wwPDB X-ray Structure Validation Summary Report (i)

Jun 24, 2024 – 02:29 PM EDT

PDB ID	:	5Y7O
Title	:	Crystal structure of folding sensor region of UGGT from Thermomyces dupontii
Authors	:	Satoh, T.; Song, C.; Zhu, T.; Toshimori, T.; Murata, K.; Hayashi, Y.; Kamikubo, H.; Uchihashi, T.; Kato, K.
Deposited on	:	2017-08-17
Resolution	:	3.10 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:

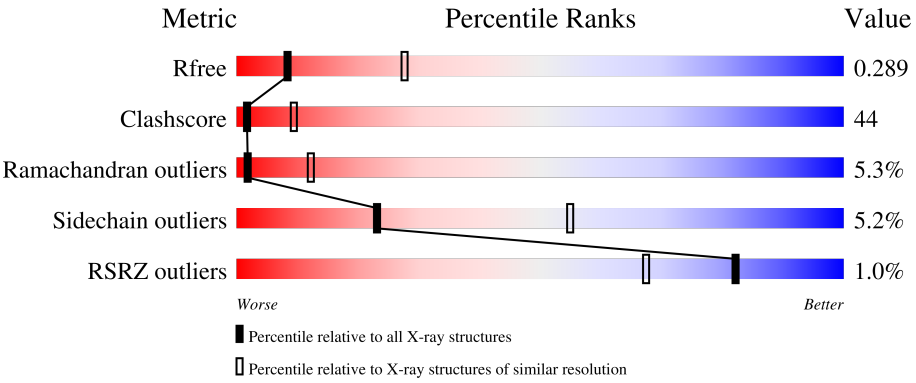
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 3.10 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	1094 (3.10-3.10)
Clashscore	141614	1184 (3.10-3.10)
Ramachandran outliers	138981	1141 (3.10-3.10)
Sidechain outliers	138945	1141 (3.10-3.10)
RSRZ outliers	127900	1067 (3.10-3.10)

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	1130	% 29% 44% 6% • 19%
1	B	1130	% 29% 45% 6% • 19%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 14318 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 UGGT.

