wwPDB X-ray Structure Validation Summary Report (i)

Aug 2, 2023 – 11:54 PM EDT

PDB ID	:	1HN1
Title	:	E. COLI (LAC Z) BETA-GALACTOSIDASE (ORTHORHOMBIC)
Authors	:	Juers, D.H.; Matthews, B.W.
Deposited on	:	2000-12-05
Resolution	:	3.00 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 (i)) were used in the production of this report:

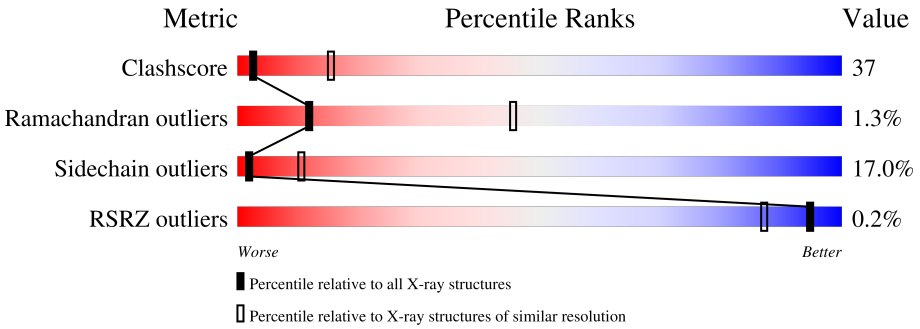
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.34
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.34

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 3.00 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))
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)

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	1023	32% 50% 16%
1	B	1023	30% 47% 20%
1	C	1023	33% 48% 16%
1	D	1023	35% 46% 15%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 32954 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	1011	Total	C	N	O	S	0	2	0
			8136	5145	1443	1510	38			
1	B	1011	Total	C	N	O	S	0	2	0
			8136	5145	1443	1510	38			
1	C	1011	Total	C	N	O	S	0	2	0
			8136	5145	1443	1510	38			
1	D	1011	Total	C	N	O	S	0	1	0
			8130	5141	1441	1510	38			

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	cloning artifact	UNP P00722
A	2	SER	-	cloning artifact	UNP P00722
A	3	HIS	-	cloning artifact	UNP P00722
A	4	MET	-	cloning artifact	UNP P00722
A	5	LEU	-	cloning artifact	UNP P00722
A	6	GLU	-	cloning artifact	UNP P00722
A	7	ASP	-	cloning artifact	UNP P00722
A	8	PRO	-	cloning artifact	UNP P00722
B	1	GLY	-	cloning artifact	UNP P00722
B	2	SER	-	cloning artifact	UNP P00722
B	3	HIS	-	cloning artifact	UNP P00722
B	4	MET	-	cloning artifact	UNP P00722
B	5	LEU	-	cloning artifact	UNP P00722
B	6	GLU	-	cloning artifact	UNP P00722
B	7	ASP	-	cloning artifact	UNP P00722
B	8	PRO	-	cloning artifact	UNP P00722
C	1	GLY	-	cloning artifact	UNP P00722
C	2	SER	-	cloning artifact	UNP P00722
C	3	HIS	-	cloning artifact	UNP P00722
C	4	MET	-	cloning artifact	UNP P00722
C	5	LEU	-	cloning artifact	UNP P00722

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	6	GLU	-	cloning artifact	UNP P00722
C	7	ASP	-	cloning artifact	UNP P00722
C	8	PRO	-	cloning artifact	UNP P00722
D	1	GLY	-	cloning artifact	UNP P00722
D	2	SER	-	cloning artifact	UNP P00722
D	3	HIS	-	cloning artifact	UNP P00722
D	4	MET	-	cloning artifact	UNP P00722
D	5	LEU	-	cloning artifact	UNP P00722
D	6	GLU	-	cloning artifact	UNP P00722
D	7	ASP	-	cloning artifact	UNP P00722
D	8	PRO	-	cloning artifact	UNP P00722

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

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

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

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	2	Total Na 2 2	0	0
3	B	2	Total Na 2 2	0	0
3	C	2	Total Na 2 2	0	0
3	D	1	Total Na 1 1	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	103	Total O 103 103	0	0

Continued on next page...

Continued from previous page...

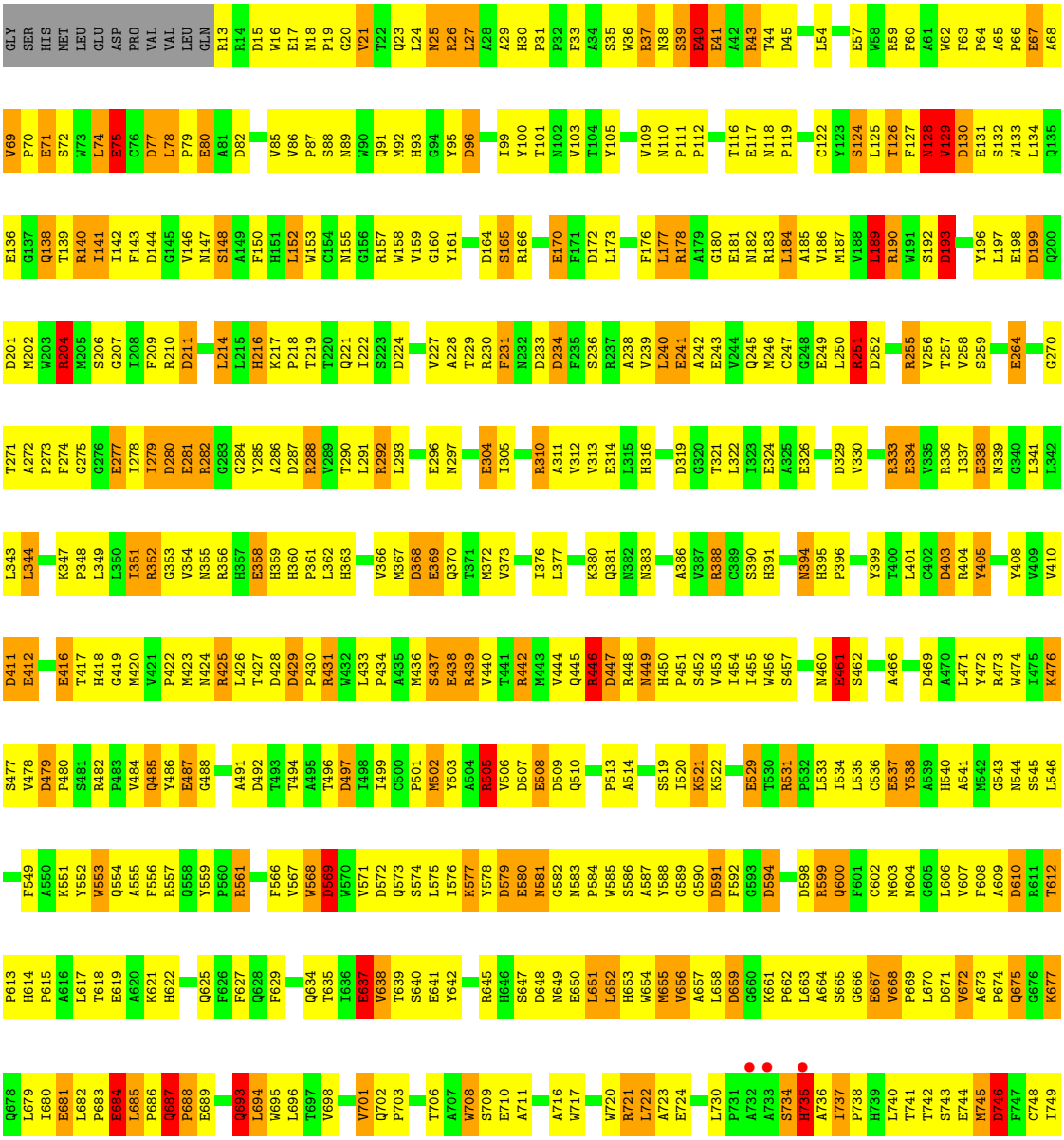
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	92	Total O 92 92	0	0
4	C	108	Total O 108 108	0	0
4	D	98	Total O 98 98	0	0

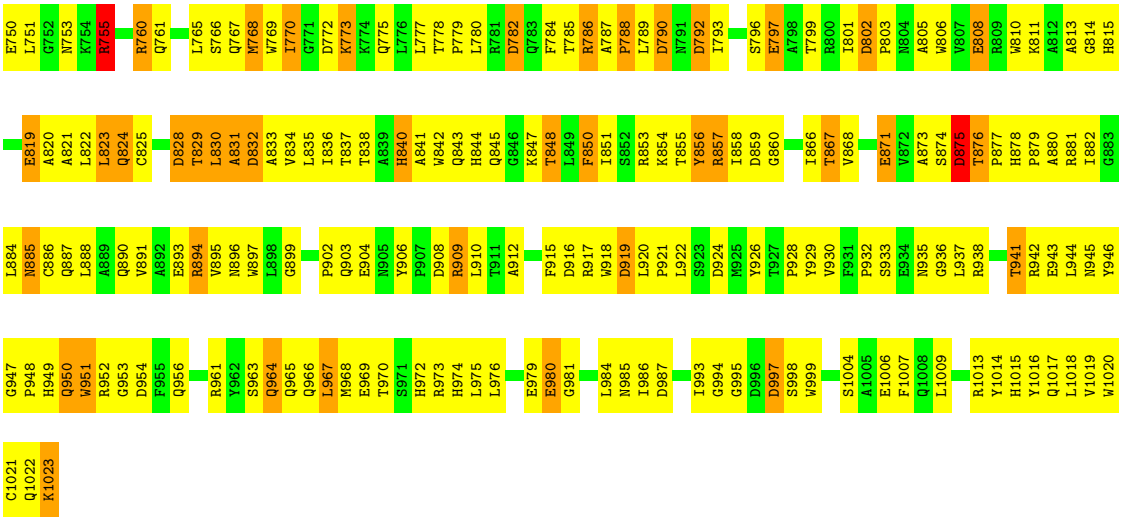
3 Residue-property plots (i)

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: BETA-GALACTOSIDASE

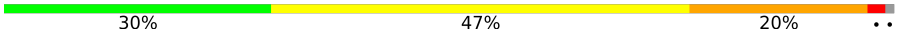
Chain A:

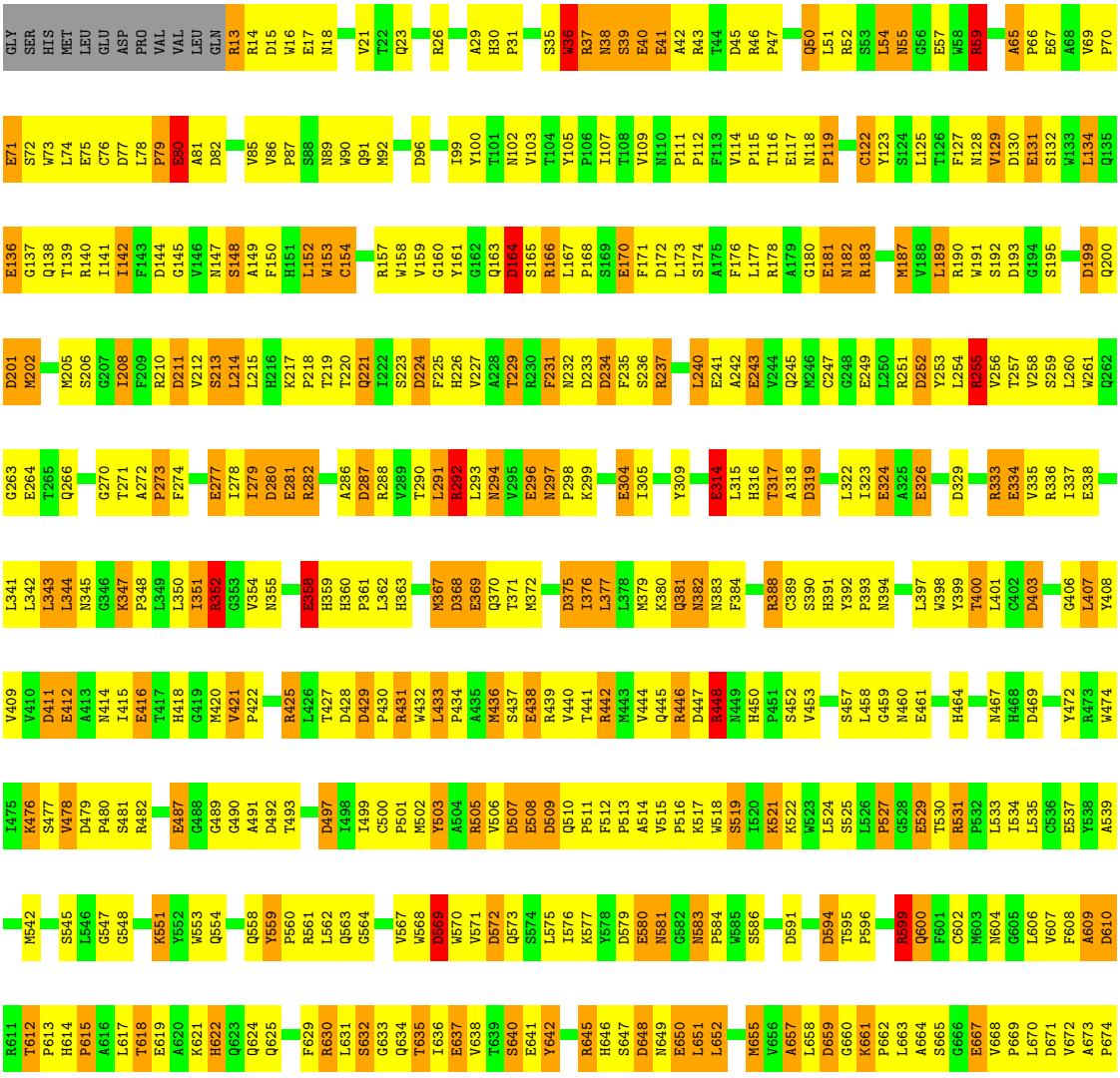


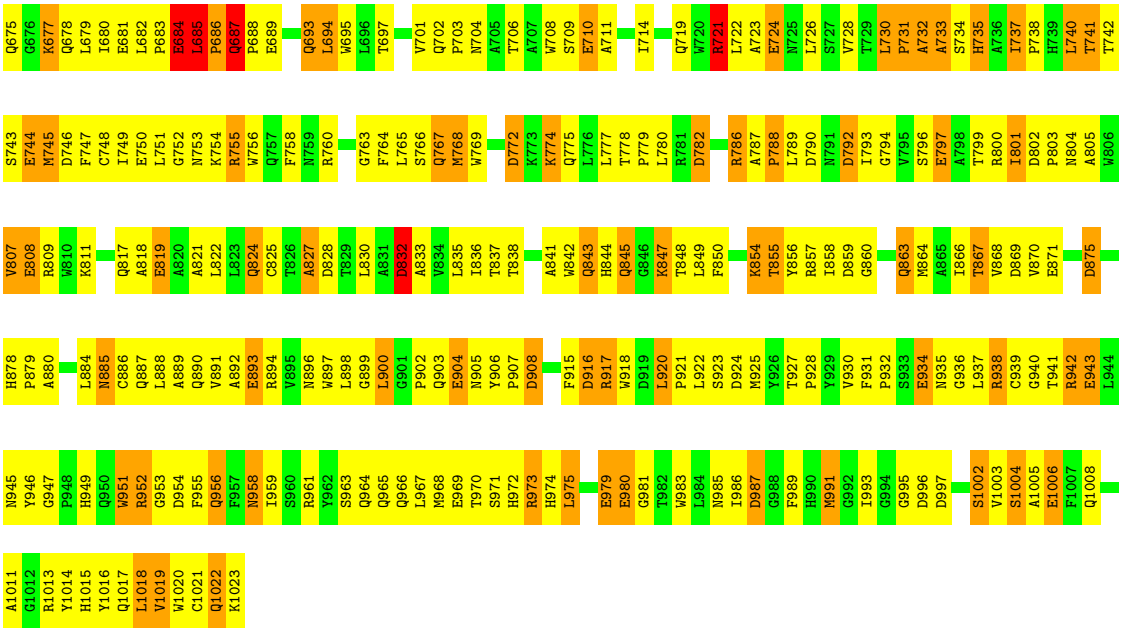


• Molecule 1: BETA-GALACTOSIDASE

Chain B:

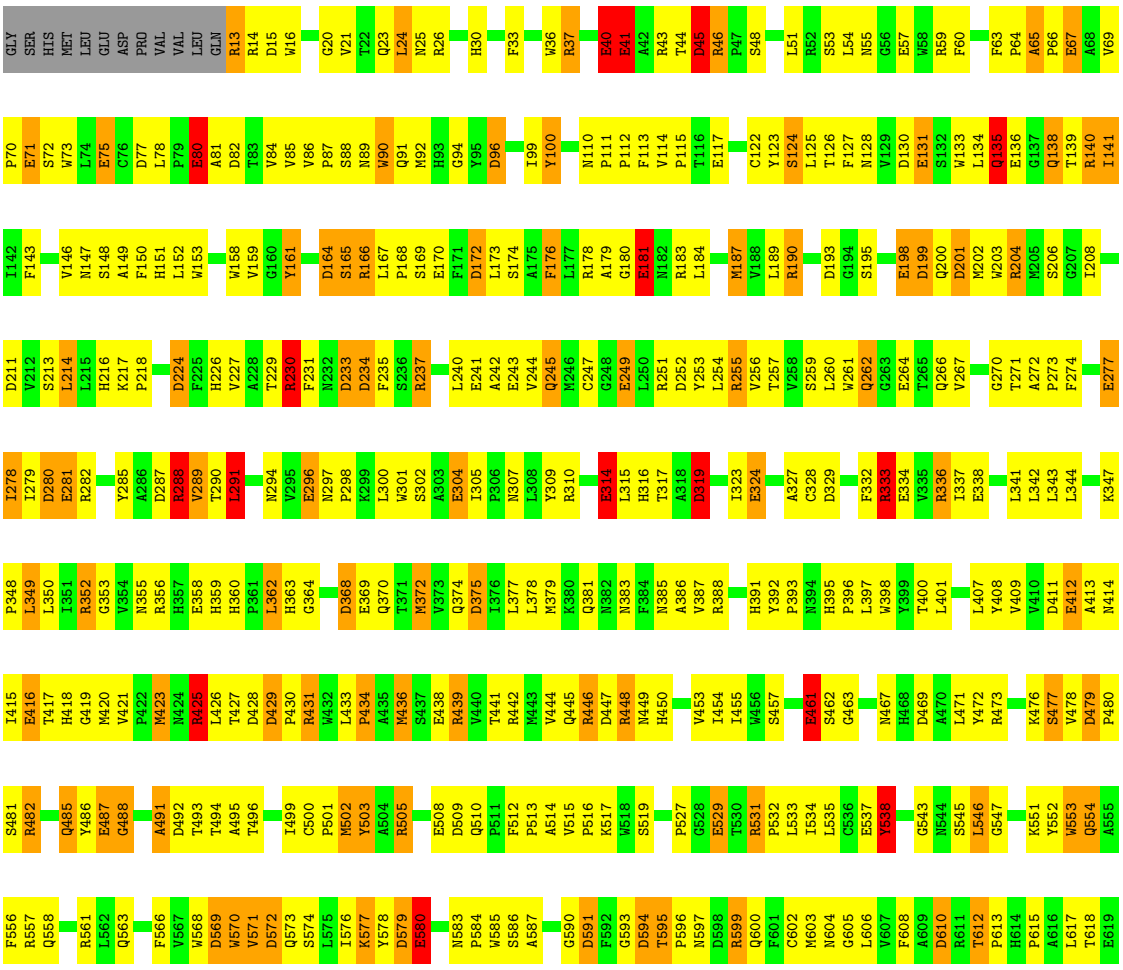


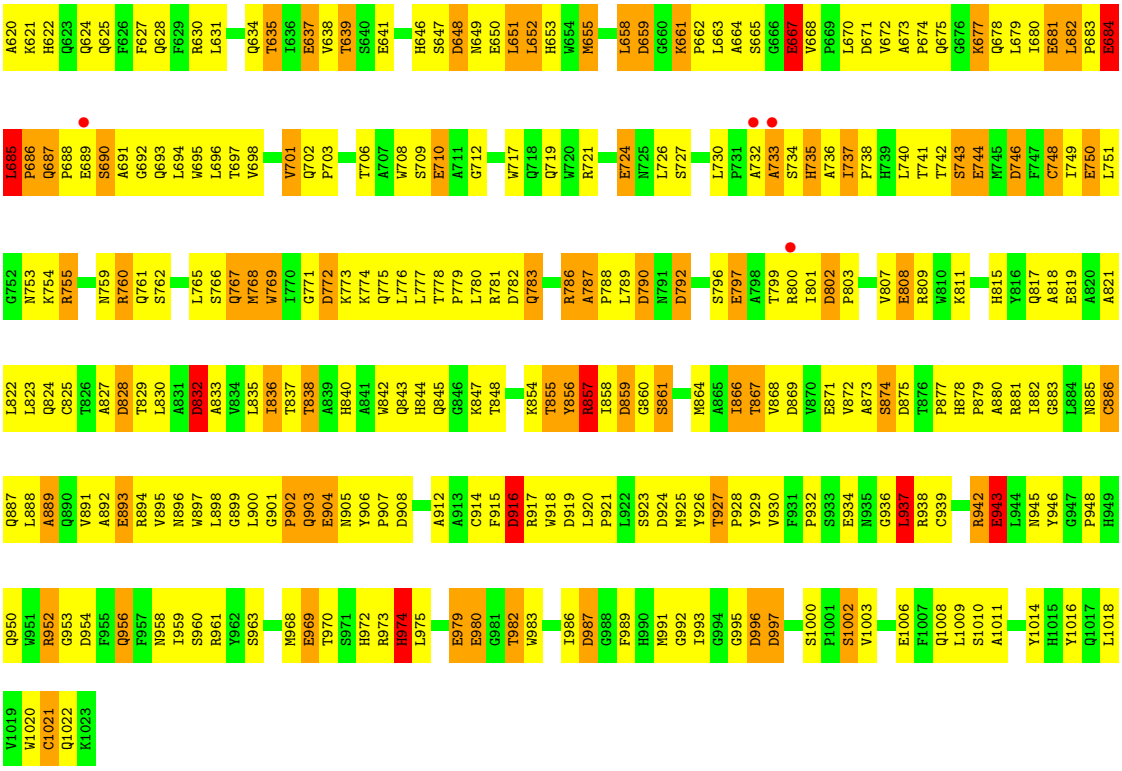




• Molecule 1: BETA-GALACTOSIDASE

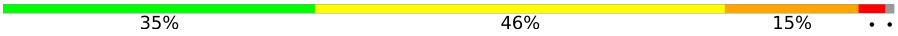
Chain C:

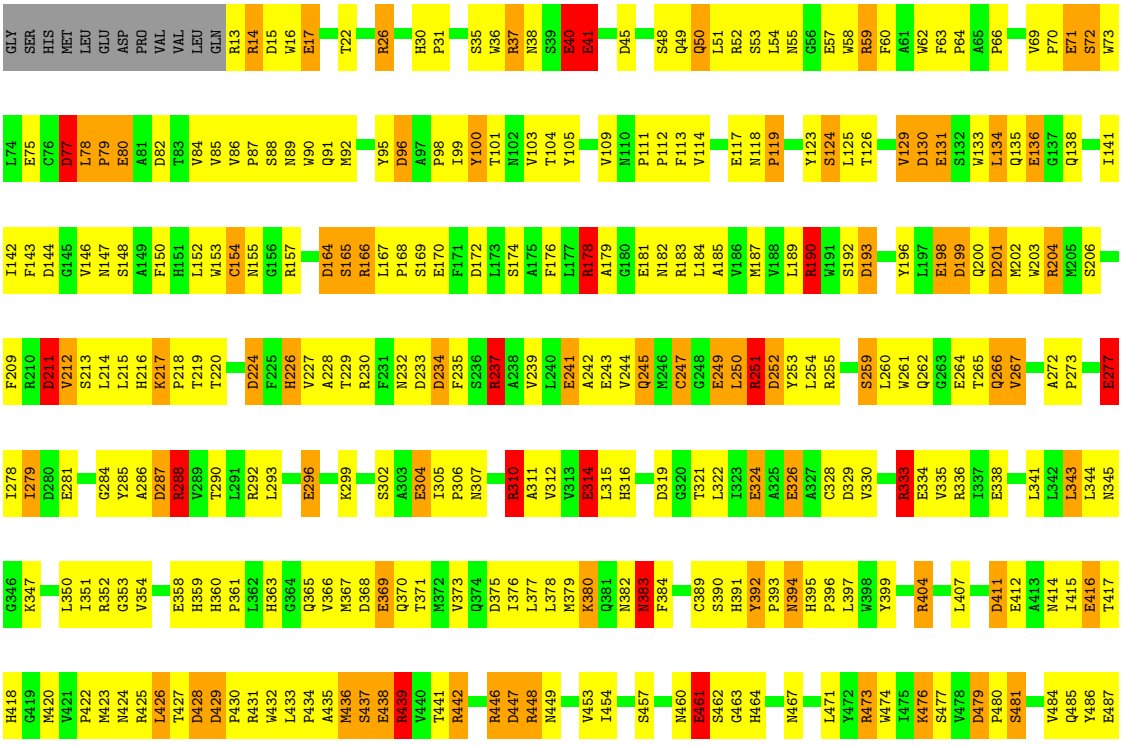


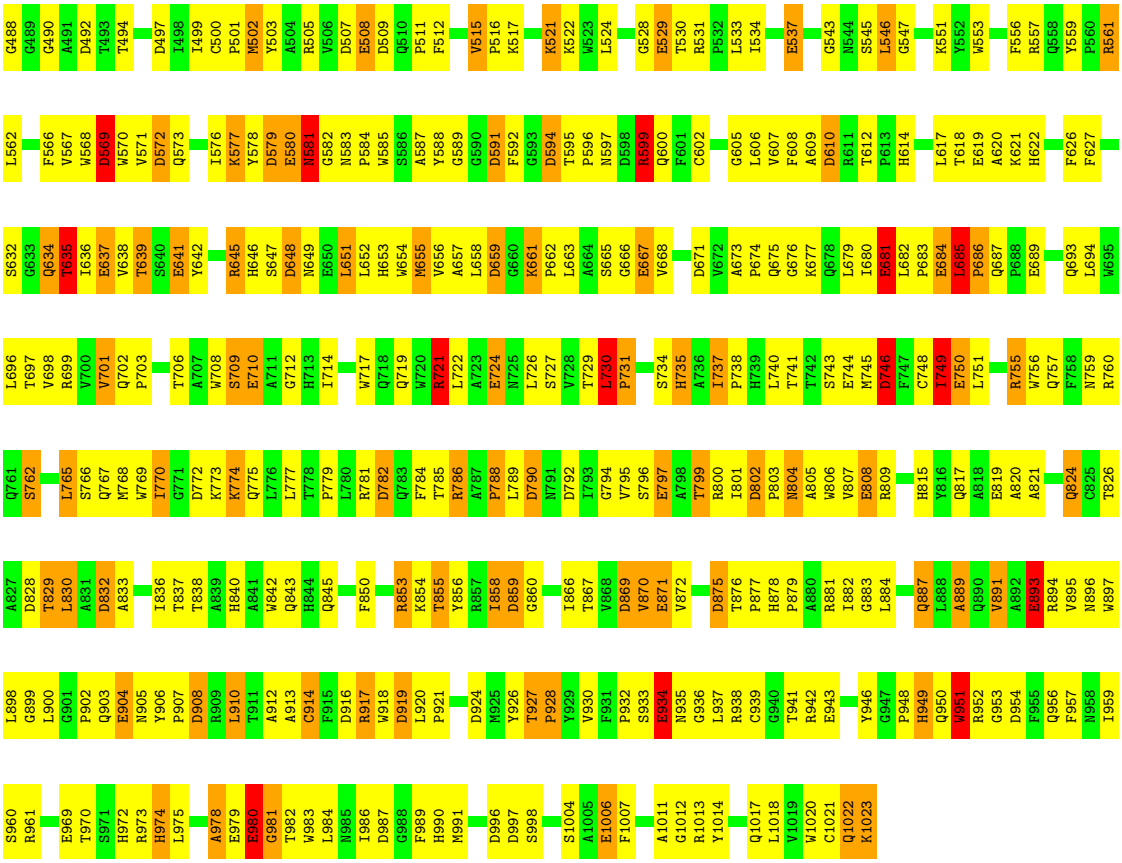


• Molecule 1: BETA-GALACTOSIDASE

Chain D:






G488 G489 G490 A491 D492 T493 T494 D497 I498 I499 C500 P501 M502 Y503 A504 R505 V506 D507 E508 D509 Q510 P511 F512 V515 P516 K517 K521 K522 W523 L524 G528 E529 T530 R531 P532 L533 I534 E537 G543 N544 S545 L546 G547 K551 Y552 W553 F556 R557 Q558 Y559 P560 R561
L562 F566 V567 W568 D569 W570 V571 D572 Q573 I576 K577 Y578 D579 E580 N581 G582 N583 P584 W585 S586 A587 Y588 G589 G590 D591 F592 G593 D594 T595 P596 N597 D598 R599 Q600 F601 C602 G605 L606 V607 F608 A609 D610 R611 T612 P613 H614 L617 T618 E619 A620 K621 H622 F626 F627
S632 G633 Q634 T635 I636 E637 V638 T639 S640 E641 Y642 R645 H646 S647 D648 N649 E650 L651 L652 H653 W654 M655 V656 A657 L658 D659 G660 K661 P662 L663 A664 S665 G666 E667 V668 D671 V672 A673 P674 Q675 G676 K677 Q678 L679 I680 E681 L682 P683 E684 L685 P686 Q687 P688 E689 Q693 L694 W695
L696 T697 V698 R699 V700 V701 Q702 P703 T706 A707 W708 S709 E710 A711 G712 H713 I714 W717 Q718 Q719 W720 R721 L722 A723 E724 N725 L726 S727 V728 T729 L730 P731 S734 H735 A736 I737 P738 H739 L740 T741 T742 S743 E744 M745 D746 F747 C748 I749 E750 L751 R755 W756 Q757 F758 N759 R760
Q761 S762 L765 S766 Q767 M768 W769 I770 G771 D772 K773 K774 Q775 L776 L777 T778 P779 L780 R781 D782 Q783 F784 T785 R786 A787 P788 L789 D790 N791 D792 I793 G794 V795 S796 E797 A798 T799 R800 I801 D802 P803 N804 A805 W806 V807 E808 R809 H815 Y816 Q817 A818 E819 A820 A821 Q824 C825 T826
A827 D828 T829 L830 A831 D832 A833 I836 T837 T838 A839 H840 A841 W842 Q843 H844 Q845 F850 R853 K854 T855 Y856 R857 I858 D859 G860 I866 T867 V868 D869 V870 E871 V872 D875 T876 P877 H878 P879 A880 R881 I882 G883 L884 Q887 L888 A889 Q890 V891 A892 E893 R894 V895 N896 W897
L898 G899 L900 G901 P902 Q903 E904 N905 Y906 P907 D908 R909 L910 T911 A912 A913 C914 F915 D916 R917 W918 D919 L920 P921 D924 W925 Y926 T927 P928 Y929 V930 F931 P932 S933 E934 N935 G936 L937 R938 C939 G940 T941 R942 E943 Y946 G947 P948 H949 Q950 W951 R952 G953 D964 F965 Q956 F957 N958 I959
S960 R961 E969 T970 S971 H972 R973 H974 L975 A978 E979 E980 G981 T982 W983 L984 N985 I986 D987 G988 F989 H990 M991 D996 D997 S998 S1004 A1005 E1006 F1007 A1011 G1012 R1013 Y1014 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	153.90Å 171.40Å 204.50Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	15.00 – 3.00 29.96 – 2.99	Depositor EDS
% Data completeness (in resolution range)	94.0 (15.00-3.00) 88.8 (29.96-2.99)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.47 (at 3.00Å)	Xtriage
Refinement program	TNT	Depositor
R, R_{free}	0.148 , 0.299 0.134 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	31.4	Xtriage
Anisotropy	0.253	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.25 , 132.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	32954	wwPDB-VP
Average B, all atoms (Å^2)	42.0	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 41.90 % 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 2.2078e-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, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality (i)

5.1 Standard geometry (i)

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

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	58/8387 (0.7%)	1.66	166/11442 (1.5%)
1	B	1.07	55/8387 (0.7%)	1.65	170/11442 (1.5%)
1	C	1.07	56/8387 (0.7%)	1.65	157/11442 (1.4%)
1	D	1.09	58/8376 (0.7%)	1.65	168/11427 (1.5%)
All	All	1.08	227/33537 (0.7%)	1.65	661/45753 (1.4%)

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	B	1	0
1	D	1	0
All	All	2	0

The worst 5 of 227 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	314	GLU	CD-OE2	8.73	1.35	1.25
1	D	304	GLU	CD-OE2	8.61	1.35	1.25
1	D	334	GLU	CD-OE2	8.13	1.34	1.25
1	A	304	GLU	CD-OE2	8.00	1.34	1.25
1	C	461	GLU	CD-OE2	7.99	1.34	1.25

The worst 5 of 661 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	685	LEU	C-N-CD	-21.41	73.50	120.60
1	B	730	LEU	C-N-CD	-21.06	74.26	120.60

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	987	ASP	CB-CG-OD2	-13.47	106.18	118.30
1	A	687	GLN	C-N-CD	-12.61	92.86	120.60
1	D	166	ARG	NE-CZ-NH2	-12.33	114.13	120.30

All (2) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	B	37	ARG	CA
1	D	951	TRP	CA

There are no planarity outliers.

