Journal of Biomolecular Structure and Dynamics, 2018
Vol. 36, No. 8, 2194–2200, https://doi.org/10.1080/07391102.2017.1345699

Insights into eukaryotic evolution from transmembrane domain lengths

Aditya Mittal* and Snigdha Singh

Kusuma School of Biological Sciences, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India

Communicated by Ramaswamy H. Sarma

(*Received 30 March 2017; accepted 15 June 2017*)

Biological membranes, comprised of proteins anchored by their trans-membrane domains (TMDs) creating a semi-permeable phase with lipid constituents, serve as 'checkposts' for not only intracellular trafficking in eukaryotic cells but also for material transactions of all living cells with external environments. Hydropathy (or hydrophobicity) plots of 'bitopic' proteins (i.e. having single alpha-helical TMDs) are routinely utilized in biochemistry texts for predicting their TMDs. The number of amino acids (i.e. TMD length) embedded as alpha-helices may serve as indicators of thickness of biological membranes in which they reside under assumptions that are universally applied for fixing window sizes for identifying TMDs using hydropathy plots. In this work we explore variations in thickness of different eukaryotic biological membranes (reflected by TMD lengths of their resident proteins) over evolutionary time scales. Rigorous *in silico* analyses of over 23,000 non-redundant membrane proteins residing in different subcellular locations from over 200 genomes of fungi, plants, non-mammalian vertebrates and mammals, reveal that differences in plasma membrane and organellar TMD lengths have decreased over time (scales) of eukaryotic cellular evolution. While earlier work has indicated decreasing differences in TMD lengths with increasing 'perceived' organismal complexity, this work is the first report on TMD length variations as a function of evolutionary time of eukaryotic cellular systems. We report that differences in TMD lengths of bitopic proteins residing in plasma membranes and other intra-cellular locations have decreased with evolutionary time, suggesting better/more avenues of intracellular trafficking in the emergence of eukaryotic organisms.

Keywords: membrane proteins; hydropathy plots; hydrophobicity; evolution; intracellular trafficking

Introduction

Compartmentalization by biological membranes resulting in formation of membrane-bound organelles is one of the key signatures of eukaryotic cells. Starting from the plasma membrane of a eukaryotic cell, different membrane-bound organelles provide localized and segregated environments/sites for specific cellular reactions. Biological membranes serve as specific boundaries with their constituents (i.e. lipids and proteins) forming localized 'checkposts' for transfer of material from one environment to another within a cell, and, for material transactions with the extra-cellular milieu. Over the last decade, chemical and structural analyses of lipid constituents of plasma membranes have provided valuable insights into not only formation of specific 'domains' in biological membranes (Bansal & Mittal, 2013; Baumgart, Hess, & Webb, 2003), but also on the origins of whole (prokaryotic and eukaryotic) cells (Bansal & Mittal, 2015). Shapes of individual lipids in the hydrophobic-effect driven assembly of membranes provide different localized geometrical characteristics (e.g. thickness, curvature) to biological membranes (Chernomordik & Kozlov, 2003, 2008; Kozlov et al., 2014) resulting in their functional specificity toward serving as a particular boundary (i.e. of different organelles or as plasma membranes). These findings on lipid constituents have obvious implications for (membrane) proteins residing in specific biological membranes, especially since their residence is determined by an anchor in the membrane called the transmembrane domain (TMD). Recognizing this, a series of recent studies (Nikolovski et al., 2012; Sharpe, Stevens, & Munro, 2010; Singh & Mittal, 2016) have reported that TMD lengths of membrane proteins residing in different subcellular locations of eukaryotic cells are different, and, more importantly statistically distinguishable. A fascinating application of this discovery has also been reported (Singh & Mittal, 2016) by identification of specific organelles of host cells involved in life cycle of some viruses based on analyses of TMD lengths extracted from viral proteomes. Clearly, the discovery of TMD lengths serving as signatures of their subcellular locations in eukaryotic cells is of substantial significance in cell biology, especially in the current era of proteomics – e.g. relatively straightforward computational analyses of eukaryotic cellular proteomes can assist in identification of membrane proteins localized to specific checkposts of intra- and subcellular trafficking based on identification of putative TMD segments and their

*Corresponding author. Email: amittal@bioschool.iitd.ac.in

respective lengths. In this work, rigorous analyses of over 23,000 membrane protein sequences from 201 eukaryotic genomes (122, 29, 17, and 33 belonging to fungi, plants, non-mammalian vertebrates and mammals respectively) reveal that variations in TMD lengths of proteins residing in plasma membranes and other organelles of different groups of eukaryotic organisms have decreased over evolutionary time (scales) estimated for the respective organisms.

