wwPDB X-ray Structure Validation Summary Report (i)

Aug 16, 2022 – 01:28 pm BST

PDB ID	:	7QX8
Title	:	Crystal structure of serine hydroxymethyltransferase, isoform 7 from Arabidopsis thaliana (SHM7)
Authors	:	Ruszkowski, M.; Grzechowiak, M.; Sekula, B.
Deposited on	:	2022-01-26
Resolution	:	2.74 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (1)) were used in the production of this report:

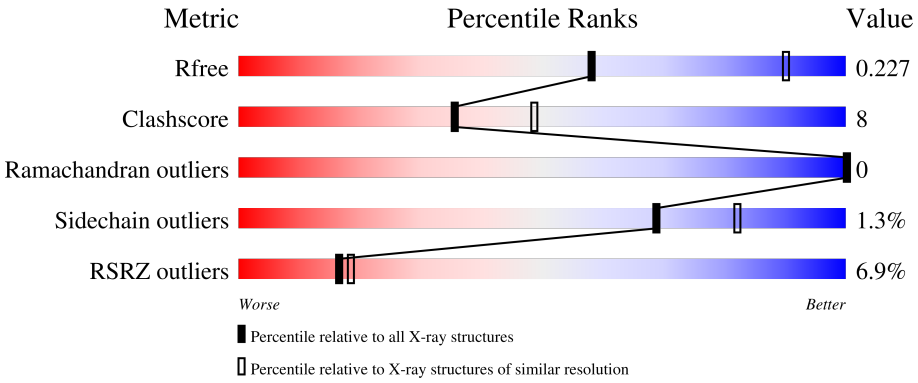
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.29
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.29

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.74 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	1271 (2.76-2.72)
Clashscore	141614	1322 (2.76-2.72)
Ramachandran outliers	138981	1297 (2.76-2.72)
Sidechain outliers	138945	1298 (2.76-2.72)
RSRZ outliers	127900	1243 (2.76-2.72)

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 of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	479	2% 79% 13% 8%
1	B	479	4% 80% 12% 8%
1	C	479	3% 80% 13% 7%
1	D	479	7% 74% 17% 9%
1	E	479	4% 79% 13% 8%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	479	12% 76% 15% 7%
1	G	479	5% 76% 17% 7%
1	H	479	11% 78% 13% 9%
1	I	479	6% 81% 11% 8%
1	J	479	2% 78% 14% 7%
1	K	479	16% 72% 19% 9%
1	L	479	5% 78% 14% 8%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 41696 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 Serine hydroxymethyltransferase 7.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	443	Total	C	N	O	S	0	0	0
			3454	2183	608	638	25			
1	B	443	Total	C	N	O	S	0	0	0
			3454	2181	608	640	25			
1	C	447	Total	C	N	O	S	0	0	0
			3486	2199	613	649	25			
1	D	438	Total	C	N	O	S	0	1	0
			3429	2167	603	634	25			
1	E	442	Total	C	N	O	S	0	1	0
			3460	2186	609	640	25			
1	F	444	Total	C	N	O	S	0	1	0
			3467	2186	612	644	25			
1	G	444	Total	C	N	O	S	0	0	0
			3460	2183	609	643	25			
1	H	438	Total	C	N	O	S	0	1	0
			3431	2169	604	633	25			
1	I	442	Total	C	N	O	S	0	1	0
			3456	2182	610	639	25			
1	J	444	Total	C	N	O	S	0	0	0
			3461	2185	609	642	25			
1	K	437	Total	C	N	O	S	0	1	0
			3419	2158	604	632	25			
1	L	443	Total	C	N	O	S	0	1	0
			3463	2187	610	641	25			

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	120	SER	-	expression tag	UNP Q84WV0
A	121	ASN	-	expression tag	UNP Q84WV0
A	122	ALA	-	expression tag	UNP Q84WV0
B	120	SER	-	expression tag	UNP Q84WV0
B	121	ASN	-	expression tag	UNP Q84WV0

