W O R L D W I D E

wwPDB

P R O T E I N D A T A B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Sep 5, 2023 – 06:13 PM EDT

PDB ID	:	8EXD
Title	:	Crystal structure of Aspergillus fumigatus sterylglucosidase A
Authors	:	Pereira de Sa, N.; Del Poeta, M.; Airola, M.V.
Deposited on	:	2022-10-25
Resolution	:	3.80 Å(reported)

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

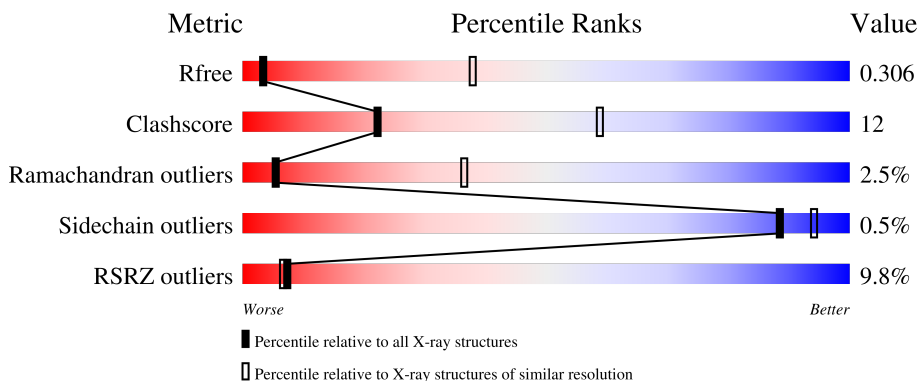|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 3.80 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	Percentile Ranks	Value
Rfree		0.306
Clashscore		12
Ramachandran outliers		2.5%
Sidechain outliers		0.5%
RSRZ outliers		9.8%

Worse   *Better*

▌ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1212 (4.00-3.60)
Clashscore	141614	1288 (4.00-3.60)
Ramachandran outliers	138981	1243 (4.00-3.60)
Sidechain outliers	138945	1237 (4.00-3.60)
RSRZ outliers	127900	1121 (4.00-3.60)

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	779	6% / 61% 23% • 14%
1	B	779	6% / 63% 20% • 15%
1	C	779	12% / 63% 20% • 16%
1	D	779	9% / 59% 20% • 20%

WORLDWIDE
wwPDB
PROTEIN DATA BANK

2 Entry composition ⓘ

There is only 1 type of molecule in this entry. The entry contains 21118 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 Sterylglucosidase A (SglA).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	670	Total	C	N	O	S	0	0	0
			5426	3486	921	1002	17			
1	B	659	Total	C	N	O	S	0	0	0
			5338	3426	908	987	17			
1	C	655	Total	C	N	O	S	0	0	0
			5309	3411	900	981	17			
1	D	623	Total	C	N	O	S	0	0	0
			5045	3248	852	929	16			

There are 80 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-19	MET	-	initiating methionine	UNP Q4WXA7
A	-18	GLY	-	expression tag	UNP Q4WXA7
A	-17	SER	-	expression tag	UNP Q4WXA7
A	-16	SER	-	expression tag	UNP Q4WXA7
A	-15	HIS	-	expression tag	UNP Q4WXA7
A	-14	HIS	-	expression tag	UNP Q4WXA7
A	-13	HIS	-	expression tag	UNP Q4WXA7
A	-12	HIS	-	expression tag	UNP Q4WXA7
A	-11	HIS	-	expression tag	UNP Q4WXA7
A	-10	HIS	-	expression tag	UNP Q4WXA7
A	-9	SER	-	expression tag	UNP Q4WXA7
A	-8	SER	-	expression tag	UNP Q4WXA7
A	-7	GLY	-	expression tag	UNP Q4WXA7
A	-6	LEU	-	expression tag	UNP Q4WXA7
A	-5	VAL	-	expression tag	UNP Q4WXA7
A	-4	PRO	-	expression tag	UNP Q4WXA7
A	-3	ARG	-	expression tag	UNP Q4WXA7
A	-2	GLY	-	expression tag	UNP Q4WXA7
A	-1	SER	-	expression tag	UNP Q4WXA7
A	0	HIS	-	expression tag	UNP Q4WXA7
B	-19	MET	-	initiating methionine	UNP Q4WXA7

