Bioinformatic and mutational studies of related toxin–antitoxin pairs in *M. tuberculosis* predict and identify key functional residues

Running title: Toxin-Antitoxin relationships in *M. tuberculosis*

Himani Tandon[1], Arun Sharma[2], Saruchi Wadhwa[2], Raghavan Varadarajan[1], Ramandeep Singh[2], Narayanaswamy Srinivasan[1*] and Sankaran Sandhya[1*]

[1]Molecular Biophysics Unit, Indian Institute of Science, Bengaluru, 560012, India.
[2]Tuberculosis Research Laboratory, Vaccine and Infectious Disease Research Centre, Translational Health Science and Technology Institute, NCR Biotech Science Cluster, 3rd Milestone, PO Box # 4, Faridabad, Haryana- 121001, India.

* To whom correspondence should be addressed: Sankaran Sandhya:

Molecular Biophysics Unit, Indian Institute of Science, Bangalore- 560012; sandhyas@iisc.ac.in; Tel: +918022932837; Fax: +918023600535.

Correspondence may also be addressed to Narayanaswamy Srinivasan.

Molecular Biophysics Unit, Indian Institute of Science, Bangalore- 560012; ns@iisc.ac.in; Tel: +918022932837; Fax: +918023600535.

Keywords: Toxin-antitoxin, homology, paralogues, structure-function, *Mycobacterium tuberculosis*, genome analysis, bacteriostasis, VapBC, PIN domain, phylogeny, molecular modelling, protein evolution

Contents

Figure S1-S15

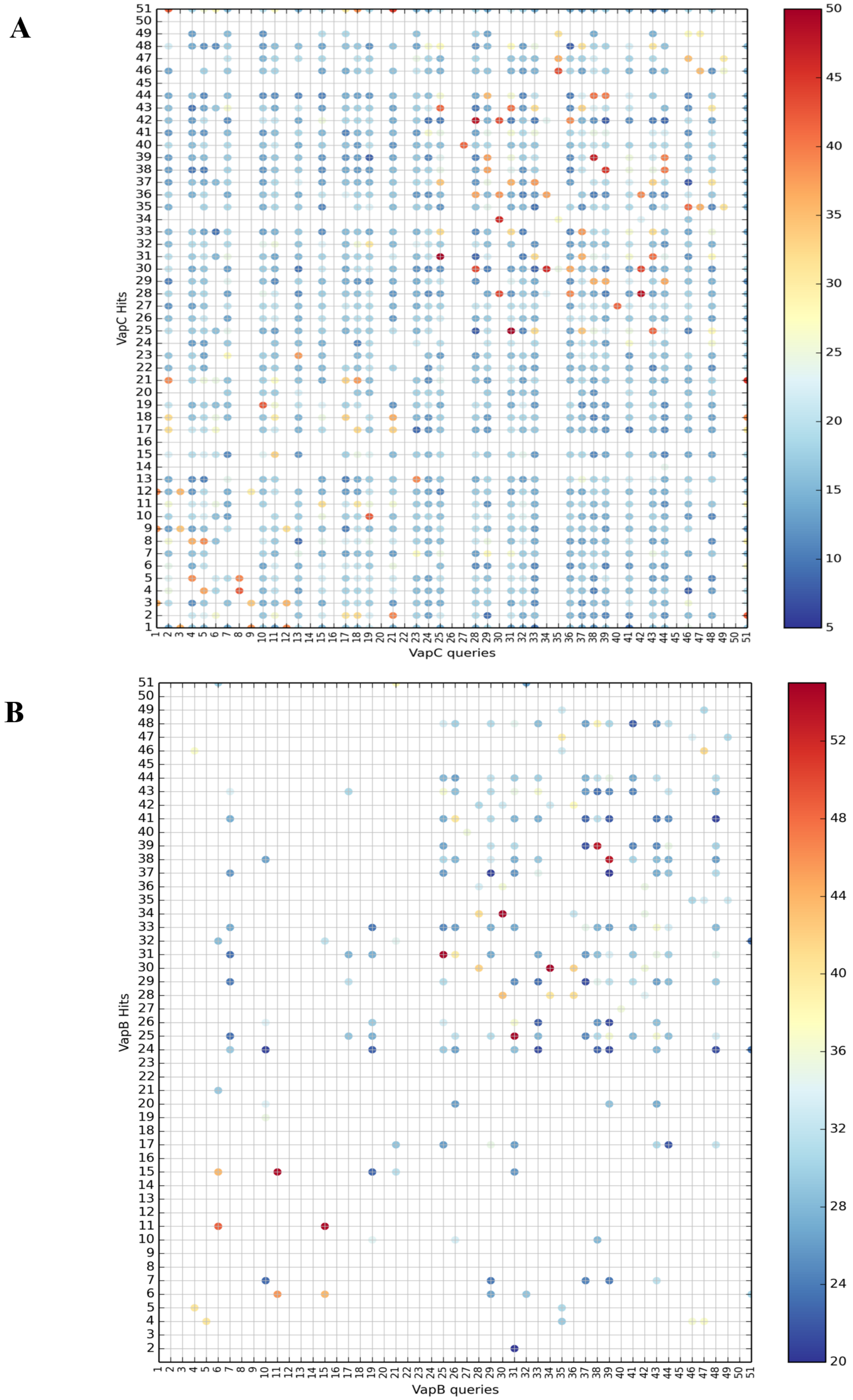


Figure S1: Paralogous VapBC TA within the *M. tuberculosis* genome: Plots show the hits for individual VapC toxin queries **(A)** and VapB antitoxin queries **(B)** in PSI-BLAST searches within the *M.tuberculosis* genome. Each dot represents a hit (on y-axis) for each query in the X-axis. For consideration as a hit, an e-value cut-off of 0.0001 was imposed in the PSI-BLAST searches. Searches were performed till convergence of the runs or a maximum of ten iterations. The color of each dot represents the percent sequence identity of a hit for the query. This ranges from blue-red depending on the sequence identity score. For the purpose of clarity, self-hits that fall along the diagonal of the plot are not reported in the figure and the sequence identity for toxins is capped at a maximum value of 50% and for antitoxins at 55%.

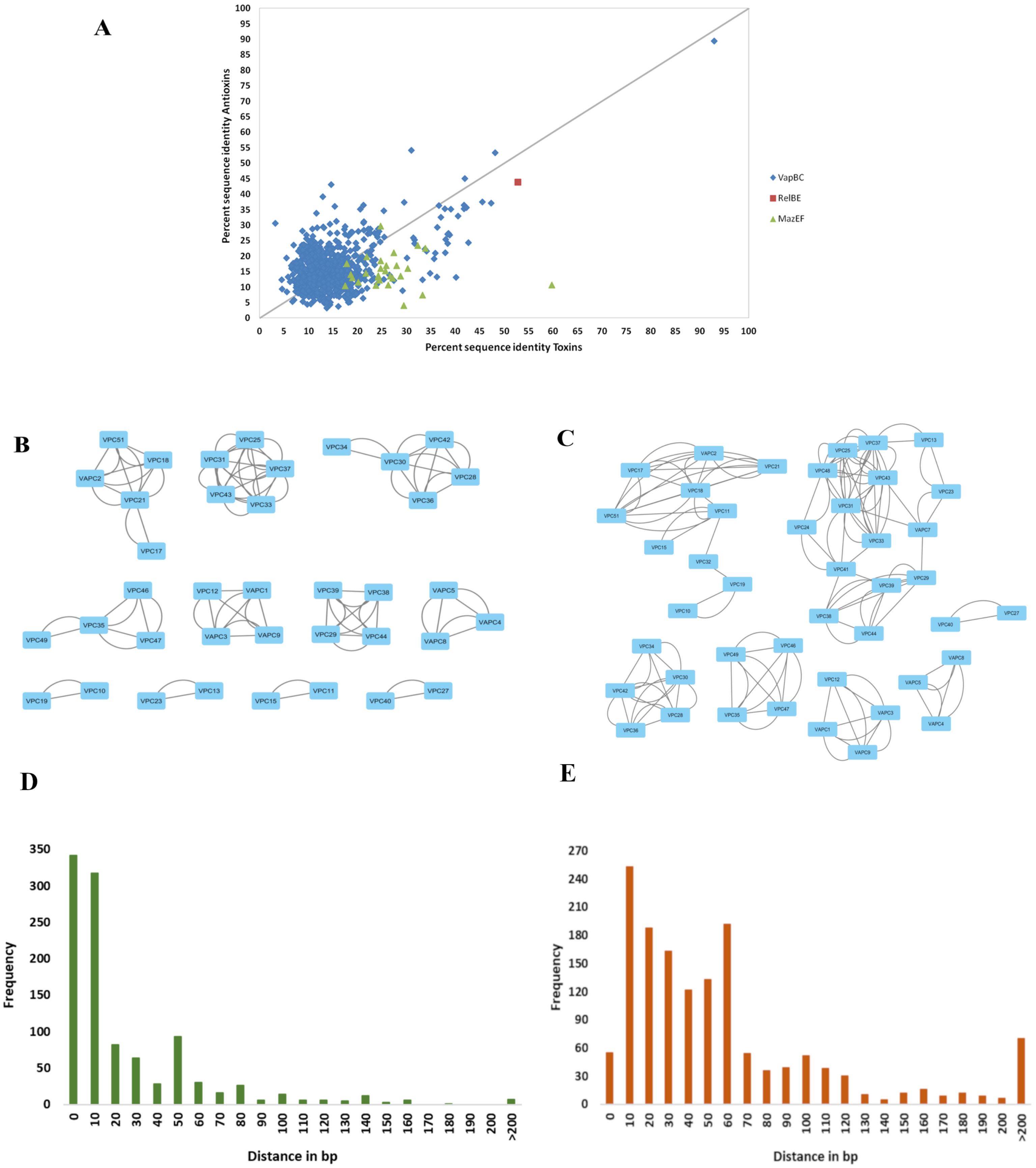


Figure S2 : **(A)** Scatter plot comparing the distribution of sequence identity between the various toxins along the X-axis and their cognate antitoxin pairs in the Y-axis. For VapBC toxins, for e.g, percentage sequence identity scores of each VapC toxin with every other VapC toxin is determined. Likewise, the sequence identity score of each VapB antitoxin with every other VapB antitoxin is determined. The points on the plot show the distribtion of scores for each such pair. VapBC scores are represented using blue diamonds, MazEF with grey triangles and RelBE with red squares. Clustering of *M. tuberculosis* VapC toxins at search thresholds of 60% query coverage and **(B)** 30% and **(C)** 25% sequence identity. 11 and 7 distinct clusters were formed separately. Histogram of distances (in bp) between the homologues of *M. tuberculosis* toxin and antitoxin detected for **(D)** mycobacteria and **(E)** other prokaryotic genomes.

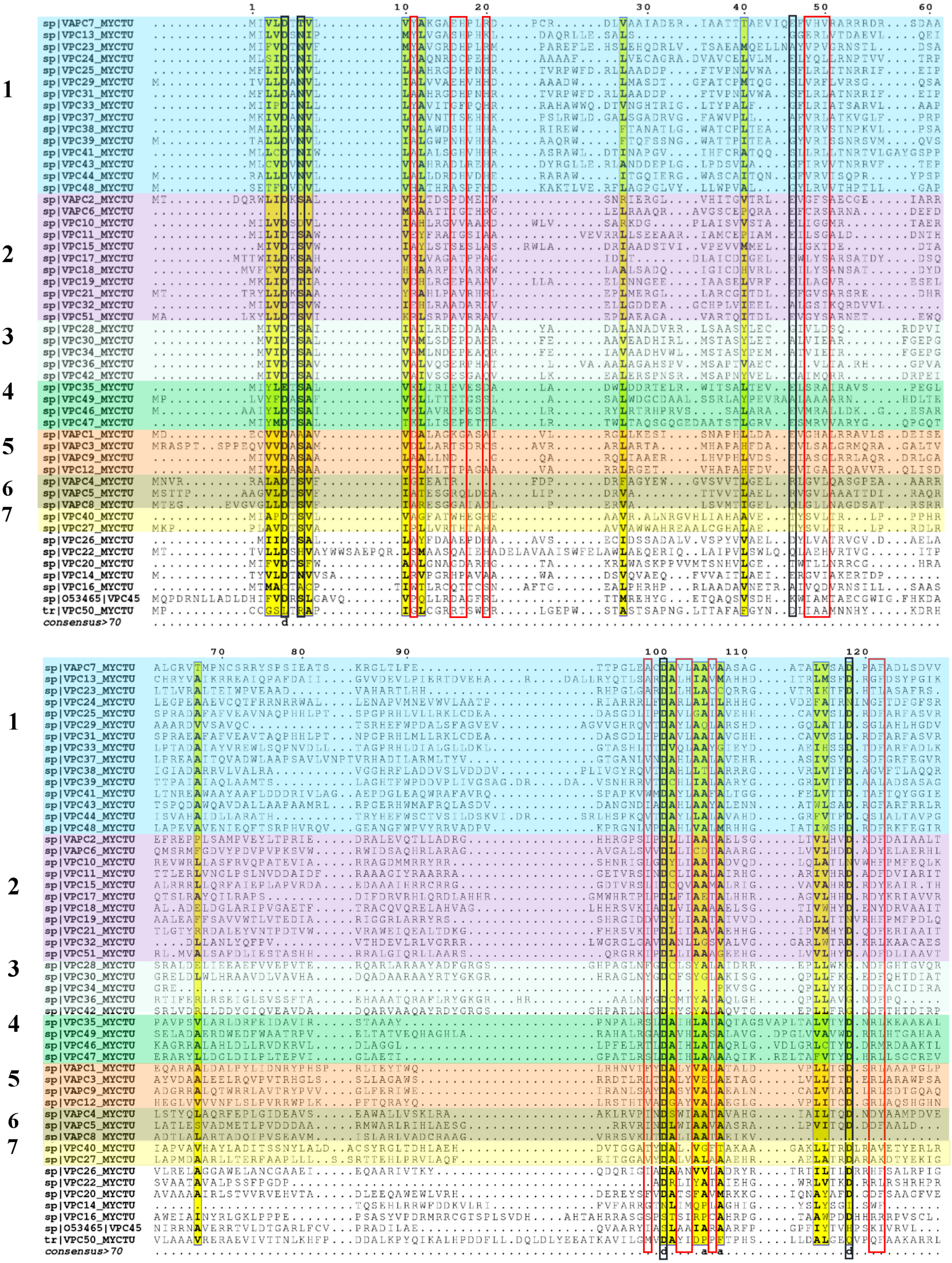
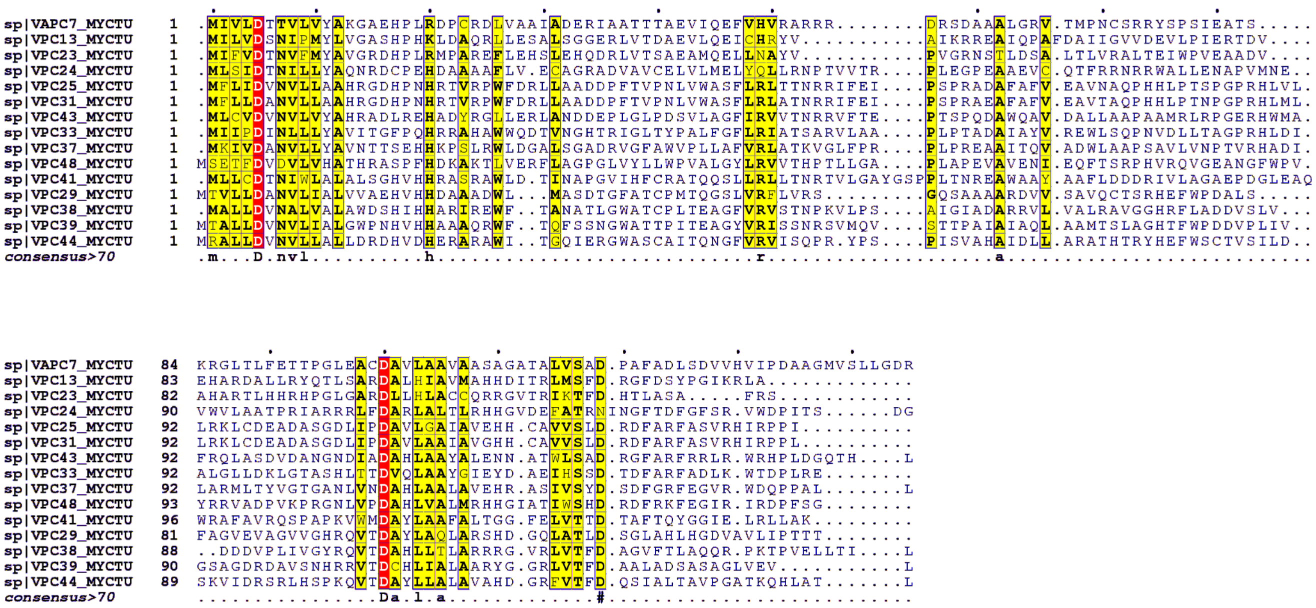


Figure S3 : Multiple sequence alignment of 51 VapC toxins of *M.tuberculosis* using MAFFT. Clusters are numbered from 1-7 and each cluster is indicated by a distinct color. The characteristic quartet of acidic residues of the PIN domain, that are well conserved across the cluster members, are shown marked in black boxes. Alignment positions, where residues that are fairly well conserved within a cluster but differ from residues in the same position in other clusters, are marked within a red box.

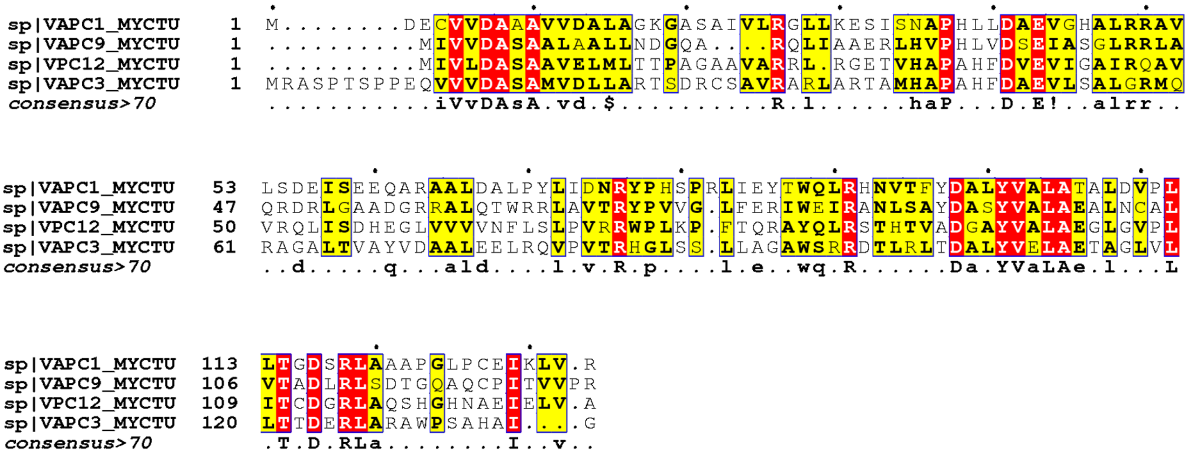


A Sub-cluster 1

```
sp|VAPC7_MYCTU    1  .MIVLDTTVLVYAKGAEHPLRDPCRDLVAAIADERIAATTTAEVIQEFVHVRARRR........DRSDAAALGRV.TMPNCSRRYSPSIEATS.....
sp|VPC13_MYCTU    1  .MILVDSNIPMYLVGASHPHKLDAQRLLESALSGGERLVTDAEVLQEICHRYV...........AIKRREAIQPAFDAIIGVVDEVLPIERTDV....
sp|VPC23_MYCTU    1  .MIFVDTNVFMYAVGRDHPLRMPAREFLEHSLEHQDRLVTSAEAMQELLNAYV...........PVGRNSTLDSA.LTLVRALTEIWPVEAADV....
sp|VPC24_MYCTU    1  .MLSIDTNILLYAQNRDCPEHDAAAAFLV.ECAGRADVAVCELVLMELYQLLRNPTVVTR....PLEGPEAAEVC.QTFRRNRRWALLENAPVMNE..
sp|VPC25_MYCTU    1  .MFLIDVNVLLAAHRGDHPNHRTVRPWFDRLLAADDPFTVPNLVWASFLRLTTNRRIFEI....PSPRADAFAFV.EAVNAQPHHLPTSPGPRHLVL.
sp|VPC31_MYCTU    1  .MFLLDANVLLAAHRGDHPNHRTVRPWFDRLLAADDPFTVPNLVWASFLRLATNRRIFEI....PSPRAEAFAFV.EAVTAQPHHLPTNPGPRHLML.
sp|VPC43_MYCTU    1  .MLCVDVNVLVYAHRADLREHADYRGLLERLANDDEPLGLPDSVLAGFIRVVTNRRVFTE....PTSPQDAWQAV.DALLAAPAAMRLRPGERHWMA.
sp|VPC33_MYCTU    1  .MIIPDINLLLYAVITGFPQHRRAHAWWQDTVNGHTRIGLTYPALFGFLRIATSARVLAA....PLPTADAIAYV.REWLSQPNVDLLTAGPRHLDI.
sp|VPC37_MYCTU    1  .MKIVDANVLLYAVNTTSEHHKPSLRWLDGALSGADRVGFAWVPLLAFVRLATKVGLFPR....PLPREAAITQV.ADWLAAPSAVLVNPTVRHADI.
sp|VPC48_MYCTU    1  MSETFDVDVLVHATHRASPFHDKAKTLVERFLAGPGLVYLLWPVALGYLRVVTHPTLLGA....PLAPEVAVENI.EQFTSRPHVRQVGEANGFWPV.
sp|VPC41_MYCTU    1  .MLLCDTNIWLALALSGHVHHRASRAWLD.TINAPGVIHFCRATQQSLLRLLTNRTVLGAYGSPPLTNREAWAAY.AAFLDDDRIVLAGAEPDGLEAQ
sp|VPC29_MYCTU    1  MTVLLDANVLIALVVAEHVHHDAAADWL...MASDTGFATCPMTQGSLVRFLVRS.........GQSAAAARDVV.SAVQCTSRHEFWPDALS.....
sp|VPC38_MYCTU    1  .MALLDVNALVALAWDSHIHHARIREWF..TANATLGWATCPLTEAGFVRVSTNPKVLPS....AIGIADARRVL.VALRAVGGHRFLADDVSLV...
sp|VPC39_MYCTU    1  MTALLDVNVLIALGWPNHVHHAAAQRWF..TQFSSNGWATTPITEAGYVRISSNRSVMQV....STTPAIAIAQL.AAMTSLAGHTFWPDDVPLIV..
sp|VPC44_MYCTU    1  MRALLDVNVLLALLDRDHVDHERARAWI..TGQIERGWASCAITQNGFVRVISQPR.YPS....PISVAHAIDLL.ARATHTRYHEFWSCTVSILD..
consensus>70         .m...D.nvl..........h............................r....................a...........................

sp|VAPC7_MYCTU   84  KRGLTLFETTPGLEACDAVLAAVAASAGATALVSAD.PAFADLSDVVHVIPDAAGMVSLLGDR
sp|VPC13_MYCTU   83  EHARDALLRYQTLSARDALHIAVMAHHDITRLMSFD.RGFDSYPGIKRLA.............
sp|VPC23_MYCTU   82  AHARTLHHRHPGLGARDLLHLACCQRRGVTRIKTFD.HTLASA....FRS.............
sp|VPC24_MYCTU   90  VWVLAATPRIARRRLFDARLALTLRHHGVDEFATRNINGFTDFGFSR.VWDPITS......DG
sp|VPC25_MYCTU   92  LRKLCDEADASGDLIPDAVLGAIAVEHH.CAVVSLD.RDFARFASVRHIRPPI..........
sp|VPC31_MYCTU   92  LRKLCDEADASGDLIPDAVLAAIAVGHH.CAVVSLD.RDFARFASVRHIRPPL..........
sp|VPC43_MYCTU   92  FRQLASDVDANGNDIADAHLAAYALENN.ATWLSAD.RGFARFRRLR.WRHPLDGQTH....L
sp|VPC33_MYCTU   92  ALGLLDKLGTASHLTTDVQLAAYGIEYD.AEIHSSD.TDFARFADLK.WTDPLRE........
sp|VPC37_MYCTU   92  LARMLTYVGTGANLVNDAHLAALAVEHR.ASIVSYD.SDFGRFEGVR.WDQPPAL.......L
sp|VPC48_MYCTU   93  YRRVADPVKPRGNLVPDAHLVALMRHHGIATIWSHD.RDFRKFEGIR.IRDPFSG........
sp|VPC41_MYCTU   96  WRAFAVRQSPAPKVWMDAYLAAFALTGG.FELVTTD.TAFTQYGGIE.LRLLAK.........
sp|VPC29_MYCTU   81  FAGVEVAGVVGHRQVTDAYLAQLARSHD.GQLATLD.SGLAHLHGDVAVLIPTTT........
sp|VPC38_MYCTU   88  ..DDDVPLIVGYRQVTDAHLLTLARRRG.VRLVTFD.AGVFTLAQQR.PKTPVELLTI....L
sp|VPC39_MYCTU   90  GSAGDRDAVSNHRRVTDCHLIALAARYG.GRLVTFD.AALADSASAGLVEV...........L
sp|VPC44_MYCTU   89  SKVIDRSRLHSPKQVTDAYLLALAVAHD.GRFVTFD.QSIALTAVPGATKQHLAT.......L
consensus>70         ................Da.l.a.............#...........................
```

