



# wwPDB NMR Structure Validation Summary Report ⓘ

Feb 12, 2017 – 05:29 pm GMT

PDB ID : 2LTQ  
Title : High resolution structure of DsbB C41S by joint calculation with solid-state NMR and X-ray data  
Authors : Tang, M.; Sperling, L.J.; Schwieters, C.D.; Nesbitt, A.E.; Gennis, R.B.; Rienstra, C.M.  
Deposited on : 2012-05-30

This is a wwPDB NMR 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

<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	:	1.7.2 (RC1), CSD as538be (2017)
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	trunk28760
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

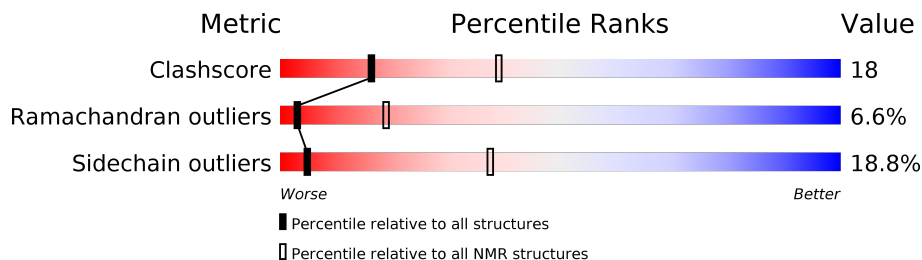
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLID-STATE NMR*

The overall completeness of chemical shifts assignment is 2%.

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	125131	11601
Ramachandran outliers	121729	10391
Sidechain outliers	121581	10367

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  $\geq 3$ , 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	176	7% 15% . 61% 16%
1	D	176	6% 10% . 67% 16%
2	B	239	50% 32% 9% 9%
2	E	239	55% 27% 8% . 9%
3	C	221	52% 34% 11% .
3	F	221	52% 36% 8% . .

## 2 Ensemble composition and analysis ⓘ

This entry contains 10 models. Model 4 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

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:15-A:29, A:53-A:55, A:73-A:93, A:151-A:152, B:21-B:132, B:134-B:239, C:1-C:100, C:105-C:119, C:121-C:221, D:53-D:62, D:72-D:91, E:21-E:132, E:134-E:239, F:1-F:119, F:121-F:215 (937)	0.07	4

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 1 clusters. No single-model clusters were found.

Cluster number	Models
1	1, 2, 3, 4, 5, 6, 7, 8, 9, 10

### 3 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 17810 atoms, of which 8819 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Disulfide bond formation protein B.

Mol	Chain	Residues	Atoms						Trace
1	A	148	Total	C	H	N	O	S	0
			2387	794	1212	183	189	9	
1	D	148	Total	C	H	N	O	S	0
			2387	794	1212	183	189	9	

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	8	ALA	CYS	ENGINEERED MUTATION	UNP P0A6M2
A	41	SER	CYS	ENGINEERED MUTATION	UNP P0A6M2
A	49	VAL	CYS	ENGINEERED MUTATION	UNP P0A6M2
D	8	ALA	CYS	ENGINEERED MUTATION	UNP P0A6M2
D	41	SER	CYS	ENGINEERED MUTATION	UNP P0A6M2
D	49	VAL	CYS	ENGINEERED MUTATION	UNP P0A6M2