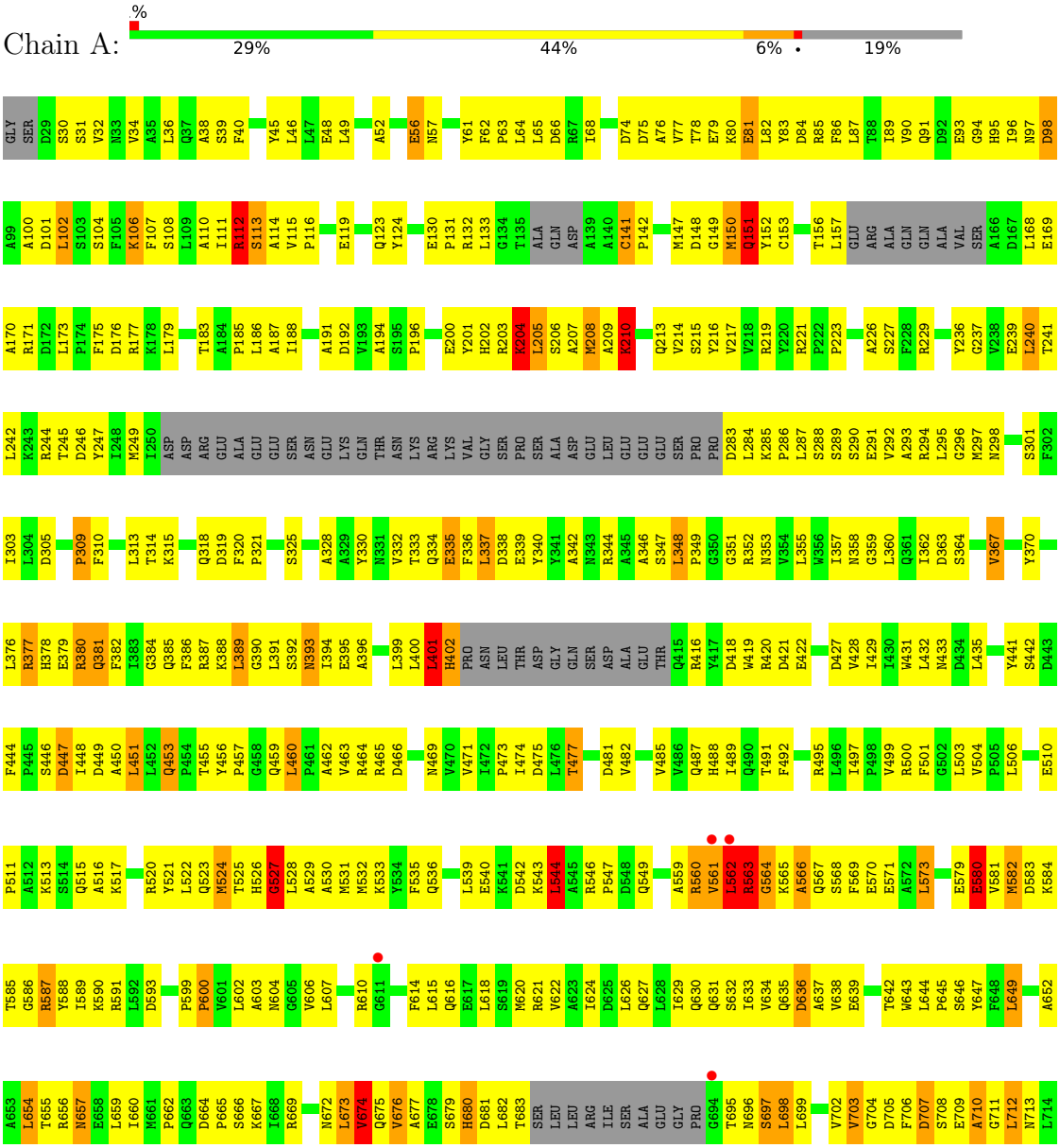
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	913	Total	C	N	O	S	Se	0	0	0
			7159	4556	1224	1362	2	15			
1	B	913	Total	C	N	O	S	Se	0	0	0
			7159	4556	1224	1362	2	15			

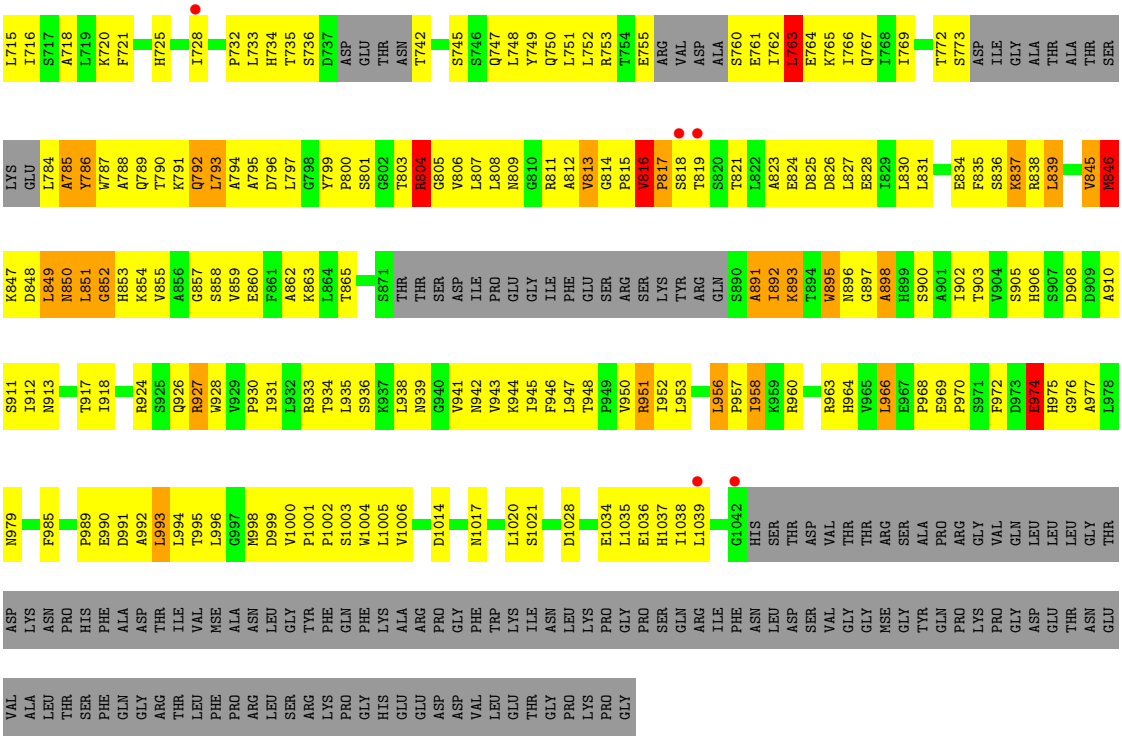
3 Residue-property plots (i)

