wwPDB X-ray Structure Validation Summary Report (i)

Aug 27, 2023 – 05:49 AM EDT

PDB ID	:	3HVD
Title	:	The Protective Antigen Component of Anthrax Toxin Forms Functional Octameric Complexes
Authors	:	Kintzer, A.F.; Dong, K.C.; Berger, J.M.; Krantz, B.A.
Deposited on	:	2009-06-15
Resolution	:	3.21 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 (1)) were used in the production of this report:

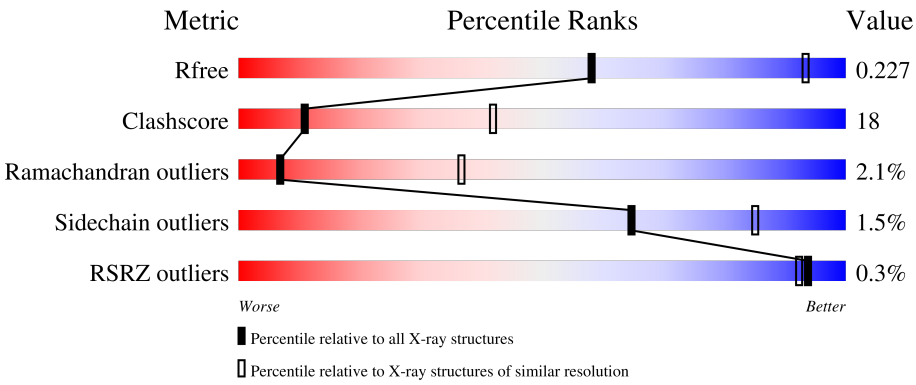
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

1 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 3.21 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	1133 (3.20-3.20)
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)
RSRZ outliers	127900	1095 (3.20-3.20)

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	548	57% 36% • 5%
1	B	548	60% 34% • 5%
1	C	548	58% 35% • 6%
1	D	548	.% 60% 33% • 5%
1	E	548	57% 36% • 5%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	548	% 60% 33% • 5%
1	G	548	56% 37% • 5%
1	H	548	59% 34% • 5%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 32701 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 Protective antigen.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	518	4067	2555	698	808	6	0	0	0
1	B	523	4075	2555	704	810	6	0	0	0
1	C	517	4078	2562	701	809	6	0	0	0
1	D	518	4066	2554	698	808	6	0	0	0
1	E	519	4091	2568	703	814	6	0	0	0
1	F	519	4098	2572	706	814	6	0	0	0
1	G	518	4094	2570	703	815	6	0	0	0
1	H	518	4108	2578	704	820	6	0	0	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	323	PRO	VAL	engineered mutation	UNP Q08G54
A	324	GLY	HIS	engineered mutation	UNP Q08G54
A	669	ARG	GLN	SEE REMARK 999	UNP Q08G54
B	303	PRO	VAL	engineered mutation	UNP Q08G54
B	304	GLY	HIS	engineered mutation	UNP Q08G54
B	669	ARG	GLN	SEE REMARK 999	UNP Q08G54
C	323	PRO	VAL	engineered mutation	UNP Q08G54
C	324	GLY	HIS	engineered mutation	UNP Q08G54
C	669	ARG	GLN	SEE REMARK 999	UNP Q08G54
D	323	PRO	VAL	engineered mutation	UNP Q08G54
D	324	GLY	HIS	engineered mutation	UNP Q08G54
D	669	ARG	GLN	SEE REMARK 999	UNP Q08G54
E	303	PRO	VAL	engineered mutation	UNP Q08G54

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	324	GLY	HIS	engineered mutation	UNP Q08G54
E	669	ARG	GLN	SEE REMARK 999	UNP Q08G54
F	323	PRO	VAL	engineered mutation	UNP Q08G54
F	324	GLY	HIS	engineered mutation	UNP Q08G54
F	669	ARG	GLN	SEE REMARK 999	UNP Q08G54
G	323	PRO	VAL	engineered mutation	UNP Q08G54
G	324	GLY	HIS	engineered mutation	UNP Q08G54
G	669	ARG	GLN	SEE REMARK 999	UNP Q08G54
H	323	PRO	VAL	engineered mutation	UNP Q08G54
H	324	GLY	HIS	engineered mutation	UNP Q08G54
H	669	ARG	GLN	SEE REMARK 999	UNP Q08G54

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Ca		
2	A	2	2	2	0	0
2	B	2	2	2	0	0
2	C	2	2	2	0	0
2	D	2	2	2	0	0
2	E	2	2	2	0	0
2	F	2	2	2	0	0
2	G	2	2	2	0	0
2	H	2	2	2	0	0

- Molecule 3 is water.