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	8136	0	7723	593	0
1	B	8136	0	7723	660	0
1	C	8136	0	7723	548	0
1	D	8130	0	7720	550	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	2	0	0	0	0
3	B	2	0	0	0	0
3	C	2	0	0	0	0
3	D	1	0	0	0	0
4	A	103	0	0	2	0
4	B	92	0	0	9	0
4	C	108	0	0	8	0
4	D	98	0	0	6	0
All	All	32954	0	30889	2319	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 37.

The worst 5 of 2319 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:502:MET:HB2	1:D:537:GLU:HB2	1.25	1.18
1:D:734:SER:HB3	1:D:860:GLY:HA3	1.20	1.11
1:B:18:ASN:HD22	1:B:21:VAL:HG23	1.09	1.10
1:B:737:ILE:HD12	1:B:738:PRO:HD2	1.26	1.09
1:A:737:ILE:HG13	1:A:832:ASP:HA	1.35	1.08

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	1011/1023 (99%)	902 (89%)	99 (10%)	10 (1%)	15	53
1	B	1011/1023 (99%)	901 (89%)	96 (10%)	14 (1%)	11	43
1	C	1011/1023 (99%)	904 (89%)	93 (9%)	14 (1%)	11	43
1	D	1010/1023 (99%)	904 (90%)	93 (9%)	13 (1%)	12	45
All	All	4043/4092 (99%)	3611 (89%)	381 (9%)	51 (1%)	12	45

5 of 51 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	129	VAL
1	A	425	ARG
1	A	688	PRO
1	B	201	ASP
1	B	647	SER

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	866/875 (99%)	723 (84%)	143 (16%)	2 11
1	B	866/875 (99%)	699 (81%)	167 (19%)	1 8
1	C	866/875 (99%)	732 (84%)	134 (16%)	2 13
1	D	865/875 (99%)	721 (83%)	144 (17%)	2 11
All	All	3463/3500 (99%)	2875 (83%)	588 (17%)	2 10

5 of 588 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	152	LEU
1	D	914	CYS
1	D	251	ARG
1	D	138	GLN
1	D	645	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 85 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	687	GLN
1	D	394	ASN
1	C	704	ASN
1	C	878	HIS
1	D	702	GLN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 15 ligands modelled in this entry, 15 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 (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	1011/1023 (98%)	-0.82	3 (0%) 94 84	10, 35, 92, 211	0
1	B	1011/1023 (98%)	-0.78	0 100 100	8, 38, 101, 212	0
1	C	1011/1023 (98%)	-0.79	4 (0%) 92 79	14, 34, 90, 214	0
1	D	1011/1023 (98%)	-0.83	0 100 100	5, 34, 93, 210	0
All	All	4044/4092 (98%)	-0.81	7 (0%) 95 87	5, 35, 94, 214	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	689	GLU	3.0
1	A	735	HIS	2.7
1	A	733	ALA	2.6
1	C	733	ALA	2.4
1	C	732	ALA	2.1

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

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. 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	B-factors(Å^2)	Q<0.9
3	NA	B	3102	1/1	0.93	0.21	32,32,32,32	0
2	MG	B	3002	1/1	0.94	0.12	30,30,30,30	0
3	NA	C	3101	1/1	0.94	0.43	49,49,49,49	0
3	NA	A	3101	1/1	0.95	0.25	52,52,52,52	0
2	MG	A	3001	1/1	0.96	0.26	29,29,29,29	0
3	NA	B	3103	1/1	0.96	0.13	45,45,45,45	0
3	NA	A	3102	1/1	0.96	0.38	24,24,24,24	0
3	NA	C	3102	1/1	0.96	0.23	24,24,24,24	0
2	MG	C	3001	1/1	0.97	0.26	25,25,25,25	0
2	MG	A	3002	1/1	0.98	0.08	29,29,29,29	0
2	MG	D	3001	1/1	0.98	0.28	26,26,26,26	0
2	MG	D	3002	1/1	0.98	0.22	30,30,30,30	0
3	NA	D	3102	1/1	0.98	0.30	47,47,47,47	0
2	MG	C	3002	1/1	0.99	0.18	25,25,25,25	0
2	MG	B	3001	1/1	0.99	0.15	13,13,13,13	0

6.5 Other polymers (i)

There are no such residues in this entry.