PCCP

PERSPECTIVE

The shape-shifting quasispecies of RNA: one sequence, many functional folds

Matthew S. Marek,[ab] Alexander Johnson-Buck[a] and Nils G. Walter*[a]

E Unus pluribum, or "Of One, Many", may be at the root of decoding the RNA sequence-structure–function relationship. RNAs embody the large majority of genes in higher eukaryotes and fold in a sequence-directed fashion into three-dimensional structures that perform functions conserved across all cellular life forms, ranging from regulating to executing gene expression. While it is the most important determinant of the RNA structure, the nucleotide sequence is generally not sufficient to specify a unique set of secondary and tertiary interactions due to the highly frustrated nature of RNA folding. This frustration results in folding heterogeneity, a common phenomenon wherein a chemically homogeneous population of RNA molecules folds into multiple stable structures. Often, these alternative conformations constitute misfolds, lacking the biological activity of the natively folded RNA. Intriguingly, a number of RNAs have recently been described as capable of adopting multiple distinct conformations that all perform, or contribute to, the same function. Characteristically, these conformations interconvert slowly on the experimental timescale, suggesting that they should be regarded as distinct native states. We discuss how rugged folding free energy landscapes give rise to multiple native states in the *Tetrahymena* Group I intron ribozyme, hairpin ribozyme, sarcin–ricin loop, ribosome, and an *in vitro* selected aptamer. We further describe the varying degrees to which folding heterogeneity impacts function in these RNAs, and compare and contrast this impact with that of heterogeneities found in protein folding. Embracing that one sequence can give rise to multiple native folds, we hypothesize that this phenomenon imparts adaptive advantages on any functionally evolving RNA quasispecies.

[a] *Department of Chemistry, 930 N. University Ave., University of Michigan, Ann Arbor, MI 48109-1055, USA. E-mail: nwalter@umich.edu; Tel: +1 734 615 2060*
[b] *Graduate Program in Cellular and Molecular Biology, 930 N. University Ave., University of Michigan, Ann Arbor, MI 48109-1055, USA*

1. Introduction

The discovery three decades ago that certain RNA molecules, termed ribozymes, catalyze chemical reactions in a manner similar to protein enzymes demonstrated an unexpected level of functional versatility of RNA that may have spawned life in

the form of an RNA world.[1-3] Accordingly, over the last decade a large number of non-protein coding RNAs (ncRNAs) have been discovered that play essential roles in all aspects of modern life.[4-7] These roles include regulation of gene expression,[8,9] post-transcriptional RNA processing,[2,3,10] protein biosynthesis,[11] and essential genomic processing in pathogens.[12-16] It was also discovered that some ribozyme motifs are broadly distributed among a wide set of organismal genomes.[17-20] Moreover, *in vitro* selection has generated ribozymes with additional activities such as aminoacyl-RNA synthesis,[21] self-replication,[22] and organic synthesis,[23] all functions postulated to have played a pivotal role in the RNA world. Finally, a rapidly increasing number of crystal structures has shed light onto the impressive complexity of the underlying RNA structures.[24-26] Clearly, RNA has the capacity to assume a wide variety of functions based on the ability of its sequence to encode versatile three-dimensional structures, yet our understanding of the RNA sequence-structure–function relationship is still in its infancy.

Self-cleaving ribozymes are ideal model systems for the study of sequence-structure–function relationships in ncRNA, since their activity (and hence, proper folding) can be quickly and easily assayed.[24,25,27-30] These ribozymes catalyze a site-specific transesterification of the phosphate-ribose backbone resulting in the formation of two cleavage products: a 5' product bearing a 2'–3' cyclic phosphate, and a 3' product bearing a 5'-OH group. The activity of these ribozymes depends on the presence of metal cations—especially divalents such as Mg^{2+}—that stabilize the active tertiary structure of an RNA and may also confer a direct chemical rate enhancement.[31] Additionally, the relatively small size of most self-cleaving ribozymes allows for convenient *in vitro* transcription of large amounts of sample for biochemical assays and chemical synthesis to incorporate site-specific modifications and labels for chemogenetic and biophysical studies.

Perhaps due to the ease of relating folding to activity, studies of ribozymes have revealed a peculiar propensity of RNA to adopt alternate conformations, a phenomenon often described as conformational heterogeneity. While the first reversible misfolding of RNA was reported in leucyl-tRNA ($tRNA^{Leu}$),[32,33] the prevalence of alternate kinetically stable structures in RNA became more apparent with studies of several self-cleaving ribozymes. In most cases, such conformational heterogeneity was attributed to "misfolded" (inactive or less active) ribozymes or long-lived folding intermediates.[34-37] In fact, virtually all ribozymes, as well as many other RNAs, are prone to this type of conformational heterogeneity, leading to a persistent view in the field that alternative folding is a nuisance to be avoided.[38,39]

A particularly intriguing example of folding heterogeneity was recently characterized in the *Tetrahymena* group I intron (TG1I) ribozyme (Fig. 1).[40] Upon binding of the substrate to its 5'-end, the TG1I ribozyme forms a helix termed P1 that subsequently swings by ~ 6 nm to dock the substrate into the preformed active site and form the active complex (Fig. 1A). Using single-molecule fluorescence resonance energy transfer (smFRET) measurements and cleavage activity assays, the authors provided evidence that subpopulations of the ribozyme exhibiting widely variable (>800-fold) docking equilibrium constants are, surprisingly, all catalytically active. In fact, 94% of all molecules within these different populations maintain the same rate constant of catalysis (Fig. 1E and F). While only a small fraction of molecules spontaneously switches between subpopulations on an experimentally accessible time scale (*i.e.*, the heterogeneity is relatively *static*), molecules can be induced to redistribute among subpopulations by refolding through the removal and subsequent reintroduction of Mg^{2+} ions (Fig. 1B–D).[40] This finding strongly suggests that several active, or "native", states arise from conformational differences rather than changes in chemistry or local environment.

This work on the TG1I ribozyme provides a strong impetus to revisit questions about the origin and possible biological function of folding heterogeneity in RNA. In fact, evidence of very similar behavior has been accruing for a number of functional RNAs over the past decade. In the following we will discuss how the physical properties of RNA give rise to a propensity for heterogeneous folding. Providing further examples, we will show that such folding behavior is commonplace, and in some cases clearly contributes to RNA function. Finally, we will speculate as to the significance of this behavior in the context of molecular adaptability and evolution.

2. The free energy landscape of RNA folding is rugged and frustrated

Among biopolymers, RNA possesses a number of characteristics that make its folding behavior unique. First, the multitude of dihedral angles in the phosphate-ribose backbone of RNA results in an immense range of possible topologies (or folds) for even relatively short RNAs. Second, the relative dominance of only a few types of base pairing interactions (Watson–Crick A·U and G·C, as well as common G·U wobble pairs) results in a "frustrated" folding landscape with a large number of nearly degenerate secondary structures. Third, the ability of RNA to form highly stable duplexes, cooperatively reinforced by a large number of hydrogen bonds and base-stacking interactions, gives its folding landscape a deeply

Nils G. Walter

Nils G. Walter is a Professor of Chemistry at the University of Michigan, Ann Arbor, and Director of the Single Molecule Analysis in Real-Time (SMART) Center. He received his PhD degree at the Max-Planck-Institute for Biophysical Chemistry, Germany, with Nobel Laureate Manfred Eigen. After a postdoc with John M. Burke at the University of Vermont, he took a faculty position at Michigan in 1999. He currently has over 100 publications, including an edited book on the biophysics of non-protein coding RNAs. While at Michigan, Dr Walter has focused on the functional dynamics of ribozymes, RNA–protein complexes, and engineered nanorobots.
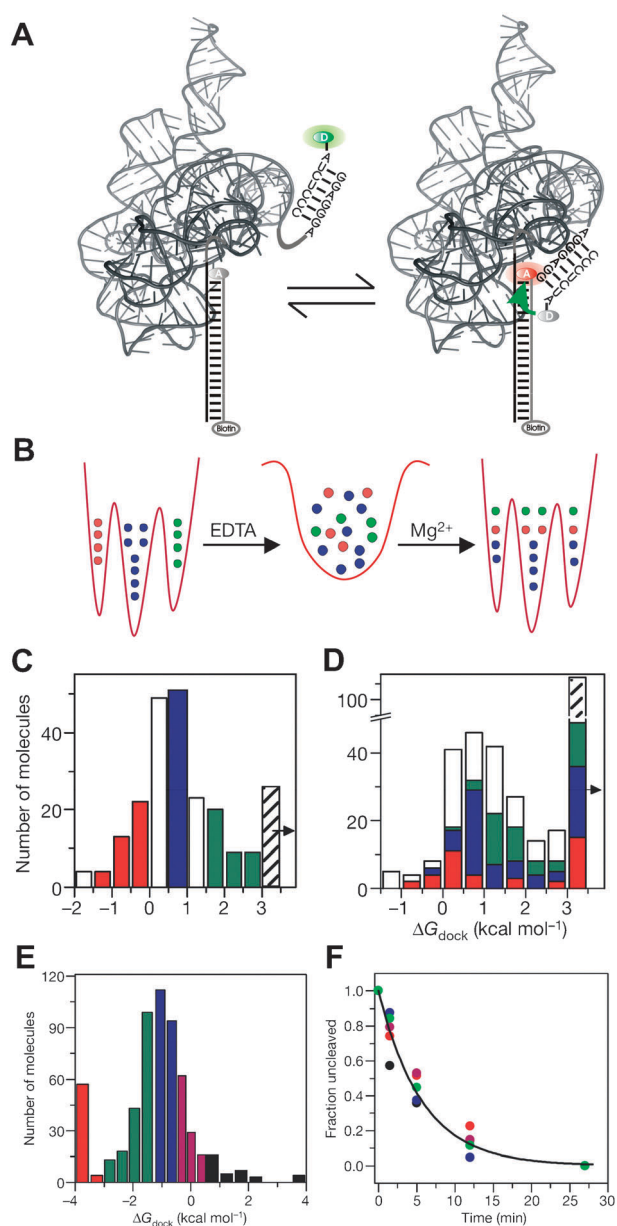
This journal is © the Owner Societies 2011

Phys. Chem. Chem. Phys., 2011, **13**, 11524–11537 | 11525

Fig. 1 Single molecule observations of the TG1I ribozyme reveal folding heterogeneity that manifests in multiple native states.[40] (A) Structural representation[143] of TG1I ribozyme docking as monitored by smFRET experiments. The sequence of helix P1, composed of the 5′-end of the ribozyme and the substrate RNA strand, is shown with donor fluorophore (D) attached to the 3′-end of the substrate strand. The TG1I ribozyme contains a 3′-extension that is hybridized to a DNA oligonucleotide (grey) with a 3′-acceptor fluorophore and 5′-biotin for surface immobilization and 3′-acceptor. (B) Schematic representation of the deeply furrowed folding landscape of the TG1I ribozyme. Molecules (red, blue, and green) redistribute between three docking conformations after partial denaturation by removal (with EDTA) and addition of Mg^{2+}. (C and D) The three energy wells manifest as single molecule distributions of vastly different docking free energy (ΔG_{dock}), between which molecules (color-coded) redistribute from before (C) to after (D) denaturation. (E) The overall distribution of docking behaviors was binned into five color-coded categories. (F) Cleavage assays of molecules representing these five docking categories (color-coded) were all shown to display kinetics similar to the global average (solid line). In part adapted with permission from ref. 40.

furrowed character, resulting in alternate secondary and tertiary structures that may take very long to interconvert once formed.[41] Together, these factors give rise to a high propensity to form kinetically trapped alternate folds.

