Sequence–structure relationships in yeast mRNAs

Andrey Chursov[1], Mathias C. Walter[2], Thorsten Schmidt[2], Andrei Mironov[3,4], Alexander Shneider[5] and Dmitrij Frishman[1,2,*]

[1]Department of Genome Oriented Bioinformatics, Technische Universität München, Wissenschaftzentrum Weihenstephan, Maximus-von-Imhof-Forum 3, D-85354, Freising, [2]Helmholtz Center Munich – German Research Center for Environmental Health (GmbH), Institute of Bioinformatics and Systems Biology, Ingolstädter Landstraße 1, D-85764 Neuherberg, Germany, [3]Department of Bioengineering and Bioinformatics, Moscow State University, Leninskie Gory, GSP-1, 119991, [4]Institute for Information Transmission Problems, Russian Academy of Sciences, Bolshoi Karetny pereulok 19, 127994, Moscow, Russia and [5]Cure Lab, Inc., 43 Rybury Hillway, Needham, MA 02492, USA

ABSTRACT

It is generally accepted that functionally important RNA structure is more conserved than sequence due to compensatory mutations that may alter the sequence without disrupting the structure. For small RNA molecules sequence–structure relationships are relatively well understood. However, structural bioinformatics of mRNAs is still in its infancy due to a virtual absence of experimental data. This report presents the first quantitative assessment of sequence–structure divergence in the coding regions of mRNA molecules based on recently published transcriptome-wide experimental determination of their base paring patterns. Structural resemblance in paralogous mRNA pairs quickly drops as sequence identity decreases from 100% to 85–90%. Structures of mRNAs sharing sequence identity below roughly 85% are essentially uncorrelated. This outcome is in dramatic contrast to small functional non-coding RNAs where sequence and structure divergence are correlated at very low levels of sequence similarity. The fact that very similar mRNA sequences can have vastly different secondary structures may imply that the particular global shape of base paired elements in coding regions does not play a major role in modulating gene expression and translation efficiency. Apparently, the need to maintain stable three-dimensional structures of encoded proteins places a much higher evolutionary pressure on mRNA sequences than on their RNA structures.

INTRODUCTION

Secondary structure elements both in the untranslated (UTR) and coding (CDS) regions of mRNAs have been implicated in a variety of regulatory functions (1). For example, riboswitches modulate gene expression through conformational changes in response to various stimuli (2). In addition, translation initiation, elongation, termination and translation efficiency all depend on higher order mRNA secondary structures in non-coding regions (3,4). Coding region hairpins have also been suggested to play a role in the regulation of translation (5). The relationship between RNA structure and gene expression in the coding regions of mRNAs has been demonstrated both computationally and experimentally (6–10). In particular, reduced mRNA stability near the start codon has been observed in a wide range of species, probably as a mechanism to facilitate ribosome binding or start codon recognition by initiator tRNA (11). Computational studies show that native mRNA sequences have lower folding energies and hence more stable structure than codon-randomized ones (5). The three mRNA functional domains—5′-UTR, CDS and 3′-UTR—form largely independent folding units, with base pairing across domain borders being rare (12). Evolutionary conserved local secondary structures have been identified in the CDS regions (13,14) and shown to be functional (15).

There is a selective pressure toward maintaining both stable RNA structures of coding regions and the three-dimensional folds of their encoded proteins (16). It has been argued that the redundancy of the genetic code plays an important role in satisfying these selection requirements (12). In general, however, sequence–structure relationships in mRNA-coding regions remain elusive; and, their spatial structure is unknown. While hundreds

---

of atomic resolution structures have been determined for smaller RNA molecules, most notably tRNAs, experimental structures of large RNAs are still rare (17). Until recently, direct experimental determination of mRNA structure has been impossible on a large scale. Furthermore, most insights into the evolutionary constraints acting on them arose from correlating predicted base paring patterns with the effects of site-directed mutagenesis on mRNA expression and degradation, as well as on the expression levels and activity of encoded protein products.

Significant progress has been made in predicting RNA secondary structure from sequence based on free-energy minimization (18), probabilistic models (19) and evolutionary information (20). However, the accuracy of current algorithms is still insufficient to model large molecules, primarily because the number of theoretically possible RNA secondary structures grows exponentially with the length of the sequence (21). Also, the free folding energy of millions of suboptimal structures is very close to the most stable structure. Lowest energy structures may not necessarily reflect folding *in vivo* (22) due to kinetic processes and protein–RNA interactions. Additionally, it is hard to model pseudoknots and unstructured regions (23).