B Sub-cluster 4

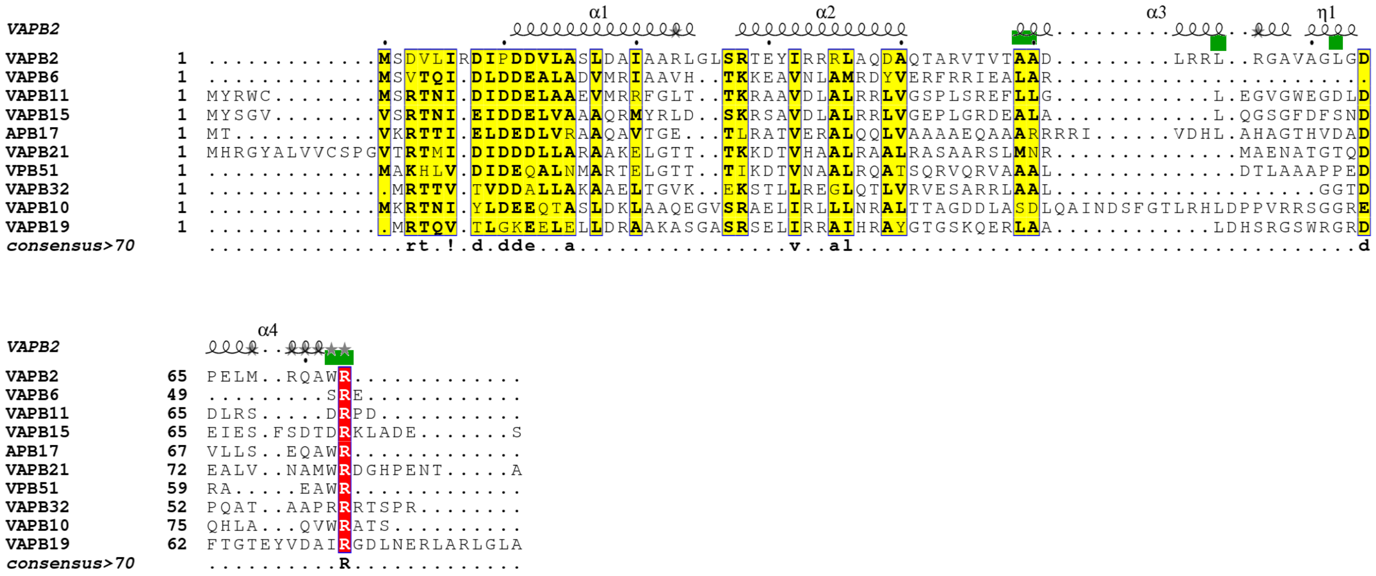
```
sp|VPC35_MYCTU    1  ..MIYLETSALVKLIRIEVESDALADWLD.......DRTELRWITSALTEVELSRAIRAV
sp|VPC46_MYCTU    1  MAAIYLDSSAIVKLAVREPESDALRRYL.........RTRHPRVSSALARAEVMRALLDK
sp|VPC47_MYCTU    1  ..MIYMDTSALTKLLISEPETTELRTWLTA.....QSGQGEDAATSTLGRVESMRVVARY
sp|VPC49_MYCTU    1  MPLVYFDASAFVKLLTTETGSSLASALWDGCDAALSSRLAYPEVRAALAAAARNHDLTES
consensus>70         ..m!Yl#.SA.vKL...E.es..l...l.........r.....v.saL...e..r.....

sp|VPC35_MYCTU   52  S.PEGLPAVPSVLARLDRFEIDAVIRSTAAAYPNP.ALRSLDAIHLATAQTAGSVAPLTA
sp|VPC46_MYCTU   52  G.ESARKAGRRALAHLDLLRVDKRVLDLAGGLLPF.ELRTLDAIHLATAQRLG..VDLGR
sp|VPC47_MYCTU   54  GQPGQTERARYLLDGLDILPLTEPVIGLAETIGPA.TLRSLDAIHLAAAAQIK..RELTA
sp|VPC49_MYCTU   61  ELADAERDWEDFWAATRPVELTATVEQHAGHLARAHALRGADAVHLASALAVG..DPGLV
consensus>70         ............la.ld.......!...A........LR.lDA!HLA.A...g....l..

sp|VPC35_MYCTU  110  LVTYDNRLKEAAEALSLAVVAPGQAR.
sp|VPC46_MYCTU  108  LCTYDDRMRDAAKTLGMAVIAPS....
sp|VPC47_MYCTU  111  FVTYDHRLLSGCREVGFVTASPGAVR.
sp|VPC49_MYCTU  119  VAVWDRRLHTGAHAAG.CRVAPAQLDP
consensus>70         ..tyD.R$...a...g...vaP.....
```



C Sub-cluster 5

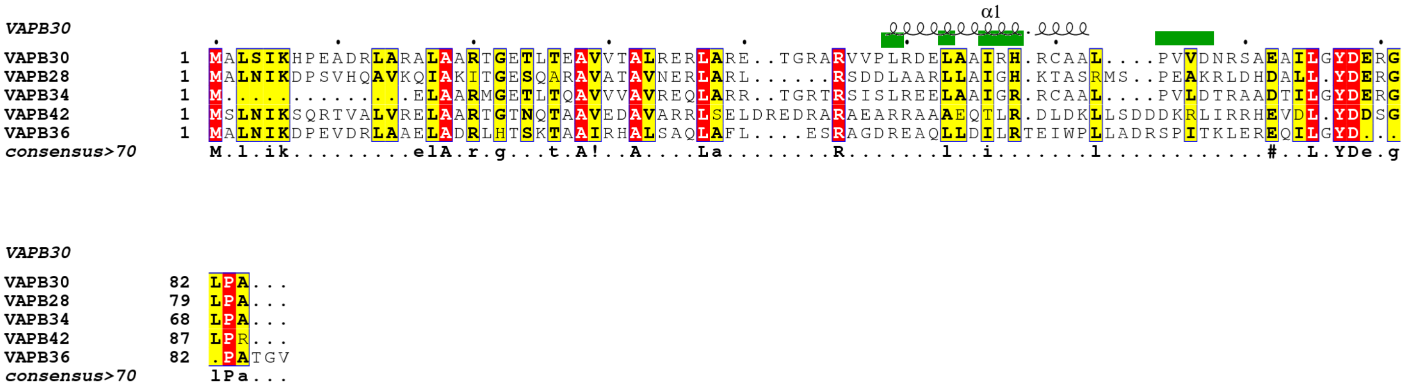
```
sp|VAPC1_MYCTU    1  M........DECVVDAAAVVDALAGKGASAIVLRGLLKESISNAPHLLDAEVGHALRRAV
sp|VAPC9_MYCTU    1  ..........MIVVDASAALAALLNDGQA....RQLIAAERLHVPHLVDSEIASGLRRLA
sp|VPC12_MYCTU    1  ..........MIVLDASAAVELMLTTPAGAAVARRL.RGETVHAPAHFDVEVIGAIRQAV
sp|VAPC3_MYCTU    1  MRASPTSPPEQVVVDASAMVDLLARTSDRCSAVRARLARTAMHAPAHFDAEVLSALGRMQ
consensus>70         ...........iVvDAsA.vd.$..........R.l......haP...D.E!..alrr..

sp|VAPC1_MYCTU   53  LSDEISEEQARAALDALPYLIDNRYPHSPRLIEYTWQLRHNVTFYDALYVALATALDVPL
sp|VAPC9_MYCTU   47  QRDRLGAADGRRALQTWRRLAVTRYPVVG.LFERIWEIRANLSAYDASYVALAEALNCAL
sp|VPC12_MYCTU   50  VRQLISDHEGLVVVVNFLSLPVRRWPLKP.FTQRAYQLRSTHTVADGAYVALAEGLGVPL
sp|VAPC3_MYCTU   61  RAGALTVAYVDAALEELRQVPVTRHGLSS.LLAGAWSRRDTLRLTDALYVELAETAGLVL
consensus>70         ..d.....q...ald....l.v.R.p....l.e..wq.R......Da.YVaLAe.l...L

sp|VAPC1_MYCTU  113  LTGDSRLAAAPGLPCEIKLV.R
sp|VAPC9_MYCTU  106  VTADLRLSDTGQAQCPITVVPR
sp|VPC12_MYCTU  109  ITCDGRLAQSHGHNAEIELV.A
sp|VAPC3_MYCTU  120  LTTDERLARAWPSAHAI....G
consensus>70         .T.D.RLa........I..v..
```

D Sub-cluster 2 (Antitoxin)



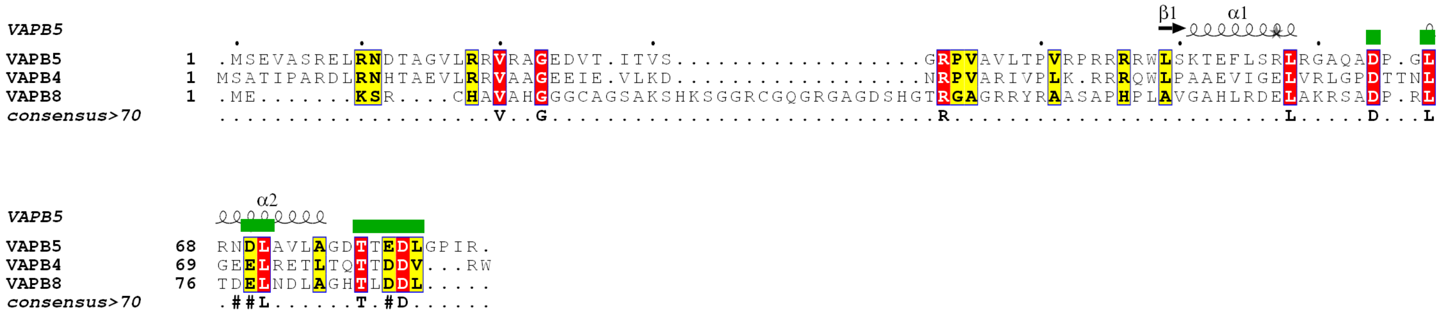
```
VAPB2
VAPB2         1 .............MSDVLIRDIPDDVLASLDAIAARLGLSRTEYIRRRLAQDAQTARVTVTAAD.........LRRL..RGAVAGLGD
VAPB6         1 .............MSVTQI.DLDDEALADVMRIAAVH..TKKEAVNLAMRDYVERFRRIEALAR........................
VAPB11        1 MYRWC........MSRTNI.DIDDELAAEVMRRFGLT..TKRAAVDLALRRLVGSPLSREFLLG............L.EGVGWEGDLD
VAPB15        1 MYSGV........VSRTNI.EIDDELVAAAQRMYRLD..SKRSAVDLALRRLVGEPLGRDEALA............L.QGSGFDFSND
APB17         1 MT...........VKRTTI.ELDEDLVRAAQAVTGE...TLRATVERALQQLVAAAAEQAAARRRRI......VDHL.AHAGTHVDAD
VAPB21        1 MHRGYALVVCSPGVTRTMI.DIDDDLLARAAKELGTT..TKKDTVHAALRAALRASAARSLMNR..............MAENATGTQD
VPB51         1 .............MAKHLV.DIDEQALNMARTELGTT..TIKDTVNAALRQATSQRVQRVAAAL..............DTLAAAPPED
VAPB32        1 ..............MRTTV.TVDDALLAKAAELTGVK..EKSTLLREGLQTLVRVESARRLAAL....................GGTD
VAPB10        1 .............MKRTNI.YLDEEQTASLDKLAAQEGVSRAELIRLLLNRALTTAGDDLASDLQAINDSFGTLRHLDPPVRRSGGRE
VAPB19        1 ..............MRTQV.TLGKEELELLDRAAKASGASRSELIRRAIHRAYGTGSKQERLAA............LDHSRGSWRGRD
consensus>70    ...............rt.!.d.dde..a................v..al......................................d

VAPB2
VAPB2        65 PELM..RQAWR.............
VAPB6        49 .........SRE............
VAPB11       65 DLRS.....DRPD...........
VAPB15       65 EIES.FSDTDRKLADE.......S
APB17        67 VLLS..EQAWR.............
VAPB21       72 EALV..NAMWRDGHPENT.....A
VPB51        59 RA.....EAWR.............
VAPB32       52 PQAT..AAPRRRTSPR........
VAPB10       75 QHLA...QVWRATS..........
VAPB19       62 FTGTEYVDAIRGDLNERLARLGLA
consensus>70    ..........R.............
```

E Sub-cluster 3 (Antitoxin)



```
VAPB30
VAPB30        1 MALSIKHPEADRLARALAARTGETLTEAVVTALRERLARE..TGRARVVPLRDELAAIRH.RCAAL....PVVDNRSAEAILGYDERG
VAPB28        1 MALNIKDPSVHQAVKQIAKITGESQARAVATAVNERLARL......RSDDLAARLLAIGH.KTASRMS..PEAKRLDHDALL.YDERG
VAPB34        1 M..............ELAARMGETLTQAVVVAVREQLARR..TGRTRSISLREELAAIGR.RCAAL....PVLDTRAADTILGYDERG
VAPB42        1 MSLNIKSQRTVALVRELAARTGTNQTAAVEDAVARRLSELDREDRARAEARRAAAEQTLR.DLDKLLSDDDKRLIRRHEVDL.YDDSG
VAPB36        1 MALNIKDPEVDRLAAELADRLHTSKTAAIRHALSAQLAFL....ESRAGDREAQLLDILRTEIWPLLADRSPITKLEREQILGYD...
consensus>70    M.l.ik.........elA.r.g...t.A!..A....La........R.......l..i.......l............#..L.YDe.g

VAPB30
VAPB30       82 LPA...
VAPB28       79 LPA...
VAPB34       68 LPA...
VAPB42       87 LPR...
VAPB36       82 .PATGV
consensus>70    lPa...
```

F Sub-cluster 6 (Antitoxin)



```
VAPB5
VAPB5         1 .MSEVASRELRNDTAGVLRRVRAGEDVT.ITVS..................GRPVAVLTPVRPRRRRWLSKTEFLSRLRGAQADP.GL
VAPB4         1 MSATIPARDLRNHTAEVLRRVAAGEEIE.VLKD..................NRPVARIVPLK.RRRQWLPAAEVIGELVRLGPDTTNL
VAPB8         1 .ME.......KSR....CHAVAHGGGCAGSAKSHKSGGRCGQGRGAGDSHGTRGAGRRYRAASAPHPLAVGAHLRDELAKRSADP.RL
consensus>70    ....................V..G............................R........................L.....D...L

VAPB5
VAPB5        68 RNDLAVLAGDTTEDLGPIR.
VAPB4        69 GEELRETLTQTTDDV...RW
VAPB8        76 TDELNDLAGHTLDDL.....
consensus>70    .##L......T.#D......
```

Figure S4: Multiple sequence alignments of VapC toxins from 4 of the 7 clusters shown in Figure 1 are shown in panels **(A)** to **(C)**. Conserved residues are highlghted in red while conservatively substituted residues are highlighted in yellow and in bold font. **(D)** to **(F)** show the alignment of VapB antitoxins as suggested by Figure 1, with toxin-binding residues indicated in green.

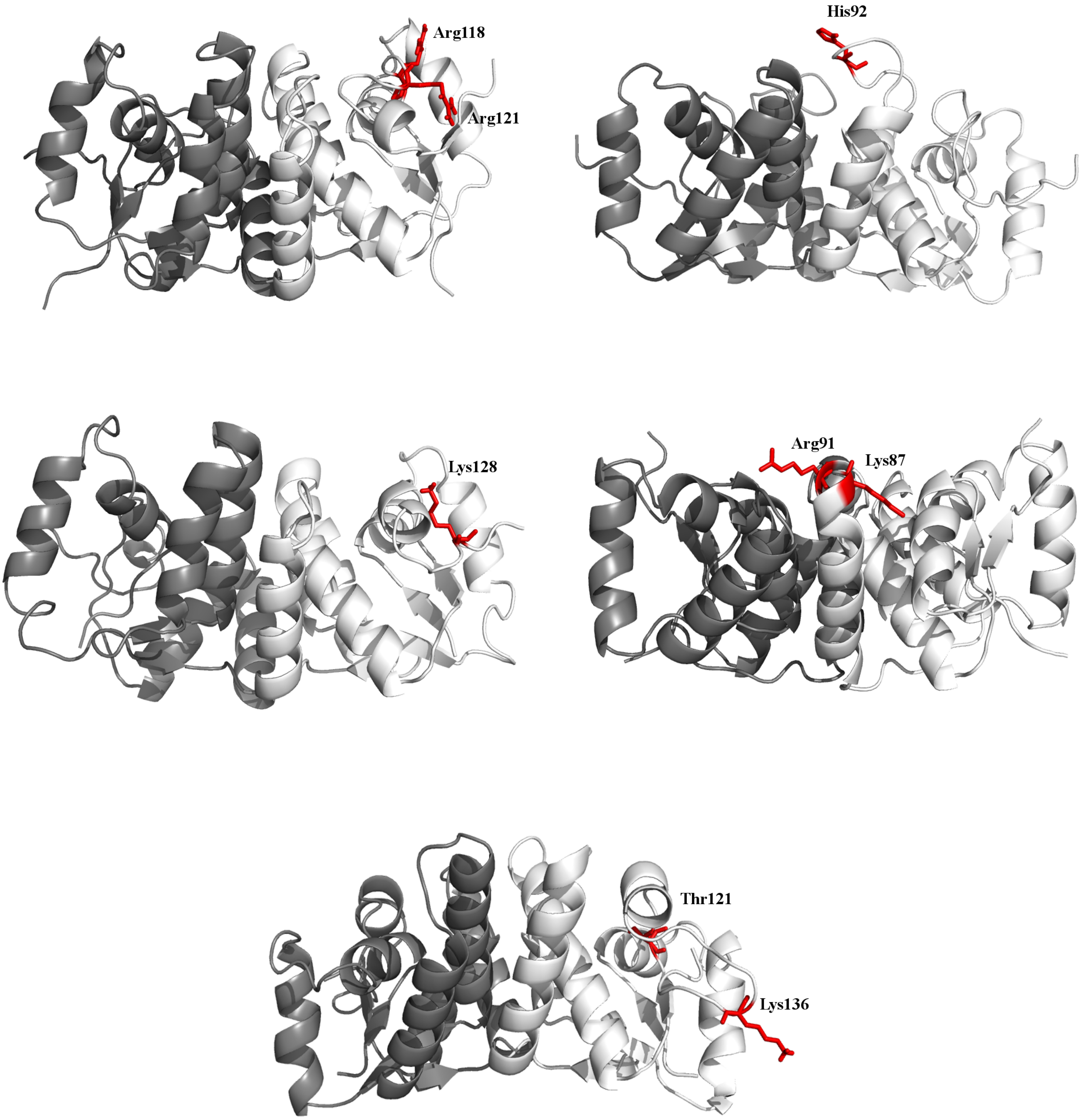


Figure S5: Structural models for various toxins used for computational alanine scanning. The models for Rv0528c, Rv0609, Rv1838c and Rv3320c were generated by modeller v9.14 using templates identified by Phyre2. The structure for Rv2757c was obtained from PDB (5sv2). The residues predicted to be important for toxin funciton that were mutated in experiments are shown in red sticks.

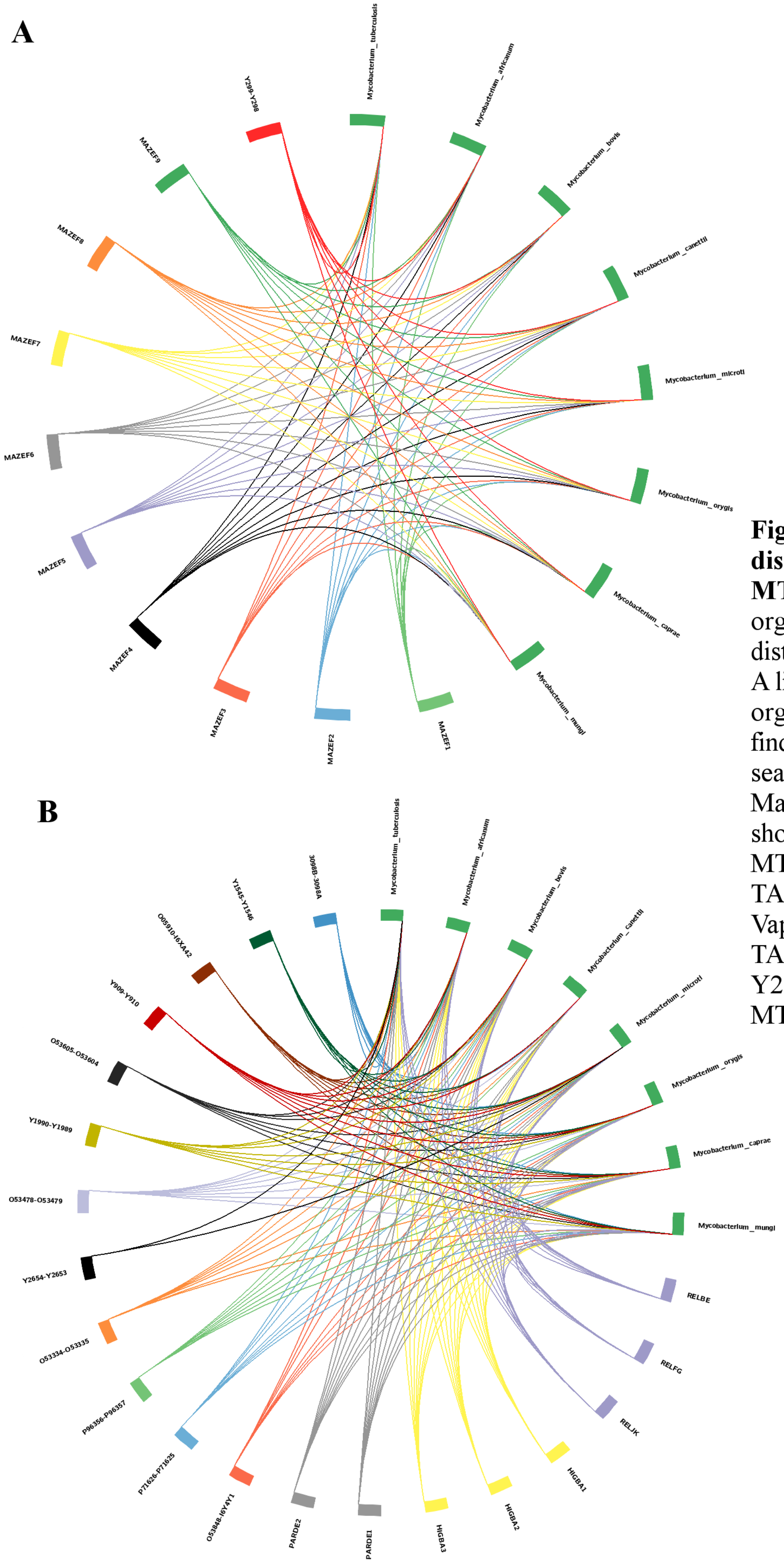


Figure S6: Circos plots showing the distribution of *M. tuberculosis* TA pairs in MTBC. In the plots, the nodes represent organism names and TA types that are grouped distinctly to facilitate interpretation of the figures. A line is drawn between a TA node and that of an organism if both the toxin and antitoxin query find a homologue in PSI-BLAST and TBLASTN searches. **(A)** shows the distribution of 10 MazEFs in the 8 MTBC of this study. The plots show that these TA are well-conserved in the MTBC (Table S6). **(B)** shows the distribution of TA systems in MTBC for query types other than VapBC and MazEF. As shown here, not all such TA are conserved as TA pairs in MTBC. E.g. Y2654-Y2653 is conserved only in 2 of the 8 MTBC in the dataset.

