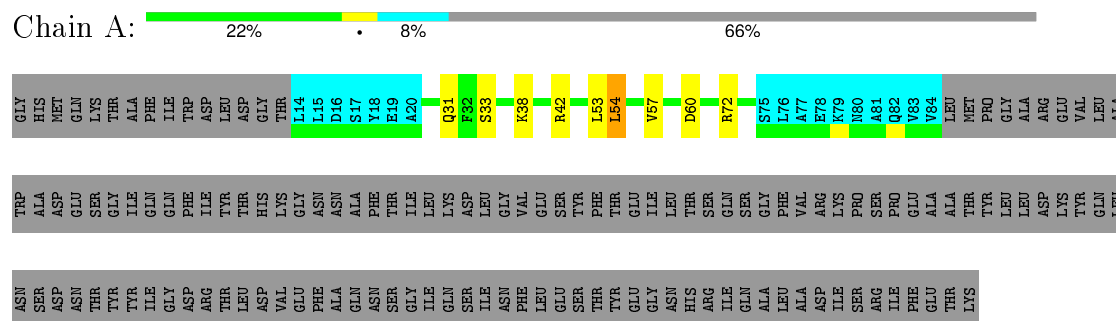
Mol	Chain	Residues	Atoms					Trace
1	A	71	Total	C	H	N	O	0
			1156	369	578	95	114	

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	EXPRESSION TAG	UNP A0A0H2URV7
A	0	HIS	-	EXPRESSION TAG	UNP A0A0H2URV7

