PCCP

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: K. McCluskey and C. Penedo, *Phys. Chem. Chem. Phys.*, 2017, DOI: 10.1039/C6CP08798A.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/pccp

View Article Online
DOI: 10.1039/C6CP08798A

Physical Chemistry Chemical Physics

PERSPECTIVE

An integrated perspective on RNA aptamer ligand-recognition models: clearing muddy waters

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

K. McCluskey[a,†] and J.C. Penedo[a,b,†]

Riboswitches are short RNA motifs that sensitively and selectively bind cognate ligands to modulate gene expression. Like protein receptor-ligand pairs, their binding dynamics are traditionally categorized as following one of two paradigmatic mechanisms: conformational selection and induced fit. In conformational selection, ligand binding stabilizes a particular state already present in the receptor's dynamic ensemble. In induced fit, ligand-receptor interactions enable the system to overcome the energetic barrier into a previously inaccessible state. In this article, we question whether a polarized division of RNA binding mechanisms truly meets the conceptual needs of the field. We will review the history behind this classification of RNA-ligand interactions, and the way induced fit in particular has been rehabilitated by single-molecule studies of RNA aptamers. We will highlight several recent results from single-molecule experimental studies of riboswitches that reveal gaps or even contradictions between common definitions of the two terms, and we will conclude by proposing a more robust framework that considers the range of RNA behaviors unveiled in recent years as a reality to be described, rather than an increasingly unwieldy set of exceptions to the traditional models.

Introduction

Our concept of the role of mRNA in biological processes, particularly in bacteria, has exploded in the past two decades. Far from being a mere messenger molecule, the untranslated regions of mRNA are rich in control sequences that modulate gene expression and splicing, and small fragments (< 30 nt) are active in gene silencing and the adaptive immune system.[1-4] The high selectivity of even small RNA molecules for their cognate sequence underpins their adaptive immune role, and their ability to fold into stable three-dimensional (tertiary) structures enables them to interact dynamically with the cell's protein machinery for genetic control.

One class of active RNAs, the riboswitches, modulate gene expression by selectively binding a particular small-molecule metabolite associated with the downstream gene (as a product, byproduct, substrate, etc.). Ligand binding to the aptamer domain shifts the equilibrium between two competing structures of the adjacent expression platform, usually by stabilizing a stem-loop often referred to as the 'switching stem'. This determines the fate of that sequence's interactions with transcription or translation machinery (Fig. 1). Riboswitches can be found modulating the interchange between transcription termination and anti-termination stems, controlling access to the Shine-Dalgarno translation initiation sequence, and controlling splicing to define the open reading frame or decide between gene expression and breakdown of the RNA. Artificial, nonregulatory aptamers have also been synthesized, and were being used as models of biomolecule folding before biological riboswitches were discovered.[5-7]

The mechanisms by which the aptamer domain binds its ligand are the subject of intense research.[8] Despite its minimal primary structure, consisting of only four nucleobases, RNA can form a wide array of secondary structures, forming strong interactions both between the canonical base pairs that define the DNA double helix, and between 'noncanonical' pairs. These base-paired regions can self-organize into complex tertiary structures including compact multi-helix junctions and stacks of helices. Base-pairing interactions between unpaired nucleotides or backbone groups pin these three-dimensional structures together. Such structures can present binding pockets for a wide array of organic and ionic ligands, and the ligands in turn provide additional contacts that stabilize the 3D fold. By incorporating specific and nonspecific binding sites for water and environmental cations, particularly Mg^{2+} and K^{+}, RNA sensors with nM affinity[9] have been found even for negatively charged ligands like fluoride.[10]

The parallels between aptamer-ligand interactions in RNA and ligand-receptor binding in proteins have been obvious to the community since its inception.[11, 12] Our current interpretation of RNA-ligand interaction models implicitly reflects some links with different stages of the study of protein-ligand interactions, and many concepts from the study of protein-ligand complexes have been adopted to describe RNA ligand-binding mechanisms. In this review, we explore how the application of single-molecule techniques to the study RNA-ligand interactions has revealed a diverse range of recognition mechanisms, discrepancies, and gaps that cannot be easily accommodated by the traditional conformational selection (CS) and induced-fit (IF) models. By examining the most recent experimental data, we propose a new framework to classify RNA-ligand interactions that moves beyond the classical CS/IF duality and clarifies existing discrepancies and gaps.

[a] Department of Physics and Astronomy, University of St. Andrews, St. Andrews, UK KY16 9SS. E-mail: km855@st-andrews.ac.uk
[b] Biomolecular Sciences Research Complex, University of St. Andrews, St. Andrews, UK KY16 9SS. E-mail: jcp10@st-andrews.ac.uk
† Corresponding authors.

Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

View Article Online
DOI: 10.1039/C6CP08798A

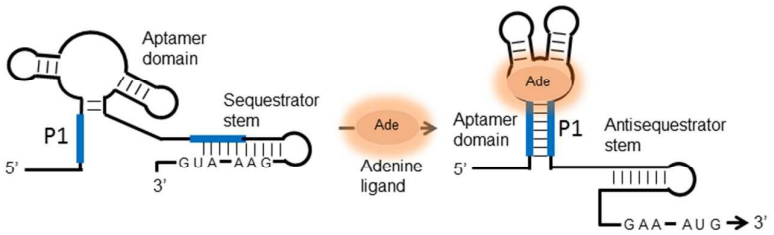

Fig. 1 An illustration of riboswitch function using an adenine-sensing aptamer. A riboswitch is composed of an aptamer domain—here, the three-helix junction at the left of each panel—and an expression platform, at right, whose sequence partially overlaps that of the aptamer domain (blue bars). In the absence of adenine (left), the P1 stem of the aptamer is partially unfolded, and the expression platform forms a sequestrator stem that includes the START codon (AUG) of the adjacent gene. In this conformation, the START codon is inaccessible and the gene is 'off'. Adenine binding (right panel) triggers the rearrangement of both the aptamer and the expression platform. In addition to a tertiary structure change, represented by the movement of the remaining stems, ligand binding stabilizes P1, forcing the expression platform to form an alternative, 'antisequestrator' configuration in which the START codon is exposed. Since the stability of the P1 helix determines the expression platform conformation, and hence the regulatory outcome, it is often called the 'switching stem'.

A historical perspective: mechanisms of protein-receptor binding

Proteins were initially conceived as rigid bodies, and the earliest model of ligand binding, the lock and key model, assumed a steric fit between correspondingly-shaped binding pockets and ligands.[13] The reaction rate was determined entirely by the encounter frequency. Koshland, Némethy, and Filmer proposed in 1966 that protein receptors do not generally match their ligands in shape, but rather than ligands induce a structural change in their receptor upon binding[14]. This induced fit theory acknowledged the growing evidence that proteins are dynamic objects.

At the same time, Monod, Wyman, and Changeux[15] proposed that the subunits of an enzyme could exist in two conformations with differing ligand affinities. The recent energy landscape picture of protein folding,[16-19] in which protein dynamics are depicted as diffusion on a high-dimensional potential energy surface, established the idea that proteins can visit a pre-existing ensemble of conformations, including the 'native' structure, even in the absence of the ligand. Conformational selection proposes that the ligand selects and stabilizes a high-affinity state out of this pre-existing ensemble.

Given this historical evolution, it is unsurprising that some authors argue that CS should replace IF as the model that best describes motion on the free energy landscape (FEL).[17, 20] Especially in older papers, it is often assumed IF implies a rigid protein, while CS implies or is even synonymous with the existence of a conformational ensemble.[17, 21-26] Others have argued that while diffusion on the FEL in configuration space is the most realistic model of protein folding, CS and IF describe not the FEL itself, but two different ways the ligand can *reshape* the FEL.[27-29] In this interpretation, CS is associated with constraining diffusion across the landscape, and IF is associated with 'forcing and tilting', adjusting the effective height of energetic barriers to local minima[30] and enabling transitions to previously energetically inaccessible subspaces.[17]

Each binding mechanism, CS or IF, can be defined in one of two ways (Fig. 2). Changes to the FEL underpin the 'kinetic' or 'energetic' definitions. Alternatively, the 'structural' or 'ordering' definition simply asks whether the protein can fold into its native state without the ligand.