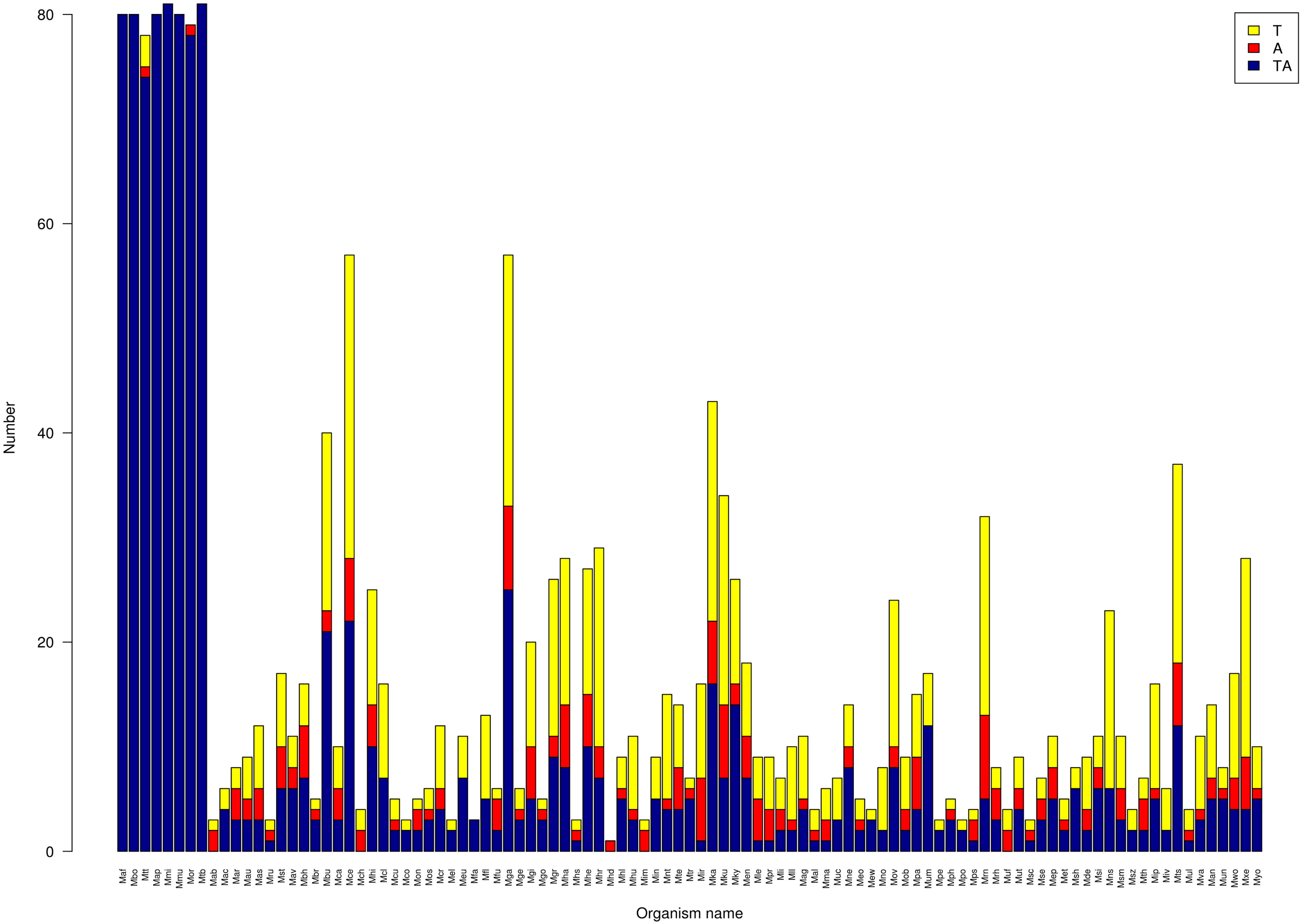
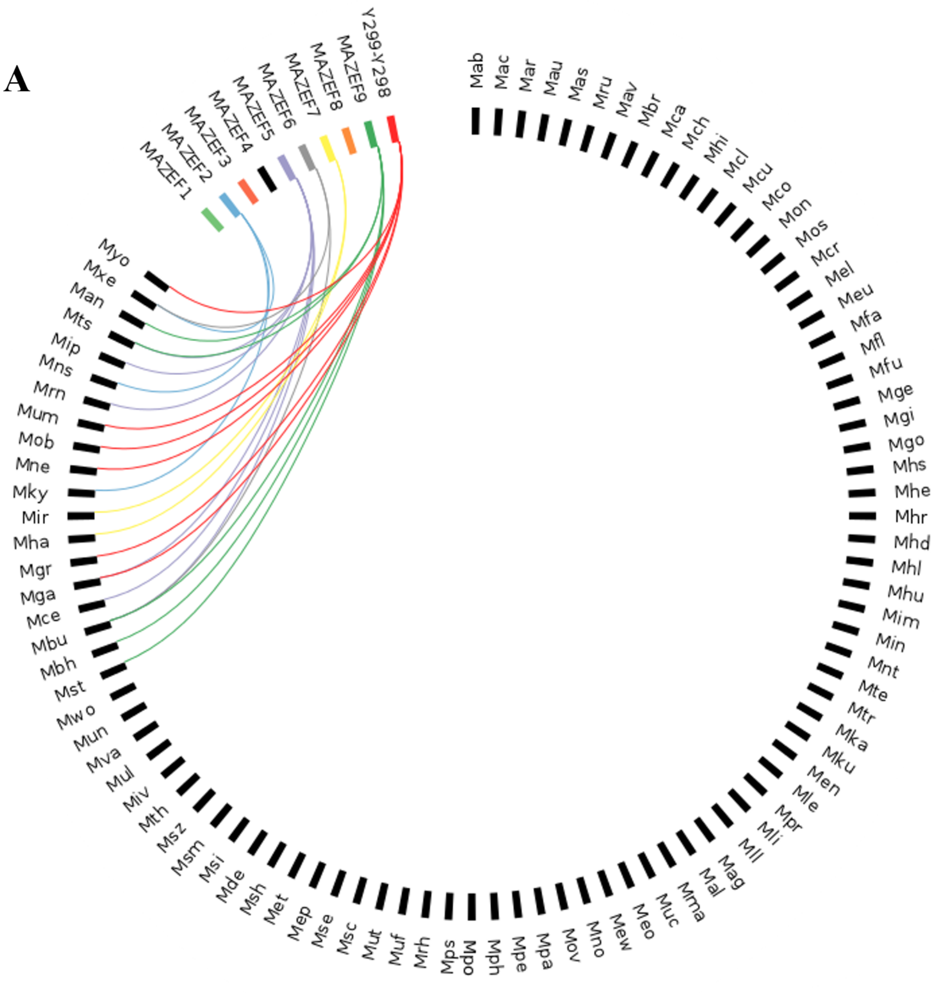


Figure S7: Plot showing the distribution of *M.tuberculosis* TA in the genus Mycobacterium. Relative abundance of the toxin and antitoxin homologues in the genus Mycobacterium. The Mycobacteria species employed in this study are listed on the X-axis using three letter codes to designate the organism (Mapping file of the code to organism name is in Table S7). The Y-axis shows the number of toxin/antitoxin/ TA pairs that were detected in the PSI-BLAST searches performed in the various genomes. Members of the MTBC complex are listed first.

A



B

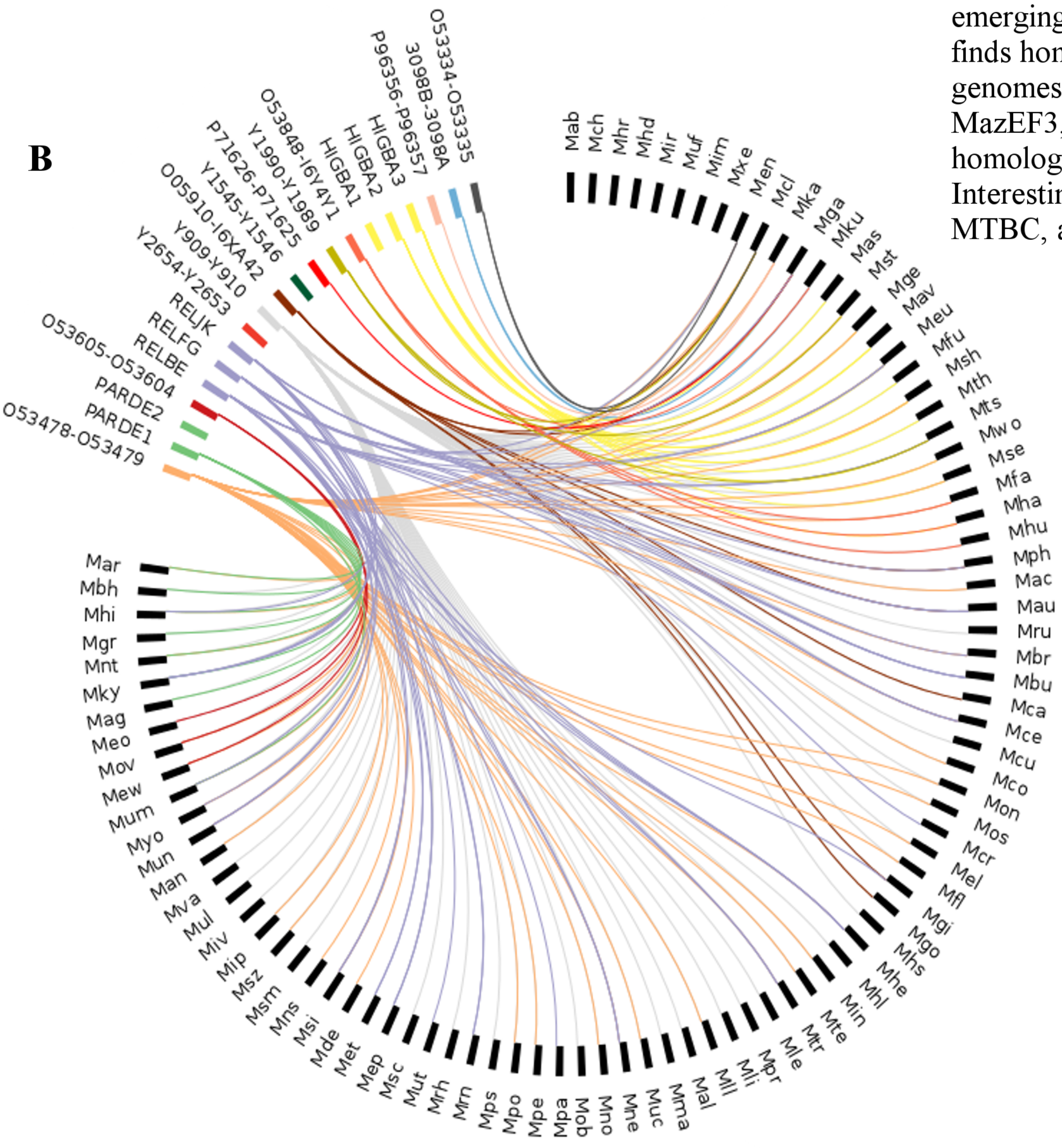


Figure S8: Circos plots showing the distribution of MazEF *M. tuberculosis* TA pairs in 93 Mycobacterial genomes excluding MTBC in **(A)** and the distribution of, all other TA in other prokaryotic genomes is shown in **(B)**.Organism names are indicated with three-letter codes as listed in Table S7. The graphs are useful to visually appreciate the spread of homologues in the various genomes studied here. For example, in (A), dense lines emerging from MazEF9 show that this TA finds homologues in a few mycobacterial genomes surveyed here. In contrast, MazEF3, MazEF8 and others do not find homologues in any of the organisms. Interestingly, all queries find homologues in MTBC, as shown in Figure S6.

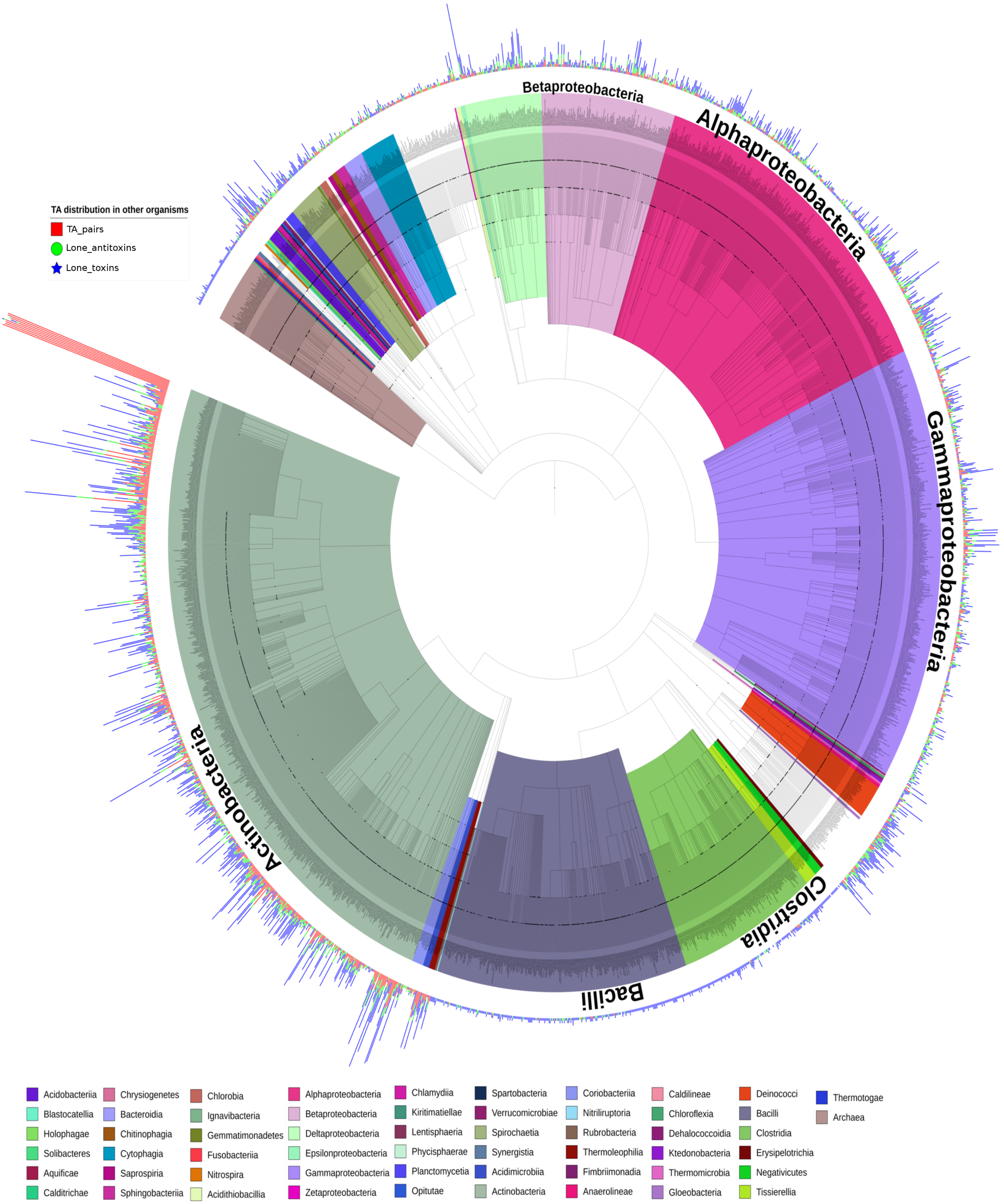


Figure S9: Phylogenetic tree showing distribution of *M.tuberculosis* TA across prokaryotic taxa. 2220 genomes containing homologues of TAs have been grouped into 54 classes shown in different colors. The organisms that are not shown in color have not been assigned a class according to NCBI taxonomy database. Outer bar-plot shows the number of *M.tuberculosis* TA homologs found in these organisms. An interactive link to this tree can be accessed at https://itol.embl.de/tree/1413912822138041518174018 where the user can hover over the tree branches for detailed data visualization. For best visibility, users are advised to visualize the tree in Firefox browser.

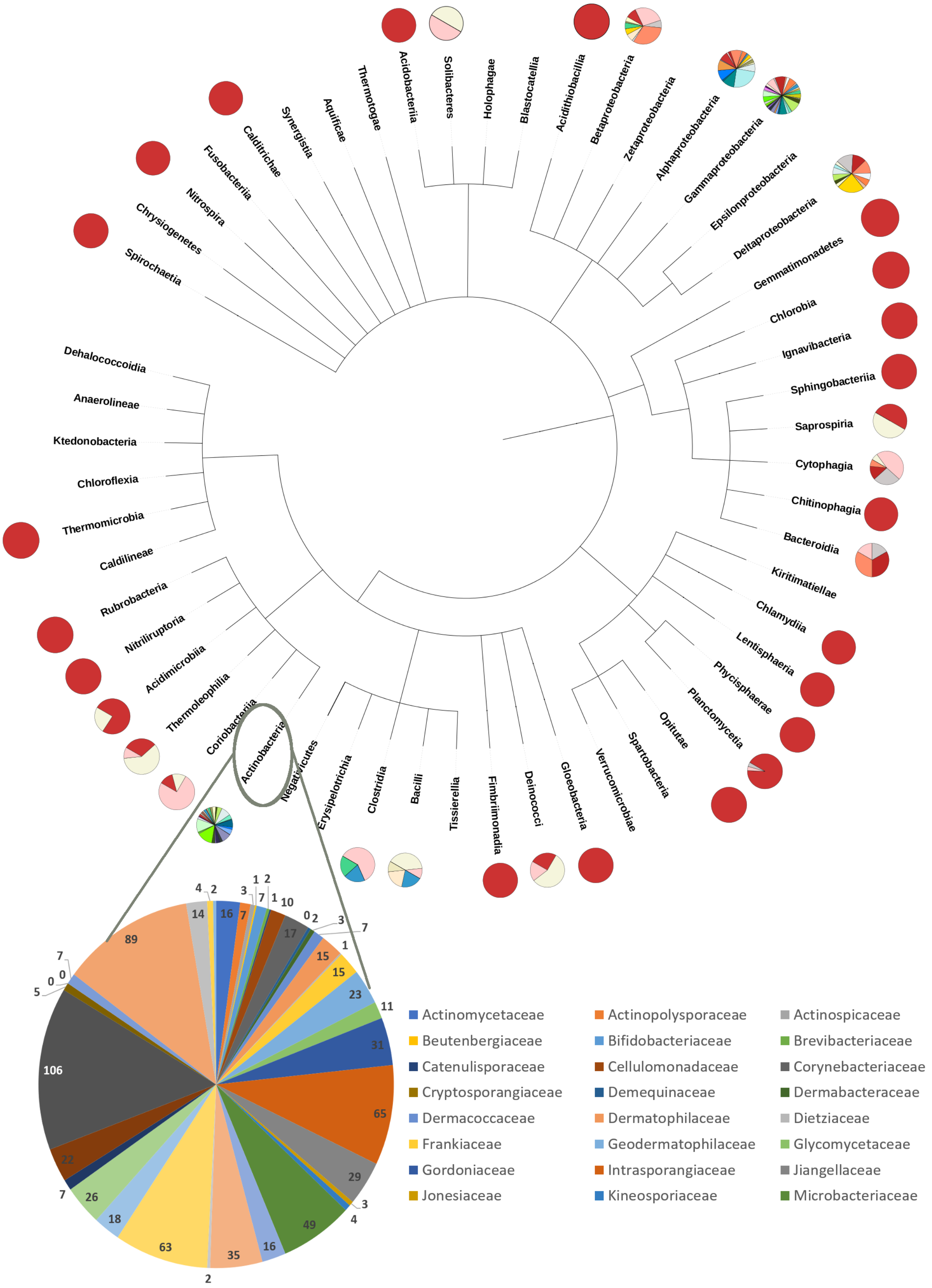
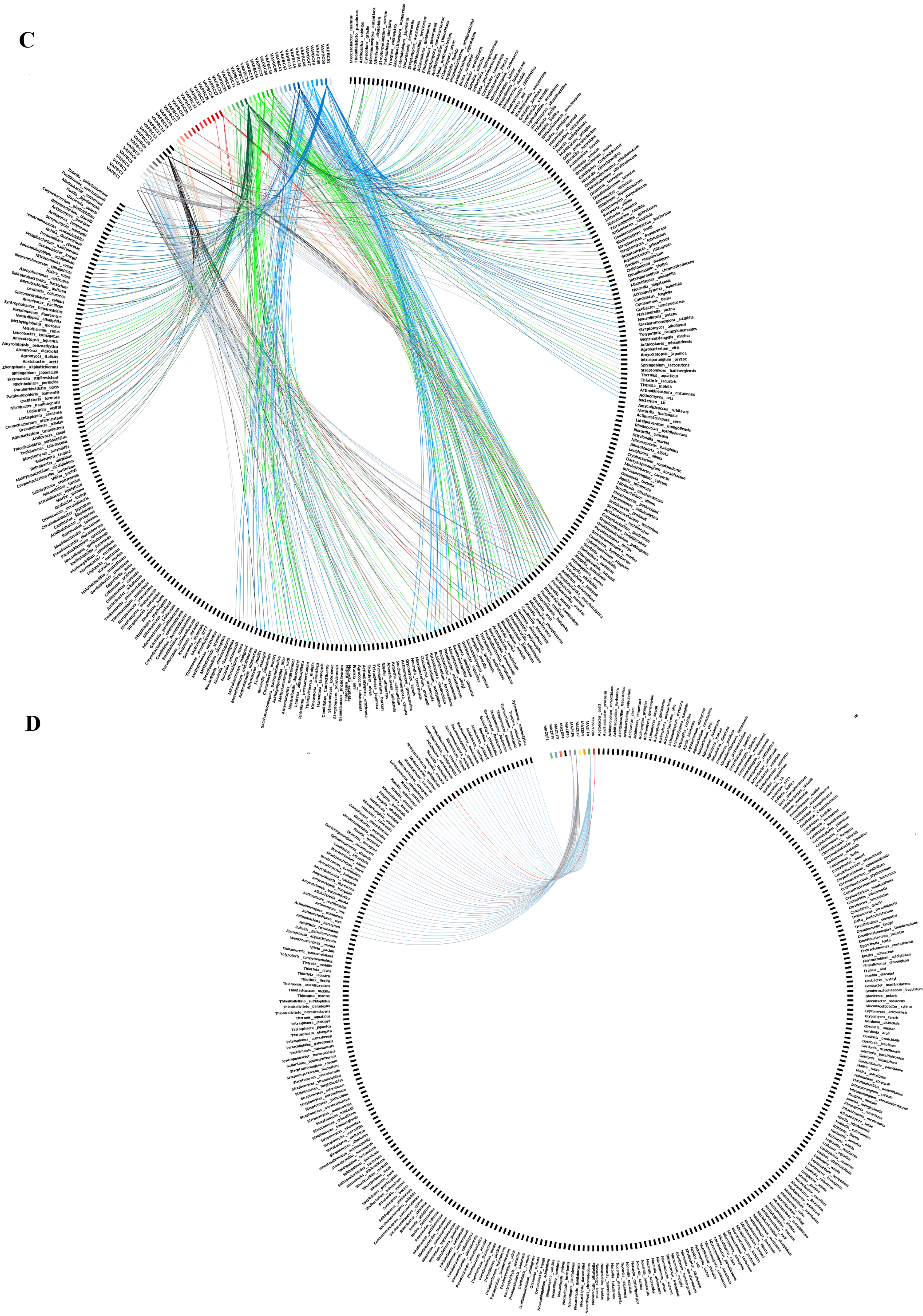


Figure S10: Distribution of *M. tuberculosis* TA in various prokaryotes, organized in a class-specific manner. Unrooted tree with class-specific pie-charts on the nodes provide the number of TA pairs in each constituent family. As an example, inset shows the number of TA pairs in various families of class Actinobacteria. An interactive link to this tree can be accessed at https://itol.embl.de/tree/1413912822265121518597108 where user can hover over the pie-charts for families in various classes. For best visibility, users are advised to visualize the tree in Firefox browser, setting the font style to 70.


