Bioinformatics Institute (BII)
*A*STAR Singapore*

Frank Eisenhaber

www.bii.a-star.edu.sg

franke@bii.a-star.edu.sg

Singapore, 13th December 2017

Bioinformatics Institute
A*STAR

Celebrating 15 Years!

New insights into TM-proteins sequence – structure - function

Wong et al., 2010, PLoS Computational Biology, 6(7), doi:10.1371/journal.pcbi.1000867

Wong et al., 2011, Biology Direct, 6(57), doi:10.1186/1745-6150-6-57

Wong et al., 2012, Nucleic Acids Research, 40, W370–W375, doi:10.1093/nar/gks379

Wong et al., 2014, BMC Bioinformatics 15, 166, doi:10.1186/1471-2105-15-166

Baker et al., 2017, BMC Biology, 15, 66, doi 10.1186/s12915-017-0404-4

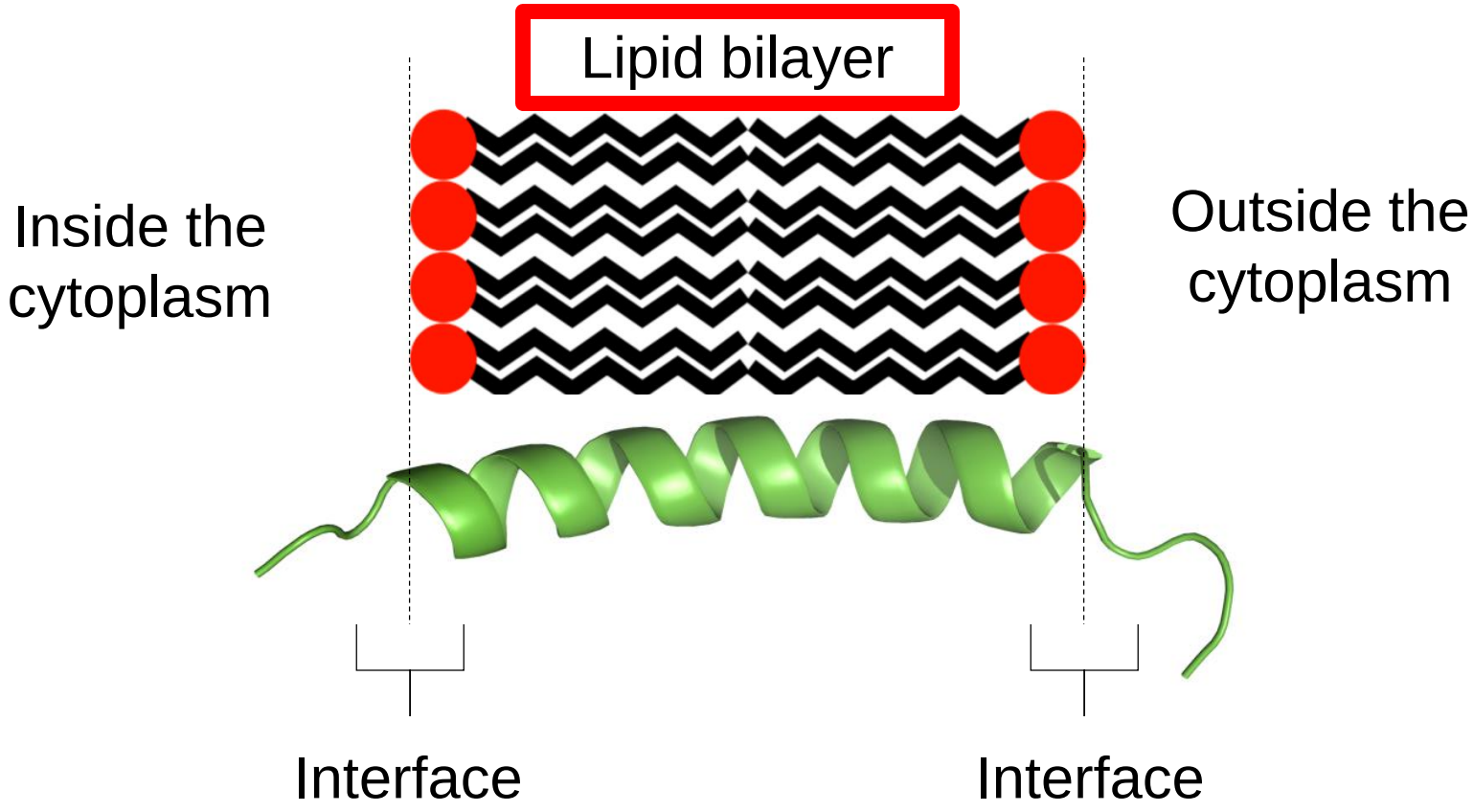
Transmembrane helices. A “negative-not-inside/negative-outside rule” complements the “positive-inside rule”.

James Baker[1,2], Wing Cheong-Wong[1], Birgit Eisenhaber[1], Jim Warwicker[2]*, Frank Eisenhaber[1]*