More accurate prediction of RNA secondary structure can be achieved by using experimental constraints obtained from oligonucleotide data to guide free-energy minimization (24). Moreover, experimental methods have been developed that allow comprehensive monitoring of RNA structure at single nucleotide resolution. One such method, fragmentation sequencing, allows for reconstructing RNA structures by sequencing fragments of single-stranded RNA resulting from nuclease digestion. Another method, known as selective 2′-hydroxyl acylation and primer extension (SHAPE) (25), exploits the sensitivity of selective acetylation of the ribose 2′-hydroxyl position to local nucleotide flexibility, thereby allowing identification of those nucleotides that are conformationally constrained by base pairing. Accurate SHAPE-directed RNA structure determination has been reported for several types of RNA molecules, including *Escherichia coli* 16S RNA and yeast tRNA[asp] (26), as well as for the entire HIV-1 genome (27). This latter work highlighted the intricate relationship between RNA sequences and protein structure of the encoded proteins. In particular, it was found that flexible loops in protein structures correspond to highly structured RNA elements, implying a functional role of mRNA structure in the modulation of ribosome processivity at domain boundaries.

In recent work, Kertesz and colleagues (28) reported the first transcriptome-wide experimental analysis of mRNA structures using the novel technology called parallel analysis of RNA structure (PARS). PARS enables the determination of base pairing probabilities at single nucleotide resolution by refolding RNAs *in vivo*, treating them with structure-specific enzymes and then sequencing the resulting fragments. Structural profiles were obtained for more than 3000 transcripts from the budding yeast *Saccharomyces cerevisiae*. The work of Kertesz *et al.* revealed higher degree of structuredness in the mRNA-coding regions compared with the 3′- and 5′-untranslated regions, implying a functional role of RNA structure in coding regions in regulating gene expression. The global data set of PARS profiles represents a true treasure trove for investigating sequence–structure and structure–function relationships in mRNAs.

This report provides the first comprehensive analysis of sequence–structure relationships in the coding regions of yeast mRNAs based on base pairing propensities measured by the PARS technology. It was found that PARS profiles of paralogous mRNAs show very strong, essentially linear, correlation sequence for identity levels upwards of 85–90%. Yet, pairs of more distantly related yeast transcripts secondary structure appear to be unrelated. Interestingly, predicted secondary structures of yeast paralogs display a similar behavior with respect to sequence identity; and, there is a significant correlation between experimental and theoretical structures, as noted previously (28). Theoretical structures of orthologous mRNA pairs from yeast and *Candida glabrata* are also uncorrelated for low sequence identity levels while for highly similar sequences no conclusion could be made due to lack of data.

MATERIALS AND METHODS

Experimental data on yeast mRNA secondary structure

Secondary structure profiles of 3000 transcripts from the budding yeast *S. cerevisiae* have recently been determined using a novel experimental strategy called PARS (28). For each individual nucleotide position of mRNAs, a PARS score reflects its likelihood to be in a double-stranded conformation. PARS scores for yeast transcripts were downloaded from http://genie.weizmann.ac.il/pubs/PARS10. 5′- and 3′-UTR regions were identified by sequence comparison with yeast amino acid sequences, and then excluded from consideration. In the following, a vector of PARS scores for a given transcript is referred to as its experimental structure.

Yeast paralogs

Data on paralogous yeast proteins were kindly provided by Martin Münsterkötter and Ulrich Güldner from the fungal genomics group at the Institute for Bioinformatics and Systems Biology (German Research Center for Environmental Health, Munich). A list of protein pairs sharing significant similarity (identity at the amino acid level >50%) was extracted from the SIMAP database (29). Additionally, the putative paralogs were required to have not >10% difference in sequence length. In total, 243 paralog pairs involving 409 different yeast genes satisfied these conditions.

Amino acid sequences of paralogous yeast proteins were globally aligned using the ggsearch program from the FASTA software suite (30). Amino acid sequence alignments were subsequently converted into mRNA sequence alignments; and, the percent identity between each pair of coding regions was calculated by dividing the number of identical nucleotides by the length of the alignment.

Orthologs from *C. glabrata*

Sequence data for *C. glabrata* were downloaded from the PEDANT genome database (31). A list of orthologous protein pairs between *S. cerevisiae* and *C. glabrata* was extracted from the eggNOG database (32). In total, we obtained 2327 ortholog pairs. The alignment procedure was the same as for paralogs, see above.