Methods

Data on TMD lengths for bitopic membrane proteins was collected as described earlier (Nikolovski et al., 2012; Sharpe et al., 2010; Singh & Mittal, 2016). The exact algorithm developed and utilized earlier (Singh & Mittal, 2016) was implemented for obtaining the final TMD lengths for further analyses in this work. Briefly: The first step was to collect primary sequences of bitopic membrane proteins (=68,281, including orthologs) from fungi, plants, non-mammalian vertebrates (referred to as vertebrates only everywhere else in the manuscript) and mammals. The second step was to orient the primary sequences of proteins from reference species (*Saccharomyces cerevisiae* for fungi, *Arabidopsis thaliana* for plants, *Gallus gallus* for vertebrates and *Homo sapiens* for mammals) from cytosolic to extracellular sides using the TMHMM server. The third step was to 'roughly' identify TMD regions along with their flanking sequences (on both sides) using hydropathy plots. The fourth step was to cluster and group sequences with more than 30% homology using Blastclust – this was done to remove redundancy and thus possible biases in further statistical analyses. The fifth and final step was to precisely identify the edges of TMDs and obtain the TMD lengths. Implementation of the above algorithm resulted in a total of 23,105 TMD lengths – fungi (8782), plants (8330), non-mammalian vertebrates (2651), and mammals (3342): this data-set is provided in Table S1 (TMD lengths of proteins with their respective accession numbers).

Here, it is important to emphasize the importance of the third, fourth, and fifth steps. While the third step estimates the individual TMD region in primary sequences (i.e. 'rough' TMD regions – TMD lengths are not important at this point) with some uncertainty due to possible differences between data sources, the fourth step specifically ensures that any possible statistical biases in the 'rough' TMD region estimates are avoided by inclusion of homologous sequences. The fifth step is a key methodological advancement toward precise identification of TMD regions (and hence TMD lengths are calculated at this point only). The accuracy of this methodology has already been demonstrated, with explicit (evaluation and) minimization of uncertainties in precise determination of TMD regions by direct comparisons of the computed results with experimental data known for different eukaryotic cells including fungi, plants (Nikolovski et al., 2012; Sharpe et al., 2010; Singh & Mittal, 2016), and even viral membrane proteins (Singh & Mittal, 2016). Therefore, earlier rigorous demonstration of the importance of the fifth step allows for deeper statistical analyses for reaching meaningful inferences.

At this stage it is important to note that, for this work Table S1 is considered as 'raw' data on TMD lengths. This 'raw' data is arranged according to subcellular location (plasma membrane, TGN/endo/Nucleus, Golgi, and ER) in the corresponding organism species (it needs to be noted that as per availability of sequences, plants have primarily Nucleus sequences in the TGN/endo/Nucleus category and other groups have primarily TGN/endo in this category – other three categories are consistently spread in all groups). Next, one of keys to analyzing the raw data in Table S1 was to select only those species for which TMD lengths were available in all the four subcellular location categories listed above (i.e. plasma membrane, TGN/endo/Nucleus, Golgi, and ER). Thus, the species, for which protein sequences were missing for any of the four subcellular location categories, were removed from the TMD length data to be analyzed. This resulted in a marginal reduction of total number of sequences to 23,018 (from 23,105 initial) with fungi (8713), plants (8312), vertebrates (2651), and mammals (3342) having 122, 29, 17, and 33 species, respectively (Table S2). Now species-wise mean TMD lengths were calculated for each of the four subcellular location categories. As an internal computational control, it was confirmed that distribution of mean TMD lengths corresponding to subcellular location categories for all species were quite tight (standard deviations were much smaller compared to respective means, see Table S2) – this was in complete agreement with conclusions of the earlier reports (Nikolovski et al., 2012; Sharpe et al., 2010; Singh & Mittal, 2016). The next step was to calculate the 'root mean squared deviation', i.e. RMSD, among the TMD lengths. Since it had been already shown that plasma membrane TMD lengths are similar for fungi, plants, vertebrates, and mammals (Singh & Mittal, 2016), RMSD was calculated for each species as $\mathrm{sqrt}([\sum(\mathrm{PM\text{-}Org})^2]/3)$, where 'PM' denotes the mean TMD length of plasma membrane of a particular species, and, 'Org' denotes the mean TMD lengths of ER or Golgi or TGN/endo or Nucleus for the same species (since there are '3' categories of 'Org,' therefore the denominator in the formula is 3). Conceptually RMSD represents the difference of mean organellar TMD lengths from the mean TMD length of proteins found in plasma membranes of a given species. Finally, distributions of RMSD values for all the species (total 201,

fungi – 122, plants – 29, vertebrates – 17, mammals – 33) were obtained and fit to normal distributions for obtaining μ and σ values corresponding to each of the eukaryotic classes (i.e. fungi, plants, vertebrates, and mammals).

