wwPDB X-ray Structure Validation Summary Report (i)

Nov 22, 2023 – 04:22 PM JST

PDB ID	:	7DRS
Title	:	Structure of SspE_40224
Authors	:	Haiyan, G.; Jinchuan, Z.; Chen, S.; Wang, L.; Wu, G.
Deposited on	:	2020-12-29
Resolution	:	3.42 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 (i)) were used in the production of this report:

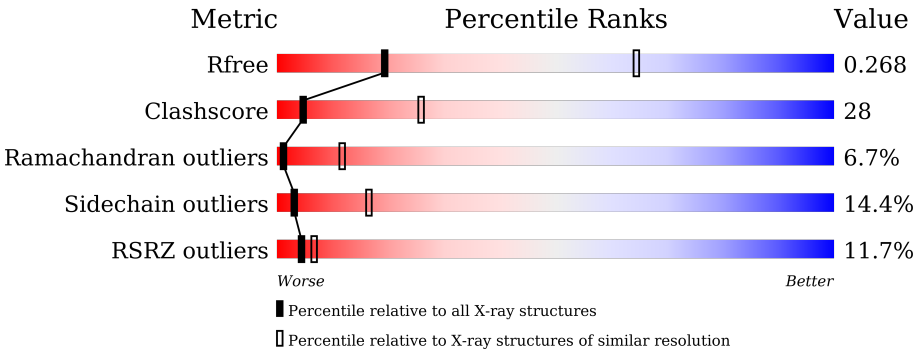
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 3.42 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	1486 (3.50-3.34)
Clashscore	141614	1572 (3.50-3.34)
Ramachandran outliers	138981	1534 (3.50-3.34)
Sidechain outliers	138945	1535 (3.50-3.34)
RSRZ outliers	127900	1395 (3.50-3.34)

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	771	7% 46% 43% 10% ·
1	B	771	23% 48% 39% 11% ·
1	C	771	7% 45% 41% 12% ·
1	D	771	9% 45% 42% 11% ·

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 24548 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 SspE protein.

