frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 30 April 2021
doi: 10.3389/fmolb.2021.626729

Abundance Imparts Evolutionary Constraints of Similar Magnitude on the Buried, Surface, and Disordered Regions of Proteins

*Benjamin Dubreuil and Emmanuel D. Levy**

Department of Structural Biology, Weizmann Institute of Science, Rehovot, Israel

An understanding of the forces shaping protein conservation is key, both for the fundamental knowledge it represents and to allow for optimal use of evolutionary information in practical applications. Sequence conservation is typically examined at one of two levels. The first is a residue-level, where intra-protein differences are analyzed and the second is a protein-level, where inter-protein differences are studied. At a residue level, we know that solvent-accessibility is a prime determinant of conservation. By inverting this logic, we inferred that disordered regions are slightly more solvent-accessible on average than the most exposed surface residues in domains. By integrating abundance information with evolutionary data within and across proteins, we confirmed a previously reported strong surface-core association in the evolution of structured regions, but we found a comparatively weak association between disordered and structured regions. The facts that disordered and structured regions experience different structural constraints and evolve independently provide a unique setup to examine an outstanding question: why is a protein's abundance the main determinant of its sequence conservation? Indeed, any structural or biophysical property linked to the abundance-conservation relationship should increase the relative conservation of regions concerned with that property (e.g., disordered residues with mis-interactions, domain residues with misfolding). Surprisingly, however, we found the conservation of disordered and structured regions to increase in equal proportion with abundance. This observation implies that either abundance-related constraints are structure-independent, or multiple constraints apply to different regions and perfectly balance each other.

Keywords: protein abundance, protein evolution, protein structure, misfolding, intrinsic disorder, contact number, misinteraction, yeast proteome

OPEN ACCESS

Edited by:
Paolo Marcatili,
Technical University of Denmark,
Denmark

Reviewed by:
Andrew James Doig,
The University of Manchester,
United Kingdom
Karen N. Allen,
Boston University, United States

****Correspondence:***
Emmanuel D. Levy
emmanuel.levy@weizmann.ac.il

Specialty section:
This article was submitted to
Structural Biology,
a section of the journal
Frontiers in Molecular Biosciences

Received: *06 November 2020*
Accepted: *29 March 2021*
Published: *30 April 2021*

Citation:
Dubreuil B and Levy ED (2021)
Abundance Imparts Evolutionary
Constraints of Similar Magnitude on
the Buried, Surface, and Disordered
Regions of Proteins.
Front. Mol. Biosci. 8:626729.
doi: 10.3389/fmolb.2021.626729

INTRODUCTION

During the course of evolution, mutations arise throughout genomes and can impact every protein at every site. However, contemplating a multiple sequence alignment of orthologous sequences typically shows widely differing levels of conservation across sites. Additionally, comparing multiple sequence alignments of different orthogroups shows even larger differences: certain groups such as those of ribosomal genes can be well conserved despite hundreds of millions of years of divergence, while others accumulate mutations much faster.

Amino-acid residues within proteins are subject to functional, biophysical, and structural constraints that are interconnected. These constraints result in different degrees of purifying selection along the sequence (i.e., purging of deleterious mutations by natural selection), which yields different levels of positional conservation. We discuss here structural aspects related to these constraints while placing an emphasis on works of Cyrus Chothia, to whom this issue is dedicated, and refer the reader to several reviews for a comprehensive overview (Liberles et al., 2012; Sikosek and Chan, 2014; Echave et al., 2016; Echave and Wilke, 2017). Following the characterization of the first few structures of proteins, their comparative analysis made it clear that the burial of non-polar residues accompanied with Van der Waals interactions and hydrogen bonding were the main contributors to the folding free energy (Chothia, 1974, 1975, 1976; Miller et al., 1987). Confirming the "hydrophobic bonding" intuition of Kauzmann (Kauzmann, 1959) and relying on calculations of molecular surfaces based on the algorithm of Lee and Richards (1971), Chothia estimated that each square Ångstrom of accessible surface removed from contact with water provides a free energy gain of 25 cal. Mol^{-1} (Chothia, 1974, 1975). At the same time, he provided universal relationships governing protein folding, e.g., on the proportion of the total accessible surface of a polypeptide chain that becomes buried upon folding (Chothia, 1975). This simple relationship has a profound meaning with respect to surface-to-volume ratios in folded proteins, notably that longer proteins should fold following a beads-on-a-string model rather than by forming larger beads (Wetlaufer, 1973) – indeed it was soon realized that beads (domains) are fundamental units of protein evolution (Chothia, 1992; Murzin et al., 1995; Bateman et al., 2002; Gough and Chothia, 2002). On top of hydrophobic bonding energy, a high degree of steric complementarity creates a well-packed protein interior (Chothia, 1975), in which mutations are incrementally accommodated by small structural changes (Lesk and Chothia, 1980). Ultimately, as sequences diverge, structures do too, albeit more slowly (Chothia and Lesk, 1986, 1987). Considering that structures are globally maintained during the course of evolution, it is intuitive that buried residues, which contribute to folding and stability more than surface residues (Creighton and Chothia, 1989; Lim and Sauer, 1989; Tokuriki et al., 2007), are more conserved (Koshi and Goldstein, 1995; Goldman et al., 1998; Guo et al., 2004; Bloom et al., 2006; Sasidharan and Chothia, 2007; Goldstein, 2008; Conant and Stadler, 2009; Franzosa and Xia, 2009; Liberles et al., 2012; Yeh et al., 2014; Echave et al., 2015; Shahmoradi and Wilke, 2016; Spielman and Wilke, 2016; Echave and Wilke, 2017; Liu et al., 2017).

We saw that the structure of a protein could help explain why certain positions – notably those buried and in contact with a large number of neighboring residues, are more conserved than others. Protein structure can also help to rationalize why certain proteins, e.g., those with more designable folds, evolve faster than others (Shakhnovich et al., 2005; Bloom et al., 2006). Globally, however, structural information only explains a small fraction of the heterogeneity in evolutionary rates seen across different proteins. Several studies have singled out other protein-centric properties associated with this heterogeneity (Zhang and Yang, 2015), including function (Wall et al., 2005; Lopez-Bigas et al., 2008; Xia et al., 2009), essentiality (Hurst and Smith, 1999; Hirsh and Fraser, 2001; Jordan et al., 2002; Liao et al., 2006), the number of interaction partners (Fraser et al., 2002; Bloom and Adami, 2004; Fraser and Hirsh, 2004; Hahn and Kern, 2005; Kim et al., 2006; Xia et al., 2009), or cellular abundance (Pal et al., 2001; Krylov et al., 2003; Rocha and Danchin, 2004; Subramanian and Kumar, 2004; Drummond et al., 2005; Bloom et al., 2006; Liao et al., 2006; Popescu et al., 2006; Pál et al., 2006; Sällström et al., 2006; Drummond and Wilke, 2008; Xia et al., 2009; Zhang and Yang, 2015). The latter is, by far, the most significant, in particular among unicellular organisms where there is no complexity added by tissue-specific expression. Several mechanistic interpretations of this abundance-conservation association have been proposed (Drummond et al., 2005; Drummond and Wilke, 2008; Cherry, 2010; Gout et al., 2010; Plata et al., 2010; Levy et al., 2012; Yang et al., 2012; Park et al., 2013; Zhang and Yang, 2015) and remain a matter of active debate (Plata and Vitkup, 2018; Razban, 2019). We will scrutinize this relationship further in the results and discussion section, in the context of the results presented.