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

Continued on next page...

Continued from previous page...

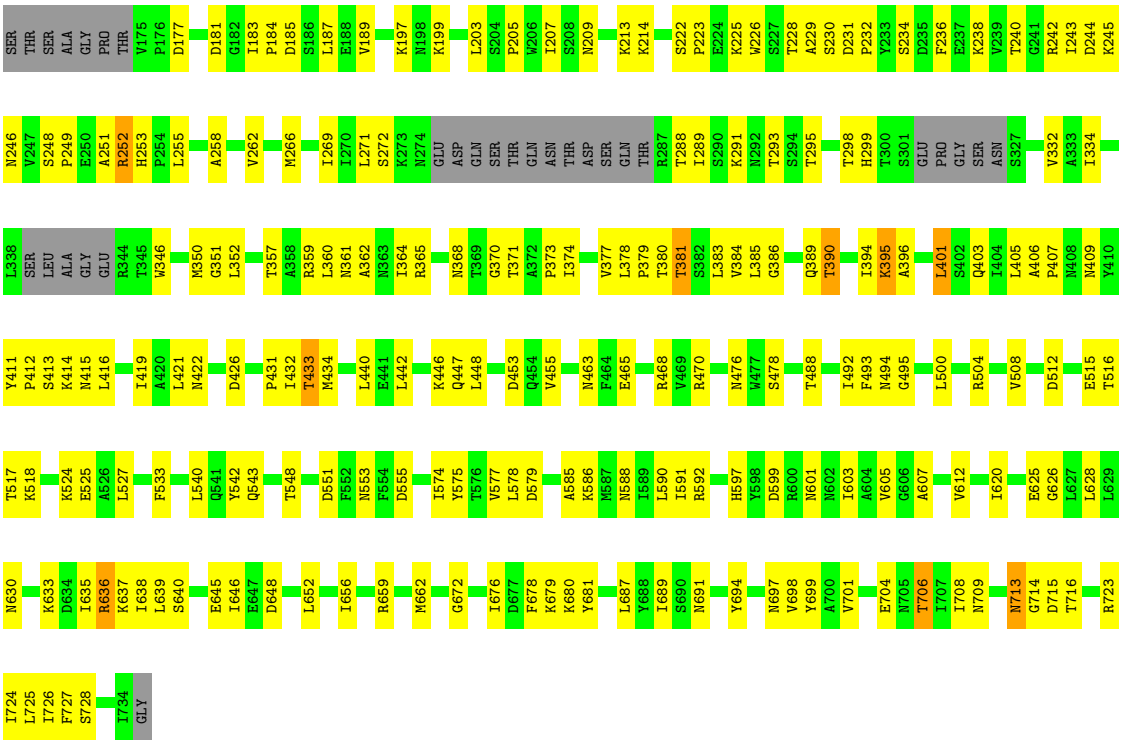
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	E	1	Total O 1 1	0	0
3	F	1	Total O 1 1	0	0
3	G	1	Total O 1 1	0	0
3	H	1	Total O 1 1	0	0

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.