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-18	GLY	-	expression tag	UNP Q4WXA7
B	-17	SER	-	expression tag	UNP Q4WXA7
B	-16	SER	-	expression tag	UNP Q4WXA7
B	-15	HIS	-	expression tag	UNP Q4WXA7
B	-14	HIS	-	expression tag	UNP Q4WXA7
B	-13	HIS	-	expression tag	UNP Q4WXA7
B	-12	HIS	-	expression tag	UNP Q4WXA7
B	-11	HIS	-	expression tag	UNP Q4WXA7
B	-10	HIS	-	expression tag	UNP Q4WXA7
B	-9	SER	-	expression tag	UNP Q4WXA7
B	-8	SER	-	expression tag	UNP Q4WXA7
B	-7	GLY	-	expression tag	UNP Q4WXA7
B	-6	LEU	-	expression tag	UNP Q4WXA7
B	-5	VAL	-	expression tag	UNP Q4WXA7
B	-4	PRO	-	expression tag	UNP Q4WXA7
B	-3	ARG	-	expression tag	UNP Q4WXA7
B	-2	GLY	-	expression tag	UNP Q4WXA7
B	-1	SER	-	expression tag	UNP Q4WXA7
B	0	HIS	-	expression tag	UNP Q4WXA7
C	-19	MET	-	initiating methionine	UNP Q4WXA7
C	-18	GLY	-	expression tag	UNP Q4WXA7
C	-17	SER	-	expression tag	UNP Q4WXA7
C	-16	SER	-	expression tag	UNP Q4WXA7
C	-15	HIS	-	expression tag	UNP Q4WXA7
C	-14	HIS	-	expression tag	UNP Q4WXA7
C	-13	HIS	-	expression tag	UNP Q4WXA7
C	-12	HIS	-	expression tag	UNP Q4WXA7
C	-11	HIS	-	expression tag	UNP Q4WXA7
C	-10	HIS	-	expression tag	UNP Q4WXA7
C	-9	SER	-	expression tag	UNP Q4WXA7
C	-8	SER	-	expression tag	UNP Q4WXA7
C	-7	GLY	-	expression tag	UNP Q4WXA7
C	-6	LEU	-	expression tag	UNP Q4WXA7
C	-5	VAL	-	expression tag	UNP Q4WXA7
C	-4	PRO	-	expression tag	UNP Q4WXA7
C	-3	ARG	-	expression tag	UNP Q4WXA7
C	-2	GLY	-	expression tag	UNP Q4WXA7
C	-1	SER	-	expression tag	UNP Q4WXA7
C	0	HIS	-	expression tag	UNP Q4WXA7
D	-19	MET	-	initiating methionine	UNP Q4WXA7
D	-18	GLY	-	expression tag	UNP Q4WXA7
D	-17	SER	-	expression tag	UNP Q4WXA7

Continued on next page...

Continued from previous page...

