wwPDB X-ray Structure Validation Summary Report (i)

Jan 4, 2023 – 01:37 pm GMT

PDB ID	:	8BBL
Title	:	SGL a GH20 family sulfoglycosidase
Authors	:	Dong, M.D.; Roth, C.R.; Jin, Y.J.
Deposited on	:	2022-10-13
Resolution	:	2.71 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:

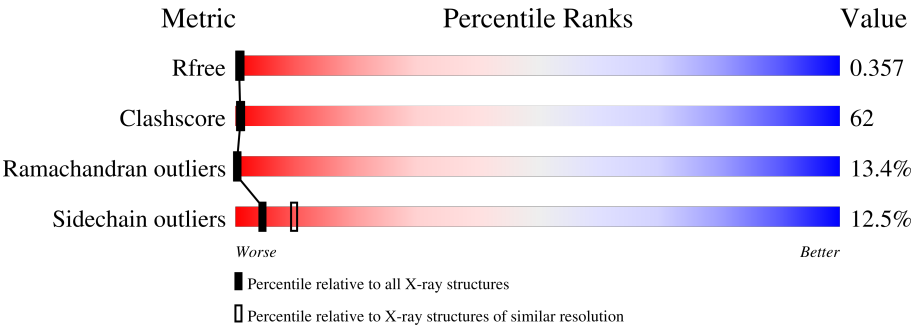
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.31.3
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.3

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 2.71 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.



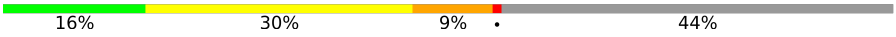
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3359 (2.74-2.70)
Clashscore	141614	3686 (2.74-2.70)
Ramachandran outliers	138981	3622 (2.74-2.70)
Sidechain outliers	138945	3623 (2.74-2.70)

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%

Mol	Chain	Length	Quality of chain
1	A	901	9% 33% 13% • 43%
1	B	901	16% 29% 9% • 44%
1	C	901	15% 29% 10% • 45%
1	D	901	13% 31% 9% • 45%
1	E	901	16% 29% 10% • 45%
1	F	901	15% 31% 9% • 45%
1	G	901	15% 31% 9% • 45%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	H	901	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 31772 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-N-acetylhexosaminidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	515	Total	C	N	O	S	0	0	0
			4061	2584	691	764	22			
1	B	501	Total	C	N	O	S	0	0	0
			3959	2521	671	745	22			
1	C	500	Total	C	N	O	S	0	0	0
			3955	2519	670	744	22			
1	D	500	Total	C	N	O	S	0	0	0
			3955	2519	670	744	22			
1	E	500	Total	C	N	O	S	0	0	0
			3955	2519	670	744	22			
1	F	500	Total	C	N	O	S	0	0	0
			3955	2519	670	744	22			
1	H	501	Total	C	N	O	S	0	0	0
			3959	2521	671	745	22			
1	G	500	Total	C	N	O	S	0	0	0
			3955	2519	670	744	22			

- Molecule 2 is water.

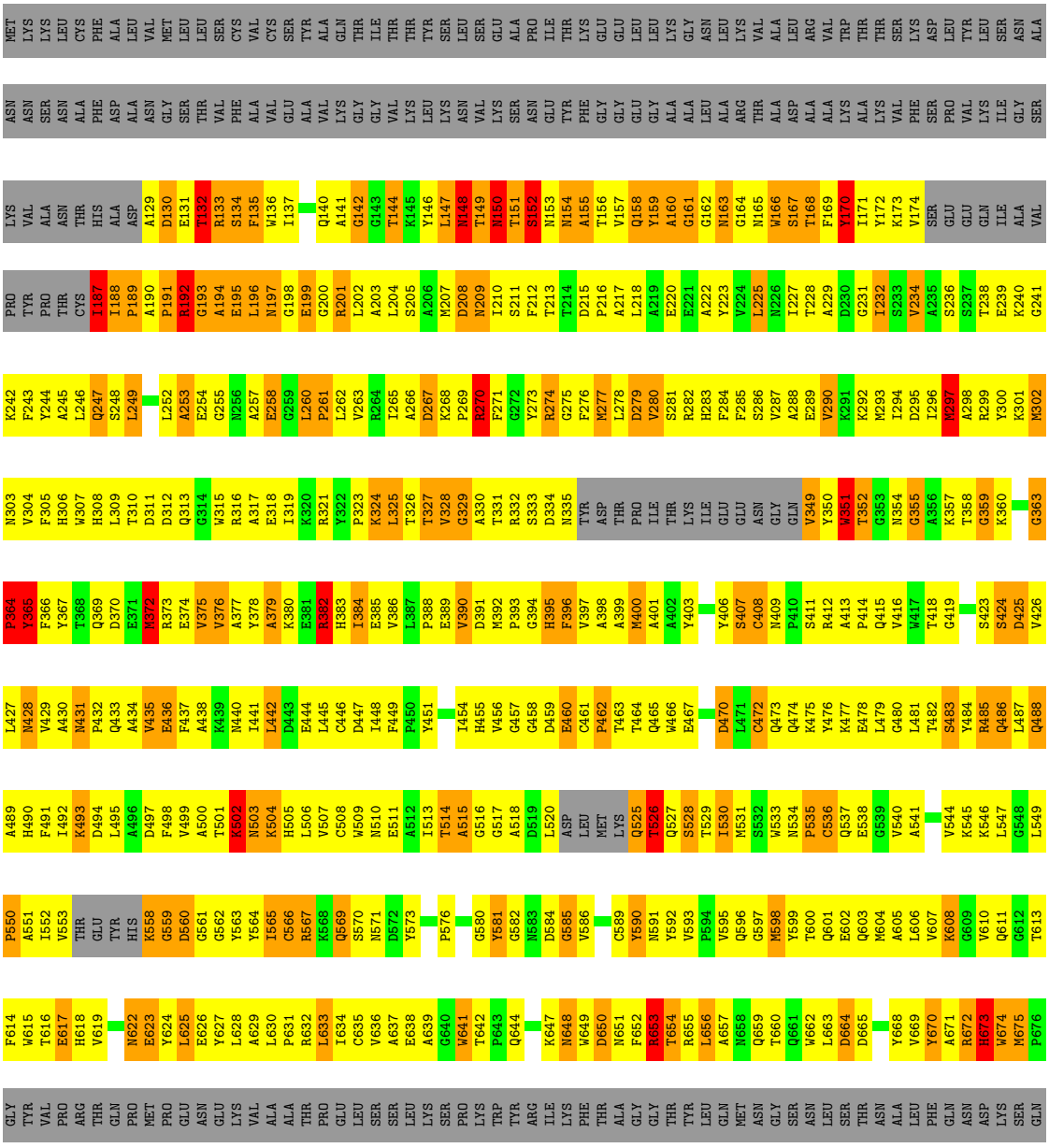
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	10	Total	O	0	0
			10	10		
2	B	4	Total	O	0	0
			4	4		
2	C	4	Total	O	0	0
			4	4		

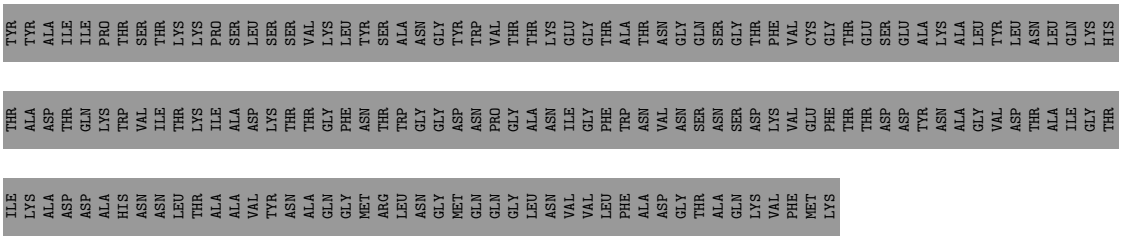
3 Residue-property plots (i)

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. 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. 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-N-acetylhexosaminidase

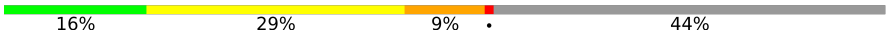
Chain A:

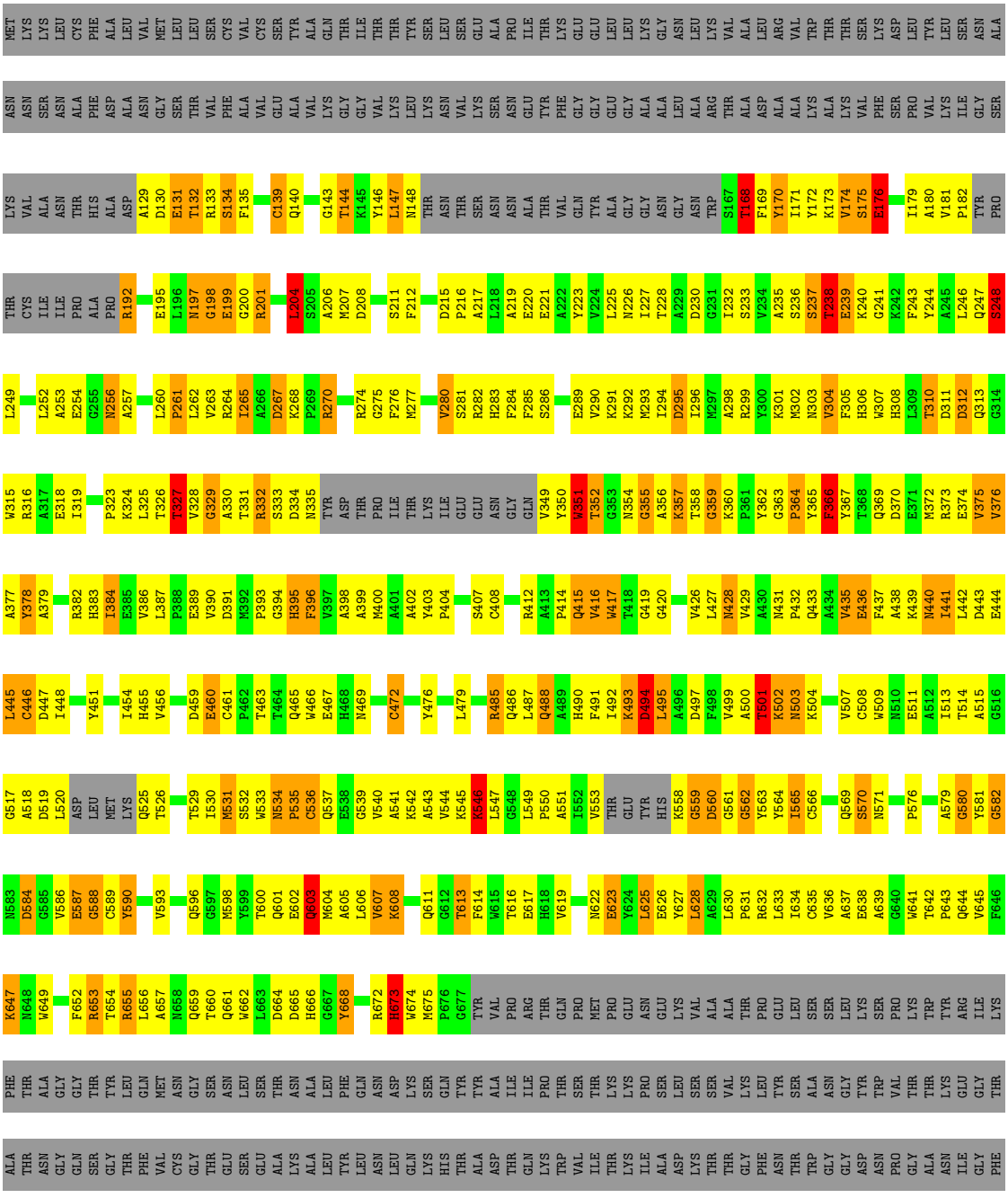


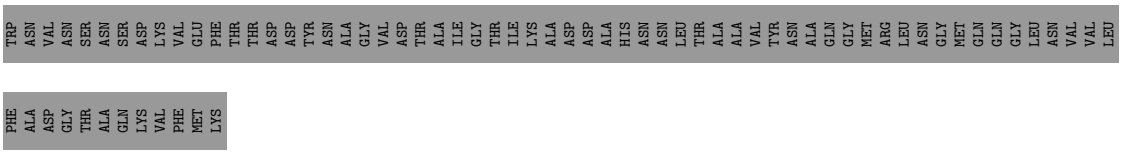


• Molecule 1: Beta-N-acetylhexosaminidase

Chain B:

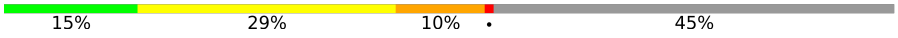


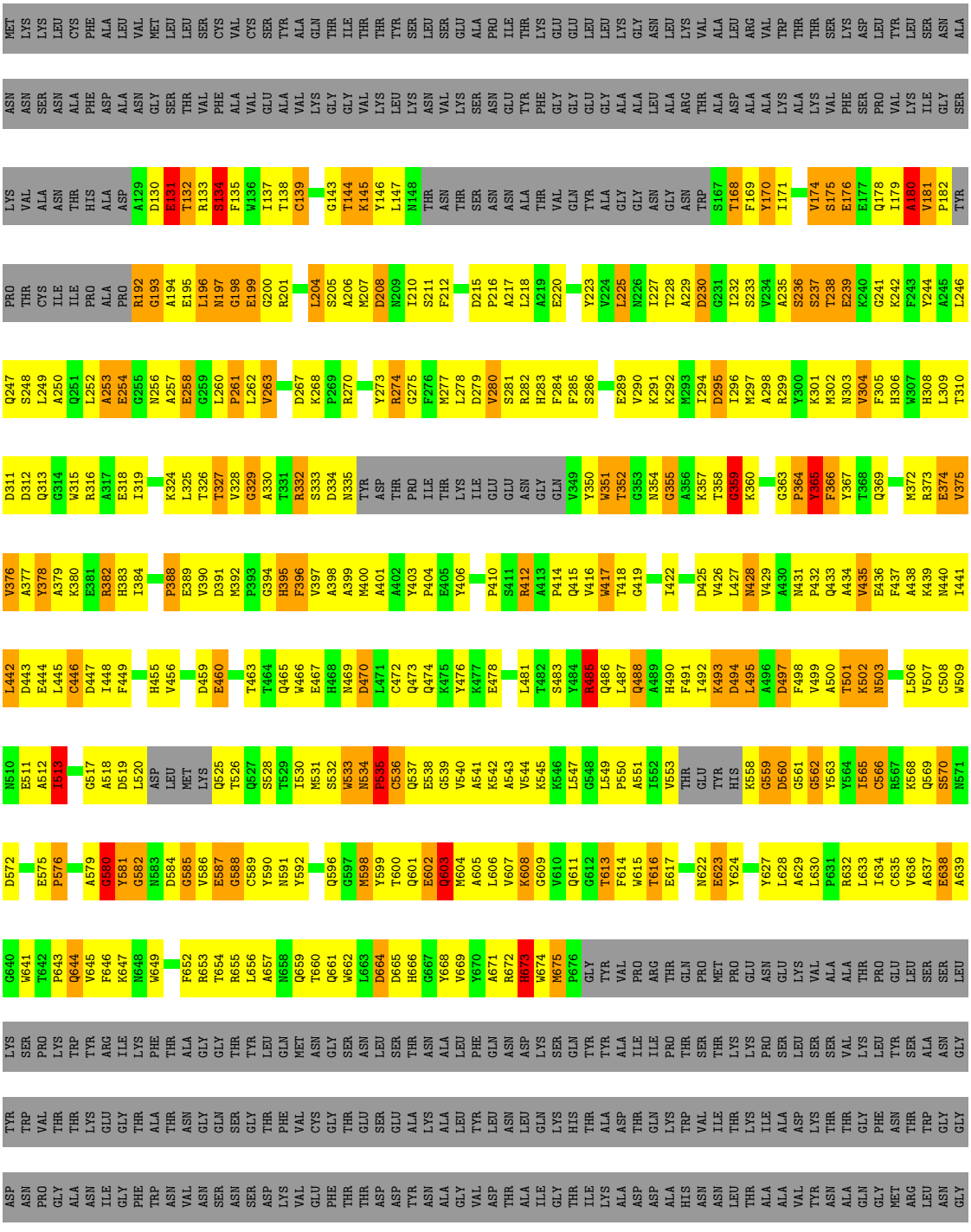




• Molecule 1: Beta-N-acetylhexosaminidase

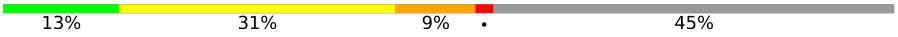
Chain C:

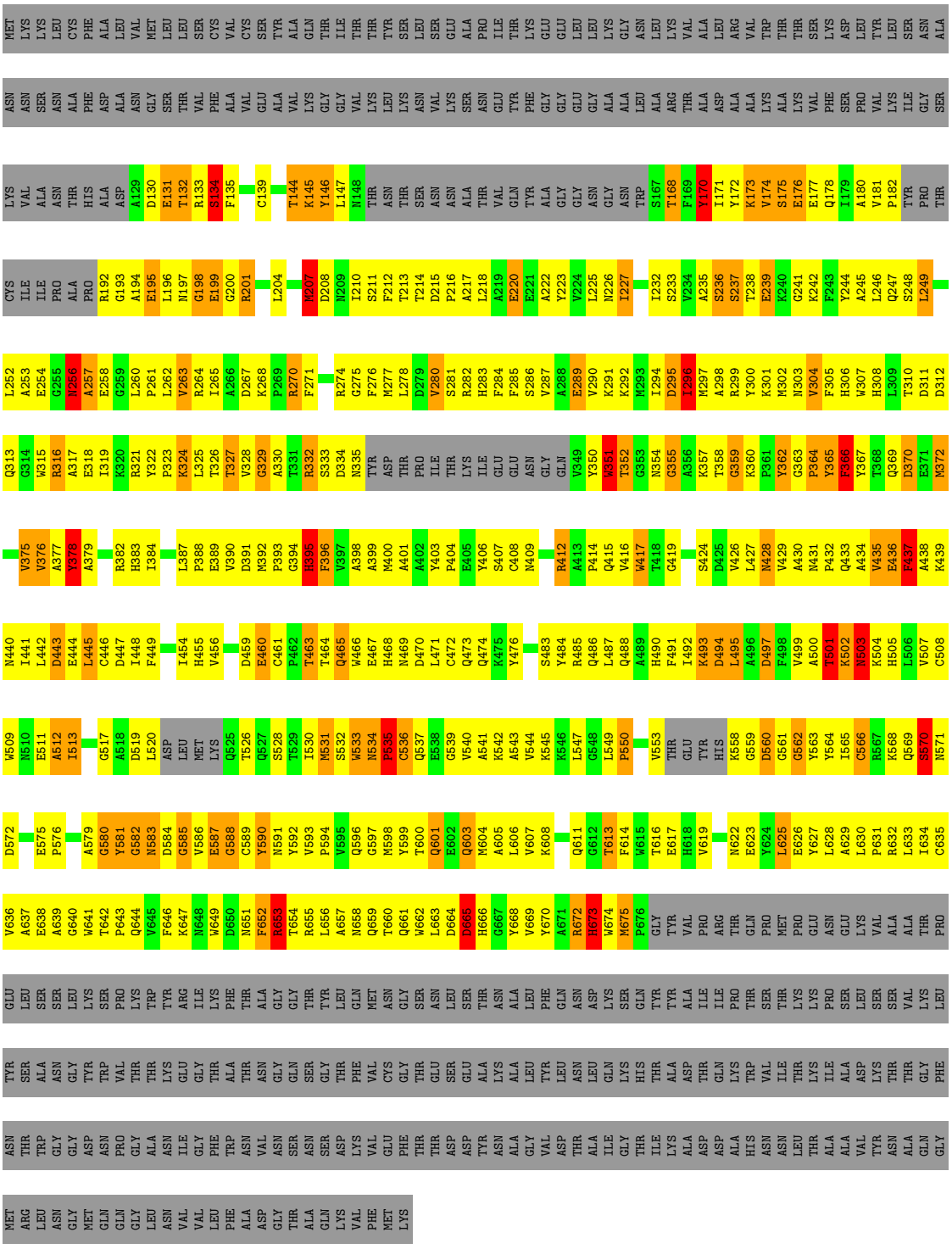




• Molecule 1: Beta-N-acetylhexosaminidase

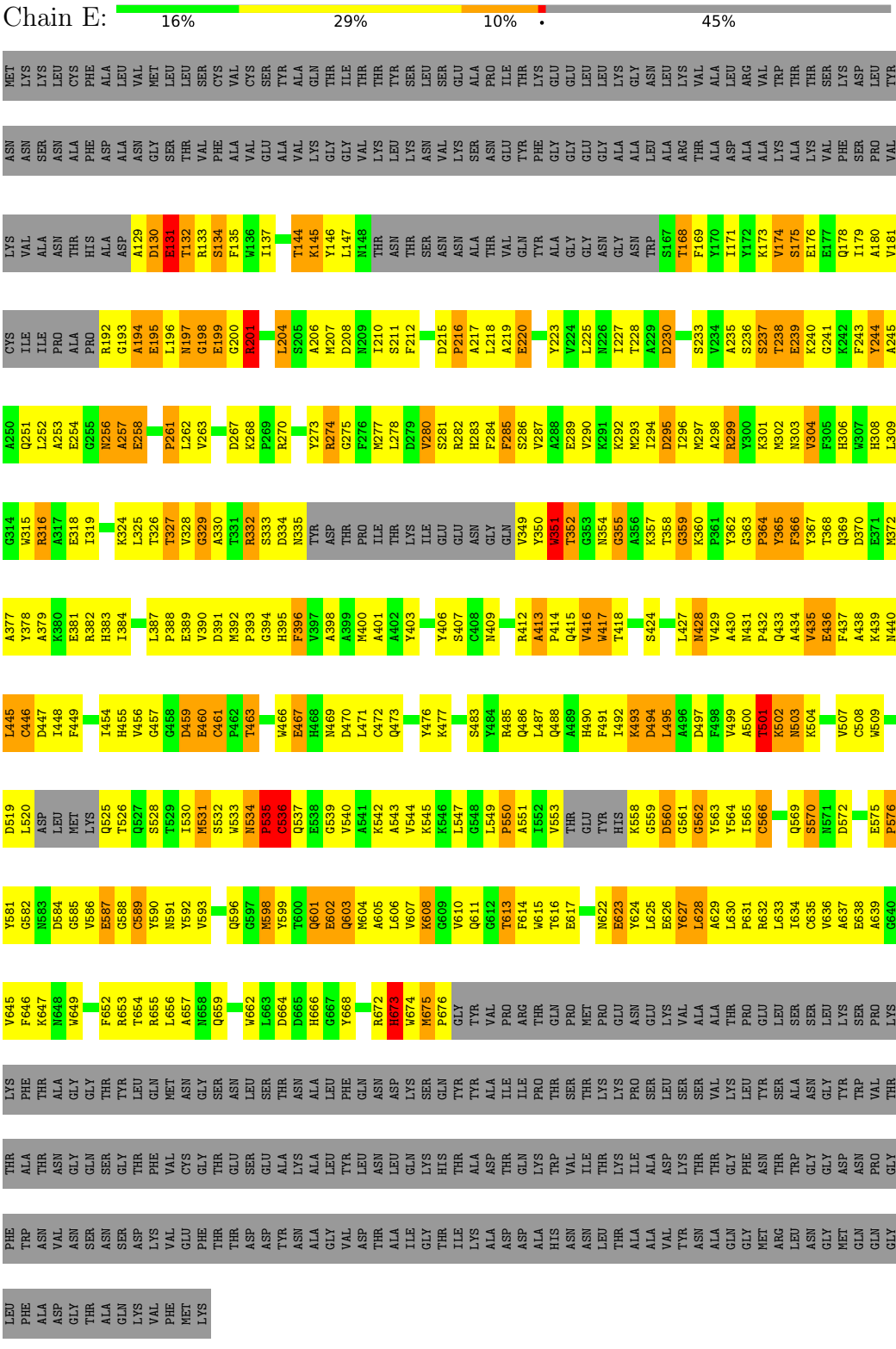
Chain D:





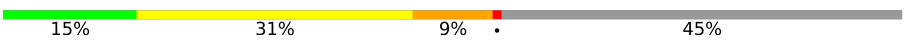
• Molecule 1: Beta-N-acetylhexosaminidase

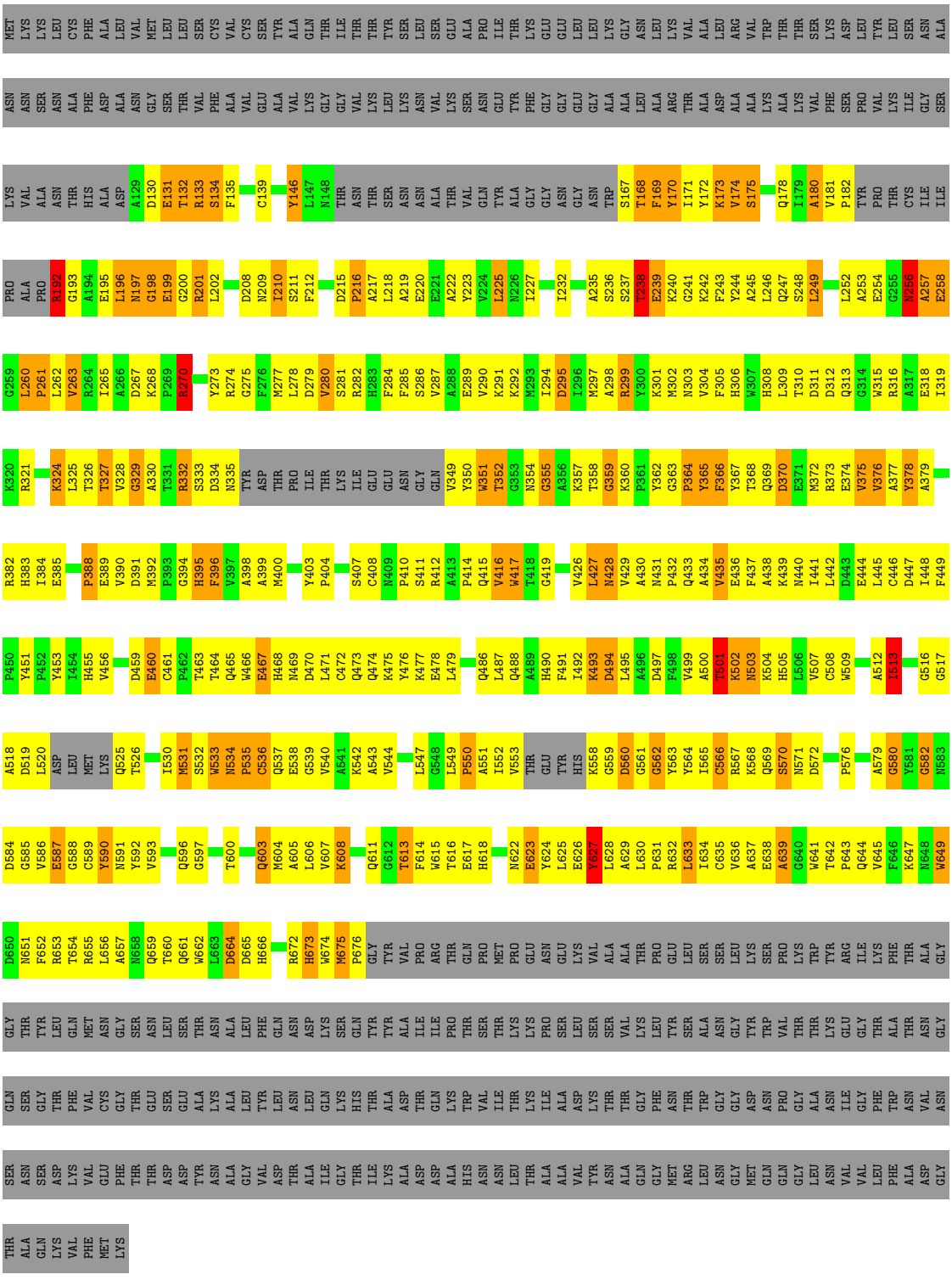
Chain E:



• Molecule 1: Beta-N-acetylhexosaminidase

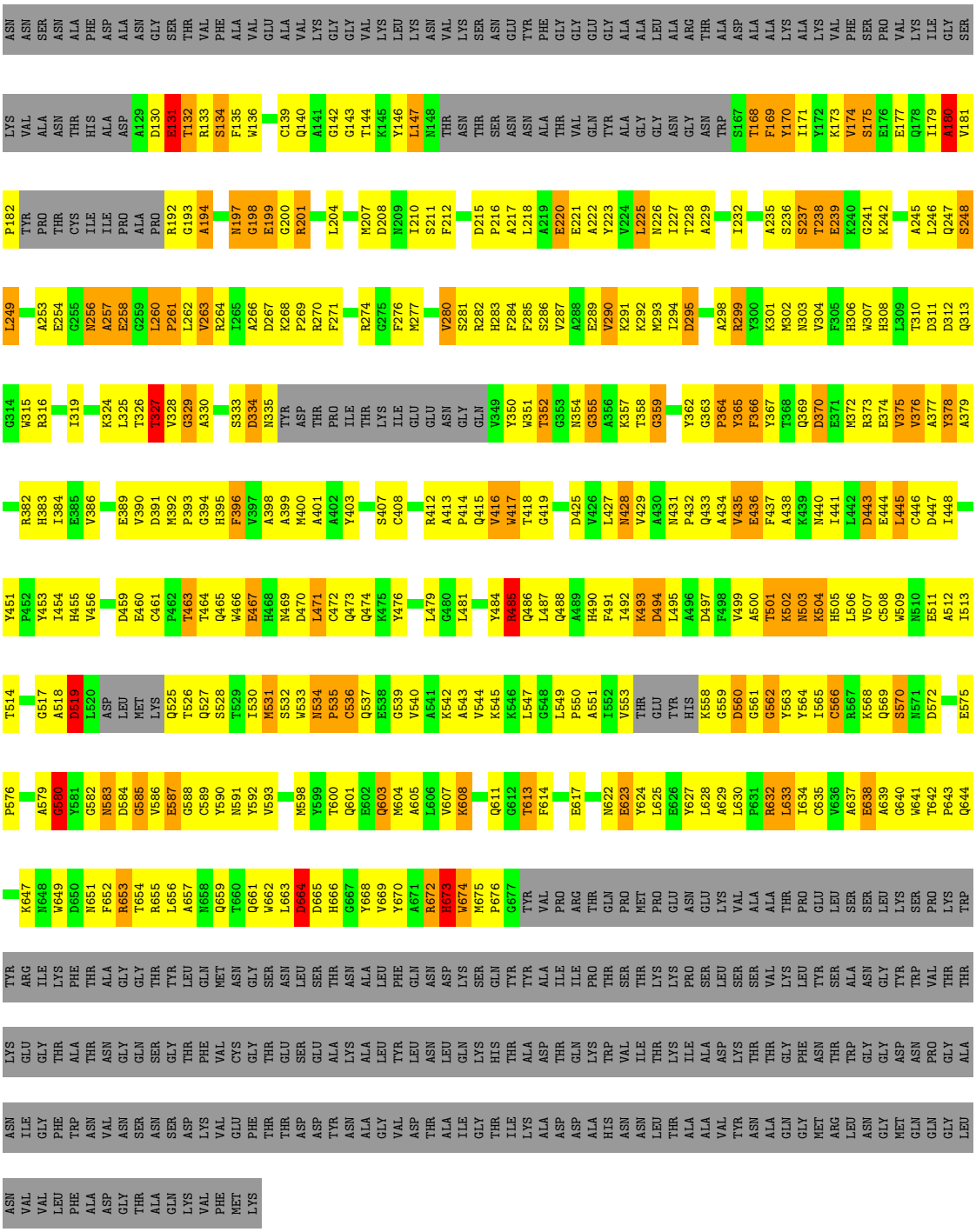
Chain F:





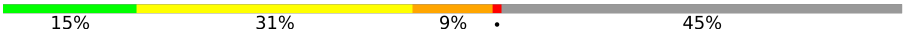
• Molecule 1: Beta-N-acetylhexosaminidase

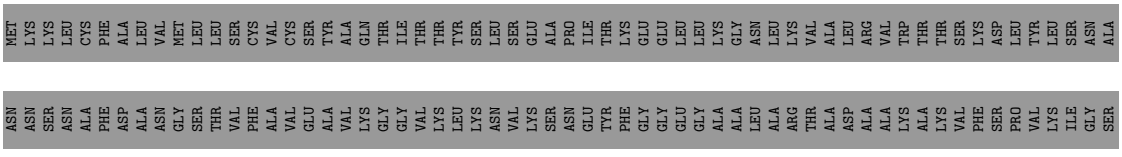
Chain H:

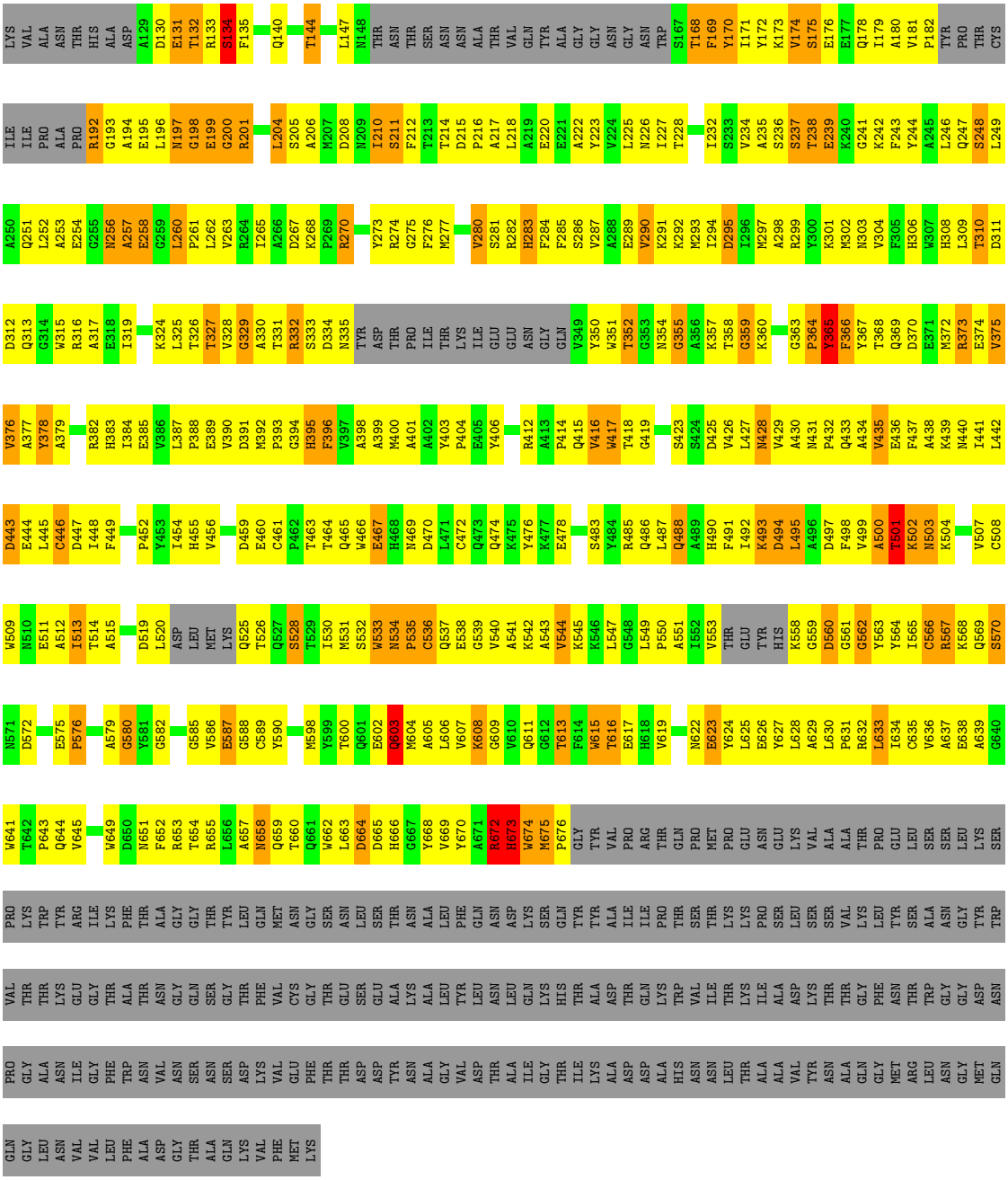


• Molecule 1: Beta-N-acetylhexosaminidase

Chain G:






LYS VAL ALA ASN THR HIS ALA ASP A129 D130 E131 T132 R133 S134 F135 Q140 T144 L147 N148 THR ASN THR SER ASN ASN ALA THR VAL GLN TYR ALA GLY GLY ASN GLY ASN TRP S167 T168 F169 Y170 I171 Y172 K173 V174 S175 E176 E177 Q178 I179 A180 V181 P182 TYR PRO THR CYS
ILE ILE PRO ALA PRO R192 G193 A194 E195 L196 N197 G198 E199 G200 R201 L204 S205 A206 M207 D208 N209 I210 S211 F212 T213 T214 D215 P216 A217 L218 A219 E220 E221 A222 Y223 V224 L225 N226 I227 T228 I232 S233 V234 A235 S236 S237 T238 E239 K240 G241 K242 F243 Y244 A245 L246 Q247 S248 L249
A250 Q251 L252 A253 E254 G255 N256 A257 E258 G259 L260 P261 L262 V263 R264 I265 A266 D267 K268 P269 R270 Y273 R274 G275 F276 M277 V280 S281 R282 H283 F284 F285 S286 V287 A288 E289 V290 K291 K292 M293 I294 D296 I296 M297 A298 R299 Y300 K301 M302 N303 V304 F305 H306 W307 H308 L309 T310 D311
D312 Q313 G314 W315 R316 A317 E318 I319 K324 L325 T326 T327 V328 G329 A330 T331 R332 S333 D334 N335 TYR ASP THR PRO ILE THR LYS ILE GLU GLU ASN GLY GLN V349 Y350 W351 T352 G353 N354 G355 A356 K357 T358 G359 K360 G363 P364 Y365 F366 Y367 T368 Q369 D370 E371 M372 R373 E374 V375
V376 A377 Y378 A379 R382 H383 I384 E385 V386 L387 P388 E389 V390 D391 M392 P393 G394 H395 F396 V397 A398 A399 M400 A401 A402 Y403 P404 E405 Y406 R412 A413 P414 Q415 Y416 W417 T418 G419 S423 S424 D425 V426 L427 N428 V429 A430 N431 P432 Q433 A434 V435 E436 F437 A438 K439 N440 I441 L442
D443 E444 L445 C446 D447 I448 F449 P452 Y453 I454 H455 V456 D459 E460 C461 P462 T463 T464 Q465 W466 E467 H468 N469 D470 L471 C472 Q473 Q474 K475 Y476 K477 E478 S483 Y484 R485 Q486 L487 Q488 A489 H490 F491 I492 K493 D494 L495 A496 D497 F498 V499 A500 T501 K502 N503 K504 V507 C508
W509 N510 E511 A512 I513 T514 A515 D519 L520 ASP LEU MET LYS Q525 T526 Q527 S528 T529 I530 M531 S532 W533 N534 P535 C536 Q537 E538 G539 V540 A541 K542 A543 V544 K545 K546 L547 G548 L549 P550 A551 I552 V553 THR GLU TYR HIS K558 G559 D560 G561 G562 Y563 Y564 I565 C566 R567 K568 Q569 S570
N571 D572 E575 P576 A579 G580 Y581 G582 G585 V586 E587 G588 C589 Y590 M598 Y599 T600 Q601 E602 Q603 M604 A605 L606 V607 K608 G609 V610 Q611 G612 T613 F614 W615 T616 E617 H618 V619 N622 E623 Y624 L625 E626 Y627 L628 A629 L630 P631 R632 L633 I634 C635 V636 A637 E638 A639 G640
W641 T642 P643 Q644 V645 W649 D650 N651 F652 R653 T654 R655 L656 A657 N658 Q659 T660 Q661 W662 L663 D664 D665 H666 G667 Y668 V669 Y670 A671 R672 H673 W674 M675 P676 GLY TYR VAL PRO ARG THR GLN PRO MET PRO GLU ASN GLU LYS VAL ALA ALA THR PRO GLU LEU SER SER LEU LYS SER
PRO LYS TRP TYR ARG ILE LYS PHE THR ALA ALA GLY GLY THR TYR LEU GLN MET ASN GLY SER ASN LEU SER THR ASN ALA LEU PHE GLN ASN ASP LYS SER GLN TYR TYR ALA ILE ILE PRO THR SER THR LYS LYS PRO SER LEU SER SER VAL LYS LEU TYR SER ALA ASN GLY TYR TRP
VAL THR THR LYS GLU GLY THR ALA THR ASN GLY GLN SER GLY THR PHE VAL CYS GLY THR GLU SER GLU ALA LYS ALA LEU TYR LEU ASN LEU GLN LYS HIS THR ILE ALA ASP THR GLN LYS TRP VAL ILE THR LYS ILE ALA ASP THR LYS THR THR TRP GLY GLY ASP ASN
PRO GLY ALA ASN ILE GLY PHE TRP ASN VAL ASN SER ASN SER ASP LYS VAL GLU PHE THR THR ASP ASP TYR ASN ALA GLY VAL ASP THR ALA ASP ASP ALA HIS ASN ASN LEU THR ALA ALA ALA VAL TYR ASN ALA GLN GLY MET ARG LEU ASN GLY MET GLN
GLN GLY LEU ASN VAL VAL LEU PHE ALA ASP GLY THR ALA GLN LYS VAL PHE MET LYS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	132.95Å 133.42Å 225.37Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	94.18 – 2.71 94.18 – 2.71	Depositor EDS
% Data completeness (in resolution range)	96.6 (94.18-2.71) 99.2 (94.18-2.71)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.02 (at 2.73Å)	Xtriage
Refinement program	REFMAC 5.8.0352	Depositor
R, R_{free}	0.283 , 0.319 0.333 , 0.357	Depositor DCC
R_{free} test set	21126 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å^2)	47.0	Xtriage
Anisotropy	0.089	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.25 , 39.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.000 for k,-h,l 0.000 for -k,h,l 0.055 for h,-k,-l 0.054 for -h,k,-l 0.040 for -h,-k,l 0.000 for k,h,-l 0.000 for -k,-h,-l	Xtriage
Reported twinning fraction	0.182 for H, K, L 0.177 for h,-k,-l 0.179 for -H, K, -L 0.177 for -h,-k,l 0.071 for K, H, -L 0.071 for k,-h,l 0.070 for -K, H, L 0.072 for -K, -H, -L	Depositor
Outliers	0 of 416904 reflections	Xtriage
F_o,F_c correlation	0.74	EDS
Total number of atoms	31772	wwPDB-VP
Average B, all atoms (Å^2)	26.0	wwPDB-VP