Classic examples of CS according to the structural definition exhibit a clear shift in the equilibrium between multiple pre-existing states in the presence of ligand. Phosphotriesterase,[31] thrombin,[32] and glucose/galactose binding protein (GGBP)[19] all have two or three identifiable structures in equilibrium. Their ligand-binding mechanism has been described as CS because the ligands shift the majority of the population into the native state. In the study of RNA, the HIV transactivation response region (TAR) has also been described as CS based on the presence of the native fold in the absence of its ligand, the Tat protein.[33, 34]

The encounter complex between an enzyme and its substrate—a ligand-bound state that precedes the chemically active configuration—is often interpreted as an instance of 'structural' IF. The PEPCK enzyme was classified as IF when the encounter complex was crystallized.[35] The LAO (lysine-arginine-ornithine) enzyme is a more complex case. Simulations have shown that it can reach the encounter complex via multiple paths with varying ligand arrival times. Interestingly, LAO folding into the native state only occurs with the ligand bound, suggesting an IF-like enzymatic step, whereas the formation of the encounter complex could be equally assigned as CS or IF.[18, 36]

Crystal structures have played an ambiguous role in the assignation of recognition models. Lidded proteins like PEPCK[35] were initially crystallized only in the ligand-bound state, as were the first RNA aptamers.[37-39] It was inferred that the bound structure could only exist in the presence of the ligand, and the mechanism was therefore IF. Even assuming the crystal represents the entire accessible ensemble, this interpretation has flaws. For lidded proteins, the ligand presumably needs to interact with some open conformation in order to enter the binding pocket, which is usually described as 'binding before folding'.[35] Furthermore, the assumption that a short-lived closed state was absent from the unfolded ensemble was unwarranted.[40] Both the encounter

View Article Online
DOI: 10.1039/C6CP08798A
Physical Chemistry Chemical Physics PERSPECTIVE

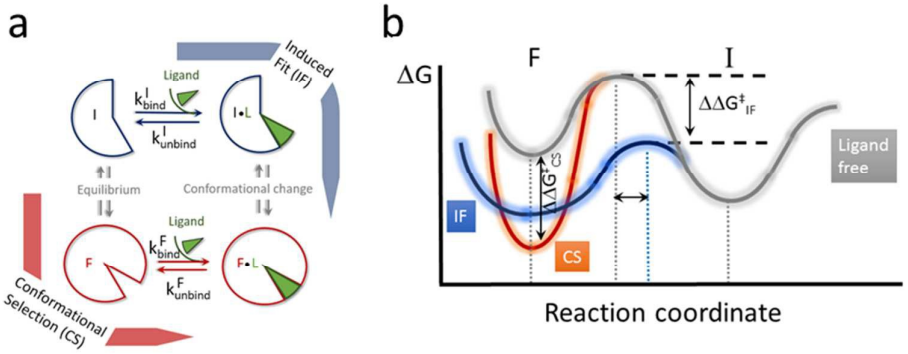

Fig. 2 Definitions of conformational selection (CS) and induced fit (IF). (a) In the 'structural' sense, CS and IF are distinguished by the ordering of binding and folding events. In CS (lower left schematics, indicated in red), an unfolded state or a manifold of them, depicted by wedge I in the upper left of the panel, exist in equilibrium with the native state, F, even in the absence of ligand. The ligand binds to F with higher affinity than to I, and the F*L complex is more stable than the unliganded state F, leading to a population shift from I to F. In IF (upper right, indicated in blue), the ligand interacts with state I and facilitates folding into the final F*L complex. The unliganded F state is rarely visited. (b) In the kinetic or energetic picture, the mechanisms are distinguished by their effect on the energy landscape. The schematic depicts the free energy ΔG as a function of some generalized reaction coordinate (extension of a molecule under force, separation of a pair of FRET dyes, etc.). A simplified landscape for a ligand-free aptamer, showing local minima corresponding to the I and F states in (a) with a barrier between them, is shown in grey. In the absence of ligand, I is more energetically favorable than F. The energy difference between them and the barrier height define their equilibrium populations and transition rates. Ligand binding via CS does not affect the rate of transitions from I to F, but it does lower the energy of the F state, stabilizing it relative to I and decreasing the rate of unfolding transitions (red curve). Ligand binding via IF also stabilizes the F state, but its most pronounced effect is to lower and shift the energy barrier between the states, enhancing the rate of transitions into F (blue curve).

complex and the kinetics of the ligand-free state[26] are, in fact, crucial to understanding the binding mechanism.

The structural definitions of CS and IF do raise questions about the kinetic effects of early or late ligand binding. A wide variety of behaviors are possible based on the balance of binding and folding rates and other environmental factors. Zhou[18] and Weikl and von Deuster[36] have separately shown that if the binding mechanism is defined purely by the order of binding and folding, then the mechanism can be changed by altering the encounter frequency relative to the rate of intrinsic protein fluctuations. Other simulations have shown that the range of the receptor-ligand interaction, its strength, and the size of the ligand all affect the mechanism along with the folding and binding rates.[41]

One implication of the energy landscape picture of protein folding is that multiple paths into the native state are possible. Different paths may include a different ordering of folding and binding and hence, entail a different mechanism by that definition. Hammes and coworkers have modeled the flux through each possible pathway into the native state for ligand binding to dihydrofolate reductase (DHFR) and flavodoxin,[42] and Daniels et al. have done the same for RNase P.[43] For DHFR and flavodoxin, each pathway was either clearly IF or clearly CS in terms of the ordering of folding and binding, but RNase P had hybrid pathways available in the sense that each of its two ligands could use a different mechanism for two different structural transitions. In all three cases, the flux through each pathway changed depending on the ligand concentration and the concentration of environmental cations like Mg^{2+}. Importantly, every protein used every available pathway, and hence every binding mechanism.

Despite the conceptual flexibility arising from a structural description of binding models, many systems exhibit mechanisms that cannot be classified as CS or IF in the structural sense. A generalized two-step mechanism sometimes termed "extended CS" involves a CS step followed by ligand binding and a subsequent induced change of varying magnitude.[18, 20, 21, 27, 29, 36, 44] As an example, the maltose-binding protein (MBP) binds maltose in two steps. The 'closed' (native) state only occurs when maltose is present, leading to a designation of IF.[27] However, the preceding transition from the open to the 'semi-closed' state appears to be a case of structural CS: the semi-closed state can form in the absence of maltose, but is significantly stabilized by the ligand. The question of whether CS should be defined relative to the selection of the native state or of a ligand-competent state[5, 45] suggests that 'extended CS' entailing a subsequent IF step could even be seen as the canonical form of CS.

The intrinsically-disordered protein (IDP) pKID fits both definitions equally.[46] The ligand forms a network of weak nonnative contacts with the protein early in the folding process, as usual for IDPs. Rather than induce folding into the native state directly, these bonds constrain the conformational search to progressively smaller regions of conformation space. The native state is still eventually reached via a random search, but with its range and rates both influenced by the ligand.

