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

Feb 1, 2016 – 07:37 AM GMT

| | | |
|---|---|---|
| PDB ID | : | 3BGA |
| Title | : | Crystal structure of beta-galactosidase from Bacteroides thetaiotaomicron VPI-5482 |
| Authors | : | Kumaran, D.; Burley, S.K.; Swaminathan, S.; New York SGX Research Center for Structural Genomics (NYSGXRC) |
| Deposited on | : | 2007-11-26 |
| Resolution | : | 2.10 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
http://wwpdb.org/validation/2016/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:

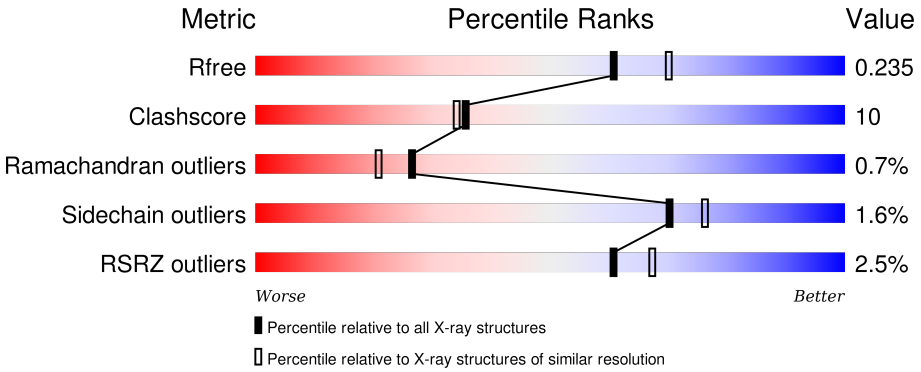
| | | |
|---|---|---|
| MolProbity | : | 4.02b-467 |
| Mogul | : | 1.7 (RC4), CSD as536be (2015) |
| Xtriage (Phenix) | : | 1.9-1692 |
| EDS | : | rb-20026688 |
| Percentile statistics | : | 20151230.v01 (using entries in the PDB archive December 30th 2015) |
| Refmac | : | 5.8.0135 |
| CCP4 | : | 6.5.0 |
| Ideal geometry (proteins) | : | Engh & Huber (2001) |
| Ideal geometry (DNA, RNA) | : | Parkinson et al. (1996) |
| Validation Pipeline (wwPDB-VP) | : | trunk26865 |

# 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 2.10 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)) |
|---|---|---|
| $R_{free}$ | 91344 | 3939 (2.10-2.10) |
| Clashscore | 102246 | 4460 (2.10-2.10) |
| Ramachandran outliers | 100387 | 4413 (2.10-2.10) |
| Sidechain outliers | 100360 | 4414 (2.10-2.10) |
| RSRZ outliers | 91569 | 3948 (2.10-2.10) |

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 on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. 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 | 1010 | 3% 79% 19% |
| 1 | B | 1010 | 2% 79% 19% |

## 2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 17243 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 Beta-galactosidase.

| Mol | Chain | Residues | Atoms | | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | A | 1003 | Total | C | N | O | S | Se | 0 | 0 | 0 |
| | | | 8064 | 5124 | 1390 | 1516 | 14 | 20 | | | |
| 1 | B | 1000 | Total | C | N | O | S | Se | 0 | 0 | 0 |
| | | | 8034 | 5109 | 1379 | 1512 | 14 | 20 | | | |

There are 22 discrepancies between the modelled and reference sequences:

| Chain | Residue | Modelled | Actual | Comment | Reference |
|---|---|---|---|---|---|
| A | 21 | MSE | - | EXPRESSION TAG | UNP Q8A799 |
| A | 22 | SER | - | EXPRESSION TAG | UNP Q8A799 |
| A | 23 | LEU | - | EXPRESSION TAG | UNP Q8A799 |
| A | 1023 | GLU | - | EXPRESSION TAG | UNP Q8A799 |
| A | 1024 | GLY | - | EXPRESSION TAG | UNP Q8A799 |
| A | 1025 | HIS | - | EXPRESSION TAG | UNP Q8A799 |
| A | 1026 | HIS | - | EXPRESSION TAG | UNP Q8A799 |
| A | 1027 | HIS | - | EXPRESSION TAG | UNP Q8A799 |
| A | 1028 | HIS | - | EXPRESSION TAG | UNP Q8A799 |
| A | 1029 | HIS | - | EXPRESSION TAG | UNP Q8A799 |
| A | 1030 | HIS | - | EXPRESSION TAG | UNP Q8A799 |
| B | 21 | MSE | - | EXPRESSION TAG | UNP Q8A799 |
| B | 22 | SER | - | EXPRESSION TAG | UNP Q8A799 |
| B | 23 | LEU | - | EXPRESSION TAG | UNP Q8A799 |
| B | 1023 | GLU | - | EXPRESSION TAG | UNP Q8A799 |
| B | 1024 | GLY | - | EXPRESSION TAG | UNP Q8A799 |
| B | 1025 | HIS | - | EXPRESSION TAG | UNP Q8A799 |
| B | 1026 | HIS | - | EXPRESSION TAG | UNP Q8A799 |
| B | 1027 | HIS | - | EXPRESSION TAG | UNP Q8A799 |
| B | 1028 | HIS | - | EXPRESSION TAG | UNP Q8A799 |
| B | 1029 | HIS | - | EXPRESSION TAG | UNP Q8A799 |
| B | 1030 | HIS | - | EXPRESSION TAG | UNP Q8A799 |

- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

| Mol | Chain | Residues | Atoms | ZeroOcc | AltConf |
|---|---|---|---|---|---|
| 2 | B | 1 | Total Mg<br>1 1 | 0 | 0 |
| 2 | A | 1 | Total Mg<br>1 1 | 0 | 0 |

- Molecule 3 is SODIUM ION (three-letter code: NA) (formula: Na).

| Mol | Chain | Residues | Atoms | ZeroOcc | AltConf |
|---|---|---|---|---|---|
| 3 | B | 2 | Total Na<br>2 2 | 0 | 0 |
| 3 | A | 2 | Total Na<br>2 2 | 0 | 0 |

- Molecule 4 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

| Mol | Chain | Residues | Atoms | ZeroOcc | AltConf |
|---|---|---|---|---|---|
| 4 | B | 1 | Total Cl<br>1 1 | 0 | 0 |
| 4 | A | 1 | Total Cl<br>1 1 | 0 | 0 |

- Molecule 5 is water.

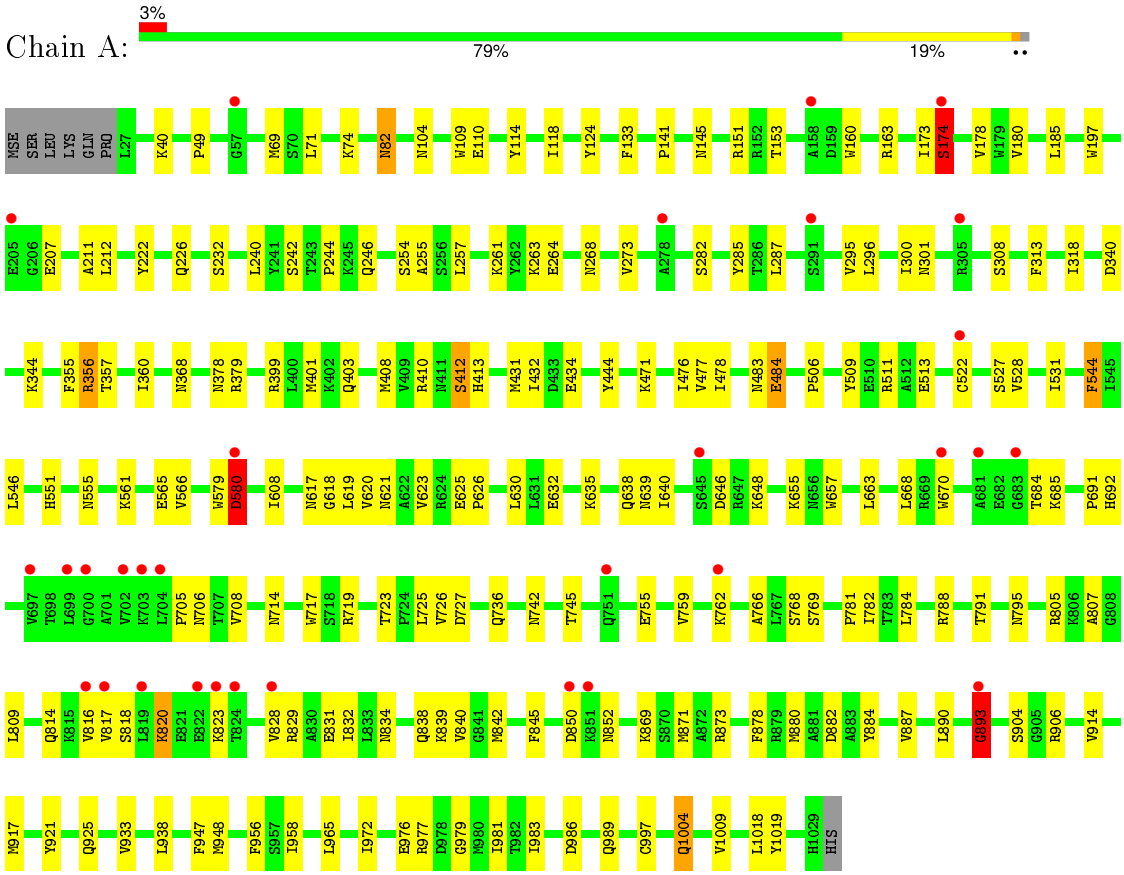
| Mol | Chain | Residues | Atoms | ZeroOcc | AltConf |
|---|---|---|---|---|---|
| 5 | A | 563 | Total O<br>563 563 | 0 | 0 |
| 5 | B | 574 | Total O<br>574 574 | 0 | 0 |

# 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 errors 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.

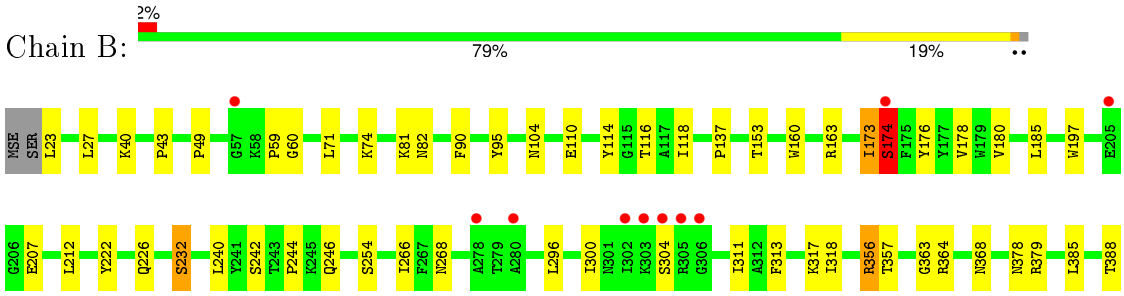
- Molecule 1: Beta-galactosidase

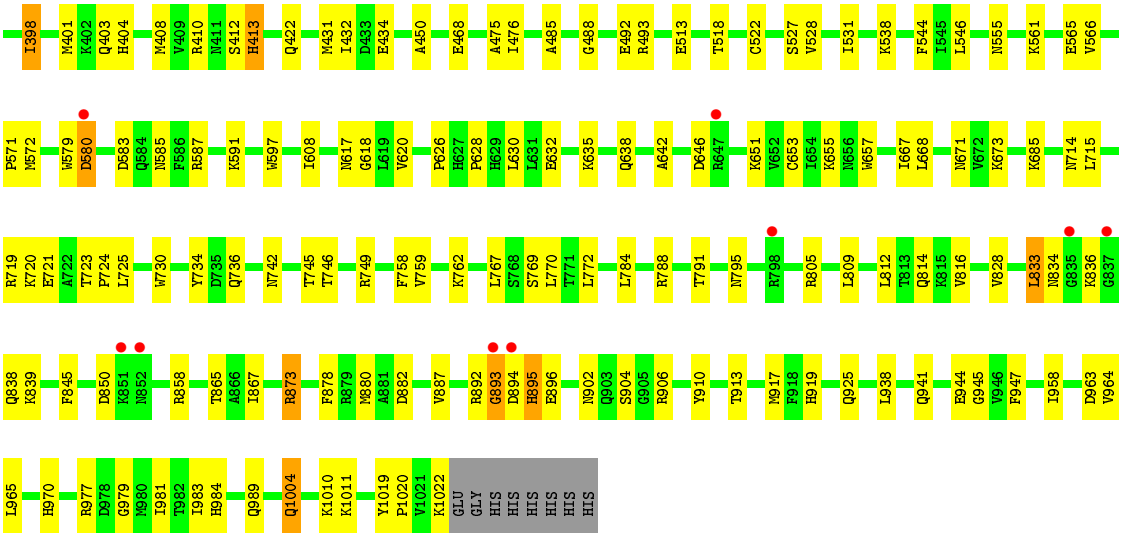
Chain A:




- Molecule 1: Beta-galactosidase

Chain B:




I398 M401 K402 Q403 H404 M408 V409 R410 M411 S412 H413 Q422 M431 I432 D433 E434 A450 E468 A475 I476 A485 G488 E492 R493 E513 T518 C522 S527 V528 I531 K538 F544 L545 L546 N555 K561 E565 V566
P571 M572 W579 D580 D583 Q584 N585 P586 R587 K591 W597 I608 N617 G618 L619 V620 P626 H627 P628 H629 L630 L631 E632 K635 Q638 A642 D646 R647 K651 V652 C653 I654 K655 N656 W657 I667 L668 N671 V672 K673 K685 N714 L715
R719 K720 E721 A722 T723 F724 L725 W730 Y734 D735 Q736 N742 T745 T746 R749 F758 V759 K762 L767 S768 S769 L770 T771 L772 L784 R788 T791 N795 R798 R805 L809 L812 T813 Q814 K815 V816 V828 L833 N834 G835 K836 G837
Q838 K839 F845 D850 K851 N852 R858 T865 A866 I867 R873 F878 R879 M880 A881 D882 V887 R892 G893 D894 H895 E896 N902 Q903 S904 G905 R906 Y910 T913 M917 F918 H919 Q925 L938 Q941 E944 G945 W946 F947 I958 D963 V964
L965 H970 R977 D978 G979 M980 I981 T982 I983 H984 Q989 Q1004 K1010 K1011 Y1019 P1020 V1021 K1022 GLU GLY HIS HIS HIS HIS HIS HIS

# 4 Data and refinement statistics (i)

| Property | Value | Source |
|---|---|---|
| Space group | P 1 21 1 | Depositor |
| Cell constants<br>a, b, c, $\alpha$, $\beta$, $\gamma$ | 88.33Å 138.86Å 102.82Å<br>90.00° 109.73° 90.00° | Depositor |
| Resolution (Å) | 49.76 – 2.10<br>49.76 – 2.09 | Depositor<br>EDS |
| % Data completeness<br>(in resolution range) | 84.7 (49.76-2.10)<br>84.0 (49.76-2.09) | Depositor<br>EDS |
| $R_{merge}$ | 0.08 | Depositor |
| $R_{sym}$ | (Not available) | Depositor |
| $< I/\sigma(I) >$ [1] | 0.50 (at 2.08Å) | Xtriage |
| Refinement program | CNS 1.1 | Depositor |
| R, $R_{free}$ | 0.200 , 0.234<br>0.200 , 0.235 | Depositor<br>DCC |
| $R_{free}$ test set | 5795 reflections (5.03%) | DCC |
| Wilson B-factor (Å$^2$) | 18.2 | Xtriage |
| Anisotropy | 0.737 | Xtriage |
| Bulk solvent $k_{sol}$(e/Å$^3$), $B_{sol}$(Å$^2$) | 0.31 , 44.7 | EDS |
| Estimated twinning fraction | No twinning to report. | Xtriage |
| L-test for twinning[2] | $< \|L\| > = 0.46$, $< L^2 > = 0.29$ | Xtriage |
| Outliers | 0 of 118127 reflections | Xtriage |
| $F_o$,$F_c$ correlation | 0.93 | EDS |
| Total number of atoms | 17243 | wwPDB-VP |
| Average B, all atoms (Å$^2$) | 25.0 | wwPDB-VP |

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.68% 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, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

# 5 Model quality

## 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: NA, MG, CL

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 | 0.36 | 0/8256 | 0.65 | 1/11168 (0.0%) |
| 1 | B | 0.35 | 0/8222 | 0.64 | 5/11122 (0.0%) |
| All | All | 0.36 | 0/16478 | 0.65 | 6/22290 (0.0%) |

There are no bond length outliers.

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

| Mol | Chain | Res | Type | Atoms | Z | Observed($^o$) | Ideal($^o$) |
|---|---|---|---|---|---|---|---|
| 1 | A | 893 | GLY | N-CA-C | 7.27 | 131.28 | 113.10 |
| 1 | B | 893 | GLY | N-CA-C | 5.46 | 126.75 | 113.10 |
| 1 | B | 81 | LYS | N-CA-C | -5.29 | 96.72 | 111.00 |
| 1 | B | 173 | ILE | CA-C-N | -5.27 | 105.60 | 117.20 |
| 1 | B | 173 | ILE | C-N-CA | 5.00 | 134.21 | 121.70 |

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 | 8064 | 0 | 7787 | 159 | 0 |
| 1 | B | 8034 | 0 | 7782 | 148 | 0 |
| 2 | A | 1 | 0 | 0 | 0 | 0 |
| 2 | B | 1 | 0 | 0 | 0 | 0 |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Non-H | H(model) | H(added) | Clashes | Symm-Clashes |
|---|---|---|---|---|---|---|
| 3 | A | 2 | 0 | 0 | 0 | 0 |
| 3 | B | 2 | 0 | 0 | 0 | 0 |
| 4 | A | 1 | 0 | 0 | 0 | 0 |
| 4 | B | 1 | 0 | 0 | 0 | 0 |
| 5 | A | 563 | 0 | 0 | 8 | 0 |
| 5 | B | 574 | 0 | 0 | 14 | 0 |
| All | All | 17243 | 0 | 15569 | 307 | 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 10.