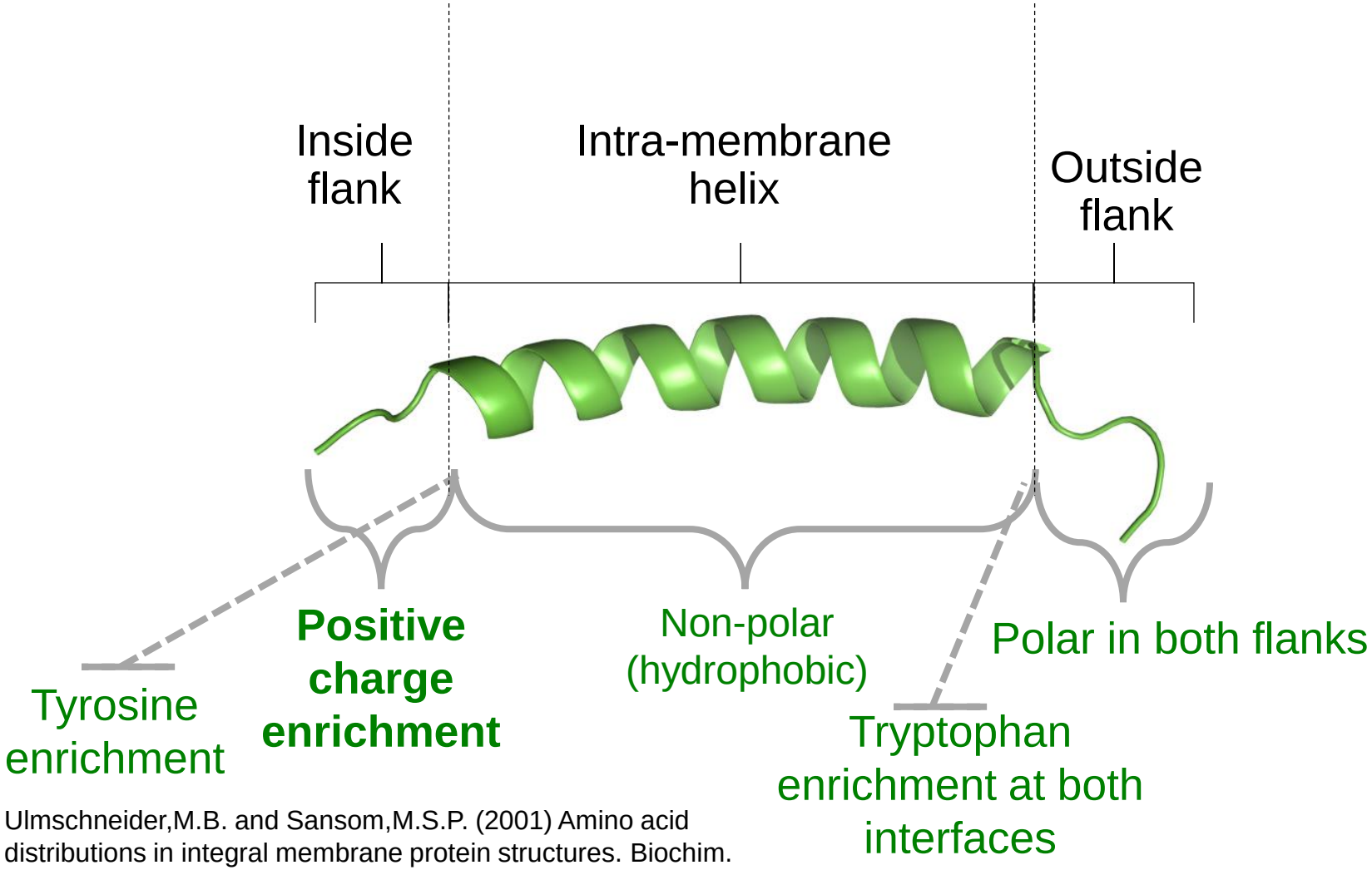
[1]BII at A*STAR, Singapore

[2]MIB at Manchester, UK

Introduction



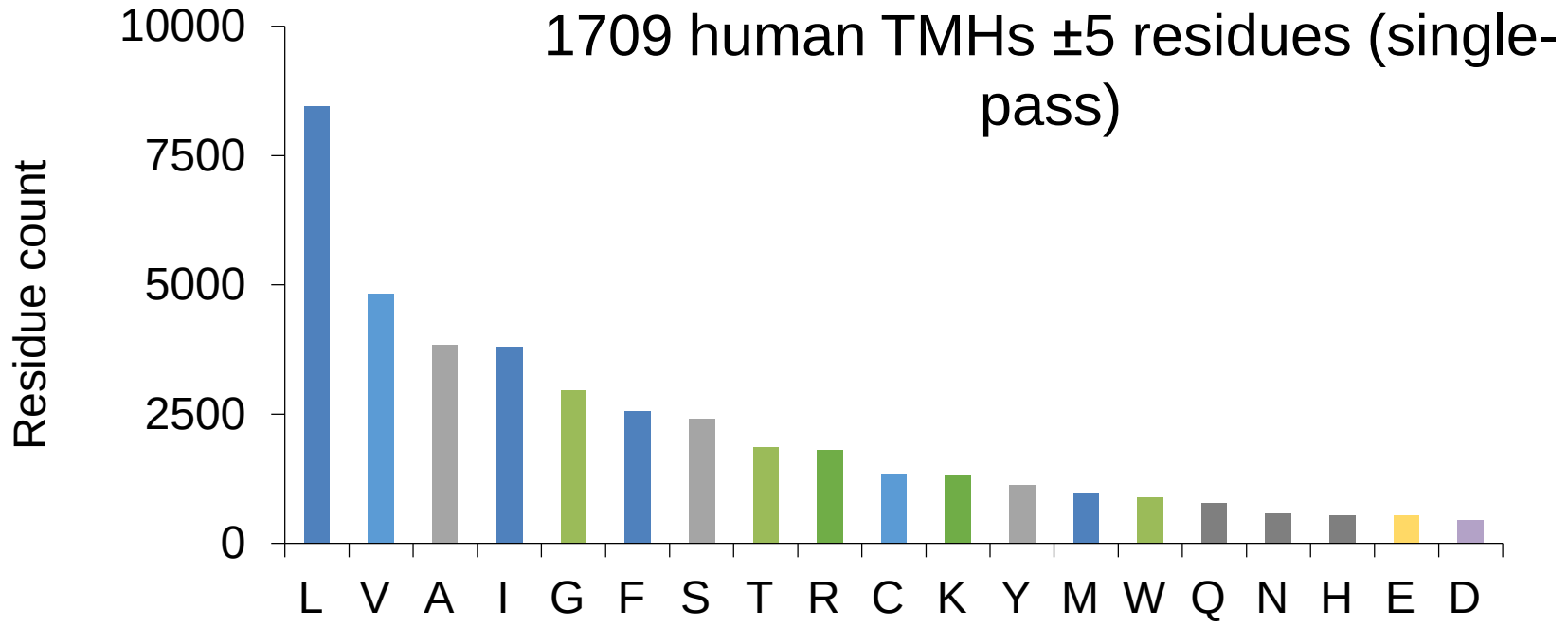
Introduction



Ulmschneider,M.B. and Sansom,M.S.P. (2001) Amino acid distributions in integral membrane protein structures. Biochim. Biophys. Acta - Biomembr., 1512, 1–14.