PARS score distances between yeast paralogs

To assess global structural similarity between pairs of aligned mRNA sequences, root mean square deviations (RMSDs) between vectors of PARS scores were calculated for all alignment positions that did not contain gaps. Additionally, for each transcript pair, profiles of local structural similarity were obtained by calculating RMSDs between PARS scores in non-gapped alignment positions within a sliding window of varying length, typically between 100 and 1000 nt.

Prediction of mRNA secondary structures

For each nucleotide position of transcript sequences, the theoretical probability to be in double-stranded conformation was calculated using the RNAfold method from the Vienna RNA package (33). As done similarly for experimental PARS scores (see above), RNAfold probability values were used to calculate global and local measures of structural similarity between aligned coding regions of mRNAs based on RMSD. For brevity, a vector of predicted probabilities of RNA bases in double-stranded conformation for a given transcript is further referred to as its theoretical structure.

Data availability

All sequence alignments together with experimentally determined and predicted structures are available in Supplementary Data.

RESULTS

By illustrating the data used in this study on a concrete example, the research results can be readily presented. Two yeast mRNA sequences, YBR092C and YBR093C, share 86.5% sequence identity, and their partial alignment is depicted in the top part of Figure 1. The position-dependent PARS scores for both sequences are shown in the middle part of Figure 1. Both graphs display a rather high degree or correlation, albeit not perfect. In the bottom part of Figure 1, theoretical structures (probabilities for individual bases to be paired) are drawn along the sequence. Figure 2 shows how distances between experimental and theoretical structures of YBR092C and YBR093C vary along the mRNA sequence dependent on sequence identity in a local sequence window. As
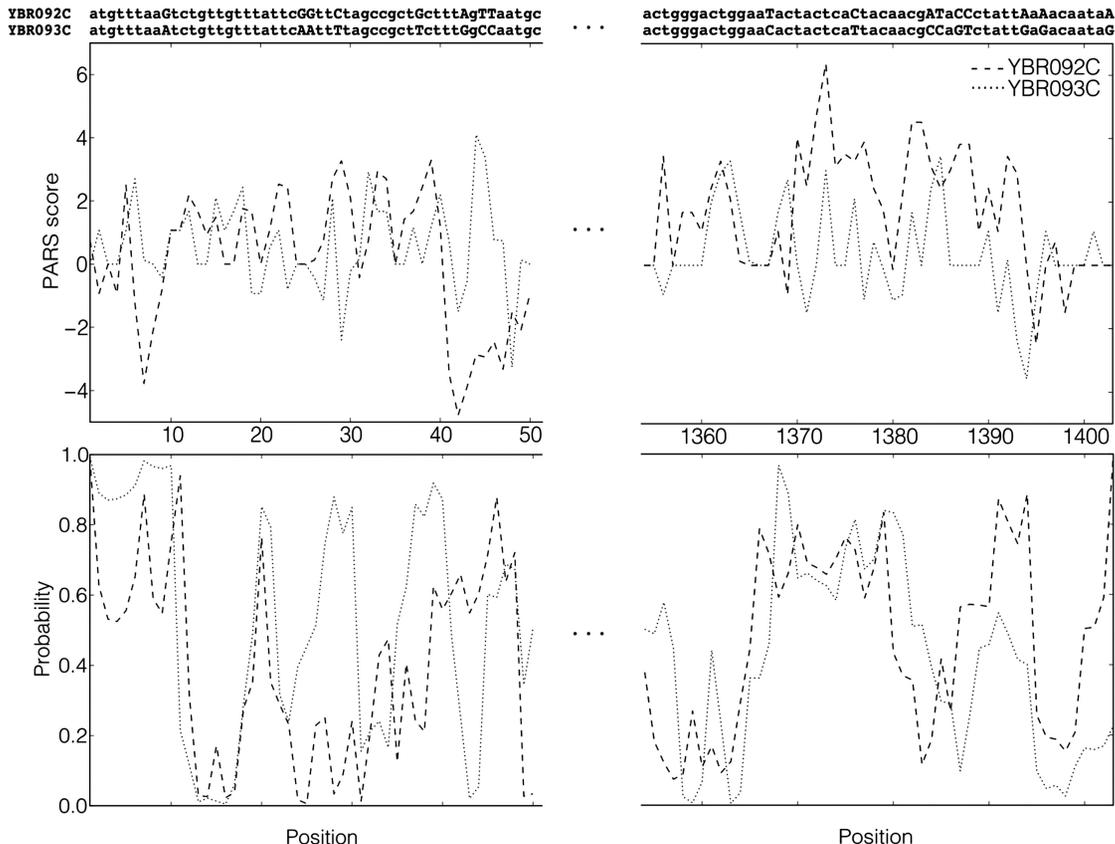

Figure 1. Sequence alignment, experimental and theoretical structures of the first and last 50 nt for the pair of yeast mRNA sequences YBR092C (dashed lines) and YBR093C (dotted lines).

expected, highly similar regions generally correspond to more similar structures.

