wwPDB X-ray Structure Validation Summary Report (i)

Sep 19, 2023 – 08:58 PM EDT

PDB ID	:	5JJ3
Title	:	Refined Structure of the Mature Virion Conformation of P22 Portal Protein
Authors	:	Lokareddy, R.K.; Sankhala, R.S.; Cingolani, G.
Deposited on	:	2016-04-22
Resolution	:	7.00 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:

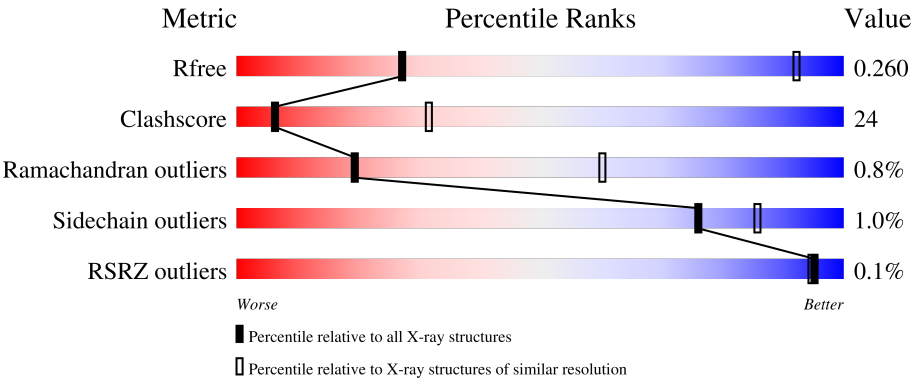
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 7.00 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))
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)
RSRZ outliers	127900	1004 (9.50-3.80)

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 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	725	49% 43% • 7%
1	B	725	48% 43% • 7%
1	C	725	47% 44% • 7%
1	D	725	50% 42% • 7%
1	E	725	48% 43% • 7%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	725	49% 43% • 7%
1	G	725	49% 42% • 7%
1	H	725	50% 41% • 7%
1	I	725	49% 43% • 7%
1	J	725	50% 42% • 7%
1	K	725	49% 43% • 7%
1	L	725	49% 43% • 7%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 64536 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.

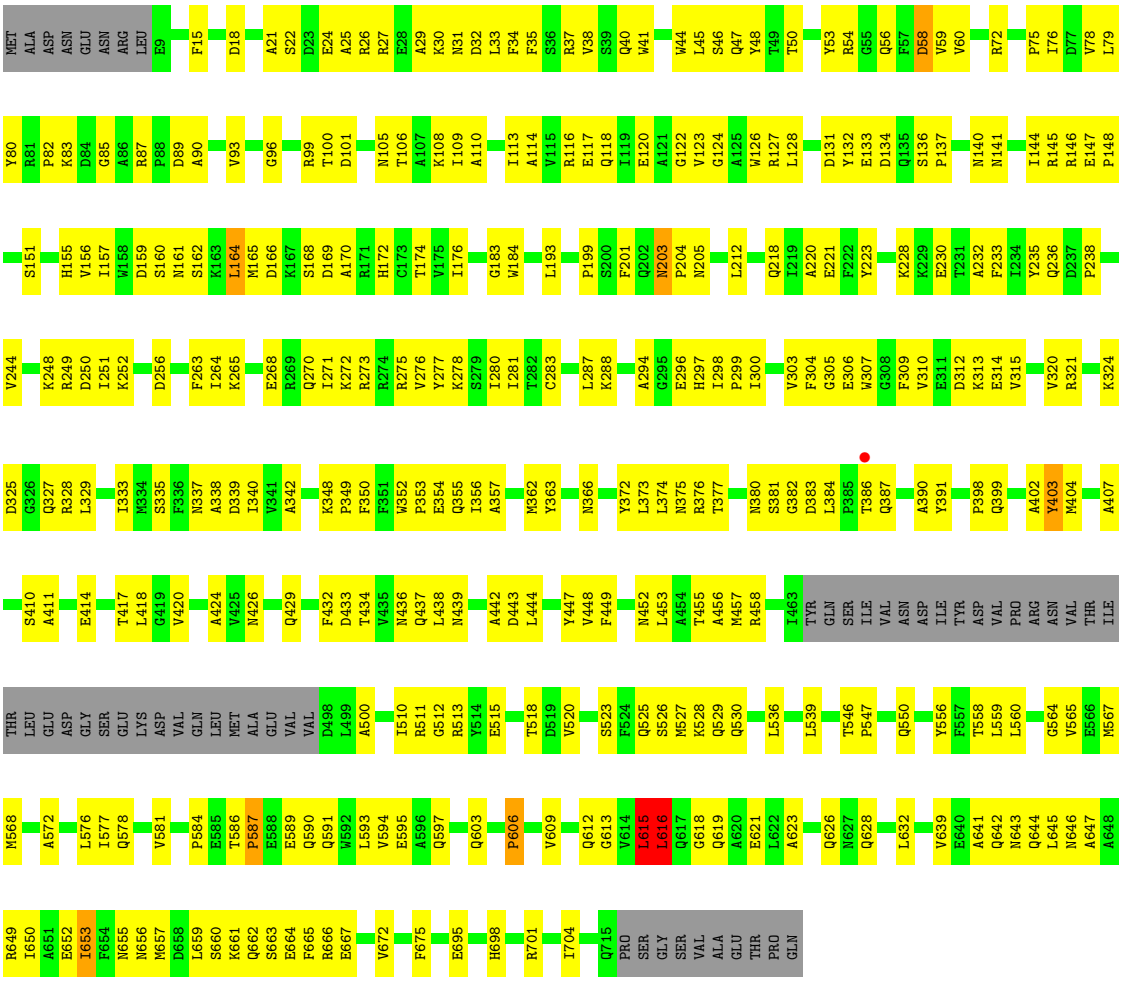
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	673	Total	C	N	O	S	0	0	0
			5378	3367	937	1051	23			
1	B	673	Total	C	N	O	S	0	0	0
			5378	3367	937	1051	23			
1	C	673	Total	C	N	O	S	0	0	0
			5378	3367	937	1051	23			
1	D	673	Total	C	N	O	S	0	0	0
			5378	3367	937	1051	23			
1	E	673	Total	C	N	O	S	0	0	0
			5378	3367	937	1051	23			
1	F	673	Total	C	N	O	S	0	0	0
			5378	3367	937	1051	23			
1	G	673	Total	C	N	O	S	0	0	0
			5378	3367	937	1051	23			
1	H	673	Total	C	N	O	S	0	0	0
			5378	3367	937	1051	23			
1	I	673	Total	C	N	O	S	0	0	0
			5378	3367	937	1051	23			
1	J	673	Total	C	N	O	S	0	0	0
			5378	3367	937	1051	23			
1	K	673	Total	C	N	O	S	0	0	0
			5378	3367	937	1051	23			
1	L	673	Total	C	N	O	S	0	0	0
			5378	3367	937	1051	23			

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.