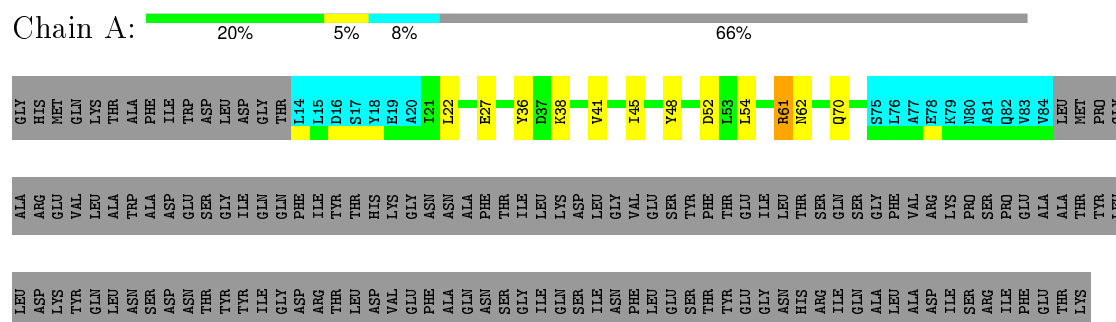
4.2.2 Score per residue for model 2

- Molecule 1: Hydrolase, haloacid dehalogenase-like family



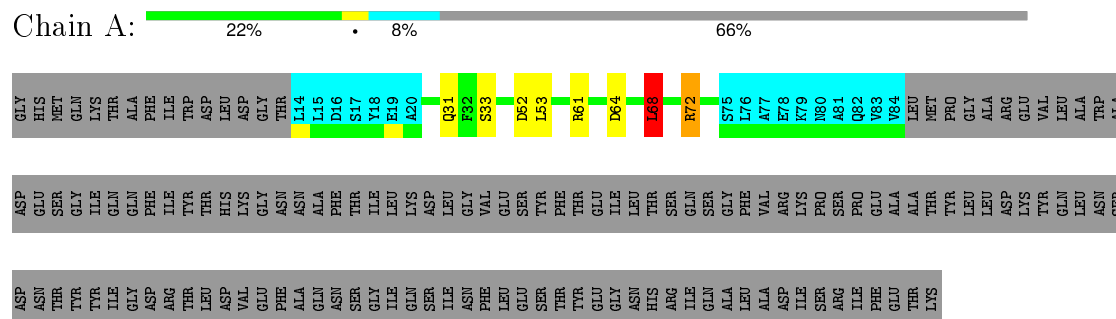
4.2.3 Score per residue for model 3

- Molecule 1: Hydrolase, haloacid dehalogenase-like family



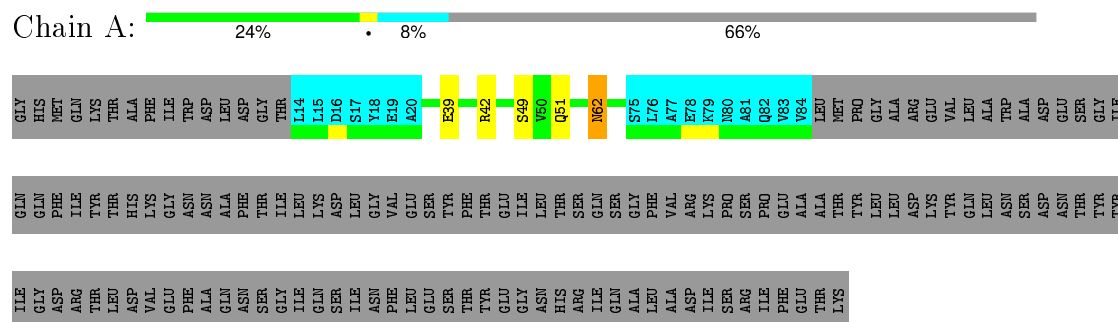
4.2.4 Score per residue for model 4

- Molecule 1: Hydrolase, haloacid dehalogenase-like family



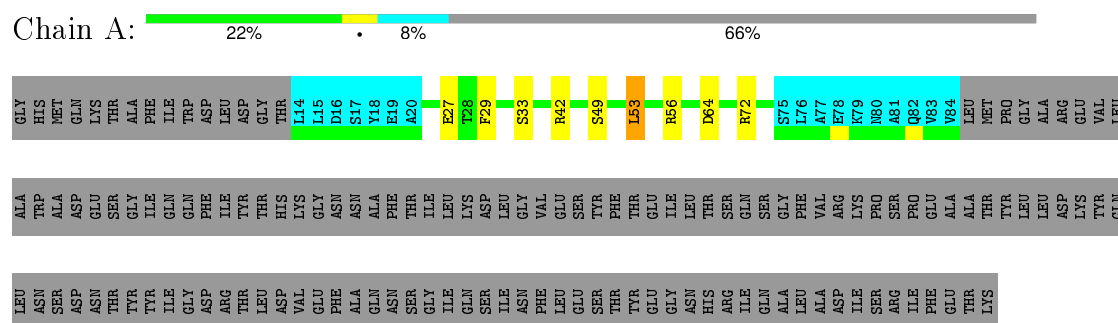
4.2.5 Score per residue for model 5

