Melting temperature highlights functionally important RNA structure and sequence elements in yeast mRNA coding regions

Fei Qi¹ and Dmitrij Frishman¹,²,*

¹Department of Bioinformatics, Technische Universität München, Wissenschaftszentrum Weihenstephan, Maximus-von-Imhof-Forum 3, D-85354 Freising, Germany and ²St Petersburg State Polytechnic University, St Petersburg 195251, Russia

Received December 20, 2016; Revised February 04, 2017; Editorial Decision February 24, 2017; Accepted February 24, 2017

ABSTRACT

Secondary structure elements in the coding regions of mRNAs play an important role in gene expression and regulation, but distinguishing functional from non-functional structures remains challenging. Here we investigate the dependence of sequence–structure relationships in the coding regions on temperature based on the recent PARTE data by Wan et al. Our main finding is that the regions with high and low thermostability (high \( T_m \) and low \( T_m \) regions) are under evolutionary pressure to preserve RNA secondary structure and primary sequence, respectively. Sequences of low \( T_m \) regions display a higher degree of evolutionary conservation compared to high \( T_m \) regions. Low \( T_m \) regions are under strong synonymous constraint, while high \( T_m \) regions are not. These findings imply that high \( T_m \) regions contain thermo-stable functionally important RNA structures, which impose relaxed evolutionary constraint on sequence as long as the base-pairing patterns remain intact. By contrast, low thermostability regions contain single-stranded functionally important conserved RNA sequence elements accessible for binding by other molecules. We also find that theoretically predicted structures of paralogous mRNA pairs become more similar with growing temperature, while experimentally measured structures tend to diverge, which implies that the melting pathways of RNA structures cannot be fully captured by current computational approaches.

INTRODUCTION

Secondary structure elements and global folding patterns play a fundamental role in the function and regulation of RNAs (1–6). For a number of years most of the attention was given to functional secondary structures of non-coding RNAs (ncRNAs), but more recently mRNA structure also moved to the spotlight of genomics and bioinformatics research. mRNA structures have been found to influence multiple stages of gene expression and protein synthesis, including transcription, splicing, RNA transport, translation initiation, elongation and termination, as well as RNA degradation (7–13). Secondary structures of the three functional mRNA domains—5′ UTR, the coding region and 3′ UTR—are largely independent, since base pairs across domain borders are rare (14). A broad variety of functional structural elements were described in UTRs (15–18). While the primary function of the coding regions is to encode amino acid sequences of proteins, they are presumed to contain even more RNA secondary structures than UTRs (1), with some of them already proven to be functional (19–24). The redundancy of the genetic code makes it possible for the coding regions to carry overlapping functions, which manifest themselves at the level of protein and RNA sequences and structures (14,25,26).

In recent years, several experimental approaches have been developed for genome-wide measurement of RNA structures. These methods, including Frag-seq (fragmentation sequencing) (27), PARS (parallel analysis of RNA structure) (1) and SHAPE-seq (selective 2′-hydroxyl acylation analyzed by primer extension sequencing) (28) combine RNA structure probing by structure-specific enzymes and chemical modifications at double- or single-stranded bases with high-throughput next generation sequencing. These approaches can probe millions of molecules at single nucleotide resolution within one experiment and therefore enable comprehensive studies of RNA structures (29). An important question, which arises in this context, is to which extent the results of high-throughput structure probing experiments are reproducible and compatible with each other. Owing to the availability of RNA structure probing data many of the classical problems in molecular evolution, which have been extensively addressed for protein molecules, can now be examined for mRNAs as well. In
particular, it is of great interest to investigate to which extent secondary and/or tertiary structure of mRNAs constrains sequence variation and how strongly mRNA structures are conserved in evolution, as was done for protein 3D structures long ago (30). Recently, Wan et al. published PARTE (Parallel Analysis of RNA structures with Temperature Elevation) experiment, in which secondary structures of yeast RNAs were probed and melting temperatures ($T_m$) were derived at single nucleotide resolution at five temperatures (13). Using these data, we demonstrate that high and low thermostability regions in the mRNA coding regions highlight functionally important RNA structures and sequence segments, respectively. We report a surprising pattern of structural divergence between sequence-similar mRNAs along the temperature ladder, which cannot be captured by the currently available computational approaches. There is a considerable reproducibility between the high-throughput RNA structure probing experiments, PARS and PARTE.