We have seen how protein structure helped to interpret and rationalize data on evolutionary conservation. Here, we invert this logic to characterize structural properties of disordered regions from data on their evolutionary conservation. First, we compared the evolutionary rate of disordered regions to that of surface residues in the same protein and found that disordered regions are equivalent to super-accessible surface residues. Second, we know that the divergence of surface and core residues is interdependent. In other words, a protein's surface can hardly diverge without mutations arising in its interior as well, and vice-versa. We confirmed this finding in showing that evolutionary rates of surface and interior regions are correlated within proteins ($R > 0.85$). In contrast, the evolutionary rates of disordered and domain regions were poorly coupled ($R \sim 0.25$), indicating that disordered regions are, for the most part, structurally independent from domains in the same sequence. Finally, the structural differences and the lack of interdependence between disordered and structured regions supports that they can be influenced differently by biophysical and structural constraints. For example, an increased purifying selection for protein stability is expected to impact buried residues more than disordered ones. This idea led us to examine whether abundance impacts the relative conservation between these regions. Surprisingly, however, the relative conservation between different regions appeared independent from abundance.

RESULTS AND DISCUSSION

Disordered Regions Are Equivalent to Super-Accessible Surface Residues in Terms of Their Conservation

Among proteins that need to fold into stable structures to function, amino-acid residues buried in the protein interior contribute the most to stability. Consequently, these residues are under stronger purifying selection than surface amino-acid residues, and are, on average, more conserved in the sequence. Two measures of residue burial have been associated with the heterogeneity of conservation in sequences: (i) solvent accessible surface area or ASA (Lee and Richards, 1971; Shrake and Rupley, 1973; Goldman et al., 1998; Bloom et al., 2006; Lin et al., 2007; Conant and Stadler, 2009; Franzosa and Xia, 2009), which measures the surface or fractional surface of an amino-acid residue that is in contact with bulk water, and (ii) the packing density of an amino-acid residue, which measure the density of its neighbors. Different metrics capture this information, including the contact number and the weighted contact number, with the latter containing longer-range information (Franzosa and Xia, 2009; Yeh et al., 2014). While not strictly equivalent, both accessible surface area and packing density correlate strongly (Echave et al., 2016), and both measures show that the less buried is a residue, the less conserved it is within a protein sequence.

This conservation-structure relationship prompts us to infer structural properties of disordered regions from their pattern of conservation within proteins. We know that disordered regions are devoid of a hydrophobic core and therefore cannot autonomously adopt a stable three-dimensional structure. However, if we consider the spectrum of solvent accessibility and packing density found among folded domains, where would disordered regions position themselves on average? Would they appear much less conserved than even the most solvent-exposed regions? Some disordered regions serve purely as linkers or entropic springs and are expected to show very weak sequence conservation (Dyson and Wright, 2005; Van der Lee et al., 2014). At the same time, disordered regions can also form secondary structure elements and bind to partners (Tompa, 2005; Vacic et al., 2007; Uversky and Dunker, 2010; Wright and Dyson, 2015; Banani et al., 2017; Dignon et al., 2019), thereby burying residues and transiently increasing their packing density. For example, p27Kip1 can wrap around the structure of Cdk2 to regulate its function (Russo et al., 1996; Galea et al., 2008).

To position disordered regions on the solvent accessibility spectrum observed in structured regions, we compared the evolutionary rate of residues in both region types. Specifically, we selected 3,350 proteins from *Saccharomyces cerevisiae*, which contain at least 20 residues in both structured regions and disordered regions. We inferred residue-level conservation using Rate4Site (Pupko et al., 2002) on multiple sequence alignments of orthologs from 14 fungal species (see section "Materials and Methods"). Evolutionary and structural information were mapped along the reference sequence from the multiple alignment as illustrated for STI1, a conserved Hsp90 co-chaperone (**Figure 1A**). We calculated a ratio per protein *i*, corresponding to the mean evolutionary rate of residues in disordered regions (R_i^{diso}) divided by the mean rate of residues in a domain (R_i^{domain}). Overall, considering 2607 proteins with known orthologs, containing both types of regions, the median ratio (R_i^{diso}/R_i^{domain}) is equal to 2.2 (**Figure 1B**). If we now consider domains of known structure (i.e., present in PDB, currently ∼670) instead of those predicted, we find a similar median ratio equal to 2.0. For those proteins, we compared the conservation of disordered regions to that of buried and surface residues separately and found ratios equal to 3.1 and 1.4, respectively. Thus, in an average protein of this dataset, disordered regions evolve 3.1 and 1.4-fold faster than buried and surface residues, respectively (**Figure 1B**).

This result is based on a definition of surface that includes residues with >25% relative ASA. As higher ASA is associated with lower conservation, we asked whether increasing the cut-off progressively from >25 to >80% would yield a point where surface residues evolve faster than disordered ones (**Figure 1C**). We did not reach such a point as the ratio remained above 1 for all values. However, the ratio did converge to a value close to 1, highlighting that in an average protein, disordered residues are almost equivalent in their conservation to the most exposed residues at the surface of structured regions.

If we assume that the differential conservation of sites within protein sequences largely reflects different structural constraints, we can infer that disordered regions are, on average, highly solvent-exposed and under weak structural constraints. In sum, our results place disordered regions in the continuum of protein structure, at the end of the solvent-accessibility spectrum. It will be interesting to refine this relationship in the future. For example, by comparing additional properties such as hydrophobicity (Kyte and Doolittle, 1982) or stickiness (Levy, 2010), by considering where disordered segments fall in the sequence (e.g., N/C-ter and inside domains), or by breaking down disorder into different types (Bellay et al., 2011).

Conservation of Disorder Versus Domains Is Poorly Correlated Among Low Abundance Proteins and the Correlation Increases With Abundance

Individual residues within a structure contribute to stability together. As a result, we can expect the evolutionary conservation of residues within a structure to be uniform. To examine this idea, we compared the average evolutionary rate of surface and buried amino-acid residues within structures. Importantly, we know that protein abundance imposes global constraints on the conservation of proteins, which may also result in a uniform evolutionary pressure across the sequence, independently of the structure. Thus, we initially focused on low abundance proteins in which such global constraints are minimized. We observed the conservation of surface and buried regions to correlate strongly ($R > 0.83$, **Figure 2A**), which is reminiscent of the surface-core association described previously (Tóth-Petróczy and Tawfik, 2011).

We next compared the association in evolutionary conservation between disordered regions and domains found in the same protein. In this case, the correlation was reduced

