REVIEW

Relative Specificity: All Substrates Are Not Created Equal

Yan Zeng *

Department of Medicine, Masonic Cancer Center, University of Minnesota, Minneapolis, MN 55455, USA

Abstract   A biological molecule, *e.g.*, an enzyme, tends to interact with its many cognate substrates, targets, or partners differentially. Such a property is termed relative specificity and has been proposed to regulate important physiological functions, even though it has not been examined explicitly in most complex biochemical systems. This essay reviews several recent large-scale studies that investigate protein folding, signal transduction, RNA binding, translation and transcription in the context of relative specificity. These results and others support a pervasive role of relative specificity in diverse biological processes. It is becoming clear that relative specificity contributes fundamentally to the diversity and complexity of biological systems, which has significant implications in disease processes as well.

Introduction

Relative specificity is defined as the characteristic whereby in a biochemical system, a molecule, symbolized as E, interacts with its numerous substrates, targets or partners (collectively symbolized as {S}) differentially, thereby impacting them distinctively depending on the identity of individual substrates, targets or partners [1]. E can be a protein, RNA or any other biological molecule, capable of interacting with other molecules, *i.e.*, {S}, through binding and/or catalysis. Some examples are hemoglobin binding to O_2, CO_2 and a few other molecules, a receptor capturing different ligands, a cytochrome P450 enzyme metabolizing diverse chemicals, an RNA-binding protein associating with its RNA targets, a protein kinase phosphorylating substrates, a protein chaperone contacting unfolded or partially folded proteins, RNA polymerases transcribing genes, and the ribosome translating mRNAs. In many cellular systems, {S} can number in the hundreds, thousands or more. Research, however, has traditionally been centered on determining whether a molecule is the substrate or not of an E of interest. For various reasons, that an E might not target all of its {S} equally in complex systems, *i.e.*, relative specificity, is rarely treated as default, and what physiological consequences relative specificity may incur is even less investigated [1].

Evidence does exist to suggest that relative specificity is functionally relevant in complex biochemical systems. For example, the RNase Drosha (E) cleaved hundreds of human primary microRNA transcripts ({S}) with different efficiencies *in vitro*, which correlated with the expression of mature microRNAs *in vivo*, and such specificity was partially explainable by the structural properties of {S} [2]. The functionality of

* Corresponding author.
  E-mail: zengx033@umn.edu (Zeng Y).

relative specificity was also detected in systems involving a transcription factor and a protein kinase in budding yeast and an RNA-binding protein in humans [1]. The phenomena were generalized to formulate the relative specificity hypothesis, which has a number of features and implications. Firstly, it focuses on complex systems where an E acts on many, *e.g.*, hundreds or thousands of substrates, since such systems are abundant in nature, yet their relative specificity has been largely ignored partially due to technical limitations. Testing the hypothesis requires that we examine and compare the interactions between an E and its numerous substrates and then correlate the preferential interactions with a phenotype, in order to filter out the effects from factors other than the E of interest and make credible references to how the E's relative specificity contributes to a biological outcome. This is critical also because an observed biochemical property, *e.g.*, the binding of an E to a target, does not automatically equate to any biological function *in vivo*.

Secondly, the hypothesis does not stipulate the nature or origin of relative specificity in myriad biochemical systems or consider the abundance and subcellular localization of E and {S} to a first approximation. An E can bind to {S} with different on and off rates, different affinities, *etc*. It can bind to {S} to induce distinct conformational changes to selectively impact downstream signaling. It can also bind to {S} before efficient or inefficient enzymatic reactions. Any of these modes and their combinations could underlie the mechanism of relative specificity.

Thirdly, biological processes frequently mandate several biochemical activities in succession or in parallel. Analogous to rate-limiting reactions, even if relative specificity is exhibited by multiple Es in a biological process, the process may still be determined largely by the specificity from one of these Es. As an example, RNases Drosha and Dicer function in the same microRNA biogenesis pathway and were shown to cleave their respective substrates of microRNA transcripts preferentially *in vitro*, yet only the selectivity by Drosha significantly correlated with mature microRNA expression *in vivo* [2,3].

Lastly, the hypothesis promotes a reevaluation of certain concepts. For example, the literature contains ample statements like this: protein X has a high specificity. What it ultimately means is that X does not have many substrates. But it is likely that X still has more than one substrate, and X does not treat them equally. Conversely, if protein Y has a low specificity, then Y has many substrates, but again, Y still differentially interacts with these {S}. Furthermore, consider the following two schemes. In the first, an E acts identically on many substrates, which have different, sometime even opposing functions, but a specific biologic outcome nevertheless results, *e.g.*, cancer. In the second, the E acts on the same {S} differentially, again leading to a specific outcome, *e.g.*, cancer. Hence, a similar phenotype might originate from two distinct mechanisms. The key to distinguish between the two possibilities is to reveal whether the E reacts with {S} differently and the phenotypes can be partially explained by this relative specificity.