**MATERIALS AND METHODS**

**Experimental data on secondary structures of yeast mRNAs**

Secondary structure profiles of 3002 yeast mRNAs determined at room temperature by PARS (Parallel Analysis of RNA Structures (1)) were downloaded from http://genie.weizmann.ac.il/pubs/PARS10. For each individual nucleotide position of mRNAs a PARS score reflects its likelihood to be in a double-stranded conformation based on the number of sequencing reads upon treatment by two structure-specific enzymes, RNase V1 and nuclease S1, which cleave at double-stranded and single-stranded regions, respectively. A total of 4 405 020 bases in the 3002 mRNAs are covered by PARS scores. For a given mRNA sequence, the vector of its PARS scores is referred to as its PARS structure.

Another mRNA structure dataset used in this work was obtained by a PARTE experiment, in which 4562 yeast mRNAs were structure-probed by RNase V1 at five temperatures (23, 30, 37, 55 and 75°C; two biological replicates were performed for each temperature) (13). PARTE reveals $T_m$ of each base, with double-stranded regions being progressively eliminated as temperature increases. V1 reads resulting from this experiment were downloaded from the GEO database (31) (GSE39680) and their counts were normalized exactly as described by Wan et al. (13): (i) for each library peaks were defined as those bases that are covered by more reads than the bases on their left and their right and whose read coverage is greater than the average coverage of bases on the same gene and the average coverage of all bases; (ii) using the PoissonSeq algorithm (32), the library size of each sequencing lane was estimated based on the high confidence peaks observed in both duplicate libraries at at least one of the five temperatures; and (iii) V1 read numbers were normalized by dividing the counts in each sample by the corresponding library size. For each RNA nucleotide position the log$_2$ value of the mean of the two normalized V1 read numbers from two duplicate samples was treated as its PARTE score (each mean V1 read number was augmented by 0.001 to avoid the undefined logarithm values for those bases where the read count is zero). A total of 7 497 468 bases in the 4562 mRNAs are covered by PARTE scores. For a given mRNA sequence, vectors of its PARTE scores are referred to as its PARTE structure.

**Predicted secondary structures of yeast mRNAs**

Sequences of 6686 yeast mRNAs were downloaded from SGD (release 57-1-1) (33). Base pairing probabilities for each yeast mRNA were calculated using the RNAfold algorithm from the ViennaRNA package (34). In order to simulate the PARTE experiment, which was carried out at five different temperatures, RNAfold was run five times for each yeast mRNA using five different values of the -T parameter (23,30,35,55,75). For a given mRNA, vectors of its theoretically predicted base pairing probabilities are referred to as its predicted structures.

**Yeast paralogous mRNAs**

We considered 246 pairs of aligned mRNA sequences of yeast paralogs, as well as the percent identity between aligned coding regions of each pair, as described previously (35). Each of these pairs of paralogous proteins shares over 50% amino acid sequence identity and <10% difference in sequence length.

**Distances between secondary structures of yeast paralogs**

For a given pair of aligned mRNA sequences, we employed root mean square deviation (RMSD) as the measure of the distance between their experimental structures. RMSD values were calculated between the vectors of PARS or PARTE scores for all aligned positions without gaps. Sequence positions not probed in the experiments (i.e. those with read number 0) were also taken into account in this calculation—their PARS score equals 0 and their PARTE score is log$_2$(0 + 0.001). Similarly, distances between predicted structures for a given pair of aligned mRNAs were calculated as RMSD between vectors of predicted base pairing probabilities.

**Paired and unpaired bases of yeast mRNAs**

We subdivided the bases of yeast mRNAs into two classes according to their PARS scores: (i) paired bases (PARS score $\geq 0$) and (ii) unpaired bases (PARS score $<0$). In all the mRNAs covered by both PARS and PARTE experiments we identified 3 514 124 paired bases and 884 030 unpaired bases.

**Melting temperatures of RNA structures in yeast mRNAs**

Data on $T_m$ of RNA structures covering over 320 000 bases in yeast mRNAs were kindly provided by Yue Wan and Howard Y. Chang from the Howard Hughes Medical Institute and the Program in Epithelial Biology at Stanford University School of Medicine. For each RNA sequence position the $T_m$ value was calculated as the mean of the two temperatures between which the position transitioned from the double-stranded to the single-stranded state in the PARTE experiment, i.e. 26.5, 33.5, 46 and 65°C for the transitions...
between 23 and 30°C, 30 and 37°C, 37 and 55°C, and 55 and 75°C, respectively. Positions that remained double-stranded at 75°C were assigned the Tₘ value of 80°C (13). We only considered 1262 mRNAs that have at least 5% of bases with probed melting temperatures.