- Molecule 1: Hydrolase, haloacid dehalogenase-like family



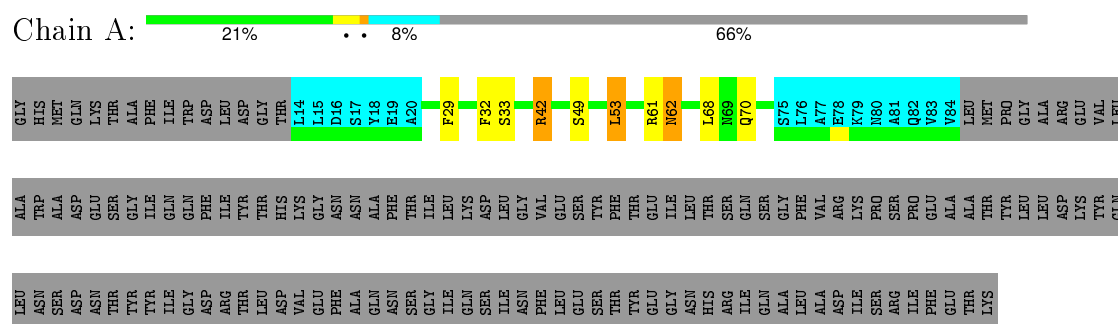
4.2.6 Score per residue for model 6

- Molecule 1: Hydrolase, haloacid dehalogenase-like family



4.2.7 Score per residue for model 7 (medoid)

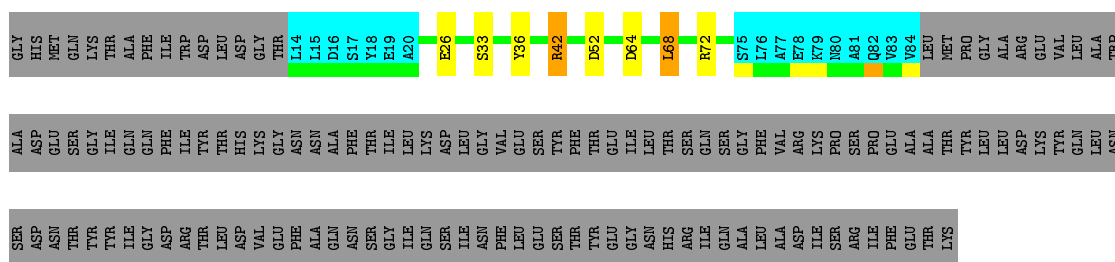
- Molecule 1: Hydrolase, haloacid dehalogenase-like family



4.2.8 Score per residue for model 8

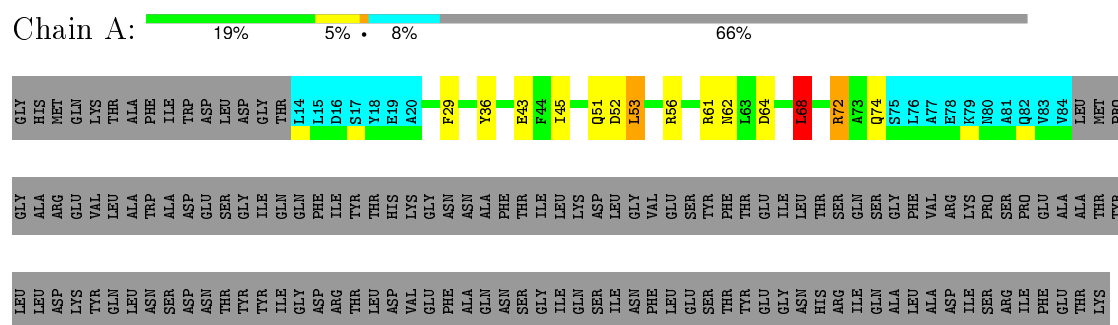
- Molecule 1: Hydrolase, haloacid dehalogenase-like family