Below, I will review several recent studies of diverse biological systems in the context of relative specificity: (1) Hsp90–client interactions, (2) protein phosphorylation by the mechanistic target of rapamycin complex 1 (mTORC1), (3) RNA stabilization by RNA binding protein, fox-1 homolog (RBFOX1), (4) the impact of N-terminal codons on translation and (5) genome-wide transcription. I will then discuss a number of issues raised by the relative specificity hypothesis.

Hsp90–client interactions

Hsp90 is a molecular chaperone that associates with a large number of client proteins ({S}) to facilitate their folding. To study how Hsp90 recognizes {S}, Taipale et al. used a reporter assay to quantify the interactions between Hsp90 and hundreds of potential clients systematically in cell cultures [4]. Hsp90 was shown to interact with the majority of human kinases. However, the interaction was not binary, *i.e.*, substrate *vs.* non-substrate, but rather, as a sign of relative specificity, varied over a 100-fold range in strength, according to the reporter readouts. Cdc37, a co-chaperone of Hsp90, also selectively interacted with human kinases in a manner highly correlated with Hsp90. Mechanistically, the thermal stability of the kinases, with still poorly defined but both localized and broadly distributed components, was proposed to be the major determinant of how Hsp90 selects and discriminates {S}.

What are the functional consequences of differential Hsp90–client interactions? Taipale et al. found a modest but significant, negative correlation between the strength of Hsp90 interaction and recombinant kinase expression ($R^2 = 0.15$; [4]). Experimentally, the stronger the interaction, the larger the extent to which a recombinant client protein might be destabilized in cell cultures when Hsp90 was inhibited pharmacologically. In addition, weak human Hsp90 client kinases were more readily overexpressed than strong clients in bacteria, which lack Hsp90. These data suggest that Hsp90 selectivity might buffer protein folding; without Hsp90, intrinsically unstable proteins would have been even less stable and expressed at a lower level.

Protein phosphorylation by mTORC1

That relative specificity is functionally relevant can be easily rationalized if it correlates with differential gene expression. For the budding yeast Cdk1, its relative specificity *in vitro* positively correlated with substrate phosphorylation during mitosis, but whether the mere fact that Cdk1 phosphorylates {S} to varying degrees would have a biological consequence is not straightforward to address [1]. This potentially novel, global form of regulation has been at least partially tackled in the mTORC1 system [5]. mTORC1 is a protein kinase that controls metabolism and growth in response to many stimuli, and its activity is altered by aging and in human disease such as cancers and can be inhibited by the drug rapamycin. Kang et al. performed *in vitro* kinase assays with recombinant mTORC1 and short peptides corresponding to mTORC1 phosphorylation sites in various substrates and their mutants [5]. Short peptides were used as the proxy for endogenous proteins because sequences immediately surrounding the phosphorylation sites in native substrates contain the most critical information necessary for kinase recognition and phosphorylation. mTORC1 was shown to phosphorylate some peptides/substrates more readily than others, indicative of relative specificity, or substrate quality as termed by Kang et al. [5]. mTORC1 activity depended partially on substrate binding affinity. Kang et al. then used a number of tests to demonstrate functional relevance [5]. For example, rapamycin, a pharmacologically

important agent, preferentially reversed the phosphorylation of poor mTORC1 substrates while had little effect on that of the strong substrates. This is because rapamycin blocks mTORC1 incompletely, and weak mTORC1 substrates would be the first or more acutely affected by reduced mTORC1 activity. Moreover, low amino acid and serum conditions differentially affected the phosphorylation of mTORC1 substrates in cell cultures, consistent with the *in vitro* kinase assay data and substrate quality. Lastly, under a starving amino acid condition, mouse embryonic fibroblasts expressing a wildtype mTORC1 substrate S6K1 grew slower than cells expressing an engineered S6K1 that was better phosphorylated by mTORC1, indicating that a change in relative specificity or substrate quality impacted cell proliferation under reduced nutrient conditions. Relative specificity, therefore, enables mTORC1 to differentially phosphorylate substrates under the same physiological inputs, which modulates cell behaviors and drug sensitivity.

