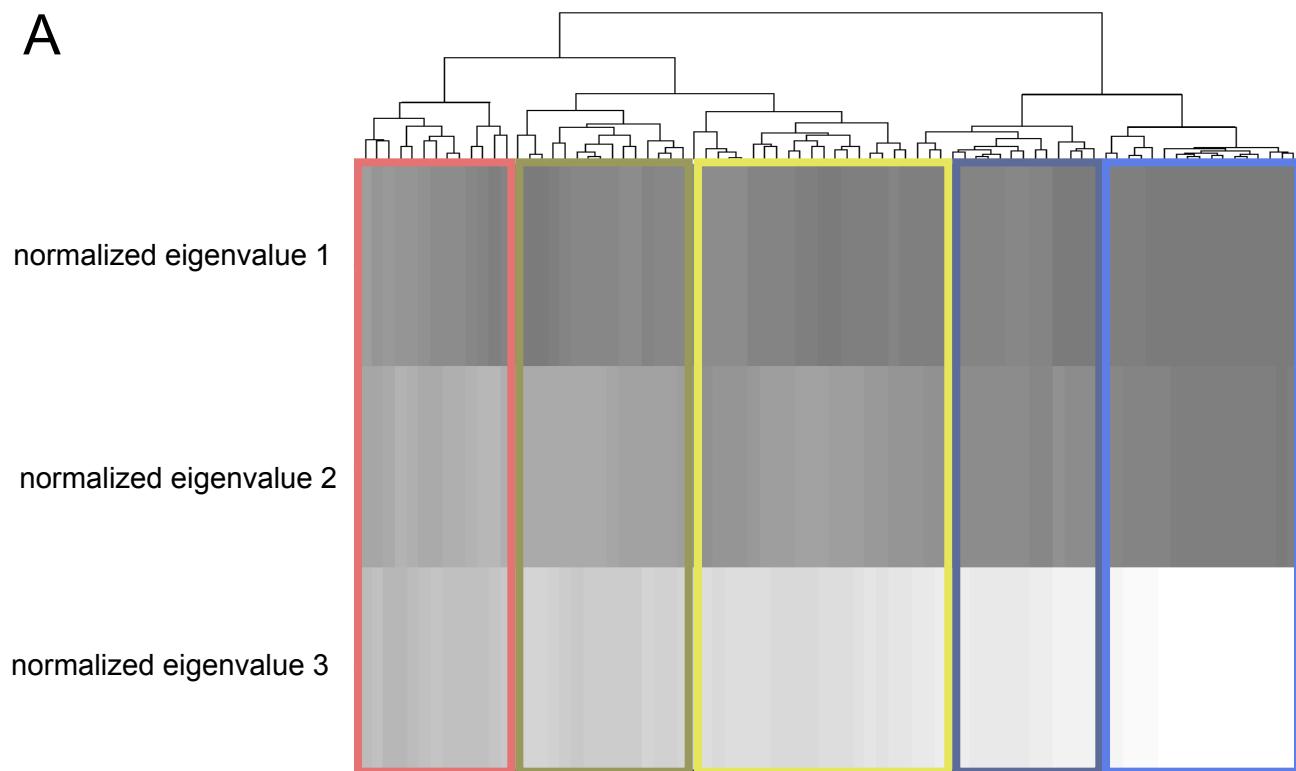


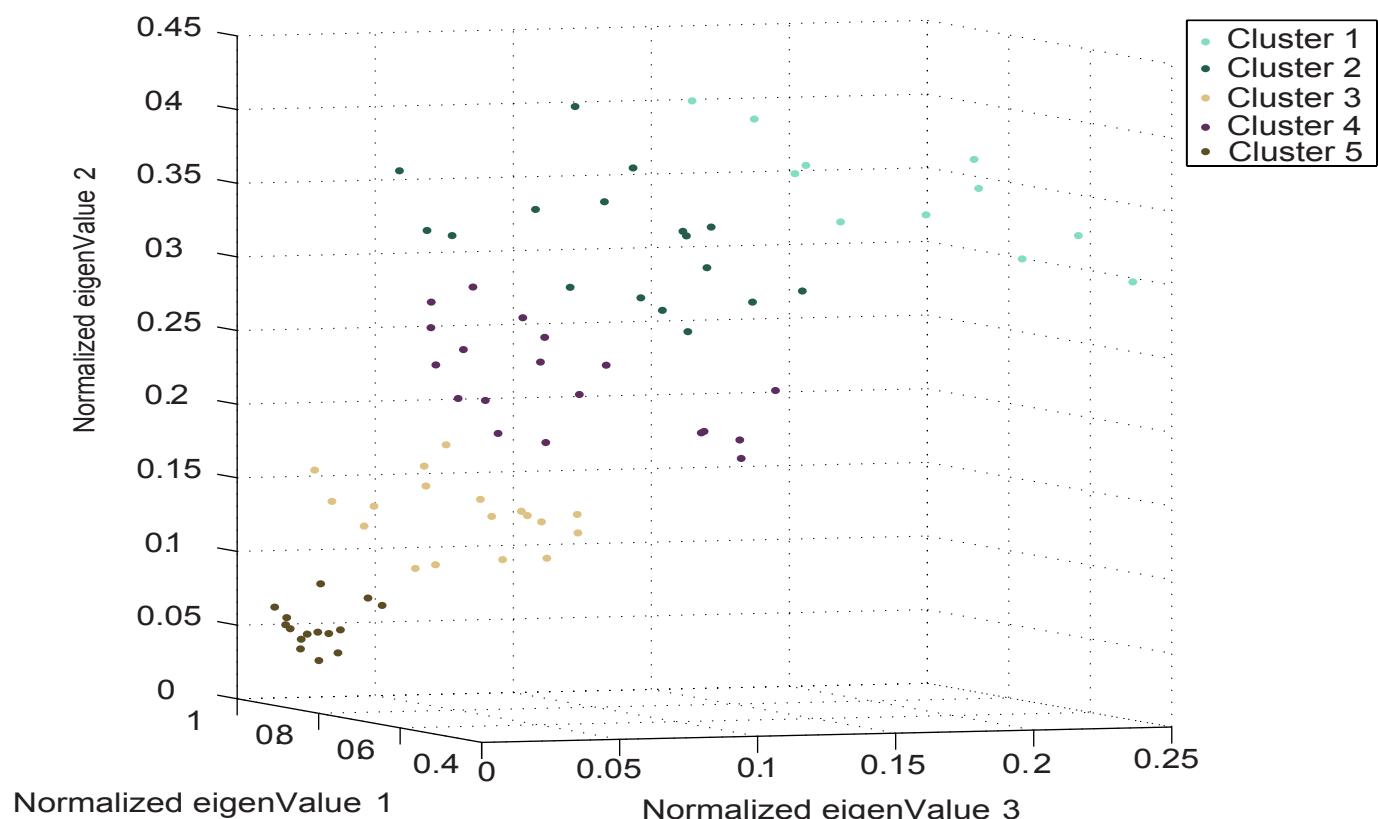
**Figure 1: Membrane Protein Subgroup Amino Acid Enrichment in Co-Evolving Sectors.**

Heatmap depicts fold enrichment of amino acid in subgroups. Lighter color represent depletion in a given amino acid and red depicts higher fold enrichment. A value of one describes no enrichment. The last row depicts the enrichment for the Transmembrane subgroup we created.

A



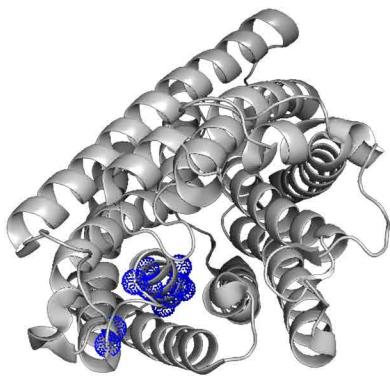
B



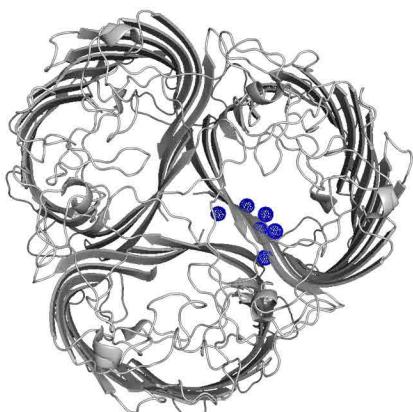
**Figure 2: Clustering of the transmembrane sector database using unsupervised hierarchical clustering (A) as well as supervised clustering (B) methods**

The transmembrane sector database did not obviously cluster, however 5 clades could be identified from the dendrogram (A). Using K-mean to separate sectors, 5 clusters could be defined (B). Each cluster contained sectors with similar shape and deformation.