Results

A computational control: variability between organellar and plasma membrane TMD lengths decreases with increasing organismal complexity

First we collected TMD lengths, expressed as number of trans-membrane residues, of proteins associated with plasma membranes of eukaryotic cells belonging to the groups of fungi, plants, non-mammalian vertebrates (referred to as vertebrates only further in the text) and mammals (see Methods for details). We also collected TMD lengths of 'organellar' proteins associated with Endoplasmic Reticulum (ER), Golgi, and Trans-Golgi-Network/endosomal-network, Nucleus in the above eukaryotic cells. Figure 1(a) shows the TMD lengths of proteins residing in organellar and plasma membranes of fungi, plants, vertebrates, and mammals. The first observation was that TMD lengths of proteins residing in eukaryotic plasma membranes are similar ($p = 0.79$ for single factor ANOVA) – in agreement with previous observations on plasma membranes based only lipid compositions (Bansal & Mittal, 2015), possibly due to common origins of eukaryotic cellular membranes (Bansal & Mittal, 2013). The second observation was that TMD lengths of organellar proteins are statistically distinguishable from those of plasma membranes for eukaryotic cells belonging to fungi, plants, vertebrates, and mammals (see Figure 1(a) and Table 1) – again in agreement with earlier reports (Nikolovski et al., 2012; Sharpe et al., 2010; Singh & Mittal, 2016). The next step was to explore actual nature of the distinctions between organellar and plasma membrane TMD lengths. To do so, we had to first devise a parameter that reflected variation between organellar and plasma membrane TMD lengths. Thus, we first calculated the 'root mean square deviation' (RMSD) of organellar TMD lengths from those of plasma membranes for all those species of eukaryotic cells for which we had complete data-sets (i.e. TMD lengths of membrane proteins associated with all the organelles and plasma membranes). Conceptually RMSD represents the difference of mean organellar TMD lengths from the mean TMD length of proteins found in plasma membranes of a given species – the lesser the difference, the smaller would be the value of RMSD.

Figure 1(b) shows the RMSD distributions obtained for fungi, plants, vertebrates, and mammals. Here it is important to emphasize that ordering 'fungi-plants-vertebrates-mammals' is as per the evolutionary appearance of these groups of eukaryotes (Heckman et al., 2001; Hedges, Blair, Venturi, & Shoe, 2004; Hedges, Marin, Suleski, Paymer, & Kumar, 2015; James et al., 2006; Simon, Bousquet, Lévesque, & Lalonde, 1993) and the perceived cellular complexity. Clearly, as we move from fungi to mammals, the RMSD distributions shift left and also appear to become narrower. This indicates that TMD lengths, reflecting of thickness of their respective biological membranes in particular, become less variable within the cells with increasing cellular complexity. Figure 1(c) illustrates these findings very explicitly – both mean RMSD and standard deviation of RMSD distributions decrease as we move from fungi to mammals. Further, the distributions are statistically very significant ($p << 0.05$) as shown in Table 2, except for the similarity in plants and vertebrates. Interestingly enough, eukaryotic life forms of terrestrial plants and vertebrates are thought to have appeared quite close to each other in the time tree of life (Hedges et al., 2015). In fact, two very appealing hypotheses emerged from the above results – (a) from Table 2, it appears that pair-wise differences in RMSD distributions for TMD lengths for eukaryotic groups directly reflect the corresponding pair-wise distance in the Tree of life in terms of evolutionary time scales and cellular complexity, and, (b) from Figure 1(c), there appears to be a directionality to evolution of eukaryotic cells from fungi to mammals toward 'homogenization' of the cellular milieu in terms of TMD lengths and hence (at least the thickness of) subcellular membranes. Most importantly, both the above hypotheses were testable from our data.