RNA stabilization by RBFOX1

Most RNA-binding proteins associate with RNA targets via short, degenerate RNA sequences or motifs. Ray et al. systematically analyzed RNA motifs recognized by over 200 RNA-binding domains/proteins *in vitro* and computationally correlated the presence of RNA motifs in transcriptomes to the functions of RNA-binding proteins [6]. A number of proteins were thereby proposed to regulate RNA expression or splicing. One of them was RBFOX1, a protein implicated in neurodevelopmental and psychiatric disorders such as autism spectrum disorder. The numbers of predicted RBFOX1 RNA binding motifs at the 3′ untranslated regions (UTRs) of mRNAs positively correlated with the abundance of mRNAs [6]. Using the data of RNA-seq following RBFOX1 knockdown in primary human neural progenitor cells [7], Ray et al. found that the number of predicted FBFOX1 binding sites in mRNAs also positively correlated with the extent to which RBFOX1 knockdown reduced the expression of the mRNAs [6], reminiscent of the finding concerning Hsp90 and clients [4]. As reduced RBFOX1 levels in the brains of autism patients had been noted [8], it was further shown that predicted RBFOX1 targets had progressively lower mRNA expression in these patients [6]. In summary, mRNAs varied in their predicted RBFOX1 binding sites, and mRNAs with a high number of predicted binding sites were more likely dependent on RBFOX1 for their expression levels than those with fewer binding sites, potentially contributing to differential gene expression under normal, physiological conditions as well as pathological conditions such as autism. Our understanding of the functions of RNA-binding proteins such as RBFOX1 would be enhanced by considering relative binding strength of the binding sites, studying the effects on splicing and incorporating information about the actual binding events through isolating protein:RNA complexes from cells followed by RNA-seq.

Impact of N-terminal codons on translation

The ribosome does not translate its {S}, *i.e.*, mRNAs equally. Translation efficiency depends on factors such as mRNA length, expression, structure, codon usage, cognate tRNA dosage and ribosome abundance [9]. In bacteria and many eukaryotes, though not necessarily in other eukaryotes such as mammals, the first ∼10 codons after the start AUG in mRNAs are enriched for low efficiency codons and the RNA secondary structure in this 5′ region is less stable than that in other regions of the mRNAs [10–15]. Because rare codons are also more A/U rich, especially at the wobble position, it was unclear whether slower translation at the N-terminus due to rare codons or reduced RNA structure ultimately led to increased protein expression. Two recent reports addressed this question [15,16].

In the first report, Bentele et al. surveyed 414 bacterial genomes and found that at the N-terminal region, codon usage favors a flexible mRNA structure [15]. Rare codons are selected only if they are A/U rich, while abundant codons disfavored only if they are G/C rich. The authors proposed that rare codons are selected for not because they are rare, but because they weaken the mRNA structure. The authors went on to confirm this conclusion by testing the expression of two reporter genes in *Escherichia coli*.

In the second report, Goodman et al. directly examined the cause and effect of N-terminal codon bias on translation by analyzing the expression of 14,234 artificial constructs that differed in their promoters, ribosome binding sites, 11 N-terminal codons with various amino acid sequences and synonymous codons in front of a common reporter gene [16]. There was a significant increase (⩾10-fold) in protein abundance from adopting the common to the rare synonymous codons, after adjusting for the effects of promoters and ribosome binding. The authors then correlated reporter expression to a number of metrics. As expected, codon rarity and reduced secondary structure were the strongest predictors of high expression. After controlling for the secondary structure changes among codons, however, there was no longer any relationship between N-terminal codon usage and expression, while secondary structure remained correlated with reporter expression after controlling for codon usage. Thus, a weak secondary structure is primarily responsible for increases in protein expression. Endogenous mRNAs differ in their N-terminal coding sequences, codon usage and secondary structures, which could greatly impart their differential translation efficiencies. Indeed, highly expressed genes might contain more low efficiency codons at this region than weakly expressed genes [17].

Genome-wide transcription

Benefiting from the availability of genome sequences and early application of genomics technologies, a number of studies of gene transcription have incorporated the analysis of relative specificity, albeit often implicitly. These include correlating transcription factor (TF) binding to target mRNA expression following computational predictions of target genes or experimental measurements of DNA binding using the chromatin immunoprecipitation (ChIP)-microarray or ChIP-seq method [18]. As reviewed by Biggin [19], animal TFs might bind to many targets in a cell over a quantitative continuum in "Continuous Networks", as opposed to the "Discrete Networks" most people or models describe. Here I will use data chiefly from the recently completed, ENcyclopedia Of DNA Elements (ENCODE) project as an example to discuss relative specificity in DNA transcription.