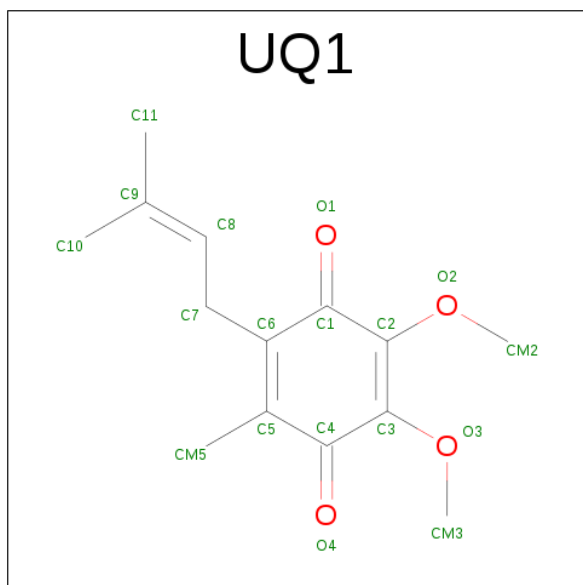
- Molecule 2 is a protein called Fab fragment light chain.

Mol	Chain	Residues	Atoms						Trace
2	B	218	Total	C	H	N	O	S	0
			3319	1052	1629	283	347	8	
2	E	218	Total	C	H	N	O	S	0
			3319	1052	1629	283	347	8	

- Molecule 3 is a protein called Fab fragment heavy chain.

Mol	Chain	Residues	Atoms						Trace
3	C	216	Total	C	H	N	O	S	0
			3189	1017	1574	264	325	9	
3	F	214	Total	C	H	N	O	S	0
			3173	1015	1563	262	324	9	

- Molecule 4 is UBIQUINONE-1 (three-letter code: UQ1) (formula: C<sub>14</sub>H<sub>18</sub>O<sub>4</sub>).



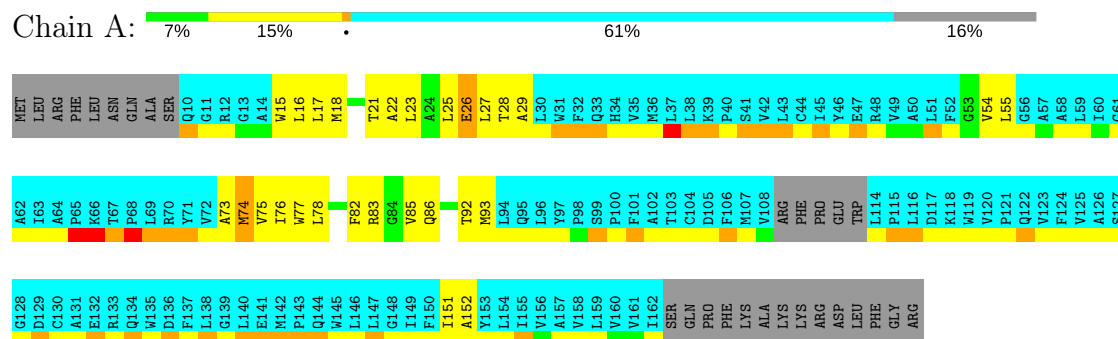
Mol	Chain	Residues	Atoms		
4	A	1	Total	C	O
			18	14	4
4	D	1	Total	C	O
			18	14	4

## 4 Residue-property plots

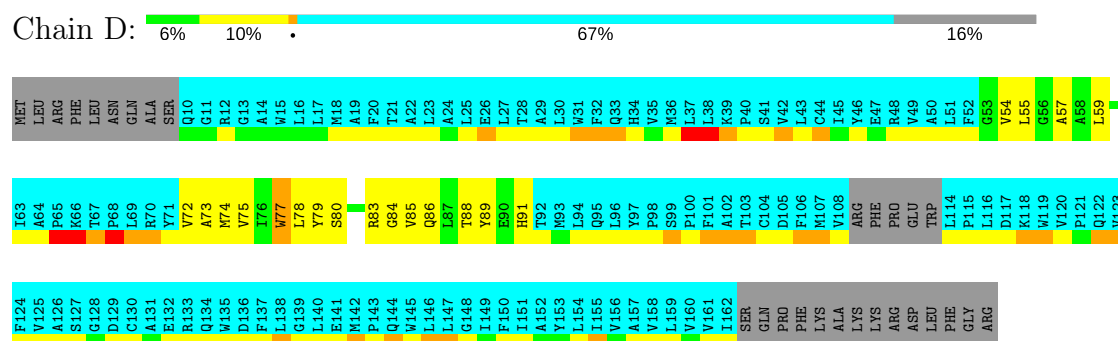
### 4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-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. Stretches of 2 or more consecutive residues without any outliers are shown as green connectors. Residues which are classified as ill-defined in the NMR ensemble, are shown in cyan with an underline colour-coded according to the previous scheme. Residues which were present in the experimental sample, but not modelled in the final structure are shown in grey.

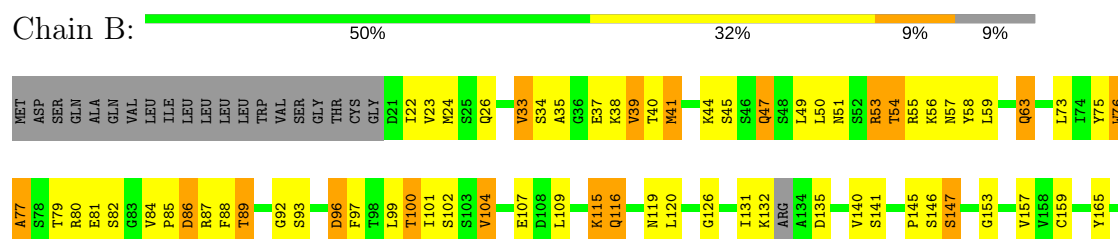
- Molecule 1: Disulfide bond formation protein B



- Molecule 1: Disulfide bond formation protein B



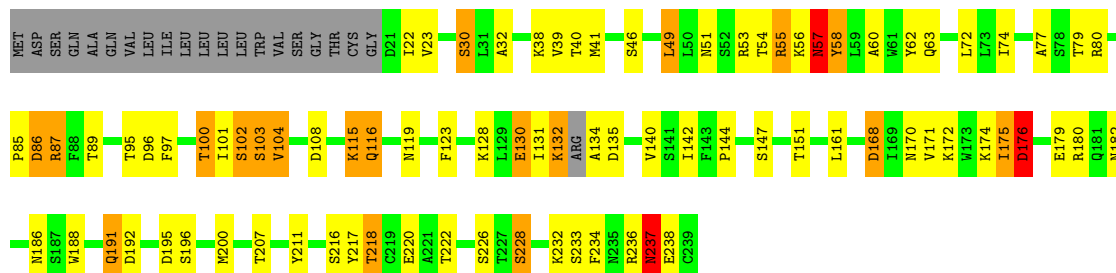
- Molecule 2: Fab fragment light chain





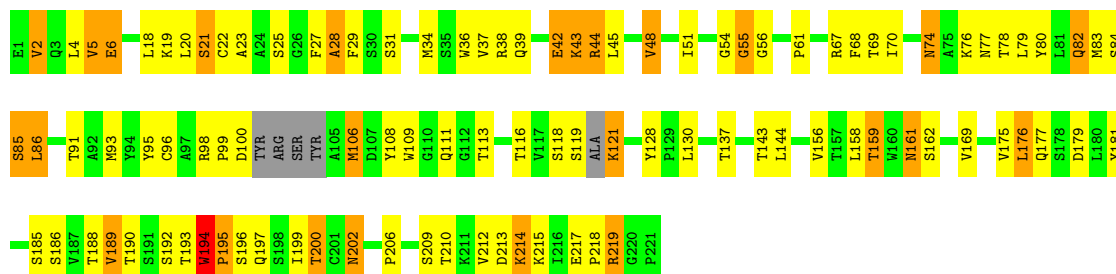
• Molecule 2: Fab fragment light chain

Chain E: 55% 27% 8% 9%



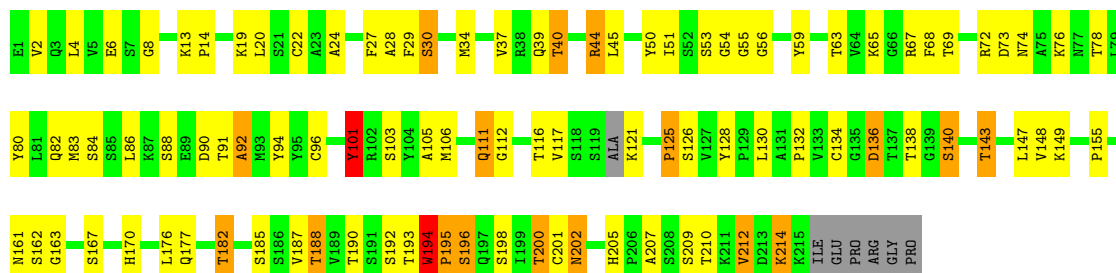
• Molecule 3: Fab fragment heavy chain

Chain C: 52% 34% 11%



• Molecule 3: Fab fragment heavy chain

Chain F: 52% 36% 8%

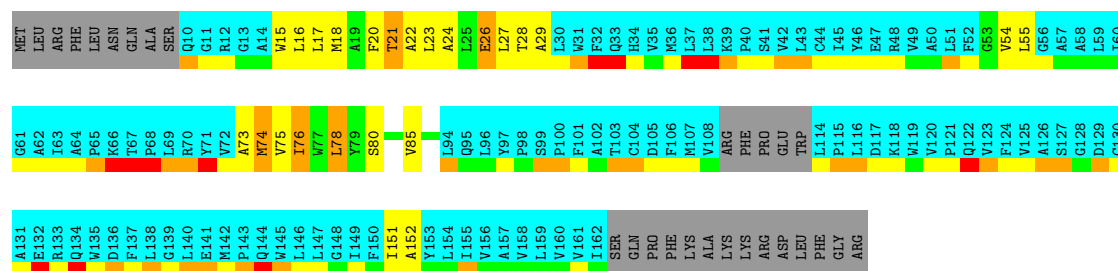


## 4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 4. Colouring as in section 4.1 above.

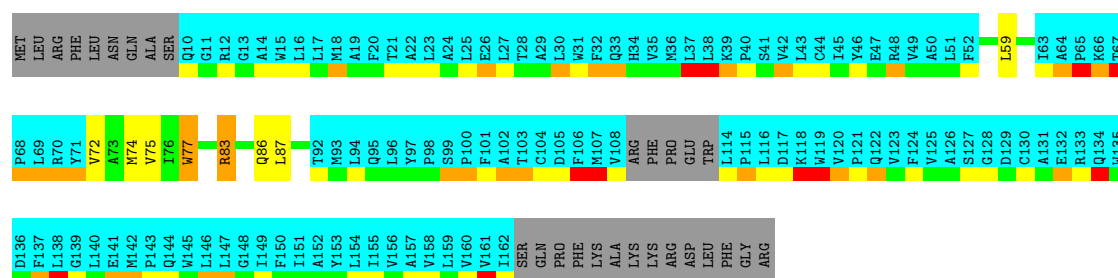
• Molecule 1: Disulfide bond formation protein B

Chain A: 10% 11% • 61% 16%



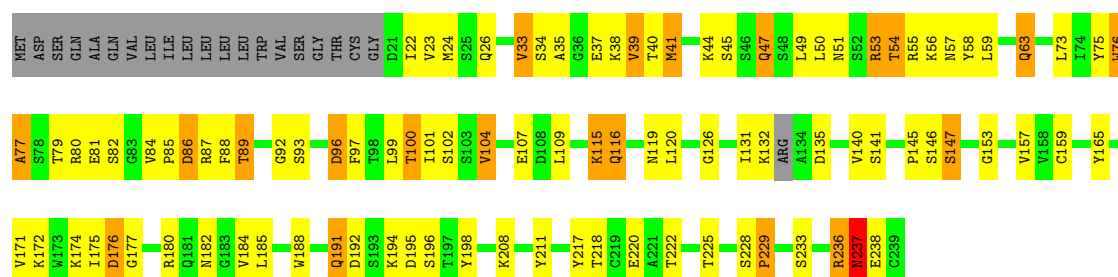
• Molecule 1: Disulfide bond formation protein B

Chain D: 13% •• 67% 16%



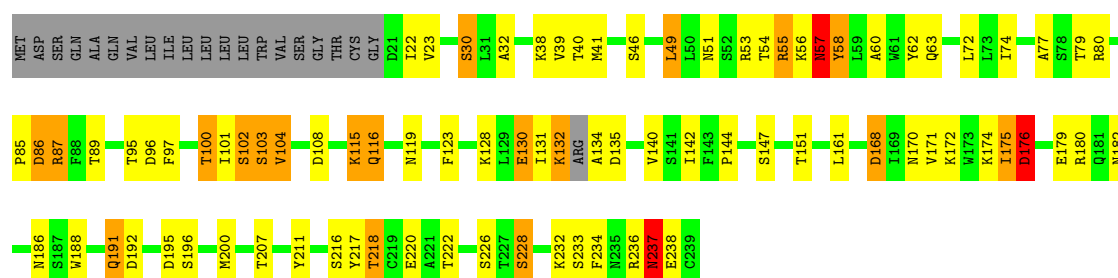
• Molecule 2: Fab fragment light chain

Chain B: 50% 32% 9% 9%



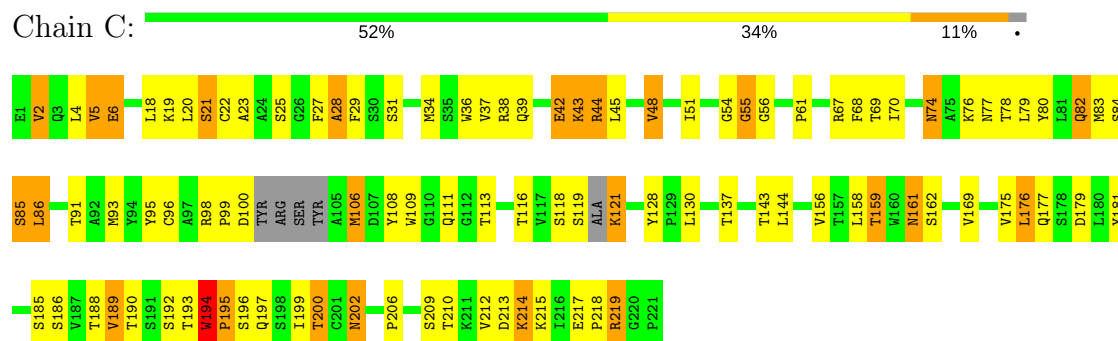
• Molecule 2: Fab fragment light chain

Chain E: 55% 27% 8% • 9%

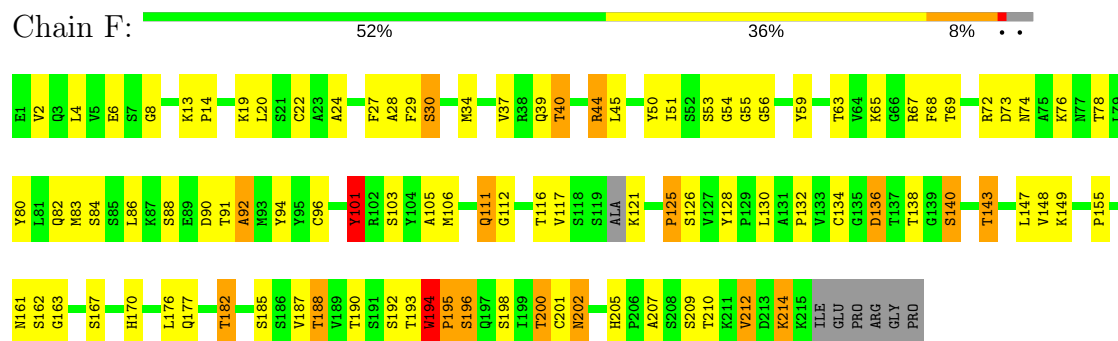


• Molecule 3: Fab fragment heavy chain





• Molecule 3: Fab fragment heavy chain



## 5 Refinement protocol and experimental data overview

The models were refined using the following method: *simulated annealing*.

Of the 200 calculated structures, 10 were deposited, based on the following criterion: *structures with the lowest energy*.

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

Software name	Classification	Version
X-PLOR NIH	refinement	
X-PLOR NIH	structure solution	

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)	2ltq_cs.str
Number of chemical shift lists	1
Total number of shifts	745
Number of shifts mapped to atoms	670
Number of unparsed shifts	74
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	1
Assignment completeness (well-defined parts)	2%

Note: This is a solid-state NMR structure, where hydrogen atoms are typically not assigned a chemical shift value, which may lead to lower completeness of assignment measure.

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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: UQ1

There are no covalent bond-length or bond-angle outliers.

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
2	E	0.0±0.0	1.0±0.0
3	C	0.0±0.0	1.0±0.0
3	F	0.0±0.0	1.0±0.0
All	All	0	30

There are no bond-length outliers.

There are no bond-angle outliers.

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)
3	C	194	TRP	Peptide	10
2	E	57	ASN	Peptide	10
3	F	194	TRP	Peptide	10

### 6.2 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 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	324	335	334	19±6
1	D	226	232	231	9±4
2	B	1690	1629	1625	66±1
2	E	1690	1629	1625	45±1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes
3	C	1615	1574	1570	66±1
3	F	1610	1563	1560	54±1
4	A	18	0	18	1±1
4	D	18	0	18	0±0
All	All	71910	69620	69810	2543

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 18.

5 of 459 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:B:115:LYS:HB2	2:B:115:LYS:NZ	1.21	1.48	6	5
2:B:115:LYS:NZ	2:B:115:LYS:HB2	1.21	1.47	8	5
2:B:44:LYS:NZ	2:B:96:ASP:OD1	1.08	1.85	6	10
2:B:26:GLN:NE2	2:B:126:GLY:H	1.07	1.46	10	10
2:B:40:THR:HG22	2:B:100:THR:HB	1.05	1.23	7	10

## 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	41/176 (23%)	38±1 (93±3%)	2±2 (5±4%)	1±1 (2±1%)	12	51
1	D	30/176 (17%)	29±1 (96±3%)	1±1 (4±3%)	0±0 (0±0%)	100	100
2	B	214/239 (90%)	171±0 (80±0%)	28±0 (13±0%)	15±0 (7±0%)	3	17
2	E	214/239 (90%)	177±0 (83±0%)	26±0 (12±0%)	11±0 (5±0%)	4	25
3	C	210/221 (95%)	165±0 (79±0%)	32±0 (15±0%)	13±0 (6±0%)	3	20
3	F	210/221 (95%)	160±0 (76±0%)	29±0 (14±0%)	21±0 (10±0%)	1	10
All	All	9190/12720 (72%)	7397 (80%)	1184 (13%)	609 (7%)	3	19

5 of 62 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
3	C	137	THR	10
3	F	8	GLY	10
3	F	14	PRO	10
3	F	92	ALA	10
2	B	86	ASP	10

### 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	32/147 (22%)	24±1 (76±4%)	8±1 (24±4%)	3	28
1	D	21/147 (14%)	16±2 (75±9%)	5±2 (25±9%)	3	25
2	B	194/212 (92%)	163±0 (84±0%)	31±0 (16±0%)	6	43
2	E	194/212 (92%)	157±0 (81±0%)	37±0 (19±0%)	5	37
3	C	184/188 (98%)	145±0 (79±0%)	39±0 (21±0%)	4	32
3	F	183/188 (97%)	151±0 (83±0%)	32±0 (17±0%)	5	40
All	All	8080/10940 (74%)	6562 (81%)	1518 (19%)	5	37

5 of 183 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)
3	C	48	VAL	10
2	B	79	THR	10
3	F	182	THR	10
2	E	100	THR	10
3	F	193	THR	10

### 6.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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

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

## 6.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 6.6 Ligand geometry [i](#)

2 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds 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 is the number of standard deviations the observed value is removed from the expected value. A bond length with  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
4	UQ1	A	201	-	18,18,18	1.89±0.00	1±0 (5±0%)
4	UQ1	D	201	-	18,18,18	1.84±0.00	1±0 (5±0%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of 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 angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
4	UQ1	A	201	-	22,25,25	1.08±0.00	0±0 (0±0%)
4	UQ1	D	201	-	22,25,25	1.02±0.00	0±0 (0±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
4	UQ1	A	201	-	-	0±0,9,33,33	0±0,1,1,1
4	UQ1	D	201	-	-	0±0,9,33,33	0±0,1,1,1

All unique bond 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(Å)	Ideal(Å)	Models	
								Worst	Total
4	A	201	UQ1	C6-C5	6.82	1.50	1.35	7	10
4	D	201	UQ1	C6-C5	6.59	1.49	1.35	1	10

There are no bond-angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 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 2% for the well-defined parts and 5% for the entire structure.

### 7.1 Chemical shift list 1

File name: 2ltq\_cs.str

Chemical shift list name: *assigned\_chemical\_shifts\_list\_DsbB*

#### 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	745
Number of shifts mapped to atoms	670
Number of unparsed shifts	74
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	1
Number of shift outliers (ShiftChecker)	1

The following errors were found when reading this chemical shift list.

- Chemical shift has been reported more than once. The only occurrence is reported below.

Shift ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
605	A	146	LEU	CA	56.463	0.300	1

- Entity instance (chain) must be specified. First 5 (of 73) occurrences are reported below.

Shift ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	?	6	ASN	C	179.258	0.145	1
2	?	6	ASN	CA	57.564	0.300	1
3	?	6	ASN	CB	38.167	0.300	1
4	?	7	GLN	C	180.404	0.092	1
5	?	7	GLN	CA	57.611	0.012	1

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- No matching atoms found in structure. The only occurrence is reported below.



Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
A	10	GLN	HN	7.835	0.3	1

### 7.1.2 Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	120	$-0.89 \pm 0.14$	Should be applied
$^{13}\text{C}_\beta$	108	$0.43 \pm 0.10$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}'$	111	$-0.46 \pm 0.25$	None needed ( $< 0.5$ ppm)
$^{15}\text{N}$	119	$1.14 \pm 0.43$	Should be applied

### 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 2%, i.e. 221 atoms were assigned a chemical shift out of a possible 10854. 0 out of 143 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	140/4597 (3%)	21/1830 (1%)	78/1874 (4%)	41/893 (5%)
Sidechain	78/5282 (1%)	0/3110 (0%)	78/1972 (4%)	0/200 (0%)
Aromatic	3/975 (0%)	0/501 (0%)	3/433 (1%)	0/41 (0%)
Overall	221/10854 (2%)	21/5441 (0%)	159/4279 (4%)	41/1134 (4%)

Note: This is a solid-state NMR structure, where hydrogen atoms are typically not assigned a chemical shift value, which may lead to lower completeness of assignment measure.

### 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	92	THR	CG2	27.48	27.15 – 15.95	5.3

### 7.1.5 Random Coil Index (RCI) plots [i](#)

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:

