wwPDB X-ray Structure Validation Summary Report (i)

Jun 12, 2024 – 07:03 PM EDT

PDB ID	:	3LJ5
Title	:	Full Length Bacteriophage P22 Portal Protein
Authors	:	Olia, A.S.; Cingolani, G.
Deposited on	:	2010-01-25
Resolution	:	7.50 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 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:

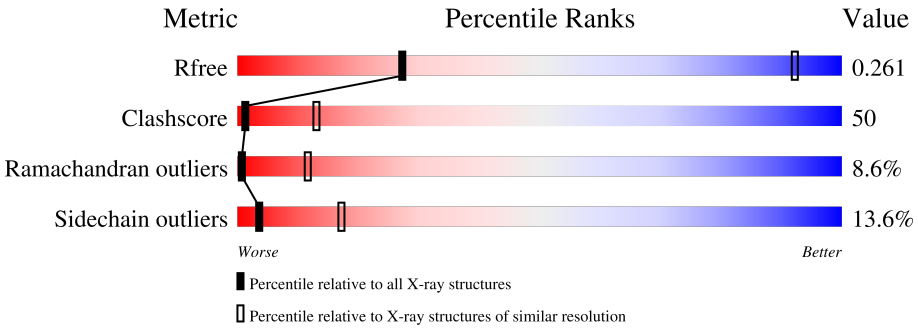
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.20.1
EDS	:	2.36.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 7.50 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	1004 (10.00-3.90)
Clashscore	141614	1069 (10.00-3.90)
Ramachandran outliers	138981	1002 (10.00-3.90)
Sidechain outliers	138945	1002 (10.00-3.86)

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	725	38% 45% 12% • 5%
1	B	725	39% 45% 12% • 5%
1	C	725	39% 44% 12% • 5%
1	D	725	38% 45% 12% • 5%
1	E	725	37% 45% 12% • 5%
1	F	725	39% 44% 12% • 5%
1	G	725	37% 47% 12% 5%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	H	725	38% 45% 12% 5%
1	I	725	39% 44% 12% 5%
1	J	725	39% 44% 12% 5%
1	K	725	39% 43% 12% 5%
1	L	725	40% 43% 12% 5%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 61512 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 Portal protein.

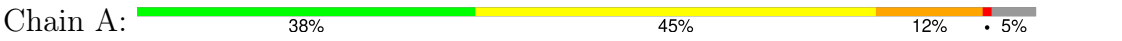
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	692	Total	C	N	O	S	0	0	0
			5126	3204	900	1002	20			
1	B	692	Total	C	N	O	S	0	0	0
			5126	3204	900	1002	20			
1	C	692	Total	C	N	O	S	0	0	0
			5126	3204	900	1002	20			
1	D	692	Total	C	N	O	S	0	0	0
			5126	3204	900	1002	20			
1	E	692	Total	C	N	O	S	0	0	0
			5126	3204	900	1002	20			
1	F	692	Total	C	N	O	S	0	0	0
			5126	3204	900	1002	20			
1	G	692	Total	C	N	O	S	0	0	0
			5126	3204	900	1002	20			
1	H	692	Total	C	N	O	S	0	0	0
			5126	3204	900	1002	20			
1	I	692	Total	C	N	O	S	0	0	0
			5126	3204	900	1002	20			
1	J	692	Total	C	N	O	S	0	0	0
			5126	3204	900	1002	20			
1	K	692	Total	C	N	O	S	0	0	0
			5126	3204	900	1002	20			
1	L	692	Total	C	N	O	S	0	0	0
			5126	3204	900	1002	20			

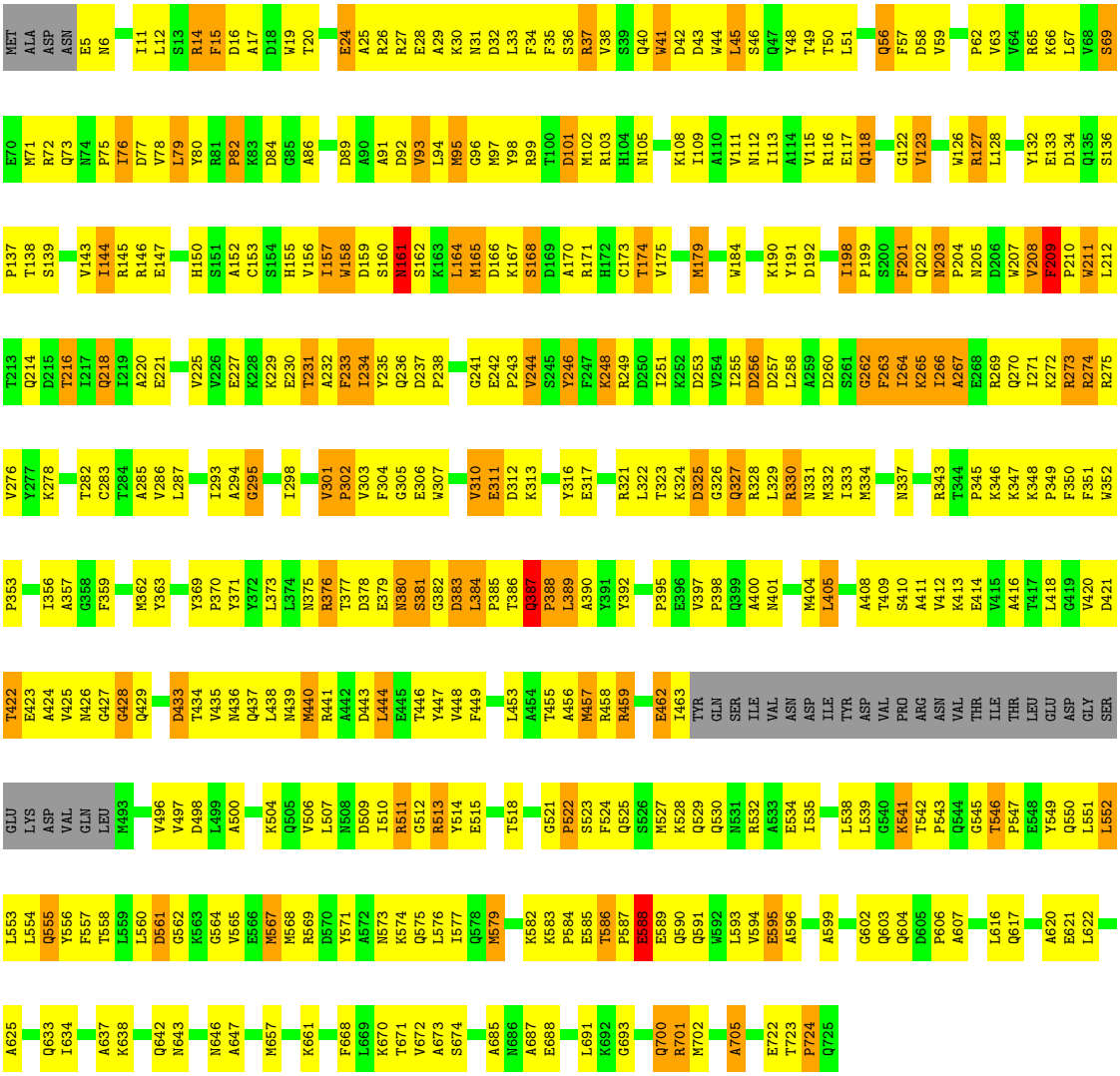
3 Residue-property plots (i)