The ENCODE project generated genome-wide DNA binding data for 119 proteins, including sequence-specific TFs, general TFs, RNA polymerases II (Pol II) and III, and histone modifying complexes, in dozens of human cell lines/types [20]. Consistent with other studies, ENCODE found that a protein typically binds to regions associated with a large number of loci with varying binding signals. How relevant are these variations, since binding as identified by ChIP does not necessitate functionality in a cell? The authors went on to show that the total TF binding signals near the transcription start sites could predict the vast differences in transcript abundance [20–22]. For example, aggregate TF binding explained at least 67% of the variance in the levels of 5′ m^7GpppN cap-containing RNAs in K562 cells, although the predictive power was weaker for RNAs prepared and sequenced in different manners [21]. Contribution by individual proteins was likewise analyzed, and TFs with more binding sites in the genome, such as Yin Yang 1 (YY1), also tended to contribute more to RNA variations in cells.

Interestingly, DNA binding by the protein REST positively correlated with the variance in transcription initiation, despite REST being a transcriptional repressor [21]. To reconcile this discrepancy, I compared the relative specificity of RE1-silencing transcription factor (REST), serum response factor (SRF), YY1 and Pol II using the publicly available ChIP-seq and RNA-seq data for human K562, HepG2, GM12878 and embryonic stem cells (ESCs). Here, protein binding to a target gene was represented by cumulative ChIP signals from two kilobases upstream of the transcription start site to two kilobases downstream of the transcription termination site. This adjustment was made to accommodate the fact that Pol II elongates transcripts along the body of genes and that although usually 60% of the ChIP peaks by sequence-specific TFs are near the transcription start sites, TFs also bind and function further away. DNA binding was then correlated to RNA-seq data. In essence, transcriptional activities along the whole gene loci were examined.

Table 1 shows the Spearman correlation between protein binding and RNA expression. In all four human cell lines, REST bound hundreds of target genes differentially, which negatively correlated with target mRNA expression. This result is consistent with the transcriptional repressor role of REST and suggests that its relative specificity contributes to differential gene suppression in human cells. The correlation was the weakest in K562 cells, perhaps partially explaining why a positive correlation was obtained by Cheng and colleagues [21]. On the other hand, REST target genes identified in K562 cells far out-numbered those in other cells (Table 1). If limited to the 2000 genes with the highest REST ChIP signals, a target number more in line with those in the other three cell lines, a stronger correlation was obtained: $\rho = -0.38$, $P = 5.9 \times 10^{-71}$. Thus, capped-RNA production at the promoter regions might not adequately reflect the functions of certain TFs or transcription of the whole genes; *e.g.*, a TF might act on transcription elongation. SRF ChIP signals correlated weakly with target mRNA expression, and not surprisingly, Pol II usually possessed the highest target gene numbers and strongest positive correlations, followed closely by YY1 (Table 1), consistent with the findings of Cheng and colleagues [21]. Notably, correlation shown in Table 1 was weaker than that reported in a few other studies in mammalian cells (*e.g.*, [21,23,24]). Explanations may lie in the differences in cells and TFs that were examined and how ChIP signals were selected and weighted for analyses. Nevertheless, the trend is clear: TFs including Pol II occupy DNA in a quantitative continuum, and for some of these proteins, their relative specificity likely plays a nontrivial role in differential mRNA expression *in vivo*.

Does relative specificity contribute to other biological phenomena?

From the examples above and others, an argument can be made that relative specificity is a prevalent, even if often overlooked, regulatory mechanism in biological processes. For example, there have been a number of puzzling discoveries: chiefly, while a protein is known to serve as a universal and essential factor in a fundamental biological process, its mutation, or gain or loss of function can yield a very specific outcome. A case in point is eIF4E, a general translation initiation factor, yet its enhanced activity promotes tumorigenesis [25] or induces an autisitic phenotype [26–28] in mammalian systems. Moreover, ribosome biogenesis factors and

Table 1   Correlation between TF occupancy and mRNA expression in several human cell lines

TF	K562 (GSM581666)			HepG2 (GSM591672)			GM12878 (GSM591661)			ESC (GSM591658)		
	N	ρ	P	N	ρ	P	N	ρ	P	N	ρ	P
REST	7822	−0.14 (GSM803440)	2.6×10^{-36}	1154	−0.33 (GSM803344)	1.5×10^{-30}	1812	−0.34 (GSM803349)	6.9×10^{-49}	3789	−0.21 (GSM803365)	6.2×10^{-40}
SRF	3076	0.1 (GSM803520)	8.6×10^{-8}	3023	0.03 (GSM803502)	0.08	1812	0.12 (GSM803350)	3.1×10^{-12}	1859	0.02 (GSM803425)	0.37
Pol II	10,067	0.51 (GSM803410)	0	9865	0.36 (GSM803368)	1.2×10^{-296}	10,589	0.53 (GSM803355)	0	10,891	0.35 (GSM803366)	5.8×10^{-304}
YY1	9988	0.4 (GSM803446)	0	10,801	0.41 (GSM803381)	0	11,117	0.47 (GSM803406)	0	10,189	0.24 (GSM803513)	4.6×10^{-129}