The worst 5 of 307 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:173:ILE:HG13 | 1:A:174:SER:N | 1.77 | 1.00 |
| 1:B:431:MSE:HE3 | 1:B:476:ILE:CG1 | 1.93 | 0.97 |
| 1:B:723:THR:HG22 | 1:B:725:LEU:H | 1.25 | 0.95 |
| 1:A:714:ASN:HD21 | 1:A:736:GLN:HE21 | 1.12 | 0.94 |
| 1:B:431:MSE:HE3 | 1:B:476:ILE:HG13 | 1.49 | 0.93 |

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 | 1001/1010 (99%) | 954 (95%) | 41 (4%) | 6 (1%) | 30 24 |
| 1 | B | 998/1010 (99%) | 948 (95%) | 43 (4%) | 7 (1%) | 26 21 |
| All | All | 1999/2020 (99%) | 1902 (95%) | 84 (4%) | 13 (1%) | 26 21 |

5 of 13 Ramachandran outliers are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | A | 412 | SER |
| 1 | B | 412 | SER |
| 1 | A | 893 | GLY |
| 1 | B | 513 | GLU |
| 1 | B | 895 | 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 | 854/840 (102%) | 844 (99%) | 10 (1%) | 78 84 |
| 1 | B | 852/840 (101%) | 834 (98%) | 18 (2%) | 61 66 |
| All | All | 1706/1680 (102%) | 1678 (98%) | 28 (2%) | 70 76 |

5 of 28 residues with a non-rotameric sidechain are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | B | 356 | ARG |
| 1 | B | 522 | CYS |
| 1 | B | 941 | GLN |
| 1 | B | 398 | ILE |
| 1 | B | 413 | HIS |

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 49 such sidechains are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | A | 989 | GLN |
| 1 | B | 301 | ASN |
| 1 | B | 929 | ASN |
| 1 | B | 85 | ASN |
| 1 | B | 328 | HIS |

### 5.3.3 RNA (i)

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

Of 8 ligands modelled in this entry, 8 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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 | 983/1010 (97%) | 0.11 | 31 (3%) | 51 | 60 | 10, 24, 43, 55 | 0 |
| 1 | B | 980/1010 (97%) | -0.05 | 19 (1%) | 70 | 75 | 11, 23, 39, 58 | 0 |
| All | All | 1963/2020 (97%) | 0.03 | 50 (2%) | 61 | 67 | 10, 23, 41, 58 | 0 |

The worst 5 of 50 RSRZ outliers are listed below:

| Mol | Chain | Res | Type | RSRZ |
|---|---|---|---|---|
| 1 | B | 305 | ARG | 7.2 |
| 1 | B | 304 | SER | 5.5 |
| 1 | B | 306 | GLY | 4.6 |
| 1 | B | 303 | LYS | 4.5 |
| 1 | A | 819 | LEU | 3.9 |

## 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

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, $95^{th}$ percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

| Mol | Type | Chain | Res | Atoms | RSCC | RSR | LLDF | B-factors(Å$^2$) | Q<0.9 |
|---|---|---|---|---|---|---|---|---|---|
| 2 | MG | A | 1 | 1/1 | 0.98 | 0.10 | -0.47 | 10,10,10,10 | 0 |
| 3 | NA | B | 5 | 1/1 | 0.99 | 0.08 | -1.05 | 18,18,18,18 | 0 |
| 3 | NA | B | 3 | 1/1 | 0.95 | 0.09 | -1.06 | 18,18,18,18 | 0 |
| 3 | NA | A | 4 | 1/1 | 0.98 | 0.07 | -1.72 | 15,15,15,15 | 0 |
| 3 | NA | A | 6 | 1/1 | 0.98 | 0.06 | -1.88 | 21,21,21,21 | 0 |
| 2 | MG | B | 2 | 1/1 | 0.98 | 0.06 | -2.84 | 20,20,20,20 | 0 |
| 4 | CL | A | 7 | 1/1 | 0.99 | 0.07 | -2.90 | 12,12,12,12 | 0 |
| 4 | CL | B | 8 | 1/1 | 0.99 | 0.04 | -3.87 | 14,14,14,14 | 0 |

## 6.5 Other polymers (i)

There are no such residues in this entry.