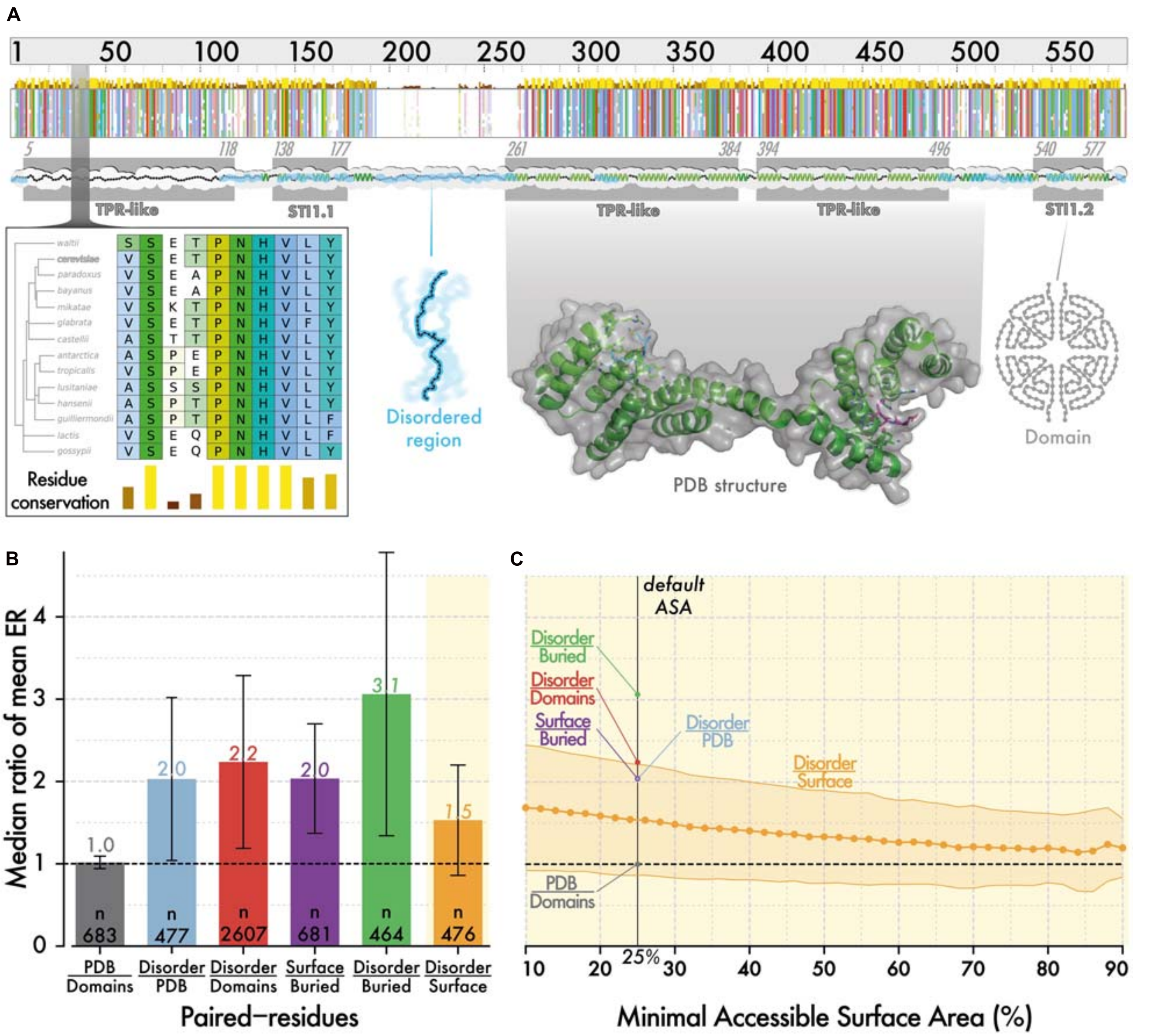

FIGURE 1 | The evolutionary rate of disordered regions is comparable to that of super-exposed regions in folded proteins. **(A)** Evolutionary information and structural features are mapped onto protein sequences from *S. cerevisiae*. The minimap represents the multiple sequence alignment of orthologous sequences to STI1. The amino acids are colored using CLUSTAL's color scale (Thompson et al., 1994) depending on residue type and conservation. The zoomed-in panel illustrates residue-level conservation, which we calculated with Rate4Site (Pupko et al., 2002). We mapped the positions of PFAM (Bateman et al., 2002) and SUPERFAMILY (Gough and Chothia, 2002) domains (gray box), and of disordered regions predicted by IUPRED (Dosztányi, 2018) (cyan ribbon). We also mapped structural information available from PDB (Rose et al., 2017; Armstrong et al., 2019) and 3DComplex (Levy et al., 2006) on sequences. For this particular sequence, structural information was partially available based on PDB code 3UQ3 (Schmid et al., 2012). **(B)** Within proteins, the evolutionary rate of residues in different regions are averaged, and we compare the ratio of these averages. We show the median of ratios with error bars corresponding to the median absolute deviation. Surface and buried residues are defined based on relative ASA of >25 and ≤25%, respectively (Levy, 2010). **(C)** We calculate the same ratio as in panel **(B)**, between disordered regions and surface regions, using an increasingly stringent relative ASA cut-off to define surface residues. As we increase the cutoff, the median ratio tends toward 1, which highlights that disordered residues evolve only slightly faster than the most exposed residues at protein surfaces.

greatly ($R = 0.25$), indicating that the structural connectivity and interdependence between disordered regions and domains are globally weak. These results are consistent with those of the previous section, which depict disordered regions as being highly solvent-accessible and structurally independent from domains. However, proteins expressed at higher levels show increasing correlation, from $R = 0.40$ among medium abundance proteins, to $R = 0.63$ in the class of proteins with the highest abundance (**Figure 2B**, lower row). This apparent coupling in evolutionary rates is unlikely to have a structural

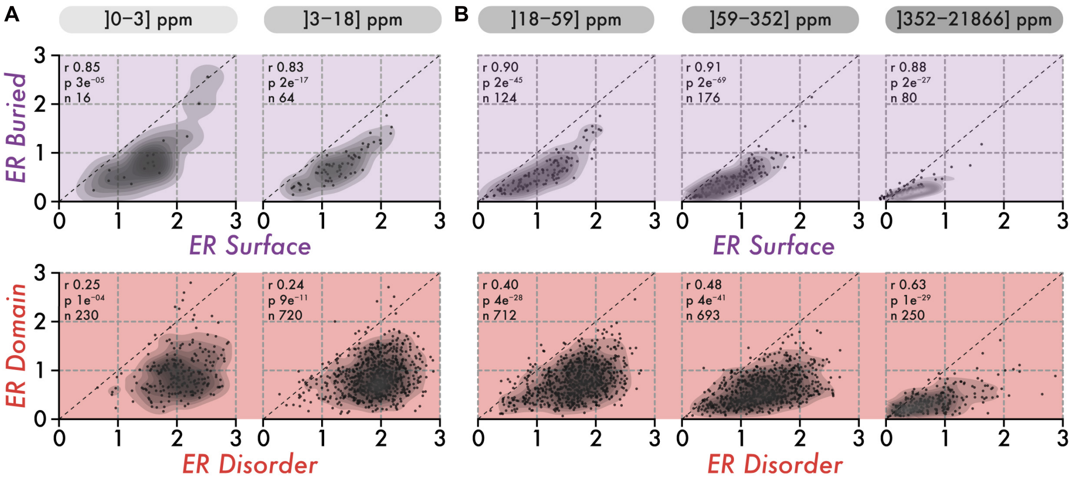

FIGURE 2 | The correlation in the conservation of disorder vs domain regions is poor among low abundance proteins and increases with abundance. **(A)** The top row shows the average evolutionary rate (ER) of surface residues (*x*-axis) vs buried residues (*y*-axis) per protein, for two classes of abundance (0–3 and 3–18 ppm or parts per millions). The lower row shows the average ER of disordered residues (*x*-axis) vs residues in domains (*y*-axis) per protein, for the same two classes of abundance. A protein falling on the diagonal (dashed line) means that residues in the two regions being compared have equal evolutionary rates (i.e., a ratio of 1). The Spearman rank correlation coefficient (r), the associated *p*-value (*p*, two-sided Spearman's rank correlation test), and the number of proteins (n) within each class of abundance are given above each scatterplot. **(B)** Same as in panel **(A)**, for three classes of abundance (18–59, 59–352, and 352–21,866 ppm or parts per million).