“Problems” in previous study

- Negative residues are especially rare, even in the flanks



New methods for this study

- Segregate single-pass and multi-pass + other segregation
- Cross reference experimental and predictive datasets
- Align from the center (removes bias)
- New normalisation – independent, percentage based
  - OLD: If we have a residue, where *and what* is it likely to be?
  - NEW: If we have a residue X, where is it likely to be?

$$q_{i,r} = \frac{100 \times a_{i,r}}{a_i}$$

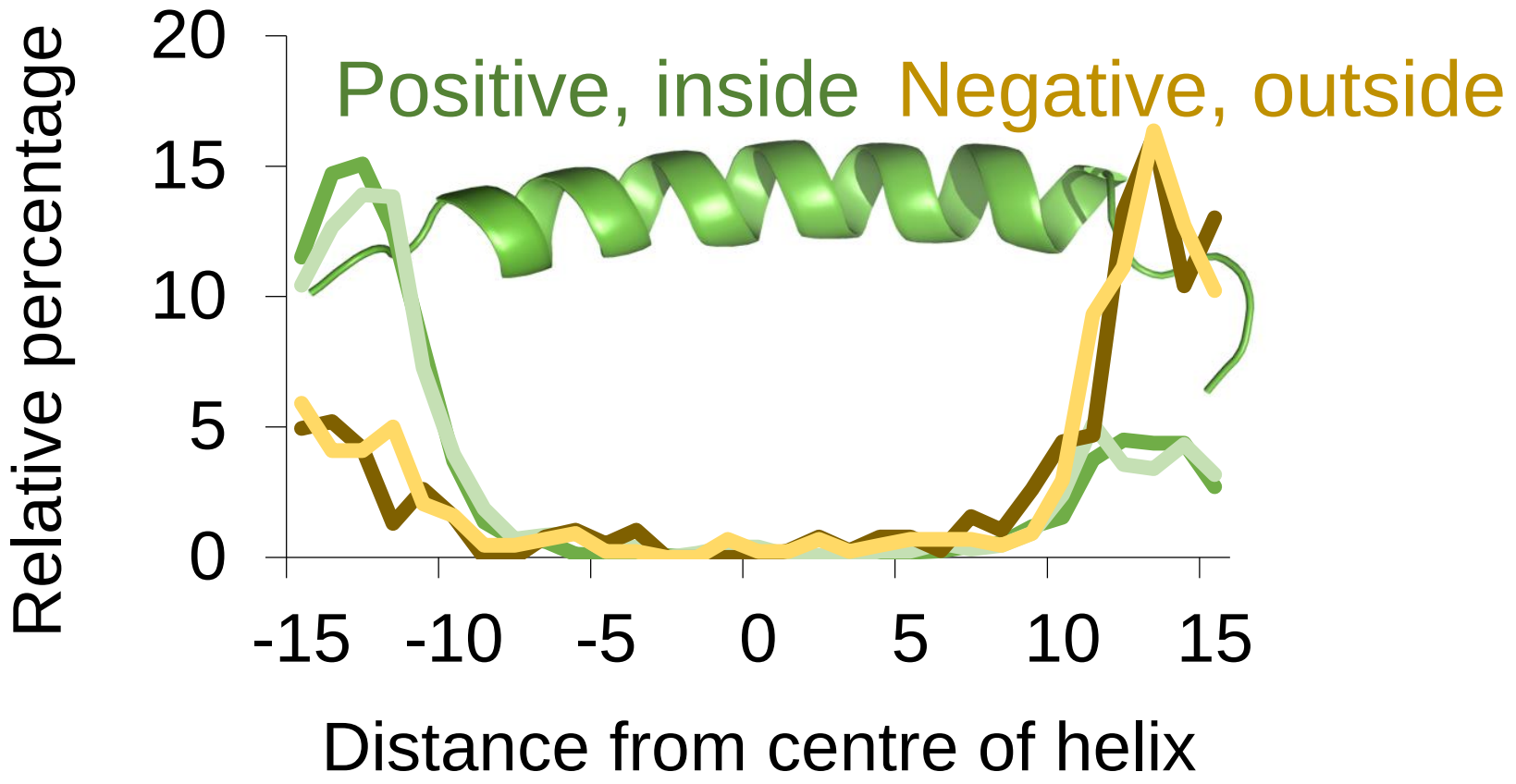
$$p_{i,r} = \frac{a_{i,r}}{\max_r \left(a_r \right)}$$

abundance = a
amino acid type = i
certain sequence position = r

Results

If we have a residue X, where is it likely to be?