The alternative, kinetic definitions of CS and IF do emphasize the kinetic effect of ligand binding. Ligand binding ought to promote folding into the native state, increasing the folding rate, k_{fwd}, with ligand concentration. CS ought to decrease the unfolding rate k_{rev} by stabilizing the native state. In ensemble measurements, it is common to extend this microscopic rate dependence to the effective reaction rate, k_{eff}, which behaves similarly in single-step reactions, and assign IF or CS accordingly. Alkaline phosphatase[47] and adenylate kinase[48] have been classified as CS in this way. Glucokinase has been studied multiple times, and conflicting results indicated that it could bind glucose using either mechanism.[49-51]

View Article Online
DOI: 10.1039/C6CP08798A

Vogt and di Cera recently challenged kinetic identification of IF, and suggested that CS may in fact be overwhelmingly dominant.[40, 52] Their careful analysis of the rate equations resulting from multi-step reactions showed that a CS mechanism, defined in the structural sense, can describe almost any k_{eff} vs. ligand curve given the appropriate combination of microscopic binding, unbinding, folding, and unfolding rates. They showed that the ligand dependence of proteins thought to be examples of IF, including glucokinase,[50, 51] can be equally well analyzed as examples of CS with different microscopic reaction rates.[52]

Since IF behavior cannot be uniquely identified using k_{eff} except in certain limits,[53] and given the large range of behaviors a CS mechanism can realize given different microscopic rate constants, Vogt and di Cera suggested that CS is in fact the predominant mechanism of ligand-receptor binding in proteins. However, while their results do reveal a genuine weakness of the ensemble treatment of ligand-binding kinetics, they do not exclude the possibility of structural IF (ligand binding preceding folding, as in the case of encounter complex formation[35]), nor do they address the question of what kinetic roles the ligand could play in such cases.

Characteristics of ligand binding by intrinsically disordered proteins and RNA aptamers

Intrinsically Disordered Proteins (IDPs) have prompted the development of entirely new ligand-binding models.[54] IDPs lack a clearly-defined folded state. They typically bind multiple substrates and participate in multiple chemical pathways. They can be found in any functional niche, but are particularly prevalent in eukaryotic cell signaling pathways.[55] Their versatility has led to several proposals for novel binding mechanisms, and they raise questions of relevance to ligand binding by RNA aptamers.

Considering IDPs' large structural ensemble, it has been hypothesized that IDPs universally use CS to recognize their substrates.[56] However, we saw that pKID, above, performs a 'constrained' conformational search after non-native contacts with the ligand have been formed.[46] Shoemaker et al. have described this process in more detail.[54] The 'fly-fishing' model relies on an initial, weakly-bound encounter complex between the IDP and its substrate. These ligand contacts enhance k_{fwd}, but not by structuring the native state, since the percent of native contacts is generally low. Instead, they decrease the entropic cost to the protein of searching the more-collapsed configurational subspace. Fly-fishing presents IF-like binding order and an IF-like effect on the folding dynamics, but via a mechanism of constrained search rather than coordinated folding.

In 2004-2006, several groups uncovered a set of secondary structure-forming motifs that are often found in IDPs.[57-60] These small regions of secondary structure increase the IDP's ligand affinity sufficiently for weak, initial binding events,[61] despite the charge and low hydrophobicity of the bulk of the IDP that prevent the native tertiary structures from stably forming without the network of ligand contacts.[59] These Molecular Recognition Elements, or MoREs,[57] broadly enhance the IDP's affinity for many potential substrates while the rest of the protein is unstructured. As examples, MoREs have been identified in MAP2c, capastatin,[58] and tumor suppressor p53.[57] They form a 'lure', so to speak, that provides a binding element to start the fly-fishing process.

If this is reminiscent of extended CS—after all, the ligand must select an intact MoRE from an ensemble[18, 36, 41, 57, 62]—it also highlights the similarities between the IDP and riboswitch folding problems. Early studies of crystallized RNA aptamers noted a potential analogy between Mg^{2+} ion interactions with the binding pocket residues and the hydrophobic core of ordered proteins.[12] A more accurate analogy would be between flexible, polyanionic RNA and charged, hydrophilic IDPs.[56, 60] In this sense, MoREs resemble the Mg^{2+}-organized conformations of riboswitches that act as binding platforms for their ligands.

These preorganized elements exclude some of the tertiary and RNA-ligand contacts found in the native structure, but they act as initial recognition elements for the ligand. For example, the 'A-comp' state of the *pbuE* aptamer includes docking of stems P2 and P3, but the binding pocket and switching stem P1 are unstructured.[63] In contrast, the P2 and P3 stems of the structurally-similar *xpt* guanine aptamer can dock without Mg^{2+} or guanine, due to the early formation of a longer and hence more stable P1.[64] The organization of P1 is also thought to create a three-helix 'interaction module' for the *thiM* thiamine phyrophosphate (TPP) riboswitch, even though the binding site itself is divided between two further stem-loops, not situated in the interhelical junction.[65]

Both fly-fishing and extended CS are reminiscent of this pattern often described in riboswitches, in which ligand binding by either mechanism is followed by local ligand-induced rearrangement of the nucleotides that form the binding site. When available, comparisons between ligand-free and ligand-bound NMR and crystal structures are particularly useful for identifying these small shifts, for example in the lysine,[66] adenine,[67] and guanine[64, 68, 69] aptamers. And just as IDPs require protein contacts with the ligand to fold, riboswitches *in vivo* are not subject to ligand-activated 'switching' per se, but rather ligand-directed *folding* that competes with processes like transcription.[5]

Accordingly, some authors have remarked on the need for kinetic as well as structural information when determining the binding mechanism. Sullivan and Holyoak have pointed out that the presence of the native structure in the ligand-free manifold does not imply CS, as ligand contacts in the encounter complex can still enhance (induce) folding.[35] Similarly, Boehr, Nussinov, and Wright have argued that encounter complex formation (i.e. ligand binding to the non-native state) does not imply induced fit, either, since the folding rate can still be dominated by the protein's intrinsic (ligand-independent) dynamics.[20] The ribose-binding protein (RBP) is a concrete picture of this ambiguity: it has been crystallized in both a ligand-free closed state and a ligand-bound semiopen state.[70]

The energy landscape picture of protein folding is the key to integrating the kinetic definitions of binding mechanisms with the notion of binding and folding order.[16, 28] Both RNA and IDPs are described as having a 'rugged' FEL.[24, 25, 69] The many, shallow energy minima correspond to kinetic traps in the folding pathway and potentially misfolded states. In IDPs, the energetic ruggedness is due to the lack of hydrophobic core and stable interactions in the absence of the ligand. In RNA, the simple primary structure of the molecule, defined by four nucleobases with a small number of potential interactions between them, leads to very many potential secondary structures in the absence of ligand or coordinating bonds with, e.g., H_2O or Mg^{2+}.

Li et al. have noted that the 'forcing and tilting' of the FEL that characterizes IF is advantageous when a folding pathway has many potential traps in it.[71] Optical force experiments have experimentally demonstrated the same for TAR RNA.[71] Ligand binding and, in the case of riboswitches, Mg^{2+} coordination, affect

View Article Online
DOI: 10.1039/C6CP08798A

folding rates and tilt the energy landscape, hence affecting the choice between CS and IF.[24, 70]

To generalize, we arrive at the impression that conformational flexibility in ordered proteins tends to imply CS, at least according to the structural definition, by allowing the protein access to the native conformation in a ligand-free state.[25, 70] IDPs achieve multiple substrate affinity through extreme conformational flexibility, at the expense of a rugged energy landscape and reduced selectivity. Ligand-free RNA aptamers, by contrast, are structurally flexible and energetically rugged while still being highly selective against ligand analogues. Flexibility and early binding to a variety of structures play a positive role in establishing the long-range tertiary contacts of the native state, suggesting that IF could be a more general mechanism in RNA.[30, 72]

Insights into aptamer-ligand interactions from single-molecule experiments

Single-molecule techniques provide access to free energy profiles and microscopic transition rates directly rather than via k_{eff}.[28] Single-molecule fluorescence can access k_{fwd} and k_{rev} directly. By strategically labeling the receptor, the ligand, or both, and resolving the dynamics of individual molecules via FRET rather than relying on the ensemble average, the ligand dependence of the rate of each individual step can be measured explicitly.[73, 74] While k_{eff} increasing with ligand concentration is not a reliable signature of IF, microscopic k_{fwd} doing the same *is* unambiguous proof that ligand interactions enhance the folding rate. Likewise for the stabilization of the folded state by the ligand, leading to a decrease in microscopic k_{rev}.