Note: Transcription factors (TFs) include Pol II here. *N* indicates number of target genes that had both ChIP-seq and RNA-seq signals in the same cell line; *ρ* represents Spearman rank correlation coefficient; and *P* stands for the *P* values calculated using SPSS, version 19 (IBM). Genes were identified by aligning the ChIP-seq data to the Human Genome hg19 using Galaxy at http://galaxyproject.org/. Listed in parentheses are the Gene Expression Omnibus accession numbers of RNA-seq datasets for the four indicated cell lines and ChIP-seq datasets for the respective TFs in these cell lines. ESC, embryonic stem cell; REST, RE1-silencing transcription factor; SRF, serum response factor; YY1, Yin Yang 1.

ribosomal proteins are necessary for the production and function of ribosomes, but mutations in some of those proteins lead to tissue-specific phenotypes and diseases, or ribosomopathies [29,30]. An example is the universal ribosomal protein L38, whose mutations cause unique patterning defects in mouse embryos, including homeotic transformations of the axial skeleton, whereas mutants of several other ribosomal proteins show no such defects [29,31].

How to explain these phenotypes? Elevated eIF4E might increase the translation of some cancer-promoting mRNAs with a long and stable 5′ UTR [25] or the expression of neuroligins, leading to neurological deficits [27]. In L38 mutant embryos, the translation of a subset of Homeobox mRNAs was altered [31]. Despite these findings, the underlying mechanisms remain incompletely understood. Multiple possibilities may be in play, including varying tissue or developmental requirements for the general factors or sensitivities to their perturbations, and potentially non-identical ribosome composition in different cells [29,31]. To gain a better understanding of ribosomopathies and, in fact, other analogous biological phenomena, however, one might also want to consider relative specificity. There is little direct experimental evidence yet, but it is not hard to imagine that even the same ribosome would differentially translate its tens of thousands of mRNAs, and that translation of some mRNAs would naturally be more affected than that of others by changes in the translation machinery. This is analogous to the aforementioned example of how N-terminal codon bias influences translation efficiency, although our understanding of the regulation of mammalian mRNA translation remains limited.

To further illustrate the functional ramifications of relative specificity, consider the scenario shown in Figure 1A. There are two systems, led by E1 and E2, which act on and only on the same substrates, S1, S2, S3,…,S100, but differ in their relative
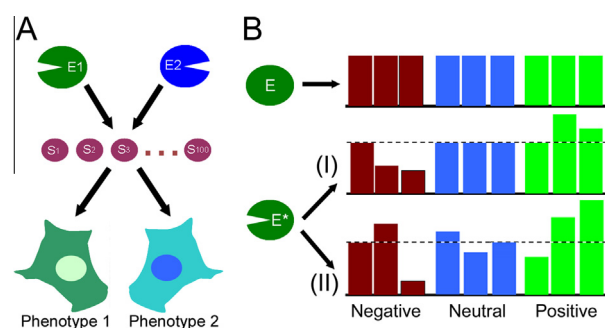
specificity. E1 through these {S} produces phenotype 1, whereas E2 through the same {S} produces phenotype 2. The question, then, is: will phenotype 1 be the same as phenotype 2? The relative specificity hypothesis would say no. An approximate situation is perhaps created by a common experimental strategy where E2 is an overexpressed version of E1. It is predicted that even if E1 and E2 still target the same {S}, {S} under E2 will not change uniformly compared to {S} under E1, and in the case of eIF4E, E2 leads to cell transformation. Further worth mentioning is a fascinating finding of Claveria and colleagues [32]. During mouse embryonic development, epiblast cells express heterogeneous levels of Myc, and cells with high Myc levels would out-compete neighboring cells with relatively lower Myc levels. If, for simplicity, Myc regulates the same sets of genes regardless of Myc levels, but differentially, then this might be a clear functional consequence as a result of Myc relative specificity. Of course, it is unknown whether the phenotypes of increased eIF4E and Myc are the direct or indirect effects of relative specificity, but it will be interesting to see if one can find or create different types of systems of Figure 1A to deliberately and directly test the relative specificity hypothesis.