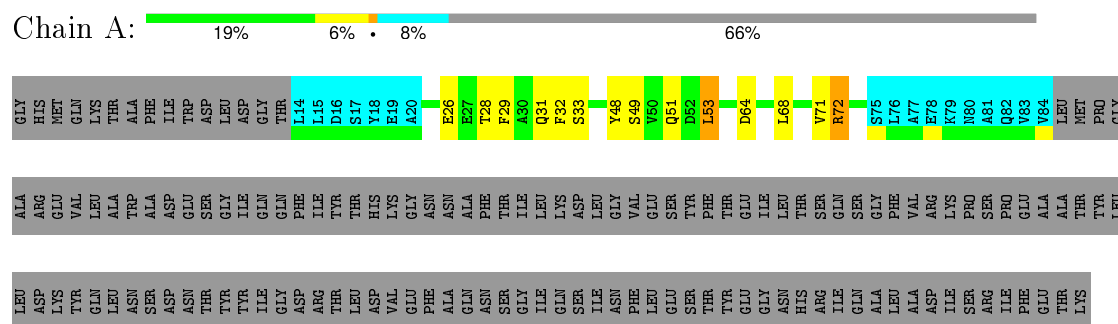
4.2.9 Score per residue for model 9

- Molecule 1: Hydrolase, haloacid dehalogenase-like family



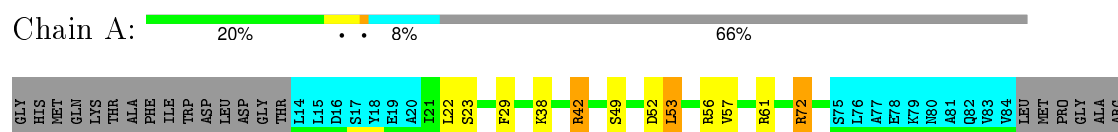
4.2.10 Score per residue for model 10

- Molecule 1: Hydrolase, haloacid dehalogenase-like family



4.2.11 Score per residue for model 11

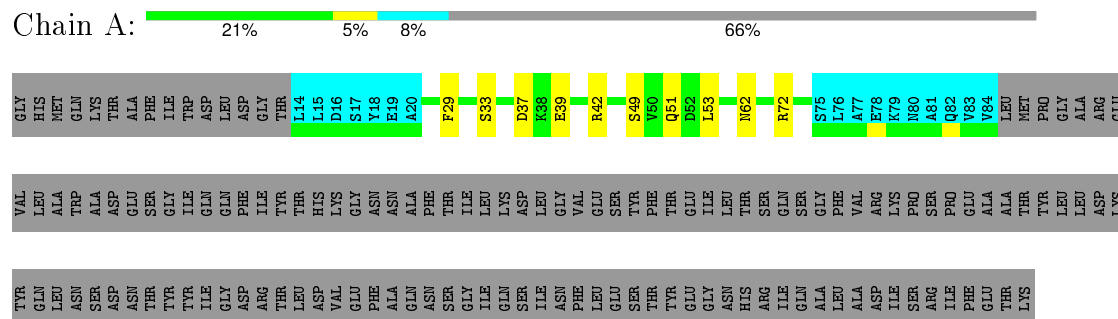
- Molecule 1: Hydrolase, haloacid dehalogenase-like family



GLU	VAL	TYR	GLN	LEU	ALA	TRP	ALA	ASP	GLU	SER	GLY	ILE	GLN	PHE	ILE	THR	THR	HIS	LYS	GLY	ASN	ALA	ASN	PHE	THR	ILE	LEU	LYS	ASP	LEU	GLY	VAL	GLU	SER	TYR	PHE	THR	THR	GLU	ILE	LEU	ASN	HIS	THR	GLN	SER	GLY	ALA	LEU	PHE	VAL	ASP	ARG	ILE	THR	LYS
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