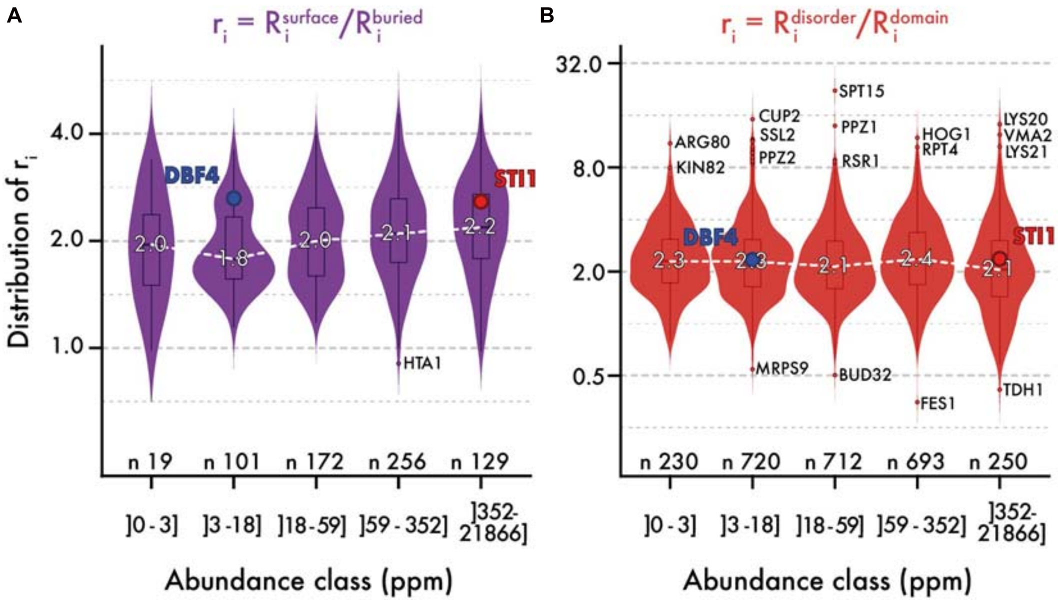

FIGURE 3 | The relative evolutionary rates of different protein regions are steady with abundance. Distribution of evolutionary rates ratio between different regions in the sequence (*y*-axis), across five classes of protein abundance (*x*-axis). A ratio is calculated by dividing the average evolutionary rate of residues found in two regions panel **(A)** surface vs. buried, panel **(B)** disorder vs. domain. The white dashed line highlights the median ratio across bins of abundance. Overlaid box plots show the interquartile range (IQR = 25 to 75% quantiles) with their whiskers extending to 1.58 × IQR. Beyond this interval, the three most extreme outlier values are annotated. The number of proteins contributing to each distribution is given. We also highlight the relative rates for a pair of proteins, one with low and one with high abundance (STI1 and DBF4). These two proteins show comparable structural features, different evolutionary rates (respectively, 0.575 and 1.34 for their full sequence), and similar ratios.

origin. Rather, it probably results from global constraints linked to abundance and exerted on the whole protein sequence. This apparent coupling also implies that different regions in a sequence all experience increasingly strong purifying selection with increasing abundance. This observation led us to quantify whether such negative selection increases equally in all regions, or whether some regions become more constrained than others.

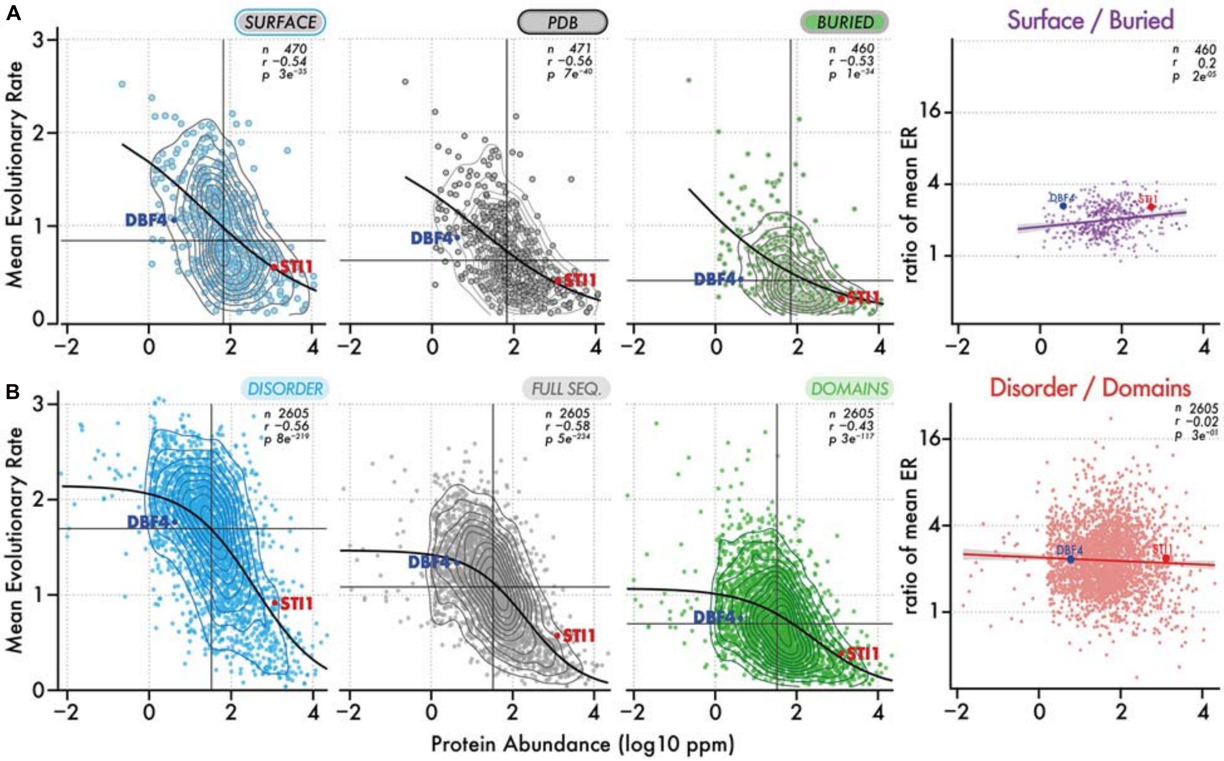

FIGURE 4 | Evolutionary rates of different regions and their ratio as a function of abundance. **(A)** Evolutionary rates (*y*-axis) as a function of protein abundance (*x*-axis) for surface regions, full-length structures, and buried regions. The ratio of evolutionary rate for surface vs buried regions is also shown as a function of abundance. Contour lines show the density of points. The median evolutionary rate and median protein abundance are shown by a vertical and horizontal line, respectively. The Spearman rank correlation coefficient and *p*-value are given with the number of proteins in each dataset. A black line shows the fitted sigmoidal regression for each plot. We highlight two proteins, one with a low and one with a high abundance (DBF4 and STI1). Both have comparable structural features but different evolutionary rates. **(B)** Same representation as in panel **(A)**, now considering disordered versus domain regions.

Evolutionary Constraints Imparted by Protein Abundance Scale Similarly Among Surface, Buried, and Disordered Regions

We saw that surface residues in a protein evolve twice as fast as buried residues on average. This difference, which has long been recognized, is mainly explained by solvent-accessibility/packing density and reflects that protein structures are more likely to be destabilized by mutations at buried positions than by mutations at the surface (Koshi and Goldstein, 1995; Goldman et al., 1998; Guo et al., 2004; Bloom et al., 2006; Sasidharan and Chothia, 2007; Goldstein, 2008; Conant and Stadler, 2009; Franzosa and Xia, 2009; Liberles et al., 2012; Yeh et al., 2014; Echave et al., 2015; Shahmoradi and Wilke, 2016; Spielman and Wilke, 2016; Echave and Wilke, 2017; Liu et al., 2017). Similarly, residues in disordered regions evolve faster than those in domains. Interestingly, this reflects that surface, buried, and disordered residues experience different structural and biophysical constraints. Thus, we propose to examine whether the ratio of their conservation is changing as a function of abundance. For example, observing that buried residues are twice more conserved than surface residues among low abundance proteins, and become four-times more conserved among high abundance proteins would suggest that stability is increasingly constrained with higher abundance.