Calculations exemplified in Figures 1 and 2 were performed for all pairs of paralogous mRNA sequences in our data set. Table 1 summarizes pair-wise correlations between the three evolutionary measures considered in this work for different ranges of sequence identities. Figure 3a shows how the difference between experimental structures depends on sequence similarity. PARS scores appear to be entirely uncorrelated for identity levels of up to ~85–90%. In this sequence identity range, the median RMSD between PARS score vectors does not differ from the median calculated for randomly selected mRNA pairs (dashed horizontal line in Figure 3a). For sequence identity levels over 85–90%, the distance between experimental structures shows essentially a linear dependence from sequence similarity (Supplementary Figure S1).

Upon conducting the same experiment with pairs of theoretical structures of yeast mRNAs, it was found that the distance between the structures also begins to depend on sequence similarity upward of roughly 85–90% identity (Figure 3b). For pairs with identity between sequences within the range from 97.5% to 100%, the median distance between theoretical structures constitutes 38% of the random level. Yet, for experimental structures, it is lower at 29%. The link between sequence and structure is thus stronger when experimental structures are considered. The distance between theoretical structures also shows a linear dependence from sequence similarity for sequence identity levels over 85–90% (Supplementary Figure S2).

Therefore, what is the significance of the sequence–structure dependence shown in Figure 3; and, how would it appear for codon-randomized mRNA sequences? Since experimental PARS scores are not available for randomly generated sequences, this issue could only be assessed for theoretical structures. For each pair of paralogs, one sequence was kept unchanged. In the second mRNA, however, mutations were randomly distributed along the sequence, keeping the encoded amino acid sequence, the codon usage and the total number of mutations between the paralogs unchanged. Overall, the divergence of structures between codon-randomized paralogs displays virtually the same dependence on sequence similarity as for native sequences (Supplementary Figure S3).

We also compared predicted structures between orthologous mRNAs from *S. cerevisiae* and the pathogenic yeast *C. glabrata* (Figure 4). Although *C. glabrata* is the most closely related organism to *S. cerevisiae* with a completely sequenced genome (34), no pair of orthologous mRNAs between these two organisms shares sequence identity >95% and thus no conclusion about structure divergence for very similar sequences could be made. However, for lower identity levels theoretical structures of orthologs are uncorrelated and thus behave the same way as paralogous structures.

DISCUSSION

In some sense, the current situation in RNA bioinformatics is reminiscent of the early days of structural
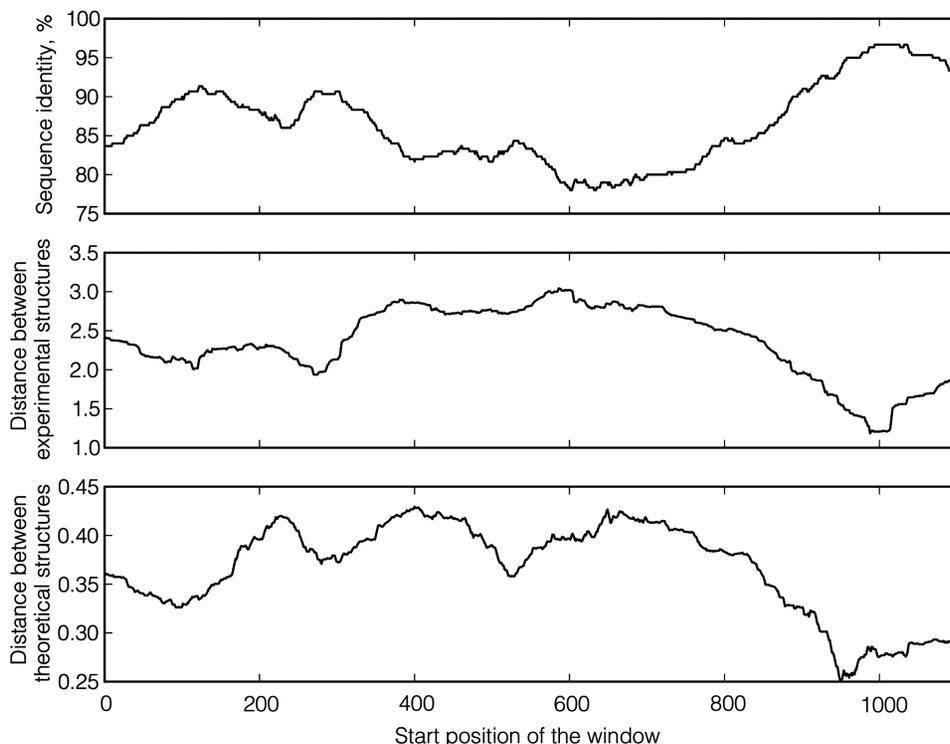