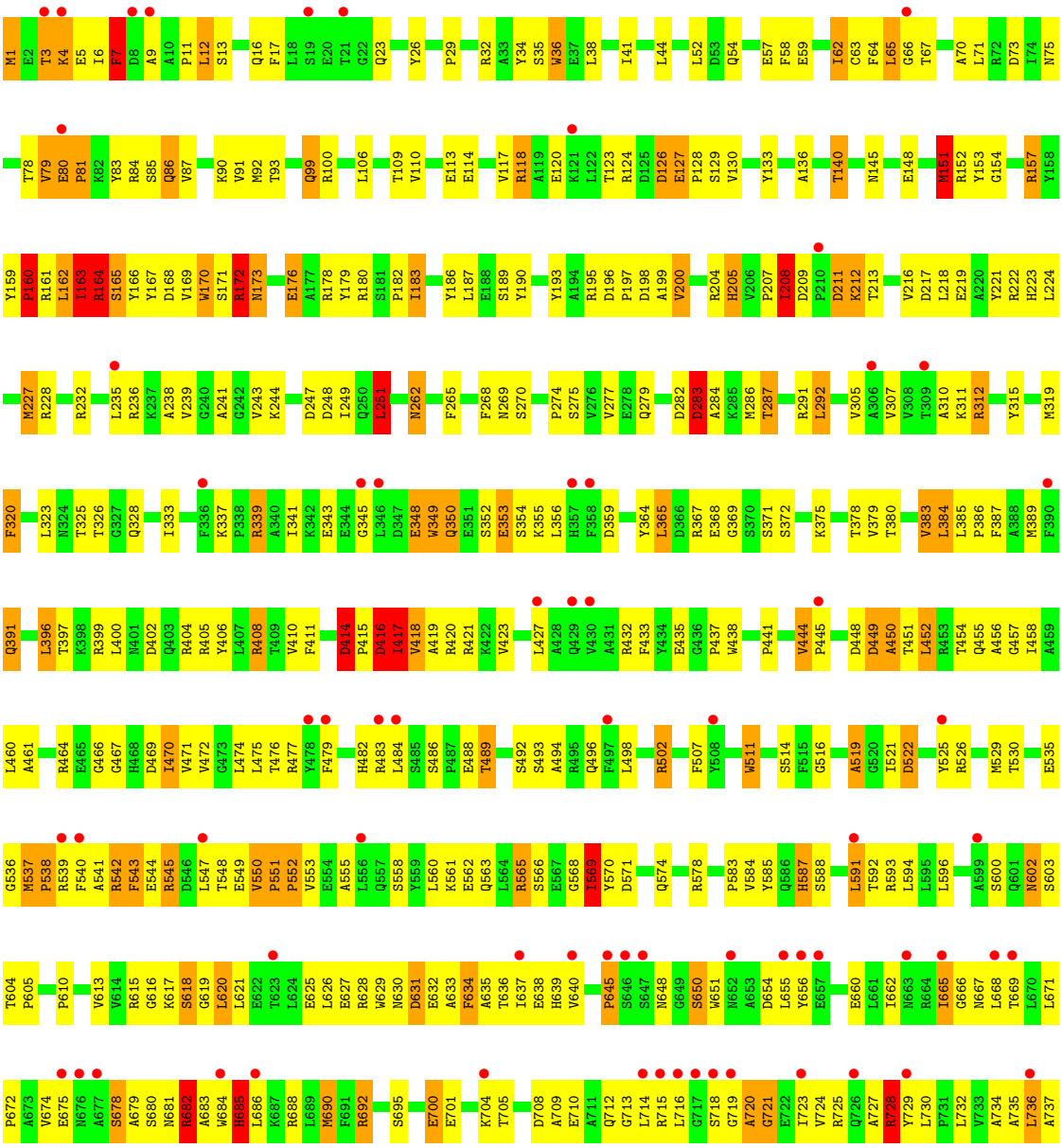
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	D	771	Total	C	N	O	S	0	0	0
			6137	3872	1079	1166	20			
1	A	771	Total	C	N	O	S	0	0	0
			6137	3872	1079	1166	20			
1	B	771	Total	C	N	O	S	0	0	0
			6137	3872	1079	1166	20			
1	C	771	Total	C	N	O	S	0	0	0
			6137	3872	1079	1166	20			

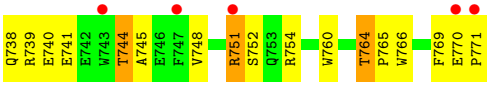
3 Residue-property plots (i)