The ribose-phosphate backbone contains six rotatable dihedral angles per nucleotide (conventionally labeled α–ζ, Fig. 2A), or even more when the constrained torsional angles of the sugar ring (ν_0–ν_4, Fig. 2A) and the glycosidic bond dihedral (χ, Fig. 2A) are considered. This number compares to only two such angles per amino acid in proteins, resulting in a much wider range of possible conformations in RNA than in peptides of comparable size. While Watson–Crick base pairing in standard A-form helical stems places strict restraints on the possible torsional angles, single-stranded regions of RNA molecules—formally junctions, loops, and bulges (Fig. 2B)—still have a multitude of possible conformations. As a result, the folding topology of an RNA can be rendered quite complex through the formation of multiple helical junctions and tertiary interactions such as pseudoknots, ribose zippers, kissing-loop interactions, and tetraloop-receptor interactions (Fig. 2C).[42–47]

Several advances have been made towards classifying combinations of dihedral angles into structural motifs based on mono- or dinucleotide units, reducing the number of empirically observed conformations considerably. In one approach, backbone dihedral angles were organized into so-called "suites", where a conformation is defined between adjacent sugar residues as a group of two sets of angles: δ–ε–ζ and α–β–γ–δ.[48] Along with careful quality-based filtering of crystallographic source data, this system enables the classification of empirically observed RNA backbone conformations into 42 suites, later expanded to 46 conformers in an effort that unified a handful of other approaches.[49] Still, the number of conformations available to even a short oligonucleotide would be staggering in the absence of other information, posing a serious obstacle to decoding the relationship between sequence and structure.

Fortunately, the complexity of RNA folding is reduced considerably by its hierarchical nature wherein the secondary structure typically folds before the tertiary structure (Fig. 2D).[39,50] Hybridization of complementary segments of an RNA sequence occurs in as little as microseconds.[28,51] Once formed, an A-form RNA helix is stabilized by base pairing and base stacking interactions worth about -1 to -3 kcal mol^{-1} per base pair.[41,50] As a result, even short RNA oligomers of ~ 10 base pairs (bp) may have half-lives of dissociation on the order of minutes or hours near room temperature.[52,53] On the other hand, tertiary interactions (loop–loop hydrogen bonding, base–phosphate and base–sugar interactions) frequently form and interconvert on the timescale of milliseconds to seconds and are comparatively weak.[54–56] Because of this difference in kinetics and stability, the secondary structure places relatively rigid constraints on accessible tertiary structures; conversely, isolated stem-loops of a larger RNA structure often fold properly even in the absence of tertiary interactions.[41] While it simplifies the prediction of the RNA structure, the stability of the secondary structure also gives rise to a deeply furrowed free energy landscape of folding with large barriers separating different folded states.

Finally, the folding free energy landscape of RNA is described as "frustrated" by a large number of possible alternative
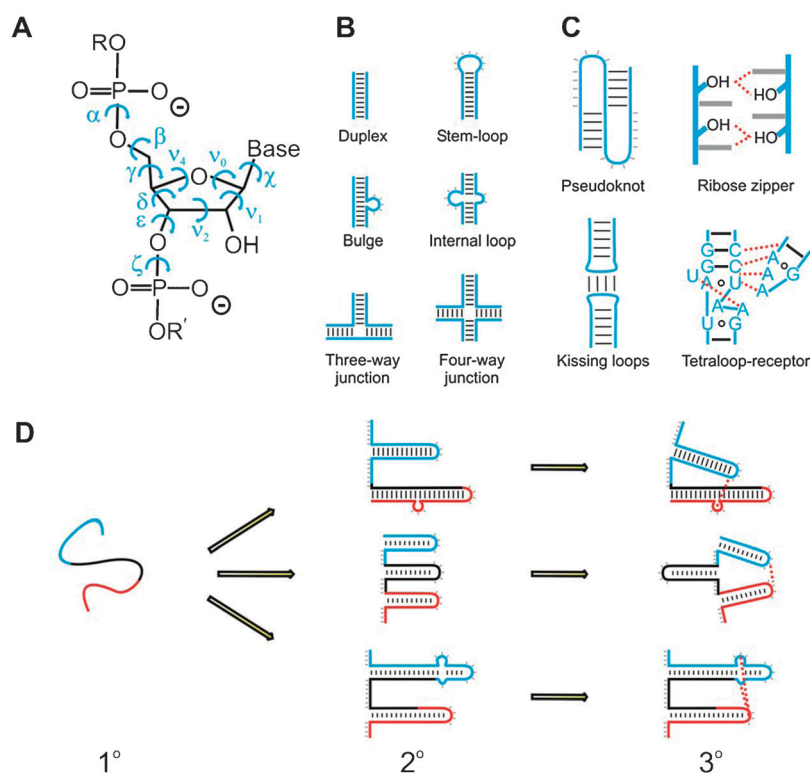
Fig. 2 Diversity of RNA structures. (A) RNA has many dihedral angles (blue) that contribute to a large number of conformational degrees of freedom per nucleotide. These include angles of torsion about the bonds of the RNA backbone (α–ζ), the nucleosidic bond (χ), and pseudo-rotation angles within the ribose ring (ν_0–ν_4). (B) Hydrogen bonding (pairing) between complementary bases gives rise to several common secondary structure motifs, consisting of the RNA backbone (blue) held in various arrangements by base pairs (black line segments) and often resulting in short segments of unpaired nucleotides (short grey lines). (C) Common tertiary structure motifs build hierarchically onto secondary structure elements. (D) The small alphabet of RNA often results in a frustrated folding landscape with many possible conformations stabilized by alternative base-pairing. In this schematic, three segments of a linear RNA molecule are marked in blue, black, and red to illustrate the variety of possible secondary ($2°$) and tertiary ($3°$) structures arising from a single primary ($1°$) sequence of nucleotides.

folds (Fig. 2D). This description derives from the tendency of RNA to become trapped in local minima and requires great energy to overcome these barriers and attain a "native" structure. Due to the diminutive 4-nucleobase alphabet of RNA, there is a high probability that any two sequences of nucleotides will have coincidentally complementary regions. Even random RNA sequences containing tens of nucleotides are predicted to be approximately 50% base-paired,[57] implying the existence of numerous possible alternative secondary structures. Modern software packages use partition function approaches to predict the RNA secondary structure, yielding more reliable predictions of base pairing and thermodynamic parameters than considering only the minimum-energy structure.[58–61] Clearly, the number and impact of possible alternate secondary structures are significant.

3. Multiple active species with slow interconversion are observed in a number of RNAs

Due to the complex and rugged folding free energy landscape of RNA, long-lived variations in the structure, or so-called static heterogeneities, arise in RNAs as diverse as those found in plant virus satellites, the eukaryotic ribosome, and artificial selections of ligand-binding aptamers.[40,54,62–64] Misfolding is perhaps a trivial example, but it is important to consider that one function's trash may be another function's treasure.

An example is the adoption of different secondary structures for the purpose of switching between active and inactive forms of a ribozyme, such as at different stages in the replication cycle of a pathogen. For instance, the hepatitis delta virus (HDV) ribozyme must be active in order to cleave concatemeric linear transcripts of HDV RNA into unit-length fragments for ligation into circular copies of the genome. Once this task is accomplished, however, the ribozyme becomes inactive by adopting alternate base pairing patterns as part of a long, 70% self-complementary, rod-like structure of the RNA genome of HDV, presumably for the purpose of packaging and delivery of the intact genome to new host cells.[16]

Here, we focus on another type of static heterogeneity: the presence of distinct species that all contribute, in varying degrees, to the same nominal function. In investigating such phenomena, several questions must be addressed:

1. What impact does the structural heterogeneity have on function?

2. How robust is the heterogeneity to changes in experimental conditions?

3. Does the heterogeneity arise from purely conformational differences, *i.e.*, can covalent modification or mutation be ruled out?

4. If the differences are conformational, do they represent alternate secondary structures or more subtle variations in tertiary structures?

Studies of heterogeneity in model systems have begun to answer these questions. Such studies often involve a combination of ensemble and single-molecule methods, utilizing the relative strengths of each approach to deduce the nature of the heterogeneity.[40,63–65] Ensemble assays of RNA are rapid, well established for many systems, and can employ a variety of detection methods, the two most common of which are fluorescence and autoradiography. While requiring relatively large amounts of sample material, ensemble assays often require minimal modification and processing of sample material. Single-molecule fluorescence based assays generally require precise fluorescent labeling, specialized equipment, and in-depth statistical analysis, but work with low amounts of sample. Additionally, single-molecule assays allow correlations to be drawn between conformational behavior and catalytic efficiency without physical separation. Only through utilizing these techniques in conjunction can the nature of these heterogeneities in RNA be addressed, as illustrated in the following with arguably the most prominent examples.