**Regions with high or low Tₘ in pairs of paralogous yeast mRNAs**

We applied a sliding window of 100 nt with a step size of 10 bases to the alignments of paralogous yeast mRNAs and calculated average Tₘ values for each sequence in a window. Only those windows in which the alignment had <10% of gaps and the two aligned sequences both had at least 10% of bases with probed Tₘ were included in the analysis. We defined two classes of windows—high Tₘ windows and low Tₘ windows—dependent on whether the two aligned sequences in a window both had Tₘ values among the top or bottom 25% of all Tₘ values. Windows with the middle range of Tₘ values were excluded from further analysis. For each window, the Tₘ values of the two aligned sequences were calculated as the arithmetic mean of all the Tₘ values of their individual bases. Overlapping windows belonging to the same class (high or low Tₘ) were merged, yielding the total of 167 regions with high Tₘ and 96 regions with low Tₘ among the 246 pairs of paralogous mRNAs.

**Thermo-stable and meltable positions in pairs of yeast paralogous mRNAs**

For each pair of yeast paralogous mRNAs, a position in the alignment was defined as thermo-stable or meltable dependent on whether both aligned bases in this position had Tₘ ≥ 65°C or Tₘ ≤ 46°C, respectively. This procedure yielded 1323 thermo-stable and 256413 meltable positions, out of which 734 thermo-stable and 20790 meltable positions were located in high Tₘ regions while low Tₘ regions contained 22765 meltable positions and no thermo-stable positions.

**Synonymous base substitutions between yeast paralogs**

The proportion of synonymous (pS) differences between yeast paralogs was estimated by the equation pS = Sd/S, where Sd is the number of observed synonymous substitutions and S is the number of potential synonymous substitutions. In this work, we employed the SNAP software (36) to calculate the pS. We compared the pS value of each high/low Tₘ region (pSregion) with the pS value of the entire alignment of yeast paralogous mRNAs (pSalignment) by calculating ΔpS = pSregion − pSalignment.

**Zipcodes in yeast mRNAs**

Positions of 12 functional motifs in yeast mRNAs responsible for binding with mRNA transport proteins—the so-called zipcodes—were obtained from the study of Jambhekar et al. (37). Five of these zipcodes (ASH1-E1min, TPO1N, ERG2N, WSC2C and SRL1C), which are located in the coding regions and covered by the PARTE melting temperature data, were included in further analysis.

**RESULTS**

**Correlation between PARS and PARTE scores**

Reproducibility of results is a crucial aspect in the evaluation of experiment strategies. To assess the correlation between the PARTE and PARS data, we computed Spearman’s rank correlation coefficients between PARS and PARTE scores over all 4,398,154 bases in those 2,995 mRNA sequences that are contained both in the PARS and in the PARTE datasets. The correlation coefficients were relatively low (0.325, 0.323, 0.321, 0.312 and 0.250 at the PARTE temperatures of 23, 30, 37, 55 and 75°C, respectively) but highly significant (all P-values < 2.2e-16). As expected, the highest correlation was detected at 23°C as it is closest to the room temperature at which the PARS experiment was carried out. The lower correlation at high temperatures can be explained by progressive unfolding of RNA structures.

Correlation between PARS and PARTE scores is not surprising given the similarity of these two experimental strategies (1,13). PARS and PARTE scores both reflect the likelihood of individual bases to be in a double-stranded conformation (1,13). The relatively low correlation coefficients are due to a key difference between the PARS and the PARTE experiments—the enzymes used to detect RNA structures. While PARS relies both on RNase V1 and nuclease S1 to probe the bases in a double- and single-stranded conformation, respectively, PARTE only probes bases in a double-stranded conformation by RNase V1 (1,13). Therefore, while PARS can capture the likelihood of bases to be in a single-stranded conformation based on reads stemming from the nuclease S1, the PARTE experiment does not deliver this information. Indeed, the correlation between PARTE and PARS scores was much stronger (correlation coefficient 0.587, P-value < 2.2e-16) when only bases in double-stranded conformation were considered (data not shown).

**Dependence of structure divergence on sequence identity**

In our previous work, we explored sequence–structure relationships in yeast mRNAs based on PARS data (35). Upon comparing secondary structures between sequence-similar paralogous yeast mRNAs we found that coding regions of mRNAs are not under strong evolutionary pressure to preserve a particular global shape, which implies that global secondary structure of the coding regions does not play a major role in gene regulation. The recent availability of PARTE data for yeast mRNAs (13) has made it possible to investigate sequence–structure divergence at different temperature levels. As seen in Supplementary Figure S1, at all five temperatures the similarity of PARTE structures shows no correlation with the sequence similarity in the range of sequence identity between 50% (the lowest level considered) and roughly 85–90%. In this range the distance between experimental structures of paralogous mRNAs displays a near linear dependence on sequence identity (Supplementary Figure S2 and Table S1).
This finding is in line with our previous analysis of structure probing data obtained by the PARS method (1), in which we found that the global structural conformation of the coding regions is not crucial for gene expression and regulation. This result is compatible with the notion that mRNA conformation depends on interactions with the solvent as well as with proteins and other ligands and that mRNAs adopt a highly dynamic ensemble of conformations instead of a single global structure (7). An important insight provided by our analysis is that interrogation of mRNA structures by PARS and PARTE leads to qualitatively similar evolutionary conclusions, indicating the reproducibility of the high-throughput RNA structure probing experiments.

**Figure 1.** Variation of the distance between secondary structures of paralogous mRNA pairs along the temperature ladder. Points are the median levels of the distance at each temperature. (A) Distance between PARTE structures. (B) Distance between predicted structures.

The availability of the PARTE structure-probing data opens up the possibility to investigate how the distance between secondary structures of similar RNA molecules varies with temperature and to obtain clues about the RNA structure unfolding pathways during the melting process. We therefore calculated structural distances between paralogous yeast mRNAs along the temperature ladder. Intuitively, one would expect the structural distance to be inversely proportional to temperature: as the temperature grows, more and more base pairs melt and an ever increasing portion of both molecules becomes single-stranded and thus more similar to each other. However, the experimentally determined PARTE structures show a strikingly different behavior. The distances between randomly selected and all paralogous mRNA pairs do not appear to vary with temperature at all, while the distances between highly similar paralogous mRNA pairs actually become larger at higher temperatures (Figure 1A). We speculate that this surprising pattern may, to some extent, be due to the limitations of the experimental approach. First, the PARTE experiment probes the *in vitro* re-folded RNA structures rather than *in vivo* structures (13). Second, as noted by Wan et al., in the PARTE data 20% of the bases show a transition for increased V1 reads at higher temperatures, which may indicate that a considerable proportion of thermo-stable RNA secondary structures became accessible to RNase V1 only upon dissolution of tertiary structures (13). This implies that the differences between these structures were only detected at higher temperatures. We also cannot rule out the possibility that the RNA unfolding pathways during the melting process are actually quite different even between similar molecules, presumably due to complex tertiary interactions and dynamic effects.

This unexpected trend could not be captured by RNAfold predictions. As seen in Figure 1B, the distances between the predicted structures behave exactly as intuitively expected. The distances between the predicted structures of randomly selected, all paralogous and highly similar paralogous mRNA pairs all become smaller as the temperature grows from 23 to 75°C. The same pattern was also obtained with the RNAfold program, which computes local base pair probabilities (Supplementary Figure S3). This may be due to a number of inherent limitations of the current computational structure prediction approaches, especially when applied to long RNA sequences and at large deviations in temperature from standard conditions. Exponential growth of the number of possible secondary structures with the sequence length necessitates the introduction of approximations into the folding algorithms (38). Modeling pseudoknots and prediction of long-range interactions continue to be an unsolved problem (39). As well, energy calculations are parametrized at 37°C and become less reliable at other temperatures (40,41).

**Melting temperature highlights functionally important structure and sequence elements in the coding regions of mRNAs**

As discussed above, the global secondary structure of the mRNA coding regions is poorly conserved in evolution and probably does not play a role in gene regulation. Instead, RNA structure is more likely to be functional at the level of local structural elements situated in the coding regions. However, it is very hard to distinguish functionally important structural elements from non-functional ones since every RNA chain tends to fold back on itself for thermodynamic reasons (13). One important and experimentally measurable feature that may be indicative of functionality is the thermostability of RNA structures. It has been demon-
strated that in ncRNAs functionally important structures have more stable structures than random RNAs of the same length and dinucleotide frequency (42,43). Many known functional structured RNA regulatory elements were identified in yeast mRNA 3' UTRs by locating thermostable base pairs (13). It is therefore conceivable that functionally important structural elements in the coding regions of mRNAs could also be discriminated by their thermostability.