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	122	ALA	-	expression tag	UNP Q84WV0
C	120	SER	-	expression tag	UNP Q84WV0
C	121	ASN	-	expression tag	UNP Q84WV0
C	122	ALA	-	expression tag	UNP Q84WV0
D	120	SER	-	expression tag	UNP Q84WV0
D	121	ASN	-	expression tag	UNP Q84WV0
D	122	ALA	-	expression tag	UNP Q84WV0
E	120	SER	-	expression tag	UNP Q84WV0
E	121	ASN	-	expression tag	UNP Q84WV0
E	122	ALA	-	expression tag	UNP Q84WV0
F	120	SER	-	expression tag	UNP Q84WV0
F	121	ASN	-	expression tag	UNP Q84WV0
F	122	ALA	-	expression tag	UNP Q84WV0
G	120	SER	-	expression tag	UNP Q84WV0
G	121	ASN	-	expression tag	UNP Q84WV0
G	122	ALA	-	expression tag	UNP Q84WV0
H	120	SER	-	expression tag	UNP Q84WV0
H	121	ASN	-	expression tag	UNP Q84WV0
H	122	ALA	-	expression tag	UNP Q84WV0
I	120	SER	-	expression tag	UNP Q84WV0
I	121	ASN	-	expression tag	UNP Q84WV0
I	122	ALA	-	expression tag	UNP Q84WV0
J	120	SER	-	expression tag	UNP Q84WV0
J	121	ASN	-	expression tag	UNP Q84WV0
J	122	ALA	-	expression tag	UNP Q84WV0
K	120	SER	-	expression tag	UNP Q84WV0
K	121	ASN	-	expression tag	UNP Q84WV0
K	122	ALA	-	expression tag	UNP Q84WV0
L	120	SER	-	expression tag	UNP Q84WV0
L	121	ASN	-	expression tag	UNP Q84WV0
L	122	ALA	-	expression tag	UNP Q84WV0

- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	32	Total O 32 32	0	0
2	B	24	Total O 24 24	0	0
2	C	35	Total O 35 35	0	0
2	D	12	Total O 12 12	0	0

Continued on next page...

Continued from previous page...

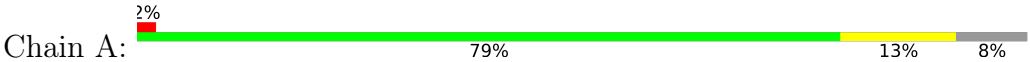
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	E	30	Total O 30 30	0	0
2	F	18	Total O 18 18	0	0
2	G	15	Total O 15 15	0	0
2	H	6	Total O 6 6	0	0
2	I	24	Total O 24 24	0	0
2	J	26	Total O 26 26	0	0
2	K	14	Total O 14 14	0	0
2	L	20	Total O 20 20	0	0

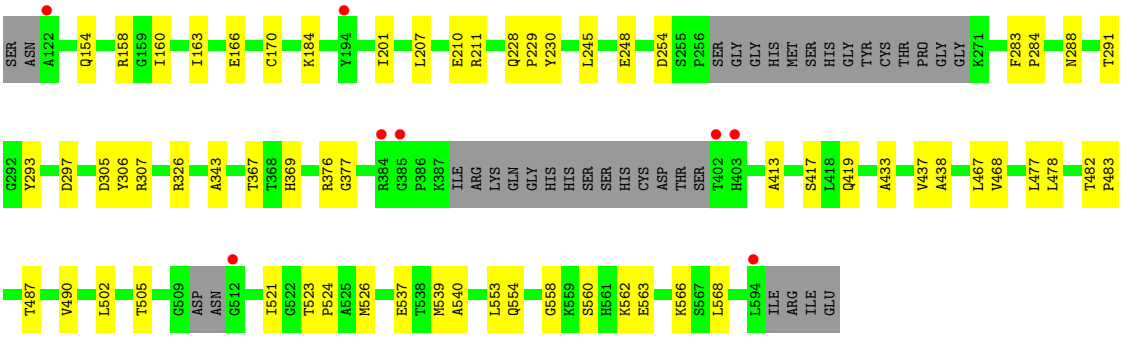
3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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: Serine hydroxymethyltransferase 7

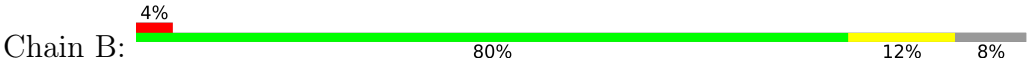
Chain A:

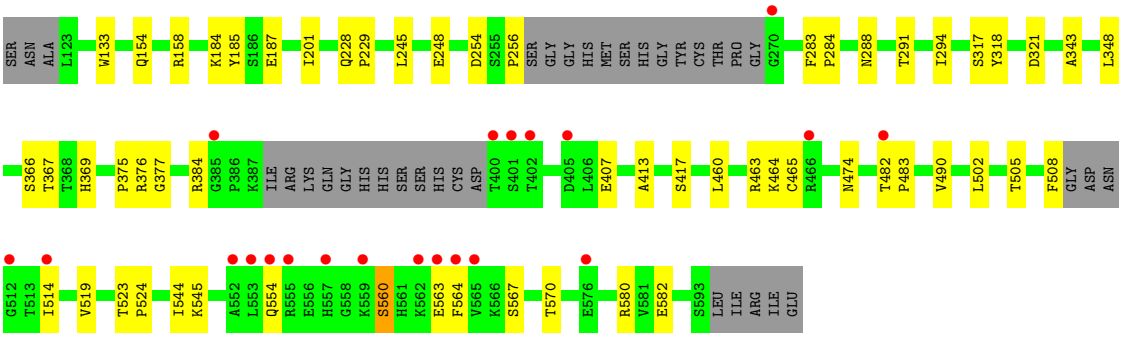




• Molecule 1: Serine hydroxymethyltransferase 7

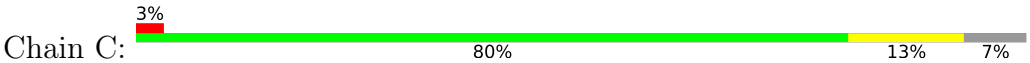
Chain B:

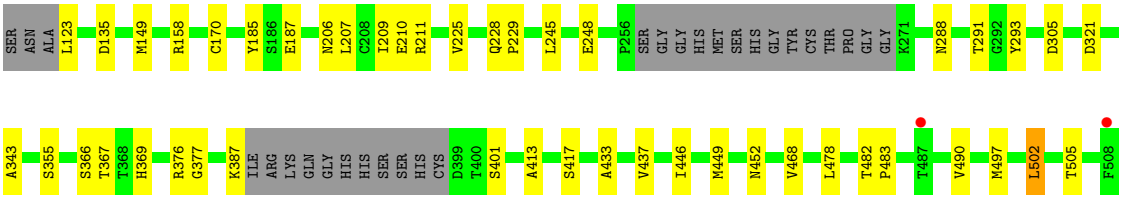


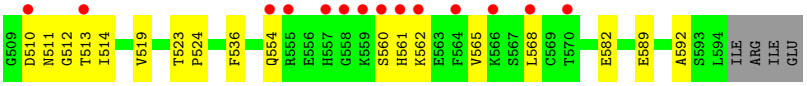


• Molecule 1: Serine hydroxymethyltransferase 7

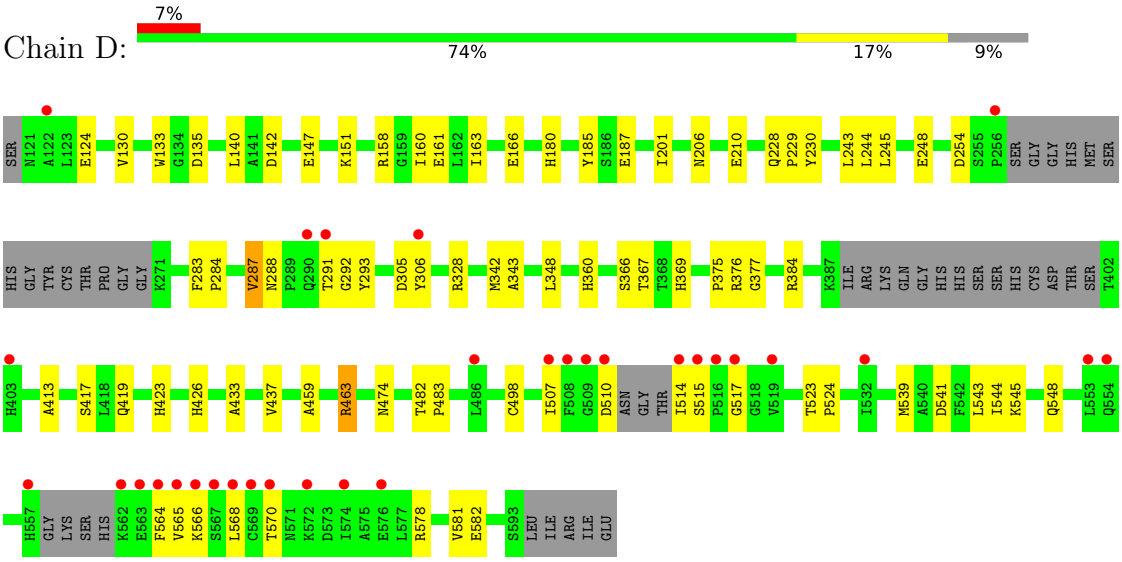
Chain C:



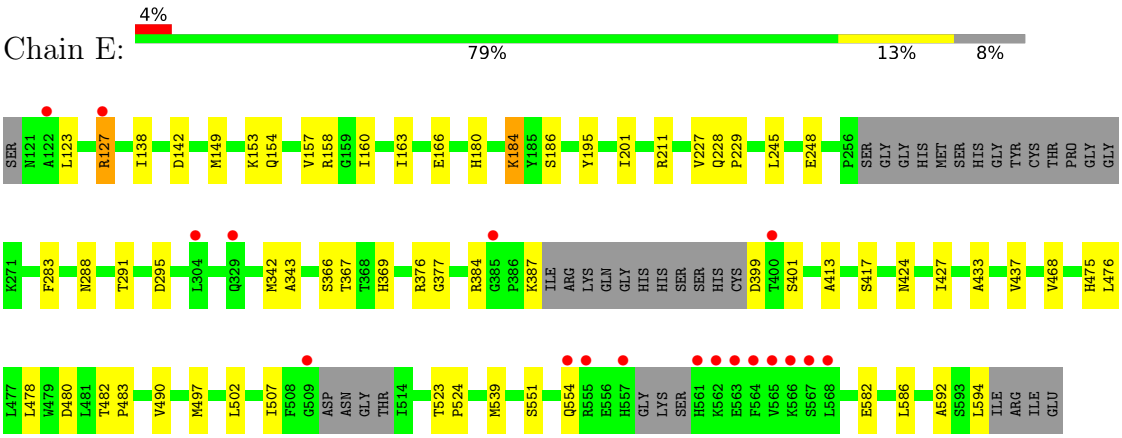




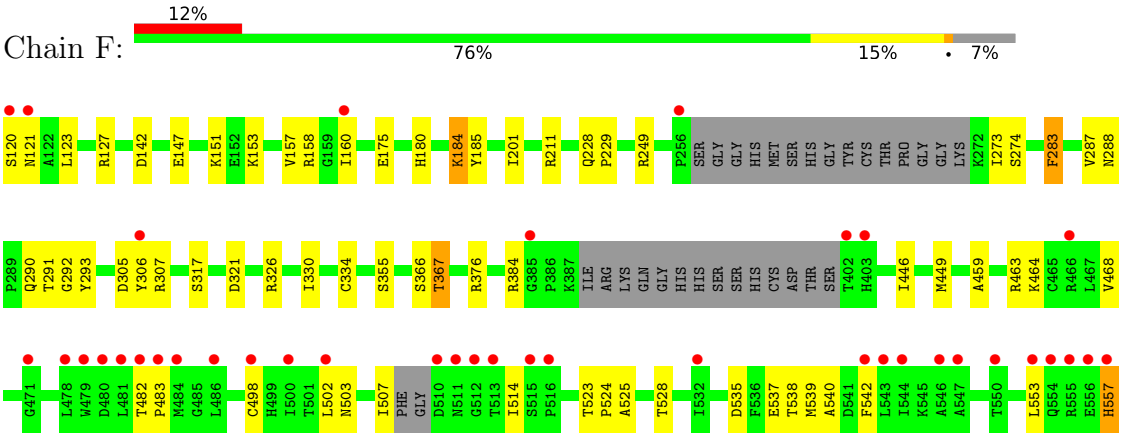
• Molecule 1: Serine hydroxymethyltransferase 7