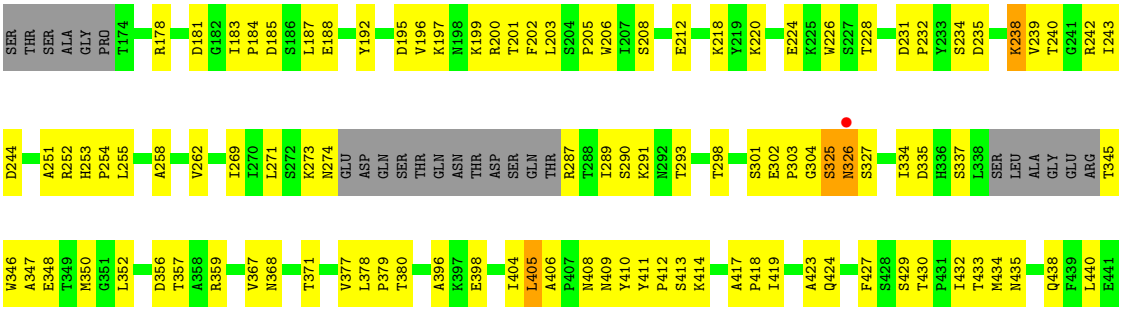
• Molecule 1: Protective antigen

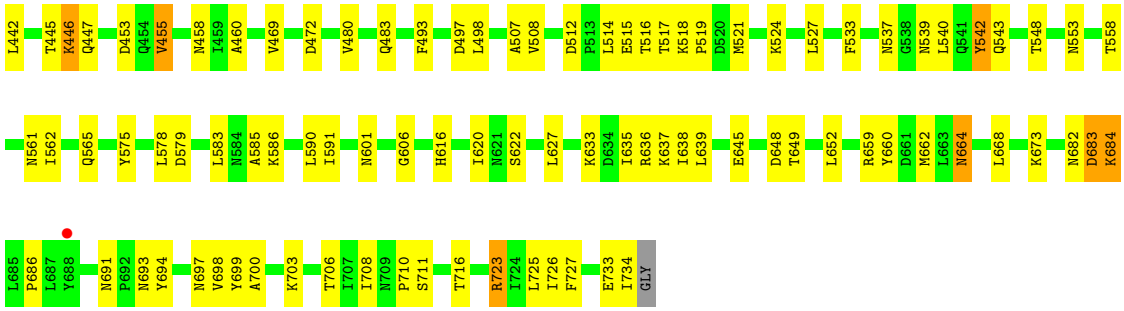
Chain A:



• Molecule 1: Protective antigen

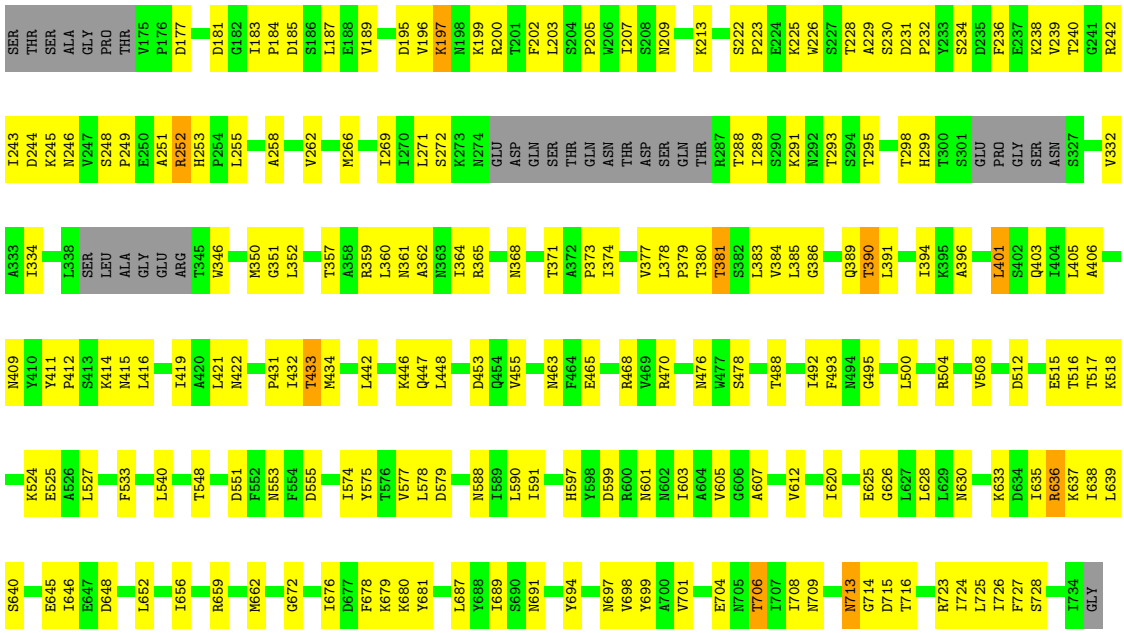
Chain B:





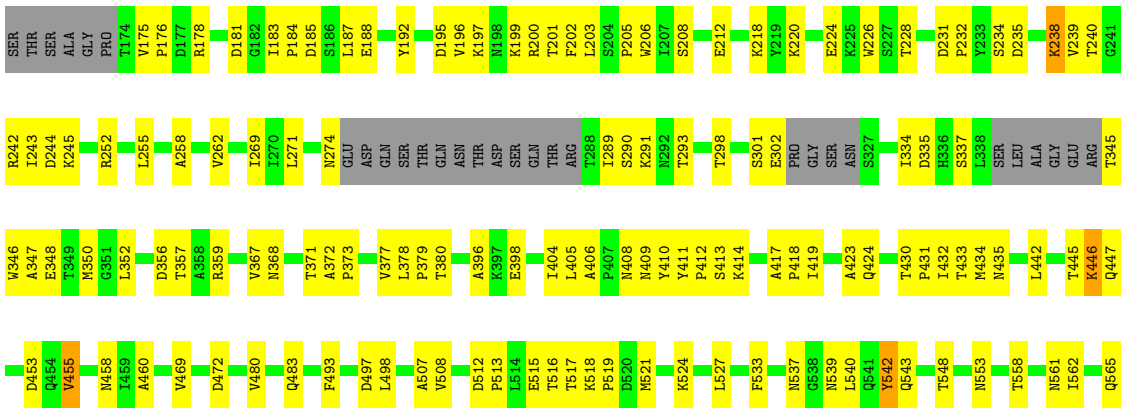
- Molecule 1: Protective antigen

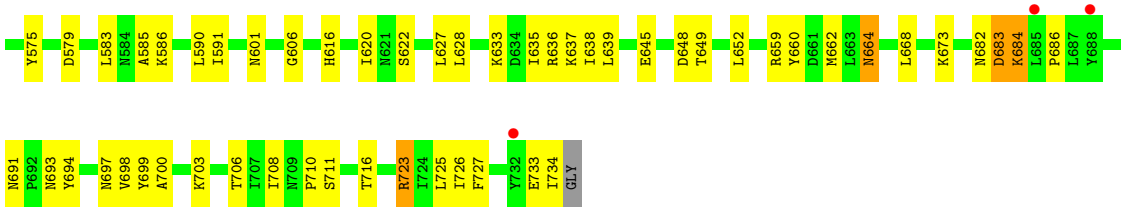
Chain C: 58% 35% • 6%



- Molecule 1: Protective antigen