Testing the relative specificity hypothesis

Evaluating relative specificity and its significance in biological processes can be challenging. Better *in vitro* assays, *e.g.*, to compare the translation of endogenous mRNAs, may have to be developed to more faithfully reflect situations especially in mammalian systems. Also needed are better measurements of biological molecules and their interactions in cells. For example, RNA-seq has the potential to more accurately reveal RNA expression than microarray that is based on hybridization to short oligonucleotide probes. Global protein expression has been profiled using mass spectrometry and calculated based on raw peptide signals, which give a poor account of endogenous protein levels. Furthermore, the majority of ChIP studies have examined TF occupancy, but TF binding dynamics may be a more appropriate predictor of gene transcription [33]. The above considerations could partly explain why the correlation between relative specificity and a phenotype, *e.g.*, gene expression, tends to be low. However, with improved technologies and modeling, relative specificity's functional contribution will be more precisely determined.

So far, the final step in testing relative specificity is to obtain a significant correlation with a cellular phenotype. Because the biochemical activity of an E in question is usually known, *e.g.*, a kinase phosphorylates substrates, an RNase cleaves RNA intermediates preceding the production of mature RNAs, the causal relationship between the E's relative specificity and a particular biological outcome can be safely inferred. Kang et al. went further to show that manipulating the substrate quality of an mTORC1 target impacted cell growth [5], although it remains possible that the manipulation of a single target affected its function irrespective of its phosphorylation by mTORC1. Nevertheless, studying phenotypic changes following alteration of relative specificity at a larger scale, as shown in Figure 1A, will fundamentally enhance our appreciation of relative specificity and its regulation of biological systems.

Figure 1   **Functional significance of relative specificity**
A. E1 and E2 act on the same set of substrates (S1, S2, S3,…,S100) but with different relative specificity to produce phenotypes 1 and 2, respectively. **B.** E has a number of substrates that are functionally negative, neutral or positive with respect to a certain phenotype. A mutant E* that produces the net phenotype may alter the substrates in two ways (I and II). In scenario I, E* represses only the negative substrates and induces only the positive substrates. In scenario II, E* can repress or induce all three types of substrates, and it is the uneven changes in these substrates that would ultimately lead to the phenotype of interest. The height of columns symbolizes expression levels, with dash lines representing expression under E.

Is relative specificity evolutionarily conserved?

Does Drosha or another microRNA processing factor in fish or flies discriminate its substrates to regulate microRNA production, and does mTORC1 differentially phosphorylate substrates in other species as in humans? We do not have the answers yet. At the sequence level, Es such as kinases, RNA-binding proteins and TFs are often conserved and recognize homologous motifs, suggesting that they might have and differentiate similar sets of {S} across species as well. On the other hand, sequence homology is not the only way to conserve biological functions. For example, rRNAs differ in their primary structures but maintain a similar fold throughout evolution, while mating gene expression and heterochromatin formation are enforced by different mechanisms in fungi. It is plausible, therefore, that relative specificity functions as a conserved mechanism even in the absence of obvious sequence homology. Related to this issue, the genome-wide DNA binding sites of TFs had been compared in multiple species and found to have undergone widespread gains and losses [34–40]. Yet, a significant number of binding site losses in one species are recovered by gains at nearby loci in another species, vice versa [38,41]. When a TF binds to orthologous genes in humans and mice, most binding sites need not be aligned [35]. These results suggest that even as genomes and TF binding sites have extensively diverged, there is a constraint to buffer the loss of differential target gene binding by TFs at a global scale. Likewise, translation efficiency across many species is regulated by mRNA codon usage and structures near the start codon, and orthologous genes are translated differentially to reflect the physiological divergence of species [15,16,42].

Implications of relative specificity in biological systems

Biological systems are diverse and complex, with variations at the species level, at the individual organisms' level, at the cellular level and at the molecular level. For example, the mammalian genomes contain over 20,000 protein-coding genes, and their mRNA levels vary by at least 4 orders of magnitude and protein levels at least 6 orders of magnitude [43]. What, then, accounts for such diversity and complexity? An obvious explanation is evolution by natural selection. Other possible answers include stochasticity and self assembly [44]. Conceivably, relative specificity could also contribute to the diversity and complexity in biological systems. Differential TF binding has been proposed to lead to individual variability [37,45].

