wwPDB X-ray Structure Validation Summary Report (i)

Jan 31, 2016 – 07:21 PM GMT

PDB ID	:	1F4A
Title	:	E. COLI (LACZ) BETA-GALACTOSIDASE (NCS CONSTRAINED MONO MER-ORTHORHOMBIC)
Authors	:	Juers, D.H.; Jacobson, R.H.; Wigley, D.; Zhang, X.J.; Huber, R.E.; Tronrud, D.E.; Matthews, B.W.
Deposited on	:	2000-06-07
Resolution	:	2.80 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:

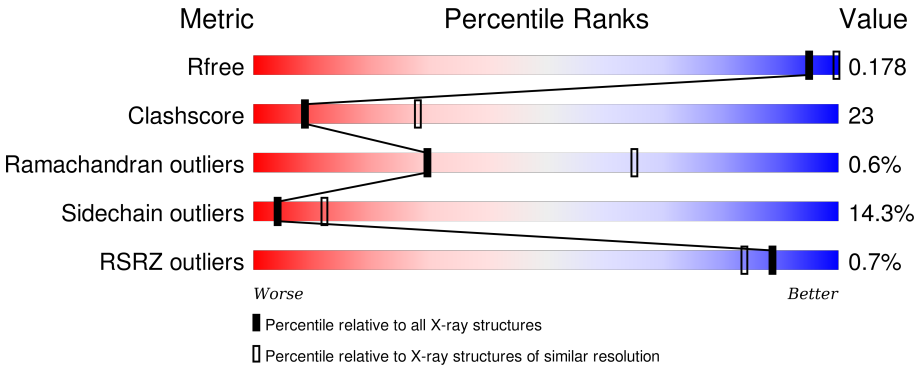
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	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 2.80 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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	2393 (2.80-2.80)
Clashscore	102246	2827 (2.80-2.80)
Ramachandran outliers	100387	2782 (2.80-2.80)
Sidechain outliers	100360	2784 (2.80-2.80)
RSRZ outliers	91569	2404 (2.80-2.80)

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	1021	% 49% 38% 11% •
1	B	1021	% 49% 38% 11% •
1	C	1021	49% 38% 11% •
1	D	1021	% 49% 38% 11% •

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 34424 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	1021	Total	C	N	O	S	84	7	0
			8238	5209	1466	1525	38			
1	B	1021	Total	C	N	O	S	84	7	0
			8238	5209	1466	1525	38			
1	C	1021	Total	C	N	O	S	84	7	0
			8238	5209	1466	1525	38			
1	D	1021	Total	C	N	O	S	84	7	0
			8238	5209	1466	1525	38			

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	2	Total	Mg	0	0
			2	2		
2	A	2	Total	Mg	0	0
			2	2		
2	D	2	Total	Mg	0	0
			2	2		
2	C	2	Total	Mg	0	0
			2	2		

- Molecule 3 is water.

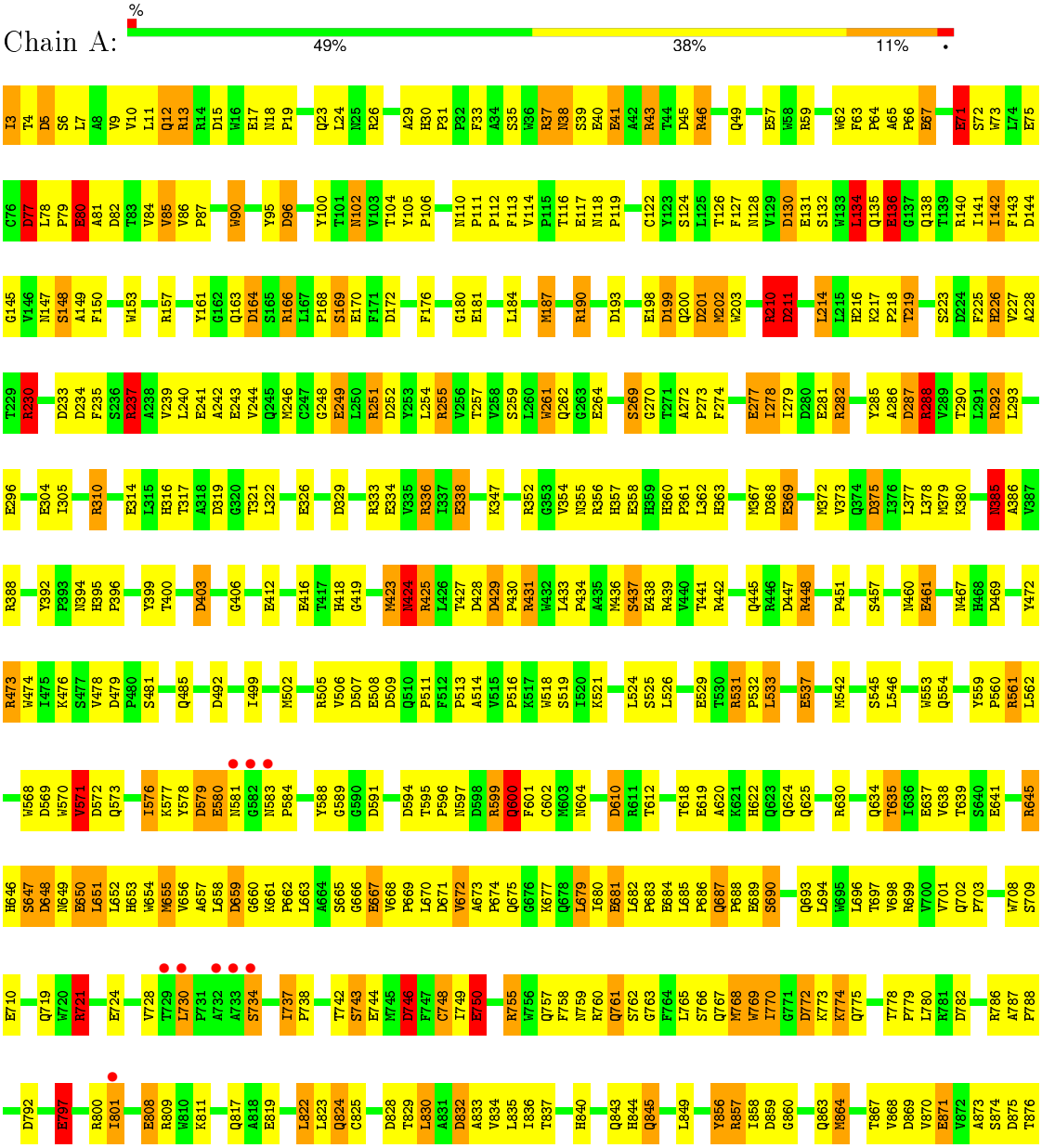
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	365	Total	O	0	0
			365	365		
3	B	366	Total	O	0	0
			366	366		
3	C	367	Total	O	0	0
			367	367		
3	D	366	Total	O	0	0
			366	366		

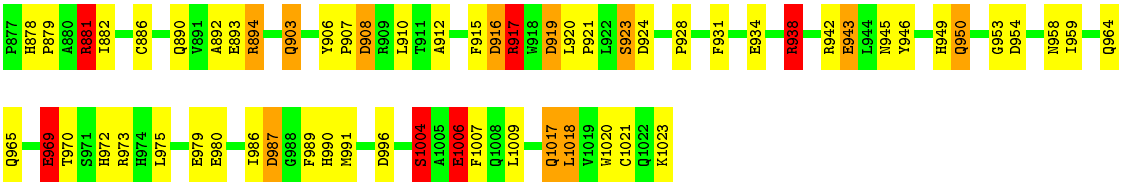
3 Residue-property plots (i)

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

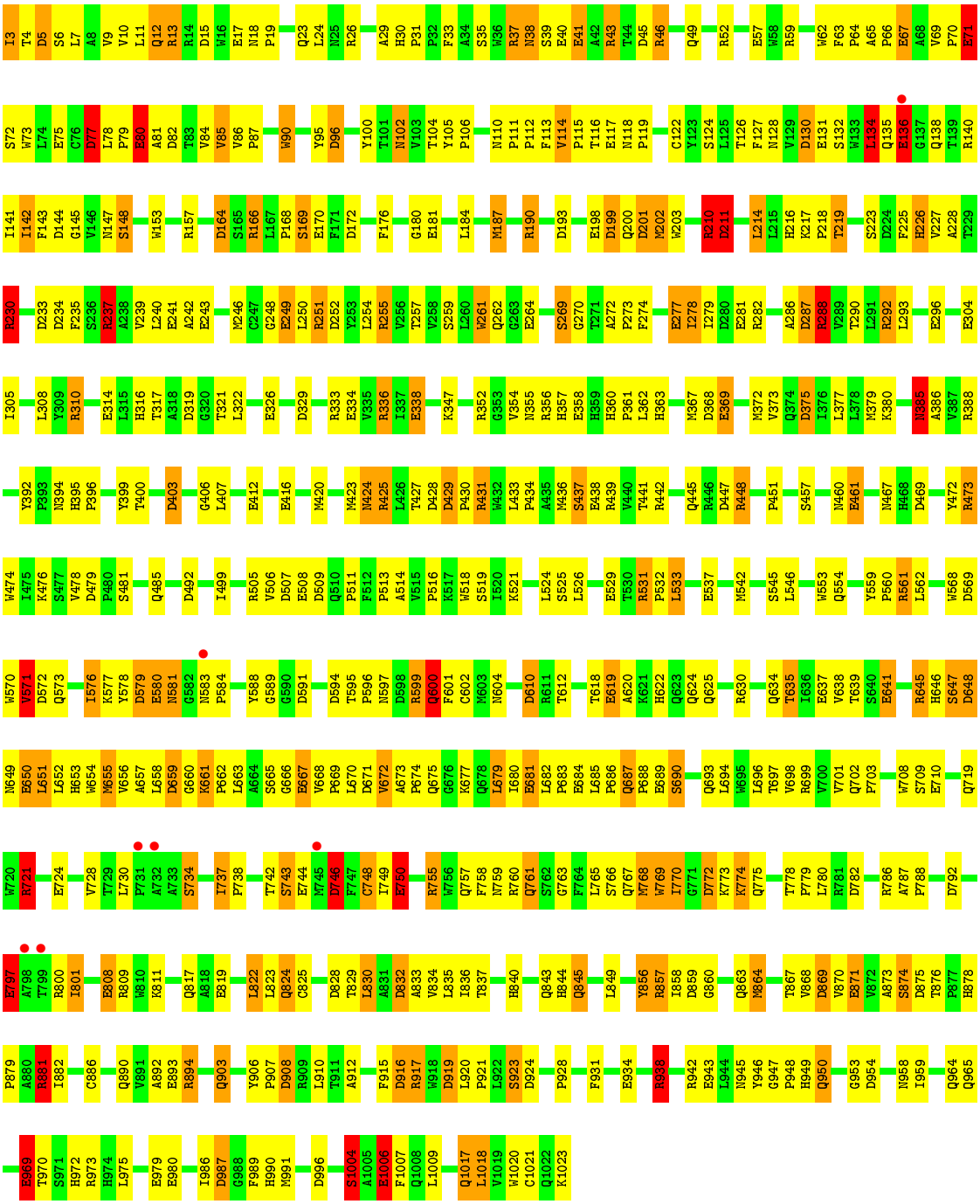
Chain A:





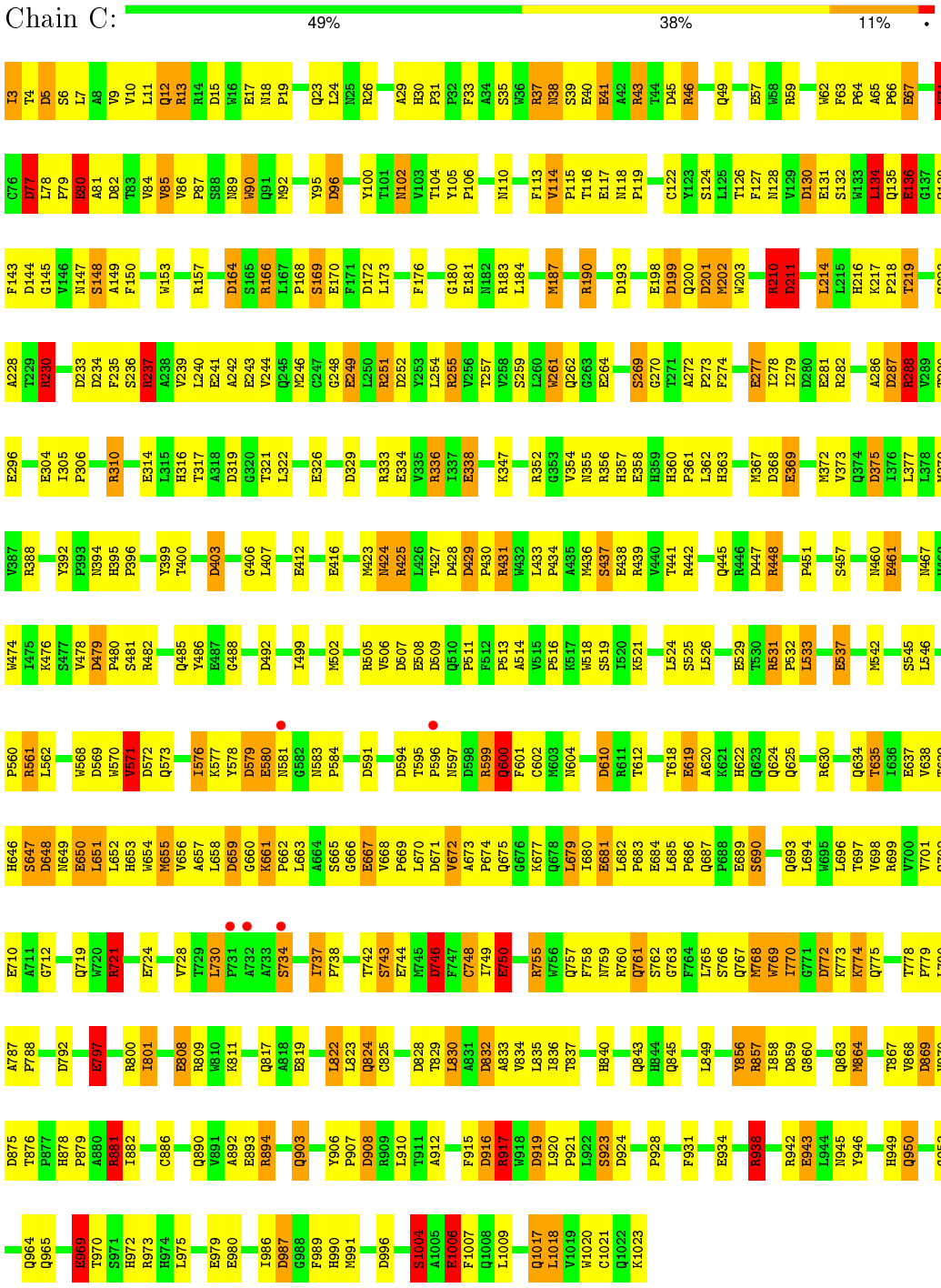
• Molecule 1: BETA-GALACTOSIDASE

Chain B:



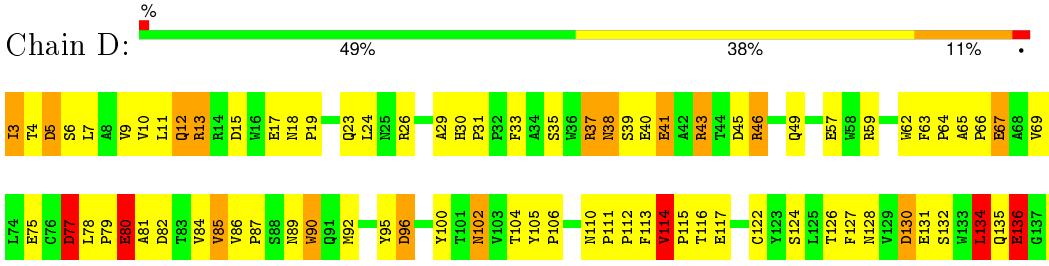
• Molecule 1: BETA-GALACTOSIDASE

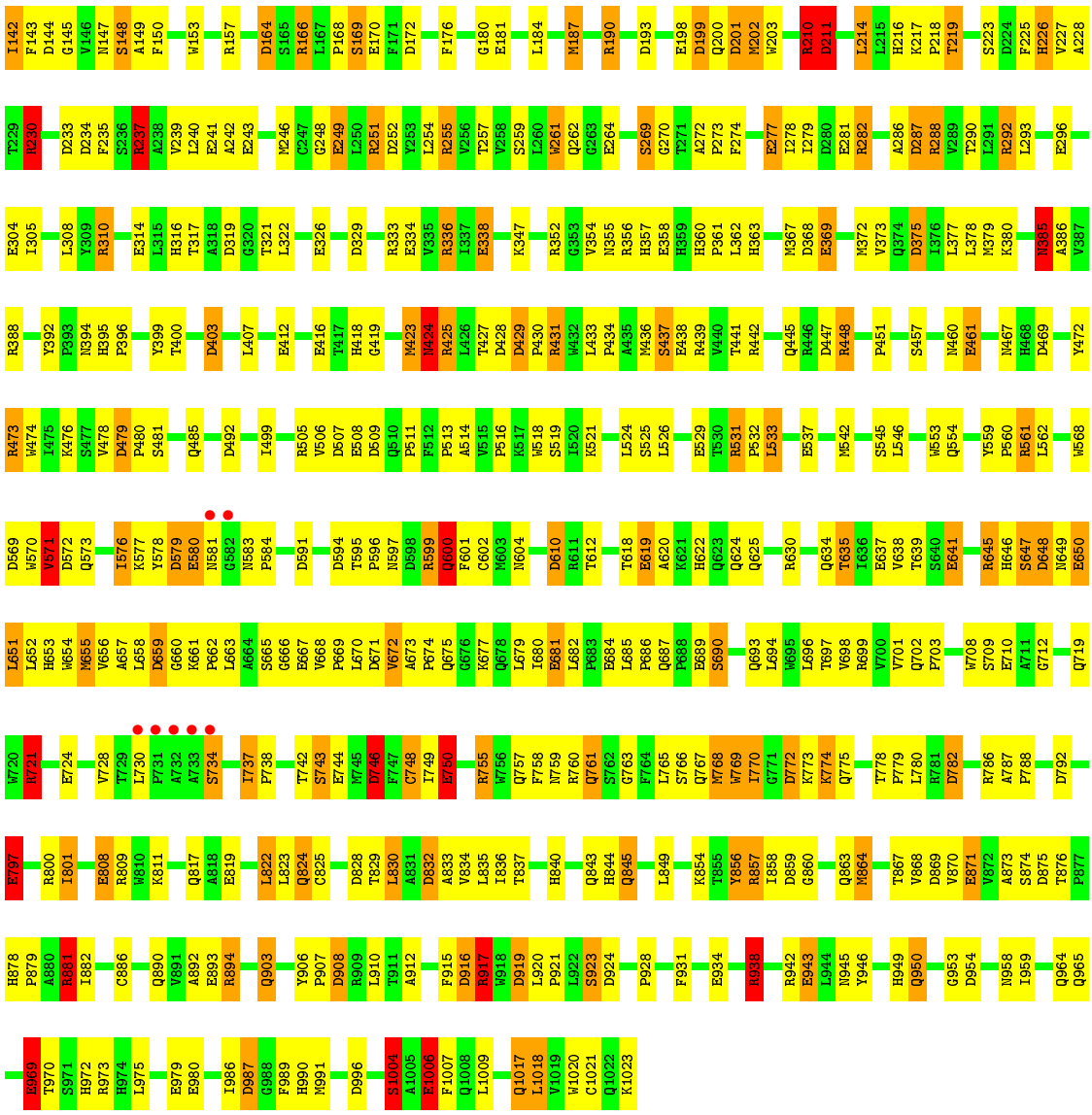
Chain C:



• Molecule 1: BETA-GALACTOSIDASE

Chain D:




I142 F143 D144 G145 V146 N147 S148 A149 F150 W153 R157 D164 S165 R166 L167 P168 S169 E170 F171 D172 F176 G180 E181 L184 M187 R190 D193 E198 D199 Q200 D201 M202 W203 R210 D211 L214 L215 H216 K217 P218 T219 S223 D224 F225 H226 V227 A228
T229 R230 D233 D234 F235 S236 R237 A238 V239 L240 E241 A242 E243 M246 C247 G248 E249 L250 R251 D252 Y253 L254 R255 V256 T257 V258 S259 L260 W261 Q262 G263 E264 S269 G270 T271 A272 P273 F274 E277 I278 I279 D280 E281 R282 A286 D287 R288 V289 T290 L291 R292 L293 E296
E304 I305 L308 Y309 R310 E314 L315 H316 T317 A318 D319 G320 T321 L322 E326 D329 R333 E334 V335 R336 I337 E338 K347 R352 G353 V354 N355 R356 H357 E358 H359 H360 P361 L362 H363 M367 D368 E369 M372 V373 Q374 D375 I376 L377 L378 M379 K380 N385 A386 V387
R388 Y392 P393 N394 H395 P396 Y399 T400 D403 L407 E412 E416 T417 H418 G419 N423 N424 R425 L426 T427 D428 D429 P430 R431 W432 L433 P434 A435 M436 S437 E438 R439 V440 T441 R442 Q445 R446 D447 R448 P451 S457 N460 E461 N467 H468 D469 Y472
R473 W474 I475 K476 S477 V478 D479 P480 S481 Q485 D492 I499 R505 V506 D507 E508 D509 Q510 P511 F512 P513 A514 V515 P516 K517 W518 S519 I520 K521 L524 S525 L526 E529 T530 R531 P532 L533 E537 M542 S545 L546 W553 Q554 Y559 P560 R561 L562 W568
D569 W570 V571 D572 Q573 I576 K577 Y578 D579 E580 N581 G582 N583 P584 D591 D594 T595 P596 N597 D598 R599 Q600 F601 G602 N603 N604 D610 R611 T612 T618 E619 A620 K621 H622 Q623 Q624 Q625 R630 Q634 T635 I636 E637 V638 T639 S640 E641 R645 H646 S647 D648 N649 E650
L651 L652 H653 W654 M655 V656 A657 L658 D659 G660 K661 P662 L663 A664 S665 G666 E667 V668 P669 L670 D671 V672 A673 P674 Q675 G676 K677 Q678 L679 I680 E681 L682 P683 E684 L685 P686 Q687 P688 E689 S690 Q693 L694 W695 L696 T697 V698 R699 V700 V701 Q702 P703 W708 S709 E710 A711 G712 Q719
W720 R721 E724 V728 T729 L730 F731 A732 A733 S734 I737 P738 T742 S743 E744 M745 D746 F747 C748 I749 E750 R755 W756 Q757 F758 M759 R760 Q761 S762 G763 F764 L765 S766 Q767 M768 W769 I770 G771 D772 K773 K774 Q775 T778 P779 L780 R781 D782 R786 A787 P788 D792
E797 R800 I801 E808 R809 W810 K811 Q817 A818 E819 L822 L823 Q824 C825 D828 T829 L830 A831 D832 A833 V834 L835 I836 T837 H840 Q843 H844 Q845 L849 K854 T855 Y856 R857 I858 D859 G860 Q863 M864 T867 V868 D869 V870 E871 Y872 A873 S874 D875 T876 P877
H878 P879 A880 R881 I882 C886 Q890 V891 A892 E893 R894 Q903 Y906 P907 D908 R909 L910 T911 A912 F915 D916 R917 W918 D919 L920 P921 L922 S923 D924 P928 F931 E934 R938 R942 E943 L944 N945 Y946 H949 Q950 G953 D954 N958 I959 Q964 Q965
E969 T970 S971 H972 R973 H974 L975 E979 E980 I986 D987 G988 F989 H990 M991 D996 S1004 A1005 E1006 F1007 Q1008 L1009 Q1017 L1018 V1019 W1020 C1021 Q1022 K1023

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	153.40Å 173.40Å 204.40Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	25.00 – 2.80 24.97 – 2.82	Depositor EDS
% Data completeness (in resolution range)	88.0 (25.00-2.80) 80.5 (24.97-2.82)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.68 (at 2.80Å)	Xtriage
Refinement program	TNT V. 5-E	Depositor
R, R_{free}	0.167 , 0.198 0.149 , 0.178	Depositor DCC
R_{free} test set	1590 reflections (1.53%)	DCC
Wilson B-factor (Å^2)	40.7	Xtriage
Anisotropy	0.148	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.26 , 100.3	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Outliers	0 of 105768 reflections	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	34424	wwPDB-VP
Average B, all atoms (Å^2)	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 41.86 % of the origin peak, indicating pseudo translational symmetry.The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 2.2179e-04.The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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: MG

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	1.08	52/8515 (0.6%)	1.61	173/11615 (1.5%)
1	B	1.08	52/8515 (0.6%)	1.61	175/11615 (1.5%)
1	C	1.08	52/8515 (0.6%)	1.61	174/11615 (1.5%)
1	D	1.08	52/8515 (0.6%)	1.61	176/11615 (1.5%)
All	All	1.08	208/34060 (0.6%)	1.61	698/46460 (1.5%)

The worst 5 of 208 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	75	GLU	CD-OE2	9.53	1.36	1.25
1	C	75	GLU	CD-OE2	9.53	1.36	1.25
1	D	75	GLU	CD-OE2	9.49	1.36	1.25
1	B	75	GLU	CD-OE2	9.46	1.36	1.25
1	D	710	GLU	CD-OE2	7.62	1.34	1.25