We analyzed the ratio of conservation (**Figures 3A**, **4A**) of surface and buried residues as a function of abundance. The distribution of these ratios showed comparable median values of about ∼2. In the highest abundance class, this ratio reached ∼2.2 (**Figure 3A**) creating a significant albeit weak ($R = 0.2$) correlation (**Figure 4A**). Overall, the ratio is relatively stable, implying that both regions are constrained to a similar extent with increasing abundance. Alternatively, a relatively constant ratio could be favored by the coupling we observed between interior and surface regions (**Figure 2**, top row). Accordingly, constraints placed on the protein surface could percolate to interior regions and vice versa (Tóth-Petróczy and Tawfik, 2011). To control for this effect, we next compared disordered and domain regions, which show minimal structural coupling. We also observed a stable ratio of ∼2 across the five same abundance classes (**Figure 3B**), and we observed no dependence of the ratio with abundance even at the highest levels ($R = -0.02$, **Figure 4B**). Additionally, focusing on disorder and domain regions increased the size of the dataset as we were not limited by the availability of atomic-resolution structures, so this observation applies to the yeast proteome.

By definition, disordered regions and domains should experience distinct structural and biophysical constraints. Thus, the fact that these two regions appear equally constrained with increasing abundance is puzzling and can be interpreted in different ways. One possible explanation is that constraints associated with abundance apply to entire sequences independently of structure. Such constraints could include translational selection (Akashi, 2003), although region-specific codon-bias constraints may exist as well (Tuller et al., 2010; Pechmann and Frydman, 2013), cost of expression (Dekel and Alon, 2005; Wagner, 2005; Cherry, 2010; Gout et al., 2010; Plata et al., 2010), as well as other functional elements and sequence properties that may impact transcription or translation (Stergachis et al., 2013; Zhou et al., 2016). Alternatively or in addition, region-specific structural and biophysical constraints associated with protein concentration could increase in similar proportions with abundance, resulting in a stable ratio. In this case, two primary constraints have been characterized: a first on protein stability (Serohijos et al., 2012, 2013) leading to selection against misfolding (Drummond et al., 2005; Drummond and Wilke, 2008), would dominate among interior residues. A second, on protein solubility (Knowles et al., 2014; Garcia-Seisdedos et al., 2017, 2018; Dubreuil et al., 2019; Foy et al., 2019; Macossay-Castillo et al., 2019; Vecchi et al., 2020), with selection against promiscuous interactions (Deeds et al., 2007; Levy et al., 2009, 2012; Liberles et al., 2011; Yang et al., 2012), would dominate among solvent-exposed residues. However, the fact that constraints on different regions scale proportionally with abundance may appear surprising and will need to be explored in future works.

CONCLUSION

We analyzed the evolutionary conservation of sites within proteins, and of proteins within proteomes. We found that disordered regions evolve about three-fold faster than buried regions, and 1.4-fold faster than surface regions. Additionally, disordered regions evolve about as fast as the most solvent-exposed surface regions, highlighting that they extend the continuum of protein structure as a "super-accessible" surface. Unlike regular surface residues, however, disordered regions evolve more independently from domains in the same sequence. This independence allowed us to examine how abundance constrains different regions that are not structurally connected in sequences. Notably, the evolution of disordered regions and domains changed in a similar proportion with abundance: on average, disordered regions evolved twice as fast as domains across the entire range of abundance. Since different regions are subject to different structural and biophysical constraints, we foresee that such comparative analyses of conservation-ratios as a function of abundance will help identify mechanisms underlying the abundance-conservation relationship. It is likely that multiple mechanisms are at play (Mehlhoff et al., 2020) and may be captured by targeted analyses of specific regions and protein subsets.

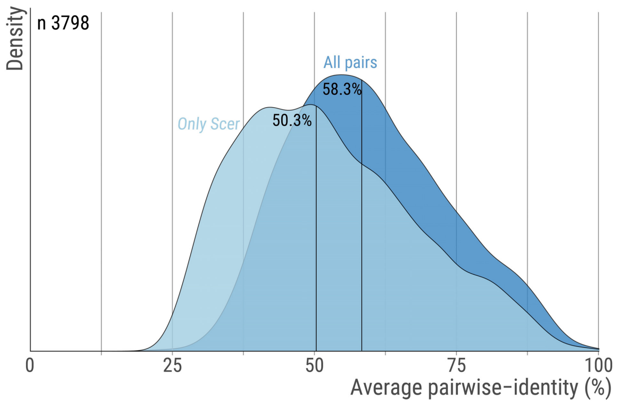

FIGURE 5 | Pairwise sequence identity across orthologs pairs. For each orthogroup we calculate the average percent sequence-identity using all ortholog pairs or only pairs that include the *S. cerevisiae* protein. The distribution for these two measures are shown with dark and light blue, respectively. Vertical lines highlight the median. The number of orthogroups is 3,798.

MATERIALS AND METHODS

Reference Proteome Sequences

The sequences were taken from the reference *S. cerevisiae* proteome maintained by SGD (Cherry et al., 2012). To facilitate data integration, we also mapped those reference sequences against the UniprotKB complete proteome for *S. cerevisiae* (Stutz et al., 2006; UniProt Consortium, 2019).

Crystallographic Structures

We relied on the 3DComplex database (Levy et al., 2006) to map UNIPROT sequences onto atomic coordinates of protein structures. For each yeast protein, the structures matching the UNIPROT sequence with the largest sequence overlap (minimum 20%) and identity above 90% were retained. Only experimentally determined crystallographic structures with resolutions below 3.0 Ångtrsoms were considered.

Cellular Abundance

Protein abundances were obtained from Pax-Db (v4.0, May 2015) (Wang et al., 2012, 2015), which provides relative abundances for unicellular and multicellular organisms including tissue-specific data. We use overall abundance inferred from all available data sets (integrated data set).

Orthologs Alignment and Position-Specific Evolutionary Rate

The orthologs' alignments were obtained from the original work by Wapinski et al. (2007). Briefly, genes sharing significant sequence similarity were clustered into putative orthogroups and their phylogeny was constructed by a modified neighbor-joining procedure based on pre-computed residues similarities and shared synteny scores. This process was repeated and optimized until each orthogroup consisted of genes that shared a single common ancestor. Here, we used 3798 groups of orthologous proteins along with their multiple sequence alignment encompassing 14 fungal species (*S.cerevisiae, Saccharomyces paradoxus, Saccharomyces mikatae, Saccharomyces bayanus, Naumovozyma castellii (Saccharomyces castellii), Candida glabrata, Kluyveromyces lactis, Debaryomyces hansenii, Yarrowia lipolytica, Eremothecium gossypii (Ashbya gossypii), Lachancea waltii (Kluyveromyces waltii), Candida albicans, Aspergillus nidulans, Fusarium graminearum, Magnaporthe grisea, Neurospora crassa, Cryptococcus neoformans, Schizosaccharomyces pombe)* were used. Only 6 orthogroups had one sequence missing and these were replaced by indels. The median pairwise sequence identity within these 3,798 orthogroups is 58.3% (**Figure 5**).

All alignments were computed using MUSCLE (Edgar, 2004) and then concatenated to estimate residue-level evolutionary rate using the software Rate4Site (Pupko et al., 2002). Additional details on how evolutionary rates were estimated are available in Landry et al. (2009).