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: UGGT

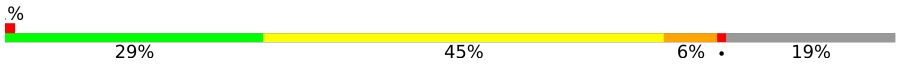
Chain A:

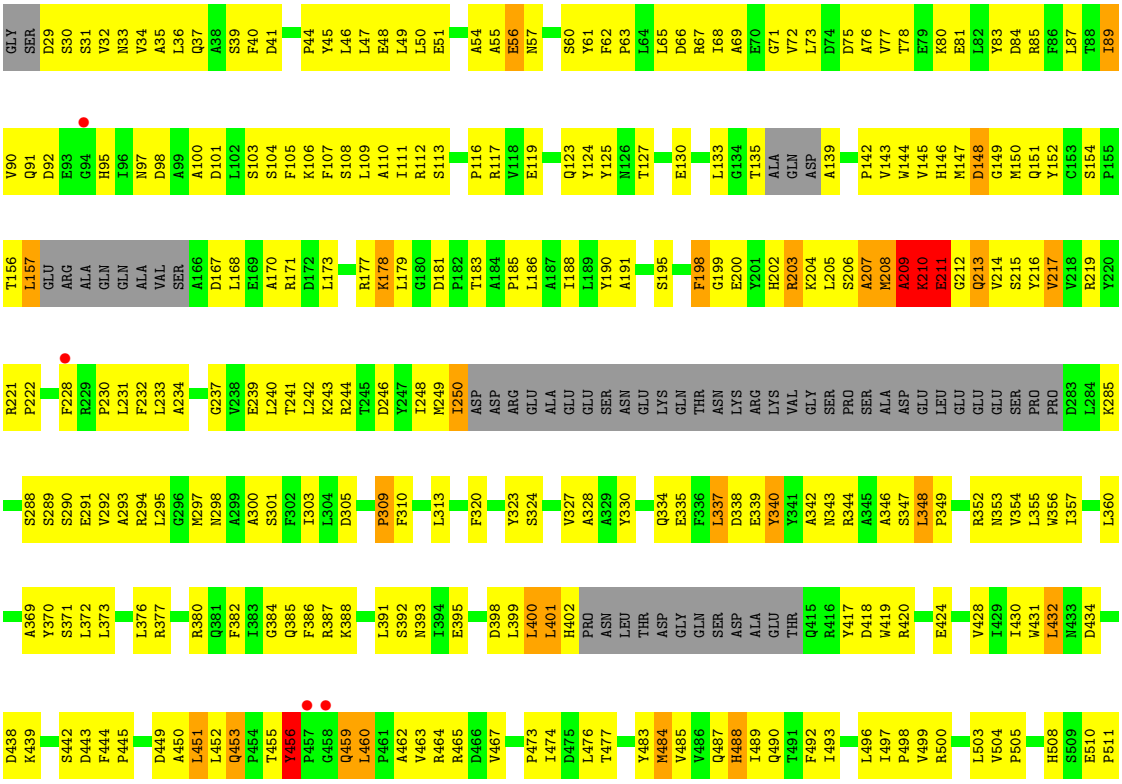


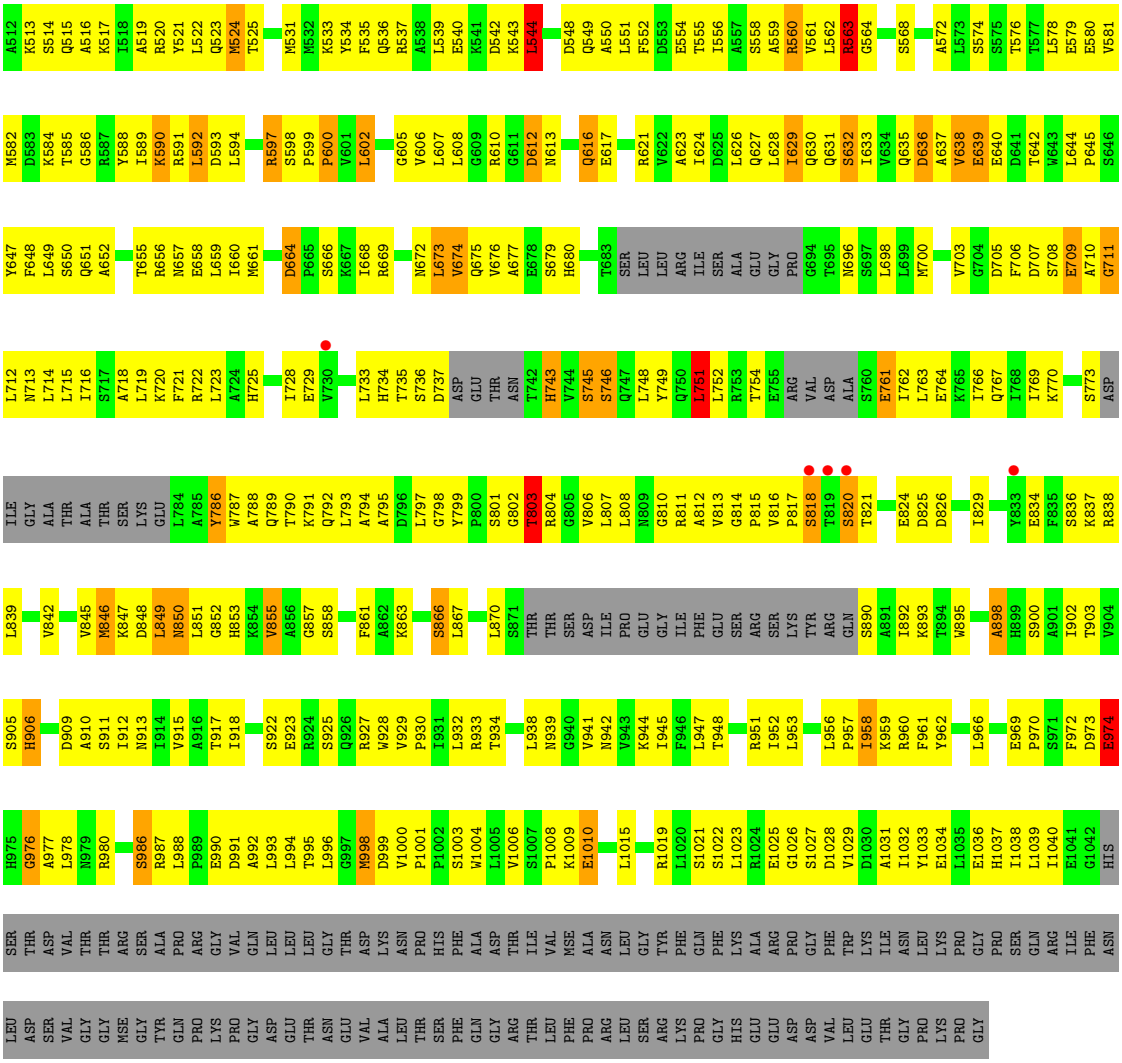


• Molecule 1: UGGT

Chain B:






A512 K513 S514 Q515 A516 K517 I518 A519 R520 Y521 L522 Q523 M524 T525 M531 M532 K533 Y534 F535 Q536 R537 A538 L539 E540 K541 D542 K543 L544 D548 Q549 A550 L551 F552 D553 E554 T555 I556 A557 S558 A559 R560 V561 L562 R563 G564 S568 A572 L573 S574 S575 T576 T577 L578 E579 E580 V581
M582 D583 K584 T585 G586 R587 Y588 I589 K590 R591 L592 D593 L594 R597 S598 P599 P600 V601 L602 G605 V606 L607 L608 G609 R610 G611 D612 N613 Q616 E617 R621 V622 A623 I624 D625 L626 Q627 L628 I629 Q630 Q631 S632 I633 V634 Q635 D636 A637 V638 E639 E640 D641 T642 W643 L644 P645 S646
Y647 F648 L649 S650 Q651 A652 T655 R656 N657 E658 L659 I660 M661 D664 P665 S666 K667 I668 R669 N672 L673 V674 Q675 V676 A677 E678 S679 H680 T683 SER LEU LEU ARG ILE SER ALA GLU GLY PRO G694 T695 N696 S697 L698 L699 M700 V703 G704 D705 F706 D707 S708 E709 A710 G711
L712 N713 L714 L715 I716 S717 A718 L719 K720 F721 R722 L723 A724 H725 I728 E729 V730 L733 H734 T735 S736 D737 ASP GLU THR ASN T742 H743 V744 S745 S746 Q747 L748 Y749 Q750 L751 L752 R753 T754 E755 ARG VAL ASP ALA S760 E761 I762 L763 E764 K765 I766 Q767 I768 I769 K770 S773 ASP
ILE GLY ALA THR ALA THR SER LYS GLU L784 A785 Y786 W787 A788 Q789 T790 K791 Q792 L793 A794 A795 D796 L797 G798 Y799 P800 S801 G802 T803 R804 G805 V806 L807 L808 N809 G810 R811 A812 V813 G814 P815 V816 P817 S818 T819 S820 T821 E824 D825 D826 I829 Y833 E834 F835 S836 K837 R838
L839 V842 V845 M846 K847 D848 L849 N850 L851 G852 H853 K854 V855 A856 G857 S858 F861 A862 K863 S866 L867 L870 S871 THR THR SER ASP ILE PRO GLU GLY ILE PHE GLU SER ARG SER LYS TYR ARG GLN S890 A891 I892 K893 T894 W895 A898 H899 S900 A901 I902 T903 V904
S905 H906 D909 A910 S911 I912 N913 Y914 V915 A916 T917 I918 S922 E923 R924 S925 Q926 R927 W928 V929 P930 I931 L932 R933 T934 L938 N939 G940 V941 N942 V943 K944 I945 F946 L947 T948 R951 I952 L953 L956 P957 I958 K959 R960 F961 Y962 L966 E969 P970 S971 F972 D973 E974
H975 G976 A977 L978 N979 R980 S986 R987 L988 P989 E990 D991 A992 L993 L994 T995 L996 G997 M998 D999 V1000 P1001 P1002 S1003 W1004 L1005 V1006 S1007 P1008 K1009 E1010 L1015 R1019 L1020 S1021 S1022 L1023 R1024 E1025 G1026 S1027 D1028 V1029 D1030 A1031 I1032 Y1033 E1034 L1035 E1036 H1037 I1038 L1039 I1040 E1041 G1042 HIS
SER THR ASP VAL THR THR ARG SER ALA PRO ARG GLY VAL GLN LEU LEU LEU GLY THR ASP LYS ASN PRO HIS PHE ALA ASP THR ILE VAL MSE ALA ASN LEU GLY TYR PHE GLN LYS PRO GLY PHE LYS ALA ARG PRO GLY PHE TRP LYS ILE ASN LEU LYS PRO GLY PRO SER GLN ARG ILE PHE ASN
LEU ASP SER VAL GLY GLY MSE GLY TYR GLN PRO LYS PRO GLY ASP GLU THR ASN GLU VAL ALA LEU THR SER PHE GLN GLY ARG THR LEU PHE PRO ARG LEU SER ARG LYS PRO GLY HIS GLU GLU ASP ASP VAL LEU GLU THR GLY PRO LYS PRO GLY

