# wwPDB X-ray Structure Validation Summary Report (i)

May 25, 2020 – 06:56 am BST

| | | |
|---|---|---|
| PDB ID | : | 5WTJ |
| Title | : | Crystal structure of an endonuclease |
| Authors | : | Liu, L.; Wang, Y. |
| Deposited on | : | 2016-12-13 |
| Resolution | : | 3.50 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 following versions of software and data (see references (1)) were used in the production of this report:

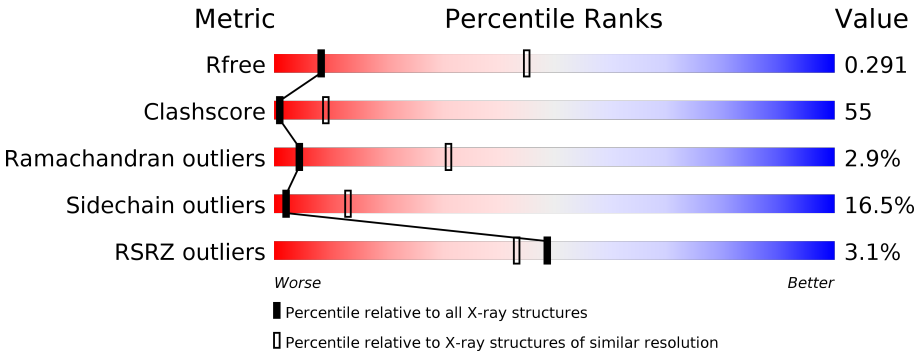
| | | |
|---|---|---|
| MolProbity | : | 4.02b-467 |
| Mogul | : | 1.8.5 (274361), CSD as541be (2020) |
| Xtriage (Phenix) | : | 1.13 |
| EDS | : | 2.11 |
| Percentile statistics | : | 20191225.v01 (using entries in the PDB archive December 25th 2019) |
| Refmac | : | 5.8.0158 |
| CCP4 | : | 7.0.044 (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 (i)

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

The reported resolution of this entry is 3.50 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 | 1659 (3.60-3.40) |
| Clashscore | 141614 | 1036 (3.58-3.42) |
| Ramachandran outliers | 138981 | 1005 (3.58-3.42) |
| Sidechain outliers | 138945 | 1006 (3.58-3.42) |
| RSRZ outliers | 127900 | 1559 (3.60-3.40) |

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 >=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 | 1397 | % 30% 33% 9% • 27% |
| 1 | B | 1397 | 3% 32% 31% 9% • 28% |

## 2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 15952 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 CRISPR-associated endoribonuclease C2c2.

| Mol | Chain | Residues | Atoms | | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | A | 1019 | Total | C | N | O | S | Se | 0 | 0 | 0 |
| | | | 8013 | 5109 | 1350 | 1538 | 5 | 11 | | | |
| 1 | B | 1012 | Total | C | N | O | S | Se | 0 | 0 | 0 |
| | | | 7938 | 5055 | 1326 | 1541 | 6 | 10 | | | |

There are 16 discrepancies between the modelled and reference sequences:

| Chain | Residue | Modelled | Actual | Comment | Reference |
|---|---|---|---|---|---|
| A | 1390 | LEU | - | expression tag | UNP P0DOC6 |
| A | 1391 | GLU | - | expression tag | UNP P0DOC6 |
| A | 1392 | HIS | - | expression tag | UNP P0DOC6 |
| A | 1393 | HIS | - | expression tag | UNP P0DOC6 |
| A | 1394 | HIS | - | expression tag | UNP P0DOC6 |
| A | 1395 | HIS | - | expression tag | UNP P0DOC6 |
| A | 1396 | HIS | - | expression tag | UNP P0DOC6 |
| A | 1397 | HIS | - | expression tag | UNP P0DOC6 |
| B | 1390 | LEU | - | expression tag | UNP P0DOC6 |
| B | 1391 | GLU | - | expression tag | UNP P0DOC6 |
| B | 1392 | HIS | - | expression tag | UNP P0DOC6 |
| B | 1393 | HIS | - | expression tag | UNP P0DOC6 |
| B | 1394 | HIS | - | expression tag | UNP P0DOC6 |
| B | 1395 | HIS | - | expression tag | UNP P0DOC6 |
| B | 1396 | HIS | - | expression tag | UNP P0DOC6 |
| B | 1397 | HIS | - | expression tag | UNP P0DOC6 |

- Molecule 2 is water.