3.1 The hairpin ribozyme

One of the most thoroughly characterized examples of static heterogeneity in an RNA is found in the hairpin ribozyme, a self-cleaving and self-ligating small ribozyme first discovered in the negative strand of the tobacco ringspot virus (TRSV) satellite RNA.[66,67] Central to catalytic activity of the ribozyme is the docking of its two helix–loop–helix domains (A and B) by tertiary interactions between the nucleotides in bulges present in each helix to form the active site. At room temperature, in the presence of Mg^{2+} ions, this docking is readily reversible (Fig. 3A). Despite its much smaller size, the docking of a helix (domain A) into the catalytic core (domain B) of the hairpin ribozyme, enabled by a flexible hinge region, superficially resembles docking of helix P1 into the TG1I ribozyme (compare Fig. 1A and 2A).

Utilizing smFRET assays to quantify the kinetics of these docking and undocking transitions, the Walter and Chu groups first uncovered kinetic heterogeneities in the hairpin ribozyme in 2002.[55] Only by utilizing smFRET could the behaviors of individual molecules be observed without the ensemble averaging inherent to bulk assays. For smFRET detection, the opposite ends of one strand of the RNA were labeled with the fluorophores Cy3 (smFRET donor) and Cy5 (acceptor), which exhibit high FRET efficiency when proximal to each other in the docked state and low FRET efficiency in the undocked state (Fig. 3A). Samples were immobilized on a derivatized quartz slide, and Cy3 and Cy5 emission intensities of single molecules were observed by total internal reflection fluorescence (TIRF) microscopy.[55,68] Although the rate constant of docking was invariant across all molecules, subpopulations of ribozyme molecules were observed to undock with four distinct rate constants spanning three orders of magnitude. Furthermore, individual molecules rarely (<5% of observed molecules) switched between these kinetic regimes even over 3 hours incubation—each molecule retained a "memory" of its undocking rate constant. Strikingly, the multiple undocking populations were observed to all convert to product from the docked state; i.e., the faster the undocking relative
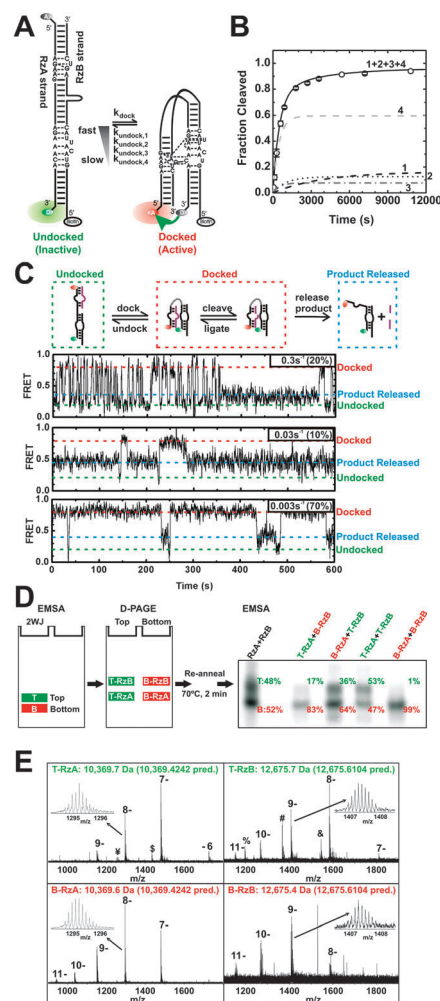

Fig. 3 Folding heterogeneity of the hairpin ribozyme.[55,65,69] (A) smFRET assays detect docking heterogeneity of the hairpin ribozyme. Donor (D) and acceptor (A) fluorophores are attached to the 3′- and 5′-ends of the RzA strand, respectively, while the RzB strand carries a 5′-biotin for surface immobilization. Whereas all active HpRz molecules display a single docking rate constant (k_{dock}), four distinct undocking rate constants ($k_{undock,1-4}$) are observed. (B) Each docking rate constant leads to a distinct cleavage time course (numbered as in panel A), which taken together account for the overall cleavage observed in ensemble assays (1 + 2 + 3 + 4). (C) Reaction pathway of the hairpin ribozyme and resulting single-molecule multiple-turnover cleavage data. The docked, undocked and product released states are indicated. Three single molecule time trajectories demonstrate catalytic proficiency of each of the distinct subpopulations they represent (undocking rate constants and the fraction of molecules undocking with this rate constant are indicated). (D) Structural heterogeneity upon EMSA in which two structural forms of the hairpin ribozyme are resolved (Top and Bottom species). Separation of the two component strands RzA and RzB of each species by denaturing polyacrylamide gel electrophoresis (D-PAGE) yields four RNAs, as indicated, that were annealed in all possible combinations and analyzed by EMSA. The resulting fractions of Top and Bottom species are given. (E) High-resolution FT-ICR mass spectrometry of each of the four RNA strands isolated by the procedure described in panel D reveals identical isotope envelopes (insets) and average masses (consistent with the predicted masses) of the corresponding strands from the Top and Bottom species. In part adapted with permission from ref. 65 and 69.

11528 | *Phys. Chem. Chem. Phys.*, 2011, **13**, 11524–11537

This journal is © the Owner Societies 2011

to the cleavage/ligation rate constants, the more docking/ undocking cycles become necessary on average for a successful substrate turnover (Fig. 3B and C). Thus, the existence of distinctly undocking, yet catalytically active, native states quantitatively explains the multi-exponential cleavage kinetics noted in ensemble assays of the hairpin ribozyme (Fig. 3A and B).[55,69]

Importantly, it was shown that this heterogeneity is not caused by the surface immobilization used in smFRET measurements. First, bulk cleavage assays of ribozymes with the same modifications (Cy3, Cy5, and biotin) yield similar kinetic parameters to those obtained by smFRET techniques.[55,69,70] Second, the same kinetic heterogeneity was observed by smFRET when the Lilley and Ha groups captured ribozymes on the slide surface by encapsulation within phospholipid vesicles, rather than through direct biotin–streptavidin interaction.[71]

Surprisingly, the level of undocking heterogeneity proved not to be affected by either changes in Mg^{2+} concentration[72] or site-specific mutations or modifications with impact on the undocking rate constants; the rate constant of each sub-population simply shifted by about the same factor.[69] In subsequent work, evidence was presented that the molecular heterogeneities of the hairpin ribozyme, while extremely long-lived, are not due to any detectable covalent modification.[65] More specifically, through the use of electrophoretic mobility shift assays (EMSAs) on polyacrylamide gels, two distinct, either slow- or fast-migrating species of the ribozyme were resolved (as top and bottom bands, respectively). Upon elution from the gel and further analysis of the fluorophore distance distributions by time-resolved FRET (trFRET), the two species showed marked differences; while less than 40% of the slow-migrating species adopts the docked conformation, greater than 80% of the fast-migrating species does. Analysis by smFRET showed that the top band is enriched in the fastest undocking subpopulation, which is expected to be less docked and compact, consistent with its lower mobility, whereas the bottom band is enriched in the most slowly undocking subpopulation, expected to reside largely in the compact docked conformation.[65]

This differential enrichment within two separable species opened up an opportunity to determine whether the hairpin ribozyme subpopulations can be induced to redistribute. To this end, the individual 5′- and 3′-segment strands (termed RzA and RzB) of each species were further separated by denaturing gel electrophoresis, removing all base pairing between them. When the RzA and RzB strands of the slow-migrating species were re-annealed and re-analyzed by EMSA, the RNA redistributed into fast- and slow-migrating species as before, but not the RzA and RzB strands originating from the fast-migrating species, which continued to preferentially form the fast-migrating species (Fig. 3D). Accordingly, mixing RzA and RzB strands originating from different species yields intermediate levels of the two EMSA bands (Fig. 3D). This asymmetry suggests that the molecular heterogeneity leading to formation of the catalytically more active fast-migrating species is maintained even upon full denaturation of all interstrand base pairs.[65]

In the same work, we were also able to show by high-resolution (<1 amu) mass spectrometry that the RzA and RzB strands from the fast- and slow-migrating strands are identical in mass (Fig. 3E).[65] In addition, both chemically synthesized and in vitro transcribed ribozyme behave similarly, and footprinting revealed only minor differences in the secondary structure between the EMSA separated species. While these observations do not rule out mass-neutral covalent modifications as the source of conformational heterogeneity, such as certain UV-induced crosslinks, they do strongly support the notion that more subtle conformational[73] or topological differences[40] are at play.

In summary, the hairpin ribozyme folds into multiple active populations with disparate global dynamics but only subtle differences in the secondary structure. Compared to the multiple native states of the TG1I ribozyme many parallels are observed, although the native states of the hairpin ribozyme are separated by larger free energy barriers and, consequently, do not interconvert quite as freely upon refolding as those of the TG1I ribozyme (Fig. 4).

3.2 The sarcin–ricin loop

Domain B of the hairpin ribozyme shares sequence homology and a version of an RNA structural motif termed an S-turn with the toxin-sensitive sarcin–ricin loop (SRL) of the large subunit ribosomal RNA (rRNA, Fig. 5A).[65] Intriguingly, conformational heterogeneity has also been observed in this stem-loop motif that is highly conserved across all kingdoms.[62] Several 27-nt versions of the sarcin–ricin loop (SRL) from rat were transcribed in vitro and analyzed by EMSA, revealing two species of identical length but with different electrophoretic mobilities (Fig. 5B), similar to our observations on the hairpin ribozyme. No heat-induced interconversion of the two species was observed, suggesting that the heterogeneity is thermodynamically quite stable. Intriguingly, the slow-migrating species (constituting 30–50% of the total SRL) is resistant to cleavage by the endoribonuclease restrictocin (a sarcin analog) when compared
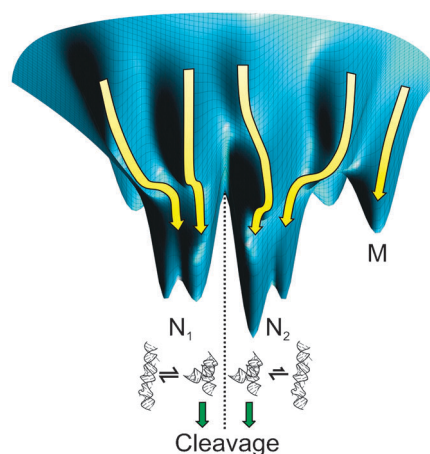