The worst 5 of 698 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	809[A]	ARG	NE-CZ-NH1	11.76	126.18	120.30
1	B	809[B]	ARG	NE-CZ-NH1	11.76	126.18	120.30
1	A	809[A]	ARG	NE-CZ-NH1	11.62	126.11	120.30
1	A	809[B]	ARG	NE-CZ-NH1	11.62	126.11	120.30
1	D	809[A]	ARG	NE-CZ-NH1	11.61	126.11	120.30

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	8238	0	7824	394	0
1	B	8238	0	7824	374	0
1	C	8238	0	7824	379	0
1	D	8238	0	7824	374	0
2	A	2	0	0	0	0
2	B	2	0	0	0	0
2	C	2	0	0	0	0
2	D	2	0	0	0	0
3	A	365	0	0	10	0
3	B	366	0	0	10	0
3	C	367	0	0	10	0
3	D	366	0	0	10	0
All	All	34424	0	31296	1496	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 23.

The worst 5 of 1496 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:427:THR:HA	1:B:436:MET:HE1	1.21	1.11
1:B:746:ASP:HA	1:B:760:ARG:HG3	1.34	1.10
1:D:746:ASP:HA	1:D:760:ARG:HG3	1.34	1.09
1:A:746:ASP:HA	1:A:760:ARG:HG3	1.34	1.09
1:C:427:THR:HA	1:C:436:MET:HE1	1.28	1.08

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	1026/1021 (100%)	953 (93%)	67 (6%)	6 (1%)	30	65
1	B	1026/1021 (100%)	953 (93%)	67 (6%)	6 (1%)	30	65
1	C	1026/1021 (100%)	953 (93%)	67 (6%)	6 (1%)	30	65
1	D	1026/1021 (100%)	953 (93%)	67 (6%)	6 (1%)	30	65
All	All	4104/4084 (100%)	3812 (93%)	268 (6%)	24 (1%)	30	65

5 of 24 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	647	SER
1	B	647	SER
1	C	647	SER
1	D	647	SER
1	A	77	ASP

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	880/873 (101%)	755 (86%)	125 (14%)	4	12
1	B	880/873 (101%)	755 (86%)	125 (14%)	4	12
1	C	880/873 (101%)	755 (86%)	125 (14%)	4	12
1	D	880/873 (101%)	755 (86%)	125 (14%)	4	12
All	All	3520/3492 (101%)	3020 (86%)	500 (14%)	4	12

5 of 500 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	822	LEU
1	C	223	SER
1	D	755	ARG

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	857	ARG
1	C	37	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 94 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	634	GLN
1	C	216	HIS
1	D	622	HIS
1	B	817	GLN
1	B	1017	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 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	1021/1021 (100%)	-0.81	9 (0%)	85	79	8, 29, 70, 100	19 (1%)
1	B	1021/1021 (100%)	-0.80	7 (0%)	89	84	8, 29, 70, 100	19 (1%)
1	C	1021/1021 (100%)	-0.84	5 (0%)	91	88	8, 29, 70, 100	19 (1%)
1	D	1021/1021 (100%)	-0.81	7 (0%)	89	84	8, 29, 70, 100	19 (1%)
All	All	4084/4084 (100%)	-0.81	28 (0%)	89	84	8, 29, 71, 100	76 (1%)

The worst 5 of 28 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	732	ALA	4.5
1	A	732	ALA	4.1
1	C	581	ASN	3.7
1	A	581	ASN	3.6
1	A	801	ILE	3.2

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	D	3002	1/1	0.96	0.12	1.95	31,31,31,31	0
2	MG	B	3002	1/1	0.94	0.11	1.60	31,31,31,31	0
2	MG	A	3002	1/1	0.98	0.11	0.98	31,31,31,31	0
2	MG	C	3002	1/1	0.96	0.12	0.97	31,31,31,31	0
2	MG	B	3001	1/1	0.99	0.06	-1.48	28,28,28,28	0
2	MG	C	3001	1/1	0.99	0.04	-2.05	28,28,28,28	0
2	MG	A	3001	1/1	0.96	0.08	-2.15	28,28,28,28	0
2	MG	D	3001	1/1	0.99	0.06	-2.52	28,28,28,28	0

6.5 Other polymers

There are no such residues in this entry.