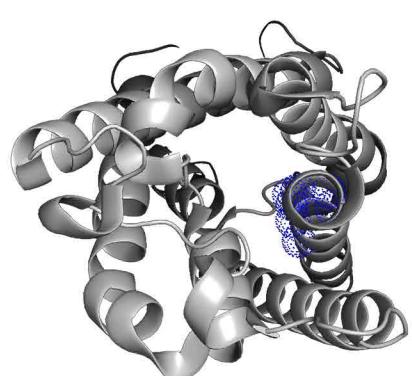
**Cluster 3**



1U7G  
Ammonia channel

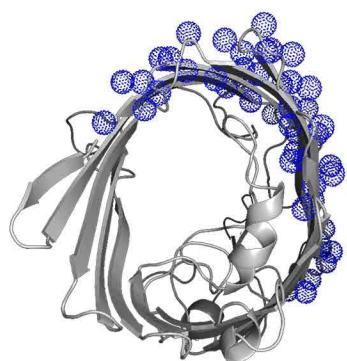


1MAL  
Maltoporin channel

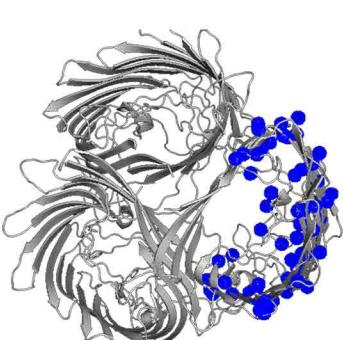


1RC2  
Aquaporin

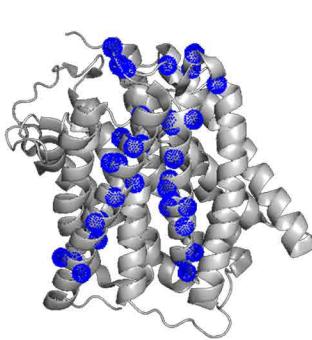
**Cluster 4**



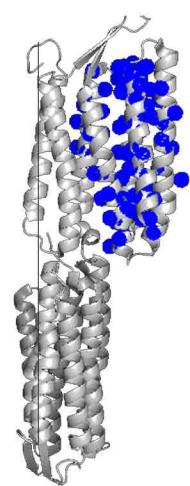
2OMF  
OmpF porin



1MAL  
Maltoporin channel

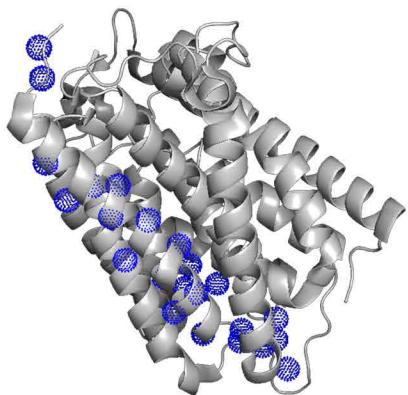


1U7G  
Ammonia channel

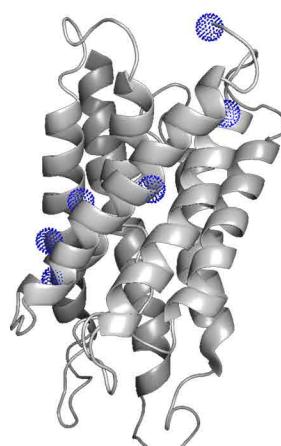


1VGO  
Archaerhodopsin

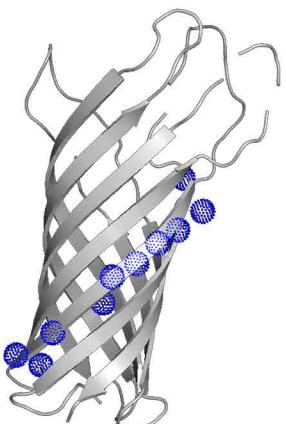
**Cluster 5**



1U7G  
Ammonia channel



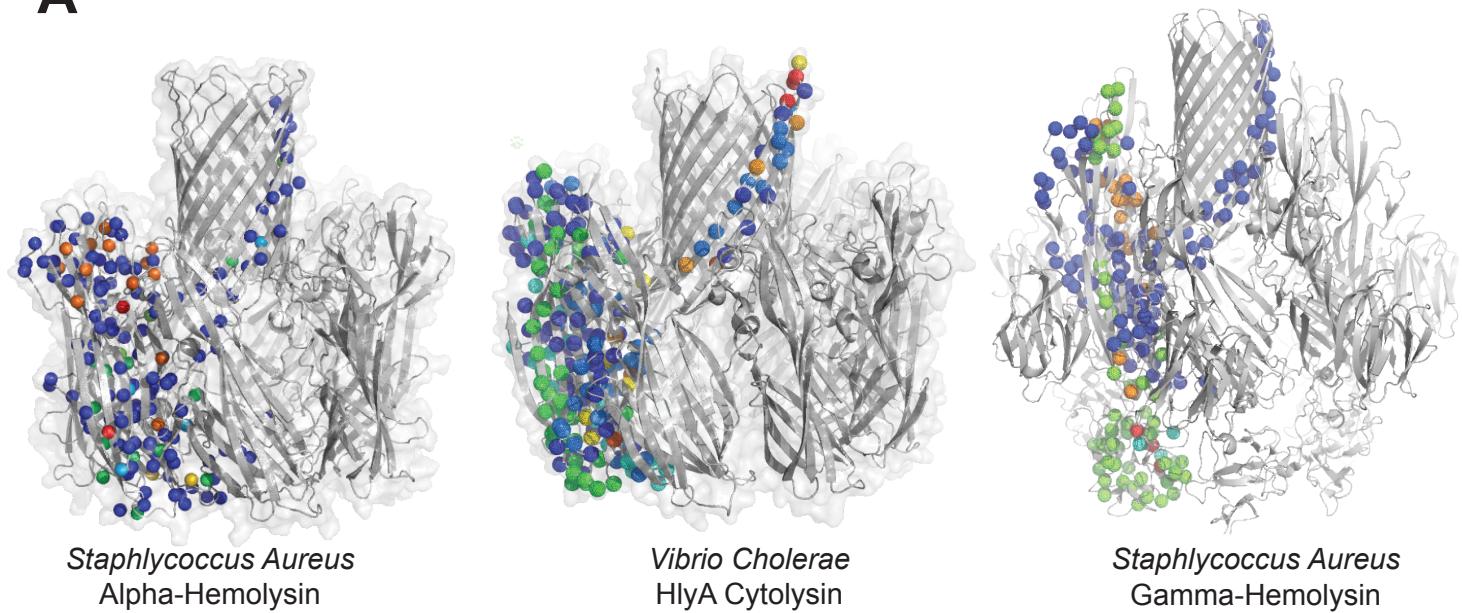
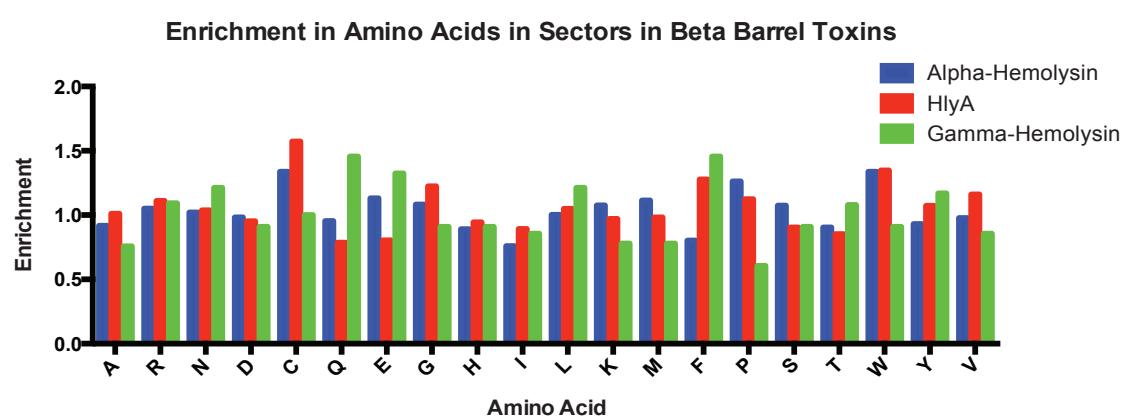
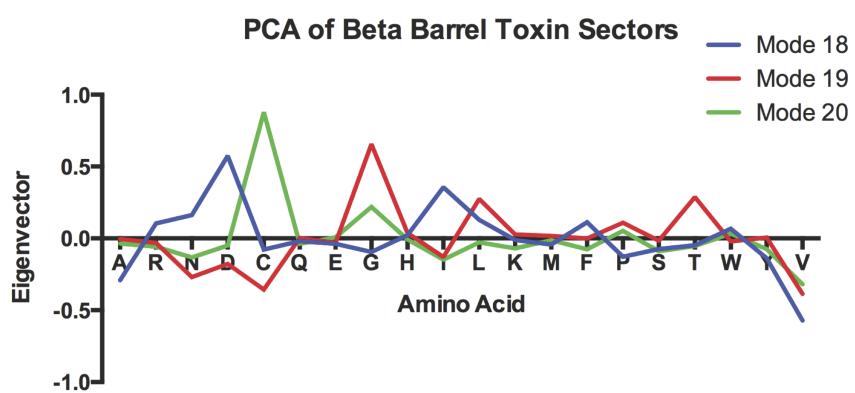
1FQY  
Aquaporin