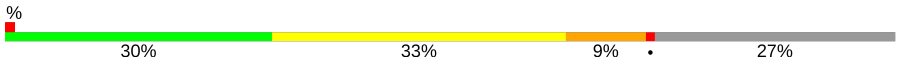
| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|
| 2 | A | 1 | Total | O | 0 | 0 |
| | | | 1 | 1 | | |

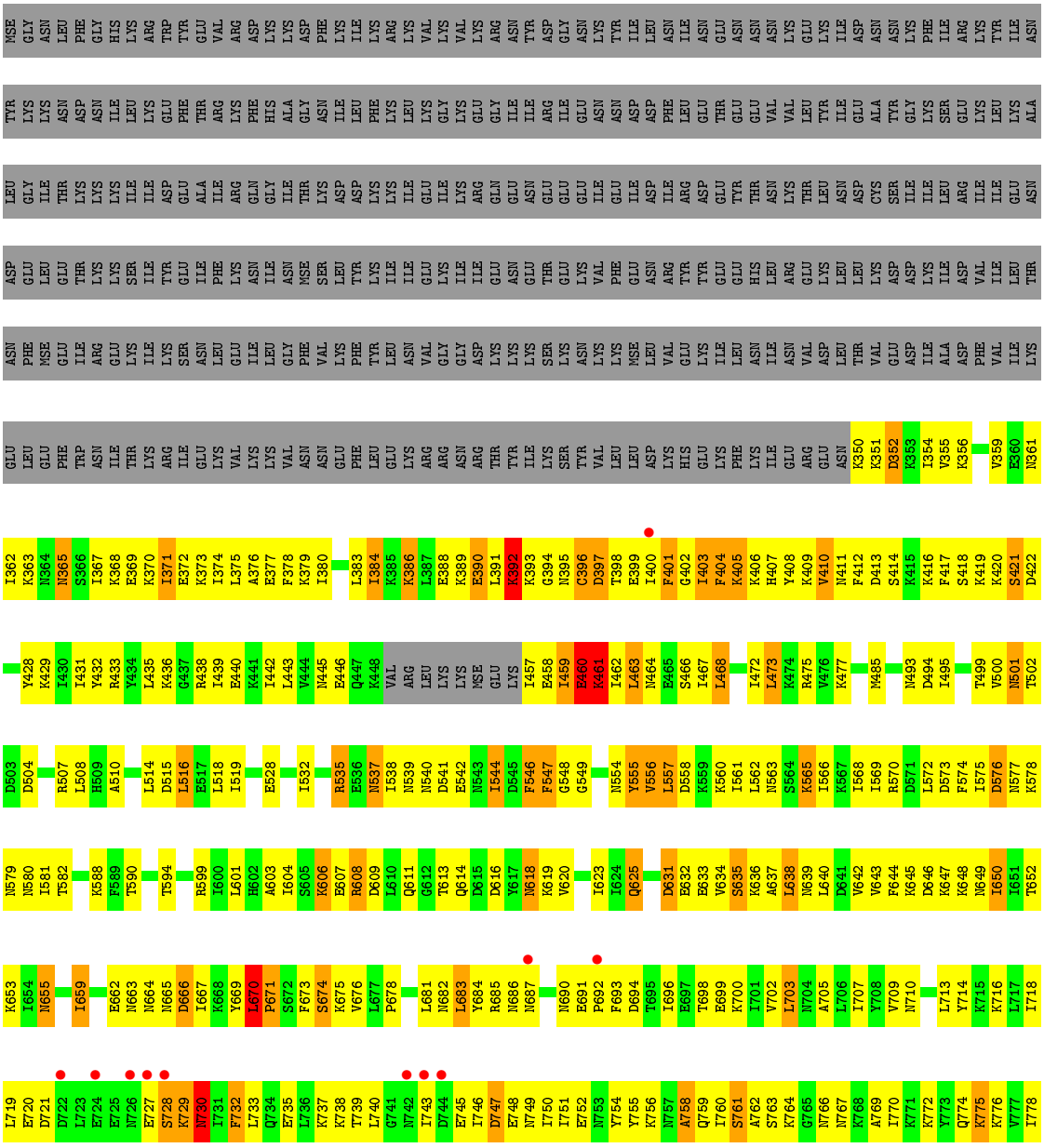
# 3 Residue-property plots (i)