Fig. 4 Schematic representation of the rugged conformational free energy landscape (blue surface) of the hairpin ribozyme. An individual molecule folds along one of many possible pathways (yellow arrows) to one of multiple native states (N_1, N_2) separated by relatively large energy barriers. These native states sample similar conformations, albeit with different kinetics, and thus possess similar cleavage activity. Alternatively, the molecule may enter a trapped misfolded state (M) that is non-functional.
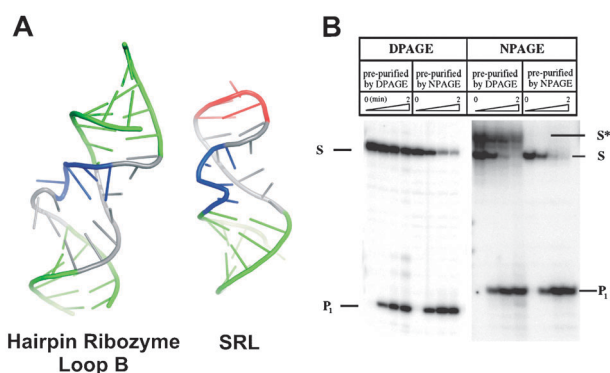
Fig. 5 Folding heterogeneity of the sarcin–ricin loop (SRL). (A) Comparative cartoon representations of crystal structures of loop B of the docked hairpin ribozyme and the SRL reveal a common, conserved S-turn motif (blue).[65] (B) *In vitro* transcribed SRL was either purified by denaturing (DPAGE) or non-denaturing polyacrylamide gel electrophoresis (NPAGE or EMSA), subjected to restrictocin cleavage, and the products over time analyzed by gel electrophoresis, as indicated.[62] EMSA in particular reveals a non-interconvertible, slow-migrating S* species that is relatively resistant to restrictocin cleavage. Adapted with permission from ref. 62 and 65.

to the native substrate (70–50% of total SRL, Fig. 5B).[62] As with the hairpin ribozyme, chemically synthesized SRL shows the same heterogeneity. While it is unknown whether this phenomenon persists *in vivo*, or how it might affect translation by the ribosome, *i.e.*, which of the species represents a native state, it is tempting to speculate that the presence of multiple folds could confer an adaptive advantage through partial resistance to different toxins. Conformational heterogeneity may thus be akin to sequence variation of rRNA; in *E. coli* alone there are seven different rRNA sequences, each with minor sequence discrepancies that may confer increased adaptability to the cell.

3.3 The ribosome

Given that a small stem-loop of the ribosome exhibits profound folding heterogeneity, it comes as no surprise that the bacterial ribosome, at ~2.4 MDa and with three RNA and >50 protein components the largest of all ribozymes, has recently been shown to display heterogenous intersubunit rotational dynamics in its pre-translocation complex (Fig. 6). Pre-translocation complexes (pre-complexes) occur when the acetyl(A)-site and peptidyl(P)-sites of the ribosome are occupied by tRNAs that have undergone a peptidyl transfer but have yet to translocate to the P- and E-sites, respectively. As shown in previous ensemble and single-molecule FRET based studies,[74–77] the small (30S) and large (50S) subunits of the ribosome spontaneously ratchet among the classic pre-complex (with the two tRNAs occupying the A/A–P/P sites in the 30S/50S subunits) and two hybrid conformations (Hy1: A/A–P/E, Hy2: A/P–P/E). Subsequently, elongation factor-G catalyzes translocation to the post-translocation complex (P/P–E/E). Heterogeneity arises in that only ~2/3 of all ribosomal complexes observed by smFRET display dynamic fluctuation between the classic and hybrid states (Fig. 6).[64,76] Of these, a roughly equal distribution of complexes exhibits transitions either between the Hy1 and the classic state or the Hy1 and the Hy2 state.

The remaining ~1/3 of static complexes are distributed among the low FRET classic and hybrid states with a high prevalence for the classic state, along with a small number of high FRET molecules occupying the POST state (Fig. 6C).[64] The authors propose a qualitative folding free energy landscape of translocation with high energy barriers preventing reverse translocation and smaller minima/maxima for the fluctuating and nonfluctuating FRET states, downwardly trending towards the Hy2 state that is adopted just before translocation.[64]

3.4 The AN58 aptamer

So far, we have discussed the static heterogeneities observed in naturally occurring RNAs, yet there exists at least one artificially selected RNA aptamer that shows similar behavior. The AN58 RNA is a truncation construct of the larger aptGluR2-99 aptamer, *in vitro* selected to block the GluR2 AMPA receptor by binding to regions thought to be essential to function.[78] When AN58 was transcribed *in vitro* and purified by gel electrophoresis, two different, non-interconvertible populations (M1 and M2) were observed by EMSA (Fig. 7). Neither of these isoforms can individually inhibit the GluR2 AMPA receptor, yet when recombined, they show inhibition comparable to that of unseparated AN58 and very similar to that of the parent aptGluR2-99 (Fig. 7B). Sequencing by primer extension and glyoxal treatment followed by gel electrophoretic analysis showed that M1 and M2 have the same length and sequence. To elucidate any differences in the secondary structure between M1 and M2, a combination of in-line probing and selective 2'-hydroxyl acetylation analyzed by primer extension (SHAPE) was utilized. Results from these footprinting assays suggested that M1 and M2 form short stem-loop structures in alternate regions of the unstructured 3'-region of the RNA (Fig. 7D).[63] Thus, in contrast to the hairpin ribozyme, in this case alternate, very stable secondary structures appear to be responsible for the heterogeneous behavior. Yet AN58 maintains its static heterogeneity in a manner similar to both hairpin ribozyme and SRL (Fig. 7C).[63]

In summary, while likely only the tip of the proverbial iceberg, the above five examples (TG1I and hairpin ribozymes, SRL, bacterial ribosome, and AN58 aptamer) arguably represent the best characterized occurrences of heterogeneous RNA behaviors. In each case, different copies of what is ostensibly a single chemical species with a defined nucleotide sequence are capable of adopting different conformations of similar or distinguishable native functionality.

4. Parallels with protein folding

Often, insights into RNA structure–function relationships trail those of proteins; it is therefore helpful to take a look at our current understanding of heterogeneities in protein folding. There are many similarities between the folding behavior of RNA and that of proteins. Like RNA, proteins fold hierarchically, with local interactions forming first and largely determining the overall structure of the folded polypeptide.[79,80] Both RNA and proteins can adopt a given fold with very few specific sequence requirements.[19,81] However, there are some important differences. Protein folding is mainly directed by fairly
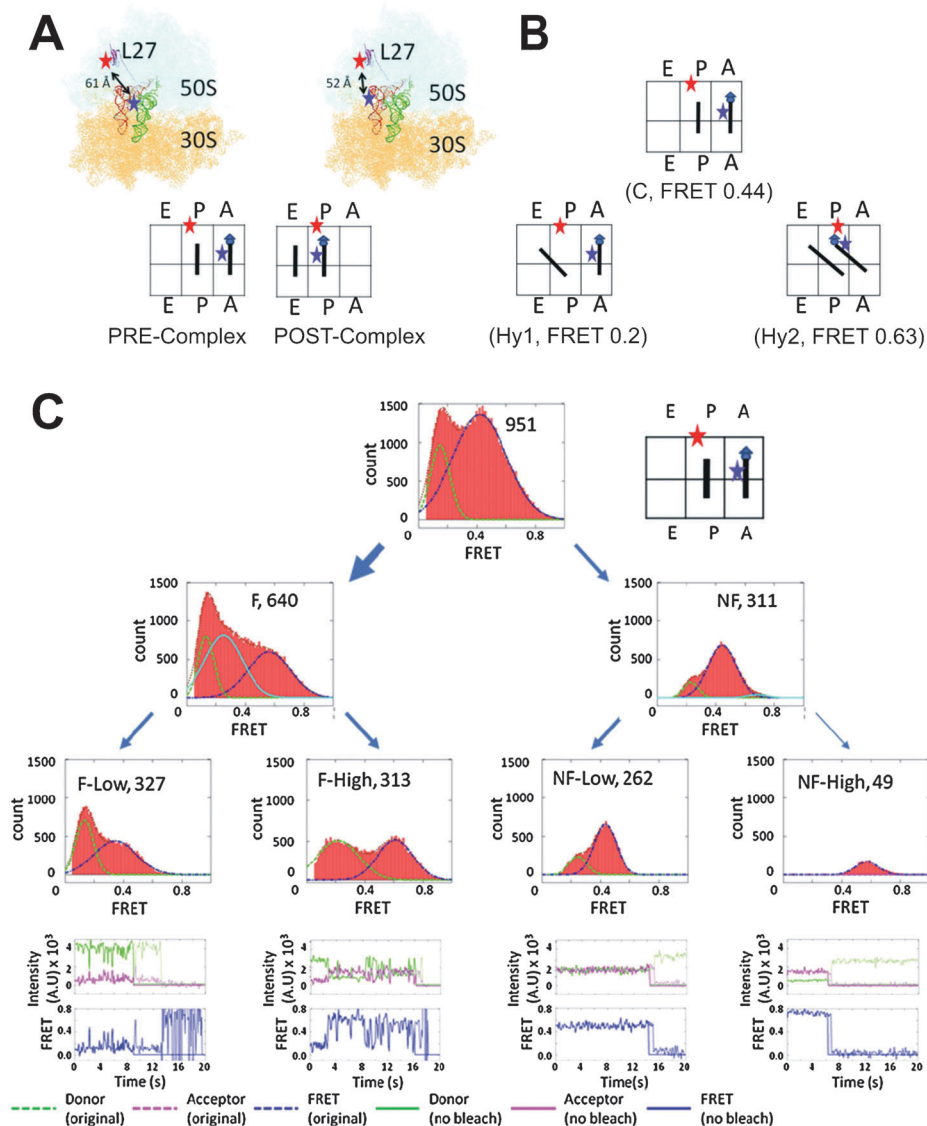
11530 | *Phys. Chem. Chem. Phys.*, 2011, **13**, 11524–11537