Figure 2. The profile of local structural similarity versus local sequence identity for the pair of yeast mRNA sequences YBR092C and YBR093C. The length of the sliding window is 300. The global sequence identity between these two sequences is 86.5%.

Table 1. Correlation coefficients and *P*-values for different ranges of sequence identity

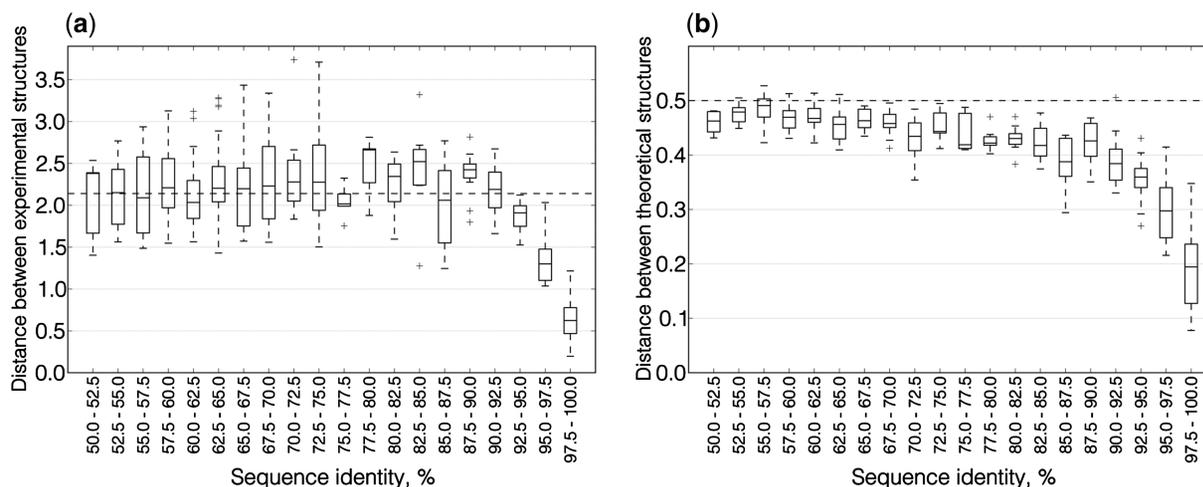| Sequence identity range (%)	Sequence identity versus RMSD between experimental structures		Sequence identity versus RMSD between theoretical structures		RMSD between experimental structures versus RMSD between theoretical structures	
	Correlation coefficient	*P*-value	Correlation coefficient	*P*-value	Correlation coefficient	*P*-value
50–60	0.12	0.39	−0.07	0.62	0.14	0.31
60–70	0.14	0.22	−0.10	0.37	−0.02	0.87
70–80	−0.08	0.67	−0.08	0.67	−0.24	0.21
80–90	0.01	0.91	−0.14	0.40	0.04	0.79
90–100	−0.92	$5.66e^{-27}$	−0.75	$1.24e^{-12}$	0.69	$3.56e^{-10}$

Figure 3. Boxplots of distances between structures of aligned paralogous mRNAs in different ranges of sequence similarity. Each box corresponds to the range of similarity 2.5%. The box extends from the lower to the upper quartile values, with a horizontal line at the median value. Whiskers demonstrate the entire range of the data. Crosses show outliers. (**a**) Distances between experimental structures. The average level of PARS score distances for alignments of random sequence pairs is 2.14 (dashed line). (**b**) Distances between theoretical structures. The average level of probability distance for alignments of random sequence pairs is 0.5 (dashed line).

bioinformatics of proteins, when the availability of a sufficiently large data set of X-ray structures allowed for the first comprehensive analysis of the relation between the divergence of sequence and structure in proteins (35). Until recently, studies of the evolutionary conservation of RNA structures were based on *in silico* predictions and largely limited to non-coding RNA. In the first large-scale study, Schudoma *et al.* (36) determined that in short RNA loops with known three-dimensional structures sequence identity >75% implies significant structural similarity. The most comprehensive investigation of sequence–structure relationships in RNA molecules to date is based on all-against-all pair-wise structural comparison of non-coding RNAs (tRNAs, rRNAs, riboswitches and riboswitches) with known spatial architectures (37). Assessment of evolutionary divergence revealed that the correlation between sequence and secondary structure conservation is highly significant for sequence identity levels in the range between just a few percentage points up to roughly 60% where this relationship saturates. Further increase of sequence similarity (60–100%) does not lead to an appreciable growth of secondary structure similarity. None of the studies mentioned

above considered mRNAs because no mRNA structures are currently known at atomic resolution.