Variations in TMD lengths are highly correlated with evolutionary time scales

We noted that p-value from a t-test between any two data-sets reflects the probability of both the data-sets belonging to the same population. Simply stated, smaller p-value indicates lower probability of two data-sets being similar, i.e. smallest p-value indicates the largest difference in two data-sets. Therefore, the next step toward testing the first hypothesis (stated above) was to plot the p-values (from Table 2) for each pair of organisms in ascending order. Figure 2(a) shows that the p-value between fungi and mammals is the smallest (indicating greatest difference). The next pair is fungi and vertebrates, i.e. the difference between fungi and vertebrates is smaller than that between fungi and mammals. Similarly, difference in plants and mammals is more than the difference between fungi and plants. Interestingly, vertebrates and mammals appear to have more difference than plants and vertebrates. Based on these observations between pairs of eukaryotic groups, we then plotted each of the p-values to the corresponding approximate time difference in evolution of each of the respective pairs

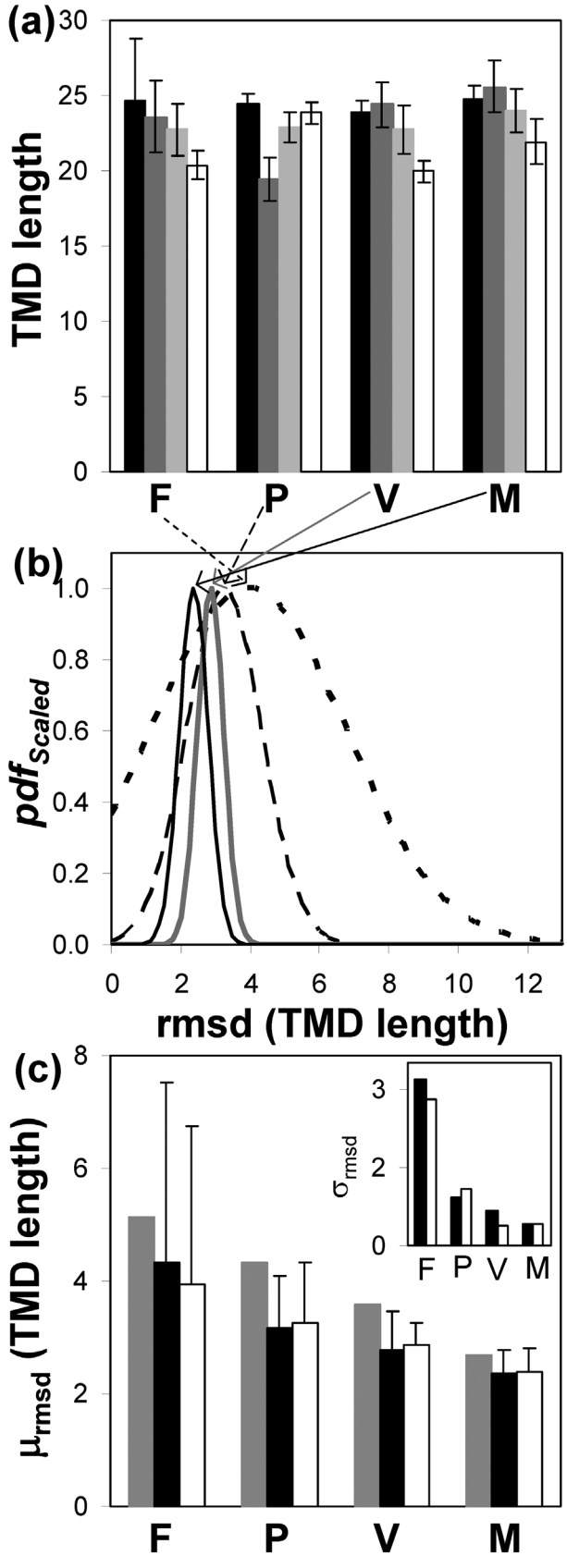

Figure 1. Transmembrane domain (TMD) lengths show decreasing variability with increasing cellular complexity. (a) TMD lengths, in terms of number of residues, for proteins residing in plasma membranes (black) and organellar membranes – TGN/endo/Nucleus (dark gray), ER (light gray) and Golgi (Empty), for eukaryotic cells belonging to fungi (F, $n = 122$), plants (P, $n = 29$), vertebrates (non-mammalian, V, $n = 17$) and mammals (M, $n = 33$). Data is shown as mean ± standard deviation. (b) Scaled Probability Density Function (pdf – normal distributions, scaled from 0 to 1 for visual clarity) of root mean square deviation (RMSD) of mean organellar TMD lengths from the mean TMD length of proteins found in plasma membranes of different eukaryotic species belonging to the groups of fungi (dotted curve, mean shown by dotted arrow), plants (dashed curve, mean shown by dashed arrow), vertebrates (non-mammalian only, gray curve, mean shown by gray arrow) and mammals (solid curve, mean shown by solid arrow). The RMSD data for each of the eukaryotic classes was fit to normal distributions and the pdf was normalized from 0 to 1 (peak frequency = 1.0) for visual clarity. (c) Mean RMSD corresponding to actual RMSD data (solid black bars) and fitted distributions (empty bars) are shown for fungi (F, $n = 8713$), plants (P, $n = 8312$), vertebrates (V, $n = 2651$) and mammals (M, $n = 3342$). The inset shows standard deviations (error bars in the main figure) of the actual RMSD data and the fitted distributions. As a computational control, earlier published data (from Figure 3(E) Singh & Mittal, 2016) on mode of RMSD between organellar and plasma membrane TMD lengths of the four eukaryotic groups as whole (i.e. without considering individual species data) is also shown (gray bars).