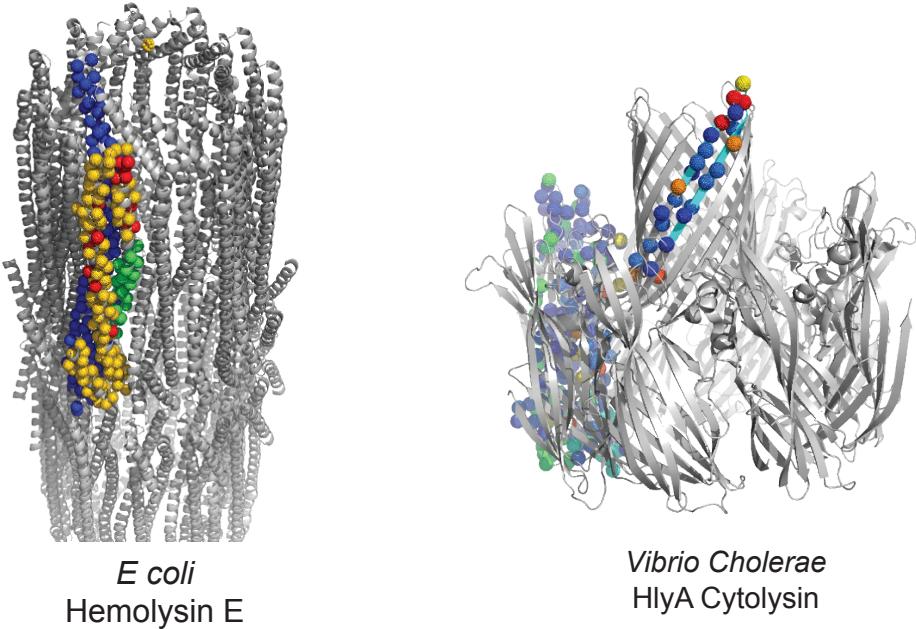
1BXW  
Outer membrane protein A

**Figure 3: Examples of sectors in cluster 3, 4 and 5**

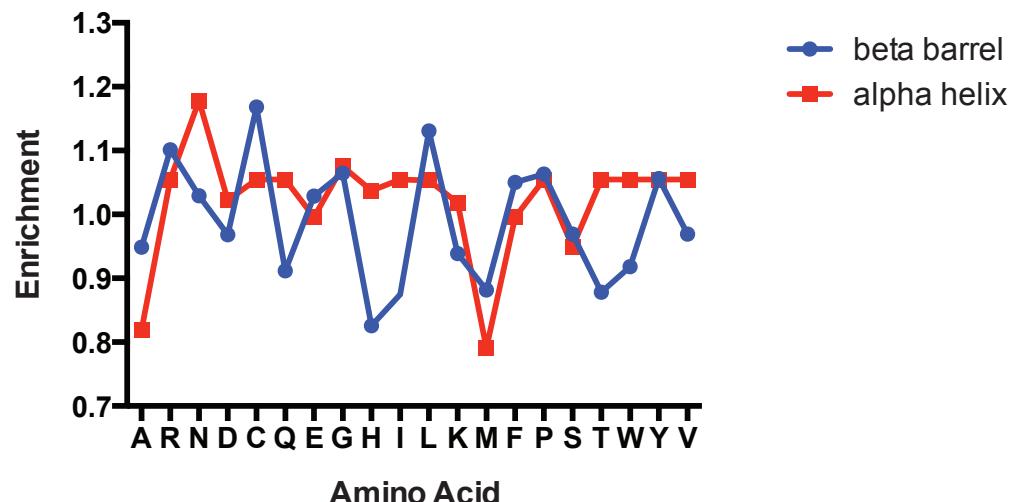
The sector structures are conserved across multiple protein families.

**A****B****C**

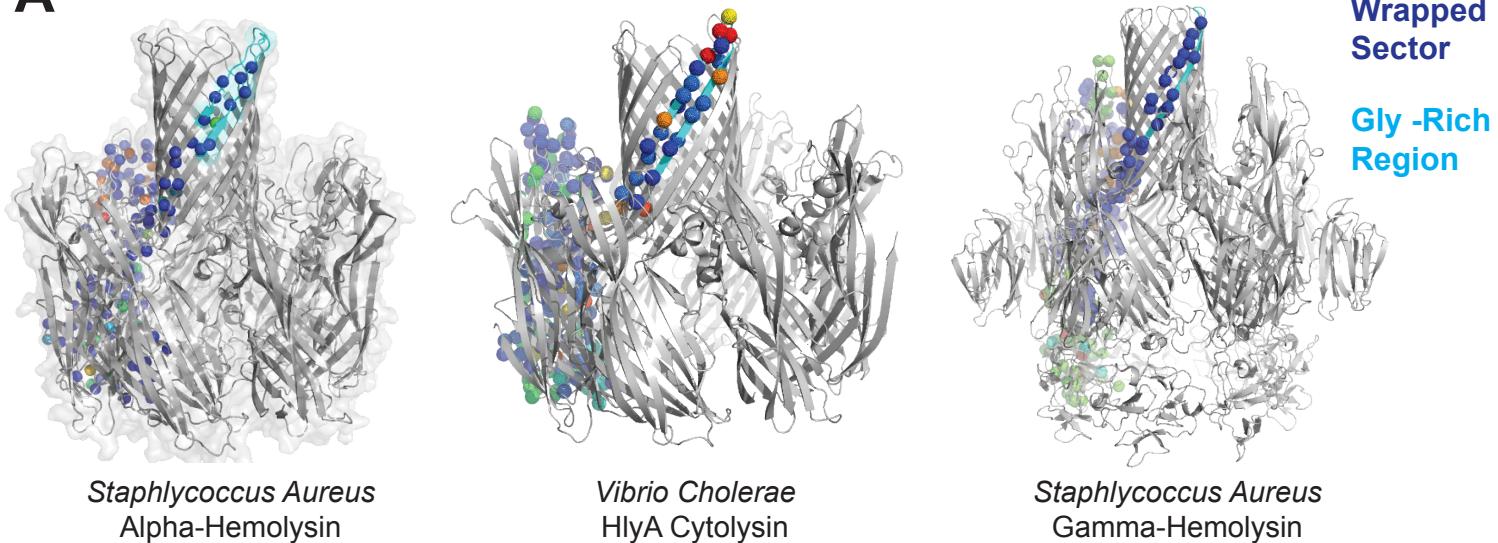
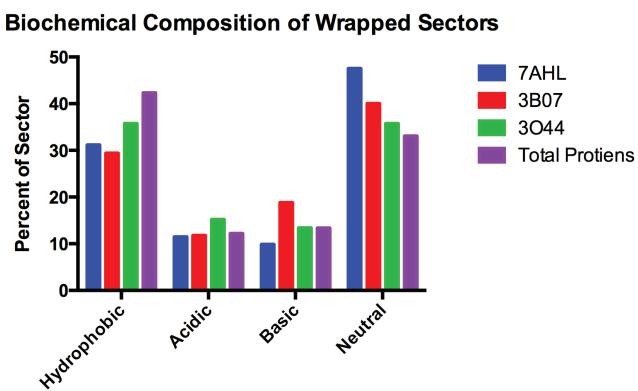
**Figure 4: Sectors in ‘Beta-Barrel Pore Forming Toxins’ have similar structure.** **A** Mapped sectors co-evolving sectors on Beta-Barrel Pore Forming Toxins. Similarly colored spheres represent the spatial coordinates of the amino acids within a single sector. PDB IDs from left: 7HAL, 3B07, 3O44. **B** Amino acid enrichment of all sectors in beta barrel toxins. There is enrichment in differential residues in the different proteins. There is an increase in Phe in all of the sectors across these toxins. **C** Residue principle components analysis was performed on all beta barrel toxins in the membrane SDB. Mode 20 accounts for 19 percent of the variance, mode 19 accounts for 14.5 percent of the variance, and mode 18 accounts for 11.3 percent of the data.