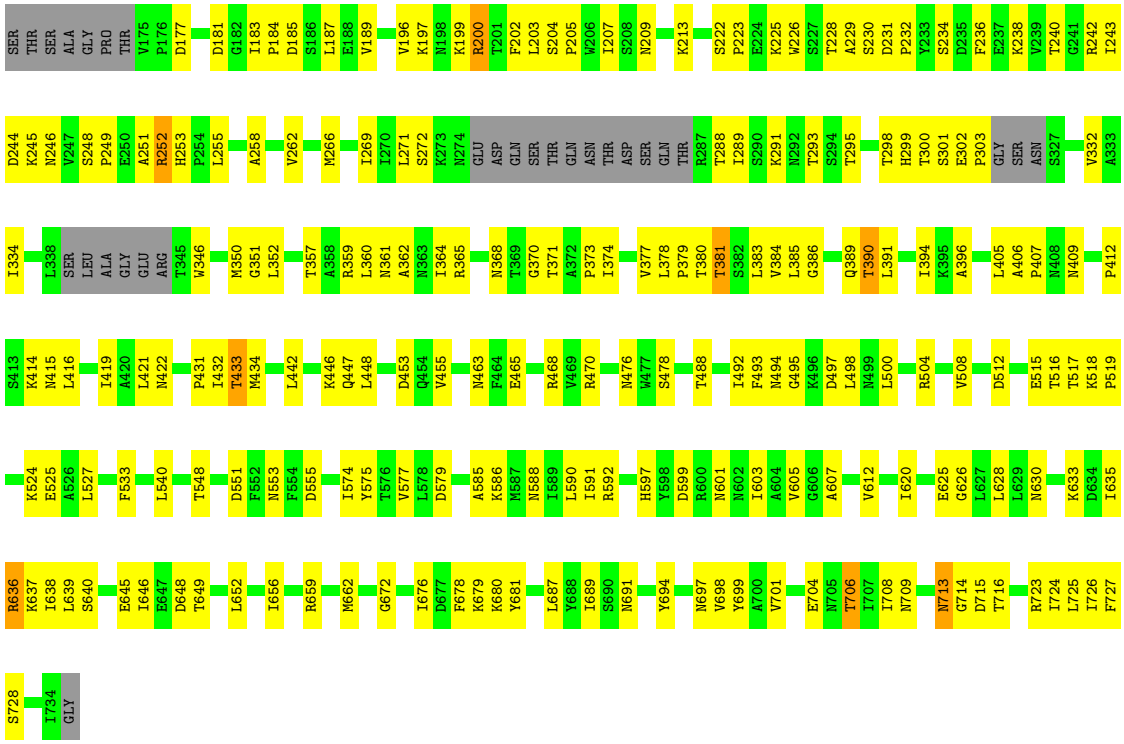
Chain D: % 60% 33% • 5%





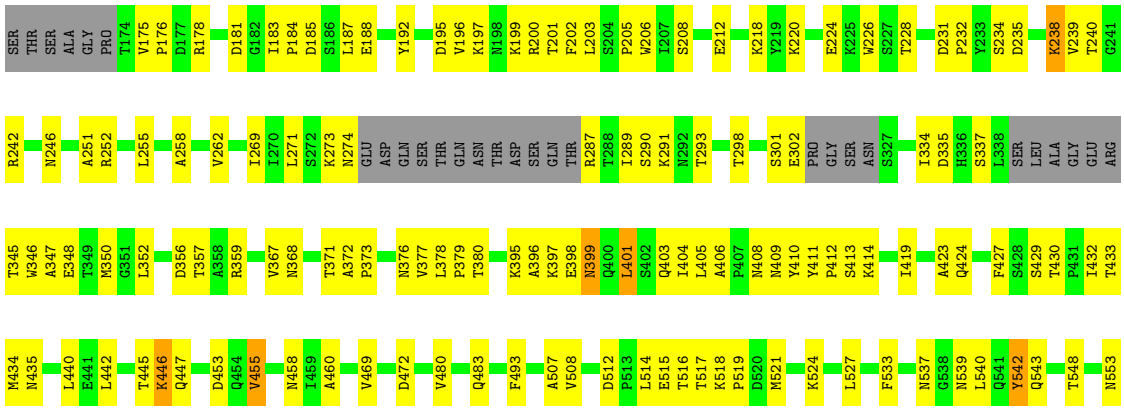
• Molecule 1: Protective antigen

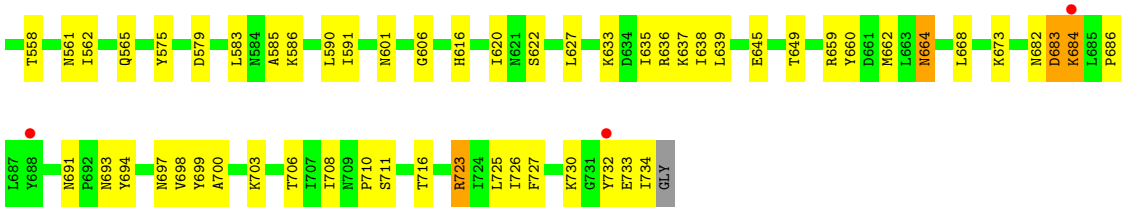
Chain E:



• Molecule 1: Protective antigen

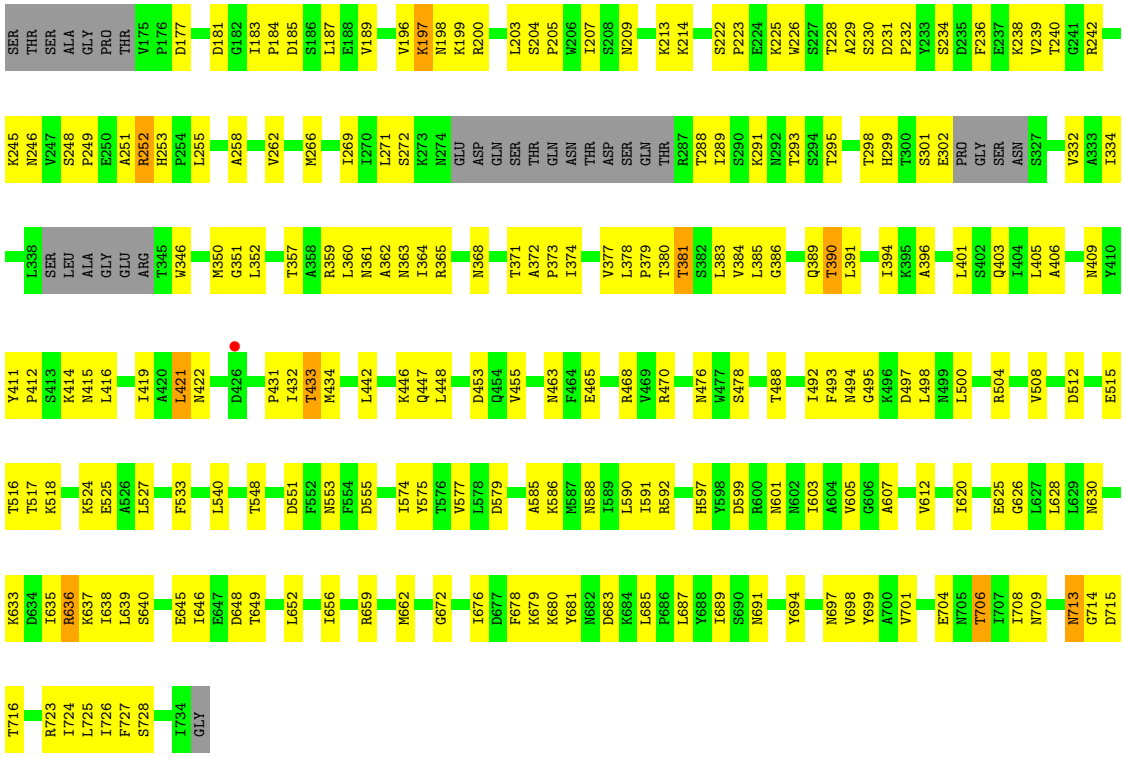
Chain F:





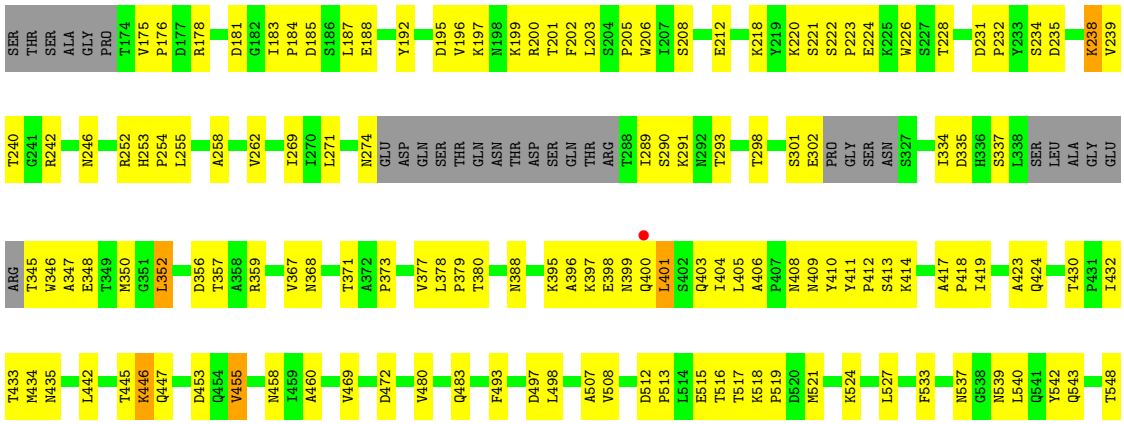
- Molecule 1: Protective antigen

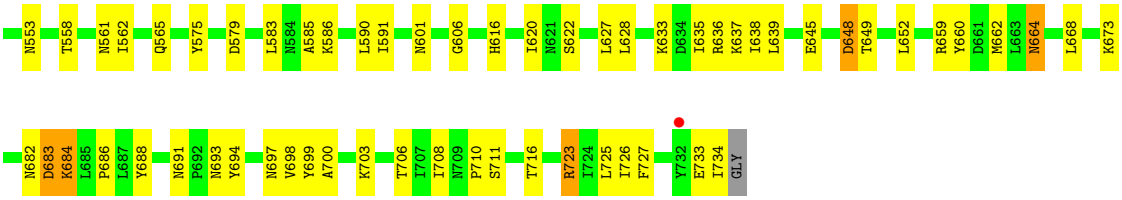
Chain G:



- Molecule 1: Protective antigen

Chain H:




N553
T558
N561
I562
Q565
Y575
D579
L583
N584
A585
K586
L590
I591
N601
G606
H616
I620
N621
S622
L627
L628
K633
D634
I635
R636
K637
I638
L639
E645
D648
T649
L652
R659
Y660
D661
M662
L663
N664
L668
K673
N682
D683
K684
L685
P686
L687
Y688
N691
P692
N693
Y694
N697
V698
Y699
A700
K703
T706
I707
I708
N709
P710
S711
T716
R723
I724
L725
I726
F727
Y732
E733
I734
GLY

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	125.60Å 125.67Å 125.82Å 106.64° 110.82° 110.98°	Depositor
Resolution (Å)	47.27 – 3.21 47.28 – 3.21	Depositor EDS
% Data completeness (in resolution range)	90.3 (47.27-3.21) 90.2 (47.28-3.21)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	0.12	Depositor
$< I/\sigma(I) >$ [1]	2.16 (at 3.19Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 2009_02_15_2320_3)	Depositor
R, R_{free}	0.210 , 0.249 0.188 , 0.227	Depositor DCC
R_{free} test set	1840 reflections (1.94%)	wwPDB-VP
Wilson B-factor (Å^2)	60.6	Xtriage
Anisotropy	0.034	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 22.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.438 for -l,-h,h+k+l 0.438 for -k,h+k+l,-h 0.032 for k,h,-h-k-l 0.458 for -h-k-l,l,k 0.033 for h+k+l,-k,-l 0.033 for l,-h-k-l,h 0.032 for -h,-l,-k	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	32701	wwPDB-VP
Average B, all atoms (Å^2)	58.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.73% 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 (i)

5.1 Standard geometry (i)

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

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	0.25	0/4136	0.45	0/5617
1	B	0.24	0/4144	0.46	0/5632
1	C	0.25	0/4147	0.45	0/5629
1	D	0.24	0/4134	0.45	0/5614
1	E	0.25	0/4161	0.45	0/5650
1	F	0.24	0/4167	0.46	0/5655
1	G	0.25	0/4163	0.46	0/5650
1	H	0.24	0/4177	0.45	0/5667
All	All	0.25	0/33229	0.45	0/45114

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

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	4067	0	4005	164	0
1	B	4075	0	3995	143	0
1	C	4078	0	4036	159	0
1	D	4066	0	4005	140	0
1	E	4091	0	4035	160	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	F	4098	0	4057	146	0
1	G	4094	0	4050	168	0
1	H	4108	0	4078	151	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
2	E	2	0	0	0	0
2	F	2	0	0	0	0
2	G	2	0	0	0	0
2	H	2	0	0	0	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
3	C	1	0	0	0	0
3	D	1	0	0	0	0
3	E	1	0	0	0	0
3	F	1	0	0	0	0
3	G	1	0	0	0	0
3	H	1	0	0	0	0
All	All	32701	0	32261	1181	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 18.

The worst 5 of 1181 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:H:648:ASP:HB3	1:H:652:LEU:HB3	1.40	1.03
1:D:648:ASP:HB3	1:D:652:LEU:HB3	1.48	0.94
1:B:274:ASN:HD21	1:B:359:ARG:HB3	1.32	0.92
1:B:187:LEU:HD21	1:B:205:PRO:HG3	1.53	0.90
1:B:231:ASP:HB2	1:B:232:PRO:HD2	1.55	0.89