The principal finding of this research is that the correlation between sequence and structure in the coding regions of yeast mRNAs is much weaker than in small non-coding RNAs. Up to ~85–90% sequence identity, the similarity of both experimental and theoretical base pairing propensities between paralogous yeast mRNAs is at random level; while, for more similar sequence pairs, sequence and structure are strongly correlated. This may imply that mRNAs do not experience a strong selective pressure to preserve a certain degree of structuredness. The fact that codon-randomized sequences display a similar behavior also indicates that there is no appreciable evolutionary pressure to preserve a particular RNA structure as long as the encoded protein remains unchanged. Taken together, these results underscore a high degree of evolutionary neutrality in yeast mRNA molecules, both at the level of primary (third codon position) and secondary (extent of base paring) structure.

On one hand, our findings are in strong contrast to many non-coding RNAs and *cis*-acting regulatory elements of mRNAs whose biological function is primarily
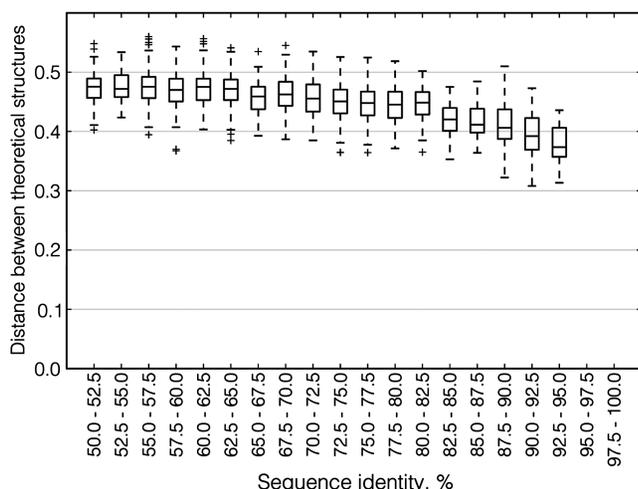
Figure 4. Boxplot of distances between theoretical structures of aligned orthologous mRNAs in different ranges of sequence similarity. Notation as in Figure 3.

mediated by their spatial architecture (38) stabilized by tertiary interactions, modified bases and interactions with proteins and small ligands. On the other hand, sequence–structure relationships observed in this work are compatible with the notion that, in general, RNA molecules do not have a single global structure. Instead, they exist as a highly dynamic ensemble of alternative conformations (39,40) that are often capable of performing different functions (41). The extent of base pairing may play a role in the regulation of pre-mRNA splicing, translation and mRNA degradation. Both experimentally determined PARS scores and computationally derived partition functions analyzed in this work are statistical measures that reflect the propensity of each nucleotide to form a base pair across a large number of metastable structures.

This analysis has several important limitations. First, PARS probes RNA structures *in vitro* rather than in the living cell and may not always reproduce functional RNA structures (42). Second, even if the base paring information obtained by the PARS technology were perfectly correct, it still merely represents a one-dimensional profile of structural propensities, a far cry from knowing the actual RNA secondary structure, let alone spatial architecture, for each individual molecule at any moment of time. Third, the findings do not rule out much stronger sequence–structure correlations in certain local structural elements of coding regions, such as reprogrammed genetic-decoding signals (43) or mRNA localization signals. We also cannot rule out the possibility that the degree of mRNA structuredness does have an important functional role in spite of quick erosion of structural similarity between paralogs with diminishing sequence similarity, and that this erosion reflects functional differentiation. However, we consider such explanation unlikely because the same behavior is observed between orthologous mRNAs. Finally, only a small subset of the PARS data constituted by pairs of sequence similar yeast mRNAs (paralogs) was explored. As a next step, it will be exciting