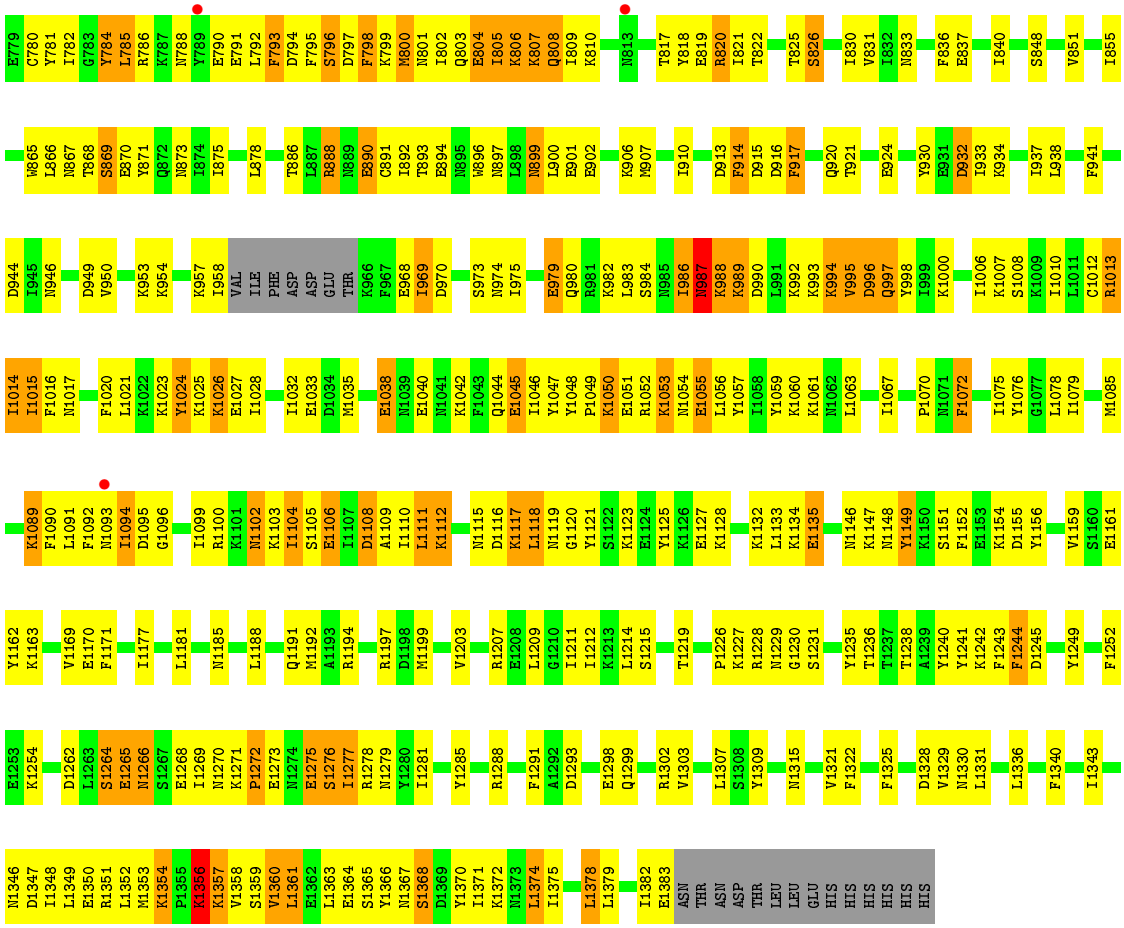
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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: CRISPR-associated endoribonuclease C2c2