The *metX* SAM-II aptamer may be an example of conformational selection according to both ordering and kinetic definitions. *MetX* is a small aptamer whose single stem-loop and tail form a pseudoknot and two double-stranded regions when the ligand is bound in the core (Fig. 3a). Haller et al. labeled the *metX* riboswitch on the loop and the tail to monitor the formation of the pseudoknot. They found that in the absence of SAM or cations, the aptamer could transiently visit a high-FRET state characteristic of pseudoknot formation. Using NMR, they confirmed that the pseudoknot and most of the native tertiary interactions outside the ligand-binding pocket were intact. Mg^{2+} stabilizes this pre-folded state by a factor of 10, and SAM stabilizes it further. Ultimately, the lifetime increases from 30 ms to 1 s.[75] According to both definitions, then, SAM-II binds its ligand using a CS-type mechanism.

MetX highlights the need to clearly define which structural transitions CS and IF are meant to describe. Although the high-FRET states with and without SAM are identical to within the resolution of FRET, Mg^{2+} does not pre-organize the binding pocket or all of the tertiary contacts between the P1 stem-loop and the P2 and P3 stems. Small changes in structure within the binding pocket are necessary and ubiquitous on ligand binding. This has led to the claim that all ligand-binding events are IF, but in most authors' usage, this is considered too general to be helpful, and the terms IF and CS refer to global tertiary structure transitions.

Single-molecule experiments point to an apparently-simple choice: characterize the kinetics of the transitions accessible to force or fluorescence vectors. These conformational changes are usually global, but can be as small as the twisting[76] or zipping[63] of a single helical element. They do, however, usually involve some change in tertiary structure, not only the rearrangement of chemical bonds in an already-compact core region to accommodate the ligand. This approach has been taken in the case of *metX*, and we will continue throughout this paper to focus on the analysis of key tertiary structure folding steps. However, *metX* undergoes additional rearrangements upon ligand binding that are invisible to long-range distance changes by FRET and to specifically positioned 2-aminopurine AP analogues.[75] That SAM captures and stabilizes the pseudoknot is without doubt, but it is worth remembering that the additional changes it induces in the process are not trivial, nor necessarily local. Which transition to emphasize may be a matter of judgment. In force experiments especially—where the shape of the FEL itself is being probed—it is worth heeding the warning that the most accessible 'pulling vector' may not correspond to a helpful reaction coordinate, and that the rupture force distributions must be interpreted accordingly.[77]

There are also examples of pure IF mechanisms among riboswitches. The *lysC* lysine aptamer is a five-helix junction that binds lysine in its central core and organizes into two clusters of stems (Fig. 3b). The three stem-loops above the central region interact via two sets of interhelical bonds, and the two below the junction stack with the upper stems, their core nucleotides caught up in ligand interactions.[66, 78] Fiegland and colleagues showed that ligand binding promotes a population shift into a previously unoccupied high-FRET state that corresponds to the close association of the bottom stem-loops, consistent with the 'ordering' definition of IF.[79] In addition, as the concentration of lysine increases, the folding rate also increases,[79] while the unfolding rate is insensitive to the ligand concentration. This is the definition of IF according to the kinetic/energetic picture.

These clearly-defined mechanisms may not be the norm. Holmstrom and colleagues' FRET analysis of the *env8* hydroxocobalamin (HyCbl) riboswitch (Fig. 3c and 4a) clearly showed that the ligand associates with the unfolded aptamer. Ligand binding before folding implies an IF mechanism.[80] However, Holmstrom et al. also found that while the addition of HyCbl increases the folding rate modestly, its most notable impact on the dynamics of folding is to decrease the *unfolding* rate by a factor of 10. The kinetics of ligand binding, then, could be interpreted as predominantly characterized by CS and the subsequent stabilization of the native state by the ligand.

In this riboswitch, it appears that the mechanisms of CS and IF have been thoroughly blended. Ligand binding to the unfolded structure plays only a modest role in determining the folding rate, which is still dominated by the aptamer's intrinsic search dynamics. In other words, early binding is *not* accompanied by strong induction of folding, but a continued conformational search and the kinetics associated with it. Is the lesson to take away that 'binding before folding' is not synonymous with IF after all? We argue that the lesson goes beyond that.

As Savinov and coworkers pointed out, the energetic definitions of CS and IF, which are experimentally demonstrated by the ligand dependence of the microscopic folding and unfolding rates, make hybrid mechanisms easy to conceptualize.[28] A ligand-binding process in which the ligand both stabilizes (selects) a configuration and promotes (induces) transitions into it through favorable bond formation seems both intuitive and general; in fact, such mechanisms have been identified in IDPs and referred to as 'extended CS'.[21] 'Selective induction' may be equally appropriate. Such general mechanisms, characterized by ligand binding to multiple states, including a pre-folded native state, and changes in the kinetics of both the on and off rates, are present in many

View Article Online
DOI: 10.1039/C6CP08798A

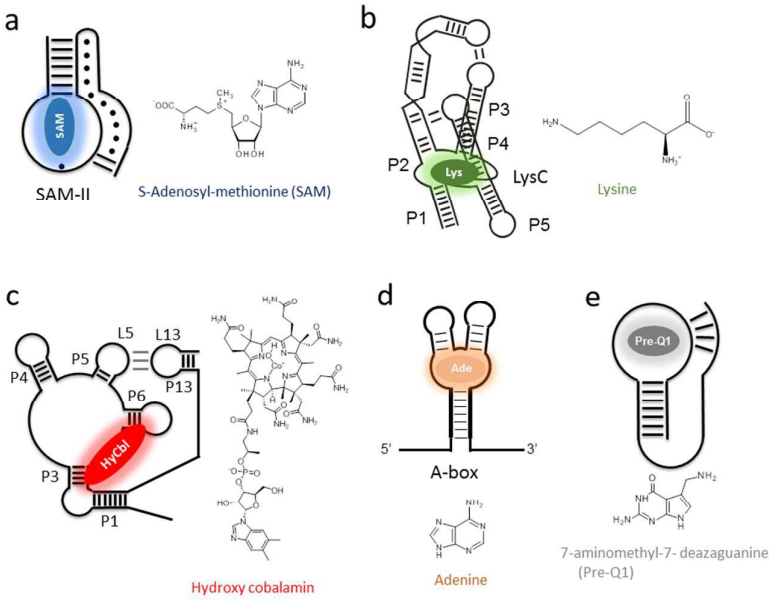

Fig. 3 The aptamer domains of the riboswitches under discussion. The chemical structure of the corresponding ligand is also shown. (a) The *metX* SAM-II aptamer. The switching stem is shown at top; SAM (S-adenosylmethionine) binding induces the formation of a pseudoknot between the loop and the single-stranded tail. (b) The *lysC* lysine aptamer is a five-helix junction. Its global structure is stabilized by two sets of tertiary contacts, P2/P3 and P2/P4. (c) The *env8* hydroxocobalamin (HyCbl) aptamer. (d) The consensus A-box aptamer. The *pbuE* and *add* aptamers are very similar in structure. (e) The *bsu* Pre-Q_1 aptamer, similar in size, structure, and switching mechanism to the SAM-II aptamer, despite the fact that its ligand is a purine closely related to adenine.

riboswitches. For instance, force extension and refolding experiments have shown that ligand binding drastically stabilizes the native-state free energy wells of both the *pbuE*[63] and *add*[81] adenine riboswitches (Fig. 3d) without a comparable effect on the folding barrier, the energetic signature of CS. The final step into the native state is the zipping of the switching stem, which occurs after the formation of this riboswitch class's characteristic loop-loop tertiary interaction.[63]

However, when the kinetics of *pbuE*'s loop-loop interaction itself are monitored via single-molecule FRET, both the on and off rates depend on ligand concentration. Because of the ligand's influence on the folding rate, the adenine aptamer's dynamics have been classified as IF, despite the fact that the native state is accessible in the absence of the ligand.[67] Furthermore, the rates are highly sensitive to the environment. Depending on the concentrations of Mg^{2+} and denaturants,[82] the overall apparent mechanism can be IF, CS, or a hybrid in the kinetic sense.[28]

The adenine riboswitch sheds further light on the potential disjoint between the timing of ligand binding and its role in the folding mechanism. NMR studies have shown that adenine can bind multiple partially-folded configurations of the *add* aptamer,[83] but that the formation of the native state still depends primarily on the concentration of Mg^{2+}, as for *pbuE*.[12] Conversely, binding of guanine to the very similar guanine riboswitch induces folding even in the absence of Mg^{2+}.[64] This indicates that the *add* aptamer has several paths through the energy landscape to its native state. The need to present a unique binding site for the ligand is not a bottleneck in the process. However, ligand binding does not significantly enhance the folding rate. This conflicts with the kinetic notion of IF, but resembles the description of 'fly-fishing' by IDPs, which presents an extended structure to maximize the likelihood of ligand capture, then fold around it.