(inset, Figure 2(a)). To our striking (and pleasant) surprise, *p*-values for pairs are highly correlated with the difference in time of evolution of the respective pairs (see inset, Figure 2(a)). This was a very interesting result – differences in TMD lengths (which almost directly represent thickness of biological membranes in which the corresponding membrane proteins reside) were providing a direct measure of the evolutionary time differences in different eukaryotic groups.

Encouraged by the above result, we then proceeded to test the second hypothesis. To do so, we plotted the mean RMSD values (marked in Figure 1(b) by arrows and shown in Figure 1(c)) as a function of the absolute time of evolution for the corresponding eukaryotic group (Figure 2(b)). Once again, regardless of whether we plotted the mode RMSD values of the eukaryotic groups as whole (gray symbols, shown as gray bars in Figure 1(c)) or we plotted the mean RMSD values (black symbols, obtained by considering individual species within each of the eukaryotic groups), the RMSD values show a remarkable correlation with the absolute evolutionary time scales for fungi, plants, vertebrates, and mammals. In fact, from Figure 2(b) the directionality of evolution toward cellular complexity, indicated by our earlier results, is demonstrated quite convincingly – even allowing the possibility of a prediction. The equations for

Table 1. *p*-values obtained from two-tailed heteroscedastic *t*-tests between TMD lengths of proteins residing in organellar and plasma membranes for fungi (F), plants (P), vertebrates (V), and mammals (M) – data is shown in Figure 1(a).

F (n = 122)	PM	ER	Golgi	TGN/endo/Nucleus
PM	–	2.90×10^{-06}	2.40×10^{-21}	9.50×10^{-03}
ER	2.90×10^{-06}	–	7.94×10^{-30}	2.10×10^{-03}
Golgi	2.40×10^{-21}	7.94×10^{-30}	–	5.80×10^{-29}
TGN/endo/Nucleus	9.50×10^{-03}	2.10×10^{-03}	5.80×10^{-29}	–
P (n = 29)				
PM	–	9.44×10^{-09}	9.35×10^{-04}	6.63×10^{-20}
ER	9.44×10^{-09}	–	1.25×10^{-04}	1.36×10^{-14}
Golgi	9.35×10^{-04}	1.25×10^{-04}	–	4.04×10^{-18}
TGN/endo/Nucleus	6.63×10^{-20}	1.36×10^{-14}	4.04×10^{-18}	–
V (n = 17)				
PM	–	1.26×10^{-02}	5.42×10^{-16}	2.38×10^{-01}
ER	1.26×10^{-02}	–	2.38×10^{-06}	3.57×10^{-03}
Golgi	5.42×10^{-16}	2.38×10^{-06}	–	8.61×10^{-11}
TGN/endo/Nucleus	2.38×10^{-01}	3.57×10^{-03}	8.61×10^{-11}	–
M (n = 33)				
PM	–	6.90×10^{-03}	4.23×10^{-13}	2.36×10^{-02}
ER	6.90×10^{-03}	–	2.94×10^{-07}	1.28×10^{-04}
Golgi	4.23×10^{-13}	2.94×10^{-07}	–	2.85×10^{-13}
TGN/endo/Nucleus	2.36×10^{-02}	1.28×10^{-04}	2.85×10^{-13}	–

Note: Number of species for each group is given in parentheses.