Chain A:

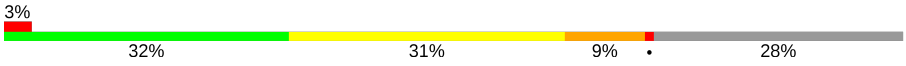


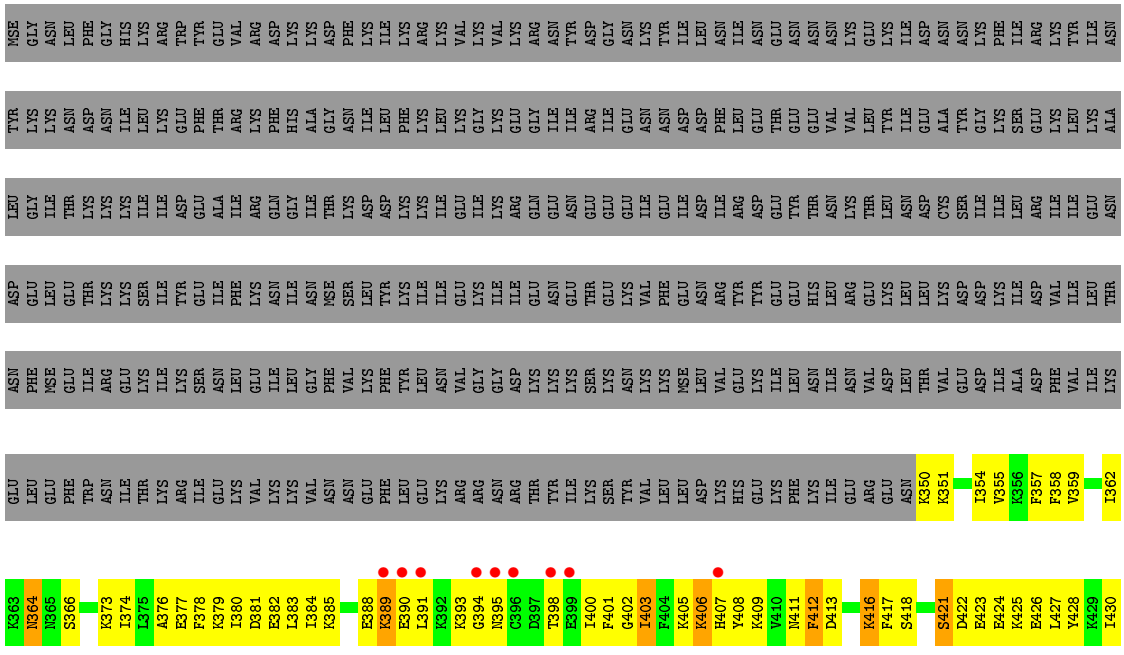