The *bsu* PreQ_1-I riboswitch takes hybrid folding mechanisms one step further (Fig. 3e and 4b). In low Mg^{2+} conditions, this small pseudoknot-based aptamer exists primarily in its unfolded state, and Suddala et al. showed that it binds the ligand entirely through an IF mechanism.[84] The folding rate increased with PreQ_1 concentration. As the background Mg^{2+} concentration was increased, however, the pseudoknot formed, much like the case of *metX*, and ligand binding decreased the unfolding rate of this structure, the kinetic hallmark of CS. Ligand-detected dispersion relaxation measurements have also confirmed that the folded aptamer is present in the ligand-free ensemble.[85] However, the presence of the ligand still enhanced the folding rate in these conditions.[84] The kinetics of PreQ_1-I are notable for two reasons: first, it exhibits a clear transition from distinctively IF kinetics in one environment to a hybrid mechanism in another. Second, the mechanism in the presence of Mg^{2+} cannot be simply categorized. The ligand interacts with a pre-organized aptamer, the structural definition of CS, but its kinetics are a true hybrid mechanism as described by Savinov and coworkers.[28]

Such hybrid mechanisms and classification paradoxes highlight the complexity, variety, and especially the adaptability of RNA ligand-binding mechanisms. It should not be surprising that multiple pathways into the ligand-bound state exist in general. The famous picture of the folding energy landscape is, after all, high-dimensional, and dozens of environmental factors, including Mg^{2+} and the ligand, influence it. While some riboswitches like *lysC* have a well-defined mechanism across a wide range of environments,[79] *bsu* can bind its ligand at different stages of folding,[84] changing its path to the native state dramatically, while *env8*[86] and the adenine

View Article Online
DOI: 10.1039/C6CP08798A
Physical Chemistry Chemical Physics PERSPECTIVE

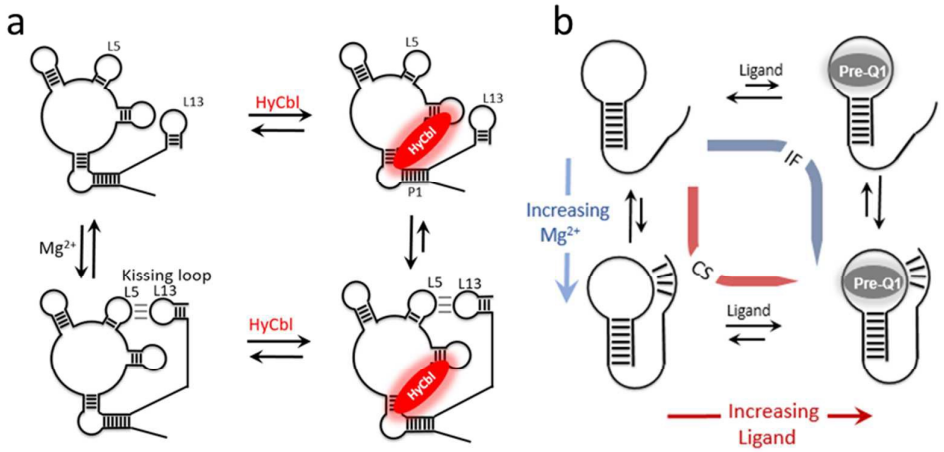

Fig. 4 Two riboswitches with hybrid ligand binding mechanisms. (a) Mg^{2+} can preorganize the *env8* hydroxocobalamin (HyCbl) aptamer, but surface-immobilized FRET showed that HyCbl interacts primarily with the unfolded state, making the *env8* aptamer an example of structural IF. However, HyCbl binding decreases the unfolding rate by a factor of ten, schematically represented by a shorter arrow, with a more modest effect on the folding rate, indicating that the aptamer behaves kinetically according to CS. (b) Screening interactions with Mg^{2+} can also preorganize the pseudoknot of the *bsu* aptamer even in the absence of ligand, shifting the binding mechanism from structural-definition IF in a Mg^{2+}-deprived environment to CS in the presence of cations. The kinetic behaviors along the two paths, represented by the relative sizes of the forward and back arrows, are consistent with the structural mechanisms. In the absence of Mg^{2+}, ligand binding enhances the folding rate (kinetic IF), while in the presence of Mg^{2+}, the preorganized state has a higher Pre-Q_1 affinity than the unfolded state (kinetic CS).

aptamers[63, 81] exhibit general mechanisms that allow both conformational search and ligand induction, which alternate importance in changing environments, recalling Hammes et al.'s notion of flux.[42]

A new conceptualization of aptamer-ligand interactions

Is it any use, then, to classify these mechanisms at all? If flux and generalized mechanisms are the norm, what role can a system of categories play? They have their uses: a clear set of categories acts as a common vocabulary for scientific discourse, but the categories must be apt. Below, we will suggest an extended classification system for RNA ligand-binding mechanisms that takes these generalities into account without losing its utility. But we hasten to emphasize that a change in the discussion is necessary: classifying the *mechanisms* may be helpful, but we should be careful of classifying *aptamers* as using or 'being' any one of them. Use of any ligand-binding mechanism is, generally speaking, contingent and environment-dependent.

The three requirements for a paradigm that is fit for purpose are (1) clarity, (2) practicality, and (3) comprehensiveness, or flexibility. We take each in turn.

(1) Clarity. What precisely are we describing by 'CS', 'IF', and other such terms? We argue that some of the assumptions carried over from the origins of these terms in early protein theory can be safely jettisoned, and that various molecular behaviors currently lumped together under the catch-all categories of CS and IF can be usefully separated and given independent and simultaneous descriptors.

(2) Practicality. If we are to use CS and IF as categories, or substitute something else in their place, we much resolve the many ambiguities in their usage. For example, on what scale are we speaking? It has been observed that ligand-binding always results, at minimum, in minute rearrangements of the nucleotides in the binding pocket that are visible to X-ray crystallography or NMR, and that accordingly, all ligand-binding events are 'IF'. However, this is plainly not what most researchers mean when we refer to an aptamer's binding mechanism. In general, global structural changes are under discussion. Perhaps the subset of transitions needs to be narrowed further: are ligand interactions that occur during the folding of secondary structure sufficiently different—or insufficiently understood—that 'CS' and 'IF' (or their replacements) should be restricted to discussions of tertiary transitions?

(3) Comprehensiveness, or flexibility. By this we mean that whatever categories we use must remain clear while avoiding overspecification. Every study that assigns an 'IF' or 'CS' mechanism to an RNA aptamer adds the caveat that these are pure, 'model' mechanisms that are never perfectly expressed in real systems. We suggest that a classification scheme that makes the continuum between poles more obvious would be beneficial, as it would allow us to describe a diverse repertoire of ligand-binding behaviors without simply assessing how much they correspond with one or the other of a pair of extremes.