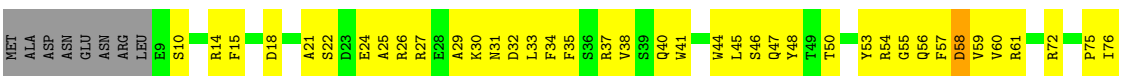
- Molecule 1: Portal protein

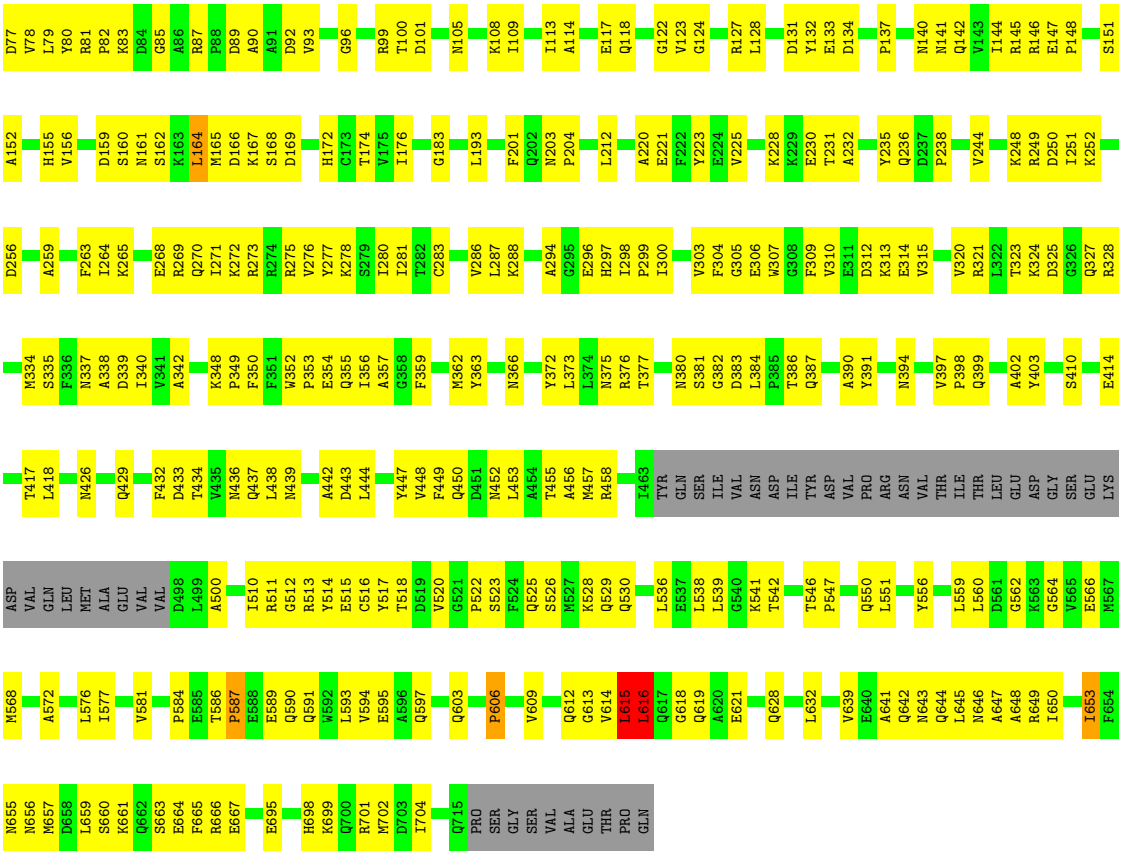
Chain A:



- Molecule 1: Portal protein

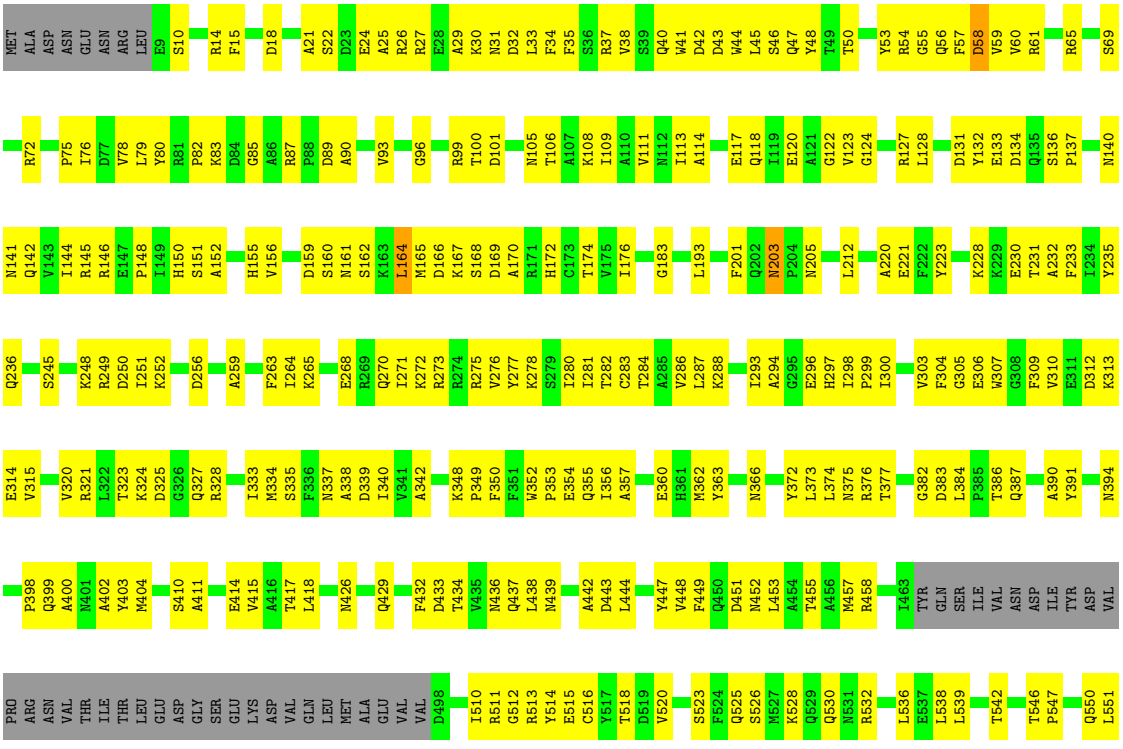
Chain B:

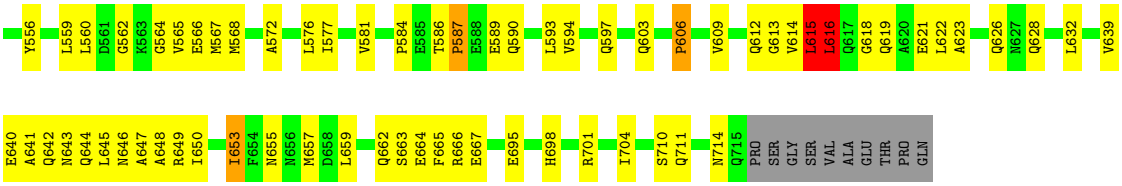




• Molecule 1: Portal protein

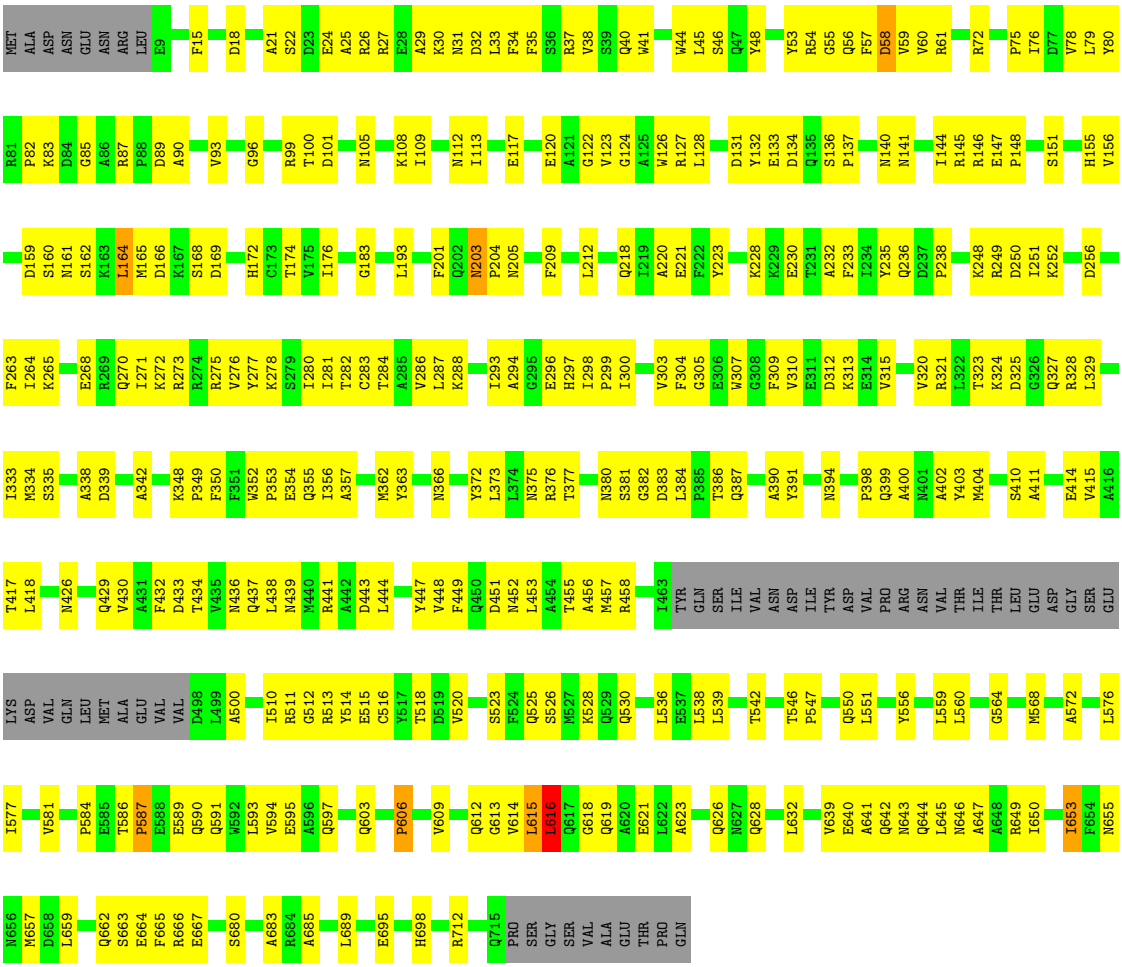
Chain C:





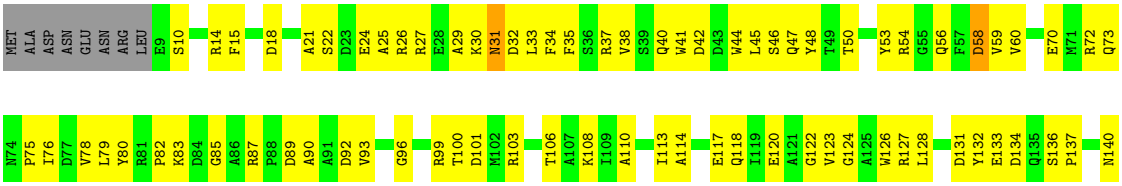
• Molecule 1: Portal protein

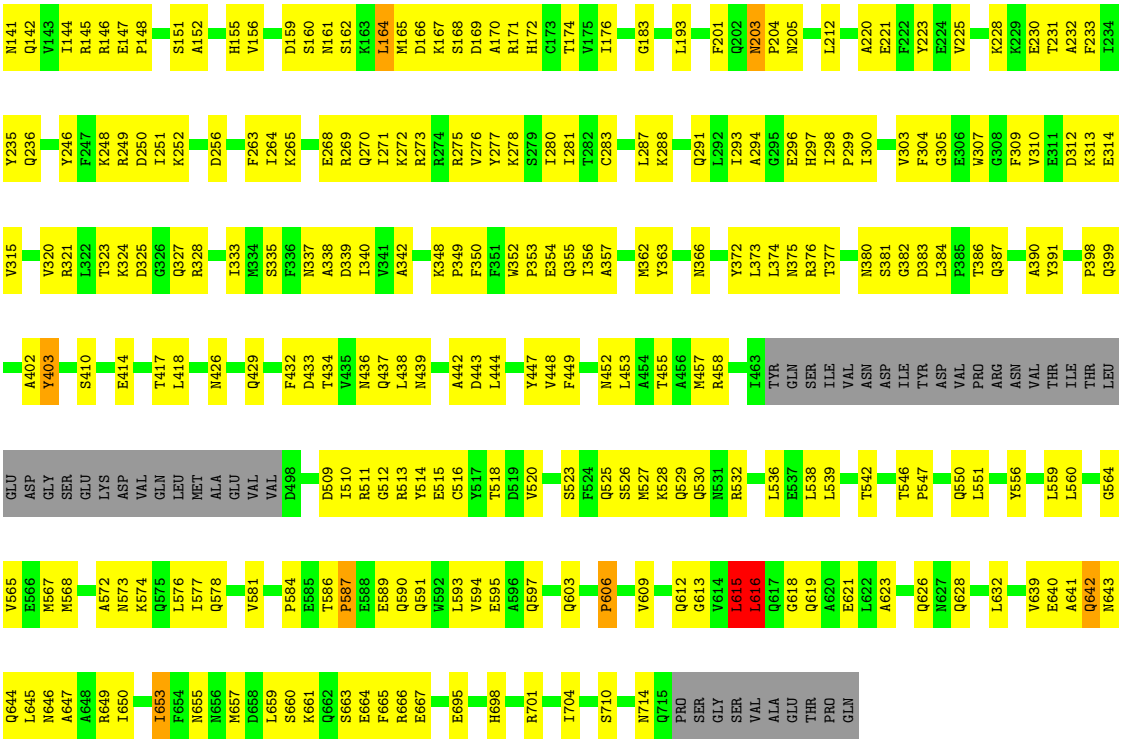
Chain D:



• Molecule 1: Portal protein

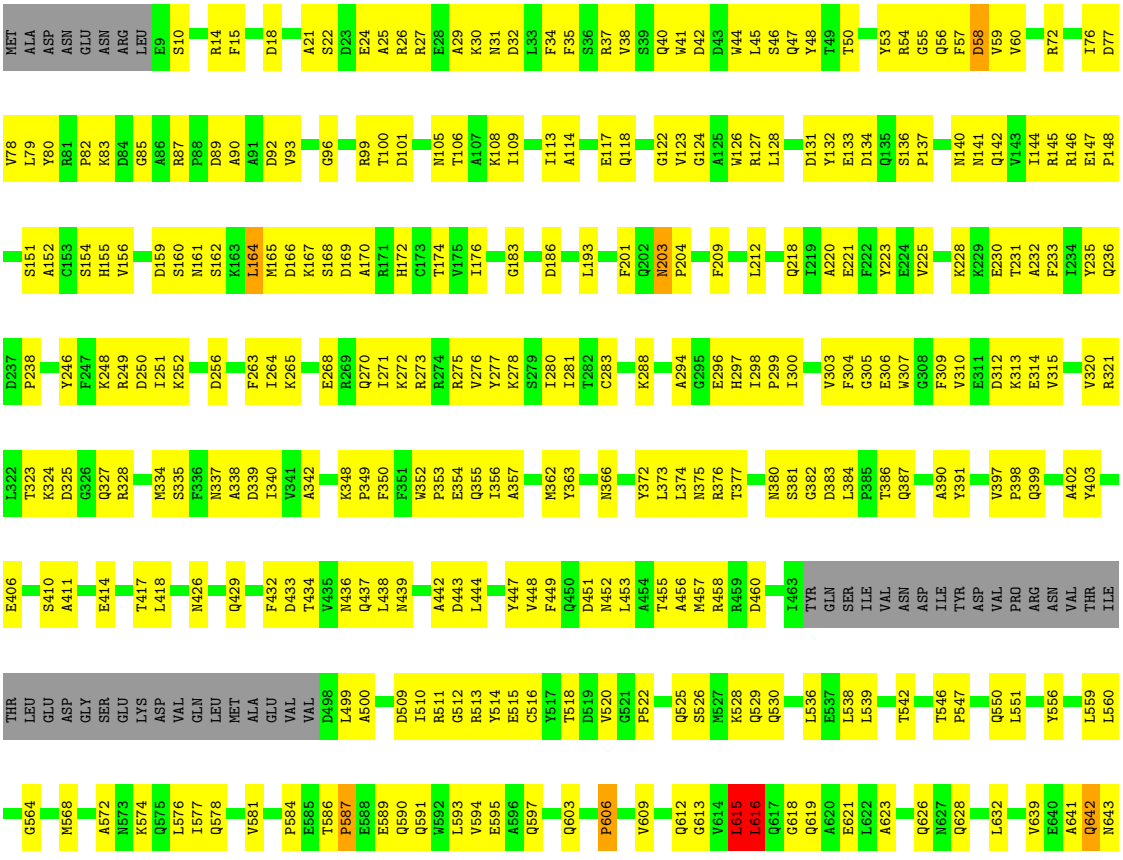
Chain E:

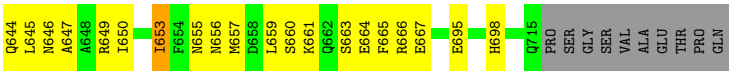




• Molecule 1: Portal protein

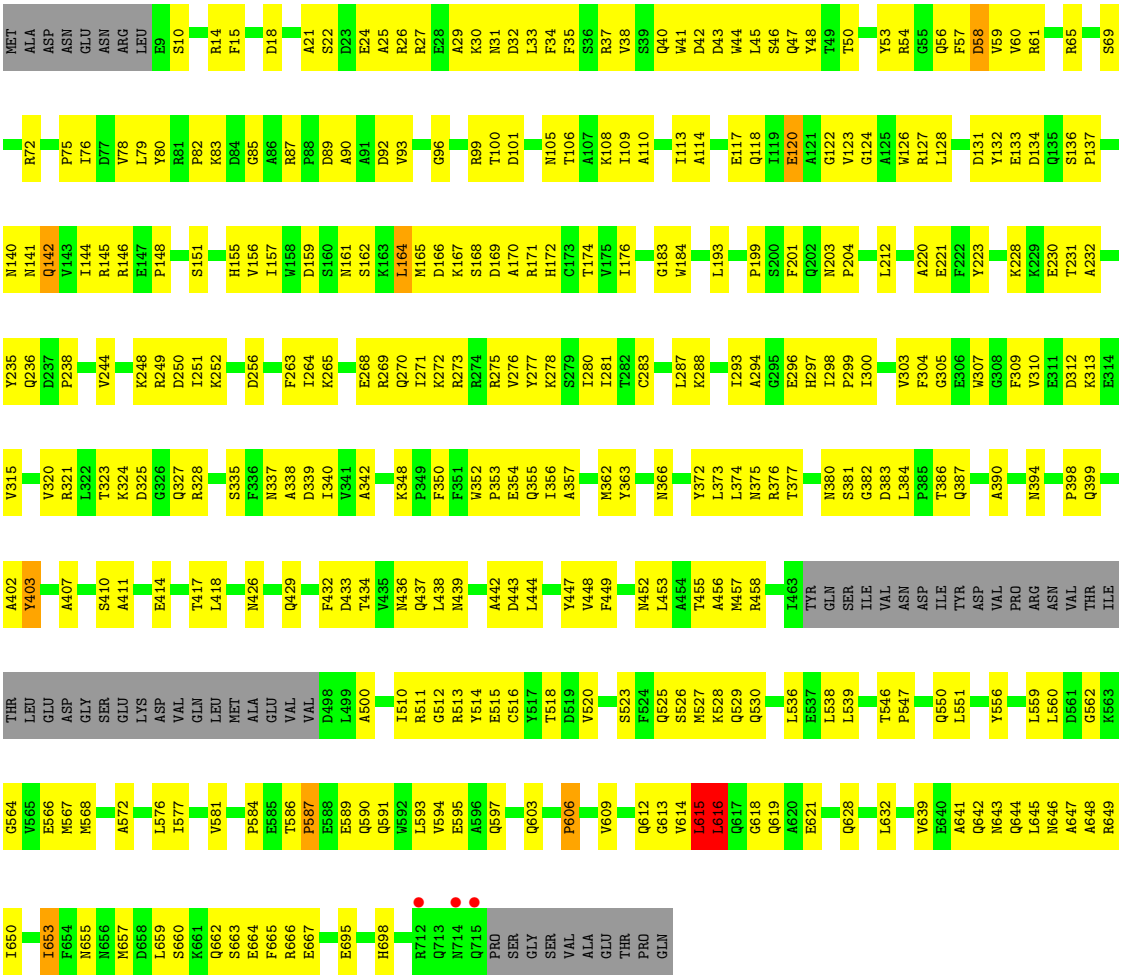
Chain F:





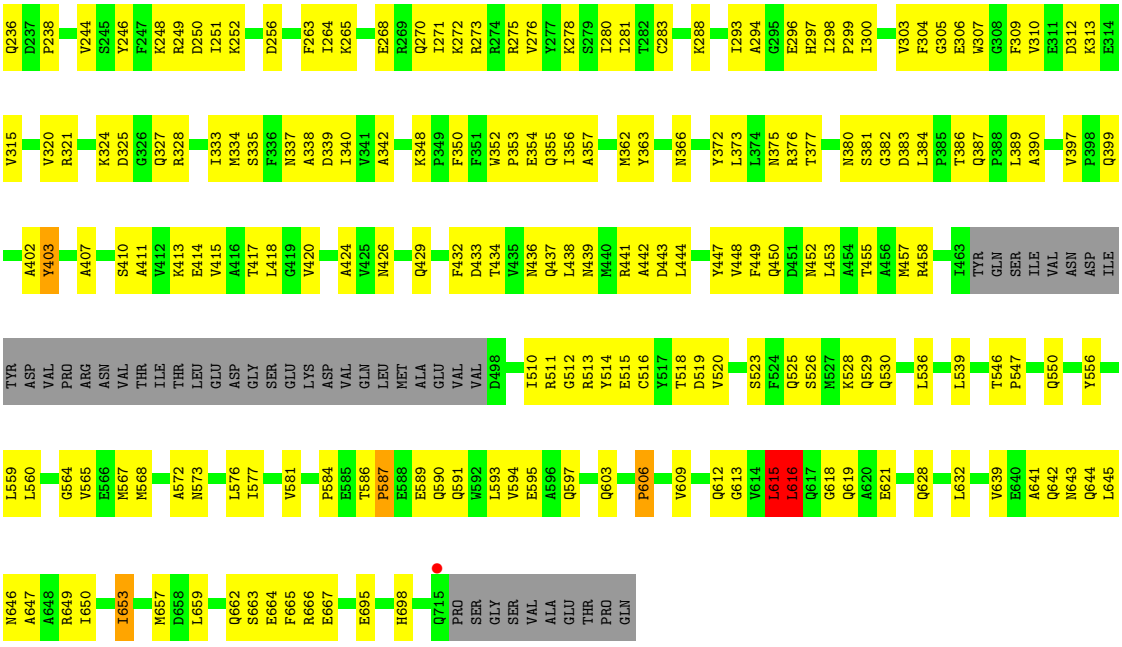
• Molecule 1: Portal protein

Chain G:



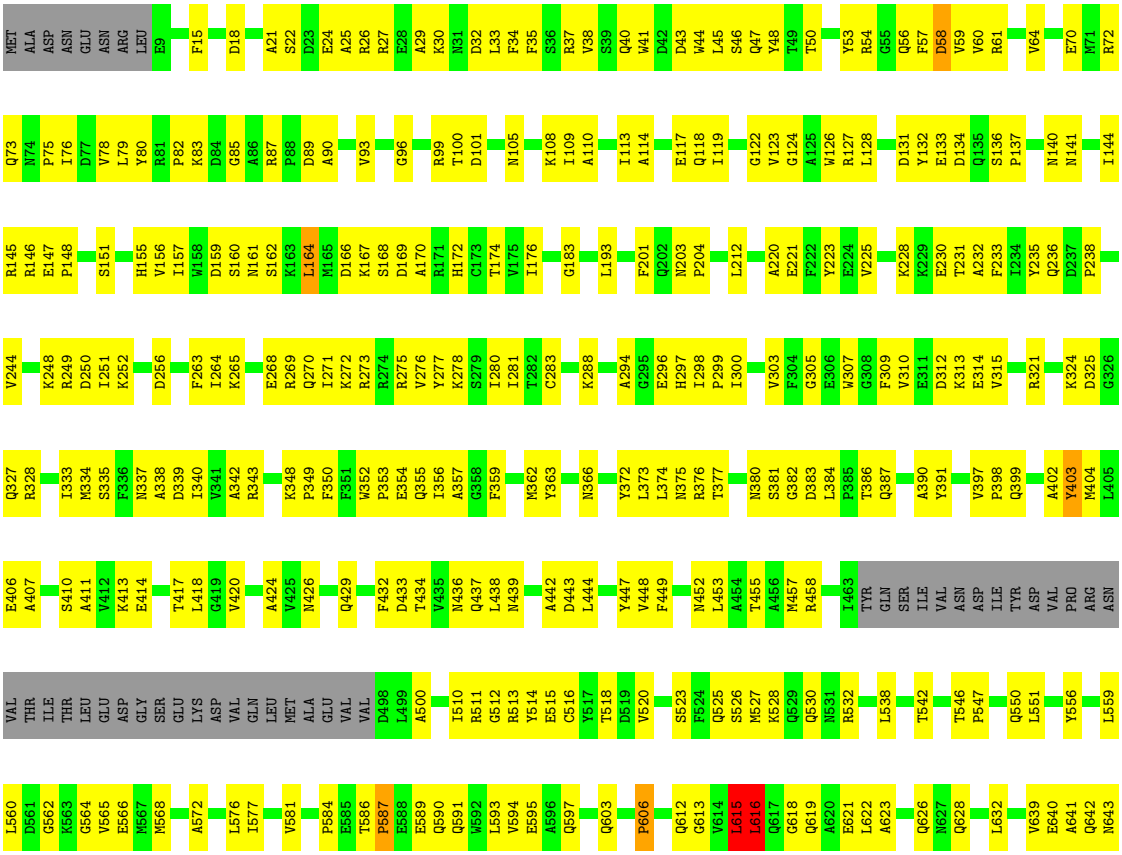
• Molecule 1: Portal protein

Chain H:



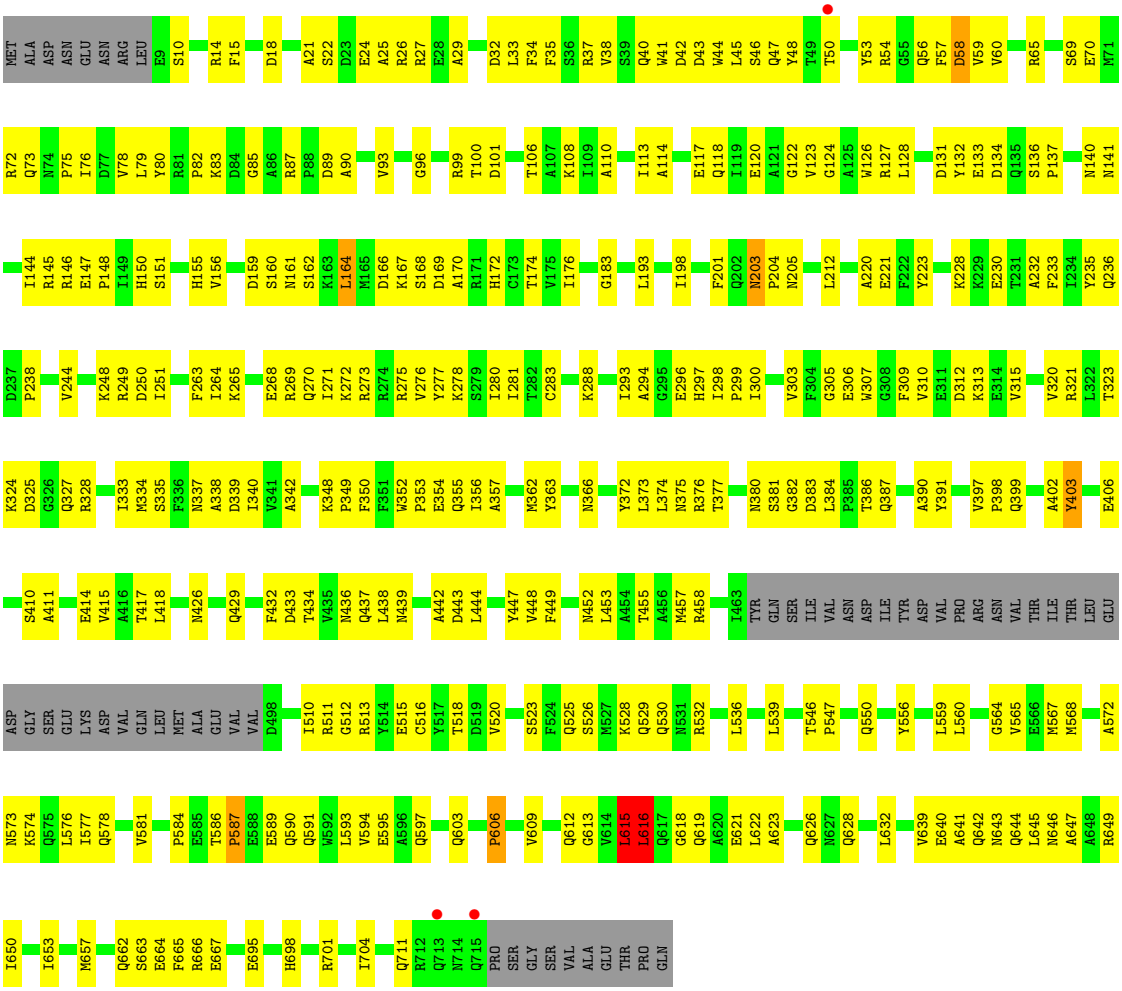
• Molecule 1: Portal protein

Chain I:



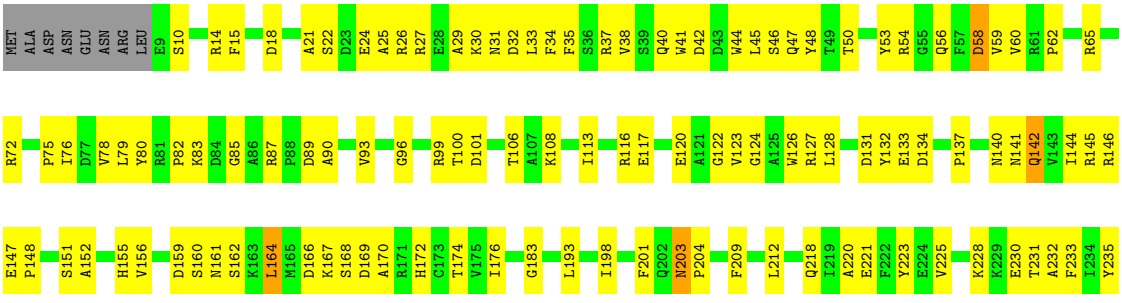
• Molecule 1: Portal protein

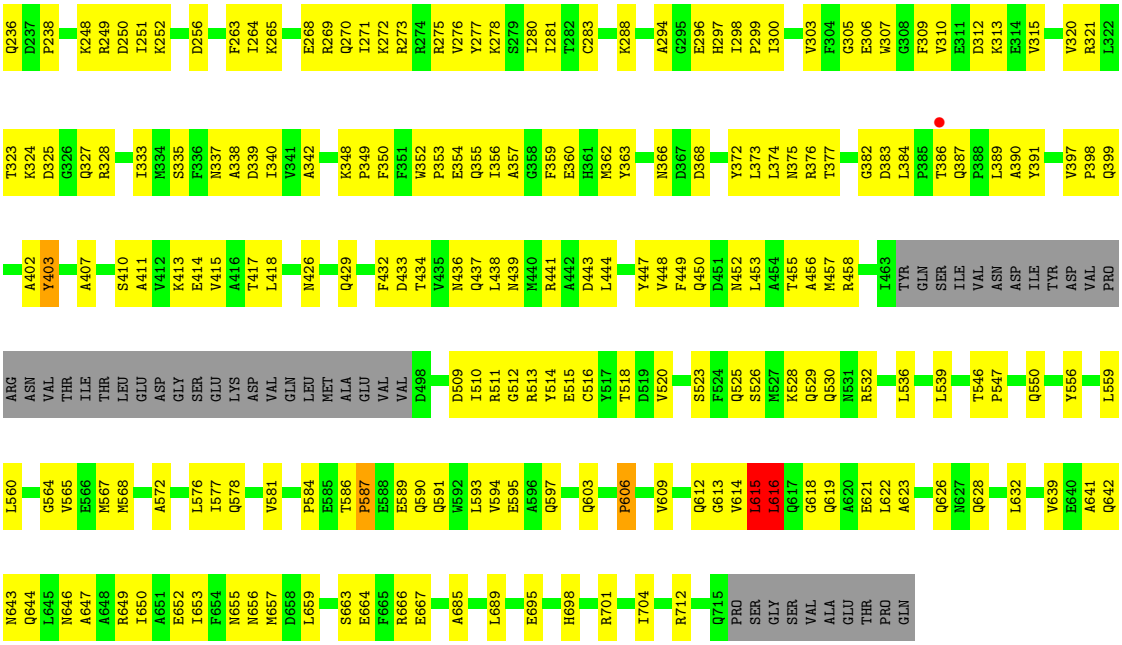
Chain J:



• Molecule 1: Portal protein

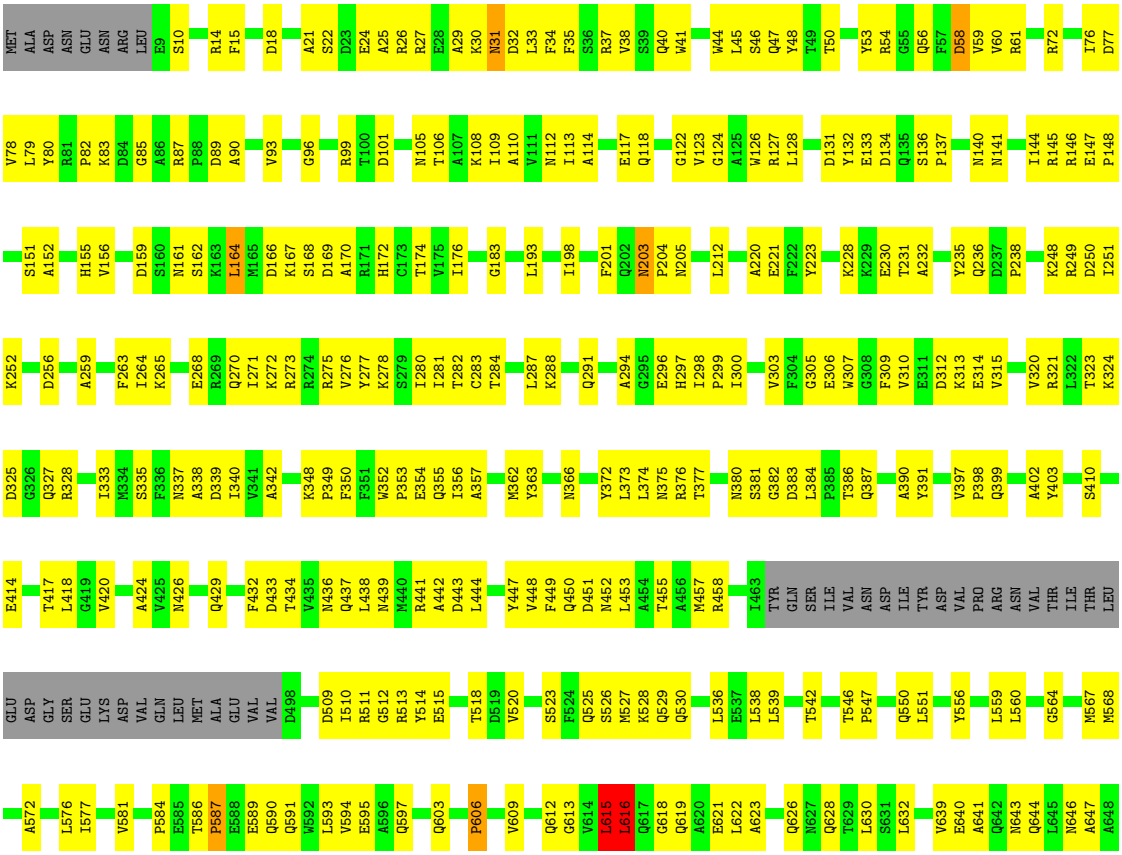
Chain K:





• Molecule 1: Portal protein

Chain L:




R649
I650
A651
E652
I653
F654
N655
N656
M657
D658
L659
Q662
S663
E664
F665
R666
E667
T671
F675
E695
H698
R701
I704
Q715
PRO
SER
GLY
SER
VAL
ALA
GLU
THR
PRO
GLN

4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 4	Depositor
Cell constants a, b, c, α, β, γ	409.04Å 409.04Å 260.00Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	14.99 – 7.00 14.99 – 7.00	Depositor EDS
% Data completeness (in resolution range)	99.0 (14.99-7.00) 99.1 (14.99-7.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.07 (at 7.36Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, R_{free}	0.239 , 0.260 0.239 , 0.260	Depositor DCC
R_{free} test set	1561 reflections (5.19%)	wwPDB-VP
Wilson B-factor (Å^2)	214.7	Xtriage
Anisotropy	0.009	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.14 , -10.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.28$	Xtriage
Estimated twinning fraction	0.045 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.88	EDS
Total number of atoms	64536	wwPDB-VP
Average B, all atoms (Å^2)	118.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.93% 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	0.28	0/5482	0.58	3/7429 (0.0%)
1	B	0.28	0/5482	0.59	2/7429 (0.0%)
1	C	0.28	0/5482	0.59	2/7429 (0.0%)
1	D	0.28	0/5482	0.57	3/7429 (0.0%)
1	E	0.27	0/5482	0.58	2/7429 (0.0%)
1	F	0.27	0/5482	0.58	3/7429 (0.0%)
1	G	0.27	0/5482	0.58	2/7429 (0.0%)
1	H	0.27	0/5482	0.57	2/7429 (0.0%)
1	I	0.28	0/5482	0.59	3/7429 (0.0%)
1	J	0.29	1/5482 (0.0%)	0.59	3/7429 (0.0%)
1	K	0.30	1/5482 (0.0%)	0.59	2/7429 (0.0%)
1	L	0.29	1/5482 (0.0%)	0.58	3/7429 (0.0%)
All	All	0.28	3/65784 (0.0%)	0.58	30/89148 (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	4
1	B	0	4
1	C	0	3
1	D	0	4
1	E	0	3
1	F	0	4
1	G	0	4
1	H	0	3
1	I	0	3
1	J	0	3
1	K	0	4
1	L	0	3
All	All	0	42