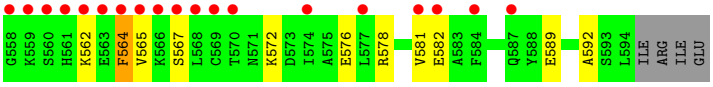


• Molecule 1: Serine hydroxymethyltransferase 7



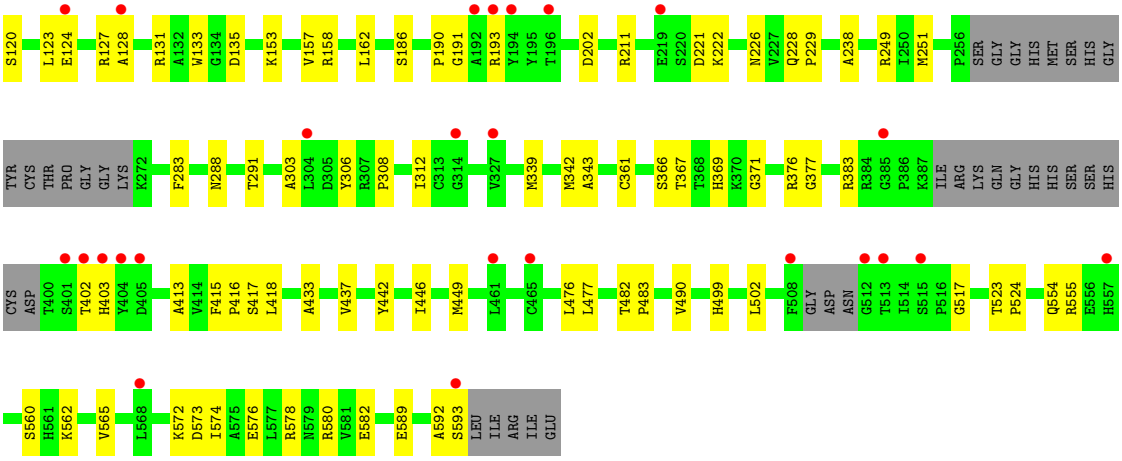
• Molecule 1: Serine hydroxymethyltransferase 7





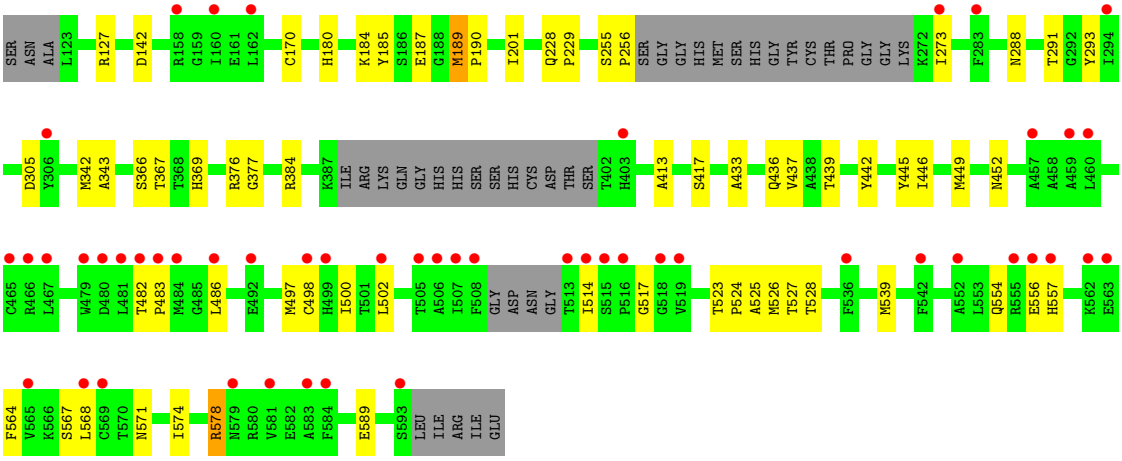
• Molecule 1: Serine hydroxymethyltransferase 7

Chain G:



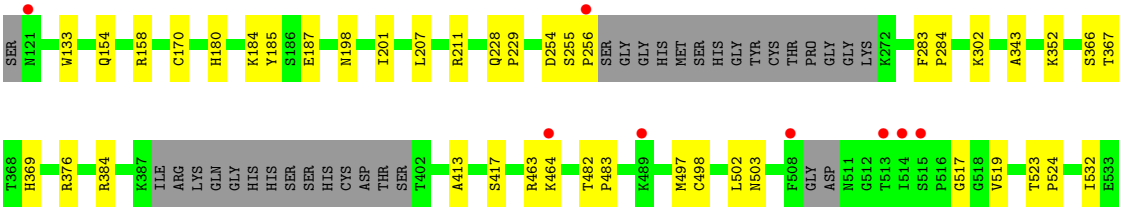
• Molecule 1: Serine hydroxymethyltransferase 7

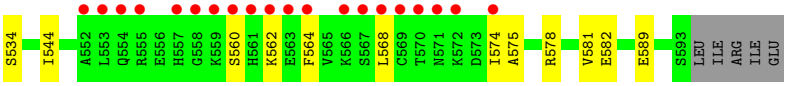
Chain H:



• Molecule 1: Serine hydroxymethyltransferase 7

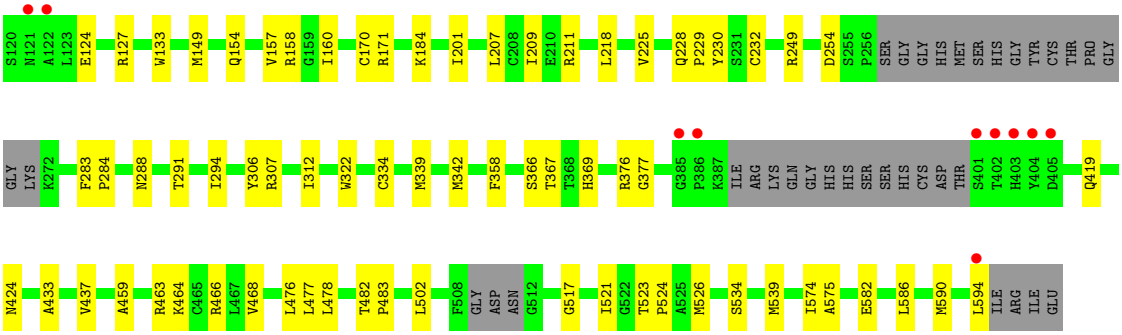
Chain I:





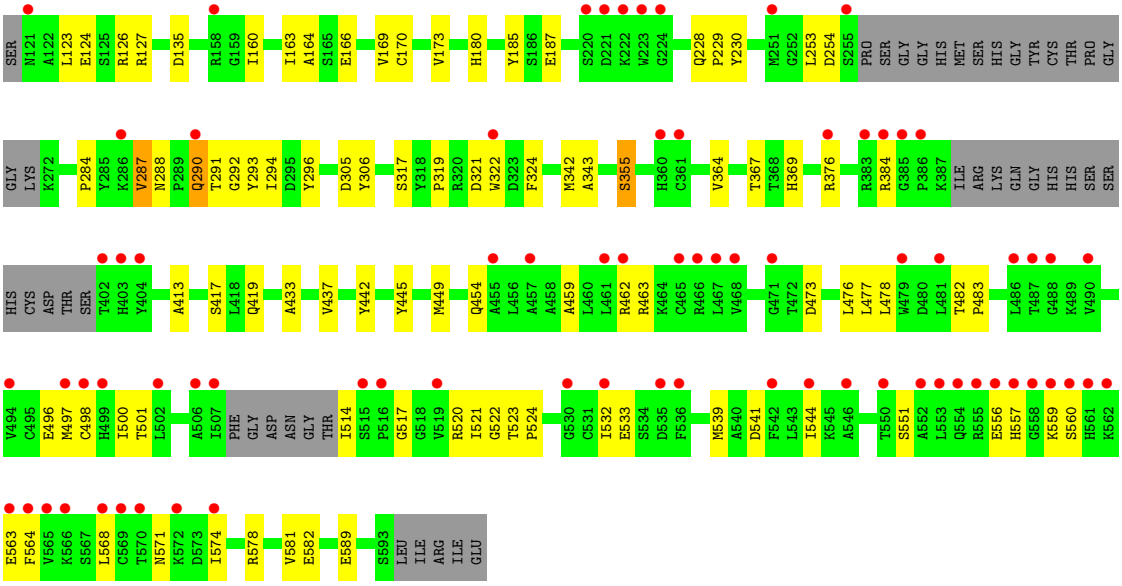
• Molecule 1: Serine hydroxymethyltransferase 7

Chain J:



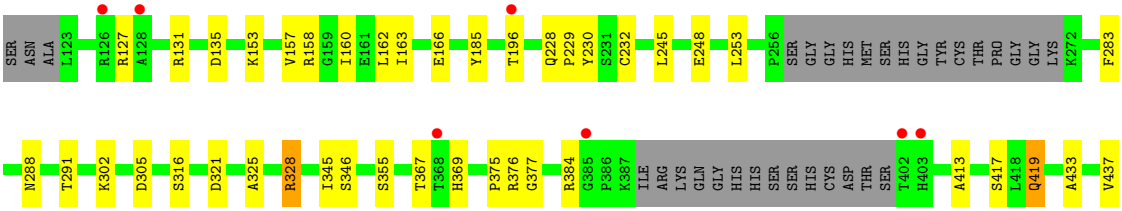
• Molecule 1: Serine hydroxymethyltransferase 7

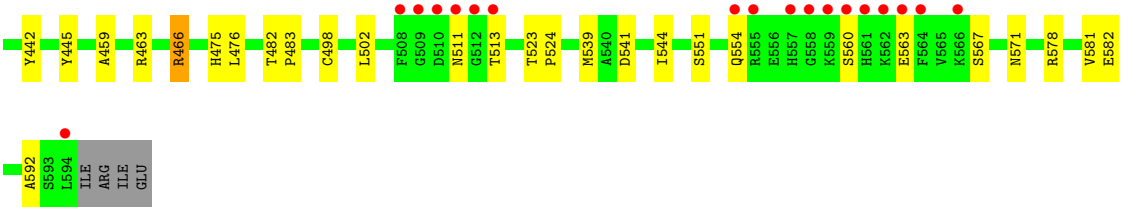
Chain K:



• Molecule 1: Serine hydroxymethyltransferase 7

Chain L:




Y442
Y445
A459
R463
R466
H475
L476
T482
P483
C498
L502
F508
G509
D510
N511
G512
T513
T523
P524
M539
A540
D541
I544
S551
Q554
R555
E556
H557
G558
K559
S560
H561
K562
E563
F564
V565
K566
S567
N571
R578
V581
E582
A592
S593
L594
ILE
ARG
ILE
GLU

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	199.83Å 123.95Å 290.54Å 90.00° 93.17° 90.00°	Depositor
Resolution (Å)	58.00 – 2.74 58.02 – 2.74	Depositor EDS
% Data completeness (in resolution range)	65.2 (58.00-2.74) 65.2 (58.02-2.74)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.76 (at 2.73Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.187 , 0.227 0.188 , 0.227	Depositor DCC
R_{free} test set	1220 reflections (1.00%)	wwPDB-VP
Wilson B-factor (Å^2)	58.1	Xtriage
Anisotropy	0.276	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	41696	wwPDB-VP
Average B, all atoms (Å^2)	82.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.59% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

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

5 Model quality

5.1 Standard geometry

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	0.74	0/3524	0.88	0/4757
1	B	0.71	0/3524	0.86	0/4757
1	C	0.74	0/3557	0.88	0/4804
1	D	0.71	0/3501	0.86	0/4726
1	E	0.73	0/3533	0.88	0/4770
1	F	0.69	0/3540	0.86	0/4781
1	G	0.68	0/3530	0.88	0/4767
1	H	0.67	0/3505	0.85	0/4733
1	I	0.73	0/3530	0.88	0/4767
1	J	0.74	0/3531	0.90	0/4768
1	K	0.67	0/3491	0.83	0/4713
1	L	0.69	0/3538	0.87	0/4779
All	All	0.71	0/42304	0.87	0/57122

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

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	3454	0	3446	50	0
1	B	3454	0	3442	44	0
1	C	3486	0	3468	55	0
1	D	3429	0	3413	75	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	E	3460	0	3443	52	0
1	F	3467	0	3449	66	0
1	G	3460	0	3442	72	0
1	H	3431	0	3418	62	0
1	I	3456	0	3438	48	0
1	J	3461	0	3446	56	0
1	K	3419	0	3406	94	0
1	L	3463	0	3446	58	0
2	A	32	0	0	2	0
2	B	24	0	0	0	0
2	C	35	0	0	2	0
2	D	12	0	0	0	0
2	E	30	0	0	0	0
2	F	18	0	0	5	0
2	G	15	0	0	1	0
2	H	6	0	0	0	0
2	I	24	0	0	1	0
2	J	26	0	0	1	0
2	K	14	0	0	1	0
2	L	20	0	0	0	0
All	All	41696	0	41257	668	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 8.

The worst 5 of 668 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:A:526:MET:HG3	1:A:539:MET:HE1	1.27	1.12
1:I:497:MET:HG3	1:I:568:LEU:HD21	1.11	1.10
1:H:452:ASN:HD22	1:H:527:THR:HG22	1.10	1.08
1:G:249:ARG:HB3	1:G:306:TYR:HE1	1.16	1.05
1:B:560:SER:OG	1:B:563:GLU:HB2	1.60	1.01

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	435/479 (91%)	417 (96%)	18 (4%)	0	100	100
1	B	435/479 (91%)	418 (96%)	17 (4%)	0	100	100
1	C	441/479 (92%)	419 (95%)	22 (5%)	0	100	100
1	D	429/479 (90%)	411 (96%)	18 (4%)	0	100	100
1	E	433/479 (90%)	415 (96%)	18 (4%)	0	100	100
1	F	437/479 (91%)	417 (95%)	20 (5%)	0	100	100
1	G	436/479 (91%)	418 (96%)	18 (4%)	0	100	100
1	H	431/479 (90%)	410 (95%)	21 (5%)	0	100	100
1	I	435/479 (91%)	418 (96%)	17 (4%)	0	100	100
1	J	436/479 (91%)	419 (96%)	17 (4%)	0	100	100
1	K	430/479 (90%)	410 (95%)	20 (5%)	0	100	100
1	L	438/479 (91%)	409 (93%)	29 (7%)	0	100	100
All	All	5216/5748 (91%)	4981 (96%)	235 (4%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	366/396 (92%)	365 (100%)	1 (0%)	92	95
1	B	367/396 (93%)	363 (99%)	4 (1%)	73	84

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	C	371/396 (94%)	365 (98%)	6 (2%)	62 78
1	D	364/396 (92%)	357 (98%)	7 (2%)	57 74
1	E	368/396 (93%)	362 (98%)	6 (2%)	62 78
1	F	369/396 (93%)	357 (97%)	12 (3%)	38 59
1	G	368/396 (93%)	366 (100%)	2 (0%)	88 92
1	H	365/396 (92%)	361 (99%)	4 (1%)	73 84
1	I	367/396 (93%)	366 (100%)	1 (0%)	92 95
1	J	368/396 (93%)	367 (100%)	1 (0%)	92 95
1	K	363/396 (92%)	354 (98%)	9 (2%)	47 67
1	L	368/396 (93%)	364 (99%)	4 (1%)	73 84
All	All	4404/4752 (93%)	4347 (99%)	57 (1%)	69 82

5 of 57 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	317	SER
1	L	419	GLN
1	F	564	PHE
1	L	384	ARG
1	K	384	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	H	579	ASN
1	I	199	GLN
1	L	369	HIS
1	K	290	GLN
1	K	557	HIS

5.3.3 RNA

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

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

6.1 Protein, DNA and RNA chains

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^{th} 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^2)	Q<0.9
1	A	443/479 (92%)	0.11	8 (1%) 68 74	43, 56, 97, 147	0
1	B	443/479 (92%)	0.23	21 (4%) 31 34	39, 70, 125, 174	0
1	C	447/479 (93%)	0.05	16 (3%) 42 47	42, 56, 124, 183	0
1	D	438/479 (91%)	0.37	32 (7%) 15 17	41, 79, 135, 179	0
1	E	442/479 (92%)	0.31	18 (4%) 37 41	41, 64, 122, 195	0
1	F	444/479 (92%)	0.70	58 (13%) 3 3	41, 89, 151, 209	0
1	G	444/479 (92%)	0.44	25 (5%) 24 27	53, 97, 142, 178	0
1	H	438/479 (91%)	0.58	51 (11%) 4 4	53, 94, 180, 219	0
1	I	442/479 (92%)	0.31	28 (6%) 20 22	41, 64, 143, 193	0
1	J	444/479 (92%)	0.15	10 (2%) 60 67	41, 66, 107, 134	0
1	K	437/479 (91%)	0.85	75 (17%) 1 1	46, 106, 156, 199	0
1	L	443/479 (92%)	0.32	25 (5%) 24 27	51, 75, 139, 182	0
All	All	5305/5748 (92%)	0.37	367 (6%) 16 18	39, 74, 146, 219	0

The worst 5 of 367 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	568	LEU	8.0
1	E	564	PHE	7.1
1	F	563	GLU	6.5
1	H	557	HIS	6.4
1	B	565	VAL	6.4

6.2 Non-standard residues in protein, DNA, RNA chains

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

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.