Locally stable structures can be gleaned from the genome-wide PARTE experiment, in which secondary structures of yeast RNAs were probed and $T_m$ were derived at single nucleotide resolution at five temperatures (13). However, proving that such local structures actually fulfill a biological function is a challenging task. One approach to this problem could be based on assessing RNA-level selective constraints acting on protein-coding regions, including synonymous constraint and compensatory mutations. These unique patterns of sequence–structure relationships are a hallmark of the functionally important RNA elements in the coding regions.

We identified 167 high $T_m$ regions and 96 low $T_m$ regions in 246 pairs of paralogous mRNAs. As seen in Figure 2, high $T_m$ regions show a much stronger sequence–structure relationship than the low $T_m$ regions in paralogous mRNA pairs. The distance between the structures of high $T_m$ regions depends linearly on their sequence similarity for the sequence identity levels over 80%, while in low $T_m$ regions this dependence only becomes apparent for sequence identities over 90% (correlation coefficient $-0.69$, $P$-value = 2.1e-11). Linear regression for each 10% range of sequence identity is shown by a dashed line with the corresponding color. PARTE structures at 23°C were used.

Low $T_m$ regions are under synonymous constraint while high $T_m$ regions exhibit relaxed sequence constraint

It is currently believed that RNA-level functions in coding regions manifest themselves by synonymous constraint (25,26). We therefore compared the synonymous substitution rate ($pS$) in the high and low $T_m$ regions with the $pS$ values calculated over the entire alignment of yeast paralogous mRNAs. Most low $T_m$ regions exhibit negative $\Delta pS$ values while most high $T_m$ regions exhibit positive $\Delta pS$ values (chi-squared test, $P$-values $< 0.01$) (Figure 6), which indicates that the low $T_m$ regions are under synonymous constraint and may harbor functionally important nucleotide sequence motifs, such as ncRNA and protein binding sites (25,26,47). This notion is compatible with the complete absence of thermo-stable nucleotide base pairs in low $T_m$ regions, ensuring good accessibility of binding sites. High concentration of synonymous substitutions in high $T_m$ regions may points to relaxed RNA sequence constraint, which may provide an evolutionary advantage for these regions in terms of accommodating functionally important RNA secondary structure elements (6,25).
Figure 3. Low \( T_m \) regions in paralogous mRNA pairs are more conserved in sequence while high \( T_m \) regions are more conserved in RNA secondary structure. (A) Sequence identity between mRNAs (Mann–Whitney–Wilcoxon test, \( P \)-value = 1.9e-21). (B) Distance between RNA secondary structures (PARTE structures at 23° C; Mann–Whitney–Wilcoxon test, \( P \)-value = 4.6e-3). The differences are significant according to Mann–Whitney–Wilcoxon test. Error bars indicate standard error. The investigation of structural distances was effected upon controlling for sequence identity. Only regions with sequence identity 85–95% were considered in this analysis, while the regions with sequence identity < 85%, for which the distance between structures does not differ from randomly selected mRNA pairs as well as the regions with sequence identity > 95%, among which almost no high \( T_m \) regions exists, were excluded from consideration.

Figure 4. Schematic illustration of the conservation levels of high/low \( T_m \) regions and thermo-stable/meltable positions. The alignment of two mRNAs is shown on the top. The low \( T_m \) region displays a higher sequence identity than the high \( T_m \) region (92 versus 60%). When all thermo-stable and meltable positions are considered together, the thermo-stable positions show a higher conservation level than the meltable positions (75 versus 71.9%). When the thermo-stable and the meltable positions located in high \( T_m \) and low \( T_m \) regions are considered separately, the meltable positions in the low \( T_m \) region are most conserved (90%), followed by the thermo-stable positions in the high \( T_m \) region (80%), while the meltable positions in the high \( T_m \) region are least conserved (40%).

Functionally important structure elements in the coding regions of yeast mRNAs tend to be thermostable