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	510/548 (93%)	450 (88%)	52 (10%)	8 (2%)	9	43
1	B	517/548 (94%)	448 (87%)	54 (10%)	15 (3%)	4	28
1	C	509/548 (93%)	446 (88%)	55 (11%)	8 (2%)	9	43
1	D	510/548 (93%)	449 (88%)	48 (9%)	13 (2%)	5	32
1	E	511/548 (93%)	447 (88%)	55 (11%)	9 (2%)	8	41
1	F	511/548 (93%)	446 (87%)	52 (10%)	13 (2%)	5	32
1	G	510/548 (93%)	448 (88%)	54 (11%)	8 (2%)	9	43
1	H	510/548 (93%)	447 (88%)	51 (10%)	12 (2%)	6	34
All	All	4088/4384 (93%)	3581 (88%)	421 (10%)	86 (2%)	7	37

5 of 86 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	713	ASN
1	B	181	ASP
1	B	325	SER
1	B	326	ASN
1	B	398	GLU

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	453/491 (92%)	443 (98%)	10 (2%)	52	79

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	450/491 (92%)	446 (99%)	4 (1%)	78	91
1	C	457/491 (93%)	448 (98%)	9 (2%)	55	80
1	D	452/491 (92%)	449 (99%)	3 (1%)	84	94
1	E	458/491 (93%)	449 (98%)	9 (2%)	55	80
1	F	460/491 (94%)	455 (99%)	5 (1%)	73	88
1	G	460/491 (94%)	452 (98%)	8 (2%)	60	83
1	H	465/491 (95%)	457 (98%)	8 (2%)	60	83
All	All	3655/3928 (93%)	3599 (98%)	56 (2%)	65	85

5 of 56 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	381	THR
1	H	723	ARG
1	F	356	ASP
1	H	688	TYR
1	H	356	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 59 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	389	GLN
1	H	485	GLN
1	E	709	ASN
1	H	483	GLN
1	G	597	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)

Of 16 ligands modelled in this entry, 16 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 (i)

6.1 Protein, DNA and RNA chains (i)

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, 95th 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	518/548 (94%)	-0.67	0 100 100	5, 45, 111, 207	0
1	B	523/548 (95%)	-0.57	2 (0%) 92 89	4, 55, 140, 335	0
1	C	517/548 (94%)	-0.66	0 100 100	8, 46, 110, 234	0
1	D	518/548 (94%)	-0.56	3 (0%) 89 83	8, 54, 136, 258	0
1	E	519/548 (94%)	-0.64	0 100 100	5, 46, 120, 198	0
1	F	519/548 (94%)	-0.55	3 (0%) 89 83	5, 55, 142, 205	0
1	G	518/548 (94%)	-0.64	1 (0%) 95 94	5, 46, 111, 264	0
1	H	518/548 (94%)	-0.57	2 (0%) 92 89	6, 54, 135, 221	0
All	All	4150/4384 (94%)	-0.61	11 (0%) 94 92	4, 50, 126, 335	0

The worst 5 of 11 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	688	TYR	3.3
1	F	732	TYR	2.8
1	B	688	TYR	2.5
1	G	426	ASP	2.5
1	H	732	TYR	2.5

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

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)

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, 95th 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
2	CA	B	736	1/1	0.95	0.17	56,56,56,56	0
2	CA	C	737	1/1	0.97	0.18	18,18,18,18	0
2	CA	H	737	1/1	0.97	0.15	21,21,21,21	0
2	CA	A	737	1/1	0.98	0.16	20,20,20,20	0
2	CA	D	736	1/1	0.98	0.15	44,44,44,44	0
2	CA	E	736	1/1	0.98	0.17	38,38,38,38	0
2	CA	F	736	1/1	0.98	0.17	47,47,47,47	0
2	CA	C	736	1/1	0.98	0.19	33,33,33,33	0
2	CA	B	737	1/1	0.99	0.15	20,20,20,20	0
2	CA	E	737	1/1	0.99	0.16	24,24,24,24	0
2	CA	A	736	1/1	0.99	0.16	35,35,35,35	0
2	CA	F	737	1/1	0.99	0.16	20,20,20,20	0
2	CA	G	736	1/1	0.99	0.20	40,40,40,40	0
2	CA	G	737	1/1	0.99	0.17	23,23,23,23	0
2	CA	H	736	1/1	0.99	0.15	37,37,37,37	0
2	CA	D	737	1/1	0.99	0.17	25,25,25,25	0

6.5 Other polymers (i)

There are no such residues in this entry.