Table 2. *p*-values obtained from two-tailed heteroscedastic *t*-tests between RMSD data for fungi, plants, vertebrates and mammals (shown in Figure 1(c); raw data is provided in the last column of Supplementary Table S2).

p – values	Fungi (122)	Plants (29)	Vertebrates (17)	Mammals (33)
Fungi (122)	–	6.47×10^{-4}	7.59×10^{-6}	7.71×10^{-10}
Plants (29)	6.47×10^{-4}	–	1.07×10^{-1}	1.02×10^{-4}
Vertebrates (17)	7.59×10^{-6}	1.07×10^{-1}	–	3.36×10^{-2}
Mammals (33)	7.71×10^{-10}	1.02×10^{-4}	3.36×10^{-2}	–

Note: Number of species for each group is given in parentheses.

regression lines (shown in legend to Figure 2(b)) predict that mean RMSD would become 'zero' at an absolute evolutionary time of 500 million years from now. This means that if eukaryotic evolution remains on course (i.e. maintains its directionality), then eukaryotic cells with no differences in (at least the thickness of) biological membranes (plasma membranes and compartments within) will emerge within next 500 million years. However, in presence of substantial evidence that evolutionary paths are not linear (rather random), this prediction, while apparently 'entertaining,' is highly unlikely because of the persistence of differences in TMD lengths and other conserved features in TMD compositions over the span of evolution.

Discussion

The first major finding emerging from this work is that TMD lengths in eukaryotic cellular systems are correlated with evolutionary time scales – an interesting feature at a cellular level emerging randomly from natural selection. Rigorous analyses of TMD lengths show that eukaryotic cells have evolved toward subcellular membrane environments becoming more uniform (at least in terms of bilayer thickness), thereby allowing higher number of intracellular trafficking operations and resulting in more cellular complexity. However, this apparent 'uniformity' is obviously constrained by the persistence of differences in TMDs of membrane proteins in different eukaryotic cells.

Here it becomes pertinent to discuss the assumptions toward TMD lengths representing thickness of biological membranes in which the TMDs reside. The first assumption is somewhat uniform orientation of TMDs (i.e. all TMD helices are positioned either normal to the membranes or are inclined in somewhat similar angles inside membranes) in all eukaryotic cells. It is very important to note that application of this simplified assumption is