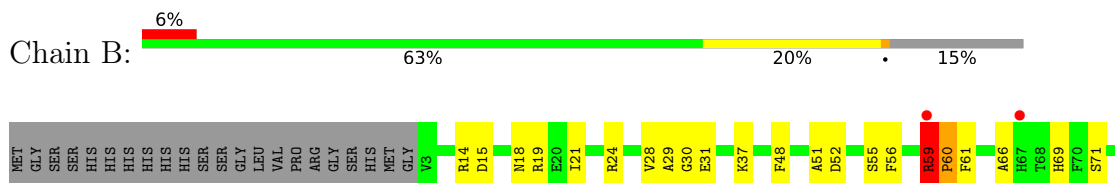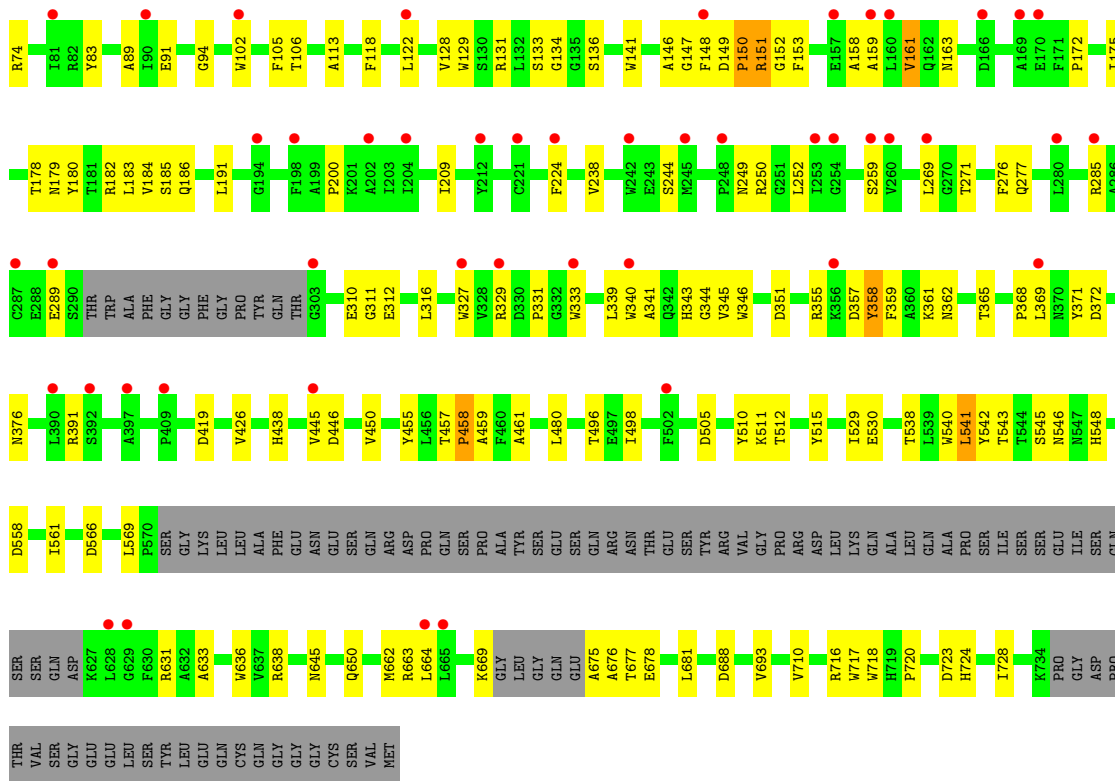| Chain	Residue	Modelled	Actual	Comment	Reference
D	-16	SER	-	expression tag	UNP Q4WXA7
D	-15	HIS	-	expression tag	UNP Q4WXA7
D	-14	HIS	-	expression tag	UNP Q4WXA7
D	-13	HIS	-	expression tag	UNP Q4WXA7
D	-12	HIS	-	expression tag	UNP Q4WXA7
D	-11	HIS	-	expression tag	UNP Q4WXA7
D	-10	HIS	-	expression tag	UNP Q4WXA7
D	-9	SER	-	expression tag	UNP Q4WXA7
D	-8	SER	-	expression tag	UNP Q4WXA7
D	-7	GLY	-	expression tag	UNP Q4WXA7
D	-6	LEU	-	expression tag	UNP Q4WXA7
D	-5	VAL	-	expression tag	UNP Q4WXA7
D	-4	PRO	-	expression tag	UNP Q4WXA7
D	-3	ARG	-	expression tag	UNP Q4WXA7
D	-2	GLY	-	expression tag	UNP Q4WXA7
D	-1	SER	-	expression tag	UNP Q4WXA7
D	0	HIS	-	expression tag	UNP Q4WXA7