Intrinsic Disorder Predictions

We predicted disordered regions in the yeast proteome by combining short and long disorder segments predicted by IUPred (Mészáros et al., 2009; Dosztányi, 2018). We considered the 20% amino-acid residues with the highest disorder probabilities among all proteins. In all analyses, we required a minimum number of 20 residues in a particular region to calculate an average evolutionary rate. When fewer residues were available, the average rate of the region was considered undefined.

Domains Assignment

To assign domains, we aligned profiles from Pfam-A (v27.0, May 2013) (Bateman et al., 2002; Finn et al., 2014) and SUPERFAMILY (v1.75, March 2013) (Gough, 2002; Oates et al., 2015) to reference proteome sequences, filtering the hits with an *E*-value score above 10^{-3}. Finally, domain residues are those that were identified as part of a hit from either Pfam, SUPERFAMILY, or both.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s. Data used in this work are available on Figshare in a tabulated format: https://doi.org/10.6084/m9.figshare.13738657.

AUTHOR CONTRIBUTIONS

BD and EL designed the analyses and experiments, analyzed the data, and wrote the manuscript. BD carried out the analyses. Both authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Israel Science Foundation (grant No. 1452/18), by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 819318), by a research grant from A.-M. Boucher, by research grants from the Estelle Funk Foundation, the Estate of Fannie Sherr, the Estate of Albert Delighter, the Merle S. Cahn Foundation, Mildred S. Gosden, the Estate of Elizabeth Wachsman, and the Arnold Bortman Family Foundation.

ACKNOWLEDGMENTS

We thank H. Greenblatt for helping with the computer infrastructure and Tal Pupko for his advice.

REFERENCES

Akashi, H. (2003). Translational selection and yeast proteome evolution. *Genetics* 164, 1291–1303.

Armstrong, D. R., Berrisford, J. M., Conroy, M. J., Gutmanas, A., Anyango, S., Choudhary, P., et al. (2019). PDBe: improved findability of macromolecular structure data in the PDB. *Nucleic Acids Res.* 48, D335–D343.

Banani, S. F., Lee, H. O., Hyman, A. A., and Rosen, M. K. (2017). Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* 18, 285–298.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., et al. (2002). The Pfam protein families database. *Nucleic Acids Res.* 30, 276–280.

Bellay, J., Han, S., Michaut, M., Kim, T., Costanzo, M., Andrews, B. J., et al. (2011). Bringing order to protein disorder through comparative genomics and genetic interactions. *Genome Biol.* 12:R14.

Bloom, J. D., and Adami, C. (2004). Evolutionary rate depends on number of protein-protein interactions independently of gene expression level: response. *BMC Evol. Biol.* 4:14. doi: 10.1186/1471-2148-4-14

Bloom, J. D., Drummond, D. A., Arnold, F. H., and Wilke, C. O. (2006). Structural determinants of the rate of protein evolution in yeast. *Mol. Biol. Evol.* 23, 1751–1761.

Cherry, J. L. (2010). Expression level, evolutionary rate, and the cost of expression. *Genome Biol. Evol.* 2, 757–769.

Cherry, J. M., Hong, E. L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E. T., et al. (2012). *Saccharomyces* genome database: the genomics resource of budding yeast. *Nucleic Acids Res.* 40, D700–D705.

Chothia, C. (1974). Hydrophobic bonding and accessible surface area in proteins. *Nature* 248, 338–339.

Chothia, C. (1975). Structural invariants in protein folding. *Nature* 254, 304–308.

Chothia, C. (1976). The nature of the accessible and buried surfaces in proteins. *J. Mol. Biol.* 105, 1–12.

Chothia, C. (1992). Proteins. One thousand families for the molecular biologist. *Nature* 357, 543–544.

Chothia, C., and Lesk, A. M. (1986). The relation between the divergence of sequence and structure in proteins. *EMBO J.* 5, 823–826.

Chothia, C., and Lesk, A. M. (1987). The evolution of protein structures. *Cold Spring Harb. Symp. Quant. Biol.* 52, 399–405.

Conant, G. C., and Stadler, P. F. (2009). Solvent exposure imparts similar selective pressures across a range of yeast proteins. *Mol. Biol. Evol.* 26, 1155–1161.

Creighton, T. E., and Chothia, C. (1989). Protein structure. Selecting buried residues. *Nature* 339, 14–15.

Deeds, E. J., Ashenberg, O., Gerardin, J., and Shakhnovich, E. I. (2007). Robust protein protein interactions in crowded cellular environments. *Proc. Natl. Acad. Sci. U.S.A.* 104, 14952–14957.

Dekel, E., and Alon, U. (2005). Optimality and evolutionary tuning of the expression level of a protein. *Nature* 436, 588–592.

Dignon, G. L., Zheng, W., and Mittal, J. (2019). Simulation methods for liquid–liquid phase separation of disordered proteins. *Curr. Opin. Chem. Eng.* 23, 92–98.

Dosztányi, Z. (2018). Prediction of protein disorder based on IUPred. *Protein Sci.* 27, 331–340.

Drummond, D. A., and Wilke, C. O. (2008). Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* 134, 341–352.

Drummond, D. A., Bloom, J. D., Adami, C., Wilke, C. O., and Arnold, F. H. (2005). Why highly expressed proteins evolve slowly. *Proc. Natl. Acad. Sci. U.S.A.* 102, 14338–14343.

Dubreuil, B., Matalon, O., and Levy, E. D. (2019). Protein abundance biases the amino acid composition of disordered regions to minimize non-functional interactions. *J. Mol. Biol.* 431, 4978–4992.

Dyson, H. J., and Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6, 197–208.

Echave, J., and Wilke, C. O. (2017). Biophysical models of protein evolution: understanding the patterns of evolutionary sequence divergence. *Annu. Rev. Biophys.* 46, 85–103.

Echave, J., Jackson, E. L., and Wilke, C. O. (2015). Relationship between protein thermodynamic constraints and variation of evolutionary rates among sites. *Phys. Biol.* 12:025002.

Echave, J., Spielman, S. J., and Wilke, C. O. (2016). Causes of evolutionary rate variation among protein sites. *Nat. Rev. Genet.* 17, 109–121.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230.

Foy, S. G., Wilson, B. A., Bertram, J., Cordes, M. H. J., and Masel, J. (2019). A shift in aggregation avoidance strategy marks a long-term direction to protein evolution. *Genetics* 211, 1345–1355.

Franzosa, E. A., and Xia, Y. (2009). Structural determinants of protein evolution are context-sensitive at the residue level. *Mol. Biol. Evol.* 26, 2387–2395.

Fraser, H. B., and Hirsh, A. E. (2004). Evolutionary rate depends on number of protein-protein interactions independently of gene expression level. *BMC Evol. Biol.* 4:13. doi: 10.1186/1471-2148-4-13

Fraser, H. B., Hirsh, A. E., Steinmetz, L. M., Scharfe, C., and Feldman, M. W. (2002). Evolutionary rate in the protein interaction network. *Science* 296, 750–752.

Galea, C. A., Nourse, A., Wang, Y., Sivakolundu, S. G., Heller, W. T., and Kriwacki, R. W. (2008). Role of intrinsic flexibility in signal transduction mediated by the cell cycle regulator, p27 Kip1. *J. Mol. Biol.* 376, 827–838.

Garcia-Seisdedos, H., Empereur-Mot, C., Elad, N., and Levy, E. D. (2017). Proteins evolve on the edge of supramolecular self-assembly. *Nature* 548, 244–247.

Garcia-Seisdedos, H., Villegas, J. A., and Levy, E. D. (2018). Infinite assembly of folded proteins in evolution, disease, and engineering. *Angew. Chem. Int. Ed. Engl.* 58, 5514–5531. doi: 10.1002/anie.201806092

Goldman, N., Thorne, J. L., and Jones, D. T. (1998). Assessing the impact of secondary structure and solvent accessibility on protein evolution. *Genetics* 149, 445–458.