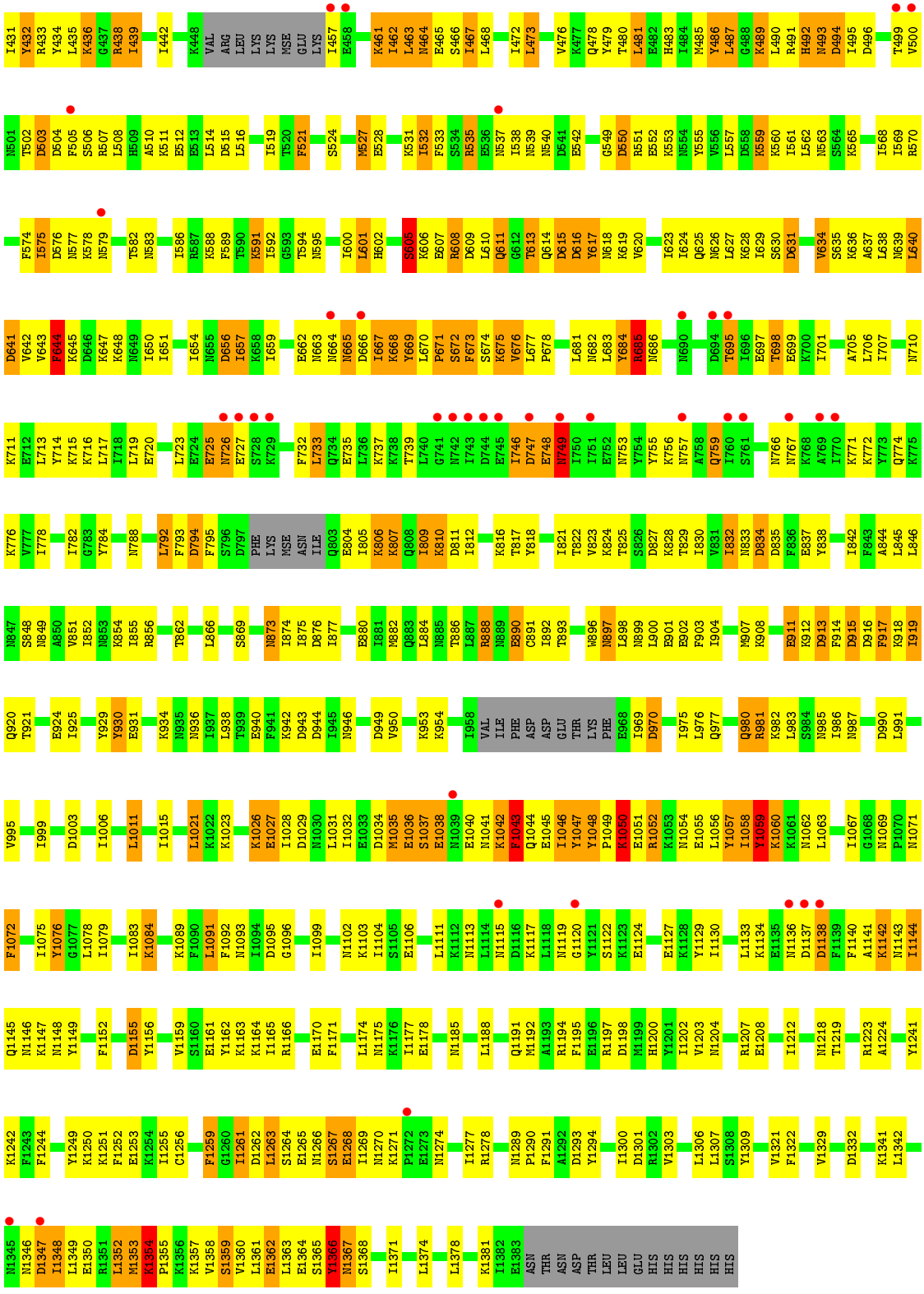


• Molecule 1: CRISPR-associated endoribonuclease C2c2

Chain B:




I431 Y432 R433 Y434 L435 K436 G437 R438 I439 I442 K448 VAL ARG LEU LYS LYS MSE GLU LYS I457 E458 K461 I462 L463 N464 E465 S466 I467 L468 I472 L473 V476 K477 Q478 Y479 T480 L481 E482 H483 I484 M485 Y486 L487 G488 K489 L490 R491 H492 N493 D494 I495 D496 T499 V500
N501 T502 D503 D504 F505 S506 R507 L508 H509 A510 K511 E512 E513 L514 D515 L516 I519 T520 F521 S524 M527 E528 K531 I532 F533 S534 R535 E536 N537 I538 N539 N540 D541 E542 G549 D550 R551 E552 K553 N554 Y555 V556 L557 D558 K559 K560 I561 L562 N563 S564 K565 I568 I569 R570
F574 I575 D576 N577 K578 N579 T582 N583 I586 R587 K588 F589 T590 K591 I592 G593 T594 N595 I600 L601 H602 S605 K606 E607 R608 D609 L610 Q611 G612 T613 Q614 D615 D616 Y617 N618 K619 V620 I623 I624 Q625 N626 L627 K628 I629 S630 D631 V634 S635 K636 A637 L638 N639 L640
D641 V642 V643 F644 K645 D646 K647 K648 N649 I650 I651 I654 N655 D656 I657 K658 I659 E662 N663 N664 N665 D666 I667 K668 Y669 L670 P671 S672 F673 S674 K675 V676 L677 P678 L681 N682 L683 Y684 R685 N686 N690 D694 T695 I696 E697 T698 E699 K700 I701 A705 L706 I707 N710
K711 E712 L713 Y714 K715 K716 L717 I718 L719 E720 L723 E724 E725 N726 E727 S728 K729 F732 L733 Q734 E735 L736 K737 K738 T739 L740 G741 N742 I743 D744 E745 I746 D747 E748 N749 I750 I751 E752 N753 Y754 Y755 K756 N757 A758 Q759 I760 S761 N766 N767 K768 A769 I770 K771 K772 Y773 Q774 K775
K776 W777 I778 I782 G783 Y784 N788 L792 F793 D794 F795 S796 D797 PHE LYS MSE ASN ILE Q803 E804 I805 K806 K807 Q808 I809 K810 D811 I812 K816 T817 Y818 I821 T822 V823 K824 R825 S826 D827 K828 T829 I830 V831 I832 N833 D834 D835 F836 E837 Y838 I842 F843 A844 L845 L846
N847 S848 N849 A850 V851 I852 N853 K854 I855 R856 T862 L866 S869 N873 I874 I875 D876 I877 E880 I881 M882 Q883 L884 N885 T886 L887 R888 N889 E890 C891 I892 T893 W896 N897 L898 N899 L900 E901 E902 F903 I904 N907 R908 E911 K912 D913 F914 D915 D916 F917 K918 I919
Q920 T921 E924 I925 Y929 Y930 E931 K934 N935 N936 I937 L938 Y939 E940 F941 K942 D943 D944 I945 N946 D949 V950 K953 K954 I958 VAL ILE PHE ASP ASP GLU THR LYS PHE E968 I969 D970 I975 L976 Q977 Q980 R981 K982 L983 S984 N985 I986 N987 D990 L991
V995 I999 D1003 I1006 L1011 I1015 L1021 K1022 K1023 K1026 E1027 I1028 D1029 N1030 L1031 I1032 E1033 D1034 M1035 E1036 S1037 E1038 N1039 E1040 N1041 K1042 F1043 Q1044 E1045 I1046 Y1047 Y1048 P1049 K1050 E1051 R1052 K1053 N1054 E1055 L1056 Y1057 I1058 Y1059 K1060 K1061 N1062 L1063 I1067 G1068 N1069 P1070 N1071
F1072 I1075 Y1076 G1077 L1078 I1079 I1083 K1084 K1089 F1090 L1091 F1092 N1093 I1094 D1095 G1096 I1099 N1102 K1103 I1104 S1105 E1106 L1111 K1112 N1113 L1114 N1115 D1116 K1117 L1118 N1119 G1120 Y1121 S1122 K1123 E1124 E1127 K1128 Y1129 I1130 L1133 K1134 E1135 N1136 D1137 D1138 F1139 F1140 A1141 K1142 N1143 I1144
Q1145 N1146 K1147 N1148 Y1149 F1152 D1155 Y1156 V1159 S1160 E1161 Y1162 K1163 K1164 I1165 R1166 E1170 F1171 L1174 N1175 K1176 I1177 E1178 N1185 L1188 Q1191 M1192 A1193 R1194 F1195 E1196 R1197 D1198 M1199 H1200 Y1201 I1202 V1203 N1204 R1207 E1208 I1212 N1218 T1219 R1223 A1224 Y1241
K1242 F1243 F1244 Y1249 K1250 K1251 F1252 E1253 K1254 I1255 C1256 F1259 G1260 I1261 D1262 L1263 S1264 E1265 N1266 S1267 E1268 I1269 N1270 K1271 P1272 E1273 N1274 I1277 R1278 N1289 P1290 F1291 A1292 D1293 Y1294 I1300 D1301 R1302 Y1303 L1306 L1307 S1308 Y1309 V1321 F1322 V1329 D1332 K1341 L1342
N1345 N1346 D1347 I1348 L1349 E1350 R1351 L1352 M1353 K1354 P1355 K1356 K1357 V1358 S1359 V1360 L1361 E1362 L1363 E1364 S1365 Y1366 N1367 S1368 I1371 L1374 L1378 K1381 I1382 E1383 ASN THR ASN ASP THR LEU LEU GLU HIS HIS HIS HIS HIS