C
D

E

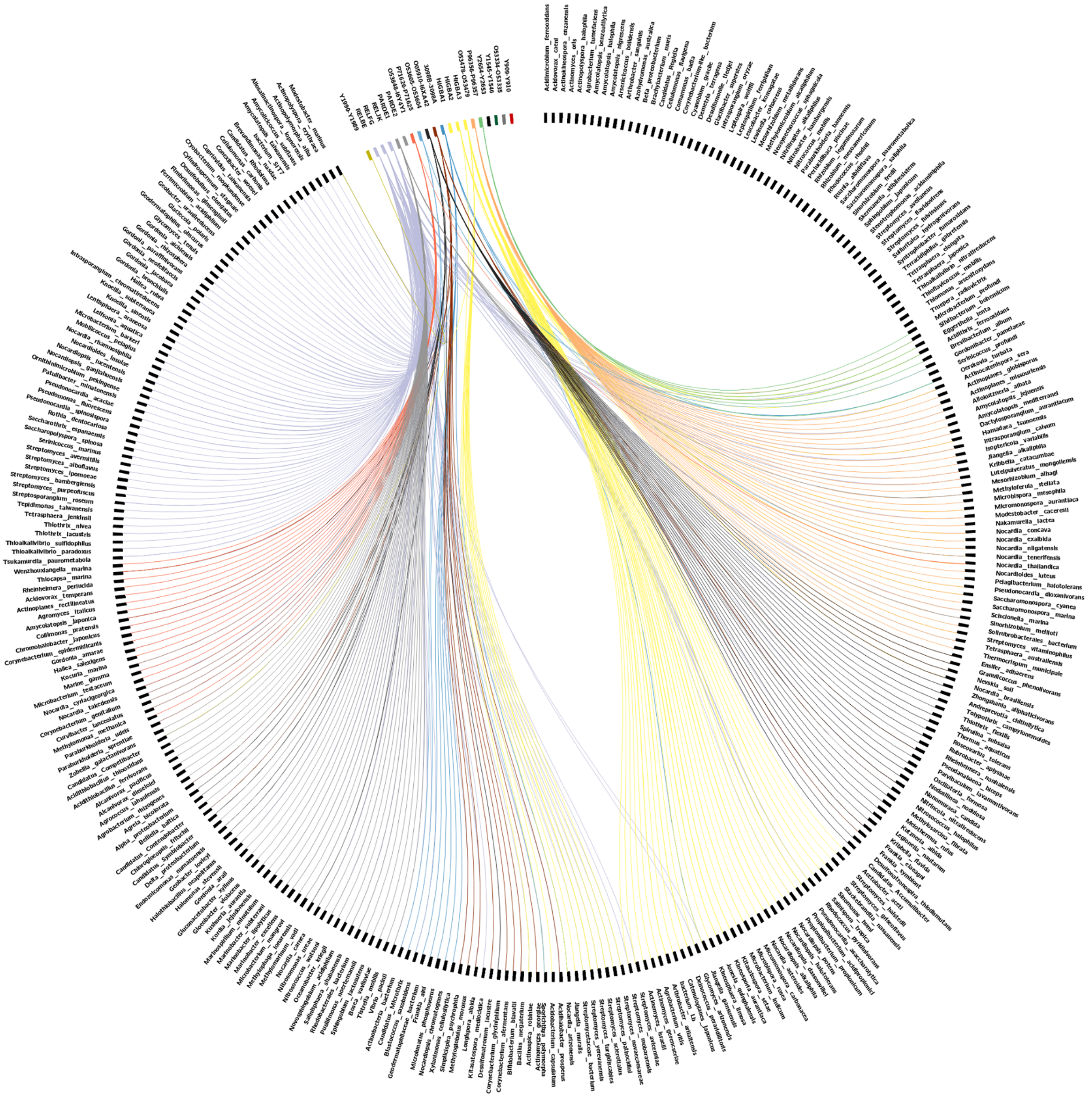


Figure S11: Circos plots showing the distribution of *M. tuberculosis* TA pairs in all 2220 prokaryotes. (A-C) show the distribution of VapBC, MazEF and other TA respectively. Plots help infer that some TA such as VapBC32 are extensively well conserved in many of the prokaryotes surveyed here. The number of homologues of MazEF TA system in the various prokaryotes is far fewer as shown in **(B)**. As shown here, Rv0909-Rv0910 is the only TA although universally present in mycobacteria is absent in prokaryotes **(C)**.

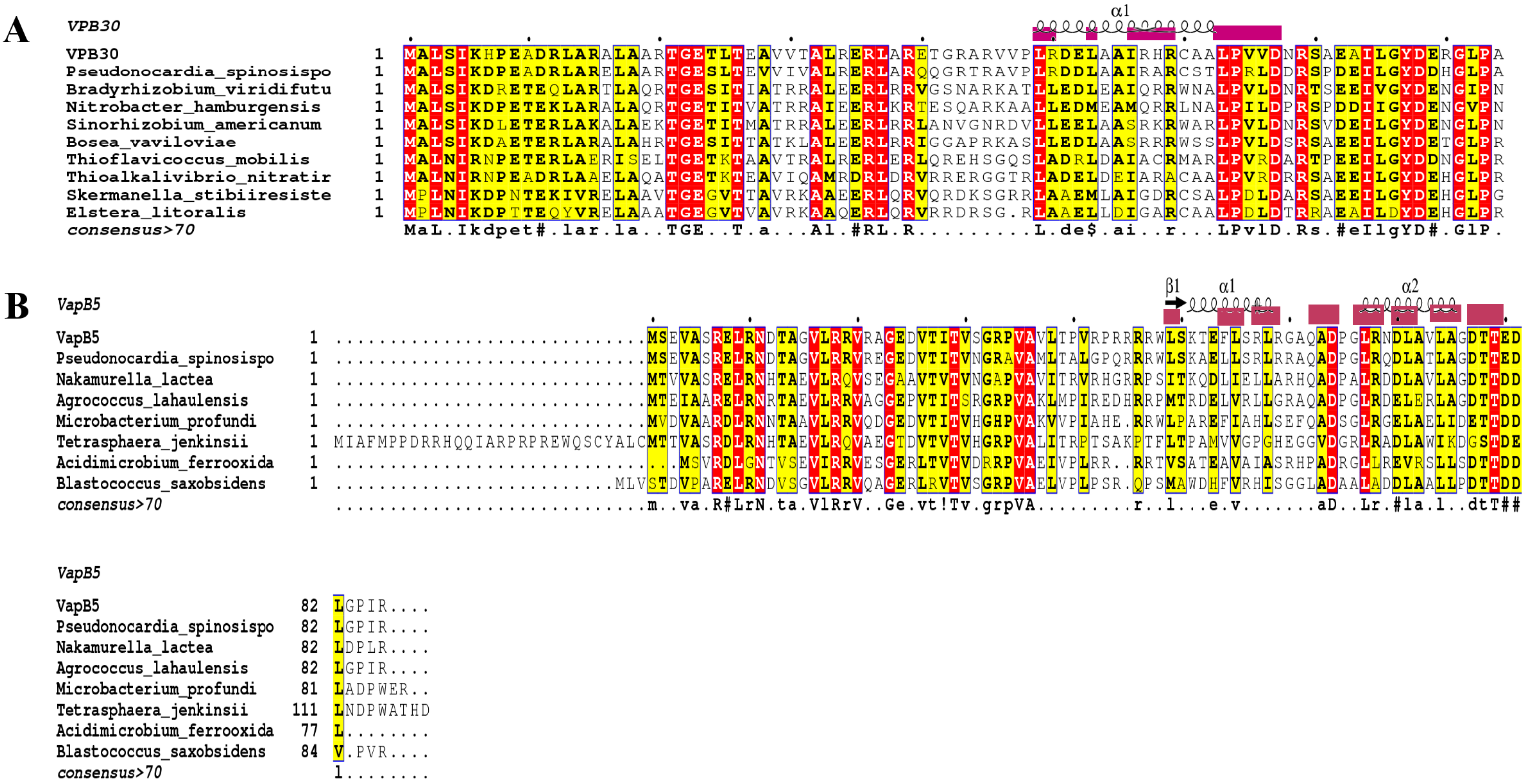


Figure S12: Multiple sequence alignment of *M.tuberculosis* **(A)** VapB30 (4xgq) with top 10 orthologues from the prokaryotic genomes. **(B)** VapB5 (3dbo) with top 10 orthologues from the prokaryotic genomes. The columns shaded dark red show conserved residues while columns shaded yellow show conservatively substituted residues. Secondary structure elements of the reference structures are depicted on top of each alignment. Pink colored boxes on top of the alignment are the toxin-binding residues in the reference structures. Consensus sequence (at a threshold > 70%) for the alignment is shown at the bottom of the alignment.

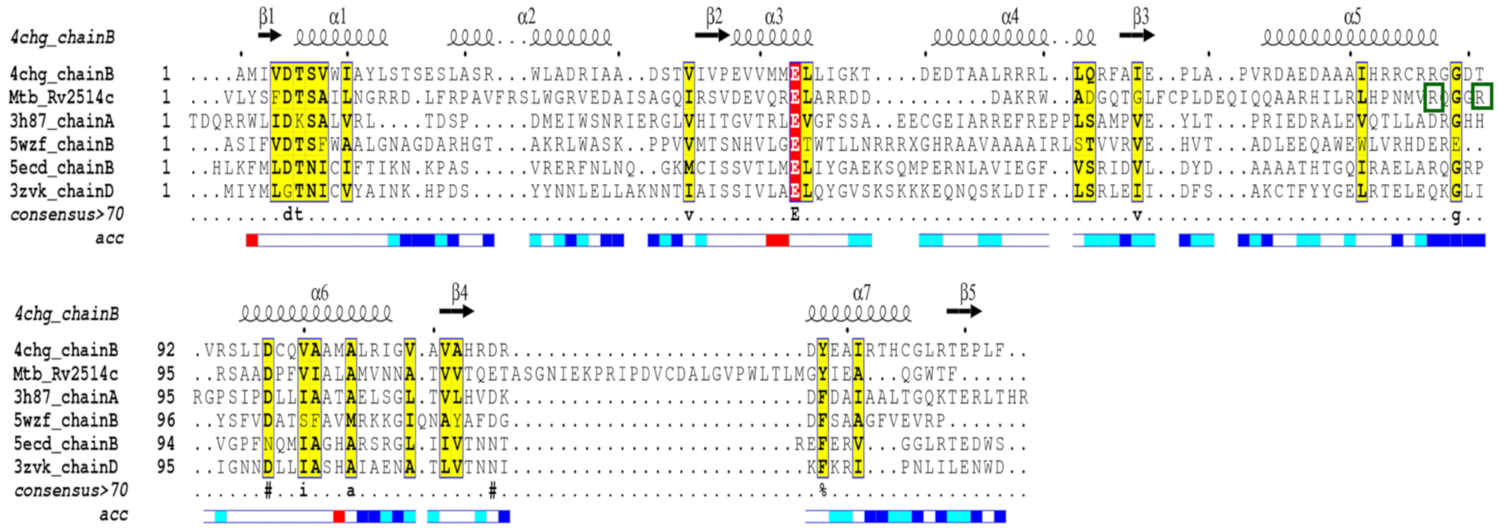
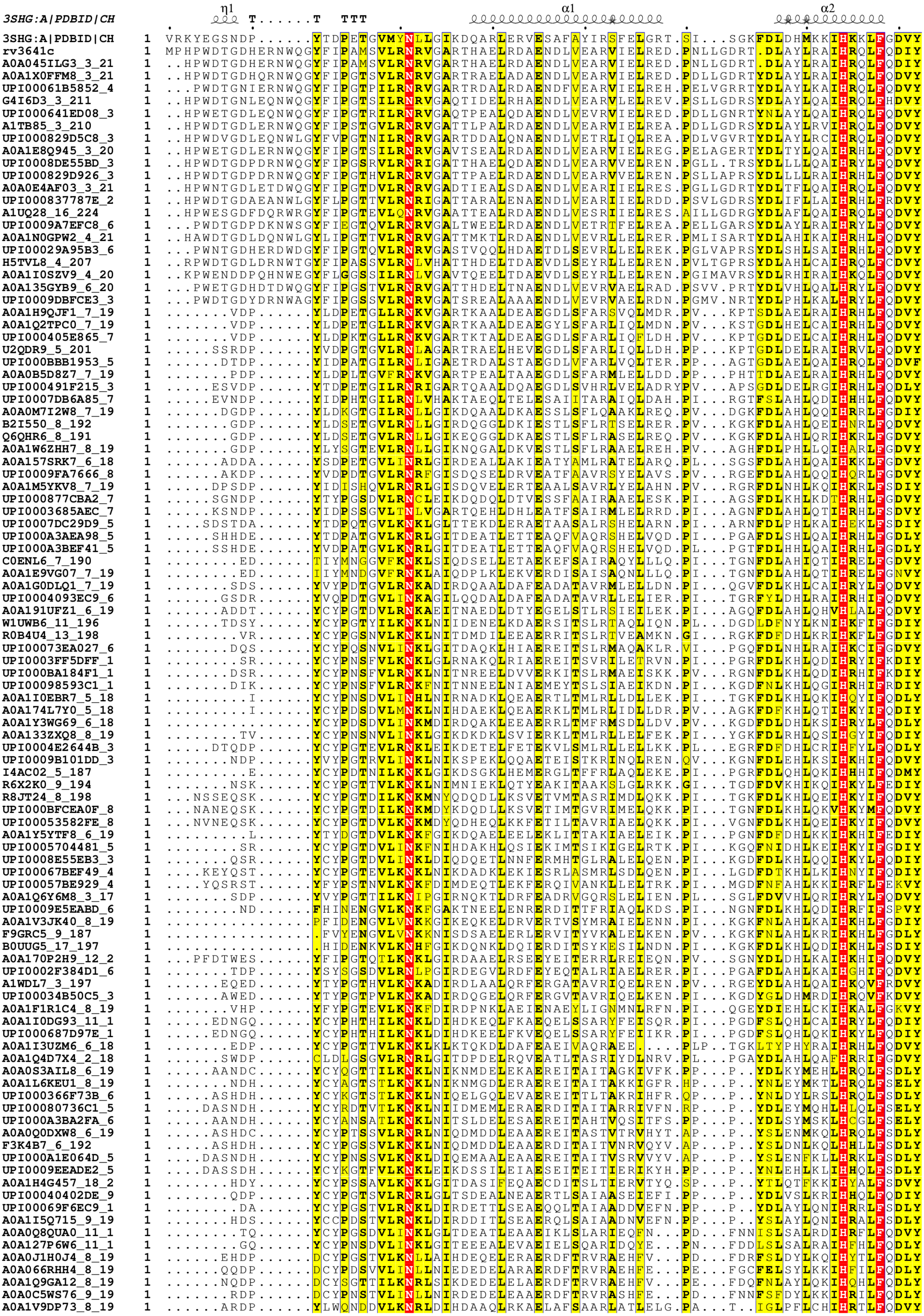
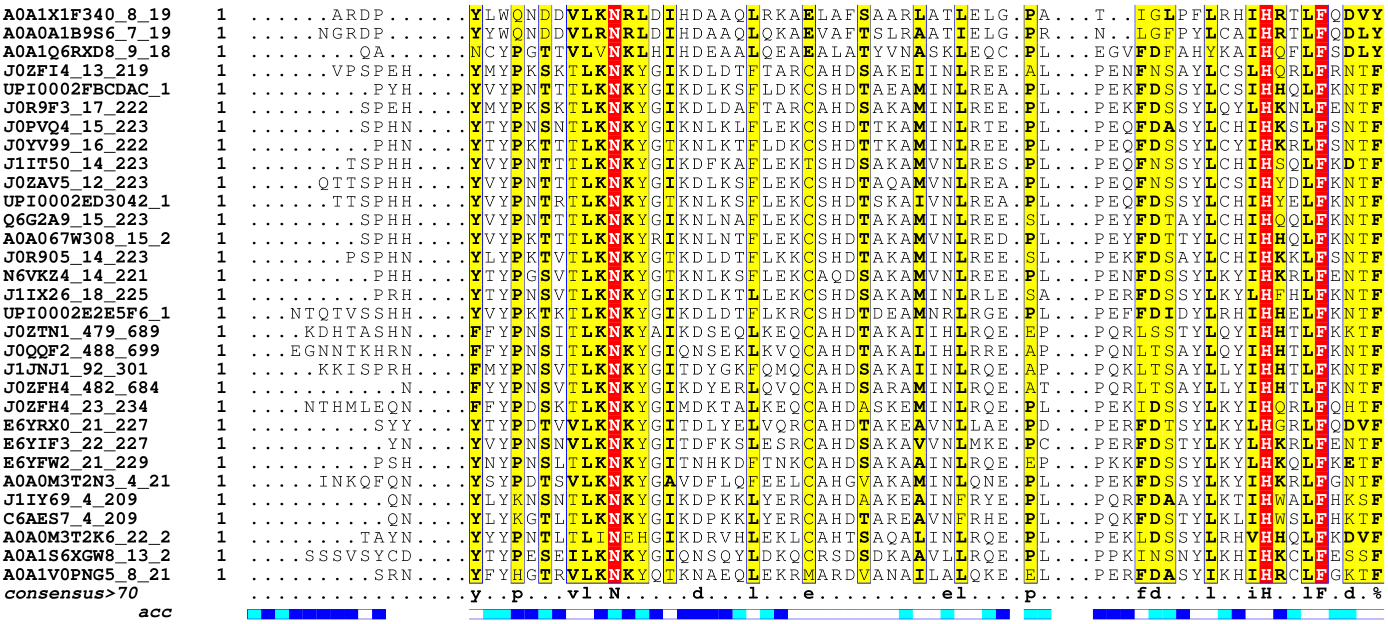


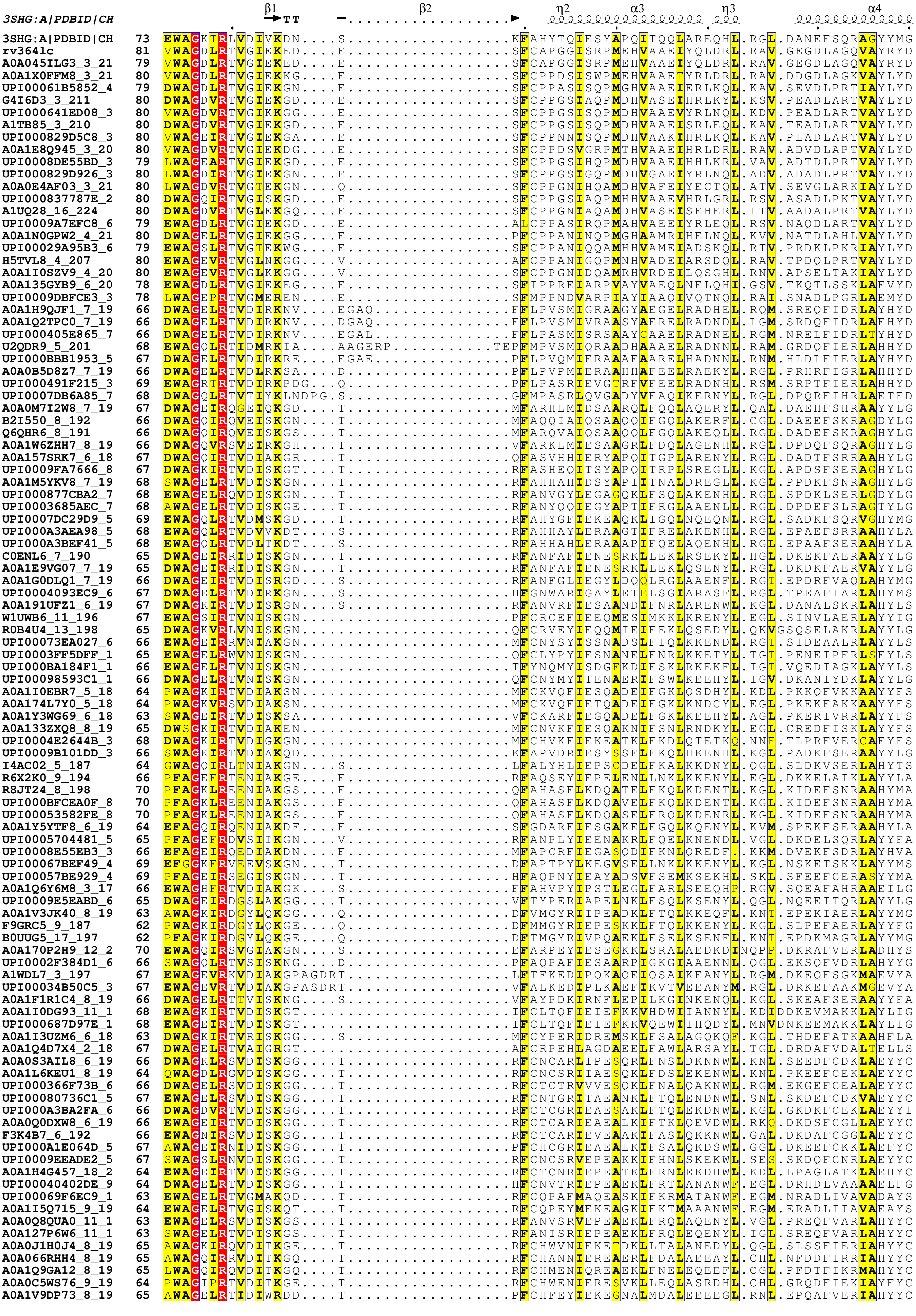
Figure S13: Alignment of Rv2514c with homologous VapC toxins from various organisms. The columns shaded dark red show conserved residues while columns shaded yellow show conservatively substituted residues. The two residues chosen for mutations are marked in green boxes. Consensus sequence (at a threshold > 70%) is shown at the bottom of the alignment. Solvent accessibility of the reference structure, shown using the Espript colour convention with dark blue indicating exposed residues, cyan showing partially exposed residues and white for buried residues (and red for residues for which this value could not be determined), is shown at the bottom of the alignment.