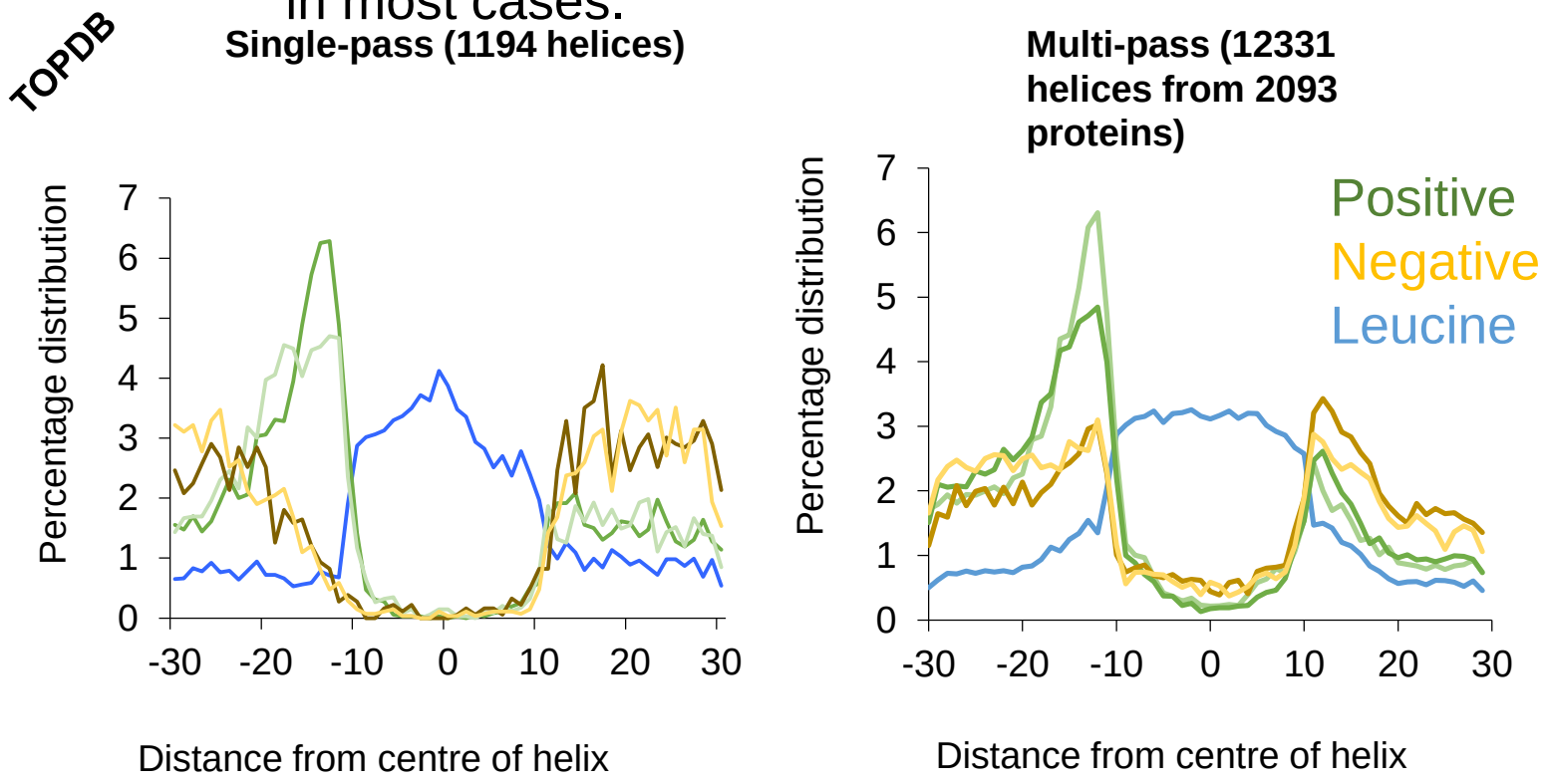
$$q_{i,r} = \frac{100 \times a_{i,r}}{a_i}$$



Results

At which membranes negative charges follow the negative-not-inside/negative-outside rule?

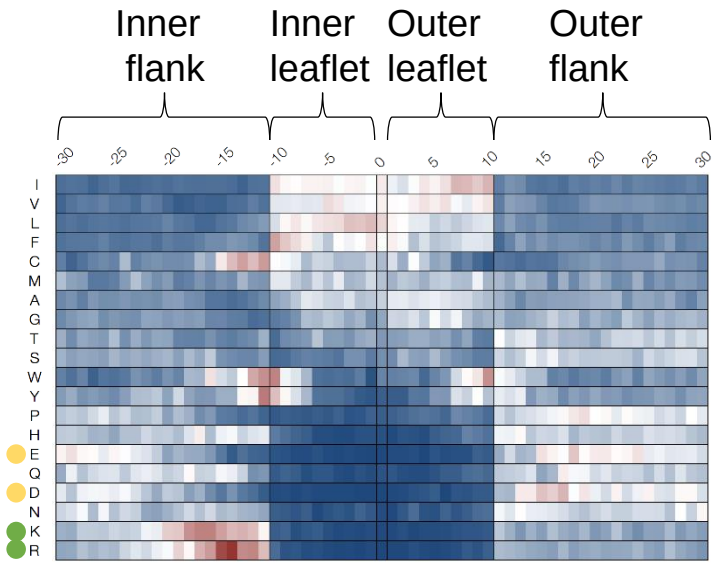
- Single-pass graphically.
- Multi-pass not graphically present, but statistically present in most cases.