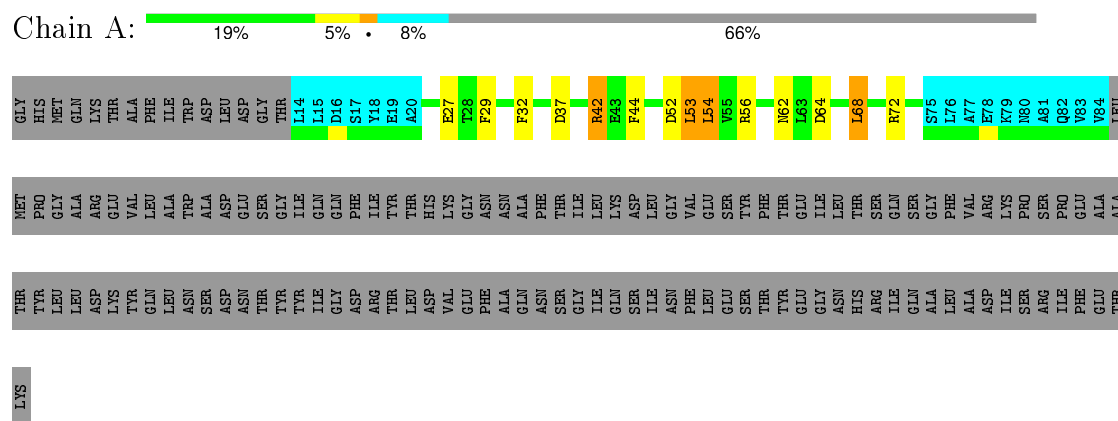
4.2.12 Score per residue for model 12

- Molecule 1: Hydrolase, haloacid dehalogenase-like family



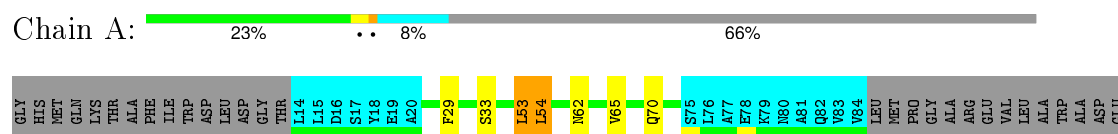
4.2.13 Score per residue for model 13

- Molecule 1: Hydrolase, haloacid dehalogenase-like family



4.2.14 Score per residue for model 14

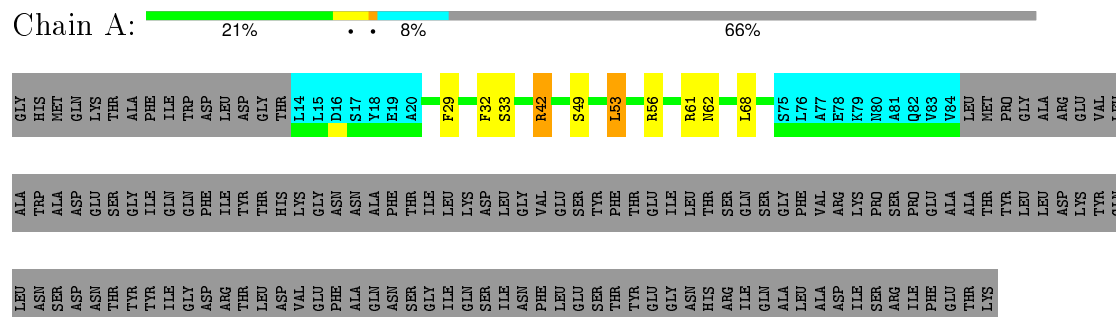
- Molecule 1: Hydrolase, haloacid dehalogenase-like family