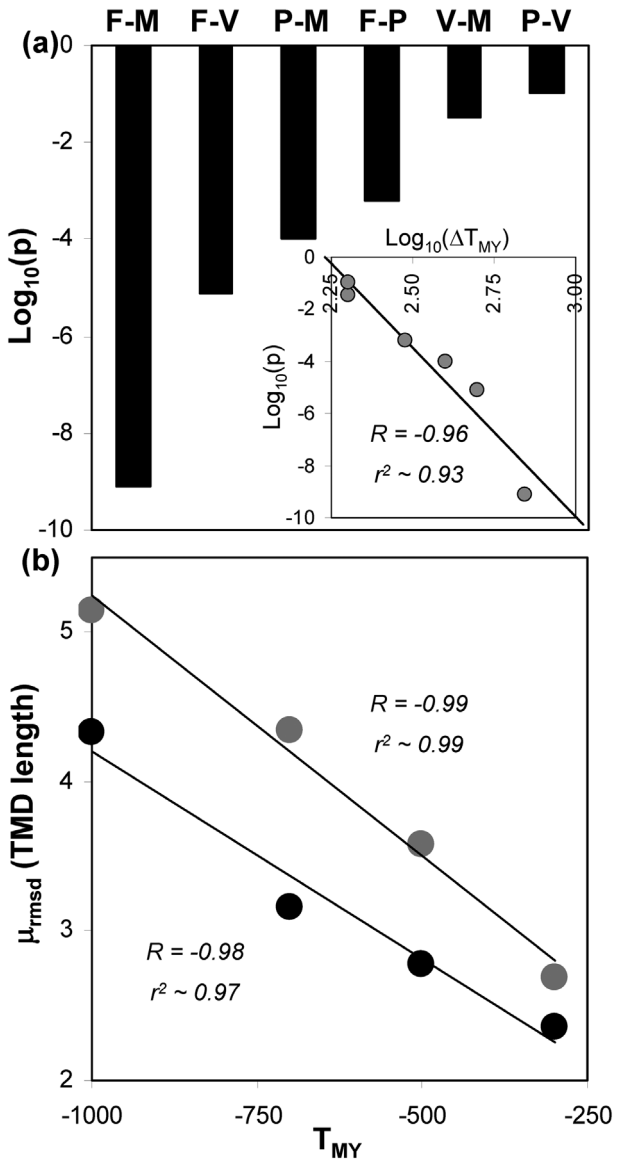

Figure 2. Variations in TMD lengths correlate very strongly with evolutionary time scales of increasing cellular complexity and may predict the next steps in eukaryotic evolution. (a) Log (to the base 10) of 'p' values (shown in Table 2) corresponding to two-tailed heteroscedastic *t*-tests between the RMSD data for fungi (F), plants (P), vertebrates (V), and mammals (M). *p*-values, indicating differences between pairs (e.g. fungi and mammals shown by F-M) among the four eukaryotic groups, were arranged in ascending order and plotted (solid bars). Pairs with smallest *p*-values (indicating highest differences) have the largest differences in evolutionary time scales (Time in Million Years – T_{MY}). The inset confirms this trend – *p*-values between each of the two groups of organisms correlate very strongly (high R and r^2 values) with the corresponding difference in evolutionary time scales (ΔT_{MY}). The regression line for the inset has the equation $Log(p) = -12.964Log(\Delta T_{MY}) + 28.919$. (b) Mean RMSD values corresponding to actual RMSD data (black, corresponding to black bars in Figure 1(c)) and whole RMSD values (gray, corresponding to gray bars in Figure 1(c)) correlate very strongly with absolute evolutionary times (T_{MY} has negative values indicating million years ago) of each eukaryotic group. The regression line for the former has the equation $RMSD = -0.0028T_{MY} + 1.4118$ and for the latter has the equation $RMSD = -0.0035T_{MY} + 1.7581$.

universal in current text books, research literature and all available computational tools in form of fixing window sizes for identifying TMDs using hydropathy plots. The second assumption is that overall 'height' of TMDs of bitopic proteins which may oligomerize (and also may undergo conformational transitions in their TMD regions) are also somewhat uniform in all eukaryotic cells. The third assumption is that in spite of lateral heterogeneities known in biological membranes (especially plasma membranes) the 'overall' or 'average' thickness of membranes of different sub-cellular environments is different. This assumption arises from several important experimental studies (several are cited clearly in Sharpe et al., 2010) that establish differences in membrane thickness of intracellular organelles and plasma membranes. In fact, it has been experimentally noted that outer monolayers of plasma membranes lead to lateral heterogeneities (this also results in membrane asymmetry) due to compositional differences. These lateral heterogeneities are observed at different scales (e.g. nano-domains or sometimes even micro-domains), while the 'overall' thickness of the plasma membranes appear uniform. However, regardless of the extent of applicability of these assumptions, the key finding of TMD lengths in eukaryotic cellular systems being correlated with evolutionary time scales certainly appears to be an important discovery and independent of interpretational biases.

A leap from understanding evolution of biological cells to emergence of human civilizations: The 'ECONOMIBS' hypothesis

An interesting (perhaps anecdotal) finding that emerges from this work is that reducing differences in TMD lengths, indicating reducing differences in thickness of subcellular membranes of eukaryotes, in terms decreasing differences in TMD lengths being correlated to evolutionary time scales, appears to be directly reflective of development and advances in human civilization (Harari, 2011). Variations in cellular 'checkposts'/boundaries in the tree of life over billions of years appear to directly reflect variations in social checkposts in establishment of human dwellings over thousands of years of civilization. Stated more explicitly, social evolution of human civilization (in some thousands of years), with decreasing differences in 'checkposts' and boundaries of societies directly reflects the biological evolution of eukaryotic cells (in a billion years) with decreasing differences in 'checkposts' and boundaries within the cellular milieu.

Thus, it may be possible to extrapolate a plethora of other variables/factors observed from biological evolution for developing sustainable societies that survive and evolve. For example, the energetic constraints limiting the size of prokaryotic cells (energetic transactions take place on the cellular surface – hence any increase in size, requiring a volumetric/cubic increase, is limited by surface area increase) are mirrored by resource constraints limiting size of initial dwellings in human civilization. Subsequent developing compartmentalization in eurkaryotic cells is mirrored by subsequent formation of specialized units in evolving societies. Similar other parallels between cellular evolution and human civilization can also be drawn. Of course, a detailed and meticulous analyses toward listing parameters that have emerged out biological evolution (e.g. optimized packaging of functional protein structures, high thermodynamic efficiencies of biological molecular motors due to close coupling of reactions, spatio-temporal coordination within and between cells to minimize losses with in-built redundancies to maximize chances of survival etc.), and may be applicable toward social evolution, would be required. For this, we propose a new field of study called 'ECONOMIBS' – Evolution of Civilization based On the Origin of Molecular Insights on Biological Self-assembly. We hope that just as '*economics'* has emerged as a key field in understanding development of materialistic transactions for societies in general, '*economibs'* will emerge as a key field in learning about the development and establishment of sustainable societies in the face of environmental challenges and changes – mimicking the survivability observed in eukaryotic cellular evolution over a billion years.