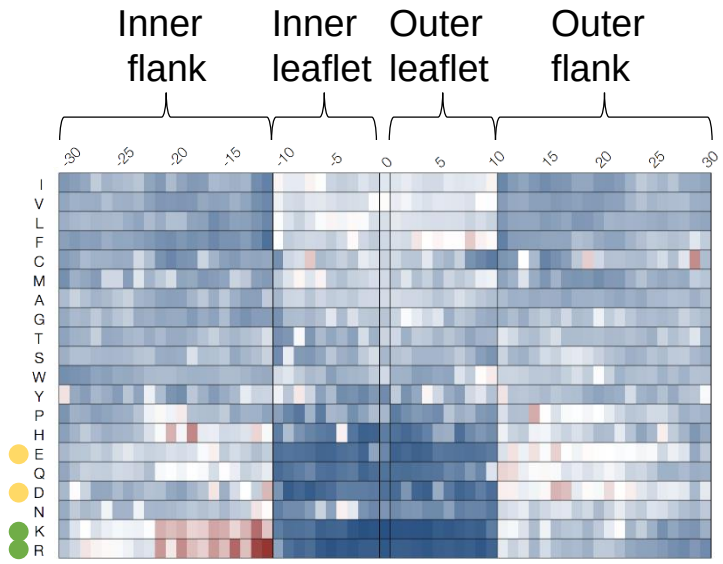


Results

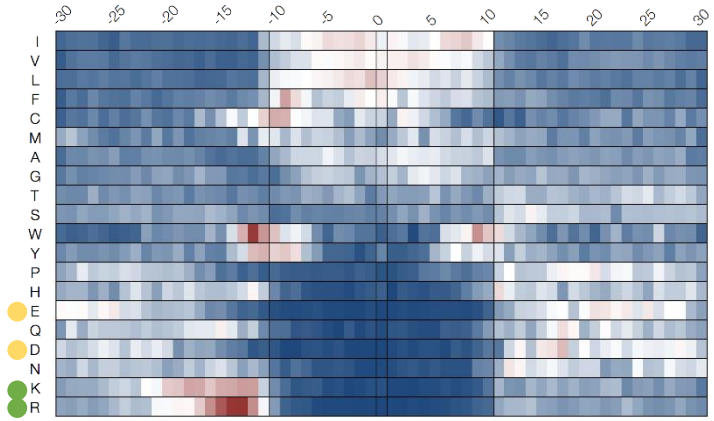
Single-pass | **Multi-pass**

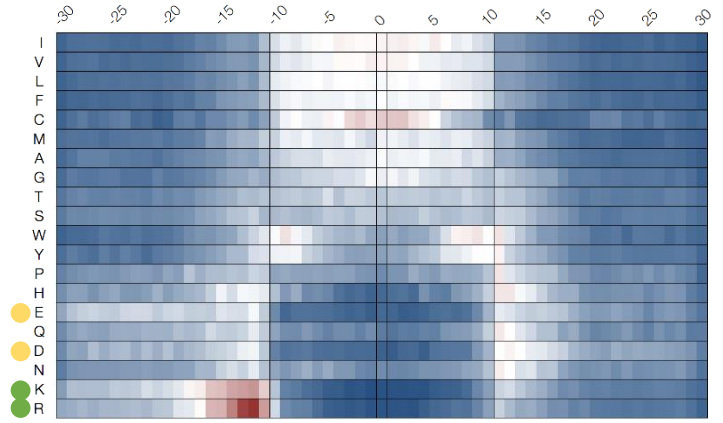
Human



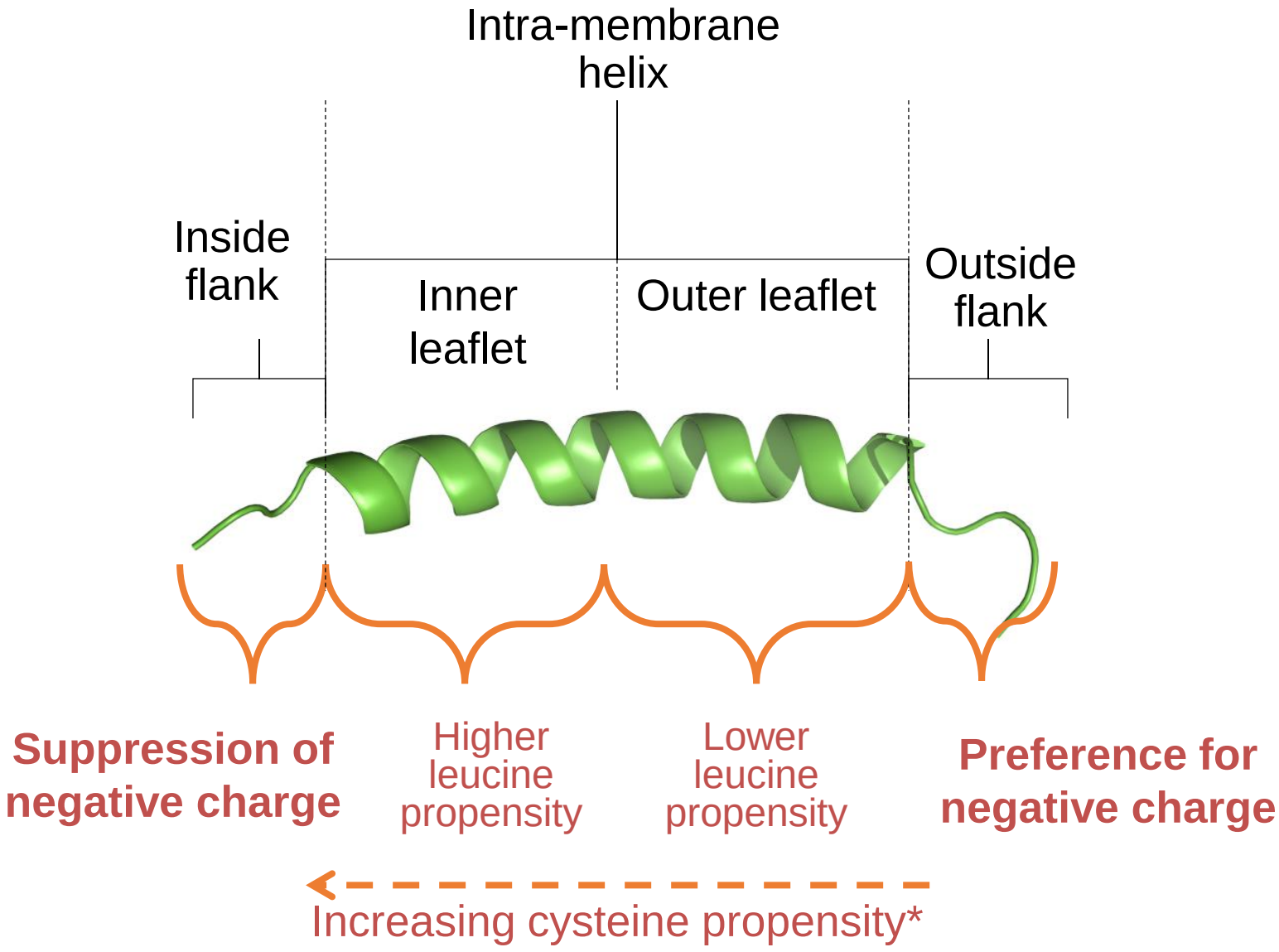


TOPDB





Our Findings

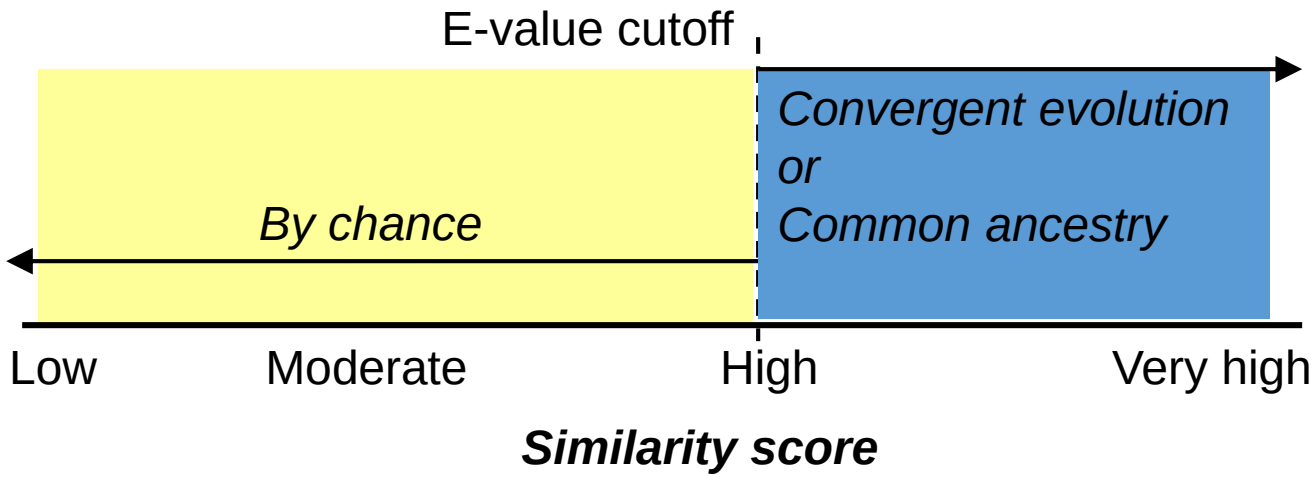


Conclusions

- A “negative-not-inside/negative-outside rule” complements the “positive-inside rule”.
- Leucine intra-helix propensity reflects leaflet asymmetry.
- Multi-pass helices are very different (on average) to single-pass helices.

Background considerations