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: SspE protein

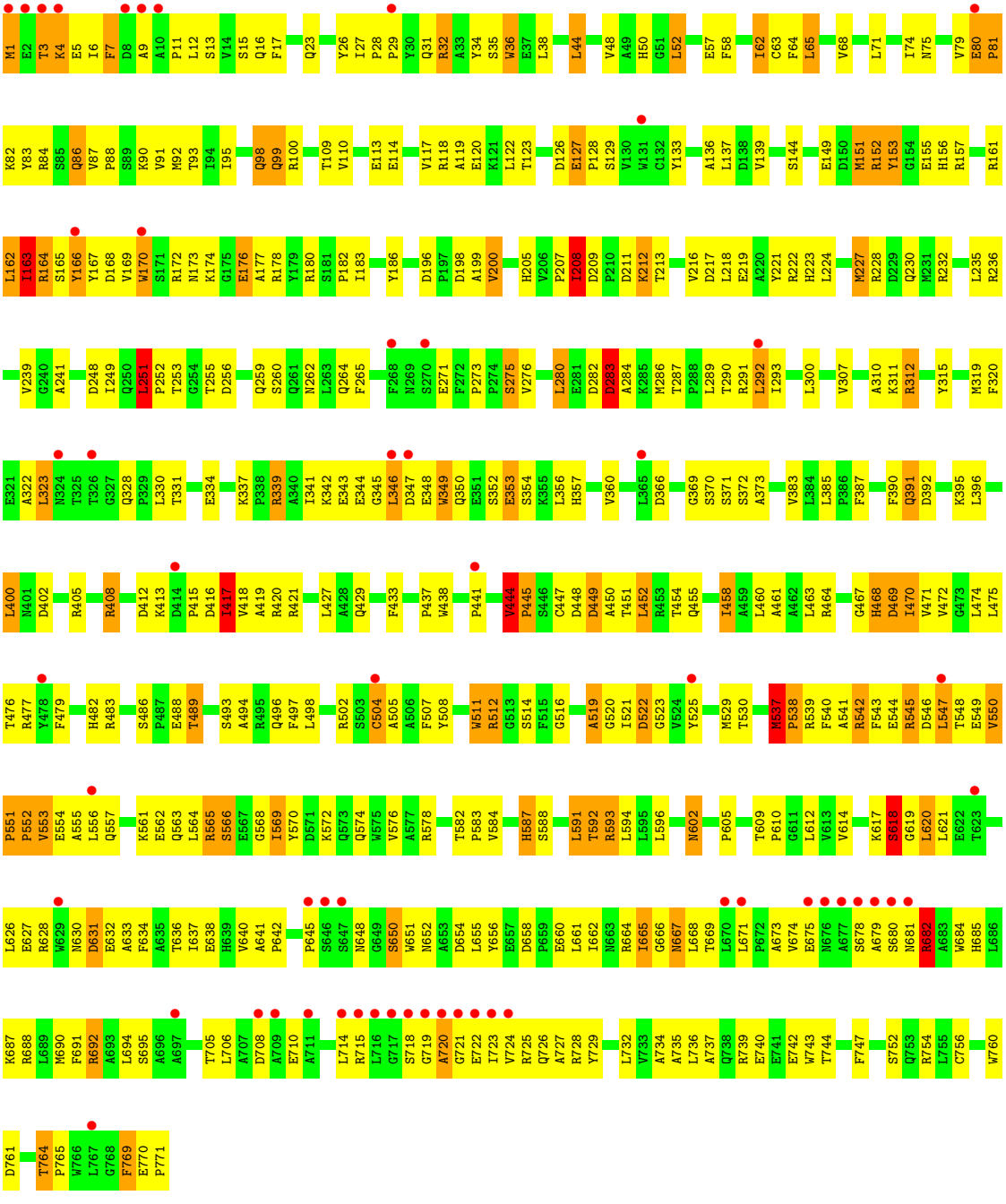
Chain D:





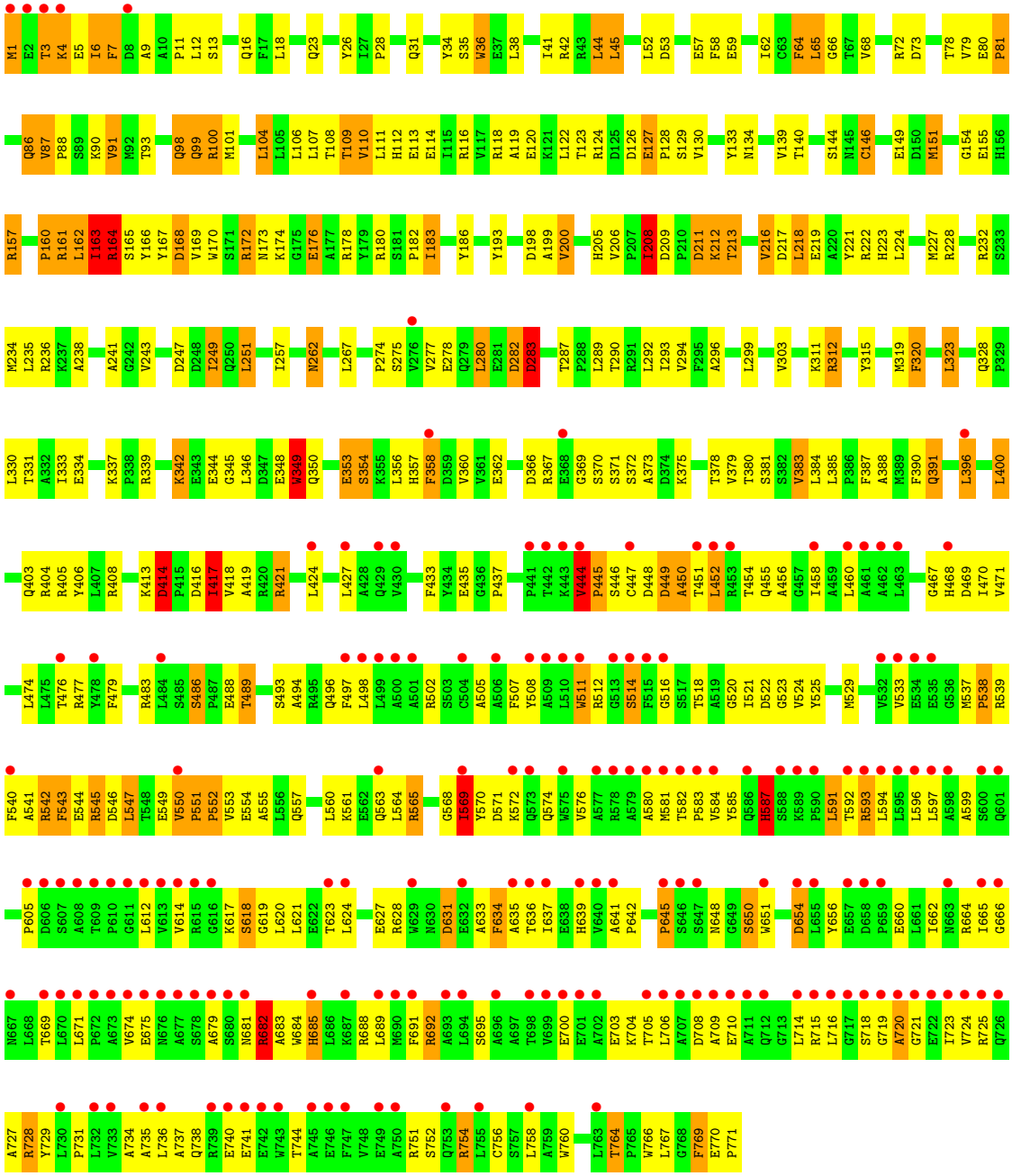
• Molecule 1: SspE protein

Chain A:



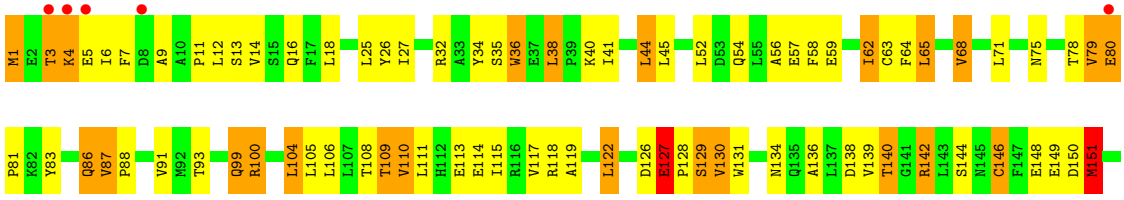
• Molecule 1: SspE protein

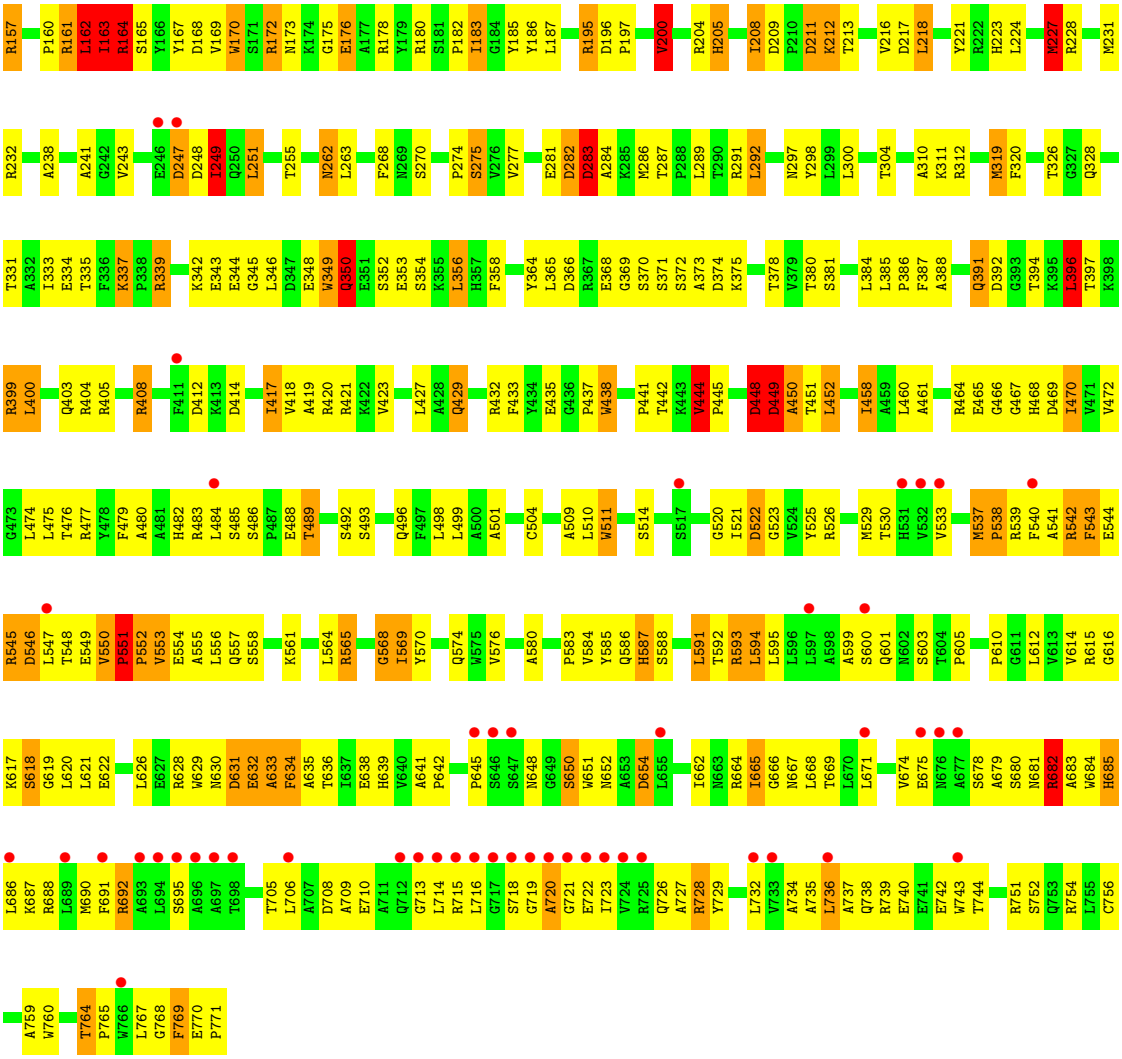
Chain B:



• Molecule 1: SspE protein

Chain C:




R157 P160 R161 L162 I163 R164 S165 Y166 Y167 D168 V169 W170 S171 R172 N173 K174 G175 E176 A177 R178 Y179 R180 S181 P182 I183 G184 Y185 Y186 L187 R195 D196 P197 V200 R204 H205 I208 D209 P210 D211 K212 T213 V216 D217 L218 Y221 R222 H223 L224 M227 R228 M231
R232 A238 A241 G242 V243 E246 D247 D248 I249 Q250 L251 T255 N262 L263 F268 N269 S270 P274 S275 V276 V277 E281 D282 D283 A284 K285 M286 T287 P288 L289 T290 R291 L292 N297 Y298 L299 L300 T304 A310 K311 R312 M319 F320 T326 G327 Q328
T331 A332 I333 E334 T335 F336 K337 P338 R339 K342 E343 E344 G345 L346 D347 E348 W349 Q350 E351 S352 E353 S354 K355 L356 H357 F358 Y364 L365 D366 R367 E368 G369 S370 S371 S372 A373 D374 K375 T378 V379 T380 S381 L384 L385 P386 F387 A388 Q391 D392 G393 T394 K395 L396 T397 K398
R399 L400 Q403 R404 R405 R408 F411 D412 K413 D414 I417 V418 A419 R420 R421 K422 V423 L427 A428 Q429 R432 F433 Y434 E435 G436 P437 W438 P441 T442 K443 V444 P445 D448 D449 A450 T451 L452 I458 A459 L460 A461 R464 E465 G466 G467 H468 D469 I470 V471 V472
G473 L474 L475 T476 R477 Y478 F479 A480 A481 H482 R483 L484 S485 S486 P487 E488 T489 S492 S493 Q496 F497 L498 L499 A500 A501 C504 A509 L510 W511 S514 S517 G520 I521 D522 G523 V524 Y525 R526 M529 T530 H531 V532 V533 M537 P538 R539 F540 A541 R542 F543 E544
R545 D546 L547 T548 E549 V550 P551 P552 V553 E554 A555 L556 Q557 S558 K561 L564 R565 G568 I569 Y570 Q574 W575 V576 A580 P583 V584 Y585 Q586 H587 S588 L591 T592 R593 L594 L595 L596 L597 A598 A599 S600 Q601 N602 S603 T604 P605 P610 G611 L612 V613 V614 R615 G616
K617 S618 G619 L620 L621 E622 L626 E627 R628 W629 N630 D631 E632 A633 F634 A635 T636 I637 E638 H639 V640 A641 P642 P645 S646 S647 N648 G649 S650 W651 N652 A653 D654 L655 I662 N663 R664 I665 G666 N667 L668 T669 L670 L671 V674 E675 N676 A677 S678 A679 S680 N681 R682 A683 W684 H685
L686 K687 R688 L689 M690 F691 R692 A693 L694 S695 A696 A697 T698 T705 L706 A707 D708 A709 E710 A711 Q712 G713 L714 R715 L716 G717 S718 G719 A720 G721 E722 I723 V724 R725 Q726 A727 R728 Y729 L732 V733 A734 A735 L736 A737 Q738 R739 E740 E741 E742 W743 T744 R751 S752 Q753 R754 L755 C756
A759 W760 T764 P765 W766 L767 G768 F769 E770 P771

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	111.08Å 138.15Å 293.69Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.14 – 3.42 51.95 – 3.42	Depositor EDS
% Data completeness (in resolution range)	99.7 (46.14-3.42) 99.7 (51.95-3.42)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.40 (at 3.40Å)	Xtriage
Refinement program	REFMAC 5.5	Depositor
R, R_{free}	0.212 , 0.264 0.221 , 0.268	Depositor DCC
R_{free} test set	3031 reflections (4.91%)	wwPDB-VP
Wilson B-factor (Å^2)	121.6	Xtriage
Anisotropy	0.343	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 100.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	24548	wwPDB-VP
Average B, all atoms (Å^2)	156.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.48% 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	0.63	2/6272 (0.0%)	0.85	8/8515 (0.1%)
1	B	0.60	3/6272 (0.0%)	0.83	11/8515 (0.1%)
1	C	0.79	9/6272 (0.1%)	0.98	24/8515 (0.3%)
1	D	0.70	5/6272 (0.1%)	0.91	13/8515 (0.2%)
All	All	0.69	19/25088 (0.1%)	0.90	56/34060 (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	2
1	B	0	1
1	C	0	5
1	D	0	3
All	All	0	11