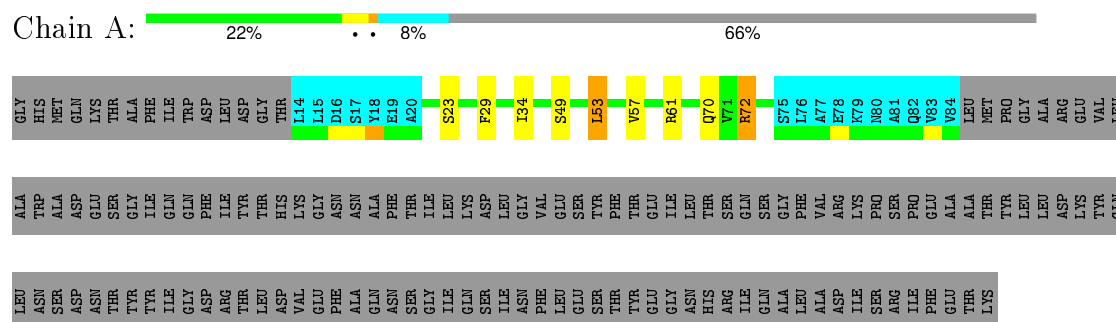
4.2.15 Score per residue for model 15

- Molecule 1: Hydrolase, haloacid dehalogenase-like family



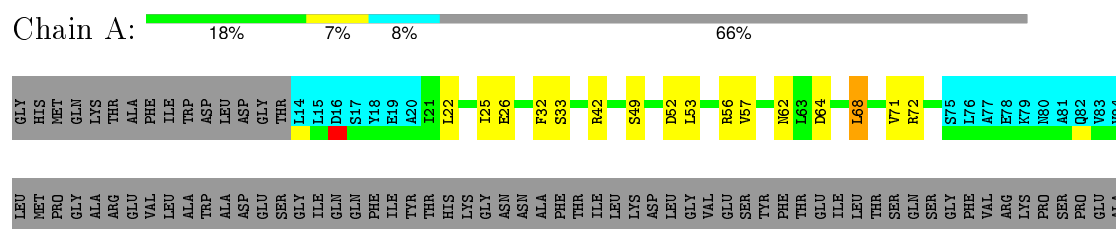
4.2.16 Score per residue for model 16

- Molecule 1: Hydrolase, haloacid dehalogenase-like family



4.2.17 Score per residue for model 17

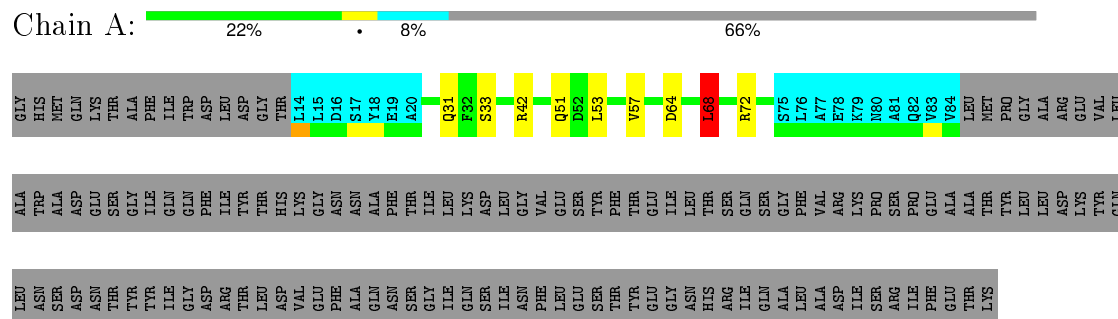
- Molecule 1: Hydrolase, haloacid dehalogenase-like family





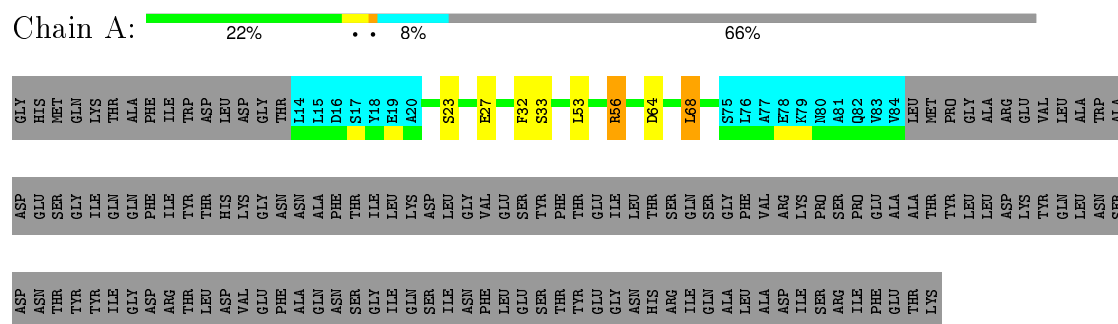
4.2.18 Score per residue for model 18

- Molecule 1: Hydrolase, haloacid dehalogenase-like family



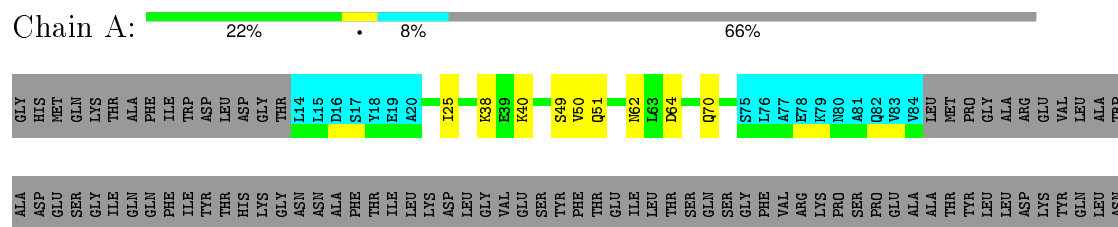
4.2.19 Score per residue for model 19

- Molecule 1: Hydrolase, haloacid dehalogenase-like family



4.2.20 Score per residue for model 20

- Molecule 1: Hydrolase, haloacid dehalogenase-like family



SER	ASP	ASN	THR	TYR	ILE	GLY	ASP	ARG	THR	LEU	ASP	VAL	GLU	PHE	ALA	GLN	ASN	SER	GLY	ILE	GLN	SER	ILE	ASN	PHE	LEU	GLU	SER	THR	TYR	GLU	GLY	ASN	HIS	ARG	ILE	GLN	ALA	LEU	ALA	ASP	ILE	SER	ARG	ILE	PHE	GLU	THR	LYS
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

5 Refinement protocol and experimental data overview

The models were refined using the following method: *torsion angle dynamics, simulated annealing*.

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *target function*.

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
CYANA	structure solution	3.0
OPAL	refinement	
UNIO-ATNOS/CANDID	structure solution	
CYANA	refinement	3.0

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	2mu2_cs.str
Number of chemical shift lists	1
Total number of shifts	2213
Number of shifts mapped to atoms	2213
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	76%

No validations of the models with respect to experimental NMR restraints is performed at this time.