This journal is © the Owner Societies 2011

Fig. 6 Folding heterogeneity of the bacterial ribosome.[64] (A) Structural and schematic representations of the fluorophore labeled *E. coli* ribosome in the pretranslocation (PRE) and posttranslocation (POST) complexes. Fluorophores attached to the L27 protein of the large subunit and A-site tRNA are used to characterize by smFRET their relative motions as the ribosome samples multiple PRE-complex conformations. (B) Ribosomes are observed in conformations including the classic state, C, along with two hybrid states, Hy1 and Hy2, with tRNA occupying various sites in the large and small subunit, each with a distinct FRET value, as indicated. (C) Ribosomes can be categorized by their dynamic (or fluctuating, F) or static (nonfluctuating, NF) occupancy of these conformations. Among the F molecules, roughly half exhibit low (0.2) to mid (0.44) FRET transitions while the other half exhibit low to high (0.63) transitions. Among the NF molecules, three categories emerge, where two categories with the low and mid FRET values of the C and Hy1 states, respectively, were described as a joint NF-Low category, whereas a high FRET category was assigned to POST complexes. Representative smFRET time trajectories accompany each of these four categories. Adapted with permission from ref. 64.

nonspecific hydrophobic collapse[80,82] rather than by the formation of specific hydrogen bonds such as in tightly aligned (stacked) RNA base pairs. Furthermore, the native states of proteins are typically only marginally stabilized by 5–10 kcal mol^{-1} relative to their denatured states, which is comparable to the stability of an RNA duplex containing a mere 8–10 base pairs.[83,84] In other words, a short RNA stem-loop folds with similar thermodynamic stability as an entire protein. This distinction is reflected by the fact that, whereas RNA secondary structure elements fold very stably in isolation,[41] the secondary structure of proteins is often strongly influenced by context

(such as tertiary interactions).[85] Thus, one might expect to observe less profoundly heterogeneous folding in proteins than in RNA.

Ensemble kinetic experiments on protein folding have frequently observed trapped conformations. Many of these are non-native folding intermediates that generally persist for only seconds or less, as for lysozyme[86] and staphylococcal nuclease.[87] In other cases, states with native-like activity have been observed. For instance, upon refolding from 8 M urea, the majority of dihydrofolate reductase folds transiently into an intermediate that efficiently binds a substrate analog.[88] Similar behavior was
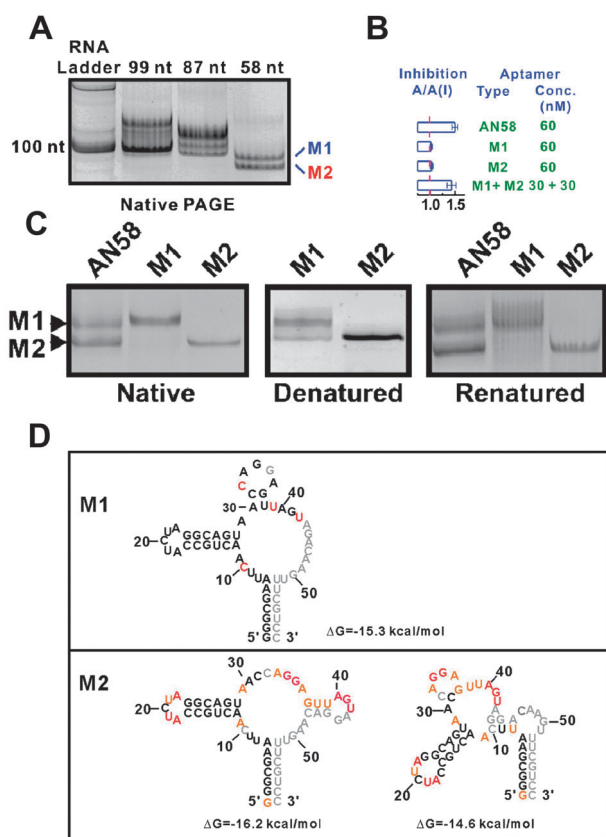
Fig. 7 Folding heterogeneity of the AN58 aptamer.[63] (A) AN58, a truncation of a larger AN99 aptamer, resolves as two discrete bands during EMSA (Native PAGE), termed M1 and M2, suggesting two different structures of chemically homogenous RNA. (B) While AN58 is capable of blocking the action of the GluR2 AMPA receptor channel, neither purified form (M1 or M2) can inhibit activity on its own, yet a mix of both M1 and M2 results in restoration of AN58 function. (C) M1 and M2 were separated from one another by EMSA and analyzed by EMSA (left panel), denatured by urea-containing polyacrylamide gel electrophoresis and visualized (middle panel, denatured), then refolded and further analyzed by EMSA (right panel, renatured), revealing a lack of interconversion between species. (D) Proposed secondary structures of M1 and M2. Adapted with permission from ref. 63.

observed for RNase A, which has a long-lived folding intermediate possessing enzymatic activity similar to that of the native state in spite of some structural differences.[89] In both of these proteins, the native state ultimately forms, even though RNase A takes from one to several minutes to fold to completion.

The advent of single-molecule fluorescence spectroscopy overcame the drawbacks of ensemble averaging and revealed numerous examples of heterogeneous activity and folding of single protein enzymes, showing that this property is the rule rather than the exception. Individual lipase molecules were shown to exhibit fluctuating substrate turnover kinetics, likely explained by conformational changes on the order of tens of milliseconds.[90] Single-molecule studies of flavoenzymes, monitored by changes in the intrinsic fluorescence of the flavin cofactor in the course of its redox chemistry or electron transfer to a nearby tyrosine, found fluctuations as slow as 1 s^{-1} in both the substrate turnover kinetics and conformations

of individual enzymes (Fig. 8A–C),[91,92] as did similar studies of horseradish peroxidase.[93] Longer-lived fluctuations
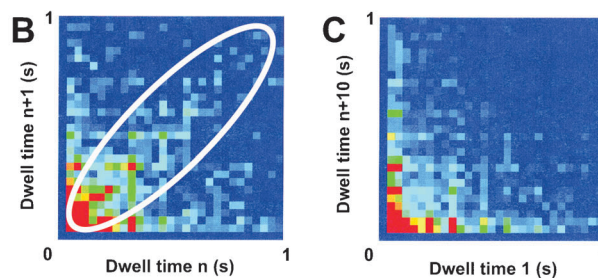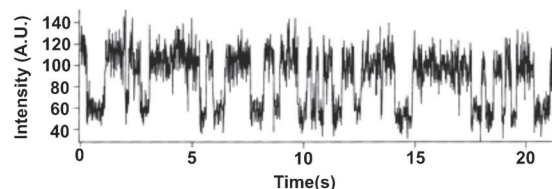
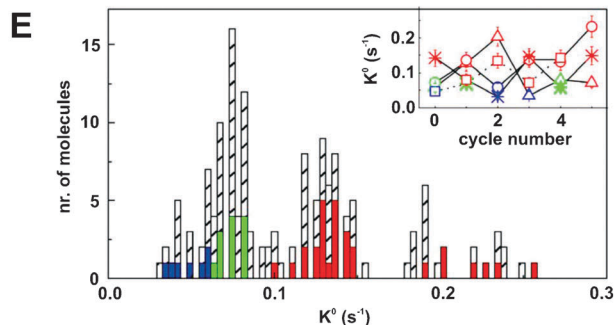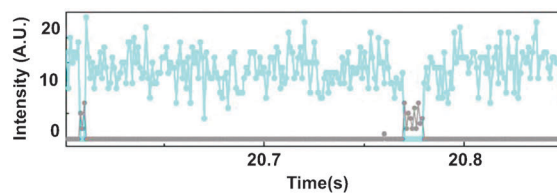**Fig. 8** Examples of folding heterogeneity in the proteins cholesterol oxidase[91] and green fluorescent protein (GFP).[103] (A) Turnover of substrates by an individual molecule of cholesterol oxidase gives rise to stochastic transitions between fluorescent and non-fluorescent states. (B) Although transitions between these two states occur stochastically, dwell times in the fluorescent state for two adjacent turnovers n and $n + 1$ are slightly correlated, as shown by the diagonal feature (encircled by a white ellipse) of a conditional probability distribution. (C) In contrast, dwell times separated by 10 turnovers are not correlated. (D) Single GFPmut2 molecules exhibit spontaneous switching between anionic (fluorescence intensity in cyan) and neutral (grey) states. (E) The frequency of switching events in folded molecules, K^0, is not uniformly distributed, but occurs in clusters. Furthermore, the value of K^0 for a molecule is predictive of the frequency of switching while unfolding in guanidinium chloride, K_{FIN}, as shown by the clustering of similar colors ($K_{FIN} \approx 440$ Hz in red, ~ 720 Hz in green, and ~ 930 Hz in blue). The value of K^0 for a given molecule is stable over 24 h, and though it may change through multiple cycles of denaturation, K^0 remains within the same cluster of the distribution (inset). Adapted with permission from ref. 91 and 103.

11532 | *Phys. Chem. Chem. Phys.*, 2011, **13**, 11524–11537

This journal is © the Owner Societies 2011

in activity, with lifetimes on the order of minutes, were observed for single molecules of bacteriophage λ exonuclease.[94] The activity of electrophoretically purified lactate dehydrogenase molecules was found to vary by a factor of four in a manner that remained constant for a given single molecule over two hours, which the authors suggested could be due to different stable arrangements of monomers in the homotetrameric enzyme.[95,96] Similar findings of static heterogeneity were made for alkaline phosphatase,[97] β-galactosidase,[98] and the DNA helicase RecBCD.[99,100] Perhaps most intriguingly, recent folding studies of a GFP triple mutant with enhanced fluorescence emission termed GFPmut2 provide evidence of multiple native states, characterized by distinct chromophore switching kinetics, that do not interconvert over several hours unless refolded from the denatured state,[101–103] providing strong evidence of long-lived conformational heterogeneity in native proteins (Fig. 8D and E).

In summary, like the RNA examples highlighted above, many native proteins exhibit conformational heterogeneity that generally lasts for milliseconds to seconds, but can persist for hours in some proteins. The commonly shorter timescale of most of these protein fluctuations may reflect the less rugged conformational landscape of proteins as compared with RNA, although there are clearly exceptions. As with RNA, the microscopic origin of this heterogeneity in proteins is generally unclear. As an exception, Polakowski et al. showed that at least some long-lived heterogeneity of enzyme activity can be attributed to covalent differences such as partial degradation of the peptide or post-translational modifications such as glycosylation, which persist in crudely purified samples.[104] However, at least in the case of the GFP mutant, GFPmut2, the differences between native states appear conformational in origin, as the states redistribute upon denaturation as is observed also for the TG1I ribozyme. The discovery of such behavior in both proteins and RNAs, in spite of their distinct biophysical properties, suggests that multiple similar native states may be a general feature of biopolymers of complex structure. The fact that wild-type GFP shows less conformational heterogeneity than GFPmut2 invokes the notion that natural evolution may in some cases select against it.