A

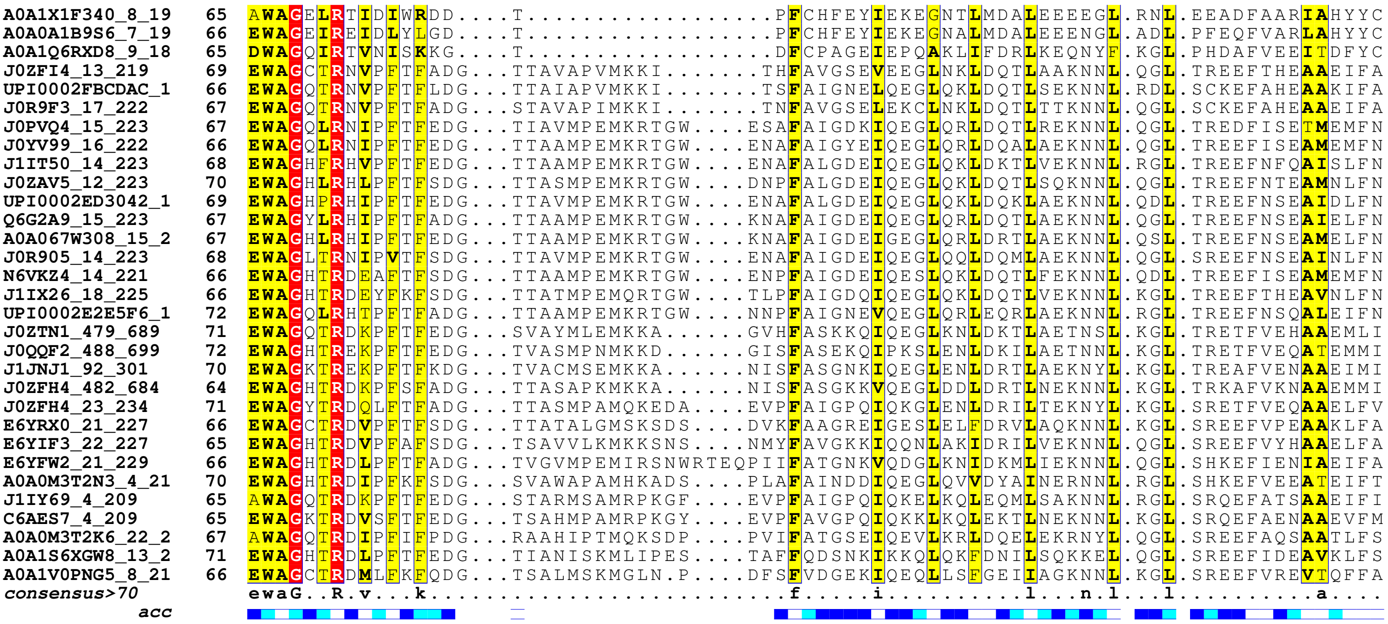




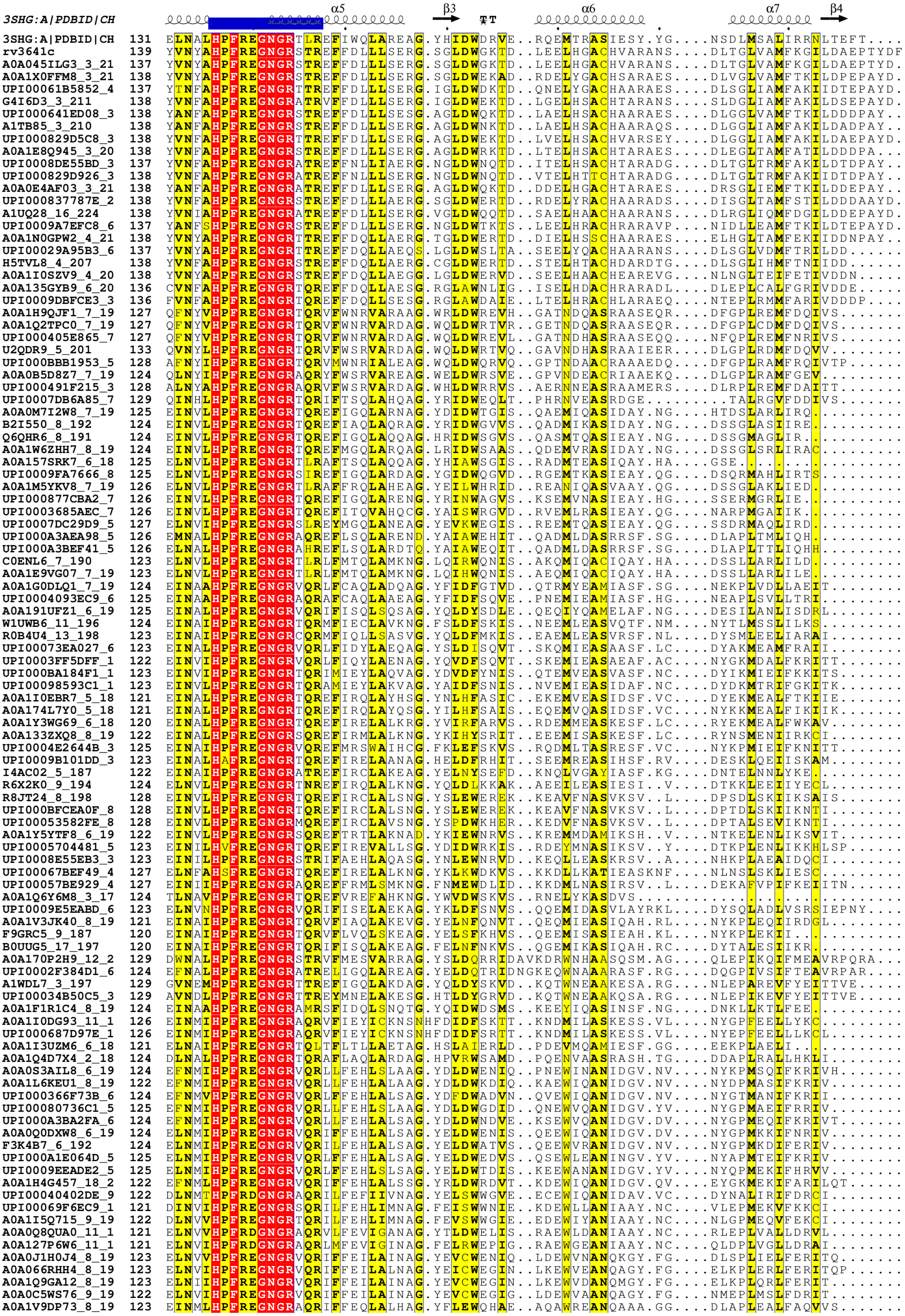
```
A0A1X1F340_8_19   1 ......ARDP......YLWQNDDVLKNRLDIHDAAQLRKAELAFSAARLATLELG.PA...T..IGLPFLRHIHRTLFQDVY
A0A0A1B9S6_7_19   1 .....NGRDP......YYWQNDDVLRNRLDIHDAAQLQKAEVAFTSLRAATIELG.PR...N..LGFPYLCAIHRTLFQDLY
A0A1Q6RXD8_9_18   1 ........QA......NCYPGTTVLVNKLHIHDEAALQEAEALATYVNASKLEQC.PL...EGVFDFAHYKAIHQFLFSDLY
J0ZFI4_13_219     1 ......VPSPEH....YMYPKSKTLKNKYGIKDLDTFTARCAHDSAKEIINLREE.AL...PENFNSAYLCSLHQRLFRNTF
UPI0002FBCDAC_1   1 .........PYH....YVYPNTTTLKNKYGIKDLKSFLDKCSHDTAEAMINLREA.PL...PEKFDSSYLCSIHHQLFKNTF
J0R9F3_17_222     1 ........SPEH....YMYPKSKTLKNKYGIKDLDAFTARCAHDSAKAMINLREE.PL...PEKFDSSYLQYLHKNLFENTF
J0PVQ4_15_223     1 ........SPHN....YTYPNSNTLKNKYGIKNLKLFLEKCSHDTTKAMINLRTE.PL...PEQFDASYLCHIHKSLFSNTF
J0YV99_16_222     1 .........PHN....YTYPKTTTLKNKYGTKNLKTFLDKCSHDTTKAMINLREE.PL...PEQFDSSYLCYIHKRLFSNTF
J1IT50_14_223     1 .......TSPHH....YVYPNTTTLKNKYGIKDFKAFLEKTSHDSAKAMVNLRES.PL...PEQFNSSYLCHIHSQLFKDTF
J0ZAV5_12_223     1 .....QTTSPHH....YVYPNTTTLKNKYGIKDLKSFLEKCSHDTAQAMVNLREA.PL...PEQFNSSYLCSIHYDLFKNTF
UPI0002ED3042_1   1 ......TTSPHH....YVYPNTRTLKNKYGTKNLKSFLEKCSHDTSKAIVNLREA.PL...PEQFDSSYLCHIHYELFKNTF
Q6G2A9_15_223     1 ........SPHH....YVYPNTTTLKNKYGIKNLNAFLEKCSHDTAKAMINLREE.SL...PEYFDTAYLCHIHQQLFKNTF
A0A067W308_15_2   1 ........SPHH....YVYPKTTTLKNKYRIKNLNTFLEKCSHDTAKAMVNLRED.PL...PEYFDTTYLCHIHHQLFKNTF
J0R905_14_223     1 .......PSPHN....YLYPKTVTLKNKYGTKDLDTFLKKCSHDTAKAMINLREE.SL...PEKFDSAYLCHIHKKLFSNTF
N6VKZ4_14_221     1 .........PHH....YTYPGSVTLKNKYGTKNLKSFLEKCAQDSAKAMVNLREE.PL...PENFDSSYLKYIHKRLFENTF
J1IX26_18_225     1 .........PRH....YTYPNSVTLKNKYGIKDLKTLLEKCSHDSAKAMINLRLE.SA...PERFDSSYLKYLHFHLFKNTF
UPI0002E2E5F6_1   1 ...NTQTVSSHH....YVYPKTKTLKNKYGIKDLDTFLKRCSHDTDEAMNRLRGE.PL...PEFFDIDYLRHIHHELFKNTF
J0ZTN1_479_689    1 ....KDHTASHN....FFYPNSITLKNKYAIKDSEQLKEQCAHDTAKAIIHLRQE.EP...PQRLSSTYLQYIHHTLFKKTF
J0QQF2_488_699    1 ...EGNNTKHRN....FFYPNSITLKNKYGIQNSEKLKVQCAHDTAKALIHLRRE.AP...PQNLTSAYLQYIHHTLFKNTF
J1JNJ1_92_301     1 .....KKISPRH....FMYPNSVTLKNKYGITDYGKFQMQCAHDSAKAIINLRQE.AP...PQKLTSAYLLYIHHTLFKNTF
J0ZFH4_482_684    1 ...........N....FYYPNSVTLKNKYGIKDYERLQVQCAHDSARAMINLRQE.AT...PQRLTSAYLLYIHHTLFKNTF
J0ZFH4_23_234     1 ....NTHMLEQN....FFYPDSKTLKNKYGIMDKTALKEQCAHDASKEMINLRQE.PL...PEKIDSSYLKYIHQRLFQHTF
E6YRX0_21_227     1 .........SYY....YTYPDTVVLKNKYGITDLYELVQRCAHDTAKEAVNLLAE.PD...PERFDTSYLKYLHGRLFQDVF
E6YIF3_22_227     1 ..........YN....YVYPNSNVLKNKYGITDFKSLESRCAHDSAKAVVNLMKE.PC...PERFDSTYLKYLHKRLFENTF
E6YFW2_21_229     1 .........PSH....YNYPNSLTLKNKYGITNHKDFTNKCAHDSAKAAINLRQE.EP...PKKFDSSYLKYLHKQLFKETF
A0A0M3T2N3_4_21   1 .....INKQFQN....YSYPDTSVLKNKYGAVDFLQFEELCAHGVAKAMINLQNE.PL...PEKFDSSYLKYIHKRLFGNTF
J1IY69_4_209      1 ..........QN....YLYKNSNTLKNKYGIKDPKKLYERCAHDAAKEAINFRYE.PL...PQRFDAAYLKTIHWALFHKSF
C6AES7_4_209      1 ..........QN....YLYKGTLTLKNKYGIKDPKKLYERCAHDTAREAVNFRHE.PL...PQKFDSTYLKLIHWSLFHKTF
A0A0M3T2K6_22_2   1 ........TAYN....YYYPNTLTLINEHGIKDRVHLEKLCAHTSAQALINLRQE.PL...PEKLDSSYLRHVHHQLFKDVF
A0A1S6XGW8_13_2   1 ....SSSVSYCD....YTYPESEILKNKYGIQNSQYLDKQCRSDSDKAAVLLRQE.PL...PPKINSNYLKHIHKCLFESSF
A0A1V0PNG5_8_21   1 .........SRN....YFYHGTRVLKNKYQTKNAEQLEKRMARDVANAILALQKE.EL...PERFDASYIKHIHRCLFGKTF
consensus>70        ................y..p...vl.N.....d...l...e.........el....p.......fd...l..iH..lF.d.%
acc
```



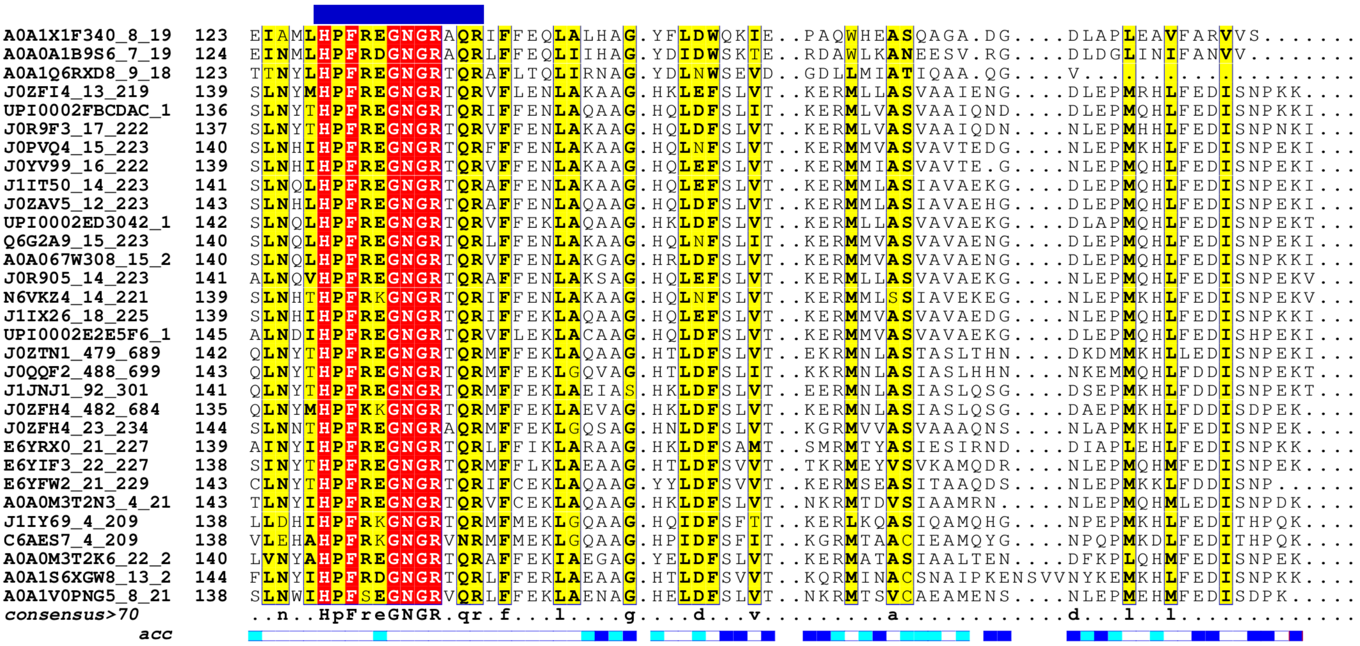
```
                                         β1              β2                η2        α3        η3          α4
3SHG:A|PDBID|CH                         →TT    ─..................►     
3SHG:A|PDBID|CH  73  EWAGKTRLVDIVKDN....S..................KFAHYTQIESYAPQITQQLAREQHL.RGL.DANEFSQRAGYYMG
rv3641c          81  VWAGDLRTVGIEKED....E..................SFCAPGGISRPMEHVAAEIYQLDRL.RAV.GEGDLAGQVAYRYD
A0A045ILG3_3_21  79  VWAGDLRTVGIEKED....E..................SFCAPGGISRPMEHVAAEIYQLDRL.RAV.GEGDLAGQVAYRYD
A0A1X0FFM8_3_21  80  VWAGDLRTVGIEKKG....E..................SFCPPDSISWPMEHVAAETYRLDRL.RAV.DDGDLAGKVAYLYD
UPI00061B5852_4  79  DWAGDLRTVGIEKGN....E..................SFCPPASISQPMGHVAAEIERLKQL.KAV.SEVDLPRTIAYLYD
G4I6D3_3_211     80  DWAGDVRTVGIEKGD....E..................SFCPPGSIAQPMDHVASEIHRLKYL.KSV.PEADLARTVAYLYD
UPI000641ED08_3  80  VWAGDVRTVGIKKGG....E..................SFCPPGSISQPMDHVAAEVYRLNQL.KDV.PEADLARTVAYLYD
A1TB85_3_210     80  DWAGDVRTVGIEKGD....E..................SFCPPGSISQPMDHVAAEISRLEQL.KAV.TEADLARTVAYLYD
UPI000829D5C8_3  80  VWAGEIRTVGIEKGA....E..................SFCPPNNISQPMDHVAAKIHRLKQL.KAV.PLADLARTVAYLYD
A0A1E8Q945_3_20  80  VWAGDVRTVGIEKDG....E..................SFCPPDSVGRPMTHVAAEIHRLDRL.RAV.ADVDLAPTVAYLYD
UPI0008DE55BD_3  79  LWAGEARTVGIEKGD....E..................SFCPPGSIHQPMDHVAAEIHHLKRL.VAV.ADTDLPRTVAYLYD
UPI000829D926_3  80  LWAGDIRTVGIEKGN....E..................SFCPPGSIGQPMDHVGAEIYRLNQL.ADV.ADADLPRTVAYLYD
A0A0E4AF03_3_21  80  LWAGDVRTVGTEKGN....Q..................SFCPPGNIHQAMDHVAFEIYECTQL.ATV.SEVGLARKIAYLYD
UPI000837787E_2  80  DWAGQIRTVGISKGD....E..................SFCPPGSIAQPMHHVAAEVHRLGRL.RSV.ASADLPKTVAYLYD
A1UQ28_16_224    80  DWAGDVRTVGLEKGQ....E..................SFCPPGNIAQAMDHVASEISEHERL.LTV.AADALPRTVAYLYD
UPI0009A7EFC8_6  79  EWAGDLRTVGIEKGD....E..................ALCPPASIRQPMAHVAEEIYRLEQL.RSV.NAQDLARTVAYLYD
A0A1N0GPW2_4_21  80  DWAGELRTVGIEKGG....E..................PFCPPANINQPMGHAAMRIHELNQL.KVV.NATDLAHLVAYLYD
UPI00029A95B3_6  79  EWAGSLRTVGTEKWG....E..................SFCPPANIQQAMHHVAMEIADSKQL.RTV.PRDKLPKRIAYLYD
H5TVL8_4_207     80  EWAGEVRTVGLNKGG....V..................AFCPPANIGQPMNHVADEIARSDQL.RAV.TPDQLPRTVAYLYD
A0A1I0SZV9_4_20  80  EWAGEVRTVGLKKGG....V..................SFCPPGNIDQAMRHVRDEILQSGHL.RNV.APSELTAKIAYLYD
A0A135GYB9_6_20  78  EWAGDLRTVGIEKEG....E..................SFIPPREIARPVAYVAEQLNELQHI.GSV.TKQTLSSKLAFVYD
UPI0009DBFCE3_3  78  LWAGEPRTVGMEREN....E..................SFMPPNDVARPIAYIAAQIVQTNQL.RAI.SNADLPGRLAEMYD
A0A1H9QJF1_7_19  66  DWAGELRTVDIRKNV....EGAQ...............FFLPVSMIGRAAGYAEGELRADHEL.RQM.SREQFIDRLAYHYD
A0A1Q2TPC0_7_19  66  DWAGELRTVDIRKNV....DGAQ...............FFLPVSMIVRAASYAERELRADNDL.LGM.SRDQFIDRLAFHYD
UPI000405E865_7  66  DWAGELRTVDIRKNV....EGAL...............FFLPASMISRSAAYCAAELRADNEL.RGM.NRELFIDRLTYHYD
U2QDR9_5_201     68  EWAGQLRTIDMRKIA....AAGERP...........TEPFMPVSMIQRAADHAAAELRADNNL.RAM.GRDQFIERLAHHYD
UPI000BBB1953_5  67  DWAGELRTVDIRKRE....EGAE...............PFLPVQMIERAAAFAARELHADNNL.RNM.HLDLFIERLAYHYD
A0A0B5D8Z7_7_19  66  DWAGELRTVDLRKSA....D..................PFLPVPMIERAAHHAFEELRAEKYL.RGL.PRHRFIGRLAHHYD
UPI000491F215_3  69  EWAGRTRTVDIRKPDG...Q..................PFLPASRIEVGTRFVFEELRADNHL.RSM.SRPTFIERLAHHYD
UPI0007DB6A85_7  68  DWAGQLRTVTIYKLNDPG.S..................GFMPASRLQVGADYVFAQIKERNYL.RGL.DPQRFIHRLAETFD
A0A0M7I2W8_7_19  67  DWAGEIRQGEIQKGD....T..................MFARHLMIDSAARQLFQQLAQERYL.QAL.DAEHFSHRAAYYLG
B2I550_8_192     66  DWAGQIRQVEISKGN....T..................MFAQQIAIQSAAQQIFGQLAKERFL.CGL.DAEEFSKRAGDYLG
Q6QHR6_8_191     66  DWAGQIRQVEISKGS....T..................MFARQVAIQSAAQQLFGQLAKEQLL.RGL.DADEFSKRAGHYLG
A0A1W6ZHH7_8_19  66  DWAGQVRSVEIRKGH....T..................VFARKLMIESAAGRLFDQLAGENHL.RGL.DPDQFSQRAGHYLG
A0A157SRK7_6_18  67  DWAGQIRTVDIAKGR....T..................QFASVHHIERYAPQITGPLARENYL.RGL.DADTFSRRAAHYLG
UPI0009FA7666_8  67  DWAGKIRTVDISKGT....T..................RFASHEQITSYAPQITRPLSREGLL.KGL.APDSFSERAGHYLG
A0A1M5YKV8_7_19  68  SWAGELRTVDISKGN....T..................RFAHHAHIDSYAPIITNALDREGLL.KGL.PPDKFSNRAGHYLG
UPI000877CBA2_7  68  EWAGELRQVDISKGN....T..................RFANVGYLEGAGQKLFSQLAKENHL.RGL.SPDKLSERLGDYLG
UPI0003685AEC_7  68  AWAGELRTVDISKGE....T..................RFANYQQIEGYAPTIFRGLAAENLL.RGL.DPDEFSLRAGTYLG
UPI0007DC29D9_5  69  EWAGQLRTVDMSKGD....T..................RFAHYGFIEKEAQKLIGQLNQEQNL.RGL.SADKFSQRVGHYMG
UPI000A3AEA98_5  68  EWAGQLRTVDVVKDT....S..................RFAHHAYLERAAGTIFRELAKENHL.RGL.EPAEFSRRAAHYLA
UPI000A3BEF41_5  68  EWAGQLRTVDLTKDT....S..................RFAHHAHLERAAAPIFQELAQENHL.RGL.EPAAFSERAAHYLA
C0ENL6_7_190     65  DWAGEIRRIDISKGN....T..................RFANFAFIENESRKLLEKLRSEKYL.HGL.DKEKFAERAAYYLG
A0A1E9VG07_7_19  65  DWAGEIRRIDISKGN....T..................RFANFAFIENESRKLLEKLQSENYL.RGL.DKDKFAERVAYYLG
A0A1G0DLQ1_7_19  66  DWAGEIRTVDISRGD....S..................RFANFGLIEGYLDQQLRGLAAENFL.RGT.EPDRFVAQLAHYMG
UPI0004093EC9_6  67  DWAGELRTVDIRKGN....S..................HFGNWARIGAYLETELSGIARASFL.RGL.TPEAFIARLAHYMS
A0A191UFZ1_6_19  67  DWAGKIRTVDISRGN....S..................RFANVRFIESAANDIFNRLARENWL.KGL.DANALSKRLAHYLS
W1UWB6_11_196    67  EWAGSIRTVNISKGN.......................PFCRCEFIEEQMESIMKKLKRENYL.EGL.SINVLAERLAYYIG
R0B4U4_13_198    65  DWAGKVRLVNISKGN.......................QFCRVEYIEQQMIEIFEKLQSEDYL.QKVGSQSELAKRLAYYLG
UPI00073EA027_6  66  EWAGEIRRVNIAKGN.......................MFCNYSYISSNADSLFIQLKKENDL.RGT.SIDEAALRLAYYLS
UPI0003FF5DFF_1  65  EWAGELRWVNISKGN.......................QFCLYPYIESNANELFNRLKKETYL.TGT.PENEIPFRLSFYLS
UPI000BA184F1_1  66  EWAGELRTVNISKGN.......................TFYNQMYISDGFKDIFSKLRKEKFL.IGT.VQEDIAGKLAYYLS
UPI00098593C1_1  66  DWAGELRTVNISKGN.......................QFCNYMYITENAERIFSWLKEEHYL.IGV.DKANIYDKLAYYLG
A0A1I0EBR7_5_18  64  PWAGKIRTVDIAKSN.......................MFCKVQFIESQADEIFGKLQSDHYL.KDL.PKKQFVKKAAYYFS
A0A174L7Y0_5_18  64  PWAGKVRSVDIAKSN.......................MFCKVQFIETQADEIFGKLKNDCYL.EGL.PKEKFAKKAAYYFS
A0A1Y3WG69_6_18  63  SWAGEIRTVDISKSA.......................VFCKARFIEGQADELFGKLKEEHYL.AGL.PKERIVRRLAYYFS
A0A133ZXQ8_8_19  65  DWSGKIRTVDIAKEN.......................MFCNVKFIESQAKNIFSNLRNDKYL.EGL.CREDFISKLAYYFS
UPI0004E2644B_3  68  DWAGKIRTVDIGKGN.......................MFCHVKFIEKEATKLFKDLQTETKQ.NNF.TLPRFVERCAFYFS
UPI0009B101DD_3  66  SWAGKIRTVDIAKQD.......................KFAPVDRIESYSSFLFKQLHKENHL.KGL.PADKFSERAAYYLG
I4AC02_5_187     64  GWAGQIRLTNIAKGN....S..................LFALYHLIEPSCDELFKALKKDNYL.QGL.SLDKVSERLAHYTS
R6X2K0_9_194     66  PFAGEFRTENIAKGE....F..................RFAQSEYIEPELENLLNKLKKENYL.EGL.DKKELAIKLAYYLA
R8JT24_8_198     70  PFAGKLREENIAKGS....F..................QFAHASFLKDQATELFKQLKDENTL.KGL.KIDEFSNRAAHYMA
UPI000BFCEA0F_8  70  PFAGKLREENIAKGS....F..................QFAHASFLKDQAVELFKQLKDENTL.KGL.KIDEFSNRAAHYMA
UPI00053582FE_8  70  PFAGKLREENIAKGS....F..................QFAHASFLKDQASELFRQLKDENRL.KGL.KIDEFSNRAAHYMA
A0A1Y5YTF8_6_19  64  EFAGQIRQENIAKDF....F..................SFGDARFIESGAKELFGQLKQENYL.KVM.SPEKFSERAAHYLA
UPI0005704481_5  65  PFAGEFRDVSIIKGN....V..................RFANPLYIEENAKRLFQELKNENDL.VGL.DKDELSQRLAYYFA
UPI0008E55EB3_3  66  EFAGEIRQEDIAKDN....F..................MFAPCRFIEGASQDIFKNLQREDF..KKM.DVEKFSDRLAYHVA
UPI00067BEF49_4  69  EFGGKFRVEEVSKGN....T..................DFAPTPYLKEGVSELLNKLKKENYL.KGL.NSKETSKKLAYYMS
UPI00057BE929_4  69  PFAGEIRSEGISKGN....T..................HFAQPNYIEAYADSVFSEMKSEKHL.SKL.SFEEFCERASYYMA
A0A1Q6Y6M8_3_17  66  EWAGHFRTVDIAKGK....S..................FFAHVPYIPSTLEGLFARLSEEQHP.RGV.SQEAFAHRAAEILG
UPI0009E5EABD_6  65  DWAGEIRDGSLAKGG....T..................VFTYPERIAPELNKLFSQLKSENYL.QGL.NQEEITQRLAYYLG
A0A1V3JK40_8_19  63  AWAGKIRDGYLQKGG....Q..................DFVMGYRIIPEADKLFTQLKKEQFL.KNT.EPEKIAERLAYYMG
F9GRC5_9_187     62  PWAGKIRDGYLQKGE....Q..................DFMMGYRIEPESKKLFTQLKKENYL.KGL.SLPEFAERLAYYFG
B0UUG5_17_197    62  PFAGKIRDGYLQKGE....Q..................DFTMGYRIVPQAEKLFSELKSENFL.KNT.EPDKMAGRLAYYMG
A0A170P2H9_12_2  70  EWAGQIRSVGIAKGN....S..................EFARPEYIESEGKKLSRALAEDKDINQPP.DKRAFVERLADHYS
UPI0002F384D1_6  66  SWAGQLRTVSISKNG....D..................PFAQPAFIESAARPIGKGIAAENNL.QGL.EKSQFVDRLAHYYG
A1WDL7_3_197     67  EWAGEVRKVDIAKGPAGDRT..................LFTFKEDIPQKAQEIQSAIKEANYL.RGM.DKEQFSGKMAEVYA
UPI00034B50C5_3  67  EWAGEVRTVDIAKGPASDRT..................VFALKEDIPLKAEFIKVTVEEANYM.RGL.DRKEFAAKMGEVYA
A0A1F1R1C4_8_19  66  DWAGELRTTVISKNG....S..................VFAYPDKIRNFLEPILKGINKENQL.KGL.SKEAFSERAAYYFA
A0A1I0DG93_11_1  68  EWAGKIRTVDISKGT.......................IFCLTQFIEIEFKKVHDWIIANNYL.KDIDDKEVMAKKLAYLIG
UPI000687D97E_1  68  EWAGEIRTVDISKGT.......................IFCLTQFIEIEFKKVQEWIIHQDYL.MNVDNKEEMAKKLAYLIG
A0A1I3UZM6_6_18  63  DWAGKIRTVRISKGG....S..................MFCYPERIDREMSKLFSALAGQKQF.KGL.THDEFATKAAHFLA
A0A1Q4D7X4_2_18  67  DWAGELRTVAIGRGT.......................AFCRPEHLAGDAEELFAWLARSAYL.TGL.DRDAFVDALTELLS
A0A0S3AIL8_6_19  66  DWAGKLRSVDISKGG....T..................RFCNCARLIPESQRLFNSLDKNNWL.KNL.SEEDFCDKLAEYYC
A0A1L6KEU1_8_19  64  QWAGDLRGVDISKGG....T..................RFCNCSRIIAESQKLFDSLEKENWL.KNL.PKEEFCEKLAEYYC
UPI000366F73B_6  66  DWAGELRSVDISKGG....T..................RFCTCTRVVVESQKLFNALQAKNWL.RGM.EKGEFCEALAEYYC
UPI00080736C1_5  67  EWAGELRSVDISKGG....T..................RFCNTGRITAEANKLFTQLENDNWL.SNL.SKDEFCDKVAEYYC
UPI000A3BA2FA_6  66  DWAGDVRTVDISKGG....T..................RFCTCGRIEAESAKLFTQLEKDNWL.KGL.GKDKFCEKLAEYYI
A0A0Q0DXW8_6_19  66  EWAGEIRTVDISKGG....T..................RFCTCARIEAEALKIFTKLQEVDWL.RKQ.DKDSFCGLLAEYYC
F3K4B7_6_192     66  EWAGNIRSVDISKGG....T..................RFCTCARIEVEAAKIFASLQKKNWL.GAL.DKDAFCGGLAEYYC
UPI000A1E064D_5  67  AWAGEIRNIDISKGG....T..................PFCHCGRIEAEALKIFSALQKDSWL.GGL.DKEEFCKKLAEFYC
UPI0009EEADE2_5  67  SWAGSLRNVDISKGG....T..................RFCNCSRVEPEAKKIFHNLSKEKWL.SES.SKDQFCNRLAEYYC
A0A1H4G457_18_2  64  EWAGEIRTVDISKGG....T..................RFCTCNRIEPEATKLFRNLEKDHWL.KDL.LPAGLATKLAEHYC
UPI00040402DE_9  64  DWAGELRTVDISKGG....T..................HFCNVTRIEPEAEKLFRTLANANWF.EGL.DRHDLVAAAAELFG
UPI00069F6EC9_1  63  EWAGELRTVGMAKQD....T..................RFCQPAFMAQEASKIFKRMATANWF.EGM.NRADLIVAVADAYS
A0A1I5Q715_9_19  64  EWAGELRTVGISKQT....T..................RFCQPEYMEKEAGKIFKTMAAANWF.EGM.ERADLIIAVAEAYS
A0A0Q8QUA0_11_1  63  EWAGELRSVDISKGS....T..................RFANVSRVEPEAEKLFRQLAQENYL.VGL.PREQFVARLAHYYC
A0A127P6W6_11_1  63  SWAGELRTVDISKGS....T..................RFANISRIEPEAEKLFRQLEQENYL.IGL.SREQFVARLAHYYC
A0A0J1H0J4_8_19  65  AWAGKIRQVDITKGE....T..................RFCHWVNIEKETDKLLTALANEDYL.QGL.SLSSFIERIAHYYC
A0A066RHH4_8_19  65  AWAGQIRRVDITKGQ....T..................RFCHANNIEREAERLFDALAEEAYL.CHL.NFDDFIRRIAHYYC
A0A1Q9GA12_8_19  65  LWAGQIRQVDITKGQ....T..................RFCHMENIQREAEKLFNELEQEEFL.CGL.PFDTFIKRMAHYYC
A0A0C5WS76_9_19  64  PWAGIPRTVDITKGE....T..................RFCHWENIERESVKLLEQLASRDHL.CHL.GRQDFIEKIAYFYC
A0A1V9DP73_8_19  65  AWAGELRTIDIWRDD....T..................PFCHFEYIEKEGNALMDALEEEEGL.RNL.EPQAFVERIAHYYC
```



```
A0A1X1F340_8_19   65  AWAGELRTIDIWRDD....T..................PFCHFEYIEKEGNTLMDALEEEEGL.RNL.EEADFAARIAHYYC
A0A0A1B9S6_7_19   66  EWAGEIREIDLYLGD....T..................PFCHFEYIEKEGNALMDALEEENGL.ADL.PFEQFVARLAHYYC
A0A1Q6RXD8_9_18   65  DWAGQIRTVNISKKG....T..................DFCPAGEIEPQAKLIFDRLKEQNYF.KGL.PHDAFVEEITDFYC
J0ZFI4_13_219     69  EWAGCTRNVPFTFADG...TTAVAPVMKKI.......THFAVGSEVEEGLNKLDQTLAAKNNL.QGL.TREEFTHEAAEIFA
UPI0002FBCDAC_1   66  EWAGQTRNVPFTFLDG...TTAIAPVMKKI.......TSFAIGNELQEGLQKLDQTLSEKNNL.RDL.SCKEFAHEAAKIFA
J0R9F3_17_222     67  EWAGQTRNVPFTFADG...STAVAPIMKKI.......TNFAVGSELEKCLNKLDQTLTTKNNL.QGL.SCKEFAHEAAEIFA
J0PVQ4_15_223     67  EWAGQLRNIPFTFEDG...TIAVMPEMKRTGW....ESAFAIGDKIQEGLQRLDQTLREKNNL.QGL.TREDFISETMEMFN
J0YV99_16_222     66  EWAGQLRNIPFTFEDG...TTAAMPEMKRTGW....ENAFAIGYEIQEGLQRLDQALAEKNNL.QGL.TREEFISEAMEMFN
J1IT50_14_223     68  EWAGHFRHVPFTFEDG...TTAAMPEMKRTGW....ENAFALGDEIQEGLQKLDKTLVEKNNL.RGL.TREEFNFQAISLFN
J0ZAV5_12_223     70  EWAGHLRHLPFTFSDG...TTASMPEMKRTGW....DNPFALGDEIQEGLQKLDQTLSQKNNL.QGL.TREEFNTEAMNLFN
UPI0002ED3042_1   69  EWAGHPRHIPFTFEDG...TTAVMPEMKRTGW....ENAFALGDEIQEGLQKLDQKLAEKNNL.QDL.TREEFNSEAIDLFN
Q6G2A9_15_223     67  EWAGYLRHIPFTFADG...TTAAMPEMKRTGW....KNAFAIGDEIQEGLQRLDQTLAEKNNL.QGL.TREEFNSEAIELFN
A0A067W308_15_2   67  EWAGHLRHIPFTFEDG...TTAAMPEMKRTGW....KNAFAIGDEIGEGLQRLDRTLAEKNNL.QSL.TREEFNSEAMELFN
J0R905_14_223     68  EWAGLTRNIPVTFSDG...TTAVMPEMKRTGW....ENAFAIGDEIQEGLQQLDQMLAEKNNL.QGL.SREEFNSEAINLFN
N6VKZ4_14_221     66  EWAGHTRDEAFTFSDG...TTAAMPEMKRTGW....ENPFAIGDEIQESLQKLDQTLFEKNNL.QDL.TREEFISEAMEMFN
J1IX26_18_225     66  EWAGHTRDEYFKFSDG...TTATMPEMQRTGW....TLPFAIGDQIQEGLQKLDQTLVEKNNL.KGL.TREEFTHEAVNLFN
UPI0002E2E5F6_1   72  EWAGQLRHTPFTFADG...TTAAMPEMKRTGW....NNPFAIGNEVQEGLQRLEQRLAEKNNL.RGL.TREEFNSQALEIFN
J0ZTN1_479_689    71  EWAGQTRDKPFTFEDG...SVAYMLEMKKA......GVHFASKKQIQEGLKNLDKTLAETNSL.KGL.TRETFVEHAAEMLI
J0QQF2_488_699    72  EWAGHTREKPFTFEDG...TVASMPNMKKD......GISFASEKQIPKSLENLDKILAETNNL.KGL.TRETFVEQATEMMI
J1JNJ1_92_301     70  EWAGKTREKPFTFKDG...TVACMSEMKKA......NISFASGNKIQEGLENLDRTLAEKNYL.KGL.TREAFVENAAEIMI
J0ZFH4_482_684    64  EWAGHTRDKPFSFADG...TTASAPKMKKA......NISFASGKKVQEGLDDLDRTLNEKNNL.KGL.TRKAFVKNAAEMMI
J0ZFH4_23_234     71  EWAGYTRDQLFTFADG...TTASMPAMQKEDA....EVPFAIGPQIQKGLENLDRILTEKNYL.KGL.SRETFVEQAAELFV
E6YRX0_21_227     66  EWAGCTRDVPFTFSDG...TTATALGMSKSDS....DVKFAAGREIGESLELFDRVLAQKNNL.KGL.SREEFVPEAAKLFA
E6YIF3_22_227     65  EWAGHTRDVPFAFSDG...TSAVVLKMKKSNS....NMYFAVGKKIQQNLAKIDRILVEKNNL.QGL.SREEFVYHAAELFA
E6YFW2_21_229     66  EWAGHTRDLPFTFADG...TVGVMPEMIRSNWRTEQPIIFATGNKVQDGLKNIDKMLIEKNNL.QGL.SHKEFIENIAEIFA
A0A0M3T2N3_4_21   70  EWAGHTRDIPFKFSDG...SVAWAPAMHKADS....PLAFAINDDIQEGLQVVDYAINERNNL.RGL.SHKEFVEEATEIFT
J1IY69_4_209      65  AWAGQTRDKPFTFEDG...STARMSAMRPKGF....EVPFAIGPQIQKELKQLEQMLSAKNNL.RGL.SRQEFATSAAEIFI
C6AES7_4_209      65  EWAGKTRDVSFTFEDG...TSAHMPAMRPKGY....EVPFAVGPQIQKKLKQLEKTLNEKNNL.KGL.SRQEFAENAAEVFM
A0A0M3T2K6_22_2   67  AWAGQTRDIPFIFPDG...RAAHIPTMQKSDP....PVIFATGSEIQEVLKRLDQELEKRNYL.QGL.SREEFAQSAATLFS
A0A1S6XGW8_13_2   71  EWAGHTRDLPFTFEDG...TIANISKMLIPES....TAFFQDSNKIKKQLQKFDNILSQKKEL.QGL.SREEFIDEAVKLFS
A0A1V0PNG5_8_21   66  EWAGCTRDMLFKFQDG...TSALMSKMGLN.P....DFSFVDGEKIQEQLLSFGEIIAGKNNL.KGL.SREEFVREVTQFFA
consensus>70          ewaG..R.v...k..........................f.....i..........l...n.l...l..........a....
              acc
```