With these requirements in mind, we present our proposed scheme for the classification of RNA-ligand binding in Fig. 5. Our basic premise is to separate the 'structural' and 'kinetic' definitions into two separate descriptors, which we call Ordering and Kinetics. We focus on describing a single tertiary structure transition, specific to the aptamer under

View Article Online
DOI: 10.1039/C6CP08798A

consideration, that occurs upon ligand binding, and require that it include elements outside the rearrangement of the ligand-binding pocket, but that it can occur after pre-organization of some elements of the native structure, e.g. by Mg^{2+}. The definition does not eliminate all ambiguities—for example, an experimenter could still choose to describe either the loop-loop docking transition or the zipping of P1 of the adenine aptamer—so the choice must be made clear.

As mentioned before, the choice of which folding event to monitor must be made explicit and perhaps justified. If the ligand can bind multiple 'unfolded' configurations, the researchers must carefully consider how to represent these paths on a single axis. Likewise, an aptamer that binds multiple ligands will have paths in which one ligand binds by each mechanism, and this may not translate to a single fraction.[43]

The extremes of the Ordering parameter, which reflects the sequence of binding/folding events, correspond to the former structural definitions of CS and IF: folding before binding and binding before folding, respectively. Likewise, the Kinetic parameter corresponds to the kinetic definitions, and we use the names IF and CS for the poles of ligand-enhanced docking and undocking transitions, in reference to the single-molecule kinetic/energetic definitions described by Savinov et al.[28]

The corners of this two-dimensional space therefore correspond to four 'pure' ligand-binding mechanisms. We refer to the intersections of the old structural and kinetic definitions as 'Pure IF' (PIF) and 'Pure CS' (PCS). In honor of the common search tactic of IDPs, we refer to the intersection of early binding and ligand-independent off rate, as in the HyCbl riboswitch, as Fly-Fishing (FF); finally, we call the combination of conformational search and ligand-enhanced folding kinetics, as in the *bsu* and adenine aptamers in some conditions, Selective Induction (SI).

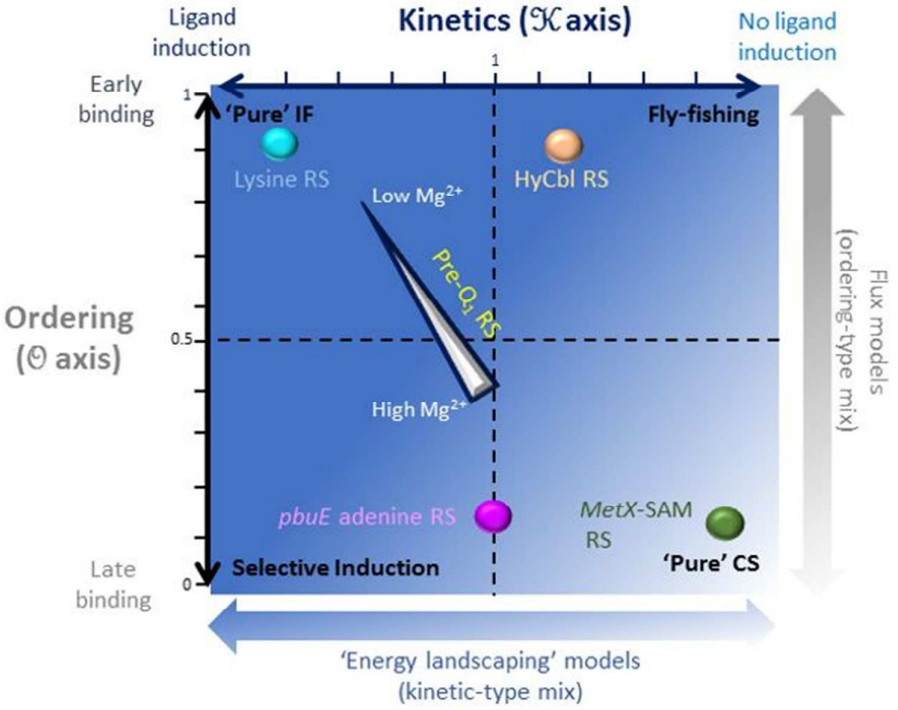

Fig. 5 Proposed scheme for classifying aptamer-ligand binding mechanisms. The two axes allow a ligand-binding transition to be classified according to both the Ordering of binding and folding (the structural definition of mechanism) and the Kinetic effect of ligand binding on the folding and unfolding rates (the kinetic definition). The space of possible mechanisms divides naturally into four quadrants. A Pure IF mechanism fits both the traditional definitions of IF: binding before folding and enhancement of the folding rate. Pure CS is analogous. Fly Fishing describes mechanisms in which the ligand binds an unfolded or semifolded state, but does not enhance the folding rate. Selective Induction describes mechanisms in which the ligand associates with a pre-folded structure and also enhances the rate of folding. Since the Ordering (O) and Kinetic (K) axes are continuous, hybrid mechanisms can be fully accounted for. The O axis can be quantified by calculating the flux through structural IF and CS pathways, while the K axis can be quantified by the ratio of changes in the folding and unfolding energy barrier heights. Five of the riboswitches discussed in this review are shown as qualitative examples. *LysC* and *metX* are located in the Pure IF and Pure CS quadrants, and the *pbuE* aptamer is shown near the border between Pure CS and Selective Induction. Since loop-loop docking and P1 zipping have the same mechanism (Table 1), they are represented by a single spot. The change in the *bsu* Pre-Q_1 aptamer's mechanism is shown by a wedge whose width indicates increasing Mg^{2+} concentration, while the *env8* HyCbl aptamer is shown near the border between Pure IF and Fly Fishing to indicate that while ligand binding has its most dramatic effect on the unfolding rate (kinetic CS), it also has a measurable effect on the folding rate, giving it a hybrid character, though 'more' on the side of kinetic CS.

View Article Online
DOI: 10.1039/C6CP08798A

Table 1 Proposed aptamer ligand-binding mechanisms

Aptamer	'Ordering' mechanism	'Kinetic' mechanism	Updated mechanism
add (P1)[81]	CS	CS	PCS
bsu (-Mg^{2+})[84]	IF	IF	PIF
bsu (+Mg^{2+})[84]	CS	hybrid	PCS/SI
env8[80]	IF	CS	FF
lysC[79]	IF	IF	PIF
metX[75]	CS	CS	PCS
pbuE (P1)[63]	CS	hybrid	PCS/SI
pbuE (loops)[63, 67]	CS	hybrid	PCS/SI

Acronyms: PIF, Pure Induced Fit; PCS, Pure Conformational Selection; FF, Fly-Fishing; and SI, Selective Induction.

We note that the 'Ordering' and 'Kinetic' axes have different meanings. An aptamer located somewhere near the middle of the Kinetic axis has a hybrid mechanism, and it is assumed that ligand binding affects both the folding and unfolding rates. However, it makes no sense for a ligand to bind 'both before and after' the folding of the native state.

Therefore, the center of the Ordering axis corresponds to an environment in which more than one folding pathway is operating, and the ligand is able to bind at multiple points in the folding process. Hammes et al.[42] have indicated that multi-pathway flux is the norm for protein-ligand interactions, and it is likely to be the case in RNA as well. A given aptamer may move vertically along the Ordering axis as the environment and the fluxes through the pathways change, and have an 'average' mechanism somewhere in the middle.

By presenting the range of ligand-binding mechanisms as a two-dimensional continuum rather than a pair of poles, we can easily situate the diverse examples discussed earlier in a coherent framework (also see Table 1). With the current experimental evidence, the lysine riboswitch could sit in the upper left corner of the Pure IF quadrant: ligand binding occurs distinctly and apparently exclusively to a partially-unfolded manifold of states, and it enhances the folding rate without affecting the stability of the native state. *MetX* occupies the corresponding position in the Pure CS quadrant.