5. RNA conformational quasispecies may be natural facilitators of molecular evolution

In the context of evolution, genetic diversity within a population of organisms correlates with the fitness of that population.[105,106] Such diversity confers upon the population resistance to parasites, toxins, and other environmental insults. Numerous examples for such effects have been observed, ranging from genetic resistance to certain human diseases,[107–109] resistance of insects towards pesticides,[110] and the appearance of antibiotic-resistant strains of bacteria.[111] While a large amount of phenotypic diversity arises from genetic mutations, many other molecular sources of phenotypic variation have been elucidated over the last several decades, including the action of transcription factors and repressors,[112] covalent modification of histones and DNA,[113] RNA interference,[114–116] riboswitches,[117–121] and alternative splicing.[122,123] These mechanisms make a great degree of phenotypic diversity possible even in populations of genetically identical cells. In some cases, phenotypic variation can arise stochastically in a population, such as through the translation of low-copy-number mRNAs in cells[124,125] or the presence of very low concentrations of transcriptional regulators.[112] Thus, stochastic single-molecule events have an important impact on the fate of an entire organism and perhaps the fitness of the entire population of organisms.[126]

In the case of rapidly replicating systems with relatively high mutation rates, such as viruses or bacteria, molecular evolution is described by the quasispecies model, in which variation and selection occur not on the level of well-defined molecular species with nearly identical genetic makeup, but rather on the level of so-called quasispecies comprising clusters of related sequences replicating according to their aggregate fitness level.[127–129] Due to high replication and mutation rates, the fitness of a single genotype, which will not likely be faithfully preserved in the offspring, becomes less important than the overall fitness of the cluster or quasispecies. In fact, the functional diversity of the quasispecies confers enhanced adaptability to dynamic environments, allowing for example viruses to rapidly evolve resistance to vaccines and antiviral drugs.[130] To maximally exploit this evolutionary advantage, most viruses and organisms are found to maintain an inverse relationship between their mutational error rate and the length of their genome, i.e., they live close to an error threshold imposed by their genome length.[127–129] Consequently, mutation-inducing drugs can cause this error threshold to be crossed, resulting in lethality.[131,132]

In light of the evidence presented here, the quasispecies model, formulated in the context of genotypic variation, may now need to be extended to conformational variability of RNA with a single sequence. Specifically, given the capability of RNA to form alternative active folds that are stable in vitro relative to the lifetime of RNA in vivo, we suggest that RNA might function and evolve as conformationally distinct, but functionally related conformational quasispecies. In this view, the source of functional variation is not solely provided by the sequence, but by the inherent ruggedness and high degree of energetic degeneracy in the RNA folding landscape. In fit RNA quasispecies, then, alternate folding may constitute another level of adaptive phenotypic variation.

What advantages might such heterogeneity confer on quasispecies of RNA? First, it would enable molecules to achieve their function even if one particular conformer becomes a target for a toxin or nuclease. The observation of a conformational species of the sarcin–ricin loop resistant to cleavage by restrictocin[62] provides a salient example of how such heterogeneity could confer an immediate advantage, provided that the resistant species is still biologically functional. Such conformational heterogeneity could constitute an attractive mode of adaptation, complementary to sequence variation, when it is necessary to respond to variable environmental challenges on short time scales because larger fractions of an RNA are immediately available with an altered folding behavior than are typically found to carry a specific (set of) mutation(s) leading to such behavior. Conversely, long-term exposure to a toxin or other insult would likely drive preferential selection of sequence variants that thermodynamically or kinetically prefer the formation of the resistant conformer(s).

This journal is © the Owner Societies 2011

Phys. Chem. Chem. Phys., 2011, **13**, 11524–11537 | 11533

Second, the capability of an RNA of a single sequence to adopt multiple conformations that directly or indirectly act in concert may enable short RNA oligomers to adopt more sophisticated functions, such as found in the AN58 aptamer. While this aptamer was artificially selected *in vitro*, this type of behavior could be useful in nature due to its sequence economy. Furthermore, Huang *et al.* suggest that this type of dual-use sequence could provide a precursor to gene duplication and phenotype divergence for functional nucleic acids.[63] Previous work, in which a single RNA sequence was designed to encode the folds and activities of both the HDV ribozyme and an RNA ligase ribozyme,[133] similarly suggests that intersections in sequence space between neutral networks of distinct functional RNAs may be common, and could give rise to new folds and functions during evolution. In fact, the simplistic single RNA–single function paradigm does not do justice to the complexity of nature, where an RNA will always have to exert multiple functions in parallel. An example is the hairpin ribozyme that, like the HDV ribozyme, needs to cleave concatemeric replication intermediates of its satellite RNA into monomers, then ligate these into circles that function as rolling-circle replication substrates and are devoid of exonuclease-sensitive 5′- and 3′-ends so as to maintain their integrity as substrates.[56] That is, catalytic activity is essential (and defines the "native" state) for one part of the replication cycle, but catalytic inactivity is critical ("native") for another part. The existence of conformational isomers of the hairpin ribozyme with different docking–undocking equilibria may then ensure that some RNA molecules are always optimally performing one function while others optimally perform another function without losing all capacity for the former. We hypothesize that such conformational adaptability endows an RNA quasispecies with enhanced functionality in the face of dynamic evolutionary selection criteria (Fig. 9).

Of course, essentially all studies demonstrating multiple functional folded states of RNA have been conducted *in vitro*, and it remains to be seen whether such behaviors will be recapitulated *in vivo*. The one example of obligate folding heterogeneity was observed for a truncated sequence of an artificially selected aptamer, and observations of multiple native states in the hairpin and TG1I ribozymes were made using *in vitro* transcribed or chemically synthesized RNA that had been purified at least once by denaturing polyacrylamide gel electrophoresis. In nature, by contrast, RNA folds as it is transcribed from 5′- to 3′-end, which influences folding in important ways. For example, the segmental co-transcriptional folding of circularly permuted variants of the *Tetrahymena* group I intron was found to yield a higher percentage of natively folded RNA than refolding the entire sequence at once.[134] Transcriptional speed and site-specific pausing were found to be important factors in the folding and function of the FMN riboswitch.[135] The Varkud satellite ribozyme, shown to exhibit folding heterogeneity by smFRET[56] and EMSA, folds into a much narrower range of conformations when purified without denaturation or refolding after transcription.[136] A bioinformatic study found evidence that sequences of natural transcripts are selected for features that promote co-transcriptional folding into the correct native secondary structure.[137] Interestingly, while the hairpin ribozyme was found to fold sequentially under kinetic control during *in vitro* transcription, the relative thermodynamic stability of competing helices was a larger determinant of folding in yeast cells,[138] though kinetic traps can persist *in vivo* if they are sufficiently stable.[139] The greater preference for thermodynamically stable structures *in vivo* could be due to RNA chaperones and other RNA-binding proteins in the cell[39] that may serve to re-equilibrate kinetically trapped species *via* ATP-driven helicase activity or nonspecific stabilization of unfolded intermediates. In the case of CYT-19, an ATP-dependent DEAD-box helicase, there even appears to be some preference for unwinding duplexes within misfolded TG1IRz molecules, perhaps based on compactness of the tertiary structure alone.[140] Another DEAD-box helicase, Mss116, has been shown to stimulate the folding of a group II intron into its near-native state by promoting the
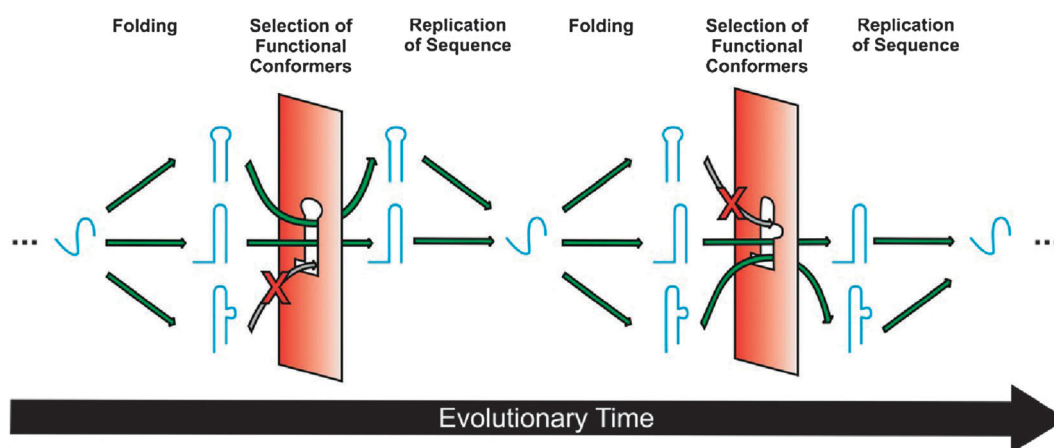

Fig. 9 Schematic representation of a possible adaptive role for conformational quasispecies of RNA under evolutionary pressure. A single RNA sequence (blue) may fold into several stable conformers, or native states, with varying functionality. Changing environmental conditions may impose certain restrictions (red) on the fitness of conformers, but the success of a subset of these conformers will enable the replication of the sequence and the evolutionary survival of all stable (and kinetically accessible) conformers. If conditions are sufficiently variable, there is a clear survival advantage to maintaining a broad quasispecies of RNA folds and functions.

11534 | *Phys. Chem. Chem. Phys.*, 2011, **13**, 11524–11537

This journal is © the Owner Societies 2011

formation of unstable intermediates and dynamic sampling of structures along the folding pathway of the intron.[141,142] While still in their infancy, these studies of co-transcriptional RNA folding and RNA chaperone action suggest that RNA folding behavior should also be studied under conditions as similar as possible to those found in the native cellular environment.