3 Residue-property plots ⓘ

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.
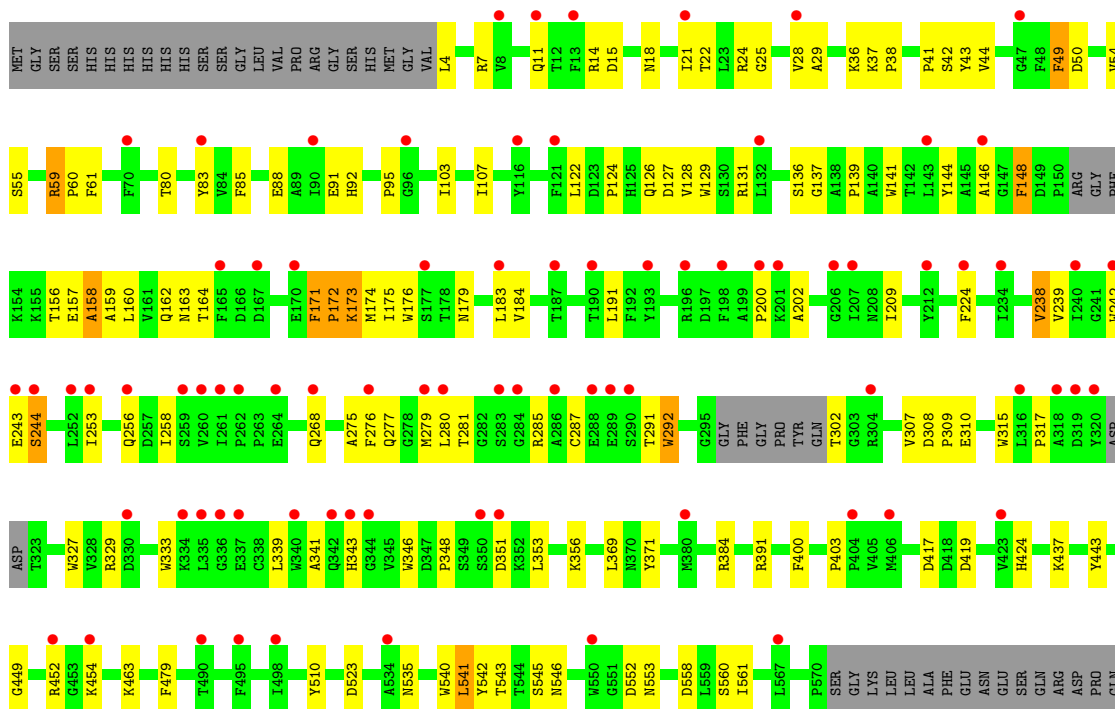