**A****B**

### Amino Acid Enrichment for Membrane Toxins

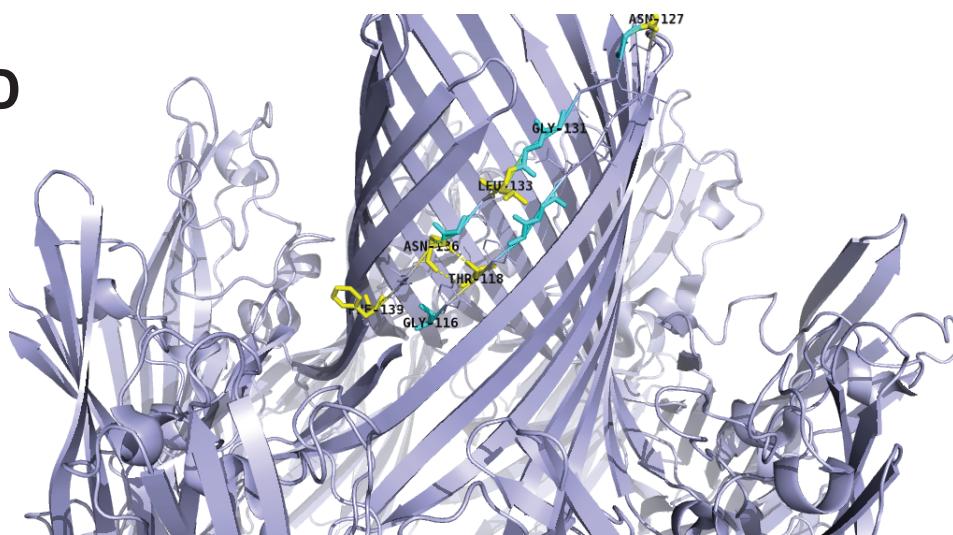


**Figure 5: Functionally similar subgroups differ in sector structure organization as well as in amino acid enrichment. A** Co-evolved sectors mapped on to Hemolysin E (left PDB: 2WCD) and HlyA (right). **B** Comparison of the median enrichment of beta-barrel toxins (blue) and alpha-helix toxins (red).

**A****B****C**

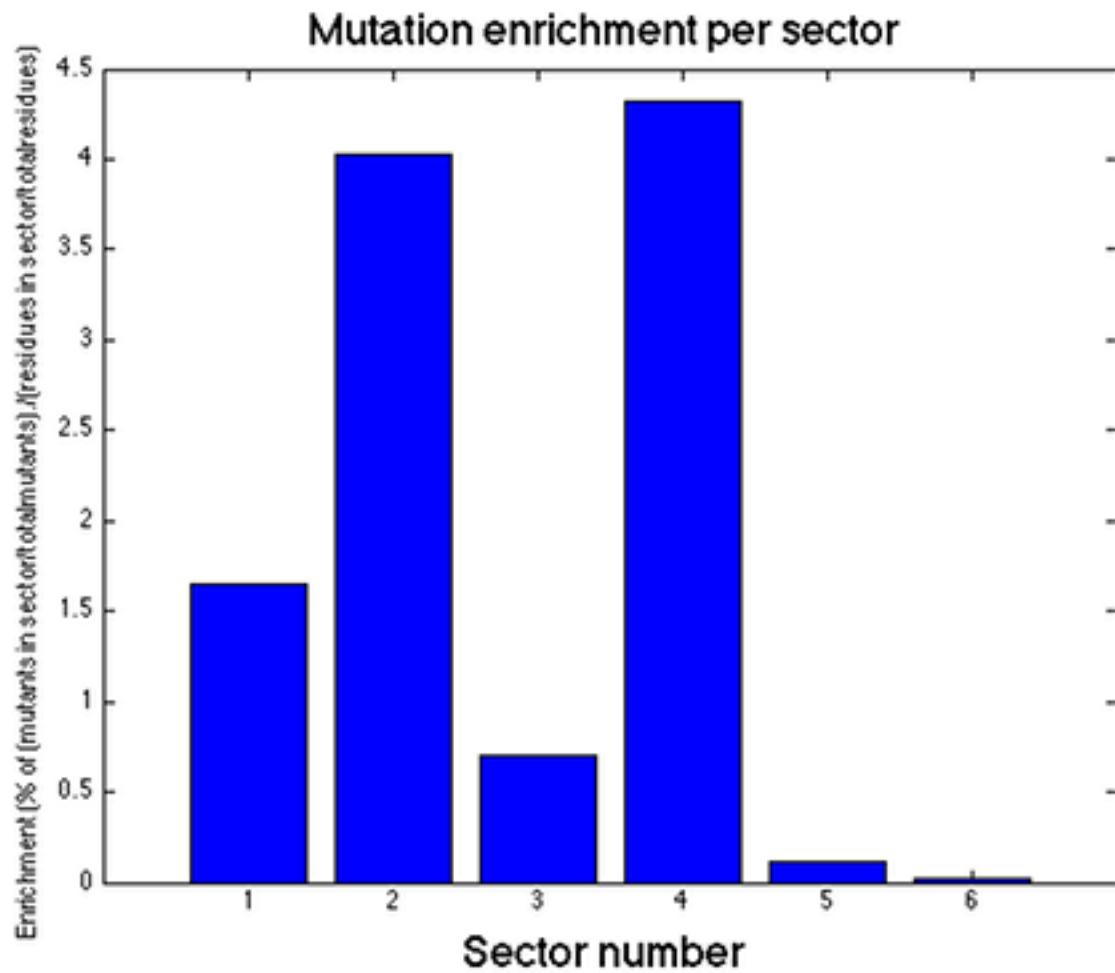
	Gly-Rich Entry Sequence
Alpha Hemolysin	<b>G<b>E</b>N<b>G</b>N<b>V</b>T<b>G</b>D<b>D</b>T<b>G</b>K<b>I</b><b>G</b>L<b>I</b><b>G</b><b>A</b>N<b>V</b>S<b>I</b><b>G</b></b>
HlyA	<b>G<b>Y</b>T<b>F</b><b>G</b><b>G</b>D<b>I</b><b>S</b><b>I</b><b>S</b><b>N</b><b>G</b><b>L</b><b>S</b><b>G</b><b>G</b><b>L</b><b>N</b><b>G</b><b>N</b><b>T</b><b>A</b><b>F</b><b>S</b></b>
Gamma Hemolysin	<b>G<b>F</b>E<b>L</b><b>G</b><b>V</b>T<b>G</b><b>G</b>V<b>E</b>V<b>S</b><b>G</b><b>D</b><b>G</b>P<b>K</b><b>A</b><b>K</b><b>L</b><b>E</b><b>A</b><b>R</b></b>