to conduct comparative analyses of mRNA structuromes [the term coined by Westhof and Romby (44)], focusing on orthologous sequences from multiple organisms and taking into account important genomic variables, such as expression level and evolutionary rate. Given the current pace of high-throughput RNA analysis technologies there is no doubt that such data will become available in the near future.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary figures S1–S3.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Bevilacqua,P.C. and Blose,J.M. (2008) Structures, kinetics, thermodynamics, and biological functions of RNA hairpins. *Annu. Rev. Phys. Chem.*, **59**, 79–103.
2. Serganov,A. and Patel,D.J. (2007) Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat. Rev. Genet.*, **8**, 776–790.
3. Gray,N.K. and Hentze,M.W. (1994) Regulation of protein synthesis by mRNA structure. *Mol. Biol. Rep.*, **19**, 195–200.
4. Kozak,M. (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene*, **361**, 13–37.
5. Katz,L. and Burge,C.B. (2003) Widespread selection for local RNA secondary structure in coding regions of bacterial genes. *Genome Res.*, **13**, 2042–2051.
6. Kudla,G., Murray,A.W., Tollervey,D. and Plotkin,J.B. (2009) Coding-sequence determinants of gene expression in Escherichia coli. *Science*, **324**, 255–258.
7. Duan,J., Wainwright,M.S., Comeron,J.M., Saitou,N., Sanders,A.R., Gelernter,J. and Gejman,P.V. (2003) Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. *Hum. Mol. Genet.*, **12**, 205–216.
8. Ilyinskii,P.O., Schmidt,T., Lukashev,D., Meriin,A.B., Thoidis,G., Frishman,D. and Shneider,A.M. (2009) Importance of mRNA secondary structural elements for the expression of influenza virus genes. *OMICS*, **13**, 421–430.
9. Carlini,D.B., Chen,Y. and Stephan,W. (2001) The relationship between third-codon position nucleotide content, codon bias, mRNA secondary structure and gene expression in the drosophilid alcohol dehydrogenase genes Adh and Adhr. *Genetics*, **159**, 623–633.
10. Nackley,A.G., Shabalina,S.A., Tchivileva,I.E., Satterfield,K., Korchynskyi,O., Makarov,S.S., Maixner,W. and Diatchenko,L. (2006) Human catechol-O-methyltransferase haplotypes modulate

protein expression by altering mRNA secondary structure. *Science*, **314**, 1930–1933.

11. Gu,W., Zhou,T. and Wilke,C.O. (2010) A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. *PLoS Comput. Biol.*, **6**, e1000664.

12. Shabalina,S.A., Ogurtsov,A.Y. and Spiridonov,N.A. (2006) A periodic pattern of mRNA secondary structure created by the genetic code. *Nucleic Acids Res.*, **34**, 2428–2437.

13. Meyer,I.M. and Miklós,I. (2005) Statistical evidence for conserved, local secondary structure in the coding regions of eukaryotic mRNAs and pre-mRNAs. *Nucleic Acids Res.*, **33**, 6338–6348.

14. Steigele,S., Huber,W., Stocsits,C., Stadler,P.F. and Nieselt,K. (2007) Comparative analysis of structured RNAs in S. cerevisiae indicates a multitude of different functions. *BMC Biol.*, **5**, 25.

15. Olivier,C., Poirier,G., Gendron,P., Boisgontier,A., Major,F. and Chartrand,P. (2005) Identification of a conserved RNA motif essential for She2p recognition and mRNA localization to the yeast bud. *Mol. Cell. Biol.*, **25**, 4752–4766.

16. White,H.B. III, Laux,B.E. and Dennis,D. (1972) Messenger RNA structure: compatibility of hairpin loops with protein sequence. *Science*, **175**, 1264–1266.

17. Holbrook,S.R. (2008) Structural principles from large RNAs. *Annu. Rev. Biophys.*, **37**, 445–464.

18. Mathews,D.H. and Turner,D.H. (2006) Prediction of RNA secondary structure by free energy minimization. *Curr. Opin. Struct. Biol.*, **16**, 270–278.

19. Dowell,R.D. and Eddy,S.R. (2006) Efficient pairwise RNA structure prediction and alignment using sequence alignment constraints. *BMC Bioinformatics*, **7**, 400.

20. Bernhart,S.H. and Hofacker,I.L. (2009) From consensus structure prediction to RNA gene finding. *Brief. Funct. Genomic Proteomic*, **8**, 461–471.

21. Meyer,I.M. and Miklós,I. (2007) SimulFold: simultaneously inferring RNA structures including pseudoknots, alignments, and trees using a Bayesian MCMC framework. *PLoS Comput. Biol.*, **3**, e149.