Goldstein, R. A. (2008). The structure of protein evolution and the evolution of protein structure. *Curr. Opin. Struct. Biol.* 18, 170–177.

Gough, J. (2002). The SUPERFAMILY database in structural genomics. *Acta Crystallogr. D Biol. Crystallogr.* 58, 1897–1900.

Gough, J., and Chothia, C. (2002). SUPERFAMILY: HMMs representing all proteins of known structure. SCOP sequence searches, alignments and genome assignments. *Nucleic Acids Res.* 30, 268–272.

Gout, J.-F., Kahn, D., Duret, L., and Paramecium Post-Genomics Consortium (2010). The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. *PLoS Genet.* 6:e1000944. doi: 10.1371/journal.pgen.1000944

Guo, H. H., Choe, J., and Loeb, L. A. (2004). Protein tolerance to random amino acid change. *Proc. Natl. Acad. Sci. U.S.A.* 101, 9205–9210.

Hahn, M. W., and Kern, A. D. (2005). Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Mol. Biol. Evol.* 22, 803–806.

Hirsh, A. E., and Fraser, H. B. (2001). Protein dispensability and rate of evolution. *Nature* 411, 1046–1049.

Hurst, L. D., and Smith, N. G. (1999). Do essential genes evolve slowly? *Curr. Biol.* 9, 747–750.

Jordan, I. K., Rogozin, I. B., Wolf, Y. I., and Koonin, E. V. (2002). Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. *Genome Res.* 12, 962–968.

Kauzmann, W. (1959). "Some factors in the interpretation of protein denaturation11the preparation of this article has been assisted by a grant from the national science foundation," in *Advances in Protein Chemistry*, eds C. B. Anfinsen, M. L. Anson, K. Bailey, and J. T. Edsall (Cambridge, MA: Academic Press), 1–63.

Kim, P. M., Lu, L. J., Xia, Y., and Gerstein, M. B. (2006). Relating three-dimensional structures to protein networks provides evolutionary insights. *Science* 314, 1938–1941. doi: 10.1126/science.1136174

Knowles, T. P., Vendruscolo, M., and Dobson, C. M. (2014). The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.* 15, 384–396.

Koshi, J. M., and Goldstein, R. A. (1995). Context-dependent optimal substitution matrices. *Protein Eng. Des. Sel.* 8, 641–645.

Krylov, D. M., Wolf, Y. I., Rogozin, I. B., and Koonin, E. V. (2003). Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. *Genome Res.* 13, 2229–2235.

Kyte, J., and Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157, 105–132.

Landry, C. R., Levy, E. D., and Michnick, S. W. (2009). Weak functional constraints on phosphoproteomes. *Trends Genet*. 25, 193–197. doi: 10.1016/j.tig.2009.03.003

Lee, B., and Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* 55, 379–400.

Lesk, A. M., and Chothia, C. (1980). How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. *J. Mol. Biol.* 136, 225–270.

Levy, E. D. (2010). A simple definition of structural regions in proteins and its use in analyzing interface evolution. *J. Mol. Biol.* 403, 660–670.

Levy, E. D., De, S., and Teichmann, S. A. (2012). Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. *Proc. Natl. Acad. Sci. U.S.A.* 109, 20461–20466.

Levy, E. D., Landry, C. R., and Michnick, S. W. (2009). How perfect can protein interactomes be? *Sci. Signal.* 2:e11.

Levy, E. D., Pereira-Leal, J. B., Chothia, C., and Teichmann, S. A. (2006). 3D complex: a structural classification of protein complexes. *PLoS Comput. Biol.* 2:e155. doi: 10.1371/journal.pcbi.0020155

Liao, B.-Y., Scott, N. M., and Zhang, J. (2006). Impacts of gene essentiality, expression pattern, and gene compactness on the evolutionary rate of mammalian proteins. *Mol. Biol. Evol.* 23, 2072–2080.

Liberles, D. A., Teichmann, S. A., Bahar, I., Bastolla, U., Bloom, J., Bornberg-Bauer, E., et al. (2012). The interface of protein structure, protein biophysics, and molecular evolution. *Protein Sci.* 21, 769–785.

Liberles, D. A., Tisdell, M. D. M., and Grahnen, J. A. (2011). Binding constraints on the evolution of enzymes and signalling proteins: the important role of negative pleiotropy. *Proc. Biol. Sci.* 278, 1930–1935.

Lim, W. A., and Sauer, R. T. (1989). Alternative packing arrangements in the hydrophobic core of lambda repressor. *Nature* 339, 31–36.

Lin, Y.-S., Hsu, W.-L., Hwang, J.-K., and Li, W.-H. (2007). Proportion of solvent-exposed amino acids in a protein and rate of protein evolution. *Mol. Biol. Evol.* 24, 1005–1011.

Liu, J.-W., Lin, J.-J., Cheng, C.-W., Lin, Y.-F., Hwang, J.-K., and Huang, T.-T. (2017). On the relationship between residue structural environment and sequence conservation in proteins. *Proteins* 85, 1713–1723.

Lopez-Bigas, N., De, S., and Teichmann, S. A. (2008). Functional protein divergence in the evolution of Homo sapiens. *Genome Biol.* 9:R33.

Macossay-Castillo, M., Marvelli, G., Guharoy, M., Jain, A., Kihara, D., Tompa, P., et al. (2019). The balancing act of intrinsically disordered proteins: enabling functional diversity while minimizing promiscuity. *J. Mol. Biol.* 431, 1650–1670. doi: 10.1016/j.jmb.2019.03.008

Mehlhoff, J. D., Stearns, F. W., Rohm, D., Wang, B., Tsou, E.-Y., Dutta, N., et al. (2020). Collateral fitness effects of mutations. *Proc. Natl. Acad. Sci. U.S.A.* 117, 11597–11607.

Mészáros, B., Simon, I., and Dosztányi, Z. (2009). Prediction of protein binding regions in disordered proteins. *PLoS Comput. Biol.* 5:e1000376. doi: 10.1371/journal.pcbi.1000376

Miller, S., Janin, J., Lesk, A. M., and Chothia, C. (1987). Interior and surface of monomeric proteins. *J. Mol. Biol.* 196, 641–656.

Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247, 536–540.

Oates, M. E., Stahlhacke, J., and Vavoulis, D. V. (2015). The SUPERFAMILY 1.75 database in 2014: a doubling of data. *Nucleic Acids Res* 43, D227–D233.

Pal, C., Papp, B., and Hurst, L. D. (2001). Highly expressed genes in yeast evolve slowly. *Genetics* 158, 927–931.

Pál, C., Papp, B., and Lercher, M. J. (2006). An integrated view of protein evolution. *Nat. Rev. Genet.* 7, 337–348.

Park, C., Chen, X., Yang, J.-R., and Zhang, J. (2013). Differential requirements for mRNA folding partially explain why highly expressed proteins evolve slowly. *Proc. Natl. Acad. Sci. U.S.A.* 110, E678–E686.

Pechmann, S., and Frydman, J. (2013). Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat. Struct. Mol. Biol.* 20, 237–243.

Plata, G., and Vitkup, D. (2018). Protein stability and avoidance of toxic misfolding do not explain the sequence constraints of highly expressed proteins. *Mol. Biol. Evol.* 35, 700–703.

Plata, G., Gottesman, M. E., and Vitkup, D. (2010). The rate of the molecular clock and the cost of gratuitous protein synthesis. *Genome Biol.* 11:R98.