Although clear evidence is lacking, if relative specificity is important for normal, physiological processes, it likely impacts pathogenesis as well. For example, the functions of Hsp90, mTORC1, RBFOX1 and many transcription factors, such as Myc and p53, have been implicated in human disease. Therefore, studying relative specificity could further help us understand disease processes, as illustrated in **Figure 1B**. Suppose an E has its normal, physiological substrates or targets, while a mutation in E results in a disease. Traditionally, one would go by identifying changes in the substrates or targets of E that are consistent with the disease phenotype, such as increased oncogene or decreased tumor suppressor expression

(scenario I, Figure 1B). Studies over the years, however, have shown that rarely a single target or a small set of targets, can sufficiently account for the mutant effects. More likely, the E has many {S}, and its mutation alters {S} differentially: *e.g.*, some growth-promoting genes have increased expression, while some growth-promoting genes have decreased expression, same with growth-inhibitory genes (scenario II, Figure 1B). It is speculated that it is these changes and the unbalanced changes in many {S}, not a dominant change in a small number of {S}, that may lead to the disease phenotype.

Conclusion

The appreciation of relative specificity in complex biochemical systems and its prevalent physiological significance has been closely linked to technological advances, which have in turn led to conceptual changes. Up until 20 years ago, to understand the function of an enzyme, for example, one could only focus on a handful of substrates at a time while ignoring a vast number of "non-substrates". The development of high-throughput techniques and bioinformatics allowed one to identify hundreds of substrates simultaneously, but typically only a small number of "the most promising" ones were subsequently scrutinized. Then, approximately 10 years ago, the importance of groups of genes in common complexes and pathways began to be widely appreciated, and research now routinely examines hundreds of genes according to their functional classifications. The acknowledgement of relative specificity could prompt us, in a more deliberate manner, to seek out a comprehensive list of {S}, investigate and compare their changes in response to perturbations, and rationalize how the uneven changes in many of the {S} ultimately contribute to the complex biological phenotypes.

Competing interests

The author declares no competing interest.

References

[1] Zeng Y. The functional consequences and implications of relative substrate specificity in complex biochemical systems. Front Genet 2011;2:65.

[2] Feng Y, Zhang X, Song Q, Li T, Zeng Y. Drosha processing controls the specificity and efficiency of global microRNA expression. Biochim Biophys Acta 2011;1809:700–7.

[3] Feng Y, Zhang X, Graves P, Zeng Y. A comprehensive analysis of precursor microRNA cleavage by human Dicer. RNA 2012;18:2083–92.

[4] Taipale M, Krykbaeva I, Koeva M, Kayatekin C, Westover KD, Karras GI, et al. Quantitative analysis of HSP90–client interactions reveals principles of substrate recognition. Cell 2012;150:987–1001.

[5] Kang SA, Pacold ME, Cervantes CL, Lim D, Lou HJ, Ottina K, et al. mTORC1 phosphorylation sites encode their sensitivity to starvation and rapamycin. Science 2013;341:1236566.

[6] Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, et al. A compendium of RNA-binding motifs for decoding gene regulation. Nature 2013;499:172–7.

[7] Fogel BL, Wexler E, Wahnich A, Friedrich T, Vijayendran C, Gao F, et al. RBFOX1 regulates both splicing and transcriptional networks in human neuronal development. Hum Mol Genet 2012;21:4171–86.

[8] Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature 2011;474:380–4.

[9] Shah P, Ding Y, Niemczyk M, Kudla G, Plotkin JB. Rate-limiting steps in yeast protein translation. Cell 2013;153:1589–601.

[10] Allert M, Cox JC, Hellinga HW. Multifactorial determinants of protein expression in prokaryotic open reading frames. J Mol Biol 2010;402:905–18.

[11] Gu W, Zhou T, Wilke CO. A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. PLoS Comput Biol 2010;6:e1000664.

[12] Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, Zaborske J, et al. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell 2010;141:344–54.

[13] Tuller T, Waldman YY, Kupiec MM, Ruppin E. Translation efficiency is determined by both codon bias and folding energy. Proc Natl Acad Sci U S A 2010;107:3645–50.

[14] Pechmann S, Frydman J. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. Nat Struct Mol Biol 2013;20:237–43.

[15] Bentele K, Saffert P, Rauscher R, Ignatova Z, Blüthgen N. Efficient translation initiation dictates codon usage at gene start. Mol Syst Biol 2013;9:675.

[16] Goodman DB, Church GM, Kosuri S. Causes and effects of N-terminal codon bias in bacterial genes. Science 2013;342:475–9.

[17] Eyre-Walker A, Bulmer M. Reduced synonymous substitution rate at the start of enterobacterial genes. Nucleic Acids Res 1993;21:4599–603.

[18] Kim HD, Shay T, O'Shea EK, Regev A. Transcriptional regulatory circuits: predicting numbers from alphabets. Science 2009;325:429–32.