Contributions

A.M. designed the study, analyzed the data and wrote the manuscript. S.S. collected the raw data on TMD lengths.

Acknowledgements

AM is grateful to B. Jayaram (Department of Chemistry and Kusuma School of Biological Sciences, IIT Delhi) for critical reading of the manuscript and to C. S. Dey (Kusuma School of Biological Sciences, IIT Delhi) for introducing the text 'Sapiens: A brief history of Humankind' in context of this work. SS is grateful to Council of Scientific and Industrial Research (CSIR), Govt. of India for fellowship support. The authors are particularly grateful for receiving constructive criticisms during the review process that allowed substantial improvement in a focused presentation of results from this work.

Disclosure statement

No potential conflict of interest was reported by the authors.

Supplemental data

The supplementary material for this paper is available online at https://doi.org/10.1080/07391102.2017.1345699.

References

Bansal, S., & Mittal, A. (2013). Extracting curvature preferences of lipids assembled in flat bilayers shows possible kinetic windows for genesis of bilayer asymmetry and domain formation in biological membranes. *The Journal of Membrane Biology, 246*, 557–570.

Bansal, S., & Mittal, A. (2015). A statistical anomaly indicates symbiotic origins of eukaryotic membranes. *Molecular Biology of the Cell, 26*, 1238–1248.

Baumgart, T., Hess, S. T., & Webb, W. W. (2003). Imaging coexisting fluid domains in biomembrane models coupling curvature and line tension. *Nature, 425*, 821–824.

Chernomordik, L. V., & Kozlov, M. M. (2003). Protein-lipid interplay in fusion and fission of biological membranes. *Annual Review of Biochemistry, 72*, 175–207.

Chernomordik, L. V., & Kozlov, M. M. (2008). Mechanics of membrane fusion. *Nature Structural & Molecular Biology, 15*, 675–683.

Harari, Y. N. (2011). *Sapiens: A brief history of humankind*. London: Vintage (Penguin Random House UK).

Heckman, D. S., Geiser, D. M., Eidell, B. R., Stauffer, R. L., Kardos, N. L., & Hedges, S. B. (2001). Molecular evidence for the early colonization of land by fungi and plants. *Science, 293*, 1129–1133.

Hedges, S. B., Blair, J. E., Venturi, M. L., & Shoe, J. L. (2004). A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evolutionary Biology, 4*, 2. doi:10.1186/1471-2148-4-2

Hedges, S. B., Marin, J., Suleski, M., Paymer, M., & Kumar, S. (2015). Tree of life reveals clock-like speciation and diversification. *Molecular Biology and Evolution, 32*, 835–845.

James, T. Y., Kauff, F., Schoch, C. L., Matheny, P. B., Hofstetter, V., Cox, C. J., … Vilgalys, R. (2006). Reconstructing the early evolution of fungi using a six-gene phylogeny. *Nature, 443*, 818–822.

Kozlov, M. M., Campelo, F., Liska, N., Chernomordik, L. V., Marrink, S. J., & McMahon, H. T. (2014). Mechanisms shaping cell membranes. *Current Opinion in Cell Biology, 29*, 53–60.

Nikolovski, N., Rubtsov, D., Segura, M. P., Miles, G. P., Stevens, T. J., Dunkley, T. P., … Dupree, P. (2012). Putative glycosyltransferases and other plant Golgi apparatus proteins are revealed by LOPIT proteomics. *Plant Physiology, 160*, 1037–1051.

Sharpe, H. J., Stevens, T. J., & Munro, S. (2010). A comprehensive comparison of transmembrane domains reveals organelle-specific properties. *Cell, 142*, 158–169.

Simon, L., Bousquet, J., Lévesque, R. C., & Lalonde, M. (1993). Origin and diversification of endomycorrhizal fungi and coincidence with vascular land plants. *Nature, 363*, 67–69.

Singh, S., & Mittal, A. (2016). Transmembrane domain lengths serve as signatures of organismal complexity and viral transport mechanisms. *Scientific Reports, 6*, 2783. doi:10.1038/srep22352