Given the profound kinetic barriers found in some RNAs it seems likely that multiple native states of certain RNAs, either naturally evolved or engineered by humans, will persist *in vivo* even when folded co-transcriptionally in the presence of nucleic acid binding proteins. For natural RNAs, such heterogeneity may depend on the balance between energy requirements to redistribute kinetically trapped species and any (dis)advantages of maintaining a homogeneous over a heterogeneous population of native RNAs. Only *in vivo* testing will determine what roles conformational heterogeneity of RNA may have in living organisms. At least in theory, a shape-shifting RNA quasispecies, as observed *in vitro*, can be expected to impart evolutionary advantages.

Acknowledgements

References

1 T. R. Cech, A. J. Zaug and P. J. Grabowski, *Cell (Cambridge, Mass.)*, 1981, **27**, 487–496.
2 K. Kruger, P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling and T. R. Cech, *Cell (Cambridge, Mass.)*, 1982, **31**, 147–157.
3 C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace and S. Altman, *Cell (Cambridge, Mass.)*, 1983, **35**, 849–857.
4 P. P. Amaral, M. E. Dinger, T. R. Mercer and J. S. Mattick, *Science*, 2008, **319**, 1787–1789.
5 N. G. Walter, S. A. Woodson and R. T. Batey, *Non-protein coding RNAs*, Springer, Heidelberg, 2009.
6 D. Moazed, *Nature*, 2009, **457**, 413–420.
7 T. R. Mercer, M. E. Dinger and J. S. Mattick, *Nat. Rev. Genet.*, 2009, **10**, 155–159.
8 C. C. Mello and D. Conte, *Nature*, 2004, **431**, 338–342.
9 W. C. Winkler, A. Nahvi, A. Roth, J. A. Collins and R. R. Breaker, *Nature*, 2004, **428**, 281–286.
10 N. Toor, K. Rajashankar, K. S. Keating and A. M. Pyle, *Nat. Struct. Mol. Biol.*, 2008, **15**, 1221–1222.
11 P. Nissen, J. Hansen, N. Ban, P. B. Moore and T. A. Steitz, *Science*, 2000, **289**, 920–930.
12 G. A. Prody, J. T. Bakos, J. M. Buzayan, I. R. Schneider and G. Bruening, *Science*, 1986, **231**, 1577–1580.
13 A. C. Forster and R. H. Symons, *Cell (Cambridge, Mass.)*, 1987, **50**, 9–16.
14 A. Hampel and R. Tritz, *Biochemistry*, 1989, **28**, 4929–4933.
15 B. J. Saville and R. A. Collins, *Cell (Cambridge, Mass.)*, 1990, **61**, 685–696.
16 M. M. C. Lai, *Annu. Rev. Biochem.*, 1995, **64**, 259–286.
17 K. Salehi-Ashtiani, A. Luptak, A. Litovchick and J. W. Szostak, *Science*, 2006, **313**, 1788–1792.
18 M. Martick, L. H. Horan, H. F. Noller and W. G. Scott, *Nature*, 2008, **454**, 899–902.
19 C. H. Webb, N. J. Riccitelli, D. J. Ruminski and A. Luptak, *Science*, 2009, **326**, 953.
20 M. de la Pena and I. Garcia-Robles, *RNA*, 2010, **16**, 1943–1950.
21 M. Illangasekare and M. Yarus, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 5470–5475.
22 T. A. Lincoln and G. F. Joyce, *Science*, 2009, **323**, 1229–1232.
23 J. J. Agresti, B. T. Kelly, A. Jaschke and A. D. Griffiths, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 16170–16175.
24 W. G. Scott, *Curr. Opin. Struct. Biol.*, 2007, **17**, 280–286.
25 A. Serganov and D. J. Patel, *Nat. Rev. Genet.*, 2007, **8**, 776–790.
26 A. Korostelev and H. F. Noller, *Trends Biochem. Sci.*, 2007, **32**, 434–441.
27 N. G. Walter, *Mol. Cell*, 2007, **28**, 923–929.
28 H. M. Al-Hashimi and N. G. Walter, *Curr. Opin. Struct. Biol.*, 2008, **18**, 321–329.
29 J. C. Cochrane and S. A. Strobel, *Acc. Chem. Res.*, 2008, **41**, 1027–1035.
30 N. G. Walter and S. Perumal, in *Non-Protein Coding RNAs*, ed. N. G. Walter, S. A. Woodson and R. T. Batey, Springer, Berlin, 2009, vol. 13, pp. 103–127.
31 A. E. Johnson-Buck, S. E. McDowell and N. G. Walter, *Met. Ions Life Sci.*, 2011, **9**, 175–196.
32 W. J. Gartland and N. Sueoka, *Proc. Natl. Acad. Sci. U. S. A.*, 1966, **55**, 948–956.
33 T. Lindahl, A. Adams and J. R. Fresco, *Proc. Natl. Acad. Sci. U. S. A.*, 1966, **55**, 941–948.
34 S. A. Walstrum and O. C. Uhlenbeck, *Biochemistry*, 1990, **29**, 10573–10576.
35 J. A. Esteban, A. R. Banerjee and J. M. Burke, *J. Biol. Chem.*, 1997, **272**, 13629–13639.
36 T. Pan and T. R. Sosnick, *Nat. Struct. Biol.*, 1997, **4**, 931–938.
37 D. M. Chadalavada, S. M. Knudsen, S. Nakano and P. C. Bevilacqua, *J. Mol. Biol.*, 2000, **301**, 349–367.
38 D. K. Treiber and J. R. Williamson, *Curr. Opin. Struct. Biol.*, 1999, **9**, 339–345.
39 R. Russell, *Front. Biosci.*, 2008, **13**, 1–20.
40 S. V. Solomatin, M. Greenfeld, S. Chu and D. Herschlag, *Nature*, 2010, **463**, 681–684.
41 P. B. Sigler, *Annu. Rev. Biophys. Bioeng.*, 1975, **4**, 477–527.
42 T. Rastogi, T. L. Beattie, J. E. Olive and R. A. Collins, *EMBO J.*, 1996, **15**, 2820–2825.
43 J. H. Cate, A. R. Gooding, E. Podell, K. Zhou, B. L. Golden, C. E. Kundrot, T. R. Cech and J. A. Doudna, *Science*, 1996, **273**, 1678–1685.
44 A. R. Ferre-D'Amare, K. H. Zhou and J. A. Doudna, *Nature*, 1998, **395**, 567–574.
45 P. B. Rupert and A. R. Ferre-D'Amare, *Nature*, 2001, **410**, 780–786.
46 D. J. Klein and A. R. Ferre-D'Amare, *Science*, 2006, **313**, 1752–1756.
47 J. Lipfert, J. Ouellet, D. G. Norman, S. Doniach and D. M. Lilley, *Structure (London)*, 2008, **16**, 1357–1367.
48 L. J. W. Murray, W. B. Arendall, D. C. Richardson and J. S. Richardson, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 13904–13909.
49 J. S. Richardson, B. Schneider, L. W. Murray, G. J. Kapral, R. M. Immormino, J. J. Headd, D. C. Richardson, D. Ham, E. Hershkovits, L. D. Williams, K. S. Keating, A. M. Pyle, D. Micallef, J. Westbrook and H. M. Berman, *RNA*, 2008, **14**, 465–481.
50 P. Brion and E. Westhof, *Annu. Rev. Biophys. Biomol. Struct.*, 1997, **26**, 113–137.
51 D. Porschke and M. Eigen, *J. Mol. Biol.*, 1971, **62**, 361–381.
52 D. H. Turner, N. Sugimoto and S. M. Freier, in *Nucleic Acids*, ed. W. Saenger, Springer, Berlin, 1990, pp. 201–227.
53 D. Herschlag, *J. Biol. Chem.*, 1995, **270**, 20871–20874.
54 X. W. Zhuang, L. E. Bartley, H. P. Babcock, R. Russell, T. J. Ha, D. Herschlag and S. Chu, *Science*, 2000, **288**, 2048–2051.
55 X. W. Zhuang, H. Kim, M. J. B. Pereira, H. P. Babcock, N. G. Walter and S. Chu, *Science*, 2002, **296**, 1473–1476.
56 M. J. B. Pereira, E. N. Nikolova, S. L. Hiley, D. Jaikaran, R. A. Collins and N. G. Walter, *J. Mol. Biol.*, 2008, **382**, 496–509.
57 J. Gralla and C. Delisi, *Nature*, 1974, **248**, 330–332.
58 M. Zuker, *Science*, 1989, **244**, 48–52.
59 J. S. Mccaskill, *Biopolymers*, 1990, **29**, 1105–1119.
60 D. H. Mathews, *RNA*, 2004, **10**, 1178–1190.
61 N. R. Markham and M. Zuker, *Nucleic Acids Res.*, 2005, **33**, W577–W581.
62 A. V. Korennykh, M. J. Plantinga, C. C. Correll and J. A. Piccirilli, *Biochemistry*, 2007, **46**, 12744–12756.

63 Z. Huang, W. M. Pei, Y. Han, S. Jayaseelan, A. Shekhtman, H. Shi and L. Niu, *Nucleic Acids Res.*, 2009, **37**, 4022–4032.

64 M. E. Altuntop, C. T. Ly and Y. H. Wang, *Biophys. J.*, 2010, **99**, 3002–3009.

65 M. A. Ditzler, D. Rueda, J. J. Mo, K. Hakansson and N. G. Walter, *Nucleic Acids Res.*, 2008, **36**, 7088–7099.

66 J. M. Buzayan, W. L. Gerlach and G. Bruening, *Proc. Natl. Acad. Sci. U. S. A.*, 1986, **83**, 8859–8862.

67 J. M. Buzayan, W. L. Gerlach, G. Bruening, P. Keese and A. R. Gould, *Virology*, 1986, **151**, 186–199.

68 N. G. Walter, C. Y. Huang, A. J. Manzo and M. A. Sobhy, *Nat. Methods*, 2008, **5**, 475–489.

69 D. Rueda, G. Bokinsky, M. M. Rhodes, M. J. Rust, X. Zhuang and N. G. Walter, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 10066–10071.

70 S. Liu, G. Bokinsky, N. G. Walter and X. Zhuang, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 12634–12639.

71 B. Okumus, T. J. Wilson, D. M. J. Lilley and T. Ha, *Biophys. J.*, 2004, **87**, 2798–2806.

72 G. Bokinsky, D. Rueda, V. K. Misra, M. M. Rhodes, A. Gordus, H. P. Babcock, N. G. Walter and X. Zhuang, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 9302–9307.