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32 1 2	Depositor
Cell constants a, b, c, α, β, γ	195.09Å 195.09Å 142.26Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	19.97 – 3.10 19.97 – 3.10	Depositor EDS
% Data completeness (in resolution range)	98.0 (19.97-3.10) 98.0 (19.97-3.10)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.37 (at 3.09Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.232 , 0.278 0.233 , 0.289	Depositor DCC
R_{free} test set	2732 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	110.4	Xtriage
Anisotropy	0.256	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 108.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.55$, $< L^2 > = 0.40$	Xtriage
Estimated twinning fraction	0.478 for -h,-k,l	Xtriage
Reported twinning fraction	0.500 for -h,-k,l	Depositor
Outliers	0 of 54601 reflections	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	14318	wwPDB-VP
Average B, all atoms (Å^2)	123.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.67% 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	0.66	3/7286 (0.0%)	1.09	30/9871 (0.3%)
1	B	0.62	3/7286 (0.0%)	1.02	29/9871 (0.3%)
All	All	0.64	6/14572 (0.0%)	1.05	59/19742 (0.3%)

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	14
1	B	0	9
All	All	0	23

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	211	GLU	CB-CG	-8.01	1.36	1.52
1	B	209	ALA	CA-CB	-7.33	1.37	1.52
1	A	895	TRP	CB-CG	-7.16	1.37	1.50
1	B	211	GLU	CG-CD	-6.12	1.42	1.51
1	A	563	ARG	CG-CD	-5.51	1.38	1.51

The worst 5 of 59 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	401	LEU	CB-CG-CD2	-18.33	79.83	111.00
1	A	763	LEU	CB-CG-CD1	-13.06	88.80	111.00
1	A	205	LEU	CA-CB-CG	-8.87	94.90	115.30
1	A	573	LEU	CB-CG-CD2	-8.71	96.20	111.00
1	A	564	GLY	N-CA-C	8.65	134.73	113.10

There are no chirality outliers.

5 of 23 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	150	MSE	Peptide
1	A	151	GLN	Peptide
1	A	204	LYS	Peptide
1	A	391	LEU	Peptide
1	A	456	TYR	Peptide

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	7159	0	7156	685	0
1	B	7159	0	7156	576	1
All	All	14318	0	14312	1254	1

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 44.

The worst 5 of 1254 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:703:VAL:HG22	1:A:804:ARG:HH11	1.15	1.10
1:A:380:ARG:HH11	1:A:912:ILE:HD12	1.02	1.07
1:B:210:LYS:HB2	1:B:211:GLU:HG2	1.11	1.06
1:A:433:ASN:HD22	1:A:500:ARG:HA	1.21	1.04
1:B:420:ARG:HE	1:B:649:LEU:HD21	1.22	1.00

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:456:TYR:OH	1:B:456:TYR:OH[5_554]	2.09	0.11

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	893/1130 (79%)	685 (77%)	155 (17%)	53 (6%)	1 10
1	B	893/1130 (79%)	696 (78%)	155 (17%)	42 (5%)	2 14
All	All	1786/2260 (79%)	1381 (77%)	310 (17%)	95 (5%)	2 12

5 of 95 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	56	GLU
1	A	61	TYR
1	A	148	ASP
1	A	348	LEU
1	A	524	MSE

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	782/948 (82%)	737 (94%)	45 (6%)	20 51
1	B	782/948 (82%)	746 (95%)	36 (5%)	27 59
All	All	1564/1896 (82%)	1483 (95%)	81 (5%)	23 55

5 of 81 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	576	THR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	804	ARG
1	B	590	LYS
1	B	709	GLU
1	B	855	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	469	ASN
1	B	488	HIS
1	B	1037	HIS
1	B	913	ASN
1	A	675	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, 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	898/1130 (79%)	-0.34	9 (1%)	82	67	70, 117, 187, 345	0
1	B	898/1130 (79%)	-0.34	9 (1%)	82	67	76, 115, 181, 279	0
All	All	1796/2260 (79%)	-0.34	18 (1%)	82	67	70, 116, 185, 345	0

The worst 5 of 18 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	819	THR	6.7
1	B	820	SER	5.7
1	B	94	GLY	5.1
1	A	728	ILE	4.5
1	A	1039	LEU	4.3

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.