# 4 Data and refinement statistics (i)

| Property | Value | Source |
|---|---|---|
| Space group | P 21 21 21 | Depositor |
| Cell constants<br>a, b, c, $\alpha$, $\beta$, $\gamma$ | 91.66Å 94.23Å 338.22Å<br>90.00° 90.00° 90.00° | Depositor |
| Resolution (Å) | 48.38 – 3.50<br>48.37 – 3.50 | Depositor<br>EDS |
| % Data completeness<br>(in resolution range) | 81.0 (48.38-3.50)<br>75.3 (48.37-3.50) | Depositor<br>EDS |
| $R_{merge}$ | 0.09 | Depositor |
| $R_{sym}$ | (Not available) | Depositor |
| $< I/\sigma(I) >$ [1] | 4.71 (at 3.48Å) | Xtriage |
| Refinement program | PHENIX (1.10.1_2155: ???) | Depositor |
| R, $R_{free}$ | 0.263 , 0.291<br>0.263 , 0.291 | Depositor<br>DCC |
| $R_{free}$ test set | 1546 reflections (5.05%) | wwPDB-VP |
| Wilson B-factor (Å$^2$) | 18.6 | Xtriage |
| Anisotropy | 0.648 | Xtriage |
| Bulk solvent $k_{sol}$(e/Å$^3$), $B_{sol}$(Å$^2$) | 0.25 , -0.8 | EDS |
| L-test for twinning[2] | $< \|L\| > = 0.44$, $< L^2 > = 0.26$ | Xtriage |
| Estimated twinning fraction | 0.053 for k,h,-l | Xtriage |
| $F_o$,$F_c$ correlation | 0.78 | EDS |
| Total number of atoms | 15952 | wwPDB-VP |
| Average B, all atoms (Å$^2$) | 57.0 | wwPDB-VP |

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.41% 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 | 0.48 | 2/8117 (0.0%) | 0.67 | 16/10925 (0.1%) |
| 1 | B | 0.42 | 2/8038 (0.0%) | 0.66 | 11/10824 (0.1%) |
| All | All | 0.45 | 4/16155 (0.0%) | 0.66 | 27/21749 (0.1%) |

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 | 7 |
| All | All | 0 | 9 |

All (4) bond length outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed(Å) | Ideal(Å) |
|---|---|---|---|---|---|---|---|
| 1 | A | 671 | PRO | N-CD | 5.24 | 1.55 | 1.47 |
| 1 | B | 1038 | GLU | CB-CG | -5.17 | 1.42 | 1.52 |
| 1 | A | 1272 | PRO | N-CD | 5.09 | 1.54 | 1.47 |
| 1 | B | 671 | PRO | N-CD | 5.09 | 1.54 | 1.47 |

The worst 5 of 27 bond angle outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed($^o$) | Ideal($^o$) |
|---|---|---|---|---|---|---|---|
| 1 | B | 634 | VAL | CB-CA-C | -11.15 | 90.21 | 111.40 |
| 1 | A | 997 | GLN | CB-CA-C | 6.94 | 124.28 | 110.40 |
| 1 | B | 1021 | LEU | CB-CG-CD1 | -6.59 | 99.80 | 111.00 |
| 1 | B | 403 | ILE | CB-CA-C | -6.57 | 98.46 | 111.60 |
| 1 | A | 758 | ALA | CB-CA-C | -6.33 | 100.60 | 110.10 |

There are no chirality outliers.