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: Portal protein

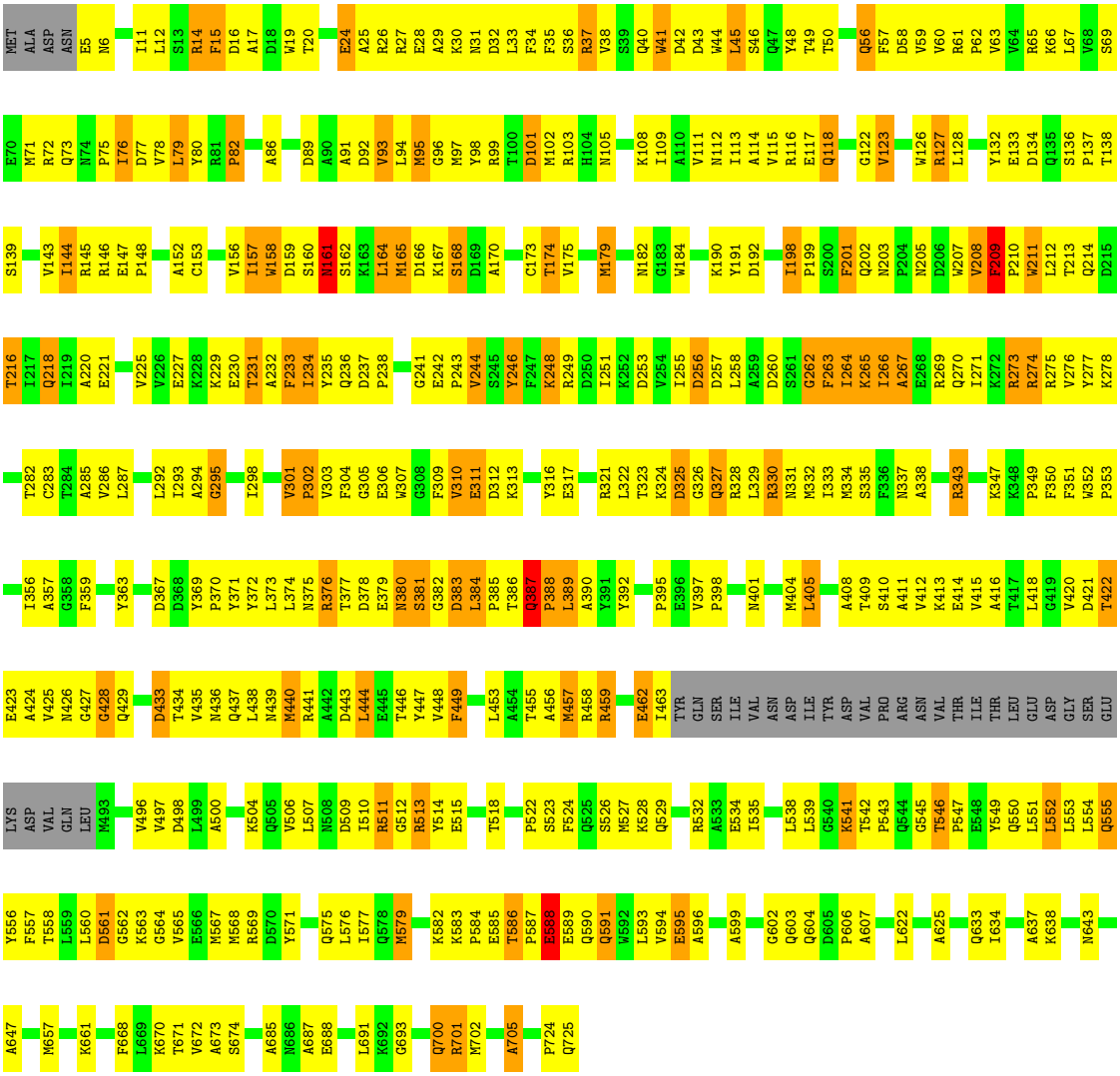
Chain A:





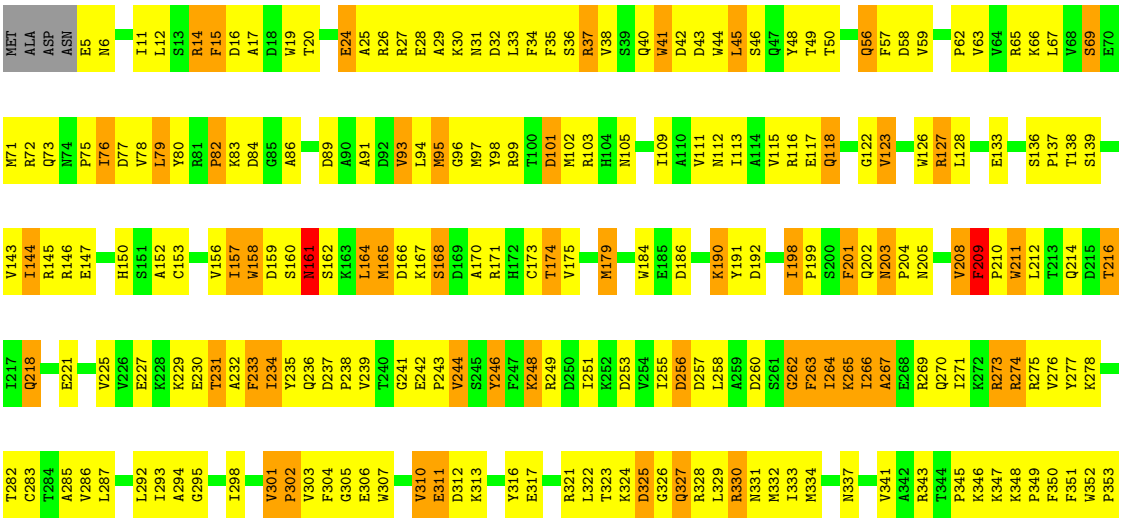
• Molecule 1: Portal protein

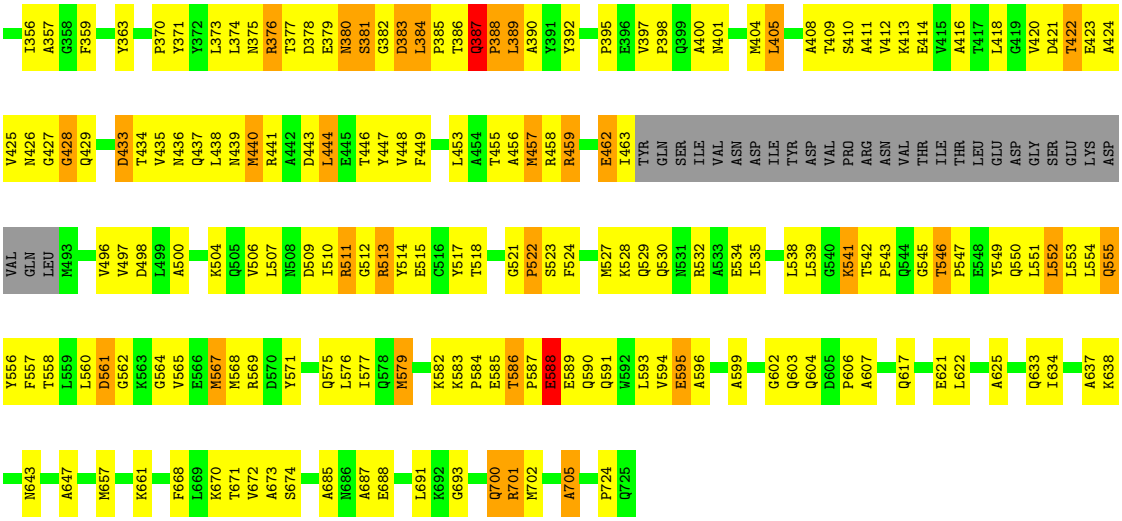
Chain B:



• Molecule 1: Portal protein

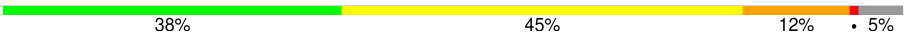
Chain C:

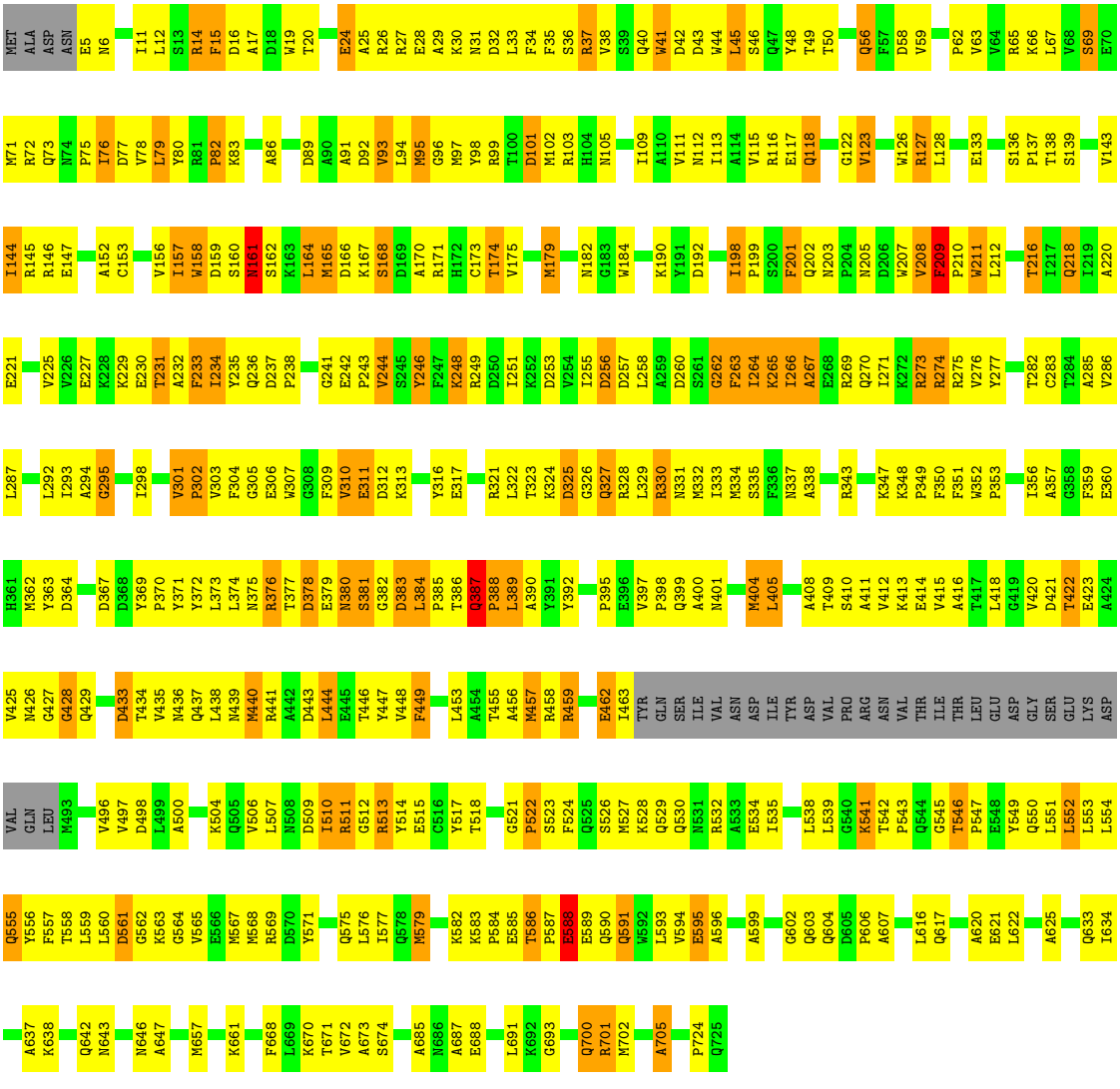




• Molecule 1: Portal protein

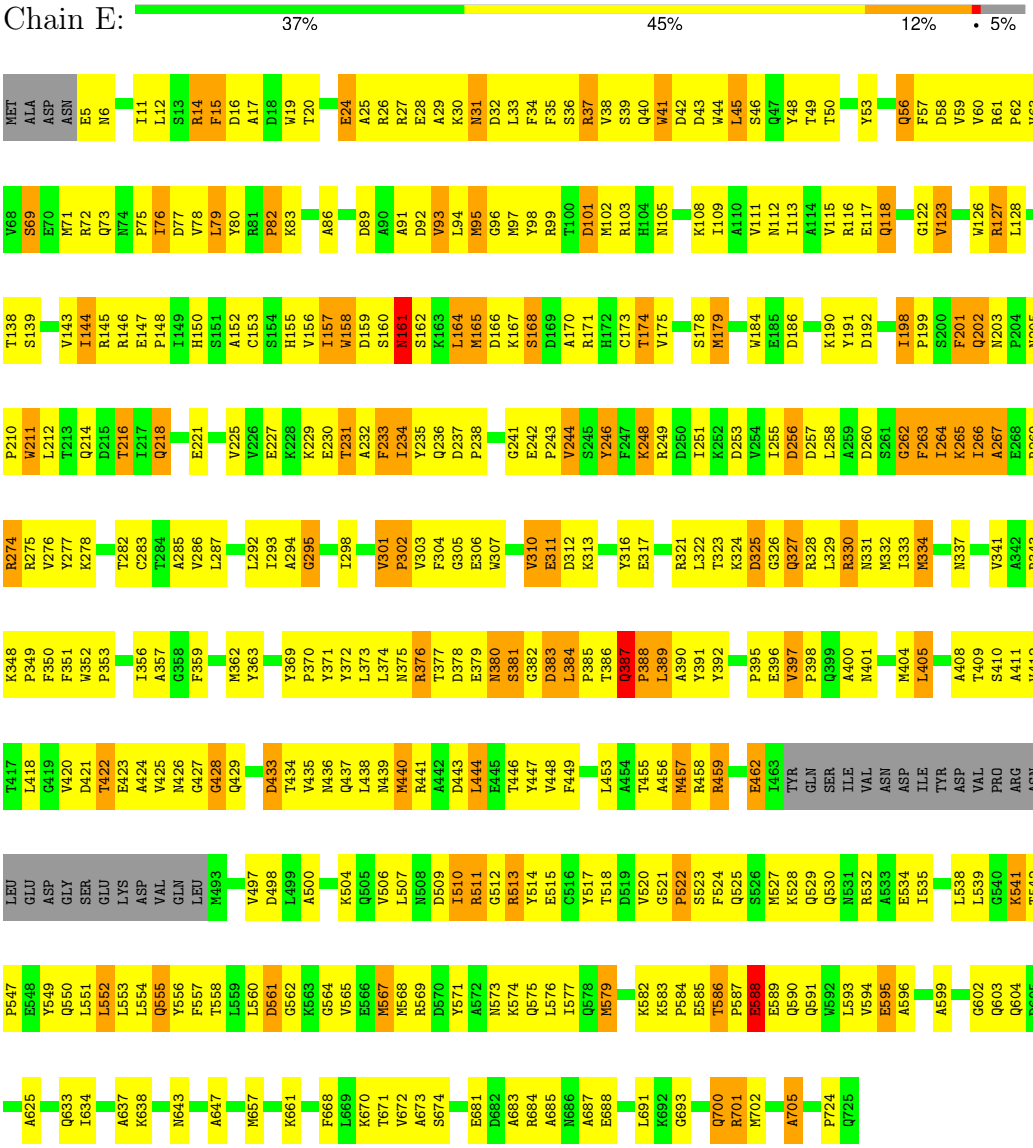
Chain D:





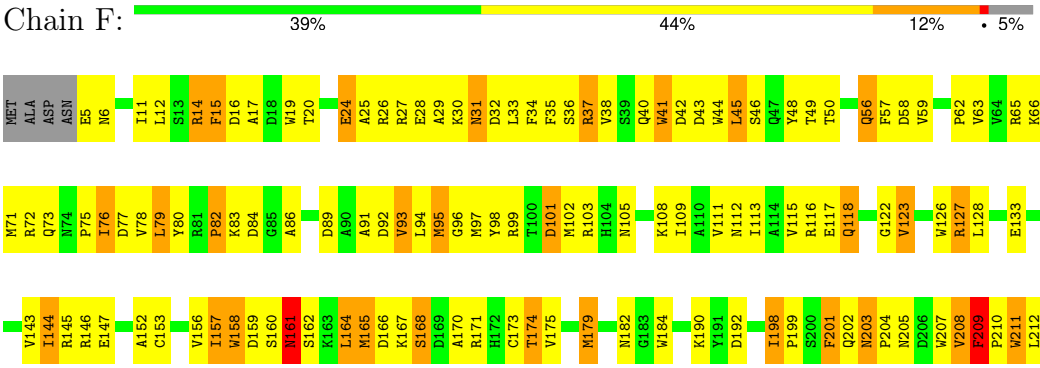
• Molecule 1: Portal protein

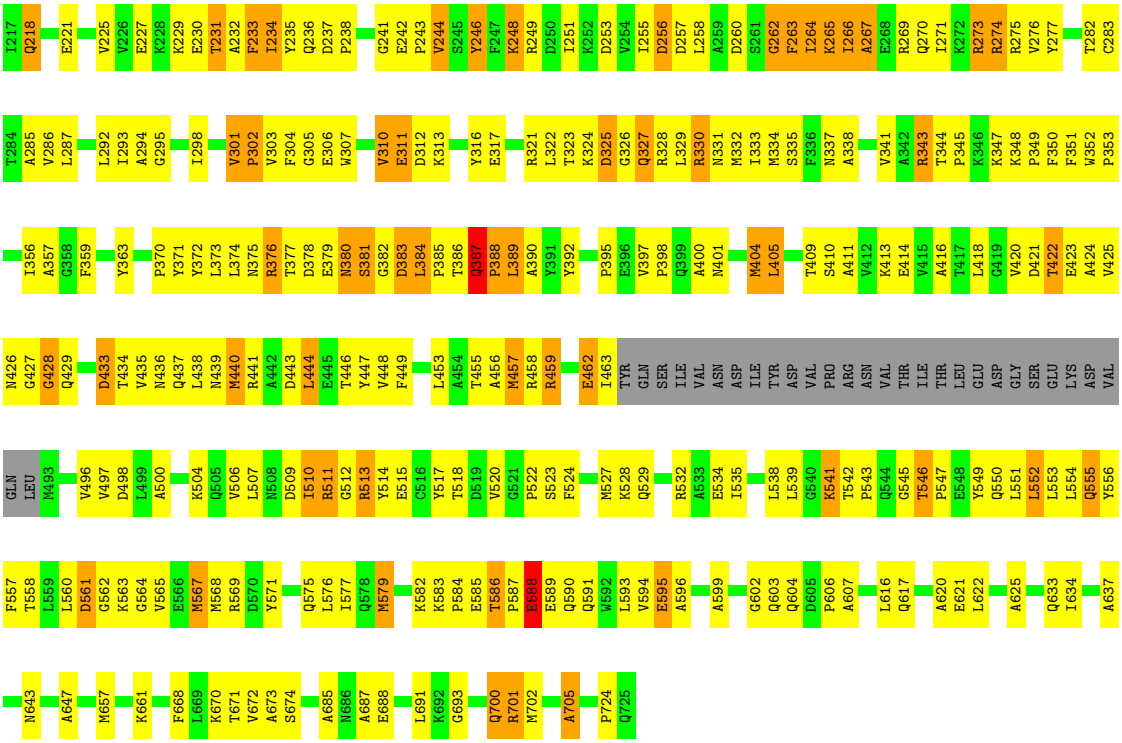
Chain E:



• Molecule 1: Portal protein

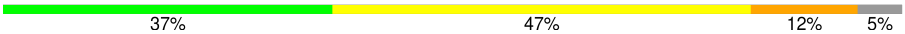
Chain F:

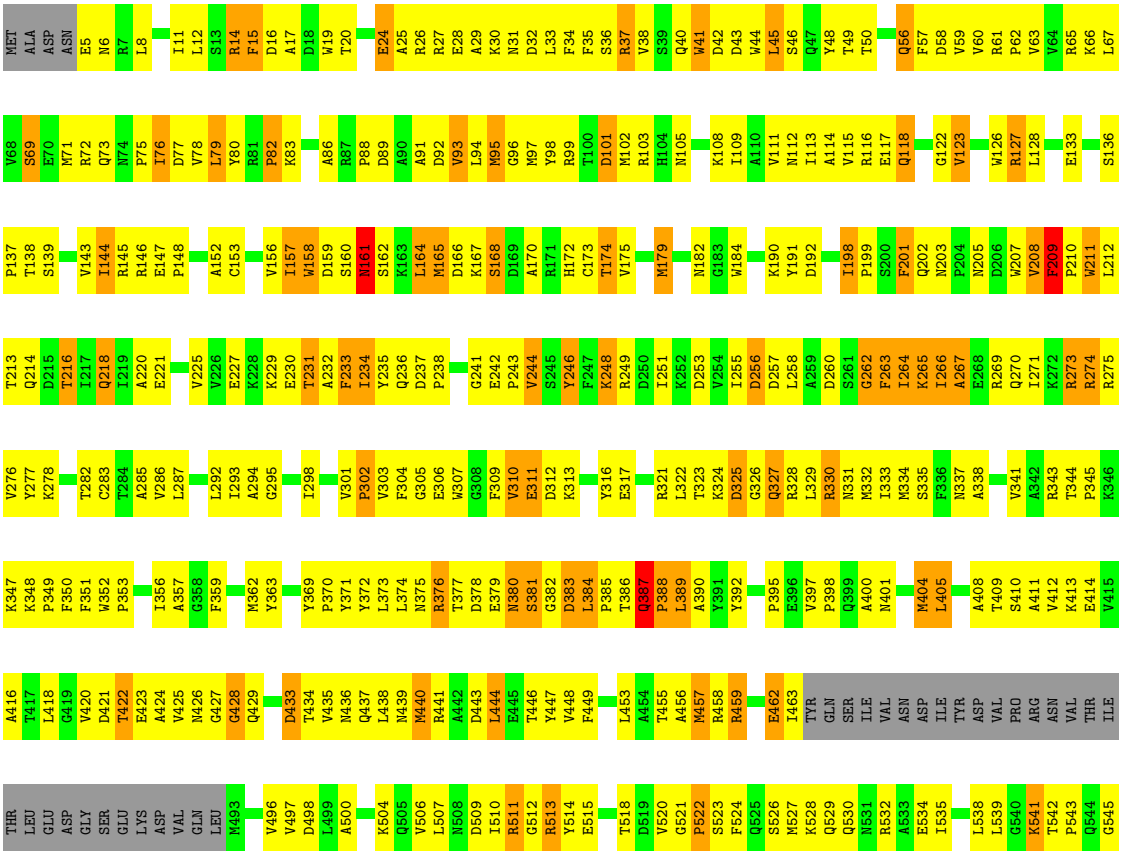


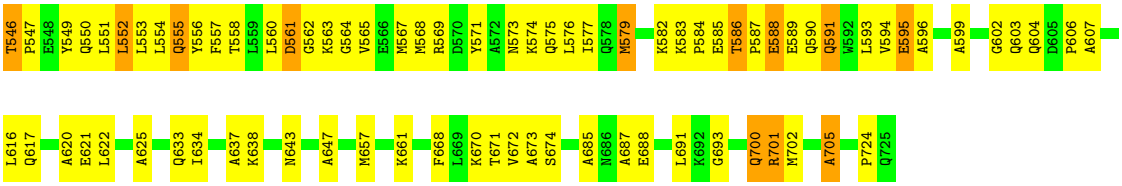


• Molecule 1: Portal protein

Chain G:

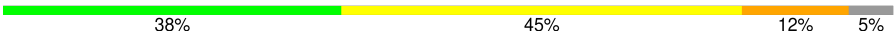


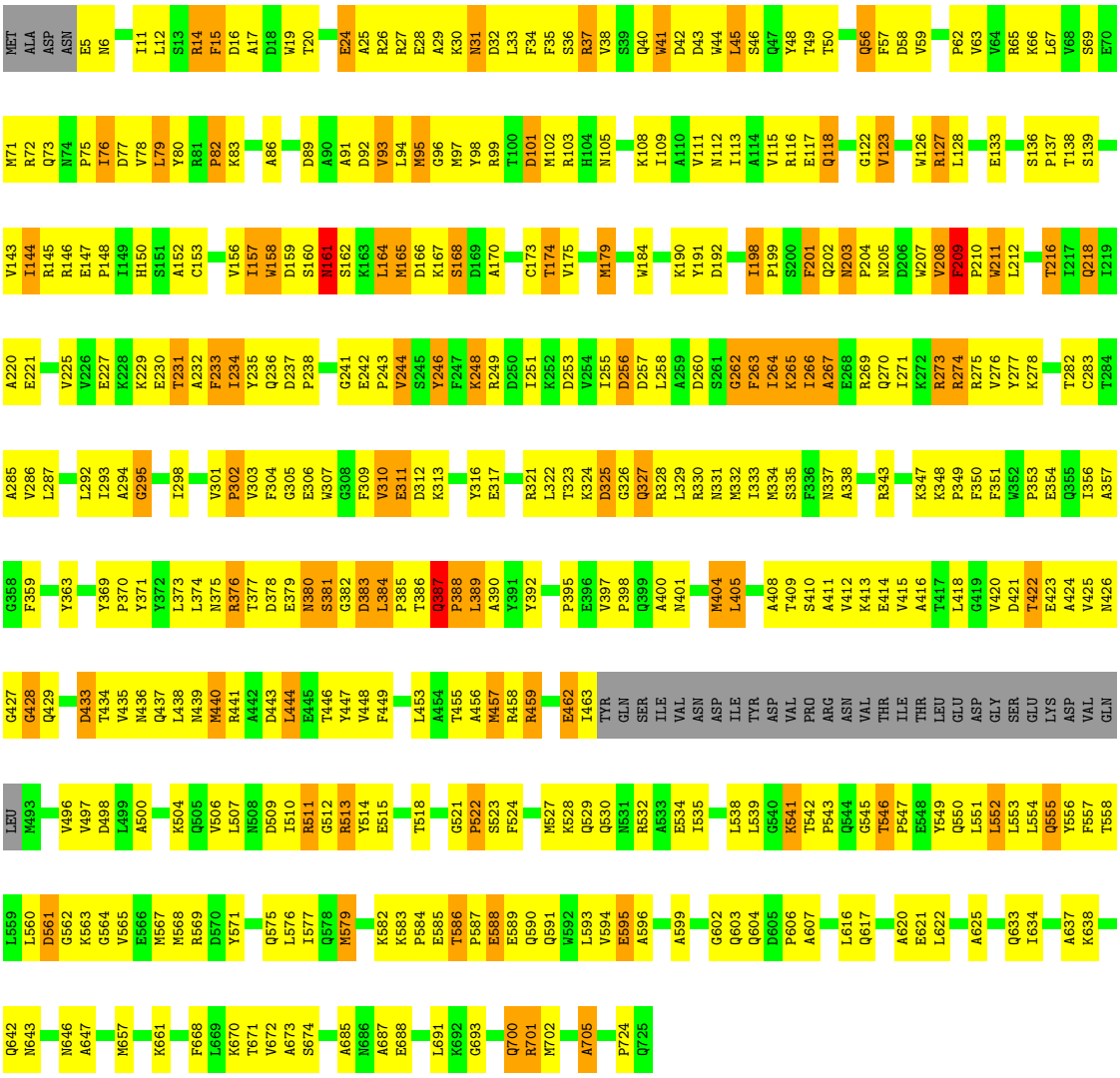




• Molecule 1: Portal protein

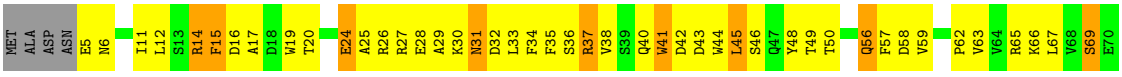
Chain H:

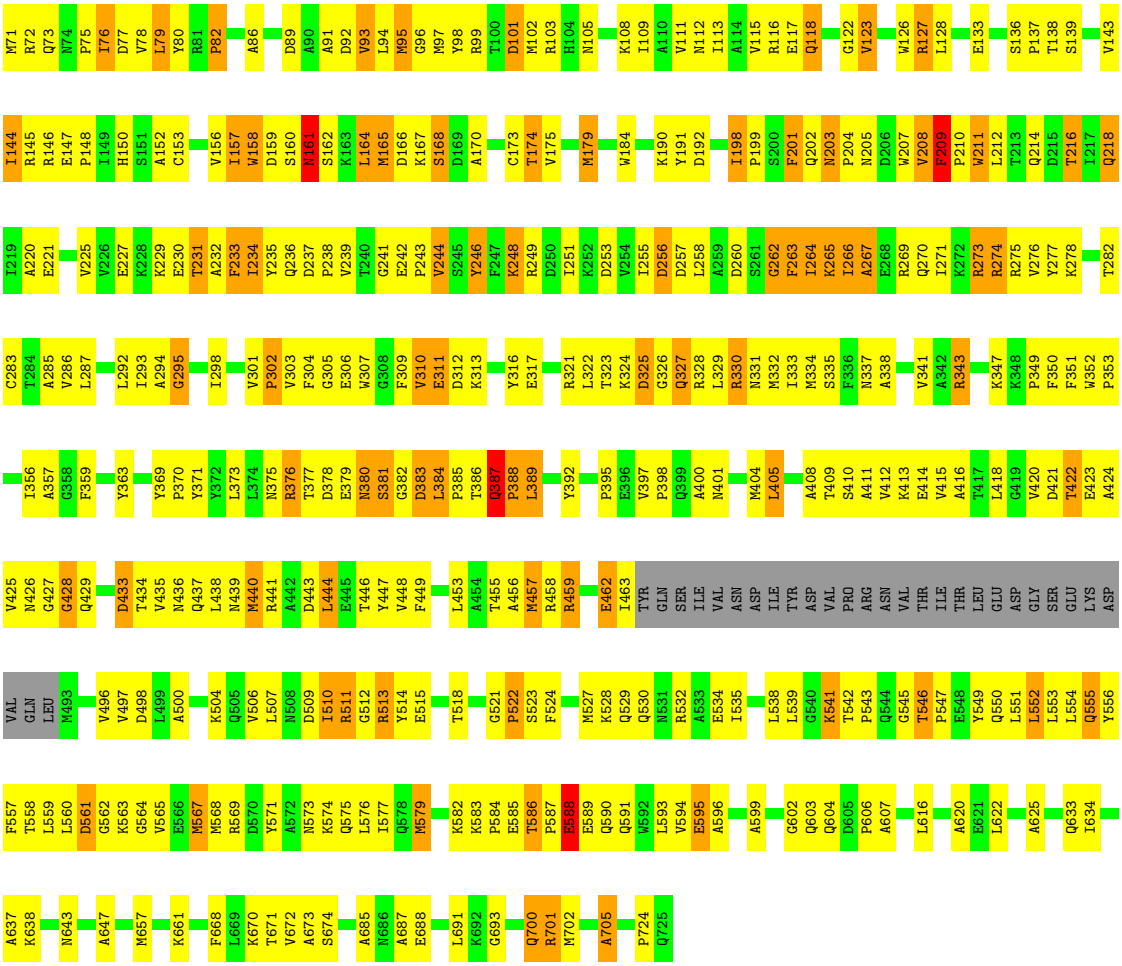




• Molecule 1: Portal protein

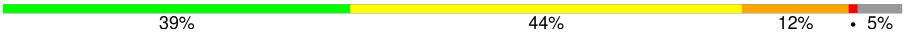
Chain I:

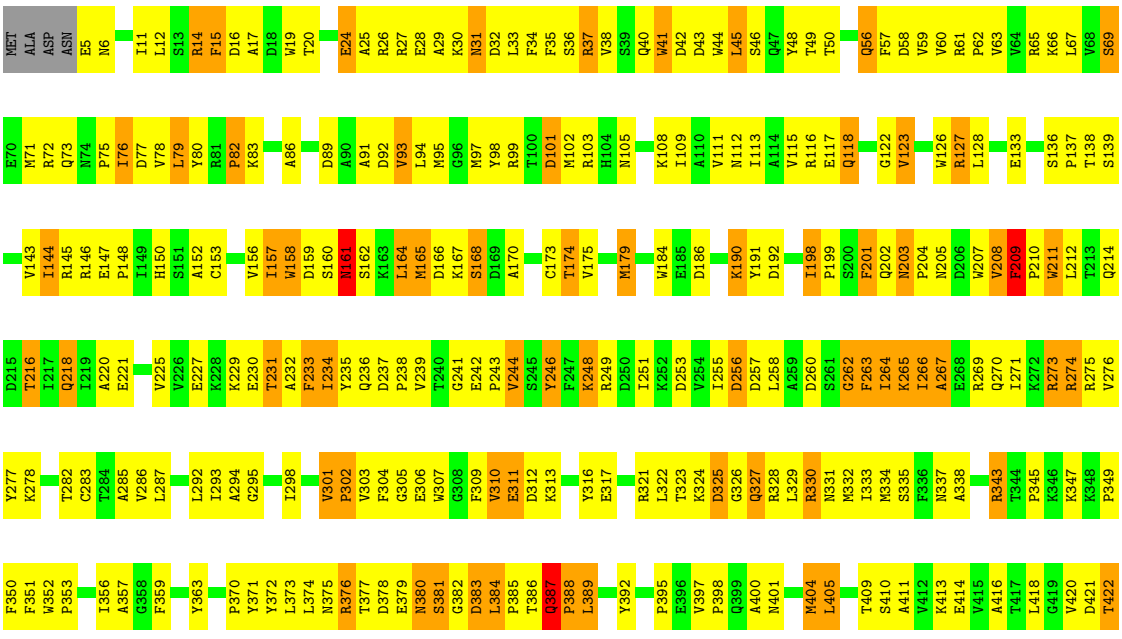


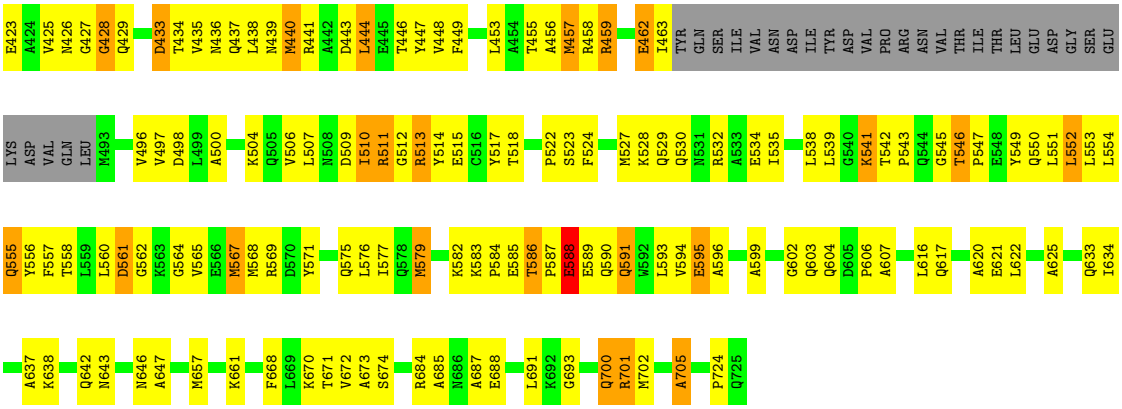


• Molecule 1: Portal protein

Chain J:

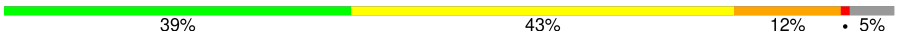


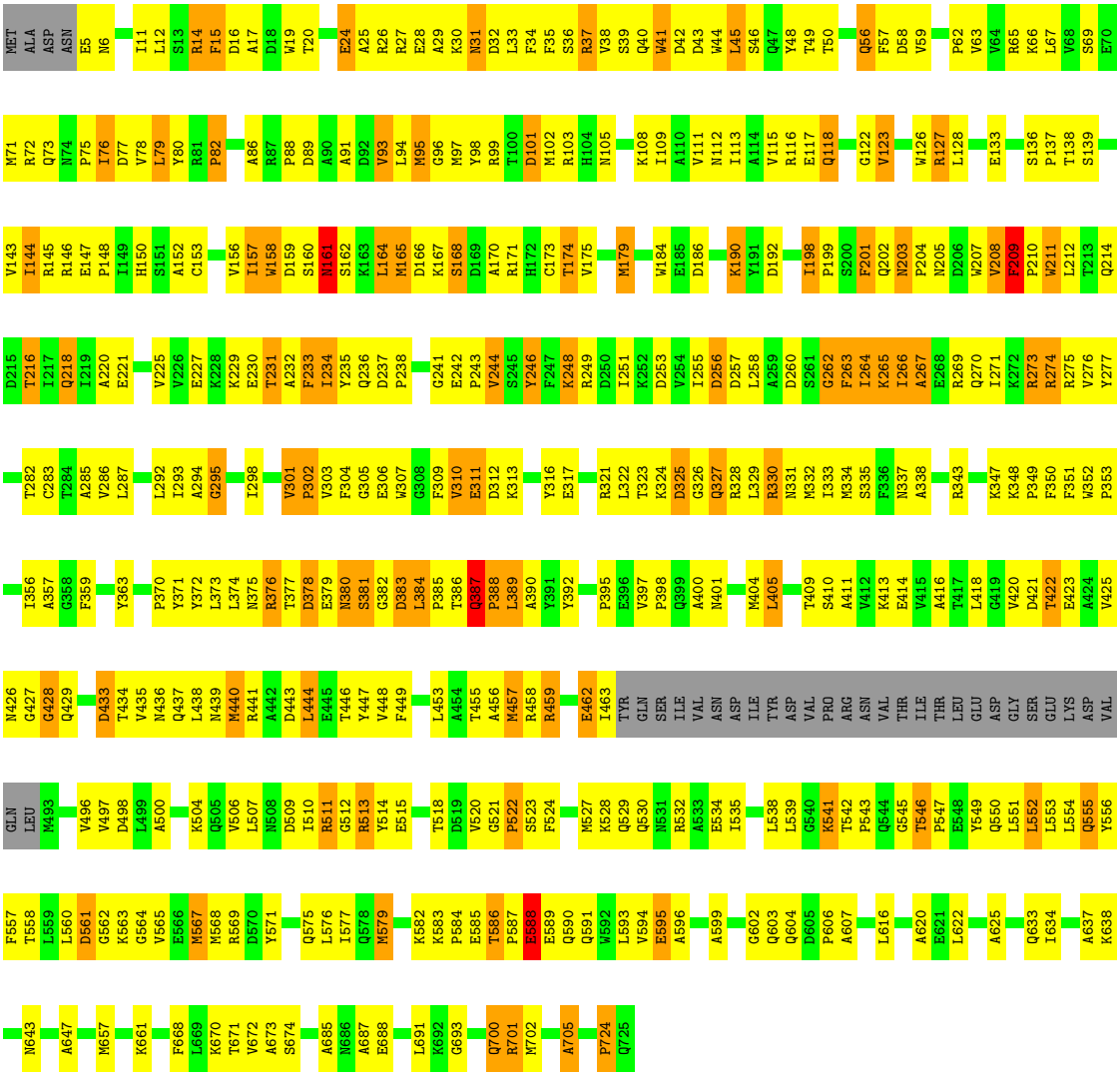




• Molecule 1: Portal protein

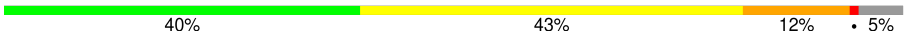
Chain K:

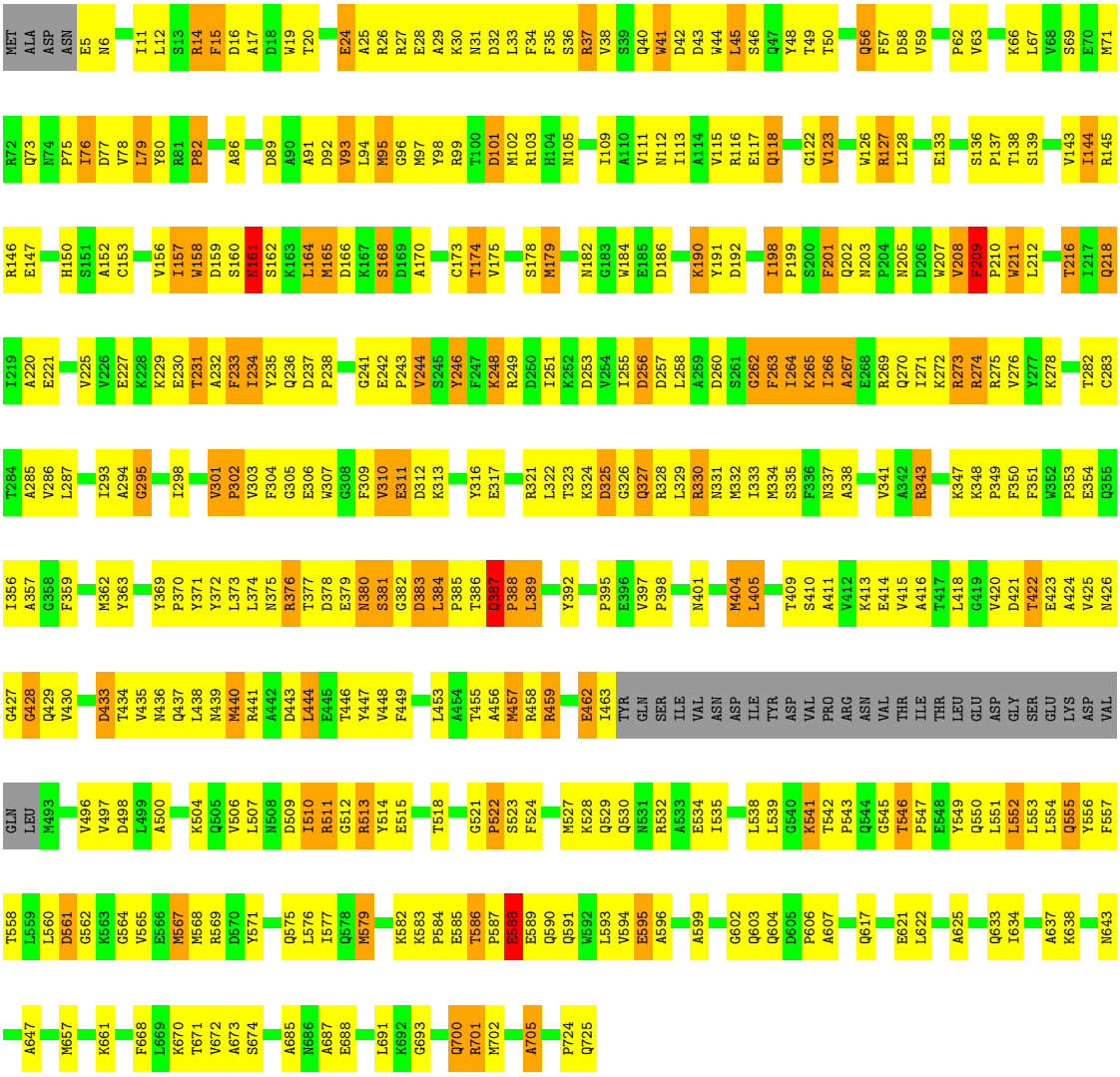




• Molecule 1: Portal protein

Chain L:


40%
43%
12%
• 5%


MET ALA ASP ASN E5 N6 I11 L12 S13 R14 F15 D16 A17 D18 W19 T20 E24 A25 R26 R27 E28 A29 K30 N31 D32 L33 F34 F35 S36 R37 V38 S39 Q40 W41 D42 D43 W44 L45 S46 Q47 Y48 T49 T50 Q56 F57 D58 V59 P62 V63 K66 L67 V68 S69 E70 M71
R72 Q73 N74 P75 I76 D77 V78 L79 Y80 R81 P82 A86 D89 A90 A91 D92 V93 L94 M95 G96 M97 Y98 R99 T100 D101 M102 R103 H104 N105 I109 A110 V111 N112 I113 A114 V115 R116 E117 Q118 G122 V123 W126 R127 L128 E133 S136 P137 T138 S139 V143 I144 R145
R146 E147 H150 S151 A152 C153 V156 I157 W158 D159 S160 N161 S162 K163 L164 M165 D166 K167 S168 D169 A170 C173 T174 V175 S178 M179 N182 G183 W184 E185 D186 K190 Y191 D192 I198 P199 S200 F201 Q202 N203 P204 N205 D206 W207 V208 F209 P210 W211 L212 T216 I217 Q218
I219 A220 E221 V225 V226 E227 K228 K229 E230 T231 A232 F233 I234 Y235 Q236 D237 P238 G241 E242 P243 V244 S245 Y246 F247 K248 R249 D250 I251 K252 D253 V254 I255 D256 D257 L258 A259 D260 S261 G262 F263 I264 K265 I266 A267 E268 R269 Q270 I271 K272 R273 R274 R275 V276 Y277 K278 T282 C283
T284 A285 V286 L287 I293 A294 G295 I298 V301 P302 V303 F304 G305 E306 W307 G308 F309 V310 E311 D312 K313 Y316 E317 R321 L322 T323 K324 D325 G326 Q327 R328 L329 R330 N331 M332 I333 M334 S335 F336 N337 A338 V341 A342 R343 K347 K348 P349 F350 F351 W352 P353 E354 Q355
I356 A357 G358 F359 M362 Y363 Y369 P370 Y371 Y372 L373 L374 N375 R376 T377 D378 E379 N380 S381 G382 D383 L384 P385 T386 Q387 P388 L389 Y392 P395 E396 V397 P398 N401 M404 L405 T409 S410 A411 V412 K413 E414 V415 A416 T417 L418 G419 V420 D421 T422 E423 A424 V425 N426
G427 G428 Q429 V430 D433 T434 V435 N436 Q437 L438 N439 M440 R441 A442 D443 L444 E445 T446 Y447 V448 F449 L453 A454 T455 A456 M457 R458 R459 E462 I463 TYR GLN SER ILE VAL ASN ASP ILE TYR ASP VAL PRO ARG ASN VAL THR ILE THR LEU GLU ASP GLY SER GLU LYS ASP VAL
GLN LEU M493 V496 V497 D498 L499 A500 K504 Q505 V506 L507 N508 D509 I510 R511 G512 R513 Y514 E515 T518 G521 P522 S523 F524 M527 K528 Q529 Q530 N531 R532 A533 E534 I535 L538 L539 G540 K541 T542 P543 Q544 G545 T546 P547 E548 Y549 Q550 L551 L552 L553 L554 Q555 Y556 F557
T558 L559 L560 D561 G562 K563 G564 V565 E566 M567 M568 R569 D570 Y571 Q575 L576 I577 Q578 M579 K582 K583 P584 E585 T586 P587 E588 E589 Q590 Q591 W592 L593 V594 E595 A596 A599 G602 Q603 Q604 D605 P606 A607 Q617 E621 L622 A625 Q633 I634 A637 K638 N643
A647 M657 K661 F668 L669 K670 T671 V672 A673 S674 A685 N686 A687 E688 L691 K692 G693 Q700 R701 M702 A705 P724 Q725