[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.

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.27% of the height of the origin peak. No significant pseudotranslation is detected.*

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	0.56	0/4169	1.11	13/5666 (0.2%)
1	B	0.48	0/4062	0.93	1/5514 (0.0%)
1	C	0.49	1/4058 (0.0%)	0.93	3/5509 (0.1%)
1	D	0.47	0/4058	0.90	2/5509 (0.0%)
1	E	0.48	0/4058	0.93	2/5509 (0.0%)
1	F	0.47	0/4058	0.92	2/5509 (0.0%)
1	G	0.47	0/4058	0.88	4/5509 (0.1%)
1	H	0.49	0/4062	0.94	5/5514 (0.1%)
All	All	0.49	1/32583 (0.0%)	0.94	32/44239 (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	33
1	B	0	13
1	C	0	14
1	D	0	18
1	E	0	13
1	F	0	12
1	G	0	11
1	H	0	16
All	All	0	130

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	131	GLU	CD-OE2	-5.24	1.19	1.25

The worst 5 of 32 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	G	658	ASN	CB-CA-C	-8.04	94.31	110.40
1	A	279	ASP	CB-CG-OD2	-7.27	111.75	118.30
1	F	627	TYR	CB-CG-CD1	7.12	125.27	121.00
1	F	627	TYR	CB-CG-CD2	-6.62	117.03	121.00
1	A	132	THR	N-CA-C	-6.31	93.95	111.00

There are no chirality outliers.

5 of 130 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	130	ASP	Peptide
1	A	132	THR	Mainchain
1	A	148	ASN	Peptide
1	A	150	ASN	Peptide
1	A	152	SER	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	4061	0	3877	681	0
1	B	3959	0	3785	454	0
1	C	3955	0	3782	463	0
1	D	3955	0	3782	460	0
1	E	3955	0	3782	453	0
1	F	3955	0	3782	478	0
1	G	3955	0	3782	434	0
1	H	3959	0	3785	469	0
2	A	10	0	0	2	0
2	B	4	0	0	2	0
2	C	4	0	0	0	0
All	All	31772	0	30357	3847	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 62.

The worst 5 of 3847 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:C:295:ASP:HA	1:C:382:ARG:NH1	1.39	1.35
1:F:537:GLN:N	1:H:192:ARG:HG2	1.37	1.34
1:A:528:SER:O	1:A:549:LEU:HD22	1.31	1.29
1:A:631:PRO:O	1:A:634:ILE:HD12	1.19	1.27
1:B:131:GLU:HG3	1:B:659:GLN:OE1	1.36	1.23

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	505/901 (56%)	297 (59%)	127 (25%)	81 (16%)	0	0
1	B	489/901 (54%)	301 (62%)	126 (26%)	62 (13%)	0	0
1	C	488/901 (54%)	295 (60%)	123 (25%)	70 (14%)	0	0
1	D	488/901 (54%)	292 (60%)	133 (27%)	63 (13%)	0	0
1	E	488/901 (54%)	294 (60%)	130 (27%)	64 (13%)	0	0
1	F	488/901 (54%)	288 (59%)	137 (28%)	63 (13%)	0	0
1	G	488/901 (54%)	300 (62%)	127 (26%)	61 (12%)	0	0
1	H	489/901 (54%)	293 (60%)	134 (27%)	62 (13%)	0	0
All	All	3923/7208 (54%)	2360 (60%)	1037 (26%)	526 (13%)	0	0

5 of 526 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	149	THR
1	A	150	ASN
1	A	151	THR
1	A	190	ALA
1	A	194	ALA

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	420/737 (57%)	354 (84%)	66 (16%)	2 6
1	B	410/737 (56%)	363 (88%)	47 (12%)	5 12
1	C	410/737 (56%)	358 (87%)	52 (13%)	4 9
1	D	410/737 (56%)	352 (86%)	58 (14%)	3 7
1	E	410/737 (56%)	361 (88%)	49 (12%)	5 11
1	F	410/737 (56%)	363 (88%)	47 (12%)	5 12
1	G	410/737 (56%)	364 (89%)	46 (11%)	6 13
1	H	410/737 (56%)	364 (89%)	46 (11%)	6 13
All	All	3290/5896 (56%)	2879 (88%)	411 (12%)	4 10

5 of 411 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	207	MET
1	F	258	GLU
1	G	501	THR
1	E	238	THR
1	E	531	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 65 such sidechains are listed below:

Mol	Chain	Res	Type
1	H	644	GLN
1	G	209	ASN
1	D	569	GLN
1	D	465	GLN
1	G	283	HIS

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)

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.