[19] Biggin MD. Animal transcription networks as highly connected, quantitative continua. Dev Cell 2011;21:611–26.

[20] ENCODE. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57–74.

[21] Cheng C, Alexander R, Min R, Leng J, Yip KY, Rozowsky J, et al. Understanding transcriptional regulation by integrative analysis of transcription factor binding data. Genome Res 2012;22:1658–67.

[22] Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, Cheng C, et al. Architecture of the human regulatory network derived from ENCODE data. Nature 2012;489:91–100.

[23] Cheng C, Gerstein M. Modeling the relative relationship of transcription factor binding and histone modifications to gene expression levels in mouse embryonic stem cells. Nucleic Acids Res 2012;40:553–68.

[24] Ouyang Z, Zhou Q, Wong WH. ChIP-Seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells. Proc Natl Acad Sci U S A 2009;106:21521–6.

[25] Sonenberg N. EIF4E, the mRNA cap-binding protein, from basic discovery to translational research. Biochem Cell Biol 2008;86:178–83.

[26] Neves-Pereira M, Müller B, Massie D, Williams JH, O'Brien PC, Hughes A, et al. Deregulation of EIF4E: a novel mechanism for autism. J Med Genet 2009;46:759–65.

[27] Gkogkas CG, Khoutorsky A, Ran I, Rampakakis E, Nevarko T, Weatherill DB, et al. Autism-related deficits via dysregulated eIF4E-dependent translational control. Nature 2013;493:371–7.

[28] Santini E, Huynh TN, MacAskill AF, Carter AG, Pierre P, Ruggero D, et al. Exaggerated translation causes synaptic and behavioural aberrations associated with autism. Nature 2013;493:411–5.

[29] Topisirovic I, Sonenberg N. Translational control by the eukaryotic ribosome. Cell 2011;145:333–4.

[30] McCann KL, Baserga SJ. Mysterious ribosomopathies. Science 2013;341:849–50.

[31] Kondrashov N, Pusic A, Stumpf C, Shimizu K, Hsieh AC, Xue S, et al. Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning. Cell 2011;145:383–97.

[32] Clavería C, Giovinazzo G, Sierra R, Torres M. Myc-driven endogenous cell competition in the early mammalian embryo. Nature 2013;500:39–44.

[33] Lickwar CR, Mueller F, Hanlon SE, McNally JG, Lieb JD. Genome-wide protein–DNA binding dynamics suggest a molecular clutch for transcription factor function. Nature 2012;484:251–5.

[34] Borneman AR, Gianoulis TA, Zhang ZD, Yu H, Rozowsky J, Seringhaus MR, et al. Divergence of transcription factor binding sites across related yeast species. Science 2007;317:815–9.

[35] Odom DT, Dowell RD, Jacobsen ES, Gordon W, Danford TW, MacIsaac KD, et al. Tissue-specific transcriptional regulation has diverged significantly between human and mouse. Nat Genet 2007;39:730–2.

[36] Tuch BB, Li H, Johnson AD. Evolution of eukaryotic transcription circuits. Science 2008;319:1797–9.

[37] Kasowski M, Grubert F, Heffelfinger C, Hariharan M, Asabere A, Waszak SM, et al. Variation in transcription factor binding among humans. Science 2010;328:232–5.

[38] Schmidt D, Wilson MD, Ballester B, Schwalie PC, Brown GD, Marshall A, et al. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. Science 2010;328:1036–40.

[39] Hemberg M, Kreiman G. Conservation of transcription factor binding events predicts gene expression across species. Nucleic Acids Res 2011;39:7092–102.

[40] Stefflova K, Thybert D, Wilson MD, Streeter I, Aleksic J, Karagianni P, et al. Cooperativity and rapid evolution of cobound transcription factors in closely related mammals. Cell 2013;154:530–40.

[41] Doniger SW, Fay JC. Frequent gain and loss of functional transcription factor binding sites. PLoS Comput Biol 2007;3:e99.

[42] Man O, Pilpel Y. Differential translation efficiency of orthologous genes is involved in phenotypic divergence of yeast species. Nat Genet 2007;39:415–21.

[43] Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, et al. Global quantification of mammalian gene expression control. Nature 2011;473:337–42.

[44] Whitesides GM, Boncheva M. Beyond molecules: self-assembly of mesoscopic and macroscopic components. Proc Natl Acad Sci U S A 2002;99:4769–74.

[45] McDaniell R, Lee BK, Song L, Liu Z, Boyle AP, Erdos MR, et al. Heritable individual-specific and allele-specific chromatin signatures in humans. Science 2010;328:235–9.