73 S. A. Mortimer and K. M. Weeks, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 15622–15627.

74 D. N. Ermolenko, Z. K. Majumdar, R. P. Hickerson, P. C. Spiegel, R. M. Clegg and H. F. Noller, *J. Mol. Biol.*, 2007, **370**, 530–540.

75 D. N. Ermolenko, P. C. Spiegel, Z. K. Majumdar, R. P. Hickerson, R. M. Clegg and H. F. Noller, *Nat. Struct. Mol. Biol.*, 2007, **14**, 493–497.

76 J. B. Munro, R. B. Altman, N. O'Connor and S. C. Blanchard, *Mol. Cell*, 2007, **25**, 505–517.

77 P. V. Cornish, D. N. Ermolenko, H. F. Noller and T. Ha, *Mol. Cell*, 2008, **30**, 578–588.

78 Z. Huang, W. Pei, S. Jayaseelan, H. Shi and L. Niu, *Biochemistry*, 2007, **46**, 12648–12655.

79 R. L. Baldwin and G. D. Rose, *Trends Biochem. Sci.*, 1999, **24**, 26–33.

80 K. A. Dill, S. B. Ozkan, M. S. Shell and T. R. Weikl, *Annu. Rev. Biophys.*, 2008, **37**, 289–316.

81 S. Dalal and L. Regan, *Protein Sci.*, 2000, **9**, 1651–1659.

82 B. W. Matthews, *Adv. Protein Chem.*, 1995, **46**, 249–278.

83 S. M. Freier, R. Kierzek, J. A. Jaeger, N. Sugimoto, M. H. Caruthers, T. Neilson and D. H. Turner, *Proc. Natl. Acad. Sci. U. S. A.*, 1986, **83**, 9373–9377.

84 N. R. Markham and M. Zuker, *Methods Mol. Biol.*, 2008, **453**, 3–31.

85 D. L. Minor, Jr. and P. S. Kim, *Nature*, 1994, **371**, 264–267.

86 T. Kiefhaber, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 9029–9033.

87 K. Kamagata, Y. Sawano, M. Tanokura and K. Kuwajima, *J. Mol. Biol.*, 2003, **332**, 1143–1153.

88 P. A. Jennings, B. E. Finn, B. E. Jones and C. R. Matthews, *Biochemistry*, 1993, **32**, 3783–3789.

89 F. X. Schmid and H. Blaschek, *Eur. J. Biochem.*, 1981, **114**, 111–117.

90 O. Flomenbom, K. Velonia, D. Loos, S. Masuo, M. Cotlet, Y. Engelborghs, J. Hofkens, A. E. Rowan, R. J. Nolte, M. Van der Auweraer, F. C. de Schryver and J. Klafter, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 2368–2372.

91 H. P. Lu, L. Y. Xun and X. S. Xie, *Science*, 1998, **282**, 1877–1882.

92 H. Yang, G. Luo, P. Karnchanaphanurach, T. M. Louie, I. Rech, S. Cova, L. Xun and X. S. Xie, *Science*, 2003, **302**, 262–266.

93 L. Edman, Z. Foldes-Papp, S. Wennmalm and R. Rigler, *Chem. Phys.*, 1999, **247**, 11–22.

94 A. M. van Oijen, P. C. Blainey, D. J. Crampton, C. C. Richardson, T. Ellenberger and X. S. Xie, *Science*, 2003, **301**, 1235–1238.

95 Q. Xue and E. S. Yeung, *Nature*, 1995, **373**, 681–683.

96 W. H. Tan and E. S. Yeung, *Anal. Chem.*, 1997, **69**, 4242–4248.

97 D. B. Craig, E. A. Arriaga, J. C. Y. Wong, H. Lu and N. J. Dovichi, *J. Am. Chem. Soc.*, 1996, **118**, 5245–5253.

98 D. B. Craig and N. J. Dovichi, *Can. J. Chem.*, 1998, **76**, 623–626.

99 N. Handa, P. R. Bianco, R. J. Baskin and S. C. Kowalczykowski, *Mol. Cell*, 2005, **17**, 745–750.

100 P. R. Bianco and S. C. Kowalczykowski, *Nature*, 2000, **405**, 368–372.

101 G. Baldini, F. Cannone and G. Chirico, *Science*, 2005, **309**, 1096–1100.

102 F. Cannone, S. Bologna, B. Campanini, A. Diaspro, S. Bettati, A. Mozzarelli and G. Chirico, *Biophys. J.*, 2005, **89**, 2033–2045.

103 G. Baldini, F. Cannone, G. Chirico, M. Collini, B. Campanini, S. Bettati and A. Mozzarelli, *Biophys. J.*, 2007, **92**, 1724–1731.

104 R. Polakowski, D. B. Craig, A. Skelley and N. J. Dovichi, *J. Am. Chem. Soc.*, 2000, **122**, 4853–4855.

105 I. Franklin, *Conservation Biology: An Evolutionary-Ecological Perspective*, ed. M. E. Soulé and B. A. Wilcox, Sinauer Associates, Sunderland, Massachusetts, 1980, pp. 135–150.

106 D. H. Reed and R. Frankham, *Conserv. Biol.*, 2003, **17**, 230–237.

107 A. C. Allison, *Exp. Parasitol.*, 1957, **6**, 418–447.

108 D. L. Robertson, B. H. Hahn and P. M. Sharp, *J. Mol. Evol.*, 1995, **40**, 249–259.

109 M. Dean, M. Carrington, C. Winkler, G. A. Huttley, M. W. Smith, R. Allikmets, J. J. Goedert, S. P. Buchbinder, E. Vittinghoff, E. Gomperts, S. Donfield, D. Vlahov, R. Kaslow, A. Saah, C. Rinaldo, R. Detels and S. J. O'Brien, *Science*, 1996, **273**, 1856–1862.

110 M. Uyenoyama, *Pesticide Resistance: Strategies and Tactics for Management*, National Academy Press, Washington, DC, 1986.

111 B. R. Lyon and R. Skurray, *Microbiol. Rev.*, 1987, **51**, 88–134.

112 P. J. Choi, L. Cai, K. Frieda and X. S. Xie, *Science*, 2008, **322**, 442–446.

113 A. P. Wolffe and M. A. Matzke, *Science*, 1999, **286**, 481–486.

114 G. J. Hannon, *Nature*, 2002, **418**, 244–251.

115 J. Han, Y. Lee, K. H. Yeom, J. W. Nam, I. Heo, J. K. Rhee, S. Y. Sohn, Y. Cho, B. T. Zhang and V. N. Kim, *Cell (Cambridge, Mass.)*, 2006, **125**, 887–901.

116 A. Grimson, K. K. Farh, W. K. Johnston, P. Garrett-Engele, L. P. Lim and D. P. Bartel, *Mol. Cell*, 2007, **27**, 91–105.

117 A. Nahvi, N. Sudarsan, M. S. Ebert, X. Zou, K. L. Brown and R. R. Breaker, *Chem. Biol.*, 2002, **9**, 1043.

118 A. S. Mironov, I. Gusarov, R. Rafikov, L. E. Lopez, K. Shatalin, R. A. Kreneva, D. A. Perumov and E. Nudler, *Cell (Cambridge, Mass.)*, 2002, **111**, 747–756.

119 W. Winkler, A. Nahvi and R. R. Breaker, *Nature*, 2002, **419**, 952–956.

120 W. C. Winkler, S. Cohen-Chalamish and R. R. Breaker, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 15908–15913.

121 M. Mandal and R. R. Breaker, *Nat. Rev. Mol. Cell Biol.*, 2004, **5**, 451–463.

122 B. R. Graveley, *Trends Genet.*, 2001, **17**, 100–107.

123 B. J. Blencowe, *Cell (Cambridge, Mass.)*, 2006, **126**, 37–47.

124 L. Cai, N. Friedman and X. S. Xie, *Nature*, 2006, **440**, 358–362.

125 Y. Taniguchi, P. J. Choi, G. W. Li, H. Chen, M. Babu, J. Hearn, A. Emili and X. S. Xie, *Science*, 2010, **329**, 533–538.

126 E. Dekel and U. Alon, *Nature*, 2005, **436**, 588–592.

127 M. Eigen, J. Mccaskill and P. Schuster, *J. Phys. Chem.*, 1988, **92**, 6881–6891.

128 P. Schuster, *Origins Life Evol. Biosphere*, 1993, **23**, 373–391.

129 C. K. Biebricher and M. Eigen, *Current Topics in Microbiology and Immunology*, 2006, **299**, 1–31.

130 A. S. Lauring and R. Andino, *PLoS Pathog.*, 2010, **6**, e1001005.

131 J. P. Anderson, R. Daifuku and L. A. Loeb, *Annu. Rev. Microbiol.*, 2004, **58**, 183–205.

132 J. J. Bull, L. A. Meyers and M. Lachmann, *PLoS Comput. Biol.*, 2005, **1**, e61.

133 E. A. Schultes and D. P. Bartel, *Science*, 2000, **289**, 448–452.

134 S. L. Heilman-Miller and S. A. Woodson, *RNA*, 2003, **9**, 722–733.

135 J. K. Wickiser, W. C. Winkler, R. R. Breaker and D. M. Crothers, *Mol. Cell*, 2005, **18**, 49–60.

136 M. J. Pereira, V. Behera and N. G. Walter, *PLoS One*, 2010, **5**, e12953.

137 I. M. Meyer and I. Miklos, *BMC Mol. Biol.*, 2004, **5**, 10.

138 E. M. Mahen, J. W. Harger, E. M. Calderon and M. J. Fedor, *Mol. Cell*, 2005, **19**, 27–37.

139 E. M. Mahen, P. Y. Watson, J. W. Cottrell and M. J. Fedor, *PLoS Biol.*, 2010, **8**, e1000307.

140 H. Bhaskaran and R. Russell, *Nature*, 2007, **449**, 1014–1018.

141 O. Fedorova, A. Solem and A. M. Pyle, *J. Mol. Biol.*, 2010, **397**, 799–813.

142 K. S. Karunatilaka, A. Solem, A. M. Pyle and D. Rueda, *Nature*, 2010, **467**, 935–939.

143 F. Guo, A. R. Gooding and T. R. Cech, *Mol. Cell*, 2004, **16**, 351–362.