Mechanisms that under the traditional definitions of CS and IF appeared to be paradoxes or hybrids can now be classified more transparently (summarized in Table 1). For instance, the HyCbl aptamer binds its ligand in a partially-folded state and ligand binding only modestly increases k_{fwd}, having a greater impact on k_{rev}. Accordingly, HyCbl is located in the Fly-Fishing quadrant of the diagram, but not quite at the far right, as it does not *only* act by stabilizing the folded state. It enhances the folding rate as well, if only slightly, making it most accurately a kinetic hybrid. This is represented by placing it right of center of the horizontal (Kinetic) axis, but not at the far end.

Changes in ligand-binding mechanism as a result of Mg^{2+} concentration or other factors can be easily depicted as trajectories in 2D space. The $PreQ_1$-I aptamer shifts from binding first + IF kinetics (the Pure IF quadrant) to folding first + hybrid kinetics (the border between the Selective Induction and Pure CS quadrants). We show this with a wedge whose larger end corresponds to the mechanism in higher Mg^{2+}.

In some cases, the binding mechanism may not be fully understood in both key dimensions. Recall the adenine aptamers, which are known to bind the ligand to multiple partially-folded structures, but which have spawned predictions of different kinetic behavior for different folding transitions and environments. A trajectory showing the mechanism with increasing Mg^{2+} may be appropriate, or two separate mechanisms, one each reflecting the dynamics of loop-loop docking and P1 zipping, may provide the most accurate description of the ligand-binding process. What is and is not known can still be compactly visualized. Here, we represent *pbuE* with a single point for simplicity, as both P1 and loop-loop dynamics can be interpreted as having a mechanism near the intersection of Pure CS and Selective Induction.

As useful as a qualitative schematic may be, both axes can be quantified to make direct and rigorous comparisons between aptamers. We will call the Ordering-axis variable O, and the Kinetic-axis variable K. By computing the flux through different folding paths, either as described by Hammes et al.[42] for simulations or by counting trajectories in single-molecule ensembles, the percent of trajectories in which binding occurs before folding fixes a value of O between 0 and 1 on the Ordering axis, with O = 0.5 indicating an environment in which the flux through both mechanisms is equal.

The Kinetic axis can be quantified in terms of the ligand-induced changes to the FEL. Since the kinetic definitions of classical CS and IF are the ligand dependence of the microscopic rates k_{rev} and k_{fwd}, respectively, we suggest quantifying the ratio of the changes in the associated free energies, $\Delta\Delta G^{\ddagger}_{rev}$ and $\Delta\Delta G^{\ddagger}_{fwd}$, as the ligand concentration changes from none to excess in a given environment. The Kinetic-axis variable K would then be

$$\Delta\Delta G^{\ddagger}_{fwd} / \Delta\Delta G^{\ddagger}_{rev} = \ln\left(k^{f}_{fwd}/k^{0}_{fwd}\right) / \ln\left(k^{f}_{rev}/k^{0}_{rev}\right).$$

In the pure IF limit, k_{rev} does not change with ligand concentration, but k_{fwd} increases, so $K \to \infty$. In the pure CS limit, $\Delta k_{fwd} = 0$ and k_{rev} increases, so $K \to 0$. $K = 1$, where the changes in $\Delta G^{\ddagger}_{fwd}$ and $\Delta G^{\ddagger}_{rev}$ are of equal magnitude, is a sensible point to divide the (logarithmically scaled) left and right halves of the diagram. In addition, we note that in single-molecule force experiments, equilibrium k_{rev} and (in the right limits) k_{fwd} can be extracted from rupture and refolding force distributions, but so can the barrier heights $\Delta G^{\ddagger}_{i}$ themselves, giving the experimenter flexibility in accessing the same data.[87, 88]

It is obvious that K will be profoundly environment-dependent. For most riboswitches, the Mg^{2+} concentration has a significant effect on the switching behavior, often stabilizing either the native state or a ligand-competent intermediate. As

View Article Online
DOI: 10.1039/C6CP08798A

such, a given aptamer may have several K values, best represented by a 'wedge' or multiple discrete points, as appropriate to the experiments conducted.

Finally, the K value cannot be defined unless a certain subset of kinetic data is available. In general, the kinetic picture of riboswitch folding is incomplete. Single-molecule reports often focus on changes in the ligand binding rate (k_{fwd}) as a function of ligand or Mg^{2+}, but do not consider k_{rev}, or emphasize population shifts rather than dynamics. Sometimes, a report of the overall stabilization of the native state can still provide enough information to qualitatively infer the mechanism, as in the case of the *metX* SAM-II aptamer. At the time of writing, though, some aptamers with well-known binding mechanisms on the ordering definition, for example, SAM-I,[89] could not be used as examples because their full kinetic profile has not yet been measured. When all the relevant data is available, however, the ordered pair (O,K), or a set of them for several structural transitions or environments, can succinctly summarize our understanding of that aptamer's ligand-binding process.

Conclusion

This unambiguous representation of what is known about the ligand-binding mechanism is all our categories are meant to do. We feel that the notions of CS and IF are not nuanced enough to describe RNA-ligand dynamics as we currently understand them. This is partially due to the potential contradictions between definitions that focus either on the order of folding and binding or on the kinetic effects of binding on folding, and partially due to the ubiquity of hybrid mechanisms. In this review we have suggested an alternative way of classifying binding mechanisms that recognizes the variety of ways these kinetic and structural behaviors can pair. This continuum can, if necessary, be divided into a core set of four binding behaviors instead of two—Pure IF, Fly-Fishing, Selective Induction, and Pure CS—which already resolves several of the paradoxes present in the literature. But the real value of these four categories lies in their representation as a quantifiable spectrum that allows general and environment-dependent mechanisms to be tracked across multiple folding pathways and kinetic profiles. The picture of riboswitch-ligand binding is rich and complex, and requires categories that have been tailored to suit the behaviors they describe, rather than the reverse.

References

1 K. V. Morris and J. S. Mattick, *Nature Reviews Genetics*, 2014, **15**, 423-437.
2 R. W. Carthew and E. J. Sontheimer, *Cell*, 2009, **136**.
3 D. Holoch and D. Moazed, *Nature Reviews Genetics*, 2015, **16**, 71-84.
4 R. A. Garret, G. Vestergaard and S. A. Shah, *Trends in Microbiology*, 2011, **19**, 549-556.
5 A. Haller, M. F. Soulière and R. Micura, *Accounts of Chemical Research*, 2011, **44**, 1339-1348.
6 S. Blouin, J. Mulhbacher, J. C. Penedo and D. A. Lafontaine, *ChemBioChem*, 2009, **10**, 400-416.
7 A. Serganov and D. J. Patel, *Biochim. Biophys. Acta, Gene Regul. Mech.*, 2009, **1789**, 592-611.
8 C. Perez-Gonzalez, D. A. Lafontaine and J. C. Penedo, *Frontiers in Chemistry*, 2016, **4**.
9 J.-F. Lemay, G. Desnoyers, S. Blouin, B. Heppell, L. Bastet, P. St-Pierre, E. Masse and D. A. Lafontaine, *PloS Genet.*, 2011, **7**.
10 J. L. Baker, N. Sudarsan, Z. Weinberg, A. Roth, R. B. Stockbridge and R. R. Breaker, *Science*, 2012, **335**, 233-235.
11 E. Westhof and D. J. Patel, *Curr. Opin. Struct. Biol.*, 1997, **7**, 305-309.
12 J. Noeske, H. Schwalbe and J. Wöhnert, *Nucleic Acids Research*, 2007, **35**, 5262-5273.
13 E. Fischer, *Berichte der deutschen chemischen Gesellschaft*, 1894, **27**, 2985-2993.
14 J. D.E. Koshland, G. Némethy and D. Filmer, *Biochemistry*, 1966, **5**, 365-385.
15 J. Monod, J. Wyman and J.-P. Changeux, *Journal of Molecular Biology*, 1965, **12**, 88-118.
16 K. A. Dill and H. S. Chan, *Nature Structural Biology*, 1997, **4**, 10-19.
17 C.-J. Tsai, S. Kumar, B. Ma and R. Nussinov, *Protein Science*, 1999, **8**, 1181-1190.
18 H. X. Kondo, N. Okimoto, G. Morimoto and M. Taiji, *Journal of Physical Chemistry B*, 2011, **115**, 7629-7636.
19 T. C. Messina and D. S. Talaga, *Biophysical Journal*, 2007, **93**, 579-585.
20 D. D. Boehr, R. Nussinov and P. E. Wright, *Nature Chemical Biology*, 2009, **5**, 789-796.
21 P. Csermely, R. Palotai and R. Nussinov, *Trends in Biochemical Sciences*, 2010, **35**, 539-546.
22 B. Ma and R. Nussinov, *Current Opinion in Chemical Biology*, 2010, **14**, 652-659.
23 D. W. Miller and K. A. Dill, *Protein Science*, 1997, **6**, 2166-2179.
24 H. R. Bosshard, *Physiology*, 2001, **16**, 171-173.
25 T. R. Weikl and C. v. Deuster, *Proteins*, 2008, **75**, 104-110.
26 S. Peuker, A. Cukkemane, M. Held, F. Noé, U. B. Kaupp and R. Seifert, *Biophysical Journal*, 2013, **104**, 63-74.
27 D. Bucher, B. J. Grant and J. A. McCammon, *Biochemistry*, 2011, **50**.
28 A. Savinov, C. F. Perez and S. M. Block, *Biochimica et Biophysica Acta*, 2014, **1839**, 1030-1045.
29 L. M. Espinoza-Fonseca, *Biochemical and Biophysical Research Communications*, 2009, **382**, 479-482.
30 F. Aboul-ela, W. Huang, M. A. Elrahman, V. Boyapati and P. Li, *WIREs RNA*, 2015, **6**, 631-650.
31 C. J. Jackson, J.-L. Foo, N. Tokuriki, L. Afriat, P. D. Carr, H.-K. Kim, G. Schenk, D. S. Tawfik and D. L. Ollis, *PNAS*, 2009, **106**, 21631-21636.
32 E. D. Cera, *Molecular Aspects of Medicine*, 2008, **29**, 203-254.
33 A. T. Frank, A. C. Stelzer, H. M. Al-Hashimi and I. Andricioaei, *Nucleic Acids Research*, 2009, **37**, 3670-3679.