In wrapped sector  
Glycine not in Sector  
Neither

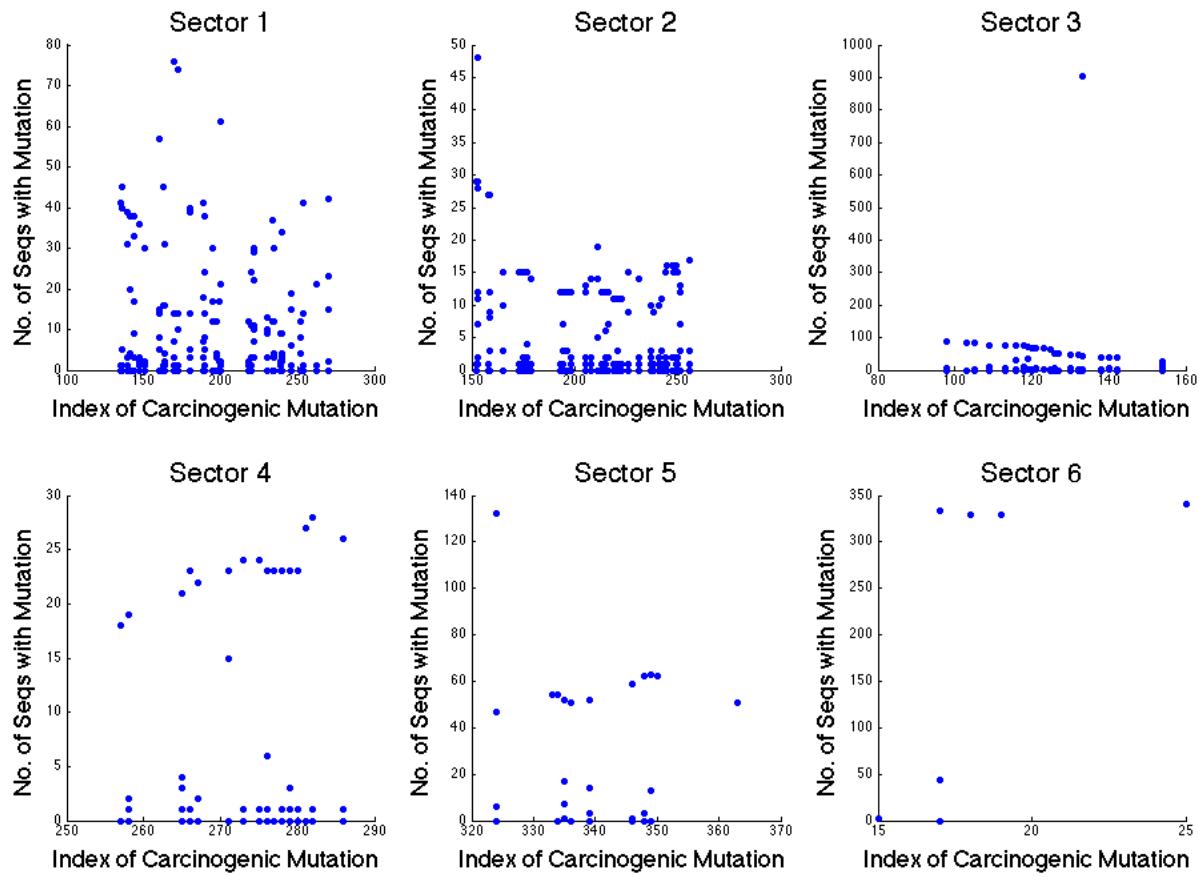
**D**

**Figure 6: 'Wrapped Sectors' reveal co-evolution of functional entry domains in beta barrel toxins.**

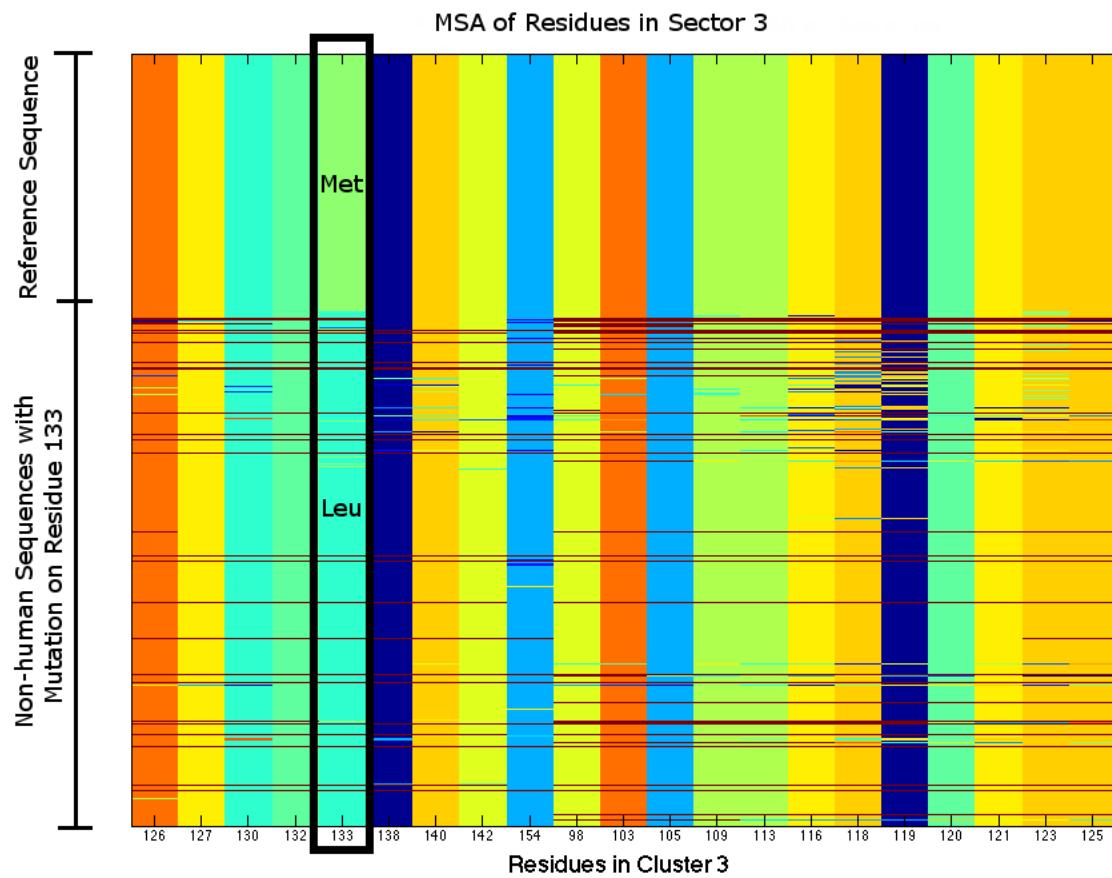
AGly rich regions (cyan) mapped onto structures of beta barrel toxin proteins. Gly rich region co-localizes with the beta barrel portion of the wrapped sector. Only the gly rich region is annotated for Alpha Hemolysin. **B** biochemical composition of wrapped sector. **C** Table of Identified residues of Gly-rich regions in wrapped sector. **D** Magnification of intersector amino acid interaction within HlyA. Cyan are glycines, yellow are sector residues, Asn136 forms a hydrogen bond with Thr118 .



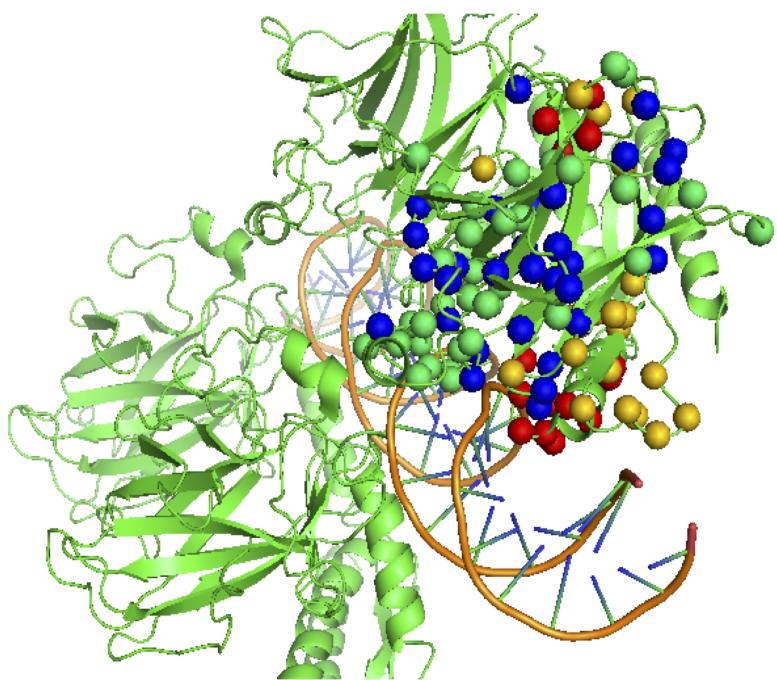
**Figure 7** Enrichment of Mutations in Sectors. Sector 1, 2 and 4 are significantly enriched in mutations. Among them, sectors 2 and 4 have odds ratios over 4. On the other hand, sectors 3, 5 and 6 are significantly depleted in mutations.