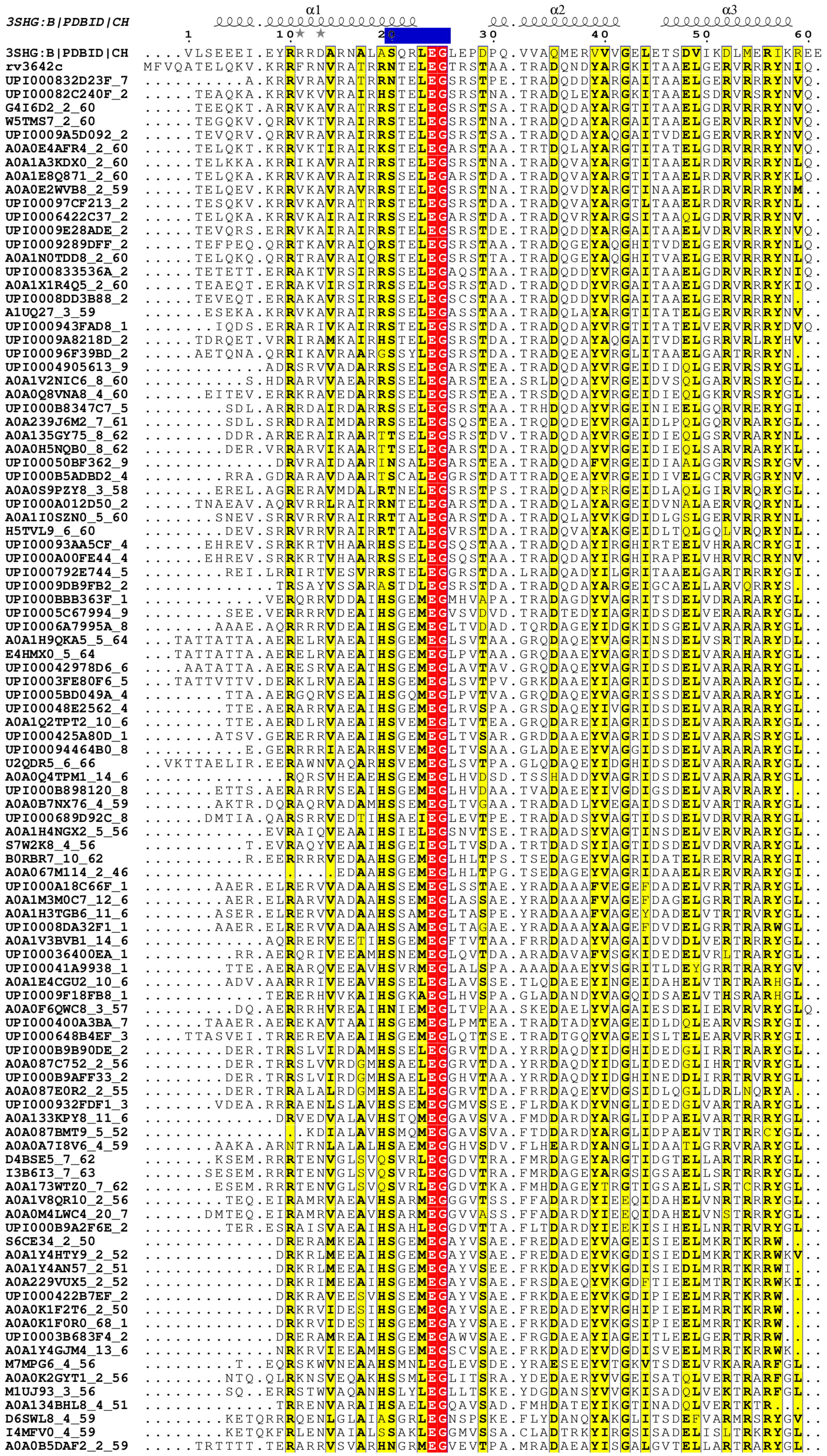


```
3SHG:A|PDBID|CH  131 ELNALHPFREGNGRTLREFIWQLAREAG.YHIDWDRVE..RQEMTRASIESY.YG....NSDLMSALIRRNLTEFT....
rv3641c          139 YVNYAHPFREGNGRSTREFFDLLLSERG.SGLDWGKTD..LEELHGACHVARANS....DLTGLVAMFKGILDAEPTYDF
A0A045ILG3_3_21  137 YVNYAHPFREGNGRSTREFFDLLLSERG.SGLDWGKTD..LEELHGACHVARANS....DLTGLVAMFKGILDAEPTYD.
A0A1X0FFM8_3_21  138 YVNYAHPFREGNGRSTREFFDLLLSERG.SGLDWEKAD..RDELYGACHSARAES....DLVGLVAMFTKILDPEPAYD.
UPI00061B5852_4  137 YTNFAHPFREGNGRTTREFFDLLLSERG.IGLDWDKTD..QNELYGACHAARAES....DLSGLIAMFAKIIDTEPAYD.
G4I6D3_3_211     138 YVNFAHPFREGNGRATRVFFDLLLSERG.IGLDWEKTD..LGELHSACHTARANS....SLDGLIAMFAKILDSEPAYD.
UPI000641ED08_3  138 YANFAHPFREGNGRSTREFFDLLLSESG.AGLDWQKTD..LAELHSACHAARAES....DLGGLVAMFTKILDDDPAY..
A1TB85_3_210     138 YANFAHPFREGNGRSTREFFDLLLSERG.AGLDWNKTD..LAELHSACHAARAQS....DLGGLIAMFAKILDADPAY..
UPI000829D5C8_3  138 YVNFAHPFREGNGRTTREFFDLLLSERG.AGLDWEKTD..LVELHSACHVARSEY....DLGGLIAMFAKILDAEPAYD.
A0A1E8Q945_3_20  138 YVNFAHPFREGNGRSTREFFDLLLSERG.AGLDWRKTD..LTELHSACHTARAES....DLEGLTAMFATILDDDPA...
UPI0008DE55BD_3  137 YVNFAHPFREGNGRATREFFNLLIAERG.NGLDWNQTD..ITELHSACHTARADG....DLTGLTRMFAKILDTDPAY..
UPI000829D926_3  138 YVNFAHPFREGNGRATREFFNLLLSERG.NGLDWNQTD..VTELHTTCHTARADG....DLTGLTRMFAKIIDTDPAY..
A0A0E4AF03_3_21  138 YANFAHPFREGNGRATREFFDLLLSERG.AGLDWQKTD..DDELHGACHTARAES....DRSGLIEMFTKILDDEPAY..
UPI000837787E_2  138 YVNFAHPFREGNGRSTREFFDLLLAERG.SGLDWERTD..RDELYRACHAARADS....DISGLTAMFSTILDDDAAYD.
A1UQ28_16_224    138 YVNIAHPFREGNGRATREFFDLLLSERG.VGLDWQQTD..SAELHAACHAARANS....DLRGLIQMFDGILDDEPAYD.
UPI0009A7EFC8_6  137 YANFSHPFREGNGRSTREFFDLLLSERG.NGLDWTKTS..LEELHRACHRARAVP....NIDLLEAMFTKILDTEPAYD.
A0A1N0GPW2_4_21  138 YVNYAHPFREGNGRTTREFFDQLLAERG.TGLDWERTD..DIELHSSCHAARAES....ELHGLTAMFEKIIDDNPAY..
UPI00029A95B3_6  137 YVNYAHPFREGNGRTTREFFDQLLAEQS.LGLDWSLTA..SEELYQACHAARANS....DLSGLVDMFTRILDD......
H5TVL8_4_207     138 YVNFAHPFREGNGRSTREFFDQLLAERG.CGLDWERTD..STELHTACHDARDTG....DLSGLIHMFTNIIDD......
A0A1I0SZV9_4_20  138 YVNYAHPFREGNGRSTREFFDQLLAEGG.LGLDWERVD..SEELHAACHEAREVG....NLNGLTEIFETIVDDN.....
A0A135GYB9_6_20  136 CVNFAHPFREGNGRTQREFFDQLLAESG.RGLAWNLIG..ISELHDACHRARADL....DLEPLCALFGRIVDDE.....
UPI0009DBFCE3_3  136 FVNFAHPFREGNGRTQREFFDQLLSESG.RGLAWDAIE..LTELHRACHLARAEQ....NTEPLRMMFARIVDDDP....
A0A1H9QJF1_7_19  127 QFNYVHPFREGNGRTQRVFWNRVARAAG.WRLDWREVH..GATNDQASRAASEQR....DFGPLREMFDQIVS.......
A0A1Q2TPC0_7_19  127 QFNYVHPFREGNGRTQRVFWNRVARDAG.WQLDWRTVR..GETNDQASRAASEQR....DFGPLCDMFDQIVS.......
UPI000405E865_7  127 QFNYVHPFREGNGRTQRVFWNRVARDAG.WQLDWREVL..GATNDHASRAASEQQ....NFQPLREMFNRIVS.......
U2QDR9_5_201     133 QVNYLHPFREGNGRTQRVFWSRVARDAG.WQLDWRTVQ..GAVNDHASRAAIEER....DLGPLRDMFDQVV........
UPI000BBB1953_5  128 AFNYIHPFREGNGRTQRVMWNRIALEAG.WQLDWQRVQ..GPTNDAACRAAAEDQ....DFGPLRAMFRQIVTP......
A0A0B5D8Z7_7_19  124 QLNYIHPFREGNGRAQRYFWSRVAREAG.WELDWRSVE..GPVNDEACRIAAEKQ....DLGPLRAMFGEV.........
UPI000491F215_3  128 ALNYAHPFREGNGRAQRVFWSRVARDAG.WHLDWRRVS..AERNNEASRAAMERS....DLRPLREMFDAITT.......
UPI0007DB6A85_7  129 QINHLHPFREGNGRTQRIFTSQLAHQAG.YRIDWEQLT..PHRNVEASRDGE.........TALRGVFDDIVS.......
A0A0M7I2W8_7_19  125 EINVLHPFREGNGRTQREFIGQLARNAG.YDIDWTGIS..QAEMIQASIDAY.NG....HTDSLARLIRQ..........
B2I550_8_192     124 EINVLHPFREGNGRTQREFIAQLAQRAG.YRIDWGVVS..QADMIKASIDAY.NG....DSSGLASIIRE..........
Q6QHR6_8_191     124 EINVLHPFREGNGRTQREFIGQLAQQAG.HRIDWSGVS..QASMTQASIEAY.NG....DSSGMAGLIR...........
A0A1W6ZHH7_8_19  124 EINVLHPFREGNGRTQREFMGQLARQAG.LRIDWSAAS..QDEMVRASIDAY.NG....DSSGLSRLIRAC.........
A0A157SRK7_6_18  125 ELNVLHPFREGNGRTLRAFTSQLAQQAG.YHIAWSGIS..RADMTEASIQAY.HA....GSE..................
UPI0009FA7666_8  125 ELNVLHPFREGNGRSIREFIGQLARDAG.YGIDWQGVD..RGEMTKASIEAY.QG....DSQRMAHLIRTS.........
A0A1M5YKV8_7_19  126 ELNVLHPFREGNGRTLRAFFRQLAHEAG.YEILWHRID..REANIQASVAAY.QG....DSSGLAKLIED..........
UPI000877CBA2_7  126 EINVLHPFREGNGRTQREFIGQLARENG.YRINWAGVS..KSEMVNASIEAY.HG....SSERMGRLIE...........
UPI0003685AEC_7  126 EINVLHPFREGNGRTQREFITQVAHQCG.YAISWRGVD..RVEMLRASIEAY.QG....NARPMGAIIK...........
UPI0007DC29D9_5  127 ELNVIHPFREGNGRSLREYMGQLANEAG.YEVKWEGIS..REEMTQASIEAY.QG....SSDRMAQLIRD..........
UPI000A3AEA98_5  126 EMNALHPFREGNGRAQREFLSQLAREND.YAIAWEHIT..QADMLDASRRSF.SG....DLAPLTMLIQH..........
UPI000A3BEF41_5  126 ELNALHPFREGNGRAHREFLSQLARDTQ.YAIAWEQIT..QADMLDASRRSF.SG....DLAPLTTLIQHH.........
C0ENL6_7_190     123 ELNVLHPFREGNGRTLRLFMTQLAMKNG.LQIRWQNIS..AKQMIQACIQAY.HA....DSSLLARLILD..........
A0A1E9VG07_7_19  123 ELNVLHPFREGNGRTLRLFMTQLAMKNG.LQIHWQNIS..AEQMIQACIQAY.HA....DSSLLARLILE..........
A0A1G0DLQ1_7_19  124 EINAAHPFREGNGRVQRLFCAQLADQAG.YFIDFGTVD..QTRMYEAMIASF.SG....NEKPLVDLLAEIT........
UPI0004093EC9_6  125 EINAAHPFREGNGRAQRAFCAQLAAEAG.YFIDFSQVE..PNEMIEAMIASF.HG....NEAPLSVLLTRI.........
A0A191UFZ1_6_19  125 EINALHPFREGNGRVQRIFISQLSQSAG.YQLDYSDLE..QEQIYQAMELAF.NG....DESILANLISDRL........
W1UWB6_11_196    124 EINAIHPFREGNGRTQRMFIECLAVKNG.FSLDFSKIS..NEEMLEASVQTF.NM....NYTLMSSLILKS.........
R0B4U4_13_198    123 EINAIHPFREGNGRCQRMFIQLLSASVG.YQLDFMKIS..EAEMLEASVRSF.NL....DYSMLEELIARAI........
UPI00073EA027_6  123 EINALHPFREGNGRVQRLFIDYLAEQAG.YSLDISQVT..SKQMIEASAASF.LC....DYAKMEAMFRAII........
UPI0003FF5DFF_1  122 EINVIHPFREGNGRVQRLFIQYLAENAG.YQVDFSQVT..SKEMIEASAEAF.AC....NYGKMDALFKRITT.......
UPI000BA184F1_1  123 EINVIHPFREGNGRTQRIMIEYLAQVAG.YDVDFYNIS..QTEMVEASVASF.NC....DYIKMIEIFKKITS.......
UPI00098593C1_1  123 EINVLHPFREGNGRTQRAMIEYLAKVAG.YAIDFSNIS..NVEMIEASIDAF.NC....DYDKMTRIFGKITK.......
A0A1I0EBR7_5_18  121 EINALHPFREGNGRTQREFIRQLAYHSG.YNLHFASIC..EKEMVEASIDSF.VC....DYEKMEALFTKIIE.......
A0A174L7Y0_5_18  121 EINALHPFREGNGRTQREFIRQLAYQSG.YILHFSAIS..EQEMVQASIDSF.VC....NYKKMEALFIKIIK.......
A0A1Y3WG69_6_18  120 EINALHPFREGNGRTQREFIRELALKRG.YVIRFARVS..RDEMMEASKESF.LC....RYEKMEALFWKAV........
A0A133ZXQ8_8_19  122 EINALHPFREGNGRSQREFIRMLALKNG.YKIHYSRIT..EEEMISASKESF.LC....RYNSMENIIRKCI........
UPI0004E2644B_3  125 EINAIHPFREGNGRVQREFMRSWAIHCG.FKLEFSKVS..RQDMLTASRESF.VC....NYKPMIEIFKNITT.......
UPI0009B101DD_3  123 EINALHAFREGNGRTQREFIRELAANAG.HELDFRHIT..SEEMREASIHSF.RV....DESKLQEIISKAM........
I4AC02_5_187     122 EINAIHPFREGNGRATREFIRCLAKEAG.YELNYSEFD..KNQLVGAYIASF.KG....DNTELNNLYKE..........
R6X2K0_9_194     124 ELNVLHPFREGNGRTNREFIRQLALKNG.YQLDLKKAK..AEEILEASIESI..V....DTKKLEEIIYECL........
R8JT24_8_198     128 EINVLHPFREGNGRTQREFIRCLALSNG.YSLEWEREK..KEAVFNASVKSV..L....DPSDLSKIIKSAIS.......
UPI000BFCEA0F_8  128 EINVLHPFREGNGRTQREFIRCLALSNG.YSLEWEREK..KEAVFNASVKSV..L....DPTDLSKIIKNTI........
UPI00053582FE_8  128 EINVLHPFREGNGRMQREFIRCLAVSNG.YSPDWKHER..KEDVFNASVKSV..L....DPTALSEVIKNTI........
A0A1Y5YTF8_6_19  122 EINVLHPFREGNGRSQREFTRTLAKNAD.YKIEWNRVS..KREMMDAMIKSH..V....NTKELENLIKSVIT.......
UPI0005704481_5  123 EINILHVFREGNGRTQREFIREVALLSG.YDIDWRKIS..RDEYMNASIKSV..Y....DTKPLENLIKKHLSP......
UPI0008E55EB3_3  123 EINILHPFREGNGRSTRIFAEHLAQASG.YNLEWNRVD..KEQLLEASKRSV..A....NHKPLAEAIDQCI........
UPI00067BEF49_4  127 ELNFAHSFREGNGRTQREFIRELALKNG.YKLKWDKVS..KKDLLKATIEASKNF....NLNSLSKLIESC.........
UPI00057BE929_4  127 EINIIHPFREGNGRAQREFFRMLSMKNG.FNMEWDLID..KKDMLNASIRSV..L....DEKAFVPIFKEIITN......
A0A1Q6Y6M8_3_17  124 TLNAVHPFREGNGRTQREFVREFAHKNG.YWVDWSKVS..RNELYKASDVS.............................
UPI0009E5EABD_6  123 ELNVNHPFREGNGRVQRIFISELAEKAG.YKLDFSNVS..QQEMIDASVLAYRKL....DYSQLADLVSRSIEPNY....
A0A1V3JK40_8_19  121 EINAIHPFREGNGRTQRVFIAQLAKEVG.YELNFQNVT..QEQMIEASIQAH.RL....NYKPLEQIIRDGL........
F9GRC5_9_187     120 EINAIHPFREGNGRTQRVFLVQLSKEAG.YELSFKHVS..QEEMISASIEAH.KL....NYKPLEKII............
B0UUG5_17_197    120 EINAIHPFREGNGRTQRIFISQLAKEAG.FELNFNKVS..QGEMIKASIQAH.RC....DYTALESIIKR..........
A0A170P2H9_12_2  129 DWNALHPFREGNGRSTRVFMESVARRAG.YSLDQRRIDAVKDRWNHAASQSM.AG....QLEPIKQIFMEAVRPQRA...
UPI0002F384D1_6  124 EFNALHPFREGNGRATRELIGQLAREAG.YELDQTRIDNGKEQWNAAARASF.RG....DQGPIVSIFTEAVRPAR....
A1WDL7_3_197     129 GVNEMHPFREGNGRTTREFIGQLAKDAG.YQLDYSKVD..KQTWNEAAKESA.RG....NLEPAREVFYEITTVE.....
UPI00034B50C5_3  129 AVNDLHPFREGNGRTTREYMNELAKESG.HTLDYGRVD..KQTWNEAAKQSA.RG....NLEPIKEVFYEITTVE.....
A0A1F1R1C4_8_19  124 EINAAHPFREGNGRAMRSFIDQLSNQTG.YQIDWDSMS..KEEYIQASINSF.NG....DNKLLKTLFEE..........
A0A1I0DG93_11_1  126 EINMIHPFREGNGRAQRVFIEYICKNSNHFDIDFSKTT..KNDMILASKESS.VL....NYGPFEELLYKC.........
UPI000687D97E_1  126 EINMIHPFREGNGRAQRVFIEYICKNSNHFDIDFSRTT..KNDMILASKESS.VL....NYEPFEELLLKCL........
A0A1I3UZM6_6_18  121 ELNAIHPFREGNGRTQLTFLTLLAETAG.HSLAIERLD..PDEVMQAMIESF.GG....EEKPLAELI............
A0A1Q4D7X4_2_18  124 DLNALHPFREGNGRTQRAFLAQLARDAG.HPVRWSAMD..PQENVAASRASH.TG....DDAPLRALLHKLI........
A0A0S3AIL8_6_19  124 EFNMIHPFREGNGRVQRLLFEHLSLAAG.YDLDWENIQ..PNEWINANIDGV.NV....NYKPMSQIFKRIVH.......
A0A1L6KEU1_8_19  122 EFNMIHPFREGNGRVQRLLFEHLALAAG.YDLDWKDIQ..KAEWIQANIDGV.NV....NYDPMASIFKRIV........
UPI000366F73B_6  124 EFNMVHPFREGNGRVQRLFFEHLALSAG.YDFDWADVN..QVEWIQANIDGV.NV....NYGTMANIFKRII........
UPI00080736C1_5  125 EFNMIHPFREGNGRVQRLLFEHLSLAAG.YDLDWGDID..QNEWVQANIDGV.YV....NYGPMAEIFRRIV........
UPI000A3BA2FA_6  124 EFNMIHPFREGNGRVQRLLFEHLALSAG.YDLDWNDVE..KSEWVQANIDGV.DV....NYGPMEQIFKRIVT.......
A0A0Q0DXW8_6_19  124 ELNMIHPFREGNGRVQRIFFEHLALSAG.YELDWENIS..QEEWIQANIDGV.DV....NYGPMKKIFNRIV........
F3K4B7_6_192     124 ELNMIHPFREGNGRVQRILFEHLALSAG.YELDWATVS..QEEWVRANIDGV.NV....NYGPMKDIFERIV........
UPI000A1E064D_5  125 ELNMIHPFREGNGRVQRLFFEHLALSAG.YELDWEDVS..QSEWLEANINGV.YV....NYAPMTEIFKRIV........
UPI0009EEADE2_5  125 ELNMIHPFREGNGRVQRLFFEHLSLSAG.YELDWTDIS..KEEWANANIDGV.YV....NYQPMEKIFHRVV........
A0A1H4G457_18_2  122 EFNMLHPFREGNGRVQRIFFEHLALSSG.YRLDWRDVT..KDEWIQANIDGV.QV....EYGKMEKIFARILQT......
UPI00040402DE_9  122 DLNMTHPFRDGNGRAQRILFEFIIVNAG.YELSWWGVE..ECEWIQANIDAV.VC....DYNALIRIFDRCI........
UPI00069F6EC9_1  121 DINVIHPFREGNGRSQRILFEHLIMNAG.FEISWWNIE..KDEWLLANIAAY.NC....VLEPMEKVFKKCI........
A0A1I5Q715_9_19  122 DLNVVHPFREGNGRTQRILFEHLIMNAG.FEVSWWGIE..EDEWIYANIAAY.NC....NLGPMEQVFEKCI........
A0A0Q8QUA0_11_1  121 ELNVVHPFRDGNGRAQRVLFEVIGANAG.YTLDWELIS..RAEWLDANIAAY.NC....RLEPLVALLDRVL........
A0A127P6W6_11_1  121 ELNVIHPFRDGNGRAQRLMFEVIGINAG.FELRWEPIG..RAEWGEANIAAY.NC....QLDPLVGLLDRAI........
A0A0J1H0J4_8_19  123 ELNVVHPFREGNGRVQRIFFEILAINAG.YEVCWENIQ..LDEWVNANQKGY.FG....DLSPLIELFERITQ.......
A0A066RHH4_8_19  123 ELNVIHPFREGNGRVQRIFFEILALNAG.YEICWEGIN..LKEWVEANQAGY.FG....ELKPLERLFERITQP......
A0A1Q9GA12_8_19  123 ELNIIHPFREGNGRVQRIFFEILAINAG.YEVCWEGIN..LKEWVDANQMGY.FG....ELKPLQRLFERIT........
A0A0C5WS76_9_19  122 ELNVIHPFREGNGRVQRIFFEILAINAG.YEVCWEGIG..LDEWVAANQAGY.FG....NMEPLMSLFDRIT........
A0A1V9DP73_8_19  123 EINMLHPFREGNGRAQRIFFEQLALHAG.YFLEWQHAE..AQAWREANQAAA.GG....DLAPLEALFAQVVS.......
```