4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 4	Depositor
Cell constants a, b, c, α, β, γ	408.95Å 408.95Å 260.01Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	59.48 – 7.50 59.48 – 7.50	Depositor EDS
% Data completeness (in resolution range)	92.8 (59.48-7.50) 93.1 (59.48-7.50)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.10	Depositor
$< I/\sigma(I) >$ [1]	2.90 (at 7.40Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.5_2)	Depositor
R, R_{free}	0.187 , 0.263 0.181 , 0.261	Depositor DCC
R_{free} test set	1994 reflections (7.33%)	wwPDB-VP
Wilson B-factor (Å^2)	236.8	Xtriage
Anisotropy	0.019	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.42 , 547.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.28$	Xtriage
Estimated twinning fraction	0.035 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.81	EDS
Total number of atoms	61512	wwPDB-VP
Average B, all atoms (Å^2)	297.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.70% 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	0.50	11/5222 (0.2%)	0.58	1/7114 (0.0%)
1	B	0.50	11/5222 (0.2%)	0.58	1/7114 (0.0%)
1	C	0.50	11/5222 (0.2%)	0.58	1/7114 (0.0%)
1	D	0.50	11/5222 (0.2%)	0.58	1/7114 (0.0%)
1	E	0.50	11/5222 (0.2%)	0.58	1/7114 (0.0%)
1	F	0.50	11/5222 (0.2%)	0.58	1/7114 (0.0%)
1	G	0.50	11/5222 (0.2%)	0.58	1/7114 (0.0%)
1	H	0.50	11/5222 (0.2%)	0.58	1/7114 (0.0%)
1	I	0.50	11/5222 (0.2%)	0.58	1/7114 (0.0%)
1	J	0.50	11/5222 (0.2%)	0.58	1/7114 (0.0%)
1	K	0.50	11/5222 (0.2%)	0.58	1/7114 (0.0%)
1	L	0.50	11/5222 (0.2%)	0.58	1/7114 (0.0%)
All	All	0.50	132/62664 (0.2%)	0.58	12/85368 (0.0%)

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	1
1	B	0	1
1	C	0	1
1	D	0	1
1	E	0	1
1	F	0	1
1	G	0	1
1	H	0	1
1	I	0	1
1	J	0	1
1	K	0	1
1	L	0	1
All	All	0	12

The worst 5 of 132 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	F	165	MET	CG-SD	6.91	1.99	1.81
1	G	165	MET	CG-SD	6.91	1.99	1.81
1	J	165	MET	CG-SD	6.91	1.99	1.81
1	H	165	MET	CG-SD	6.90	1.99	1.81
1	I	165	MET	CG-SD	6.89	1.99	1.81

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	G	97	MET	CB-CG-SD	-6.24	93.67	112.40
1	E	97	MET	CB-CG-SD	-6.24	93.68	112.40
1	B	97	MET	CB-CG-SD	-6.24	93.68	112.40
1	A	97	MET	CB-CG-SD	-6.24	93.69	112.40
1	I	97	MET	CB-CG-SD	-6.24	93.69	112.40

There are no chirality outliers.

5 of 12 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	231	THR	Peptide
1	B	231	THR	Peptide
1	C	231	THR	Peptide
1	D	231	THR	Peptide
1	E	231	THR	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	5126	0	4586	548	9
1	B	5126	0	4586	522	0
1	C	5126	0	4586	499	0
1	D	5126	0	4586	559	0
1	E	5126	0	4586	551	8
1	F	5126	0	4586	522	1
1	G	5126	0	4586	538	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	H	5126	0	4586	509	3
1	I	5126	0	4586	516	0
1	J	5126	0	4586	516	3
1	K	5126	0	4586	510	0
1	L	5126	0	4586	538	0
All	All	61512	0	55032	5771	12

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 50.

The worst 5 of 5771 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:58:ASP:OD2	1:L:322:LEU:CD1	1.73	1.35
1:G:322:LEU:HD12	1:H:58:ASP:OD2	1.23	1.34
1:I:322:LEU:HD12	1:J:58:ASP:OD2	1.26	1.34
1:F:322:LEU:HD12	1:G:58:ASP:OD2	1.19	1.32
1:H:322:LEU:HD12	1:I:58:ASP:OD2	1.19	1.31

The worst 5 of 12 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:H:49:THR:O	1:J:684:ARG:O[8_554]	1.56	0.64
1:A:49:THR:O	1:E:684:ARG:O[3_555]	1.86	0.34
1:A:48:TYR:O	1:E:687:ALA:CB[3_555]	1.89	0.31
1:H:49:THR:OG1	1:J:687:ALA:CB[8_554]	1.93	0.27
1:A:49:THR:CB	1:E:687:ALA:CB[3_555]	1.94	0.26

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	688/725 (95%)	507 (74%)	122 (18%)	59 (9%)	1 11
1	B	688/725 (95%)	507 (74%)	122 (18%)	59 (9%)	1 11
1	C	688/725 (95%)	508 (74%)	121 (18%)	59 (9%)	1 11
1	D	688/725 (95%)	506 (74%)	123 (18%)	59 (9%)	1 11
1	E	688/725 (95%)	507 (74%)	122 (18%)	59 (9%)	1 11
1	F	688/725 (95%)	507 (74%)	122 (18%)	59 (9%)	1 11
1	G	688/725 (95%)	507 (74%)	122 (18%)	59 (9%)	1 11
1	H	688/725 (95%)	506 (74%)	123 (18%)	59 (9%)	1 11
1	I	688/725 (95%)	507 (74%)	122 (18%)	59 (9%)	1 11
1	J	688/725 (95%)	507 (74%)	122 (18%)	59 (9%)	1 11
1	K	688/725 (95%)	508 (74%)	121 (18%)	59 (9%)	1 11
1	L	688/725 (95%)	507 (74%)	122 (18%)	59 (9%)	1 11
All	All	8256/8700 (95%)	6084 (74%)	1464 (18%)	708 (9%)	1 11

5 of 708 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	41	TRP
1	A	82	PRO
1	A	208	VAL
1	A	256	ASP
1	A	263	PHE

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	474/630 (75%)	410 (86%)	64 (14%)	4 17
1	B	474/630 (75%)	409 (86%)	65 (14%)	3 17
1	C	474/630 (75%)	410 (86%)	64 (14%)	4 17
1	D	474/630 (75%)	409 (86%)	65 (14%)	3 17

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	E	474/630 (75%)	409 (86%)	65 (14%)	3	17
1	F	474/630 (75%)	409 (86%)	65 (14%)	3	17
1	G	474/630 (75%)	409 (86%)	65 (14%)	3	17
1	H	474/630 (75%)	410 (86%)	64 (14%)	4	17
1	I	474/630 (75%)	410 (86%)	64 (14%)	4	17
1	J	474/630 (75%)	410 (86%)	64 (14%)	4	17
1	K	474/630 (75%)	409 (86%)	65 (14%)	3	17
1	L	474/630 (75%)	410 (86%)	64 (14%)	4	17
All	All	5688/7560 (75%)	4914 (86%)	774 (14%)	3	17

5 of 774 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	H	79	LEU
1	I	518	THR
1	H	174	THR
1	H	69	SER
1	I	30	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 298 such sidechains are listed below:

Mol	Chain	Res	Type
1	J	161	ASN
1	L	327	GLN
1	J	297	HIS
1	K	236	GLN
1	D	555	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)

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.