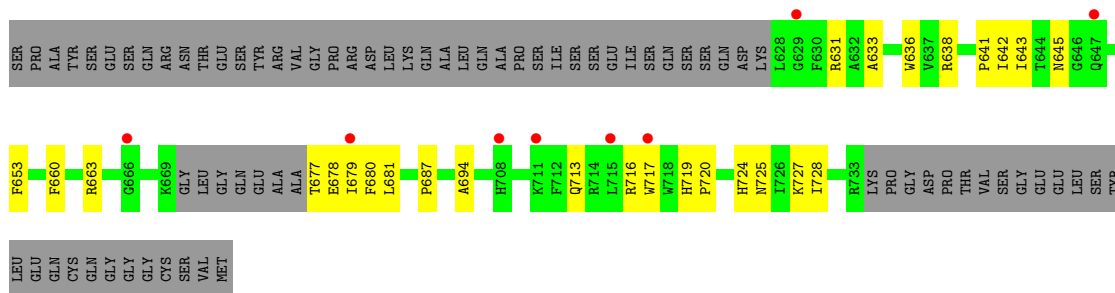
• Molecule 1: Sterylglucosidase A (SglA)
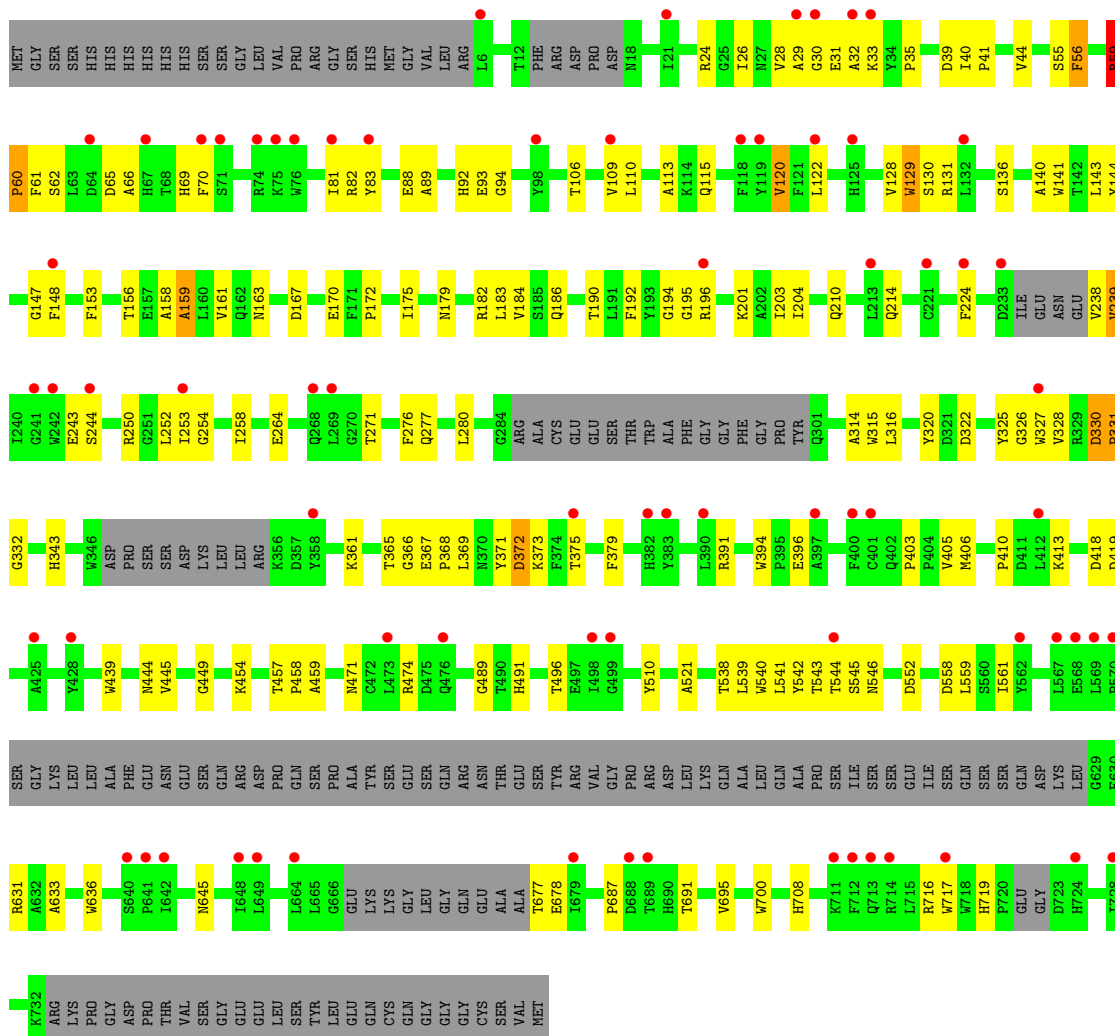