6 Model quality i

6.1 Standard geometry i

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 (average) 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	0.61±0.01	0±0/456 (0.0±0.0%)	1.05±0.06	2±1/616 (0.2±0.2%)
All	All	0.61	0/9120 (0.0%)	1.05	30/12320 (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	Planarity
1	A	0.0±0.0	0.8±0.8
All	All	0	16

There are no bond-length outliers.

All unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	A	42	ARG	NE-CZ-NH2	-8.12	116.24	120.30	15	4
1	A	68	LEU	CB-CG-CD1	7.00	122.89	111.00	13	7
1	A	72	ARG	NE-CZ-NH2	-6.72	116.94	120.30	4	3
1	A	61	ARG	CD-NE-CZ	6.34	132.47	123.60	3	1
1	A	71	VAL	CG1-CB-CG2	-6.27	100.86	110.90	17	1
1	A	42	ARG	NE-CZ-NH1	6.25	123.42	120.30	12	4
1	A	72	ARG	NE-CZ-NH1	5.97	123.29	120.30	10	2
1	A	56	ARG	NE-CZ-NH2	-5.66	117.47	120.30	11	3
1	A	36	TYR	CB-CG-CD2	-5.31	117.81	121.00	8	1
1	A	61	ARG	NE-CZ-NH1	5.25	122.92	120.30	11	1
1	A	54	LEU	CB-CG-CD1	-5.18	102.19	111.00	13	1
1	A	72	ARG	CD-NE-CZ	5.14	130.79	123.60	11	2

There are no chirality outliers.

All unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	42	ARG	Sidechain	6
1	A	72	ARG	Sidechain	3
1	A	36	TYR	Sidechain	2
1	A	61	ARG	Sidechain	2
1	A	48	TYR	Sidechain	2
1	A	56	ARG	Sidechain	1

6.2 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 each 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 averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
1	A	449	450	450	2±1
All	All	8980	9000	9000	40

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 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:54:LEU:HD11	1:A:65:VAL:HG23	0.58	1.76	14	1
1:A:29:PHE:CZ	1:A:53:LEU:HD11	0.54	2.36	12	10
1:A:34:ILE:HD13	1:A:57:VAL:HG11	0.54	1.79	16	1
1:A:32:PHE:CZ	1:A:68:LEU:HD12	0.53	2.39	17	3
1:A:54:LEU:HD12	1:A:54:LEU:H	0.52	1.63	3	1
1:A:54:LEU:HD12	1:A:54:LEU:C	0.49	2.28	2	1
1:A:32:PHE:CE1	1:A:68:LEU:HD23	0.49	2.43	10	3
1:A:25:ILE:HD11	1:A:50:VAL:HG13	0.49	1.84	20	1
1:A:53:LEU:O	1:A:57:VAL:HG23	0.49	2.07	17	5
1:A:54:LEU:C	1:A:54:LEU:HD12	0.48	2.29	14	1
1:A:64:ASP:O	1:A:68:LEU:HD22	0.47	2.09	9	4
1:A:22:LEU:HD23	1:A:25:ILE:HD12	0.44	1.88	17	1
1:A:29:PHE:HZ	1:A:53:LEU:HD11	0.44	1.72	12	1
1:A:22:LEU:HD21	1:A:45:ILE:HD13	0.43	1.90	3	1
1:A:44:PHE:CD1	1:A:53:LEU:HB2	0.43	2.49	13	1
1:A:22:LEU:HD22	1:A:41:VAL:CG1	0.42	2.44	3	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:28:THR:HG23	1:A:71:VAL:CG1	0.42	2.45	10	1
1:A:32:PHE:CE2	1:A:68:LEU:HD12	0.41	2.50	13	1
1:A:34:ILE:HD13	1:A:57:VAL:CG1	0.40	2.47	16	1
1:A:22:LEU:HD13	1:A:38:LYS:HD2	0.40	1.93	11	1