Similarity measure as a proxy to homology and its limitation

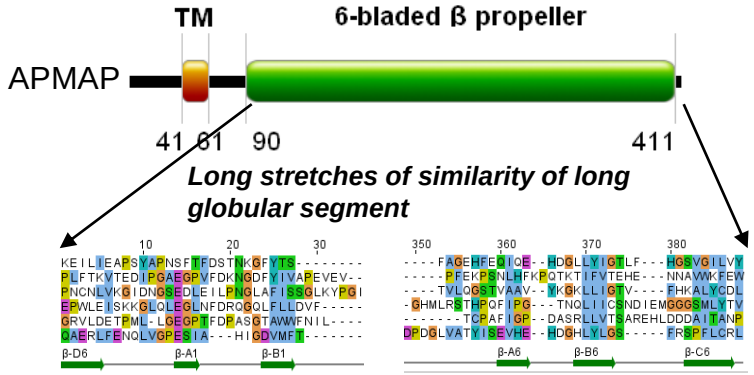


- ***Homology is a hypothesis about common evolutionary origin***
- ***Similarity is a measurable fact***
- ***Long stretches of similarity versus local resemblances (physiologically constrained to form rudimentary structure)***

Background considerations

Issues with non-globular sequences

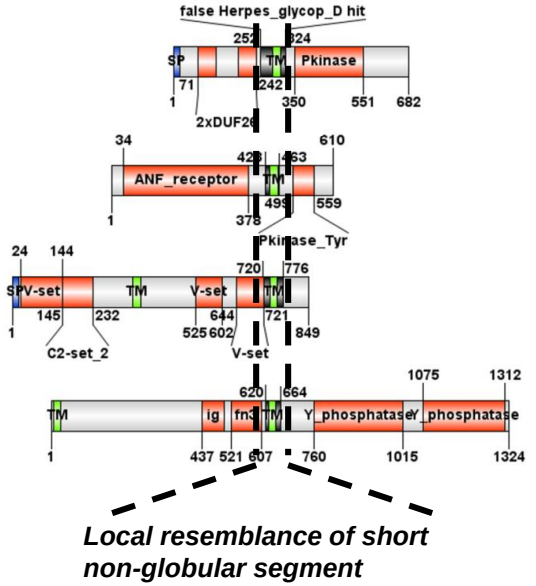
Common ancestry



Alignment of homologous structures

- *Strictosidine synthase*
- *Dissopropyflurophosphatase*
- *Serum paraoxonase*
- *Drp35*
- *Regucalcin*