Popescu, C. E., Borza, T., Bielawski, J. P., and Lee, R. W. (2006). Evolutionary rates and expression level in *Chlamydomonas*. *Genetics* 172, 1567–1576.

Pupko, T., Bell, R. E., Mayrose, I., Glaser, F., and Ben-Tal, N. (2002). Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics* 18(Suppl. 1), S71–S77.

Razban, R. M. (2019). Protein melting temperature cannot fully assess whether protein folding free energy underlies the universal abundance–evolutionary rate correlation seen in proteins. *Mol. Biol. Evol.* 36, 1955–1963.

Rocha, E. P., and Danchin, A. (2004). An analysis of determinants of amino acids substitution rates in bacterial proteins. *Mol. Biol. Evol.* 21, 108–116.

Rose, P. W., Prlic, A., Altunkaya, A., Bi, C., Bradley, A. R., Christie, C. H., et al. (2017). The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.* 45, D271–D281.

Russo, A. A., Jeffrey, P. D., Patten, A. K., Massagué, J., and Pavletich, N. P. (1996). Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A-Cdk2 complex. *Nature* 382, 325–331.

Sällström, B., Arnaout, R. A., Davids, W., Bjelkmar, P., and Andersson, S. G. E. (2006). Protein evolutionary rates correlate with expression independently of synonymous substitutions in *Helicobacter* pylori. *J. Mol. Evol.* 62, 600–614.

Sasidharan, R., and Chothia, C. (2007). The selection of acceptable protein mutations. *Proc. Natl. Acad. Sci. U.S.A.* 104, 10080–10085.

Schmid, A. B., Lagleder, S., Gräwert, M. A., Röhl, A., Hagn, F., Wandinger, S. K., et al. (2012). The architecture of functional modules in the Hsp90 co-chaperone Sti1/Hop. *EMBO J.* 31, 1506–1517.

Serohijos, A. W. R., Lee, S. Y. R., and Shakhnovich, E. I. (2013). Highly abundant proteins favor more stable 3D structures in yeast. *Biophys. J.* 104, L1–L3.

Serohijos, A. W. R., Rimas, Z., and Shakhnovich, E. I. (2012). Protein biophysics explains why highly abundant proteins evolve slowly. *Cell Rep.* 2, 249–256.

Shahmoradi, A., and Wilke, C. O. (2016). Dissecting the roles of local packing density and longer-range effects in protein sequence evolution. *Proteins Struct. Funct. Bioinf.* 84, 841–854.

Shakhnovich, B. E., Deeds, E., Delisi, C., and Shakhnovich, E. (2005). Protein structure and evolutionary history determine sequence space topology. *Genome Res.* 15, 385–392.

Shrake, A., and Rupley, J. A. (1973). Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* 79, 351–371.

Sikosek, T., and Chan, H. S. (2014). Biophysics of protein evolution and evolutionary protein biophysics. *J. R. Soc. Interface* 11:20140419.

Spielman, S. J., and Wilke, C. O. (2016). Extensively parameterized mutation–selection models reliably capture site-specific selective constraint. *Mol. Biol. Evol.* 33, 2990–3002.

Stergachis, A. B., Haugen, E., Shafer, A., Fu, W., Vernot, B., Reynolds, A., et al. (2013). Exonic transcription factor binding directs codon choice and affects protein evolution. *Science* 342, 1367–1372.

Stutz, A., Bairoch, A., and Estreicher, A. (2006). UniProtKB/Swiss-Prot: the protein sequence knowledgebase. *FEBS J.* 273, 62–62.

Subramanian, S., and Kumar, S. (2004). Gene expression intensity shapes evolutionary rates of the proteins encoded by the vertebrate genome. *Genetics* 168, 373–381.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673

Tokuriki, N., Stricher, F., Schymkowitz, J., Serrano, L., and Tawfik, D. S. (2007). The stability effects of protein mutations appear to be universally distributed. *J. Mol. Biol.* 369, 1318–1332.

Tompa, P. (2005). The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett.* 579, 3346–3354.

Tóth-Petróczy, A., and Tawfik, D. S. (2011). Slow protein evolutionary rates are dictated by surface-core association. *Proc. Natl. Acad. Sci. U.S.A.* 108, 11151–11156.

Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., et al. (2010). An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* 141, 344–354.

UniProt Consortium (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506–D515.

Uversky, V. N., and Dunker, A. K. (2010). Understanding protein non-folding. *Biochim. Biophys. Acta* 1804, 1231–1264.

Vacic, V., Oldfield, C. J., Mohan, A., Radivojac, P., Cortese, M. S., Uversky, V. N., et al. (2007). Characterization of molecular recognition features, MoRFs, and their binding partners. *J. Proteome Res.* 6, 2351–2366.

Van der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., et al. (2014). Classification of intrinsically disordered regions and proteins. *Chem. Rev.* 114, 6589–6631.

Vecchi, G., Sormanni, P., Mannini, B., Vandelli, A., Tartaglia, G. G., Dobson, C. M., et al. (2020). Proteome-wide observation of the phenomenon of life on the edge of solubility. *Proc. Natl. Acad. Sci. U.S.A.* 117, 1015–1020.

Wagner, A. (2005). Energy constraints on the evolution of gene expression. *Mol. Biol. Evol.* 22, 1365–1374.

Wall, D. P., Hirsh, A. E., Fraser, H. B., Kumm, J., Giaever, G., Eisen, M. B., et al. (2005). Functional genomic analysis of the rates of protein evolution. *Proc. Natl. Acad. Sci. U.S.A.* 102, 5483–5488.

Wang, M., Herrmann, C. J., Simonovic, M., Szklarczyk, D., and von Mering, C. (2015). Version 4.0 of PaxDb: protein abundance data, integrated across model organisms, tissues, and cell-lines. *Proteomics* 15, 3163–3168. doi: 10.1002/pmic.201400441

Wang, M., Weiss, M., Simonovic, M., Haertinger, G., Schrimpf, S. P., Hengartner, M. O., et al. (2012). PaxDb, a database of protein abundance averages across all three domains of life. *Mol. Cell. Proteomics* 11, 492–500.

Wapinski, I., Pfeffer, A., Friedman, N., and Regev, A. (2007). Natural history and evolutionary principles of gene duplication in fungi. *Nature* 449, 54–61.

Wetlaufer, D. B. (1973). Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. Natl. Acad. Sci. U. S. A.* 70, 697–701. doi: 10.1073/pnas.70.3.697

Wright, P. E., and Dyson, H. J. (2015). Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* 16, 18–29.

Xia, Y., Franzosa, E. A., and Gerstein, M. B. (2009). Integrated assessment of genomic correlates of protein evolutionary rate. *PLoS Comput. Biol.* 5:e1000413. doi: 10.1371/journal.pcbi.1000413

Yang, J. R., Liao, B. Y., Zhuang, S. M., and Zhang, J. Z. (2012). Protein misinteraction avoidance causes highly expressed proteins to evolve slowly. *Proc. Natl. Acad. Sci. U.S.A.* 109, E831–E840.

Yeh, S.-W., Huang, T.-T., Liu, J.-W., Yu, S.-H., Shih, C.-H., Hwang, J.-K., et al. (2014). Local packing density is the main structural determinant of the rate of protein sequence evolution at site level. *Biomed. Res. Int.* 2014: 572409.

Zhang, J., and Yang, J. R. (2015). Determinants of the rate of protein sequence evolution. *Nat. Rev. Genet.* 16, 409–420.

Zhou, Z., Dang, Y., Zhou, M., Li, L., Yu, C.-H., Fu, J., et al. (2016). Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proc. Natl. Acad. Sci. U.S.A.* 113, E6117–E6125.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.