22. Mahen,E.M., Watson,P.Y., Cottrell,J.W. and Fedor,M.J. (2010) mRNA secondary structures fold sequentially but exchange rapidly in vivo. *PLoS Biol.*, **8**, e1000307.

23. Reeder,J. and Giegerich,R. (2004) Design, implementation and evaluation of a practical pseudoknot folding algorithm based on thermodynamics. *BMC Bioinformatics*, **5**, 104.

24. Duan,S., Mathews,D.H. and Turner,D.H. (2006) Interpreting oligonucleotide microarray data to determine RNA secondary structure: application to the 3′ end of Bombyx mori R2 RNA. *Biochemistry*, **45**, 9819–9832.

25. Merino,E.J., Wilkinson,K.A., Coughlan,J.L. and Weeks,K.M. (2005) RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.*, **127**, 4223–4231.

26. Low,J.T. and Weeks,K.M. (2010) SHAPE-directed RNA secondary structure prediction. *Methods*, **52**, 150–158.

27. Watts,J.M., Dang,K.K., Gorelick,R.J., Leonard,C.W., Bess,J.W. Jr, Swanstrom,R., Burch,C.L. and Weeks,K.M. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature*, **460**, 711–716.

28. Kertesz,M., Wan,Y., Mazor,E., Rinn,J.L., Nutter,R.C., Chang,H.Y. and Segal,E. (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature*, **467**, 103–107.

29. Rattei,T., Tischler,P., Götz,S., Jehl,M.A., Hoser,J., Arnold,R., Conesa,A. and Mewes,H.W. (2010) SIMAP–a comprehensive database of pre-calculated protein sequence similarities, domains, annotations and clusters. *Nucleic Acids Res.*, **38**, D223–D226.

30. Pearson,W.R. (1990) Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol.*, **183**, 63–98.

31. Walter,M.C., Rattei,T., Arnold,R., Güldener,U., Münsterkötter,M., Nenova,K., Kastenmüller,G., Tischler,P., Wölling,A., Volz,A. *et al.* (2009) PEDANT covers all complete RefSeq genomes. *Nucleic Acids Res.*, **37**, D408–D411.

32. Muller,J., Szklarczyk,D., Julien,P., Letunic,I., Roth,A., Kuhn,M., Powell,S., von Mering,C., Doerks,T., Jensen,L.J. *et al.* (2010) eggNOG v2.0: extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups, species and functional annotations. *Nucleic Acids Res.*, **38**, D190–D195.

33. Gruber,A.R., Lorenz,R., Bernhart,S.H., Neuböck,R. and Hofacker,I.L. (2008) The Vienna RNA websuite. *Nucleic Acids Res.*, **36**, W70–W74.

34. Dujon,B. (2010) Yeast evolutionary genomics. *Nat. Rev. Genet.*, **11**, 512–524.

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

36. Schudoma,C., May,P., Nikiforova,V. and Walther,D. (2010) Sequence-structure relationships in RNA loops: establishing the basis for loop homology modeling. *Nucleic Acids Res.*, **38**, 970–980.

37. Capriotti,E. and Marti-Renom,M.A. (2010) Quantifying the relationship between sequence and three-dimensional structure conservation in RNA. *BMC Bioinformatics*, **11**, 322.

38. Gruber,A.R., Bernhart,S.H., Hofacker,I.L. and Washietl,S. (2008) Strategies for measuring evolutionary conservation of RNA secondary structures. *BMC Bioinformatics*, **9**, 122.

39. Mironov,A.A., Dyakonova,L.P. and Kister,A.E. (1985) A kinetic approach to the prediction of RNA secondary structures. *J. Biomol. Struct. Dyn.*, **2**, 953–962.

40. Danilova,L.V., Pervouchine,D.D., Favorov,A.V. and Mironov,A.A. (2006) RNAKinetics: a web server that models secondary structure kinetics of an elongating RNA. *J. Bioinform. Comput. Biol.*, **4**, 589–596.

41. Zhao,P., Zhang,W.B. and Chen,S.J. (2010) Predicting secondary structural folding kinetics for nucleic acids. *Biophys. J.*, **98**, 1617–1625.

42. Mauger,D.M. and Weeks,K.M. (2010) Toward global RNA structure analysis. *Nat. Biotechnol.*, **28**, 1178–1179.

43. Namy,O., Rousset,J.P., Napthine,S. and Brierley,I. (2004) Reprogrammed genetic decoding in cellular gene expression. *Mol. Cell*, **13**, 157–168.

44. Westhof,E. and Romby,P. (2010) The RNA structurome: high-throughput probing. *Nat. Methods*, **7**, 965–967.