Convergent evolution

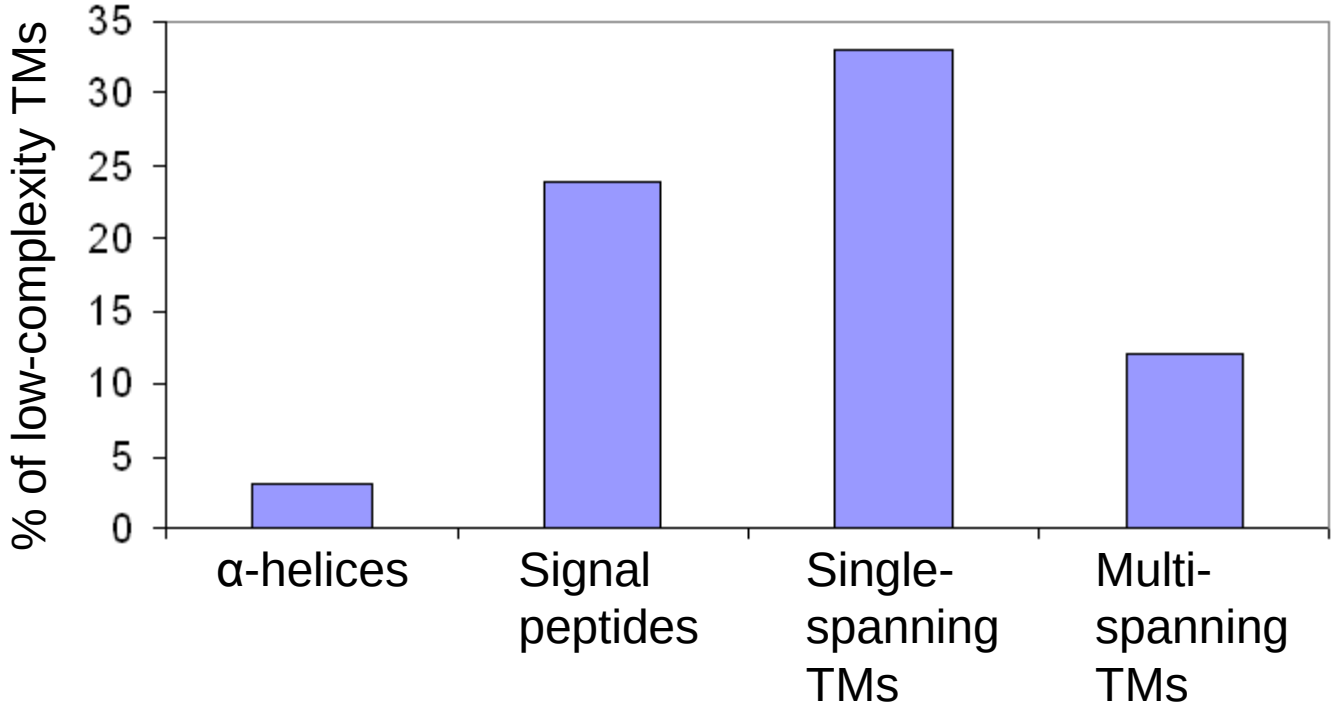


Unrelated hits with a similar TM segment

- ***Sequence homology concept is not directly applicable to non-globular sequences.***
- ***Signal-peptides/transmembrane helices (SP/TM) belong to this class***
- ***Mimics the appearance of hydrophobic core match***

Sequence complexity of SP/TM

Results of SEG (12/2.2/2.5) :

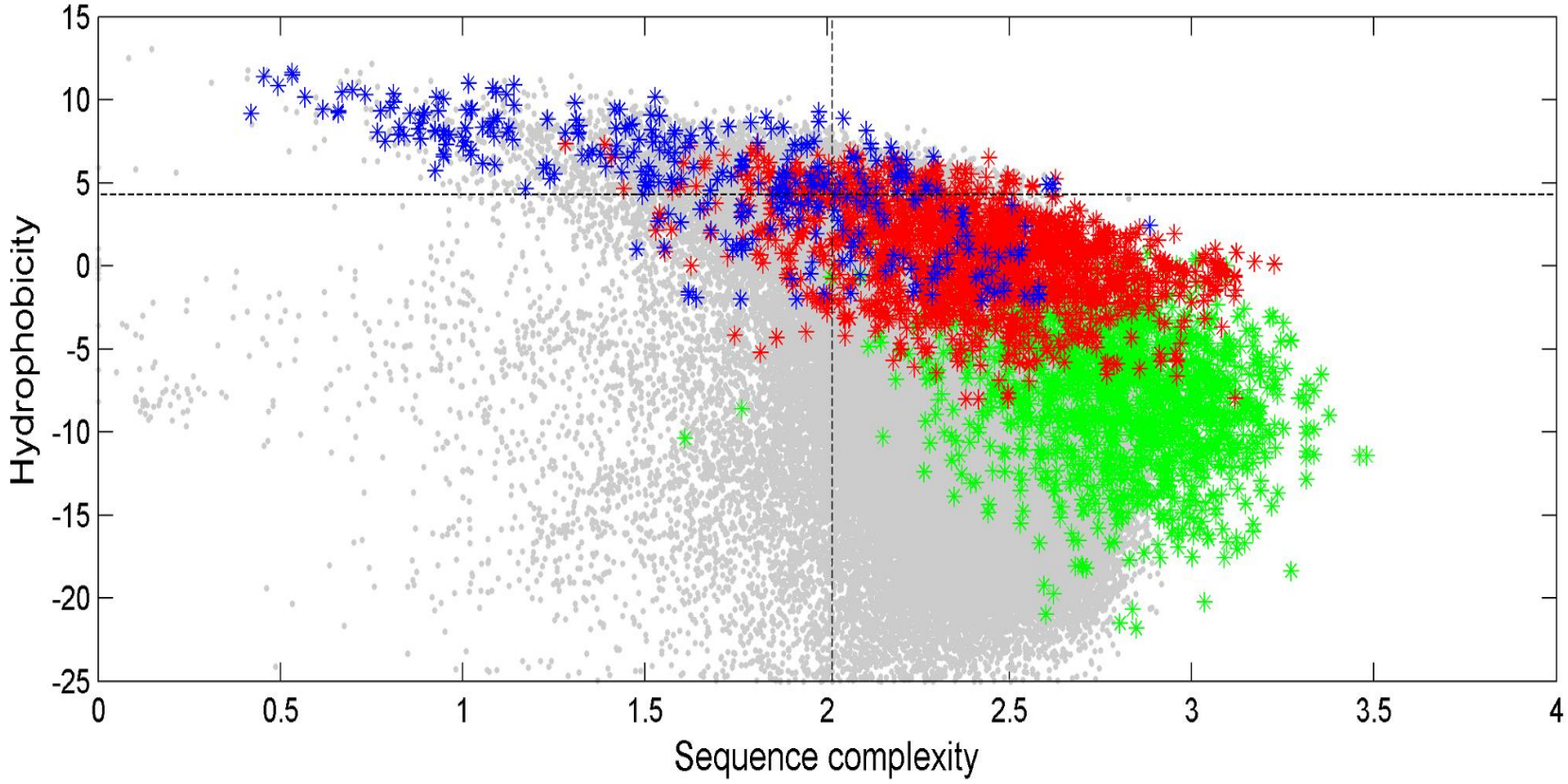


SP/TM have lower complexity than α-helices (12~33% versus 3%)