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	K	198	ILE	C-N	8.54	1.50	1.34
1	J	198	ILE	C-N	8.02	1.49	1.34
1	L	198	ILE	C-N	6.80	1.47	1.34

The worst 5 of 30 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	I	616	LEU	CA-CB-CG	13.49	146.32	115.30
1	J	616	LEU	CA-CB-CG	13.31	145.92	115.30
1	C	615	LEU	CA-CB-CG	12.84	144.83	115.30
1	B	616	LEU	CA-CB-CG	12.57	144.21	115.30
1	I	615	LEU	CA-CB-CG	12.56	144.19	115.30

There are no chirality outliers.

5 of 42 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	212	LEU	Peptide
1	A	456	ALA	Peptide
1	A	59	VAL	Peptide
1	A	606	PRO	Peptide
1	B	59	VAL	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	5378	0	5178	281	0
1	B	5378	0	5178	286	0
1	C	5378	0	5178	295	0
1	D	5378	0	5178	276	0
1	E	5378	0	5178	290	0
1	F	5378	0	5178	278	0
1	G	5378	0	5178	277	0
1	H	5378	0	5178	279	0
1	I	5378	0	5178	280	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	J	5378	0	5178	280	0
1	K	5378	0	5178	287	0
1	L	5378	0	5178	283	0
All	All	64536	0	62136	3062	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 24.

The worst 5 of 3062 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:J:220:ALA:HB3	1:J:281:ILE:O	1.55	1.05
1:B:220:ALA:HB3	1:B:281:ILE:O	1.58	1.03
1:E:220:ALA:HB3	1:E:281:ILE:O	1.58	1.02
1:I:220:ALA:HB3	1:I:281:ILE:O	1.59	1.02
1:K:220:ALA:HB3	1:K:281:ILE:O	1.59	1.01

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	669/725 (92%)	562 (84%)	102 (15%)	5 (1%)	22	63
1	B	669/725 (92%)	561 (84%)	103 (15%)	5 (1%)	22	63
1	C	669/725 (92%)	560 (84%)	104 (16%)	5 (1%)	22	63
1	D	669/725 (92%)	561 (84%)	103 (15%)	5 (1%)	22	63
1	E	669/725 (92%)	562 (84%)	101 (15%)	6 (1%)	17	57
1	F	669/725 (92%)	561 (84%)	103 (15%)	5 (1%)	22	63
1	G	669/725 (92%)	563 (84%)	100 (15%)	6 (1%)	17	57

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	H	669/725 (92%)	561 (84%)	103 (15%)	5 (1%)	22	63
1	I	669/725 (92%)	560 (84%)	105 (16%)	4 (1%)	25	66
1	J	669/725 (92%)	559 (84%)	106 (16%)	4 (1%)	25	66
1	K	669/725 (92%)	560 (84%)	104 (16%)	5 (1%)	22	63
1	L	669/725 (92%)	562 (84%)	101 (15%)	6 (1%)	17	57
All	All	8028/8700 (92%)	6732 (84%)	1235 (15%)	61 (1%)	19	60

5 of 61 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	58	ASP
1	B	58	ASP
1	C	58	ASP
1	D	58	ASP
1	E	58	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	572/630 (91%)	567 (99%)	5 (1%)	78	87
1	B	572/630 (91%)	566 (99%)	6 (1%)	76	86
1	C	572/630 (91%)	566 (99%)	6 (1%)	76	86
1	D	572/630 (91%)	568 (99%)	4 (1%)	84	90
1	E	572/630 (91%)	565 (99%)	7 (1%)	71	83
1	F	572/630 (91%)	565 (99%)	7 (1%)	71	83
1	G	572/630 (91%)	566 (99%)	6 (1%)	76	86
1	H	572/630 (91%)	568 (99%)	4 (1%)	84	90
1	I	572/630 (91%)	567 (99%)	5 (1%)	78	87
1	J	572/630 (91%)	567 (99%)	5 (1%)	78	87
1	K	572/630 (91%)	566 (99%)	6 (1%)	76	86

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	L	572/630 (91%)	567 (99%)	5 (1%)	78 87
All	All	6864/7560 (91%)	6798 (99%)	66 (1%)	76 86

5 of 66 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	K	277	TYR
1	K	615	LEU
1	L	616	LEU
1	E	403	TYR
1	E	277	TYR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 72 such sidechains are listed below:

Mol	Chain	Res	Type
1	J	550	GLN
1	L	597	GLN
1	J	643	ASN
1	K	550	GLN
1	D	643	ASN

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

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	673/725 (92%)	-0.71	1 (0%) 95 95	46, 96, 204, 282	0
1	B	673/725 (92%)	-0.69	0 100 100	31, 97, 169, 284	0
1	C	673/725 (92%)	-0.70	0 100 100	47, 81, 181, 288	0
1	D	673/725 (92%)	-0.68	0 100 100	51, 97, 160, 240	0
1	E	673/725 (92%)	-0.68	0 100 100	54, 100, 176, 292	0
1	F	673/725 (92%)	-0.68	0 100 100	73, 110, 184, 317	0
1	G	673/725 (92%)	-0.62	3 (0%) 92 87	71, 117, 190, 362	0
1	H	673/725 (92%)	-0.66	1 (0%) 95 95	77, 127, 205, 345	0
1	I	673/725 (92%)	-0.64	0 100 100	90, 134, 222, 333	0
1	J	673/725 (92%)	-0.62	3 (0%) 92 87	70, 139, 211, 312	0
1	K	673/725 (92%)	-0.67	1 (0%) 95 95	65, 113, 182, 250	0
1	L	673/725 (92%)	-0.70	0 100 100	63, 106, 203, 303	0
All	All	8076/8700 (92%)	-0.67	9 (0%) 95 95	31, 112, 196, 362	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	714	ASN	4.0
1	G	715	GLN	3.7
1	G	712	ARG	2.7
1	J	50	THR	2.5
1	H	715	GLN	2.4

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 (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.