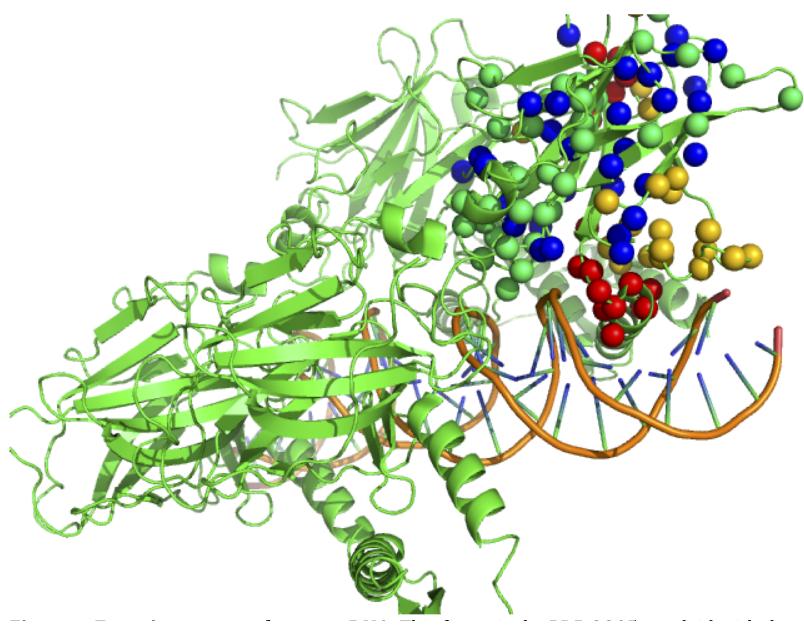
**Figure 8** Count of mutations existing in non-human species: The numbers of sequences with mutations vary with the sector. Although they are not significantly different from each other, extreme values exist in sector 3 and 6, with at least a third of the non-human sequences containing the mutations.



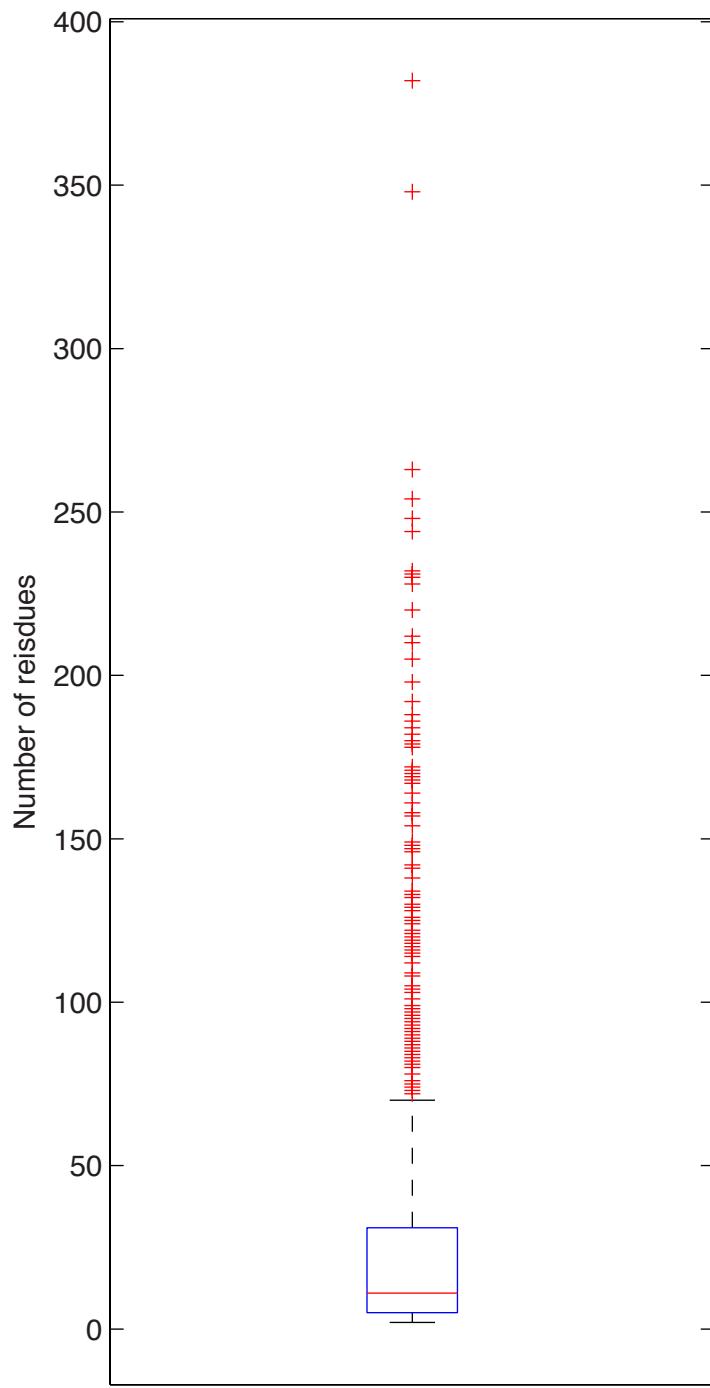
**Figure 9** MSA of Sector 3 Residues. Based on this visualization of residues in Sector 3, there is no other mutation in the sector with abundant existence in the non-human species.