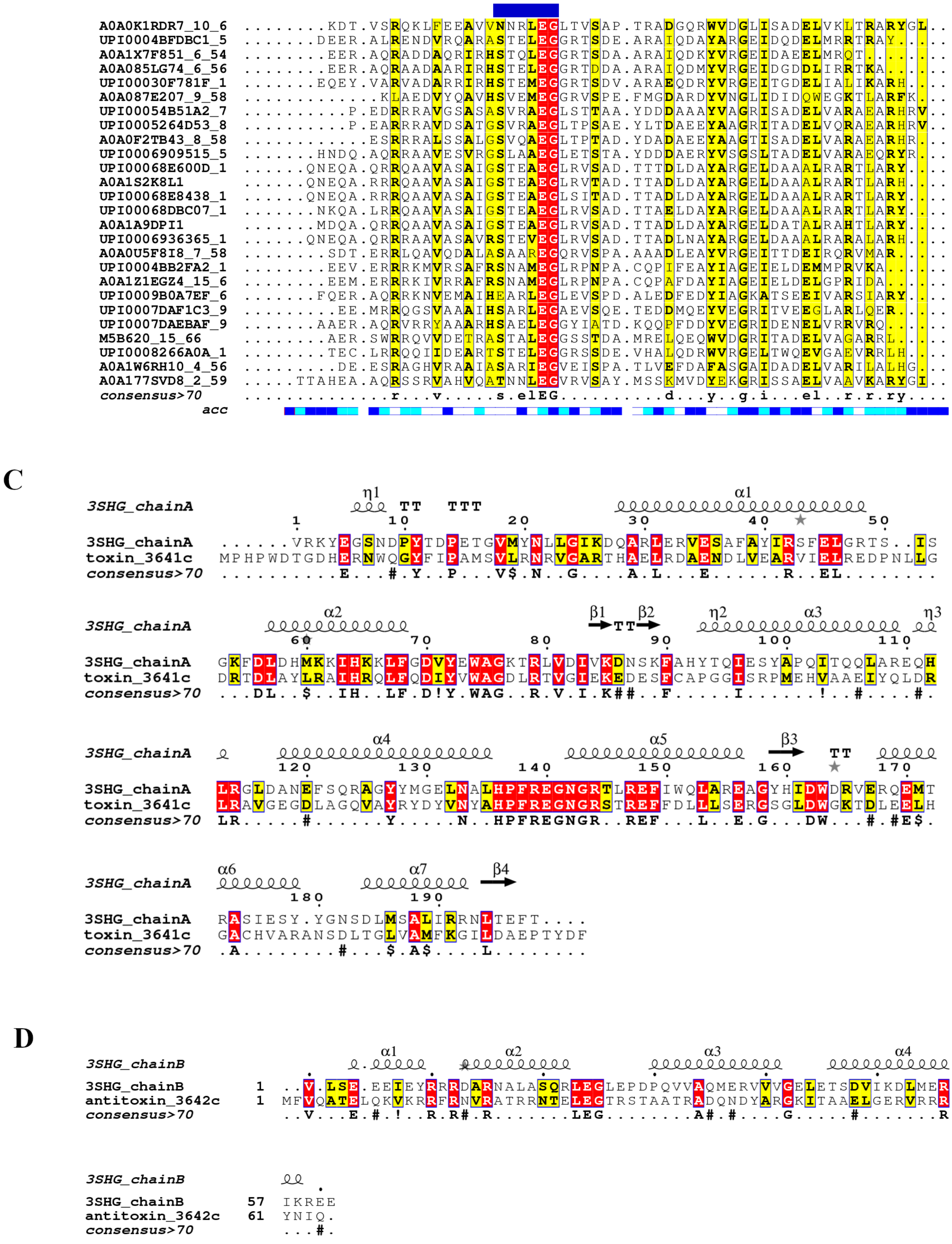


```
A0A1X1F340_8_19   123  EIAMLHPFREGNGRAQRIFFEQLALHAG.YFLDWQKIE..PAQWHEASQAGA.DG....DLAPLEAVFARVVS.......
A0A0A1B9S6_7_19   124  EINMLHPFRDGNGRAQRLFFEQLIIHAG.YDIDWSKTE..RDAWLKANEESV.RG....DLDGLINIFANVV........
A0A1Q6RXD8_9_18   123  TTNYLHPFREGNGRTQRAFLTQLIRNAG.YDLNWSEVD..GDLLMIATIQAA.QG....V....................
J0ZFI4_13_219     139  SLNYMHPFREGNGRTQRVFLENLAKAAG.HKLEFSLVT..KERMLLASVAAIENG....DLEPMRHLFEDISNPKK....
UPI0002FBCDAC_1   136  SLNYTHPFREGNGRTQRIFFENLAQAAG.HQLDFSLIT..KERMLVASVAAIQND....DLEPMQHLFEDISNPKKI...
J0R9F3_17_222     137  SLNYTHPFREGNGRTQRVFFENLAKAAG.HQLDFSLVT..KERMLVASVAAIQDN....NLEPMHHLFEDISNPNKI...
J0PVQ4_15_223     140  SLNHIHPFREGNGRTQRFFFENLAKAAG.HQLNFSLVT..KERMMVASVAVTEDG....NLEPMKHLFEDISNPEKI...
J0YV99_16_222     139  SLNHIHPFREGNGRTQRVFFENLAKAAG.HQLEFSLVT..KERMMIASVAVTE.G....NLEPMQHLFEDISNPEKI...
J1IT50_14_223     141  SLNQLHPFREGNGRTQRAFFENLAKAAG.HQLEFSLVT..KERMMLASIAVAEKG....DLEPMQHLFEDISNPEKI...
J0ZAV5_12_223     143  SLNHLHPFREGNGRTQRAFFENLAQAAG.HQLDFSLVT..KERMMLASIAVAEHG....DLEPMQHLFEDISNPEKI...
UPI0002ED3042_1   142  SLNQLHPFREGNGRTQRVFFEKLAQAAG.HKLDFSLVT..KERMLLASVAVAEKG....DLAPMQHLFEDISNPEKT...
Q6G2A9_15_223     140  SLNQLHPFREGNGRTQRLFFENLAKAAG.HQLNFSLIT..KERMMVASVAVAENG....DLEPMQHLFEDISNPEKI...
A0A067W308_15_2   140  SLNQLHPFREGNGRTQRVFFENLAKGAG.HRLDFSLVT..KERMMVASVAVAENG....DLEPMQHLFEDISNPKKI...
J0R905_14_223     141  ALNQVHPFREGNGRTQRAFFENLAKSAG.HQLEFSLVT..KERMLLASVAVAEKG....NLEPMQHLFEDISNPEKV...
N6VKZ4_14_221     139  SLNHTHPFRKGNGRTQRIFFENLAKAAG.HQLNFSLVT..KERMMLSSIAVEKEG....NLEPMKHLFEDISNPEKV...
J1IX26_18_225     139  SLNHIHPFREGNGRTQRIFFEKLAQAAG.HQLEFSLVT..KERMMVASVAVAEDG....NLEPMQHLFEDISNPKKI...
UPI0002E2E5F6_1   145  ALNDIHPFREGNGRTQRVFLEKLACAAG.HQLDFSLVT..KERMMVASVAVAEKG....DLEPMQHLFEDISHPEKI...
J0ZTN1_479_689    142  QLNYTHPFREGNGRTQRMFFEKLAQAAG.HTLDFSLVT..EKRMNLASTASLTHN....DKDMMKHLLEDISNPEKI...
J0QQF2_488_699    143  QLNYTHPFREGNGRTQRMFFEKLGQVAG.HTLDFSLIT..KKRMNLASIASLHHN....NKEMMQHLFDDISNPEKT...
J1JNJ1_92_301     141  QLNYTHPFREGNGRTQRMFFEKLAEIAS.HKLDFSLVT..EERMNLASIASLQSG....DSEPMKHLFDDISNPEKT...
J0ZFH4_482_684    135  QLNYMHPFKKGNGRTQRMFFEKLAEVAG.HKLDFSLVT..KERMNLASIASLQSG....DAEPMKHLFDDISDPEK....
J0ZFH4_23_234     144  SLNNTHPFREGNGRAQRMFFEKLGQSAG.HNLDFSLVT..KGRMVVASVAAAQNS....NLAPMKHLFEDISNPEK....
E6YRX0_21_227     139  AINYIHPFREGNGRTQRLFFIKLARAAG.HKLDFSAMT..SMRMTYASIESIRND....DIAPLEHLFEDISNPKK....
E6YIF3_22_227     138  SINYTHPFREGNGRTQRMFFLKLAEAAG.HTLDFSVVT..TKRMEYVSVKAMQDR....NLEPMQHMFEDISNPEK....
E6YFW2_21_229     143  CLNYTHPFREGNGRTQRIFCEKLAQAAG.YYLDFSVVT..KERMSEASITAAQDS....NLEPMKKLFDDISNP......
A0A0M3T2N3_4_21   143  TLNYIHPFREGNGRTQRVFCEKLAQAAG.HKLDFSLVT..NKRMTDVSIAAMRN.....NLEPMQHMLEDISNPDK....
J1IY69_4_209      138  LLDHIHPFRKGNGRTQRMFMEKLGQAAG.HQIDFSFTT..KERLKQASIQAMQHG....NPEPMKHLFEDITHPQK....
C6AES7_4_209      138  VLEHAHPFRKGNGRVNRMFMEKLGQAAG.HPIDFSFIT..KGRMTAACIEAMQYG....NPQPMKDLFEDITHPQK....
A0A0M3T2K6_22_2   140  LVNYAHPFREGNGRTQRAFFEKIAEGAG.YELDFSLVT..KERMATASIAALTEN....DFKPLQHMFEDISNPEK....
A0A1S6XGW8_13_2   144  FLNYIHPFRDGNGRTQRLFFERLAEAAG.HTLDFSVVT..KQRMINACSNAIPKENSVVNYKEMKHLFEDISNPKK....
A0A1V0PNG5_8_21   138  SLNWIHPFSEGNGRVQRLFFEKLAENAG.HELDFSLVT..NKRMTSVCAEAMENS....NLEPMEHMFEDISDPK.....
consensus>70           ..n..HpFreGNGR.qr.f...l....g....d...v.........a............d...l..l.............
acc
```