The worst 5 of 19 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	80	GLU	CB-CG	7.56	1.66	1.52
1	C	146	CYS	CB-SG	-6.49	1.71	1.82
1	C	185	TYR	CG-CD1	-6.38	1.30	1.39
1	C	63	CYS	CB-SG	-6.34	1.71	1.82
1	B	349	TRP	CB-CG	6.06	1.61	1.50

The worst 5 of 56 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	251	LEU	CA-CB-CG	10.36	139.12	115.30
1	B	346	LEU	CA-CB-CG	9.05	136.11	115.30
1	C	251	LEU	CA-CB-CG	8.93	135.85	115.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	251	LEU	CA-CB-CG	8.93	135.83	115.30
1	A	346	LEU	CA-CB-CG	8.91	135.79	115.30

There are no chirality outliers.

5 of 11 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	207	PRO	Peptide
1	A	553	VAL	Peptide
1	D	160	PRO	Peptide
1	D	173	ASN	Peptide
1	D	79	VAL	Peptide

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	6137	0	6046	338	1
1	B	6137	0	6046	345	1
1	C	6137	0	6045	353	2
1	D	6137	0	6046	351	2
All	All	24548	0	24183	1365	3

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 1365 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:628:ARG:HA	1:C:631:ASP:HB2	1.44	0.99
1:A:418:VAL:H	1:A:421:ARG:HD2	1.25	0.98
1:B:126:ASP:HB3	1:B:129:SER:HB3	1.45	0.98
1:A:126:ASP:HB3	1:A:129:SER:HB3	1.44	0.96
1:D:126:ASP:HB3	1:D:129:SER:HB3	1.47	0.95

All (3) symmetry-related close contacts are listed below. The label for Atom-2 includes the sym-

metry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:85:SER:N	1:C:211:ASP:OD2[4_457]	1.94	0.26
1:D:84:ARG:N	1:C:211:ASP:OD2[4_457]	2.13	0.07
1:A:236:ARG:NH1	1:B:278:GLU:OE1[4_447]	2.18	0.02

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	769/771 (100%)	613 (80%)	105 (14%)	51 (7%)	1	12
1	B	769/771 (100%)	605 (79%)	117 (15%)	47 (6%)	1	13
1	C	769/771 (100%)	612 (80%)	104 (14%)	53 (7%)	1	11
1	D	769/771 (100%)	609 (79%)	105 (14%)	55 (7%)	1	10
All	All	3076/3084 (100%)	2439 (79%)	431 (14%)	206 (7%)	1	12

5 of 206 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	4	LYS
1	D	36	TRP
1	D	127	GLU
1	D	200	VAL
1	D	205	HIS

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	651/651 (100%)	560 (86%)	91 (14%)	3 17
1	B	651/651 (100%)	558 (86%)	93 (14%)	3 17
1	C	651/651 (100%)	553 (85%)	98 (15%)	3 15
1	D	651/651 (100%)	558 (86%)	93 (14%)	3 17
All	All	2604/2604 (100%)	2229 (86%)	375 (14%)	3 16

5 of 375 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	323	LEU
1	C	75	ASN
1	B	358	PHE
1	B	547	LEU
1	C	164	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	401	ASN
1	C	391	GLN
1	A	667	ASN
1	C	455	GLN
1	C	223	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)

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	771/771 (100%)	0.52	57 (7%) 14 17	84, 138, 228, 331	0
1	B	771/771 (100%)	1.33	178 (23%) 0 1	74, 180, 334, 405	0
1	C	771/771 (100%)	0.56	54 (7%) 16 19	57, 129, 243, 347	0
1	D	771/771 (100%)	0.63	71 (9%) 9 11	67, 143, 228, 341	0
All	All	3084/3084 (100%)	0.76	360 (11%) 4 7	57, 142, 295, 405	0

The worst 5 of 360 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	722	GLU	30.8
1	B	637	ILE	16.5
1	D	718	SER	15.5
1	C	696	ALA	14.8
1	B	721	GLY	13.5

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

There are no monosaccharides in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.