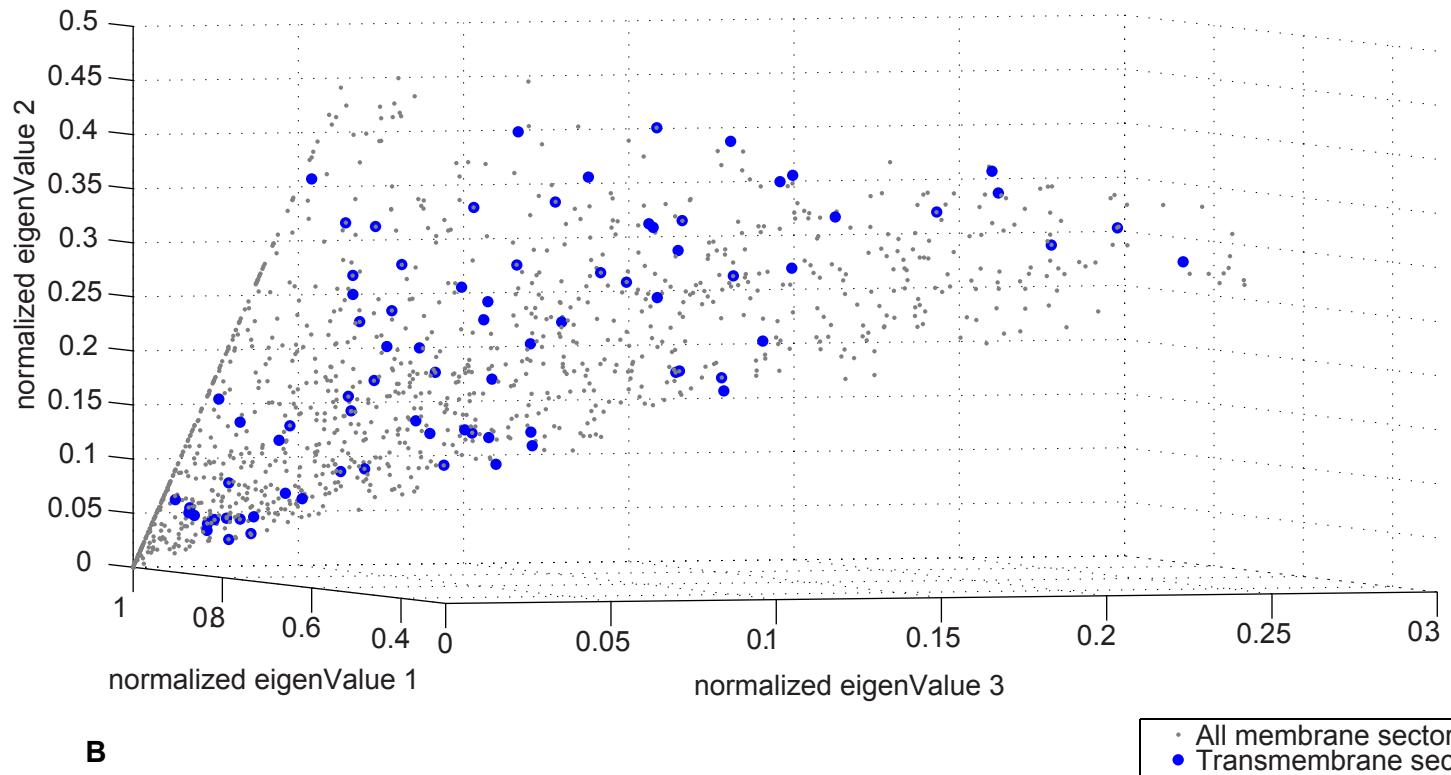
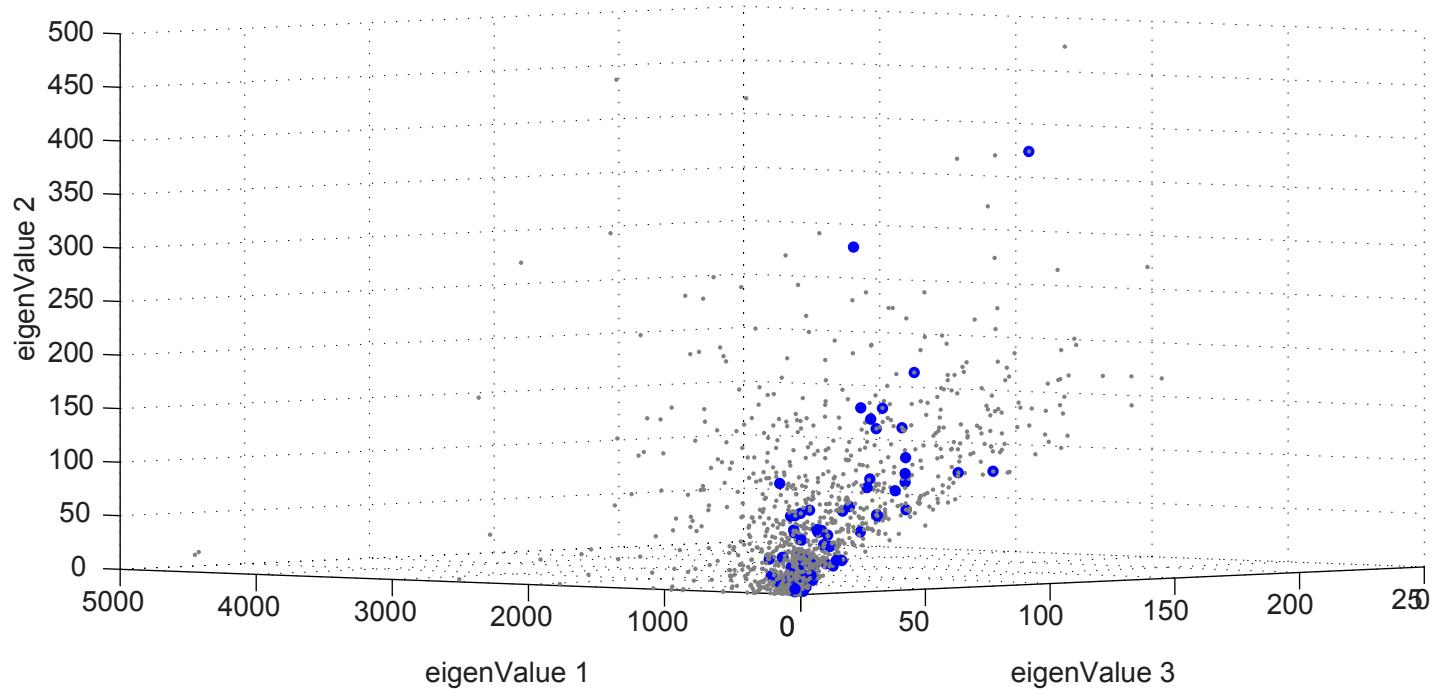
**Figure 10** Sector overlay of p53 sectors 1-4 on PDB 3Q05 complexed with DNA.  
Sectors 1-4 are overlaid onto the DNA binding domain of p53. Sector 1 is colored blue, sector 2 is green, sector 3 is green, and sector 4 is red.



**Figure 11** This figure is the PDB 3Q05 overlaid with the first four p53 sectors. Sector one is blue, sector 2 is green, sector 3 is yellow, and sector 4 is red. The figure show that sector 4 and sector 2 have a higher concentration of residues next to the DNA.

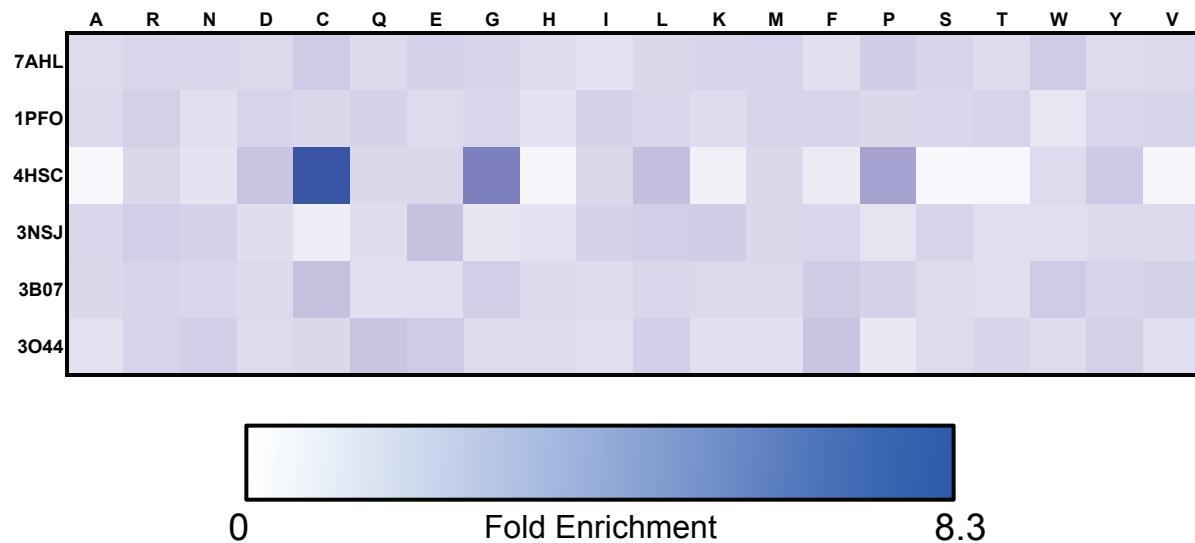


**Figure S1: Sector length distribution in the membrane sector database**

**A****B**

**Figure S2: The normalized eigenvalue (A) and raw eigenvalue (B) distributions from the principal component analysis of the sector coordinates can be used as a proxy for the shape of the sectors**

Sectors in the membrane and transmembrane sector databases have a smooth distribution of eigenvalues, no group of sectors sharing the same relative deformation can be easily detected.