Open-ended questions :

- *Should all TMs be excluded? What about multi-membrane proteins like GPCR?*
- *Should all single-spanning TM be excluded?*
- *What about those with 'a few' TMs?*

Relationship among the TM helices, functional α-helices and low-complexity segments

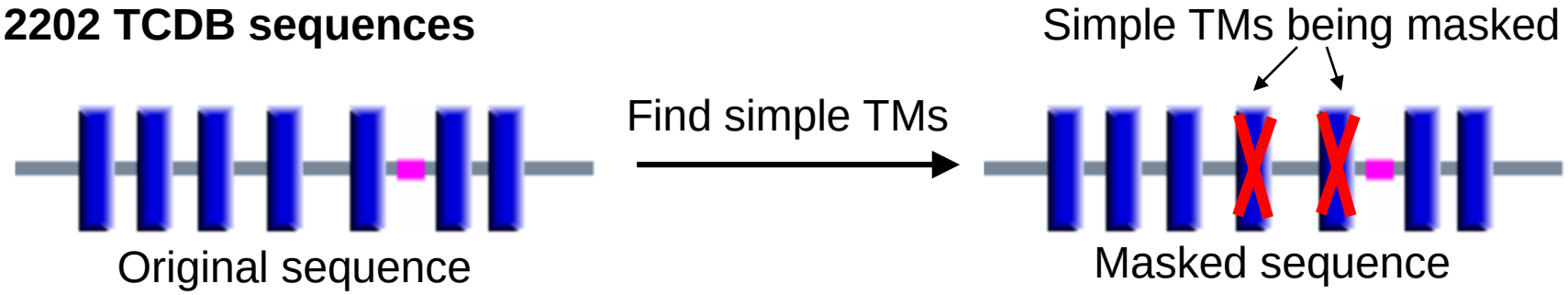


Membrane anchors, functional TMs, α-helices, low-complexity segments

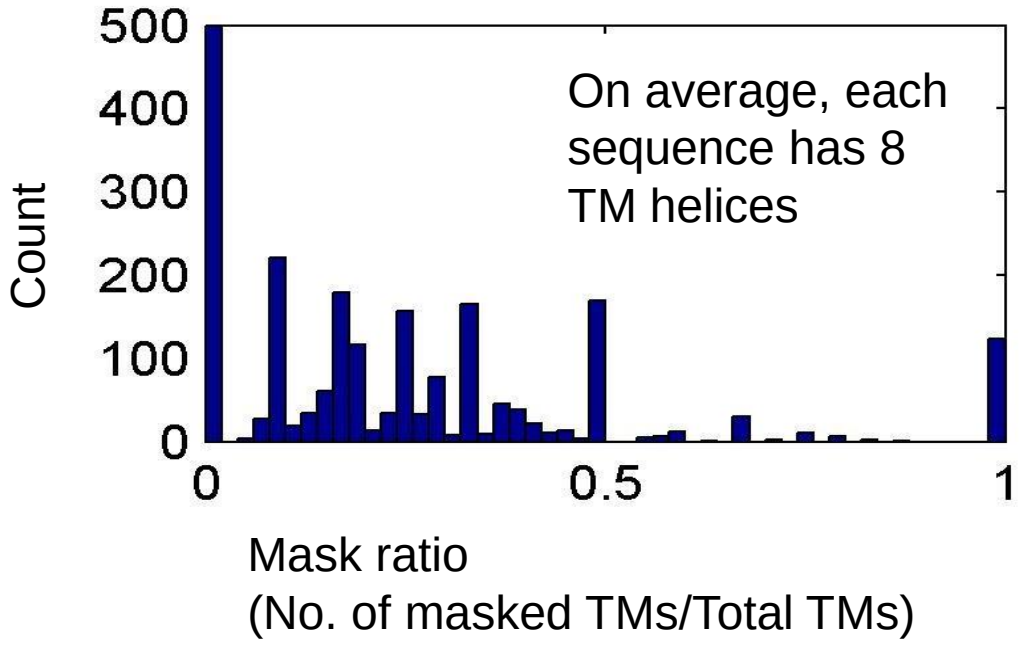
- ***Overlap of functional α- and TM- helices extents the sequence homology concept for membrane proteins***
- ***SEG samples low hydrophobicity space and hence insufficient to distinguish 'simple' or 'complex' TMs***

TM properties in multi-spanning membrane proteins

For 2202 TCDB sequences



Mask ratio for 2202 sequences



Multi-spanning membrane proteins can harbor simple TM helices

Conclusions

- ***TMs are either simple (likely of convergent evolution) or complex (likely of common ancestry).***

- ***Signal peptides and simple TMs can attract unrelated hits. Simple TMs should be quantitatively excluded from similarity searches using the z-score criteria.***

- ***Complex TMs embody ancestry information and justified for the application of sequence homology concept.***

- ***Simple TMs are found in membrane proteins regardless of membrane topology. The caveat is that it occurs more frequently in low-spanning ones.***

BII Yearbook 2017

- Thanks to Betty and all contributors
- Timeline of BII's history

Bioinformatics Institute: Status in 2017

Thank you !!

Bioinformatics Institute
A*STAR

Celebrating 15 Years!