5 of 9 planarity outliers are listed below:

| Mol | Chain | Res | Type | Group |
|---|---|---|---|---|
| 1 | A | 1051 | GLU | Peptide |
| 1 | A | 1356 | LYS | Peptide |
| 1 | B | 412 | PHE | Peptide |
| 1 | B | 605 | SER | Peptide |
| 1 | B | 644 | PHE | 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 | 8013 | 0 | 7662 | 834 | 3 |
| 1 | B | 7938 | 0 | 7580 | 876 | 1 |
| 2 | A | 1 | 0 | 0 | 0 | 0 |
| All | All | 15952 | 0 | 15242 | 1710 | 4 |

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 55.

The worst 5 of 1710 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:B:976:LEU:HD23 | 1:B:977:GLN:N | 1.22 | 1.45 |
| 1:A:696:ILE:CB | 1:A:699:GLU:HB3 | 1.55 | 1.35 |
| 1:B:817:THR:CG2 | 1:B:818:TYR:HA | 1.54 | 1.35 |
| 1:B:914:PHE:HA | 1:B:915:ASP:CB | 1.47 | 1.34 |
| 1:B:914:PHE:CA | 1:B:915:ASP:HB2 | 1.54 | 1.32 |

All (4) 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:426:GLU:OE1 | 1:B:1037:SER:OG[4_446] | 1.51 | 0.69 |
| 1:A:776:LYS:CB | 1:A:1348:ILE:CD1[3_655] | 1.91 | 0.29 |
| 1:A:398:THR:CB | 1:A:907:MSE:CE[3_655] | 2.08 | 0.12 |

*Continued on next page...*

*Continued from previous page...*

| Atom-1 | Atom-2 | Interatomic distance (Å) | Clash overlap (Å) |
|---|---|---|---|
| 1:A:776:LYS:CD | 1:A:1348:ILE:CD1[3_655] | 2.10 | 0.10 |

## 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 | 1013/1397 (72%) | 908 (90%) | 79 (8%) | 26 (3%) | 5 33 |
| 1 | B | 1004/1397 (72%) | 888 (88%) | 84 (8%) | 32 (3%) | 4 29 |
| All | All | 2017/2794 (72%) | 1796 (89%) | 163 (8%) | 58 (3%) | 4 31 |

5 of 58 Ramachandran outliers are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | A | 392 | LYS |
| 1 | A | 410 | VAL |
| 1 | A | 458 | GLU |
| 1 | A | 460 | GLU |
| 1 | A | 461 | LYS |

### 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 | 824/1316 (63%) | 681 (83%) | 143 (17%) | 2 11 |
| 1 | B | 822/1316 (62%) | 693 (84%) | 129 (16%) | 2 15 |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Analysed | Rotameric | Outliers | Percentiles |
|---|---|---|---|---|---|
| All | All | 1646/2632 (62%) | 1374 (84%) | 272 (16%) | 2 13 |

5 of 272 residues with a non-rotameric sidechain are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | A | 1185 | ASN |
| 1 | B | 438 | ARG |
| 1 | B | 1084 | LYS |
| 1 | A | 1244 | PHE |
| 1 | A | 1368 | SER |

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 30 such sidechains are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | B | 464 | ASN |
| 1 | B | 501 | ASN |
| 1 | B | 1175 | ASN |
| 1 | B | 492 | HIS |
| 1 | B | 539 | ASN |

#### 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 carbohydrates in this entry.

## 5.6 Ligand geometry

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 | 1008/1397 (72%) | -0.36 | 14 (1%) | 75 | 69 | 2, 42, 106, 153 | 0 |
| 1 | B | 1002/1397 (71%) | 0.08 | 48 (4%) | 30 | 27 | 12, 67, 134, 199 | 0 |
| All | All | 2010/2794 (71%) | -0.14 | 62 (3%) | 49 | 43 | 2, 55, 128, 199 | 0 |

The worst 5 of 62 RSRZ outliers are listed below:

| Mol | Chain | Res | Type | RSRZ |
|---|---|---|---|---|
| 1 | B | 726 | ASN | 8.0 |
| 1 | B | 744 | ASP | 6.9 |
| 1 | B | 728 | SER | 6.7 |
| 1 | B | 743 | ILE | 6.7 |
| 1 | B | 727 | GLU | 6.0 |

## 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 carbohydrates 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.