**Figure S3: Heatmap of amino acid enrichment in all Beta Barrel toxins.**

	Number of proteins	Number of sectors		Number of proteins	Number of sectors
<b>ALPHA-HELICAL PROTEINS</b>	<b>154</b>	<b>653</b>			
G Protein-Coupled Receptors (GPCRs)	15	40	Light-Harvesting Complexes	2	5
Channels: Aquaporins and Glyceroporins	10	47	Cys-Loop Receptor Family	2	11
ATP Binding Cassette (ABC) Transporters	10	55	Betaine/Choline/Carnitine Transporter (BCCT) Family	2	10
Channels: Potassium, Sodium & Proton Ion-Selective	10	65	Cytochromes P450	1	5
Major Facilitator Superfamily (MFS) Transporters	9	62	Electron Transport Chain Complexes: Complex III (Cytochrome bc)	1	2
Bacterial and Algal Rhodopsins	7	14	Electron Transport Chain Complexes: Complex IV (Cytochrome C Oxidase)	1	1
Channels: Other Ion Channels	7	45	Solute Sodium Symporter (SSS) Family	1	4
Multi-Drug Efflux Transporters	6	32	Energy-Coupling Factor (ECF) Transporters	1	5
Intramembrane Proteases	6	20	Outer Membrane Proteins	1	5
Amino Acid Secondary Transporters	5	6	Channels: Transient Receptor Potential (TRP)	1	15
F-type ATPase	4	5	Phosphoenolpyruvate-Dependent Phosphotransferases (PTPs)	1	1
P-type ATPase	4	22	Channels: Intercellular	1	1
Oxidoreductases	4	12	Bacterial V-type ATPase	1	4
Channels: Amt/Rh proteins	3	15	Ca <sup>2+</sup> Cation Antiporter (CaCA) Family	1	2
Channels: Urea Transporters	3	8	Sec and Translocase Proteins	1	3
Electron Transport Chain Complexes: Complex II	3	8	Nucleobase-Cation-Symport-2 (NCS2) Family	1	4
Amino Acid/Polyamine/Organocation (APC) Superfamily	3	18	Antiports	1	7
Channels : Formate/Nitrite Transporter (FNT) Family	3	13	SNARE Protein Family	1	1
Adventitious Membrane Proteins: Alpha-helical Pore-forming Toxins.	2	17	Oxygenases	1	1
CorA Superfamily Ion Transporters	2	13	Claudins	1	2
Solute Carrier (SLC) Transporter Superfamily	2	6	Channels: Gap Junctions	1	3
H <sup>+</sup> /Cl <sup>-</sup> Exchange Transporters	2	10	Membrane-Bound Metalloproteases	1	1
Channels: Calcium Ion-Selective	2	6	Photosystems	1	2
Membrane-Associated Proteins in Eicosanoid and Glutathione Metabolism (MAPEG)	2	6	Superfamily of K <sup>+</sup> Transporters (SKT proteins)	1	8
Apical Sodium-Dependent Bile Acid Transporters (ASBT)	2	2	Cation Diffusion Facilitator (CDF) Family	1	3
<b>MONOTOPIC MEMBRANE PROTEINS</b>	<b>32</b>	<b>187</b>			
Lipoxygenases	5	54	<b>BETA-BARREL PROTEINS</b>	<b>64</b>	<b>365</b>
Dihydroorotate Dehydrogenases (DHODH, class 2)	4	17	Beta-Barrel Membrane Proteins: Monomeric/Dimeric	25	123
Glycosyltransferases	4	18	Beta-Barrel Membrane Proteins: Porins and Relatives	10	59
Oxidases	4	28	Outer Membrane Carboxylate Channels (Occ)	9	26
Peptidases	3	13	Outer Membrane Autotransporters	7	62
Hydrolases	3	4	Adventitious Membrane Proteins: Beta-sheet Pore-forming Toxins	6	48
Dehydrogenases	2	10	Omp85-TpsB Outer Membrane Transporter Superfamily	5	37
Oxidoreductases (Monotopic)	2	12	Beta-Barrel Membrane Proteins: Mitochondrial Outer Membrane	2	10
Cyclooxygenases	2	17			
Squalene-Hopene Cyclases	1	1			
Polymerases	1	1			
Isomerasers	1	12			

**Table 1: Distribution of the proteins and sectors by structural group and functional subgroups in the membrane sector database**  
The membrane sector database contained 250 unique proteins and 1205 sectors.

Sector	1	2	3	4	5	6
Total Mutations	3690	1038	1036	4827	85	9
Expected Number	2233.7	2582.8	1465.9	116.9	767.84	628.24
Chi-squared Value	949.41	23551	126.07	12325	607.25	610.37
p value (df = 1)	$7.80 \times 10^{-202}$	0	$8.64 \times 10^{-25}$	0	$6.35 \times 10^{-128}$	$1.35 \times 10^{-128}$

**Table 2.** Chi-squared test of Mutation Enrichment in Sectors. Chi squared tests were performed on each sector to test the null hypothesis that mutations occur evenly across residues. The expected counts of mutation were generated from the proportions of the sectors in the whole sequence multiplied with the number of total mutations.

Source	SS	df	MS	Chi-sq	p value
Groups	$2.66 \times 10^6$	5	$5.32 \times 10^5$	56.46	$6.54 \times 10^{-11}$
Error	$3.52 \times 10^7$	799	$4.41 \times 10^4$		
Total	$3.79 \times 10^7$	804			

**Table 3.** Kruskal-Wallis Test of Somatic Mutation Abundance in Non-Human Sequences. We used the Kruskal-Wallis test because the data was not parametric. The single factor was sector index, and this table validated that the sectors were a factor on the distribution of mutations in non-human sequences.

Sector	Sector	p value
1	2	0.0012*
1	3	0.984
1	4	$7.49 \times 10^{-6}*$
1	5	0.210
1	6	0.0692
2	3	0.126
2	4	0.376
2	5	$5.53 \times 10^{-4}*$
2	6	0.0037*
3	4	0.0024*
3	5	0.118
3	6	0.0447*
4	5	$1.29 \times 10^{-5}*$
4	6	$5.28 \times 10^{-4}*$
5	6	0.683

**Table 4.** Post-hoc tests of Mutation Abundance in Sectors. Tukey's HSD test is used to compare the average counts of mutated sequences in sectors. Although the average counts from low to high is sector 4 < 2 < 3 < 1 < 5 < 6, the results do not show a clear alternative hypothesis.

Somatic Mutation	Residue	WT AA	Mutant AA	Sector	Compensating Mutation	Residue	WT AA	Mutant AA	Sector	Number of Sequences with Compensatory Mutation Pair
	180	Glu	Gln	1		49	Asn	Gap	0	26
	200	Asn	Gap	1		56	Thr	Asp	0	25
	200	Asn	Gap	1		56	Thr	Gap	0	17
	200	Asn	Gap	1		59	Asn	Thr	0	23
	200	Asn	Gap	1		108	Val	Gly	0	36
	200	Asn	Gap	1		291	Val	Pro	0	35
	220	Tyr	His	1		108	Val	Gly	0	15
	220	Tyr	His	1		291	Val	Pro	0	18
	235	Asn	Ser	1		100	Val	Thr	0	21
	240	Ser	Gly	1		81	Thr	Pro	0	22
	240	Ser	Gly	1		111	Met	Ile	0	24
	270	Phe	Val	1		76	Met	Thr	0	18
	270	Phe	Val	1		80	Ser	Pro	0	27
	270	Phe	Val	1		81	Thr	Pro	0	27
	270	Phe	Val	1		94	Gln	Ile	0	19
	270	Phe	Val	1		111	Met	Ile	0	29
	270	Phe	Val	1		131	Asp	Lys	0	29
	363	Arg	Lys	5		73	Cys	Ser	0	17
	363	Arg	Lys	5		73	Cys	Gap	0	15
	363	Arg	Lys	5		76	Met	Thr	0	21
	363	Arg	Lys	5		80	Ser	Pro	0	35
	363	Arg	Lys	5		81	Thr	Pro	0	26
	363	Arg	Lys	5		90	Val	Met	0	15
	363	Arg	Lys	5		94	Gln	Ile	0	19
	363	Arg	Lys	5		111	Met	Ile	0	23
	363	Arg	Lys	5		111	Met	Leu	0	20
	363	Arg	Lys	5		131	Asp	Asn	0	23
	363	Arg	Lys	5		131	Asp	Lys	0	22

**Table 1.** Compensatory Mutation Candidates. This list of compensatory mutation candidates was generated from the coevolution matrix and somatic mutation list. There is no mutation in sector 2, 3, 4, or 6 that is compensated by other mutations, based on the criteria we used for screening.