• Molecule 1: Sterylglucosidase A (SglA)

• Molecule 1: Sterylglucosidase A (SglA)

Chain C:

SER PRO ALA TYR SER GLU SER GLN ARG ASN THR GLU SER TYR ARG VAL GLY PRO ARG ASP LEU LYS GLN ALA LEU GLN ALA PRO SER ILE SER SER GLU ILE SER GLN SER SER ASP LYS L628 G629 F630 R631 A632 A633 W636 V637 R638 P641 I642 I643 T644 N645 G646 Q647

F653 F660 R663 G666 K669 GLY LEU GLY GLU ALA ALA T677 E678 I679 F680 L681 P687 A694 H708 K711 F712 Q713 R714 L715 R716 W717 W718 H719 P720 H724 N725 I726 K727 I728 R733 LYS PRO GLY ASP PRO THR VAL SER GLY GLU LEU SER TYR

LEU GLU GLN CYS GLN GLY GLY CYS SER VAL MET

● Molecule 1: Sterylglucosidase A (SglA)

Chain D:

9%

59% 20% · 20%

MET GLY SER HIS HIS HIS HIS HIS HIS SER SER GLY LEU VAL PRO ARG GLY SER HIS MET GLY VAL LEU ARG L6 T12 PHE ARG ASP PRO ASP N18 I21 R24 G25 I26 N27 V28 A29 G30 E31 A32 K33 Y34 P35 D39 I40 P41 V44 S55 F56 R59

P60 F61 S62 L63 D64 D65 A66 H67 T68 H69 F70 S71 R74 K75 W76 I81 R82 Y83 E88 A89 H92 E93 G94 Y98 T106 V109 L110 A113 K114 Q115 F118 Y119 V120 F121 L122 H125 V128 W129 S130 R131 L132 S136 A140 W141 T142 L143 Y144

G147 F148 F153 T156 E157 A158 A159 L160 V161 Q162 N163 D167 E170 F171 P172 I175 N179 R182 L183 V184 S185 Q186 T190 L191 F192 Y193 G194 G195 R196 K201 A202 I203 I204 Q210 L213 Q214 C221 F224 D233 ILE GLU ASN GLU V238 V239

I240 G241 W242 E243 S244 R250 G251 L252 I253 G254 I258 E264 Q268 L269 G270 T271 F276 Q277 L280 G284 ARG ALA CYS GLU GLU SER THR TRP ALA PHE GLY PHE GLY PRO TYR Q301 A314 W315 L316 Y320 D321 D322 Y325 G326 W327 V328 R329 D330 P331

G332 H343 W346 ASP PRO SER SER ASP LYS LEU LEU ARG K356 D357 Y358 K361 T365 G366 E367 P368 L369 N370 Y371 D372 K373 F374 T375 F379 H382 Y383 L390 R391 W394 A395 E396 A397 F400 C401 Q402 P403 P404 V405 M406 P410 D411 L412 K413 D418 D419

A425 Y428 W439 N444 V445 G449 K454 T457 P458 A459 N471 C472 L473 R474 D475 Q476 G489 T490 H491 T496 E497 I498 G499 Y510 A521 T538 L539 W540 L541 Y542 T543 T544 S545 N546 D552 D558 L559 S560 I561 Y562 L567 E568 L569 P570

SER GLY LYS LEU LEU ALA ALA PHE GLU ASN GLU SER GLN ARG ASP PRO GLN PRO ALA TYR SER GLU SER GLN ARG ASN THR GLU SER TYR ARG VAL GLY PRO ARG ASP LEU LYS ALA LEU GLN ALA LEU ALA PRO SER ILE SER SER GLU ILE SER GLN SER SER GLN ASP LYS LEU G629 F630

R631 A632 A633 W636 S640 P641 I642 N645 I648 L649 L664 L665 G666 GLU LYS LYS GLY LEU GLY GLN GLU ALA ALA T677 E678 I679 P687 D688 T689 H690 T691 V695 W700 H708 K711 F712 Q713 R714 L715 R716 W717 W718 H719 P720 GLU GLY D723 H724 I728