B



```
3SHG:B|PDBID|CH  ....VLSEEEI.EYRRRDARNALASQRLEGLEPDPQ.VVAQMERVVVGELETSDVIKDLMERIKREE
rv3642c          MFVQATELQKV.KRRFRNVRATRRNTELEGTRSTAA.TRADQNDYARGKITAAELGERVRRRYNIQ.
UPI000832D23F_7  ..........A.KRRVRAVRATRRNTELEGSRSTPE.TRADQDDYARGAITAAELGERVRRRYNVQ.
UPI00082C240F_2  .....TEAQKA.KRRVKVVRAIRHSTELEGSRSTNA.TRADQLEYARGKITAAELSDRVRSRYNVQ.
G4I6D2_2_60      .....TEEQKT.ARRVKAVRATRRSTELEGSRSTSA.TRADQDAYARGTITAAELGDRVRRRYNVQ.
W5TMS7_2_60      .....TEGQKV.QRRVKTVRAIRRSTELEGSRSTNA.TRADQDAYARGAITAAELGERVRRRYNVQ.
UPI0009A5D092_2  .....TEVQRV.QRRVRAVRAIRRSTELEGSRSTSA.TRADQDAYAQGAITVDELGERVRRRYNVQ.
A0A0E4AFR4_2_60  .....TELQKT.KRRVKTIRAIRKSTELEGARSTAA.TRTDQLAYARGTITATELGDRVRRRYNVQ.
A0A1A3KDX0_2_60  .....TELKKA.KRRIKAVRAIRRSTELEGSRSTNA.TRADQVAYARGTITAAELRDRVRRRYNLQ.
A0A1E8Q871_2_60  .....TELQKA.KRRVKAVRAIRRSSELEGARSTDA.TRADQVAYARGTITAAELRDRVRRRYKLQ.
A0A0E2WVB8_2_59  .....TELQEV.KRRVKAVRAVRRSTELEGSRSTNA.TRADQVDYARGTINAAELRDRVRRRYNM..
UPI00097CF213_2  .....TESQKV.KRRVKAVRATRRSTELEGSRSTDA.TRADQVAYARGTLTAAELRDRVRRRYNL..
UPI0006422C37_2  .....TELQKA.KRRVKAIRAIRRSSELEGARSTDA.TRADQVRYARGSITAAQLGDRVRRRYNV..
UPI0009E28ADE_2  .....TEVQRS.ERRVKAIRAIRRSTELEGSRSTDE.TRADQDDYARGAITAAELVDRVRRRYNVQ.
UPI0009289DFF_2  .....TEFPEQ.QRRTKAVRAIQRSTELEGSRSTAA.TRADQGEYAQGHITVDELGERVRRRYNLQ.
A0A1N0TDD8_2_60  .....TELQKQ.QRRTRAVRAIQRSTELEGSRSTTA.TRADQGEYAQGHITVAELGERVRRRYNLQ.
UPI000833536A_2  .....TETETT.ERRAKTVRSIRRSSELEGAQSTAA.TRADQDDYVRGAITAAELGDRVRRRYNIQ.
A0A1X1R4Q5_2_60  .....TEAEQT.ERRAKVIRSIRRSSELEGSRSTAD.TRADQDDYVRGAITAAELGDRVRRRYKIQ.
UPI0008DD3B88_2  .....TEVEQT.ERRAKAVRSIRRSSELEGSCSTAA.TRADQDDYVRGAITAAELGERVRRRYN...
A1UQ27_3_59      ......ESEKA.KRRVKAVRAIRRSSELEGASSTAA.TRADQLAYARGTITATELGERVRRRYNV..
UPI000943FAD8_1  .......IQDS.ERRARIVKAIRRSTELEGSRSTDA.TRADQDAYVRGTITATELVERVRRRYDVQ.
UPI0009A8218D_2  .....TDRQET.VRRIRAMKAIRHSTELEGSRSTDA.TRADQDAYAQGAITVDELGRRVRLRYHV..
UPI00096F39BD_2  .....AETQNA.QRRIKAVRAARGSSYLEGSRSTDA.TRADQEAYVRGLITAAELGARTRRRYN...
UPI0004905613_9  ............ADRSRVVADARRSSELEGARSTEA.TRADQDAYVRGEIDIDQLGERVRSRYGL..
A0A1V2NIC6_8_60  ..........S.HDRARVVAEARRSSELEGARSTEA.SRLDQDAYVRGEIDVSQLGARVRSRYGL..
A0A0Q8VNA8_4_60  ......EITEV.ERRKRAVEDARRSSELEGSRSTEA.TRSDQDAYVRGEIDIEQLGKRVRSRYGL..
UPI000B8347C7_5  ........SDL.ARRRDAIRDARRSSELEGSRSTAA.TRHDQDAYVRGEINIEELGQRVRERYGI..
A0A239J6M2_7_61  ........SDL.SRRDRAIMDARRSSELEGSQSTEA.TRQDQDEYVRGAIDLPQLGQRVRARYGL..
A0A135GY75_8_62  ........DDR.ARRERAIRAARTTSELEGSRSTDV.TRADQDAYVRGEIDIEQLGSRVRARYNL..
A0A0H5NQB0_8_62  ........DER.VRRARVIKAARTTSELEGARSTDA.TRADQDAYVRGEIDIEQLSARVRARYKL..
UPI00050BF362_9  .............DRVRAIDAARINSALEGARSTEA.TRADQDAFVRGEIDIAALGGRVRSRYGV..
UPI000B5ADBD2_4  ........RRA.GDRARAVDAARTSCALEGGRSTDATTRADQEAYVRGEIDIVELGCRVRGRYNV..
A0A0S9PZY8_3_58  .......EREL.AGRERAVMDALRTNELEGSRSTPS.TRADQDAYRRGEIDLAQLGIRVRQRYGL..
UPI000A012D50_2  .....TNAEAV.AQRVRRLRAIRRNTELEGARSTDA.TRADQLAYARGEIDVNALAERVRQRYNL..
A0A1I0SZN0_5_60  .......SNEV.SRRVRRVRAIRRTTALEGARSTDA.TRADQLAYVKGDIDLGSLGERVRRRYNL..
H5TVL9_6_60      ........DEV.SRRVRRVRAIRRTTALEGVRSTDA.TRADQVAYAKGEIDLTQLGQLVRQRYNL..
UPI00093AA5CF_4  ......EHREV.SRRKRTVHAARHSSELEGSQSTAA.TRADQDAYIRGHIRTEELVHRARCRYGI..
UPI000A00FE44_4  ......EHREV.SRRKRTVHAARRSSELEGSQSTAA.TRADQDAYIRGHIRAPELVHRVRCRYNV..
UPI000792E744_5  ........REI.LRRIRTVESVRRSTELEGGRSTDA.LRADQADYILGRITAAELGARTRRRYGI..
UPI0009DB9FB2_2  .............TRSAYVSSARASTDLEGSRSTDA.TRADQDAYARGEIGCAELLARVQRRYS...
UPI000BBB363F_1  ............VERQRRVDDAIHSGEMEGMHVAPA.TRADAGDYVAGRITSDELVDRARARYGL..
UPI0005C67994_9  ........SEE.VERRRRVDEAIHSGEMEGVSVDVD.TRADTEDYIAGRIDSDELVERARARYGL..
UPI0006A7995A_8  .......AAAE.AQRRRRVDEAIHSGEMEGLTVDAD.TQRDAGEYIDGKIDSGELVARARARYGL..
A0A1H9QKA5_5_64  ...TATTATTA.AERELRVAEAIHSGEMEGLSVTAA.GRQDAQEYVAGRINSDELVSRTRARYDL..
E4HMX0_5_64      ...TATTATTA.AERELRVAEAIHSGEMEGLPVTAA.GRQDAAEYVAGRIDSDELVARAHARYGL..
UPI00042978D6_6  ....AATATTA.AERESRVAEATHSGEMEGLAVTAV.GRQDAQEYVAGRIDSDELVARARARYGL..
UPI0003FE80F6_5  ...TATTVTTV.DERKLRVAEAIHSGEMEGLSVTPV.GRKDAAEYIAGRIDSDELVARARARYGL..
UPI0005BD049A_4  ........TTA.AERGQRVSEAIHSGQMEGLPVTVA.GRQDAQEYVAGRISSDELVARARARYGL..
UPI00048E2562_4  ........TTE.AERARRVAEAIHSGEMEGLRVSPA.GRSDAQEYIVGRIDSDELVARARARYGL..
A0A1Q2TPT2_10_6  ........TTE.AERDLRVAEAIHSVEMEGLTVTEA.GRQDAREYIAGRIDSDELVARARARYGL..
UPI000425A80D_1  .......ATSV.GERERRVAEAIHSGEMEGLTVSAR.GRDDAAEYVAGRIDSDELVARARSRYGL..
UPI00094464B0_8  ..........E.GERRRRIAEARHSVEMEGLTVSAA.GLADAEEYVAGGIDSDELVARARARYGL..
U2QDR5_6_66      ..VKTTAELIR.EERAWNVAQARHSVEMEGLSVTPA.GLQDAQEYIDGHIDSDELVARARARYGL..
A0A0Q4TPM1_14_6  ..............RQRSVHEAEHSGEMEGLHVDSD.TSSHADDYVAGRIDSDELVARARARYGL..
UPI000B898120_8  .......ETTS.AERARRVSEAIHSGEMEGLHVDAA.TVADAEEYIVGDIGSDELVARVRARY....
A0A0B7NX76_4_59  .......AKTR.DQRAQRVADAMHSSEMEGLTVGAA.TRADADLYVEGAIDSDELVARTRARYGL..
UPI000689D92C_8  ......DMTIA.QARSRRVEDTIHSAEIEGLEVTPE.TRADADSYVAGTIDSDELVDRVRARYGL..
A0A1H4NGX2_5_56  ............EVRAIQVEAAIHSIELEGSNVTSE.TRADADSYVAGTINSDELVERVRARYGL..
S7W2K8_4_56      ..........T.EVRAQYVEAAIHSGEIEGLTVSDA.TRTDADSYIAGTIDSDELVDRVRARYGL..
B0RBR7_10_62     ..........R.EERRRRVEDAAHSGEMEGLHLTPS.TSEDAGEYVAGRIDATELRERVRARYGI..
A0A067M114_2_46  ...................EDAAHSGEMEGLHLTPG.TSEDAGEYVAGRIDAAELRERVRARYGI..
UPI000A18C66F_1  .......AAER.ELRERVVADAAHSSEMEGLSSTAE.YRADAAAFVEGEFDADELVRRTRARYGL..
A0A1M3M0C7_12_6  ........AER.ELRARVVADAAHSSEMEGLASSAE.YRADAAAFVAGEFDAGELGRRTRARYGL..
A0A1H3TGB6_11_6  .......ASER.ELRERVVADAAHSSAMEGLTASPE.YRADAAAFVAGEYDADELVTRTRVRYGL..
UPI0008DA32F1_1  .......AAER.ELRERVVADAAHSSAMEGLTAGAE.YRADAAAYAAGEFDVDELGRRTRARWGL..
A0A1V3BVB1_14_6  ............AQRRERVEETIHSGEMEGFTVTAA.FRRDADAYVAGAIDVDDLVERTRRRYGL..
UPI00036400EA_1  .........RR.AERQRIVEEAMHSNEMEGLQVTDA.ARADAVAFVSGKIDEDELVRLTRARYGL..
UPI00041A9938_1  ........TTE.AERARQVEEAVHSVRMEGLALSPA.AAADAAEYVSGRITLDEYGRRTRARYGV..
A0A1E4CGU2_10_6  ........ADV.AARRRIVEEAVHSSELEGLAVSSE.TLQDAEEYINGEIDAHELVTRTRARHGL..
UPI0009F18FB8_1  ............TERERHVVKAIHSGKAEGLHVSPA.GLADANDYVAGQIDSAELVTHSRARHGL..
A0A0F6QWC8_3_57  .........DQ.AERRRHVREAIHNIEMEGLTVPAA.SKEDAELYVAGAIESDELIERVRVRYGLQ.
UPI000400A3BA_7  ......TAAER.AEREKAVTAAIHSGEMEGLPMTEA.TRADTADYVAGEIDLDQLEARVRSRYGI..
UPI000648B4EF_3  ....TTASVEI.TRREREVAEAIHSGEMEGLQTTSE.TRADTGQYVAGEISLTELEARVRRRYGI..
UPI000B9B90DE_2  ........DER.TRRSLVIRDAMHSSELEGGRVTDA.YRRDAQDYIDGHIDEDGLIRRTRARYGL..
A0A087C752_2_56  ........DER.TRRSLVVRDGMHSAELEGGRVTDA.YRRDAQDYIDGLIDEDGLIHRTRVRYGL..
UPI000B9AFF33_2  ........DER.TRRSLVIRDGMHSAELEGGHVTAA.YRRDAQDYIDGHINEDDLIRRTRVRYGL..
A0A087E0R2_2_55  ........DER.TRRALALRDGMHSAELEGGRVTDA.FRRDAQEYVDGSIDLQGLLDRLNQRYA...
UPI000932FDF1_3  .......VDEA.RRRAENLSLAVHSSEMEGGMVSSE.FLRDAKDYVNGLIDEDGLVARTRARYGL..
A0A133KPY8_11_6  .............DRVEDVALAVHSTQMEGGAVSVA.FMDDARDYARGLIDPAELVARTRRRYGL..
A0A087BMT9_5_52  ................KDIALAVHSMQMEGGAVSVA.FTDDARDYARGLIDPAELVTRTRRCYGL..
A0A0A7I8V6_4_59  .......AAKA.ARNTRNLALALHSAEMEGGHVSDV.FLHEARDYANGLIDAATLGRRVRARYGL..
D4BSE5_7_62      .......KSEM.RRRTENVGLSVQSVRLEGGDVTRA.FMRDAGEYARGTIDGTELLARTRRRYGL..
I3B6I3_7_63      ......SESEM.RRRTENVGLSVQSVRLEGGDVTRA.FMRDAGEYARGSIGSAELLSRTRRRYGL..
A0A173WTZ0_7_62  .......ESEM.RRRTENVGLSVQSVRLEGGDVTKA.FMHDAGEYTRGTIGSAELLSRTCRRYGL..
A0A1V8QR10_2_56  ........TEQ.EIRAMRVAEAVHSARMEGGGVTSS.FFADARDYIEEQIDAHELVNRTRRRYGL..
A0A0M4LWC4_20_7  ......DMTEQ.EIRAMRVAEAVHSARMEGGVVASS.FFADARDYIEEQIDAHELVNSTRRRYGL..
UPI000B9A2F6E_2  ........TER.ESRAISVAEAIHSAHLEGGDVTTA.FLTDARDYIEEKISIHELLNRTRVRYGL..
S6CE34_2_50      .............DRERAMKEAIHSGEMEGAYVSAE.FREDADEYVAGEISIEELMKRTKRRW....
A0A1Y4HTY9_2_52  .............DRKRLMEEAIHSGEMEGAYVSAE.FREDADEYVKGDISIEDLMKRTKRRWKV..
A0A1Y4AN57_2_51  .............DRKRLMEEAIHSGEMEGAYVSEE.FRKDADEYVKGDISIEDLMRRTKRRWI...
A0A229VUX5_2_52  .............DRKRIMEEAIHSGEMEGAYVSAE.FRSDAEQYVKGDFTIEELMTRTKRRWKI..
UPI000422B7EF_2  .............DRKRAVEESVHSSEMEGAYVSAE.FRKDAEEYVKGDIPIEELMRRTKRRW....
A0A0K1F2T6_2_50  .............DRKRVIDESIHSGEMEGAYVSAE.FREDADEYVRGHIPIEELMRRTKRRW....
A0A0K1F0R0_68_1  .............DRKRVIDESIHSGEMEGAYVSAE.FRKDADEYVKGSIPIEELMRRTKRRW....
UPI0003B683F4_2  .............DRERAMREAIHSGEMEGAWVSAE.FRGDAEAYIAGEITLEELGERTRRRW....
A0A1Y4GJM4_13_6  .............NRKRVIEEAMHSGEMEGSCVSAE.FRKDADEYVDGDISVEELMRRTKRRWK...
M7MPG6_4_56      .........T..EQRSKWVNEAAHSMNLEGLEVSDE.YRAESEEYVTGKVTSDELVRKARARFGL..
A0A0K2GYT1_2_56  ........NTQ.QLRKNSVEQAKHSSMLEGLITSRA.YDEDAERYVVGEISATQLVERTRARFGL..
M1UJ93_3_56      ........SQ..ERRSTWVAQANHSLYLEGLLTSKE.YDGDANSYVVGKISADQLVEKTRARFGL..
A0A134BHL8_4_51  .............TRRENVARAIHSAELEGQAVSAD.FMHDATEYIKGTINADQLVERTKTR.....
D6SWL8_4_59      ........KETQRRRQENLGLAIASGRLEGNSPSKE.FLYDANQYAKGLITSDEFVARMRSRYGV..
I4MFV0_4_59      ........KETQKFRLENVAIALSSAKLEGGMVSSA.CLADTRKYIRGSISADELISLTRKRYGL..
A0A0B5DAF2_2_59  .....TRTTTT.TERARRVSVARHNGRMEGVEVTPA.MRADAEAYISGALGVTELLARTRARYGL..
```



```
A0A0K1RDR7_10_6    ........KDT.VSRQKLFEEAVVNNRLEGLTVSAP.TRADGQRWVDGLISADELVKLTRARYGL..
UPI0004BFDBC1_5    .......DEER.ALRENDVRQARASTELEGGRTSDE.ARAIQDAYARGEIDQDELMRRTRAY.....
A0A1X7F851_6_54    ........EER.AQRADDAQRIRHSTQLEGGRTSDA.ARAIQDKYVRGEIDAEELMRQT........
A0A085LG74_6_56    ........EER.AQRAADAARIRHSTELEGGRTDDA.ARAIQDMYVRGEIDGDDLIRRTKA......
UPI00030F781F_1    .......EQEY.VARVADARRIRHSTEMEGGRTSDV.ARAEQDRYVRGEITGDELIALIKARH....
A0A087E207_9_58    .............KLAEDVYQAVHSVEMEGGRVSPE.FMGDARDYVNGLIDIDQWEGKTLARFK...
UPI00054B51A2_7    ..........P.EDRRRAVGSASASVRAEGLSTTAA.YDDDAAAYVAGRISADELVARAEARHRV..
UPI0005264D53_8    ..........P.EARRRAVDSATGSVRAEGLTPSAE.YLTDAEEYAAGRITADELVQRAEARHRV..
A0A0F2TB43_8_58    ............ESRRRALSSALGSVQAEGLTPTAD.YDADAEEYAAGTISADELVARAEARHR...
UPI0006909515_5    .......HNDQ.AQRRAAVESVRGSLAAEGLETSTA.YDDDAERYVSGSLTADELVARAEQRYR...
UPI00068E600D_1    ......QNEQA.QRRQAAVASAIGSTEAEGLRVSAD.TTTDLDAYARGELDAAALRARTLARY....
A0A1S2K8L1         ......QNEQA.RRRQAAVASAIGSTEAEGLRVTAD.TTADLDAYARGELDAAALRARTLARH....
UPI00068E8438_1    ......QNEQA.RRRQAAVASAIGSTEAEGLSITAD.TTADLDAYARGELDAAALRARTLARY....
UPI00068DBC07_1    .......NKQA.LRRQAAVASAIGSTEAEGLRVSAD.TTAELDAYARGELDAAALRARTLARY....
A0A1A9DPI1         .......MDQA.QRRQAAVASAIGSTEAEGLRVSAD.TTADLDAYARGELDATALRAHTLARY....
UPI0006936365_1    ......QNEQA.QRRRAAVASAVRSTEVEGLRVSAD.TTADLNAYARGELDAAALRARALARH....
A0A0U5F8I8_7_58    ........SDT.ERRLQAVQDALASAAREGQRVSPA.AAADLEAYVRGEITTDEIRQRVMAR.....
UPI0004BB2FA2_1    ........EEV.ERRRKMVRSAFRSNAMEGLRPNPA.CQPIFEAYIAGEIELDEMMPRVKA......
A0A1Z1EGZ4_15_6    ........EEM.ERRRKIVRRAFRSNAMEGLRPNPA.CQPAFDAYIAGEIELDELGPRIDA......
UPI0009B0A7EF_6    .......FQER.AQRRKNVEMAIHEARLEGLEVSPD.ALEDFEDYIAGKATSEEIVARSIARY....
UPI0007DAF1C3_9    .........EE.MRRQGSVAAAIHSARLEGAEVSQE.TEDDMQEYVEGRITVEEGLARLQER.....
UPI0007DAEBAF_9    .......AAER.AQRVRRYAAARHSAELEGGYIATD.KQQPFDDYVEGRIDENELVRRVRQ......
M5B620_15_66       ........AER.SWRRQVVDETRASTALEGGSSTDA.MRELQEQWVDGRITADELVAGARRL.....
UPI0008266A0A_1    ........TEC.LRRQQIIDEARTSTELEGGRSSDE.VHALQDRWVRGELTWQEVGAEVRRLH....
A0A1W6RH10_4_56    .........DE.ESRAGHVRAAIASARIEGVEISPA.TLVEFDAFASGAIDADELVARILALHG...
A0A177SVD8_2_59    .....TTAHEA.AQRSSRVAHVQATNNLEGVRVSAY.MSSKMVDYEKGRISSAELVAAVKARYGI..
consensus>70       ..............r...v.....s.elEG..........d...y..g.i...el..r.r.ry....
```

C

```
                   1         10        20        30        40        50
3SHG_chainA        ......VRKYEGSNDPYTDPETGVMYNLLGIKDQARLERVESAFAYIRSFELGRTS..IS
toxin_3641c        MPHPWDTGDHERNWQGYFIPAMSVLRNRVGARTHAELRDAENDLVEARVIELREDPNLLG
consensus>70       ..........E...#.Y..P...V$.N..G....A.L...E......R..EL........

                    60        70        80        90        100       110
3SHG_chainA        GKFDLDHMKKIHKKLFGDVYEWAGKTRLVDIVKDNSKFAHYTQIESYAPQITQQLAREQH
toxin_3641c        DRTDLAYLRAIHRQLFQDIYVWAGDLRTVGIEKEDESFCAPGGISRPMEHVAAEIYQLDR
consensus>70       ...DL..$..IH..LF.D!Y.WAG..R.V.I.K##..F.....I......!..#....#.

                      120       130       140       150       160       170
3SHG_chainA        LRGLDANEFSQRAGYYMGELNALHPFREGNGRTLREFIWQLAREAGYHIDWDRVERQEMT
toxin_3641c        LRAVGEGDLAGQVAYRYDYVNYAHPFREGNGRSTREFFDLLLSERGSGLDWGKTDLEELH
consensus>70       LR.....#......Y.....N..HPFREGNGR..REF...L..E.G...DW...#.#E$.

                          180       190
3SHG_chainA        RASIESY.YGNSDLMSALIRRNLTEFT....
toxin_3641c        GACHVARANSDLTGLVAMFKGILDAEPTYDF
consensus>70       .A........#...$.A$....L........
```

D

```
3SHG_chainB        1   ..V.LSE.EEIEYRRRDARNALASQRLEGLEPDPQVVAQMERVVVGELETSDVIKDLMER
antitoxin_3642c    1   MFVQATELQKVKRRFRNVRATRRNTELEGTRSTAATRADQNDYARGKITAAELGERVRRR
consensus>70           ..V...E.#.!..R.R#.R.......LEG........A#.#....G.....#.......R

3SHG_chainB        57  IKREE
antitoxin_3642c    61  YNIQ.
consensus>70           ...#.
```

Figure S14: Multiple sequence alignment of VbhT sequences from PFAM (PF02661) with Rv3641c in **(A)** and VbhA with Rv3642c in **(B)** shows the conservation of signature motifs characteristic of these two proteins. VbhT (3shg) from *Bartonella Schoenbuchensis* serves as the reference structure in the alignment that was generated using Promals3D. Signature motifs in the two alignments are marked with a blue box on top of the alignment. Consensus sequence (at a threshold > 70%) is shown at the bottom of the alignment. Solvent accessibility of the reference structure, shown using the Esprit colour convention with dark blue indicating exposed residues, cyan showing partially exposed residues and white for buried residues (and red for residues for which this value could not be determined), is shown at the bottom of the alignment. In **(C)** and **(D)** are shown the alignments for the toxin (Rv3641c) and antitoxin (Rv3642c) with template 3shg. This alignment was employed to model the protein.

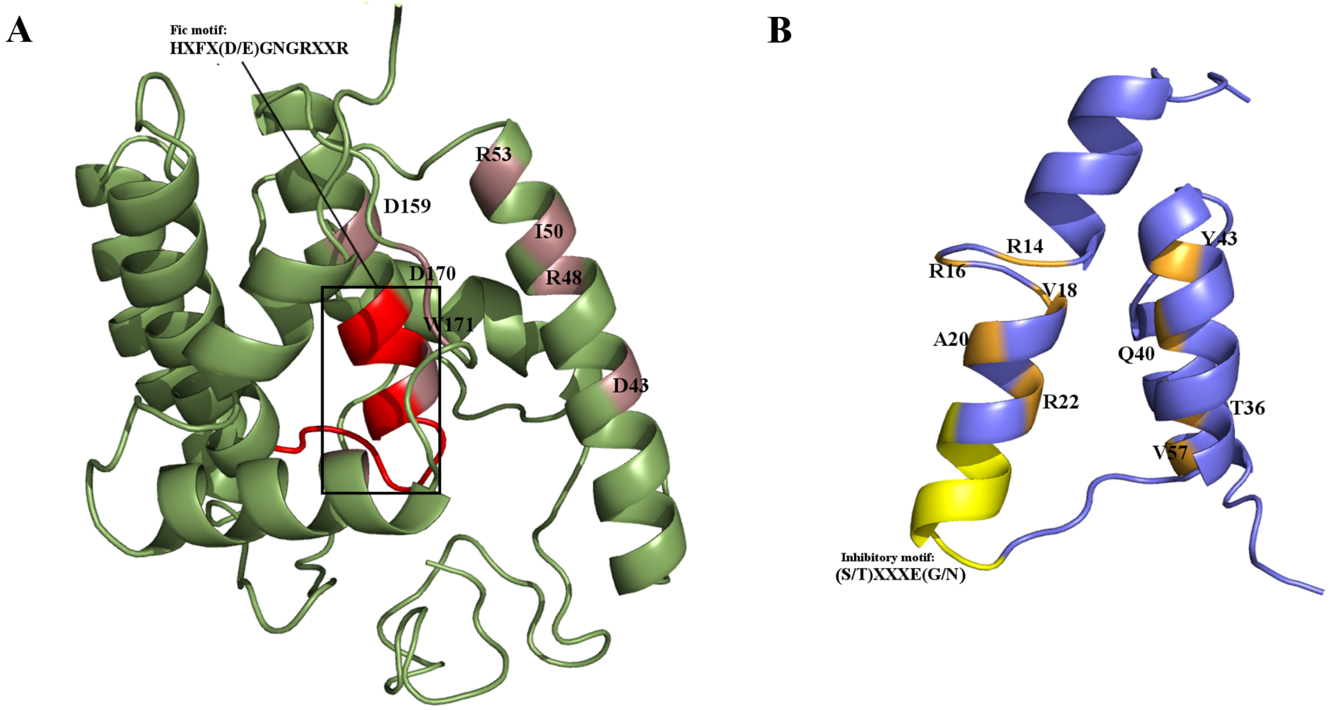


Figure S15: Structural model of putative VbhTA in *M. tuberculosis*. Cartoon representation of the comparative models of Rv3641c in green **(A)** and Rv3642c in blue **(B)** were derived using the template VbhTA (3shg) from *Bartonella schoenbuchensis*. Predicted Fic motif and inhibitory motif are shown in red and yellow respectively. Residues predicted to form the interaction interface between Rv3641c-Rv3642c are shown and colored in pink on the toxin and orange on the antitoxin.