A typical class of functionally important structure elements in yeast mRNAs is constituted by the so called zipcodes—regions of mRNAs recognized by the RNA-binding protein She2p (5,37,48). Localized mRNAs are transported to the bud tip of the daughter cell by the She protein complex depending on the interaction between She2p and the loop-stem-loop structure of the zipcode (5,37,48). Out of the 12 functional zipcodes in yeast mRNAs identified in a previous study (37), 5 zipcodes (ASH1-E1min, TPO1N, ERG2N, WSC2C and SRL1C) are located in the coding regions and covered by the PARTE melting temperature data. These 5 zipcodes range from 49 (ASH1-E1min) to 178 (SRL1C) nucleotides in length. Another functionally important structure element in yeast mRNA coding regions is the URE2 IRES (internal ribosome entry site) element, which locates between nucleotides 205 and 309 in the URE2 coding region and folds into a stem-loop structure (49). This IRES element mediates the cap-independent internal initiation of translation resulting in the expression of an N-terminal truncated form of the Ure2p protein (50). We calculated the \( T_m \) of each structure element by averaging the \( T_m \) values of every PARTE-probed base within the element. All six structure el-
Figure 5. Conservation levels of thermo-stable and meltable positions in high/low Tm regions. Positions in high Tm and low Tm regions are examined separately. In high Tm regions the thermo-stable positions exhibit higher sequence conservation than meltable positions (Z-test for two proportions, P-value = 1.3e-6), while in low Tm regions meltable positions display a very high conservation level (Z-test for two proportions, P-value = 2.3e-37) and thermo-stable position are completely absent.

Figure 6. Most low Tm regions exhibit negative ΔpS values while most high Tm regions exhibit positive ΔpS values. The differences are significant according to chi-squared test. Error bars indicate standard error.

DISCUSSION

Figure 8 summarizes our findings about high/low Tm regions as well as our inference based on these findings. High Tm regions exhibit a stronger sequence-structure relationship, conserved and thermo-stable RNA secondary structures and relatively divergent nucleotide sequences, while low Tm regions display a weaker sequence-structure relationship, divergent and less thermostable RNA secondary structures and highly conserved nucleotide sequences. These findings suggest that high Tm regions are under high evolutionary pressure to preserve RNA secondary structure, whereas low Tm regions are under high evolutionary pressure to preserve primary RNA sequence. We therefore hypothesize that high Tm regions may contain thermostable functionally important RNA structure elements and thus experience relatively high evolutionary pressure to preserve the RNA structure and a relaxed evolutionary constraint on the nucleotide sequence, as long as the thermo-stable nucleotide base pairs which are crucial for the RNA structure remain intact. Considering the highly conserved nucleotide sequence and low thermostability of low Tm regions, we hypothesize that low Tm regions may contain functionally important RNA sequence elements, for example, binding sites which are conserved in sequence and require a good accessibility to interact with ligands. High and low thermostability is, respectively, indicative of functionally important RNA structures and sequence segments in mRNA coding regions. We therefore speculate that the melting temperature is a crucial parameter for the identification of functionally important RNA structure and sequence elements. We have been able to verify the association of high thermostability with functional importance for two types of RNA structure elements—zipcodes in the yeast mRNA coding regions and URE2 IRES. The lack of experimentally determined and precisely characterized sequence motifs in the coding regions of yeast mRNAs prevented us from directly assessing the functional implications of low thermostability. While in previous research coding regions carrying RNA-level functions were associated with synonymous constraint elements and relaxed protein structure constraints, we find that synonymous constraint is only apparent in functionally important sequence regions (e.g. binding sites) and that functionally important RNA structures are not under synonymous constraint. This finding may prove useful in future investigations of functionally important elements in mRNA coding regions, as the overall attention to mRNAs structure grows. A typical example is constituted by RNA thermometers—temperature-sensitive...
RNA structural elements that are typically located in the 5′ UTR of mRNAs and form a secondary structure that traps the ribosome binding site and/or the translation initiation codon (57). In response to temperature changes an RNA thermometer undergoes a conformational transition, which impacts translation efficiency and eventually regulates gene expression (58). RNA thermometers have recently attracted growing attention (59) and efforts have been made to discover new elements of this type (40). Secondary structures of RNA thermometers are more conserved than their primary sequence (60), which is analogous to the high T\textsubscript{m} regions described in this work. We therefore speculate that our findings may facilitate the search for new RNA thermometers in mRNA coding regions. The characteristics of high T\textsubscript{m} regions, including strong sequence–structure relationships, conservation patterns of thermo-stable and meltable positions, and relaxed sequence constraint could serve as features to narrow down the search space in RNA thermometer discovery.

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR Online.

**ACKNOWLEDGEMENTS**

We thank Yue Wan and Howard Y. Chang for kindly providing data on melting temperatures of yeast mRNAs.

**FUNDING**

China Scholarship Council (No. 201206740006 to F.Q.). This work was supported by the German Research Foundation (DFG) and the Technical University of Munich (TUM) in the framework of the Open Access Publishing Program.

**Conflict of interest statement.** None declared.

**REFERENCES**


**Figure 8.** (A) Summary of the findings about high T\textsubscript{m} and low T\textsubscript{m} regions. (B) Inferences based on these findings.