6.3 Torsion angles [i](#)

6.3.1 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 PDB entries followed by that with respect to all NMR entries. 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	54/208 (26%)	49±1 (90±2%)	5±1 (9±2%)	0±1 (0±1%)	38	79
All	All	1080/4160 (26%)	974 (90%)	101 (9%)	5 (0%)	38	79

All 3 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	62	ASN	3
1	A	49	SER	1
1	A	33	SER	1

6.3.2 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 PDB entries followed by that with respect to all NMR entries. 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	50/181 (28%)	44±1 (88±3%)	6±1 (12±3%)	10	52
All	All	1000/3620 (28%)	875 (88%)	125 (12%)	10	52

All 26 unique residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	33	SER	12
1	A	53	LEU	11
1	A	62	ASN	10
1	A	49	SER	10
1	A	72	ARG	9
1	A	52	ASP	8
1	A	64	ASP	6
1	A	51	GLN	6
1	A	70	GLN	5
1	A	68	LEU	5
1	A	31	GLN	5
1	A	56	ARG	5
1	A	27	GLU	4
1	A	42	ARG	4
1	A	61	ARG	4
1	A	38	LYS	3
1	A	26	GLU	3
1	A	54	LEU	3
1	A	23	SER	3
1	A	37	ASP	2
1	A	39	GLU	2
1	A	60	ASP	1
1	A	45	ILE	1
1	A	43	GLU	1
1	A	74	GLN	1
1	A	40	LYS	1

6.3.3 RNA ⓘ

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

6.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

6.6 Ligand geometry [i](#)

There are no ligands in this entry.

6.7 Other polymers [i](#)

There are no such molecules in this entry.

6.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

7 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 76% for the well-defined parts and 77% for the entire structure.

7.1 Chemical shift list 1

File name: 2mu2_cs.str

Chemical shift list name: *assigned_chem_shift_list_1*

7.1.1 Bookkeeping

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	2213
Number of shifts mapped to atoms	2213
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	7

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction \pm precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	205	-0.02 ± 0.11	None needed (< 0.5 ppm)
$^{13}\text{C}_\beta$	196	0.40 ± 0.06	None needed (< 0.5 ppm)
$^{13}\text{C}'$	0	—	—
^{15}N	196	0.32 ± 0.32	None needed (< 0.5 ppm)

7.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 76%, i.e. 541 atoms were assigned a chemical shift out of a possible 712. 12 out of 12 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	159/268 (59%)	53/107 (50%)	54/108 (50%)	52/53 (98%)
Sidechain	343/392 (88%)	210/227 (93%)	123/144 (85%)	10/21 (48%)

Continued on next page...

Continued from previous page...

	Total	¹H	¹³C	¹⁵N
Aromatic	39/52 (75%)	27/28 (96%)	12/24 (50%)	0/0 (—%)
Overall	541/712 (76%)	290/362 (80%)	189/276 (68%)	62/74 (84%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 77%, i.e. 694 atoms were assigned a chemical shift out of a possible 907. 17 out of 17 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	¹H	¹³C	¹⁵N
Backbone	210/353 (59%)	70/141 (50%)	71/142 (50%)	69/70 (99%)
Sidechain	439/494 (89%)	268/285 (94%)	159/185 (86%)	12/24 (50%)
Aromatic	45/60 (75%)	31/32 (97%)	14/28 (50%)	0/0 (—%)
Overall	694/907 (77%)	369/458 (81%)	244/355 (69%)	81/94 (86%)

7.1.4 Statistically unusual chemical shifts ⓘ

The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, ppm	Expected range, ppm	Z-score
1	A	162	ILE	HD12	-1.47	2.13 – -0.77	-7.4
1	A	162	ILE	HD11	-1.47	2.13 – -0.77	-7.4
1	A	162	ILE	HD13	-1.47	2.13 – -0.77	-7.4
1	A	200	SER	HB2	2.12	5.18 – 2.58	-6.8
1	A	151	LYS	HG2	-0.37	2.67 – 0.07	-6.7
1	A	173	GLN	HB2	0.59	3.30 – 0.80	-5.9
1	A	93	LEU	HB2	-0.14	3.32 – -0.08	-5.2

7.1.5 Random Coil Index (RCI) plots ⓘ

The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.

Random coil index (RCI) for chain A:

