



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 19, 2020 – 10:29 am BST

PDB ID : 3W1J  
Title : Crystal structure of the N-terminal truncated selenocysteine synthase Sela in complex with thiosulfate  
Authors : Itoh, Y.; Sekine, S.; Yokoyama, S.  
Deposited on : 2012-11-15  
Resolution : 3.25 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

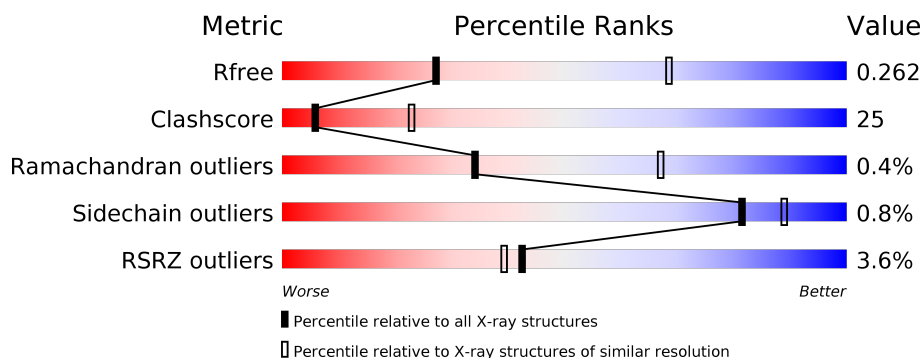
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.25 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	1191 (3.30-3.22)
Clashscore	141614	1251 (3.30-3.22)
Ramachandran outliers	138981	1229 (3.30-3.22)
Sidechain outliers	138945	1228 (3.30-3.22)
RSRZ outliers	127900	1154 (3.30-3.22)

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	392	<div> <div>0%</div> <div> <div></div> <div>58%</div> <div>41%</div> <div>.</div> </div> </div>
1	B	392	<div> <div>3%</div> <div> <div></div> <div>60%</div> <div>39%</div> <div>..</div> </div> </div>
1	C	392	<div> <div>4%</div> <div> <div></div> <div>51%</div> <div>47%</div> <div>.</div> </div> </div>
1	D	392	<div> <div>2%</div> <div> <div></div> <div>58%</div> <div>41%</div> <div>..</div> </div> </div>
1	E	392	<div> <div>2%</div> <div> <div></div> <div>51%</div> <div>49%</div> <div>.</div> </div> </div>
1	F	392	<div> <div>7%</div> <div> <div></div> <div>59%</div> <div>41%</div> <div>.</div> </div> </div>

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	G	392	
1	H	392	
1	I	392	
1	J	392	

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	LLP	B	285	-	-	X	-
1	LLP	E	285	-	-	X	-
2	K	G	501	-	-	-	X
3	THJ	A	2003	-	-	X	-
3	THJ	E	502	-	-	X	-
3	THJ	H	501	-	-	X	-
3	THJ	J	501	-	-	X	-

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 30888 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 L-seryl-tRNA(Sec) selenium transferase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	392	Total	C	N	O	P	S	0	0	0
			3086	1962	535	577	1	11			
1	B	390	Total	C	N	O	P	S	0	0	0
			3069	1951	532	575	1	10			
1	C	392	Total	C	N	O	P	S	0	0	0
			3086	1962	535	577	1	11			
1	D	390	Total	C	N	O	P	S	0	0	0
			3069	1951	532	575	1	10			
1	E	392	Total	C	N	O	P	S	0	0	0
			3086	1962	535	577	1	11			
1	F	392	Total	C	N	O	P	S	0	0	0
			3086	1962	535	577	1	11			
1	G	389	Total	C	N	O	P	S	0	0	0
			3063	1948	531	573	1	10			
1	H	392	Total	C	N	O	P	S	0	0	0
			3086	1962	535	577	1	11			
1	I	392	Total	C	N	O	P	S	0	0	0
			3086	1962	535	577	1	11			
1	J	392	Total	C	N	O	P	S	0	0	0
			3086	1962	535	577	1	11			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	61	MET	-	EXPRESSION TAG	UNP O67140
B	61	MET	-	EXPRESSION TAG	UNP O67140
C	61	MET	-	EXPRESSION TAG	UNP O67140
D	61	MET	-	EXPRESSION TAG	UNP O67140
E	61	MET	-	EXPRESSION TAG	UNP O67140
F	61	MET	-	EXPRESSION TAG	UNP O67140
G	61	MET	-	EXPRESSION TAG	UNP O67140
H	61	MET	-	EXPRESSION TAG	UNP O67140
I	61	MET	-	EXPRESSION TAG	UNP O67140

*Continued on next page...*

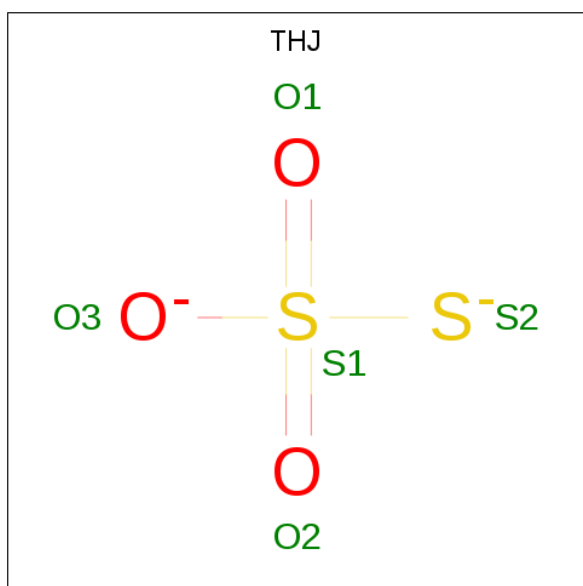
*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
J	61	MET	-	EXPRESSION TAG	UNP O67140

- Molecule 2 is POTASSIUM ION (three-letter code: K) (formula: K).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	G	1	Total K 1 1	0	0
2	I	1	Total K 1 1	0	0
2	A	1	Total K 1 1	0	0
2	D	1	Total K 1 1	0	0
2	E	1	Total K 1 1	0	0

- Molecule 3 is THIOSULFATE (three-letter code: THJ) (formula: O<sub>3</sub>S<sub>2</sub>).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total O S 5 3 2	0	0
3	A	1	Total O S 5 3 2	0	0
3	A	1	Total O S 5 3 2	0	0
3	B	1	Total O S 5 3 2	0	0

*Continued on next page...*

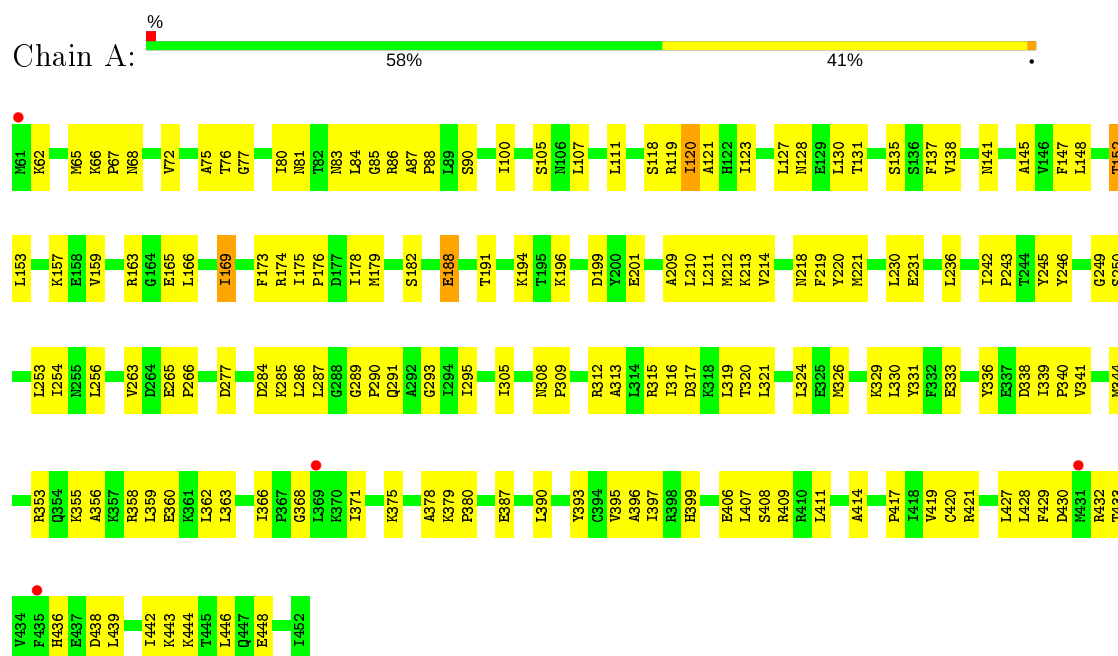
*Continued from previous page...*

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	C	1	Total	O	S	0	0
			5	3	2		
3	D	1	Total	O	S	0	0
			5	3	2		
3	D	1	Total	O	S	0	0
			5	3	2		
3	E	1	Total	O	S	0	0
			5	3	2		
3	E	1	Total	O	S	0	0
			5	3	2		
3	E	1	Total	O	S	0	0
			5	3	2		
3	H	1	Total	O	S	0	0
			5	3	2		
3	I	1	Total	O	S	0	0
			5	3	2		
3	I	1	Total	O	S	0	0
			5	3	2		
3	J	1	Total	O	S	0	0
			5	3	2		
3	J	1	Total	O	S	0	0
			5	3	2		
3	J	1	Total	O	S	0	0
			5	3	2		

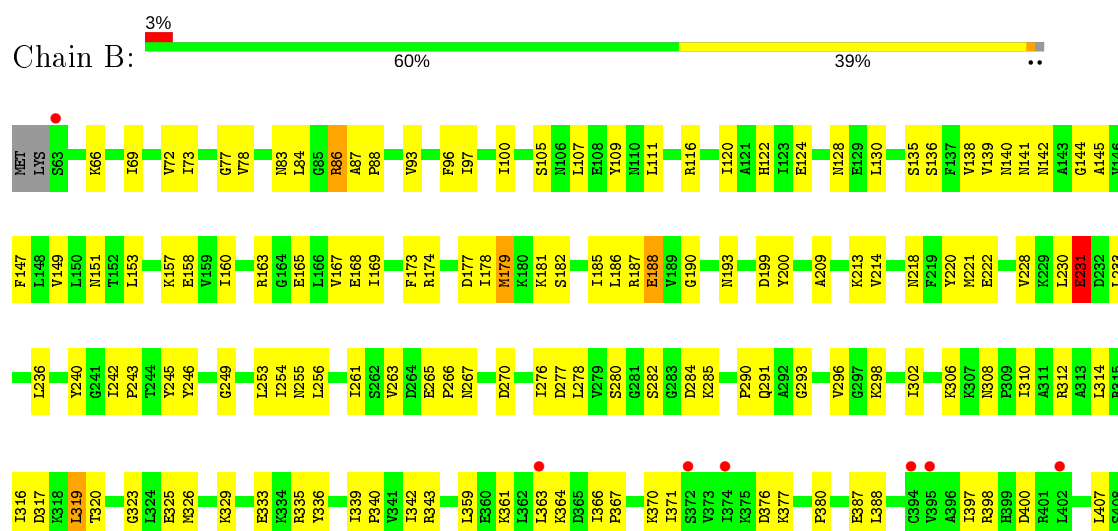
### 3 Residue-property plots

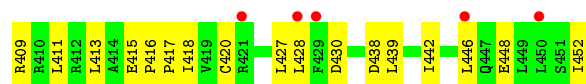
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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: L-seryl-tRNA(Sec) selenium transferase

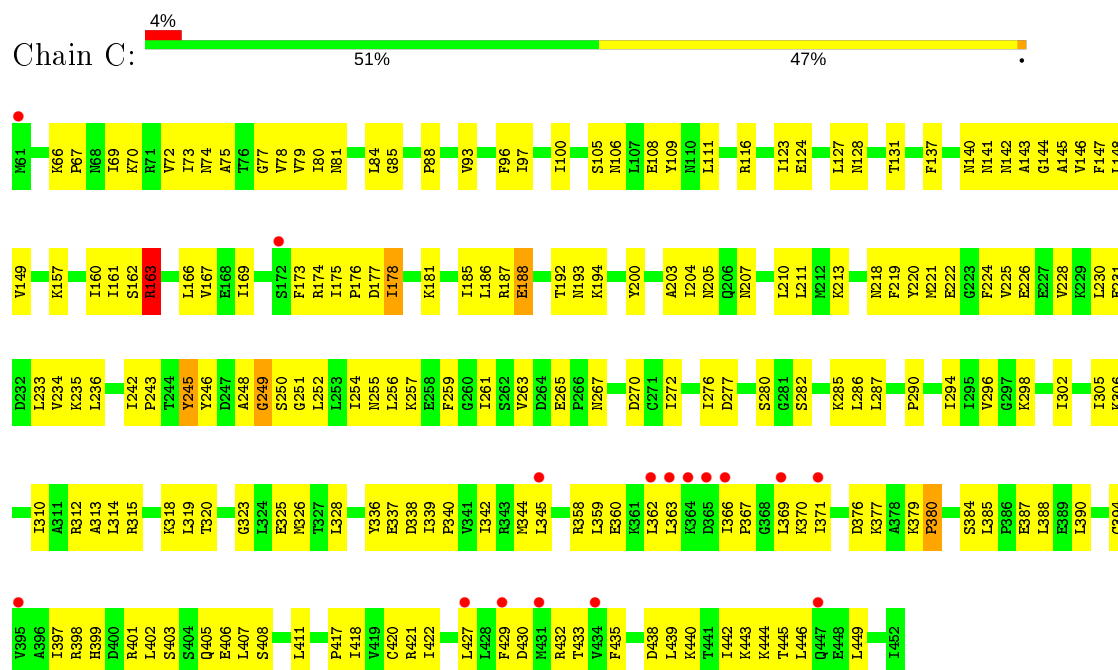


- Molecule 1: L-seryl-tRNA(Sec) selenium transferase

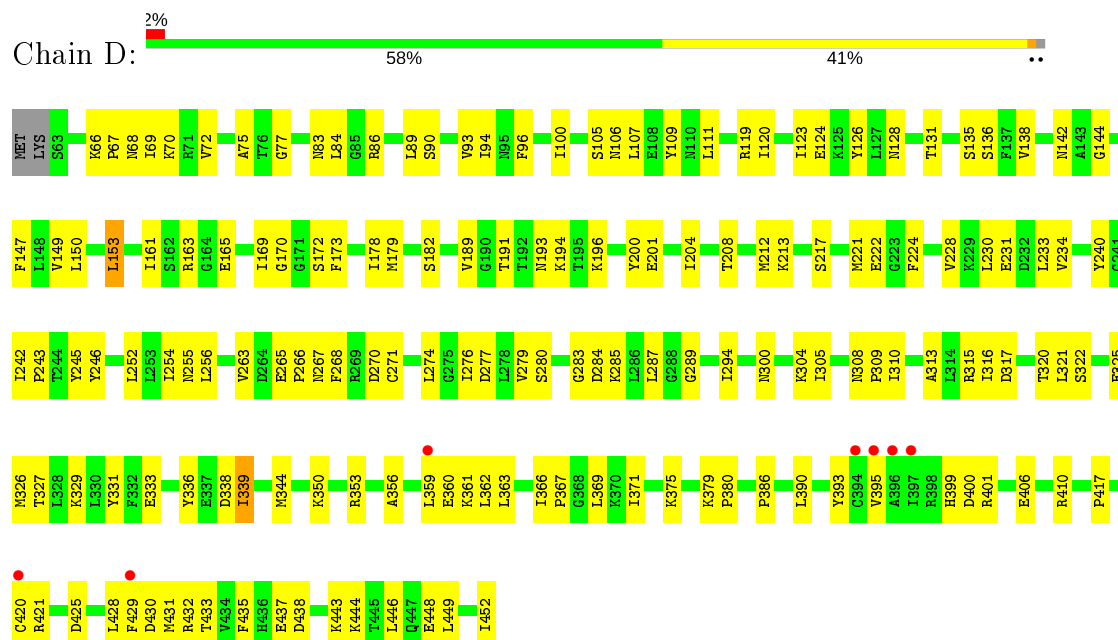




- Molecule 1: L-seryl-tRNA(Sec) selenium transferase



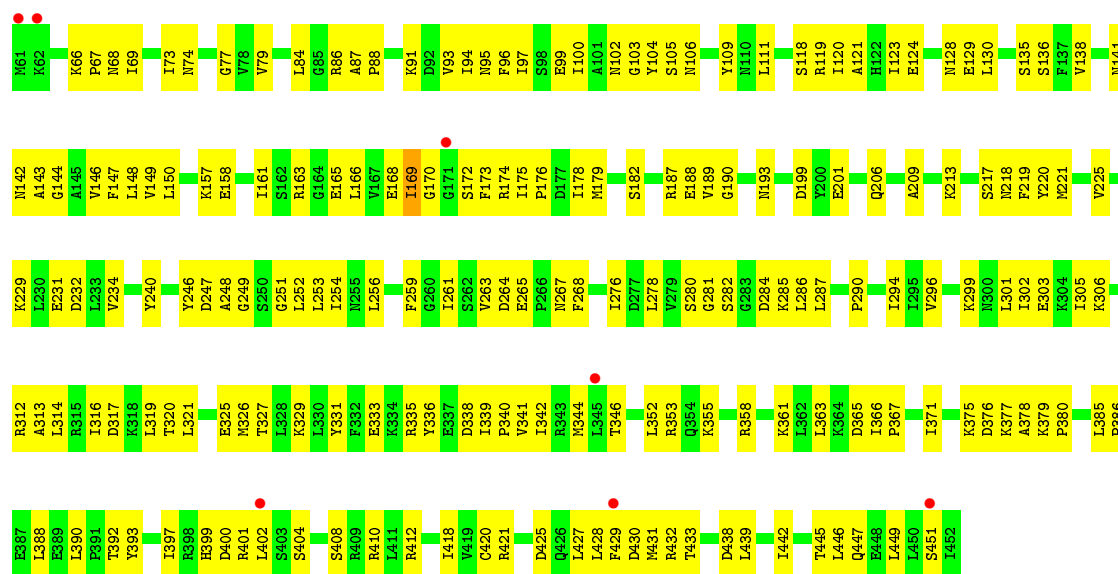
- Molecule 1: L-seryl-tRNA(Sec) selenium transferase



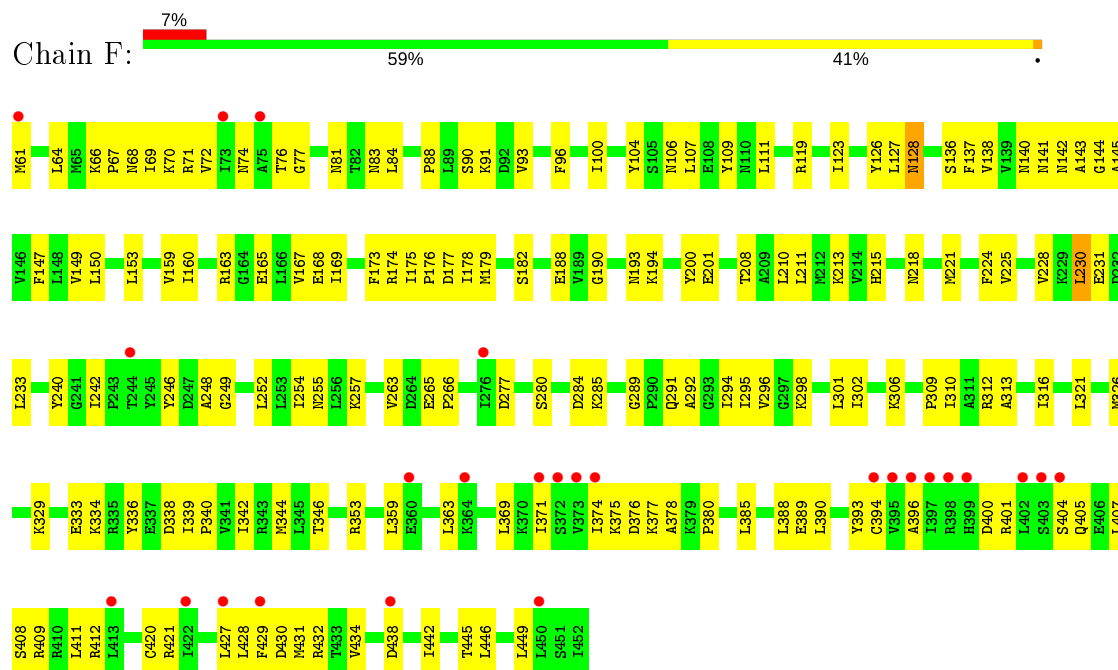
- Molecule 1: L-seryl-tRNA(Sec) selenium transferase



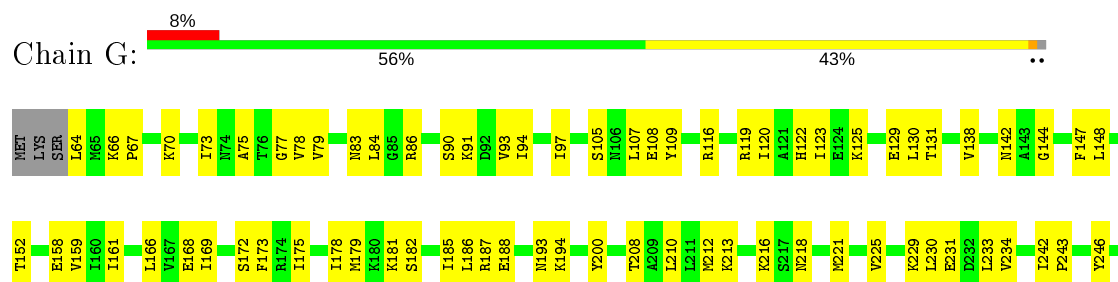


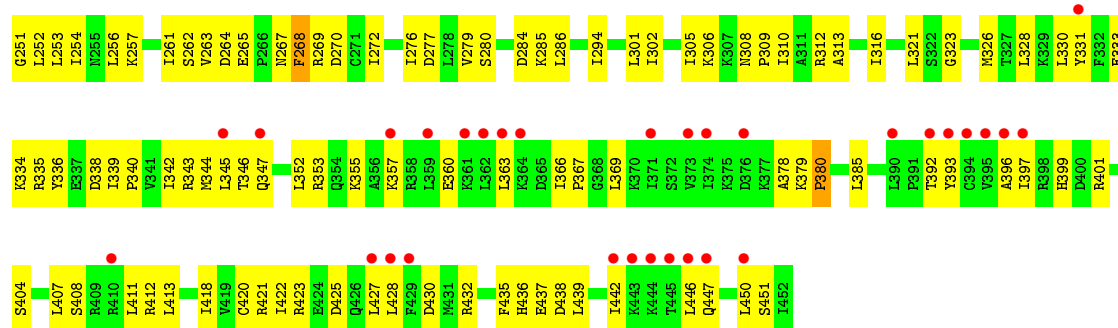


• Molecule 1: L-seryl-tRNA(Sec) selenium transferase

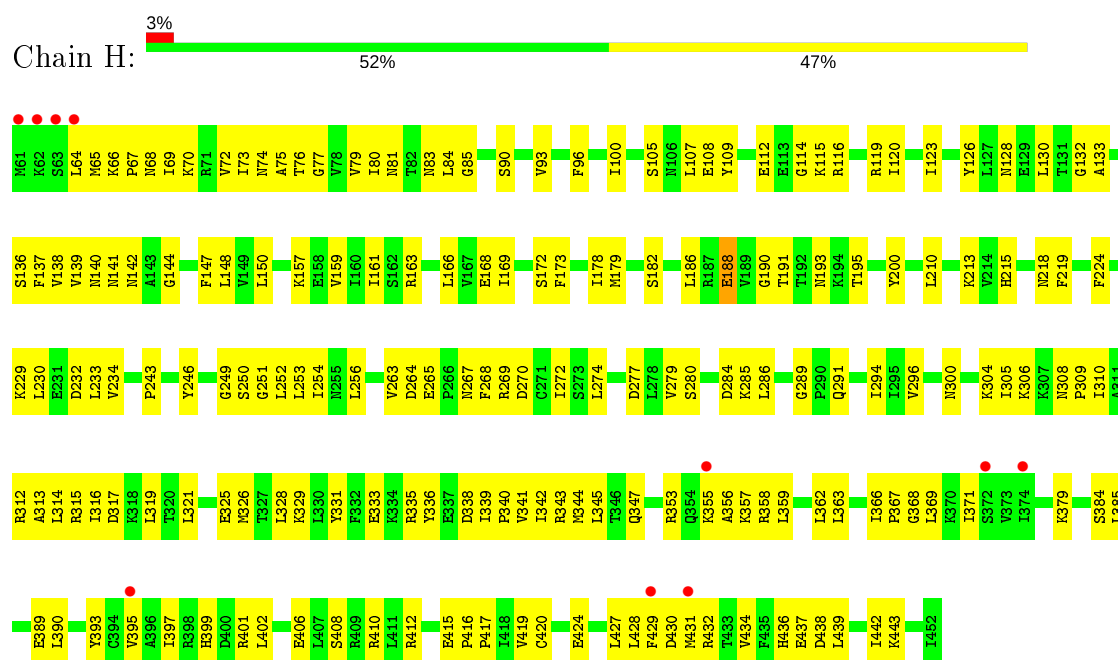


• Molecule 1: L-seryl-tRNA(Sec) selenium transferase

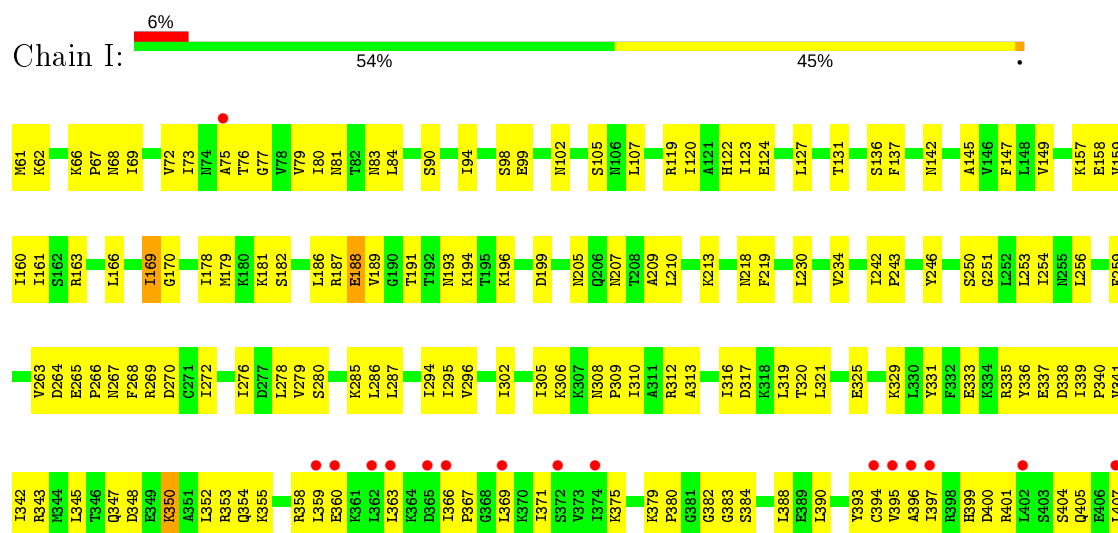




• Molecule 1: L-seryl-tRNA(Sec) selenium transferase

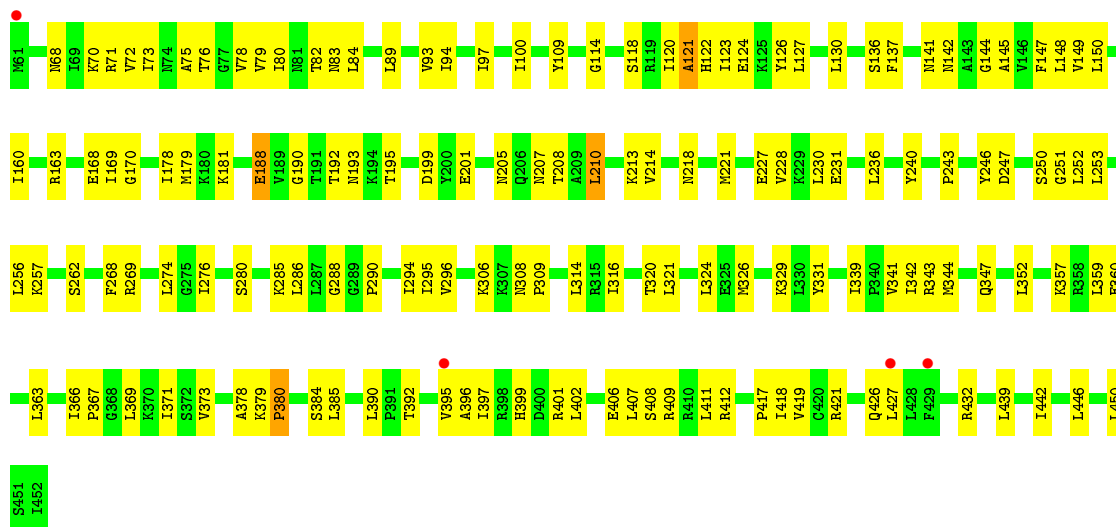


• Molecule 1: L-seryl-tRNA(Sec) selenium transferase





- Molecule 1: L-seryl-tRNA(Sec) selenium transferase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	92.44Å 116.37Å 124.79Å 102.09° 93.39° 106.06°	Depositor
Resolution (Å)	38.52 – 3.25 49.84 – 3.25	Depositor EDS
% Data completeness (in resolution range)	98.9 (38.52-3.25) 99.0 (49.84-3.25)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	0.13	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.01 (at 3.25Å)	Xtriage
Refinement program	PHENIX 1.7.1 _743	Depositor
R, $R_{free}$	0.200 , 0.266 0.196 , 0.262	Depositor DCC
$R_{free}$ test set	3814 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	91.5	Xtriage
Anisotropy	0.708	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.26 , 58.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	30888	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	123.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.59% 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: K, LLP, THJ

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	0.47	0/3101	0.67	0/4165
1	B	0.50	0/3084	0.70	3/4144 (0.1%)
1	C	0.48	0/3101	0.69	1/4165 (0.0%)
1	D	0.47	0/3084	0.69	1/4144 (0.0%)
1	E	0.43	0/3101	0.63	0/4165
1	F	0.42	0/3101	0.60	1/4165 (0.0%)
1	G	0.42	0/3078	0.57	0/4136
1	H	0.41	0/3101	0.60	0/4165
1	I	0.41	0/3101	0.59	0/4165
1	J	0.45	0/3101	0.65	1/4165 (0.0%)
All	All	0.45	0/30953	0.64	7/41579 (0.0%)

There are no bond length outliers.

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	153	LEU	CB-CG-CD2	-5.97	100.85	111.00
1	B	319	LEU	CA-CB-CG	-5.54	102.56	115.30
1	B	312	ARG	NE-CZ-NH2	-5.50	117.55	120.30
1	C	163	ARG	NE-CZ-NH1	-5.18	117.71	120.30
1	F	230	LEU	CA-CB-CG	-5.15	103.45	115.30

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	3086	0	3237	178	0
1	B	3069	0	3215	159	0
1	C	3086	0	3237	187	0
1	D	3069	0	3215	166	0
1	E	3086	0	3237	196	0
1	F	3086	0	3237	167	0
1	G	3063	0	3210	175	0
1	H	3086	0	3237	178	0
1	I	3086	0	3237	186	0
1	J	3086	0	3237	161	0
2	A	1	0	0	0	0
2	D	1	0	0	0	0
2	E	1	0	0	0	0
2	G	1	0	0	0	0
2	I	1	0	0	0	0
3	A	15	0	0	5	0
3	B	5	0	0	1	0
3	C	5	0	0	0	0
3	D	10	0	0	1	0
3	E	15	0	0	4	0
3	H	5	0	0	4	0
3	I	10	0	0	2	0
3	J	15	0	0	4	0
All	All	30888	0	32299	1585	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 25.

The worst 5 of 1585 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:405:GLN:HE22	1:C:422:ILE:HG21	0.92	1.04
1:C:405:GLN:NE2	1:C:422:ILE:HG21	1.73	1.03
1:I:102:ASN:HB3	1:J:71:ARG:HH22	1.25	1.02
1:I:61:MET:HG2	1:I:62:LYS:H	1.25	1.01
1:J:72:VAL:HG22	1:J:417:PRO:HG2	1.40	1.01

There are no symmetry-related clashes.

## 5.3 Torsion angles ⓘ

### 5.3.1 Protein backbone ⓘ

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	389/392 (99%)	362 (93%)	24 (6%)	3 (1%)	19	52
1	B	387/392 (99%)	369 (95%)	16 (4%)	2 (0%)	29	62
1	C	389/392 (99%)	369 (95%)	16 (4%)	4 (1%)	15	47
1	D	387/392 (99%)	359 (93%)	28 (7%)	0	100	100
1	E	389/392 (99%)	374 (96%)	15 (4%)	0	100	100
1	F	389/392 (99%)	376 (97%)	12 (3%)	1 (0%)	41	72
1	G	386/392 (98%)	372 (96%)	12 (3%)	2 (0%)	29	62
1	H	389/392 (99%)	371 (95%)	18 (5%)	0	100	100
1	I	389/392 (99%)	376 (97%)	13 (3%)	0	100	100
1	J	389/392 (99%)	371 (95%)	16 (4%)	2 (0%)	29	62
All	All	3883/3920 (99%)	3699 (95%)	170 (4%)	14 (0%)	34	67

5 of 14 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	120	ILE
1	C	337	GLU
1	B	86	ARG
1	A	152	THR
1	C	249	GLY

### 5.3.2 Protein sidechains ⓘ

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	339/339 (100%)	336 (99%)	3 (1%)	78	87
1	B	337/339 (99%)	333 (99%)	4 (1%)	71	83
1	C	339/339 (100%)	332 (98%)	7 (2%)	53	75
1	D	337/339 (99%)	335 (99%)	2 (1%)	86	91
1	E	339/339 (100%)	338 (100%)	1 (0%)	92	96
1	F	339/339 (100%)	338 (100%)	1 (0%)	92	96
1	G	336/339 (99%)	335 (100%)	1 (0%)	92	96
1	H	339/339 (100%)	338 (100%)	1 (0%)	92	96
1	I	339/339 (100%)	334 (98%)	5 (2%)	65	80
1	J	339/339 (100%)	338 (100%)	1 (0%)	92	96
All	All	3383/3390 (100%)	3357 (99%)	26 (1%)	81	89

5 of 26 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	222	GLU
1	D	70	LYS
1	I	420	CYS
1	C	245	TYR
1	C	342	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 24 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	142	ASN
1	G	218	ASN
1	J	122	HIS
1	F	405	GLN
1	F	426	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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	LLP	I	285	1	23,24,25	1.60	5 (21%)	25,32,34	1.17	2 (8%)
1	LLP	J	285	1	23,24,25	1.53	5 (21%)	25,32,34	1.46	5 (20%)
1	LLP	E	285	1	23,24,25	1.66	6 (26%)	25,32,34	1.60	3 (12%)
1	LLP	F	285	1	23,24,25	1.64	5 (21%)	25,32,34	1.34	3 (12%)
1	LLP	G	285	1	23,24,25	1.74	6 (26%)	25,32,34	1.27	3 (12%)
1	LLP	H	285	1	23,24,25	1.54	5 (21%)	25,32,34	1.21	2 (8%)
1	LLP	A	285	1	23,24,25	1.55	4 (17%)	25,32,34	1.40	5 (20%)
1	LLP	B	285	1	23,24,25	1.90	8 (34%)	25,32,34	1.34	3 (12%)
1	LLP	C	285	1	23,24,25	1.79	5 (21%)	25,32,34	1.44	3 (12%)
1	LLP	D	285	1	23,24,25	1.59	4 (17%)	25,32,34	1.11	3 (12%)

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	LLP	I	285	1	-	3/16/17/19	0/1/1/1
1	LLP	J	285	1	-	2/16/17/19	0/1/1/1
1	LLP	E	285	1	-	4/16/17/19	0/1/1/1
1	LLP	F	285	1	-	2/16/17/19	0/1/1/1
1	LLP	G	285	1	-	5/16/17/19	0/1/1/1
1	LLP	H	285	1	-	1/16/17/19	0/1/1/1
1	LLP	A	285	1	-	3/16/17/19	0/1/1/1
1	LLP	B	285	1	-	5/16/17/19	0/1/1/1
1	LLP	C	285	1	-	6/16/17/19	0/1/1/1
1	LLP	D	285	1	-	5/16/17/19	0/1/1/1

The worst 5 of 53 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	285	LLP	C4-C4'	3.98	1.54	1.46
1	F	285	LLP	C2'-C2	3.94	1.57	1.50
1	F	285	LLP	C4-C4'	3.86	1.54	1.46
1	G	285	LLP	C2'-C2	3.60	1.56	1.50
1	D	285	LLP	C2'-C2	3.56	1.56	1.50

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	E	285	LLP	OP4-C5'-C5	5.79	120.38	109.35
1	C	285	LLP	OP4-C5'-C5	4.00	116.98	109.35
1	A	285	LLP	OP4-C5'-C5	3.88	116.74	109.35
1	B	285	LLP	C5-C6-N1	-3.71	117.63	123.82
1	J	285	LLP	OP4-C5'-C5	3.48	115.99	109.35

There are no chirality outliers.

5 of 36 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	I	285	LLP	C4-C5-C5'-OP4
1	I	285	LLP	C6-C5-C5'-OP4
1	I	285	LLP	O-C-CA-CB
1	J	285	LLP	O-C-CA-CB
1	E	285	LLP	O-C-CA-CB

There are no ring outliers.

10 monomers are involved in 66 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	I	285	LLP	8	0
1	J	285	LLP	8	0
1	E	285	LLP	12	0
1	F	285	LLP	4	0
1	G	285	LLP	6	0
1	H	285	LLP	4	0
1	A	285	LLP	5	0
1	B	285	LLP	11	0
1	C	285	LLP	3	0
1	D	285	LLP	5	0

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 21 ligands modelled in this entry, 5 are monoatomic - leaving 16 for Mogul analysis.

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	THJ	J	501	-	2,4,4	0.05	0	2,6,6	0.37	0
3	THJ	D	502	-	2,4,4	0.16	0	2,6,6	0.65	0
3	THJ	A	2004	-	2,4,4	0.02	0	2,6,6	0.27	0
3	THJ	B	501	-	2,4,4	0.15	0	2,6,6	0.54	0
3	THJ	A	2002	-	2,4,4	0.12	0	2,6,6	0.35	0
3	THJ	J	503	-	2,4,4	0.10	0	2,6,6	0.50	0
3	THJ	I	502	-	2,4,4	0.07	0	2,6,6	0.43	0
3	THJ	C	501	-	2,4,4	0.10	0	2,6,6	0.39	0
3	THJ	E	502	-	2,4,4	0.15	0	2,6,6	0.69	0
3	THJ	H	501	-	2,4,4	0.07	0	2,6,6	0.56	0
3	THJ	E	503	-	2,4,4	0.08	0	2,6,6	0.42	0
3	THJ	J	502	-	2,4,4	0.15	0	2,6,6	0.44	0
3	THJ	I	503	-	2,4,4	0.16	0	2,6,6	0.52	0
3	THJ	D	503	-	2,4,4	0.09	0	2,6,6	0.37	0
3	THJ	E	504	-	2,4,4	0.16	0	2,6,6	0.54	0
3	THJ	A	2003	-	2,4,4	0.16	0	2,6,6	0.51	0

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.

13 monomers are involved in 21 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	J	501	THJ	2	0
3	D	502	THJ	1	0
3	A	2004	THJ	1	0
3	B	501	THJ	1	0
3	A	2002	THJ	1	0
3	J	503	THJ	1	0
3	I	502	THJ	1	0
3	E	502	THJ	3	0
3	H	501	THJ	4	0
3	E	503	THJ	1	0
3	J	502	THJ	1	0
3	I	503	THJ	1	0
3	A	2003	THJ	3	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	391/392 (99%)	-0.22	4 (1%) 82 82	56, 93, 146, 182	0
1	B	389/392 (99%)	-0.11	12 (3%) 49 47	52, 97, 178, 213	0
1	C	391/392 (99%)	-0.05	16 (4%) 37 34	55, 106, 194, 235	0
1	D	389/392 (99%)	-0.06	7 (1%) 68 65	75, 110, 165, 195	0
1	E	391/392 (99%)	-0.03	7 (1%) 68 65	73, 116, 176, 204	0
1	F	391/392 (99%)	0.17	26 (6%) 18 17	73, 125, 202, 235	0
1	G	388/392 (98%)	0.32	31 (7%) 12 11	83, 142, 208, 251	0
1	H	391/392 (99%)	0.01	10 (2%) 56 52	85, 135, 196, 236	0
1	I	391/392 (99%)	0.19	24 (6%) 21 20	62, 122, 196, 221	0
1	J	391/392 (99%)	-0.18	4 (1%) 82 82	58, 102, 175, 212	0
All	All	3903/3920 (99%)	0.00	141 (3%) 42 39	52, 115, 191, 251	0

The worst 5 of 141 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	395	VAL	7.7
1	H	61	MET	7.6
1	G	443	LYS	6.0
1	C	363	LEU	5.7
1	I	429	PHE	5.5

### 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	LLP	G	285	24/25	0.93	0.21	97,107,129,165	0
1	LLP	F	285	24/25	0.94	0.22	80,96,111,118	0
1	LLP	E	285	24/25	0.94	0.25	69,98,126,134	0
1	LLP	D	285	24/25	0.95	0.21	61,89,117,130	0
1	LLP	I	285	24/25	0.96	0.22	76,94,111,120	0
1	LLP	H	285	24/25	0.96	0.23	96,114,127,131	0
1	LLP	B	285	24/25	0.96	0.22	51,71,90,97	0
1	LLP	C	285	24/25	0.96	0.24	67,83,106,112	0
1	LLP	J	285	24/25	0.96	0.23	56,77,98,145	0
1	LLP	A	285	24/25	0.97	0.19	46,74,89,91	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	THJ	E	502	5/5	0.73	0.35	173,175,178,333	0
2	K	G	501	1/1	0.75	0.58	170,170,170,170	0
3	THJ	J	501	5/5	0.85	0.30	182,184,187,187	0
2	K	E	501	1/1	0.85	0.53	142,142,142,142	0
3	THJ	E	503	5/5	0.85	0.24	123,135,148,148	0
3	THJ	I	503	5/5	0.86	0.37	188,189,192,226	0
2	K	I	501	1/1	0.87	0.72	136,136,136,136	0
3	THJ	C	501	5/5	0.89	0.27	152,161,167,168	0
3	THJ	H	501	5/5	0.89	0.15	137,207,207,208	0
2	K	D	501	1/1	0.89	0.48	140,140,140,140	0
3	THJ	A	2003	5/5	0.89	0.33	113,168,172,175	0
3	THJ	E	504	5/5	0.90	0.17	143,181,184,186	0
3	THJ	J	503	5/5	0.90	0.21	148,157,160,161	0
3	THJ	A	2004	5/5	0.90	0.28	171,175,180,224	0
2	K	A	2001	1/1	0.91	0.25	117,117,117,117	0
3	THJ	D	502	5/5	0.93	0.12	117,130,149,151	0
3	THJ	B	501	5/5	0.94	0.14	112,125,134,139	0
3	THJ	A	2002	5/5	0.94	0.16	127,139,151,160	0
3	THJ	J	502	5/5	0.95	0.28	115,121,128,168	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
3	THJ	I	502	5/5	0.97	0.24	164,190,194,196	0
3	THJ	D	503	5/5	0.97	0.28	78,94,109,110	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.