K732 ARG LYS PRO GLY ASP PRO THR VAL SER GLY GLU GLU LEU SER TYR LEU GLU GLU GLN CYS GLN GLY GLY CYS SER VAL MET

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	107.73Å 112.79Å 139.10Å 90.00° 102.30° 90.00°	Depositor
Resolution (Å)	58.21 – 3.80 76.95 – 3.50	Depositor EDS
% Data completeness (in resolution range)	98.7 (58.21-3.80) 99.2 (76.95-3.50)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.42 (at 3.49Å)	Xtriage
Refinement program	PHENIX 1.17.1_3660	Depositor
R, R_{free}	0.279 , 0.324 0.285 , 0.306	Depositor DCC
R_{free} test set	2000 reflections (4.88%)	wwPDB-VP
Wilson B-factor (Å^2)	134.4	Xtriage
Anisotropy	0.554	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 136.1	EDS
L-test for twinning[2]	$< |L| > = 0.45$, $< L^2 > = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	21118	wwPDB-VP
Average B, all atoms (Å^2)	177.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.86% 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	0.25	0/5596	0.45	0/7613
1	B	0.24	0/5501	0.44	1/7481 (0.0%)
1	C	0.26	0/5472	0.45	1/7443 (0.0%)
1	D	0.28	0/5201	0.45	0/7076
All	All	0.26	0/21770	0.45	2/29613 (0.0%)

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	3
1	B	0	1
1	C	0	1
1	D	0	1
All	All	0	6

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	541	LEU	CA-CB-CG	5.20	127.25	115.30
1	C	541	LEU	CA-CB-CG	5.07	126.96	115.30

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	293	ALA	Peptide
1	A	301	GLN	Peptide
1	A	540	TRP	Peptide

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
1	B	59	ARG	Peptide
1	C	59	ARG	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	5426	0	5152	123	0
1	B	5338	0	5080	114	0
1	C	5309	0	5041	126	0
1	D	5045	0	4778	120	0
All	All	21118	0	20051	483	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 12.

The worst 5 of 483 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:D:192:PHE:CZ	1:D:214:GLN:HG2	1.97	0.99
1:B:30:GLY:H	1:B:541:LEU:HD21	1.38	0.88
1:A:291:THR:H	1:A:302:THR:HG21	1.41	0.86
1:B:329:ARG:HD3	1:B:333:TRP:HD1	1.45	0.80
1:A:645:ASN:H	1:A:677:THR:HG22	1.49	0.78

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	664/779 (85%)	597 (90%)	47 (7%)	20 (3%)	4	33
1	B	651/779 (84%)	593 (91%)	41 (6%)	17 (3%)	5	36
1	C	643/779 (82%)	588 (91%)	44 (7%)	11 (2%)	9	43
1	D	607/779 (78%)	550 (91%)	41 (7%)	16 (3%)	5	36
All	All	2565/3116 (82%)	2328 (91%)	173 (7%)	64 (2%)	5	36

5 of 64 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	52	ASP
1	A	149	ASP
1	A	294	PHE
1	A	300	TYR
1	A	323	THR

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	569/661 (86%)	565 (99%)	4 (1%)	84	91
1	B	562/661 (85%)	562 (100%)	0	100	100
1	C	559/661 (85%)	553 (99%)	6 (1%)	73	85
1	D	530/661 (80%)	529 (100%)	1 (0%)	93	97
All	All	2220/2644 (84%)	2209 (100%)	11 (0%)	88	94

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	172	PRO
1	C	292	TRP
1	D	56	PHE
1	C	371	TYR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	61	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	92	HIS
1	B	362	ASN
1	C	427	HIS
1	C	555	ASN
1	D	402	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

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, 95th 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(Å²)	Q<0.9
1	A	670/779 (86%)	0.41	44 (6%)	18	14	89, 179, 220, 239	0
1	B	659/779 (84%)	0.44	49 (7%)	14	11	86, 179, 223, 238	0
1	C	655/779 (84%)	0.63	90 (13%)	3	3	30, 180, 218, 249	0
1	D	623/779 (79%)	0.51	72 (11%)	4	5	93, 190, 224, 243	0
All	All	2607/3116 (83%)	0.50	255 (9%)	7	7	30, 182, 223, 249	0

The worst 5 of 255 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	629	GLY	7.8
1	B	159	ALA	7.7
1	D	570	PRO	6.9
1	C	350	SER	6.4
1	C	260	VAL	6.3

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 ⓘ

There are no such residues in this entry.