View Article Online
DOI: 10.1039/C6CP08798A

34 M. H. Bailor, X. Sun and H. M. Al-Hashimi, *Science*, 2010, **327**, 202-206.
35 S. M. Sullivan and T. Holyoak, *PNAS*, 2008, **105**, 13829-13834.
36 D.-A. Silva, G. R. Bowman, A. Sosa-Peinado and X. Huang, *PLoS Computational Biology*, 2011, **7**, e1002054.
37 Y. Xin, G. Gadda and D. Hamelberg, *Biochemistry*, 2009, **48**, 9599-9605.
38 J. A. Sigrell, A. D. Cameron and S. L. Mowbray, *Journal of Molecular Biology*, 1999, **290**, 1009-1018.
39 E. Westhof and D. J. Patel, *Current Opinion in Structural Biology*, 1997, **7**, 305-309.
40 A. D. Vogt and E. D. Cera, *Biochemistry*, 2013, **52**, 5723-5729.
41 K.-i. Okazaki and S. Takada, *PNAS*, 2008, **105**, 11182-11187.
42 G. G. Hammes, Y.-C. Chang and T. G. Oas, *PNAS*, 2009, **106**, 13737-13741.
43 K. G. Daniels, N. K. Tonthat, D. R. McClure, Y.-C. Chang, X. Liu, M. A. Schumacher, C. A. Fierke, S. C. Schmidler and T. G. Oas, *Journal of the American Chemical Society*, 2014, **136**, 822-825.
44 T. Wlodarski and B. Zagrovic, *PNAS*, 2009, **106**, 19346-19351.
45 R. V. Swift and J. A. McCammon, *Journal of the American Chemical Society*, 2009, **131**, 5126-5133.
46 K. Sugase, H. J. Dyson and P. E. Wright, *Nature Letters*, 2007, **447**, 1021-1025.
47 S. E. Halford, *Biochem. J.*, 1971, **125**, 319-327.
48 K. Arora and C. L. B. III, *PNAS*, 2007, **104**, 18496-18501.
49 Y. B. Kim, S. S. Kalinowski and J. Marcinkeviciene, *Biochemistry*, 2007, **46**, 1423-1431.
50 M. Antoine, J. A. Boutin and G. Ferry, *Biochemistry*, 2009, **48**, 5466-5482.
51 V. V. Heredia, J. Thomson, D. Nettleton and S. Sun, *Biochemistry*, 2006, **45**, 7553-7562.
52 A. D. Vogt and E. D. Cera, *Biochemistry*, 2012, **51**, 5894-5902.
53 S. Gianni, J. Dogan and P. Jemth, *Biophysical Chemistry*, 2014, **139**, 33-39.
54 B. A. Shoemaker, J. J. Portman and P. G. Wolynes, *PNAS*, 2000, **97**, 8868-8873.
55 A. K. Dunker, C. J. Brown, J. D. Lawson, L. M. Iakoucheva and Z. Obradović, *Biochemistry*, 2002, **41**, 6573-6582.
56 C.-J. Tsai, B. Ma, Y. Y. Sham, S. Kumar and R. Nussinov, *Proteins: Structure, Function, and Genetics*, 2001, **44**, 418-427.
57 C. J. Oldfield, Y. Cheng, M. S. Cortese, P. Romero, V. N. Uversky and A. K. Dunker, *Biochemistry*, 2005, **44**, 12454-12470.
58 V. Csizmók, M. Bokor, P. Bánki, É. Klement, K. F. Medzihradszky, P. Friedrich, K. Tompa and P. Tompa, *Biochemistry*, 2005, **44**, 3955-3964.
59 M. Fuxreiter, I. Simon, P. Friedrich and P. Tompa, *Journal of Molecular Biology*, 2004, **338**, 1015-1026.
60 A. Mohan, C. J. Oldfield, P. Radivojac, V. Vacic, M. S. Cortese, A. K. Dunker and V. N. Uversky, *Journal of Molecular Biology*, 2006, **362**, 1043-1059.
61 M. Onitsuka, H. Kamikubo, Y. Yamazaki and M. Kataoka, *Proteins*, 2008, **72**, 837-847.
62 H.-X. Zhou, *Biophysical Journal*, 2010, **98**, L15-L17.
63 W. J. Greenleaf, K. L. Frieda, D. A. N. Foster, M. T. Woodside and S. M. Block, *Science*, 2008, **319**, 630-633.
64 J. Buck, A. Wacker, E. Warkentin, J. Wöhnert, J. Wirmer-Bartoschek and H. Schwalbe, *Nucleic Acids Research*, 2011, **39**, 9768-9778.
65 A. Haller, R. B. Altman, M. F. Soulière, S. C. Blanchard and R. Micura, *PNAS*, 2013, **110**, 4188-4193.
66 A. D. Garst, A. Heroux, R. P. Rambo and R. T. Batey, *J. Biol. Chem.*, 2008, **283**, 22347-22351.
67 J.-F. Lemay, J. C. Penedo, R. Tremblay, D. M. J. Lilley and D. A. Lafontaine, *Chemistry and Biology*, 2006, **13**, 857-868.
68 J. Noeske, J. Buck, B. Fürtig, H. R. Nasiri, H. Schwalbe and J. Wöhnert, *Nucleic Acids Research*, 2007, **35**, 572-583.
69 O. M. Ottink, S. M. Rampersad, M. Tessari, G. J. R. Zaman, H. A. Heus and S. S. Wijmenga, *RNA*, 2007, **13**, 2202-2212.
70 K. P. Ravindranathan, E. Gallicchio and R. M. Levy, *Journal of Molecular Biology*, 2005, **353**, 196-210.
71 P. T. X. Li, C. Bustamante and I. Tinoco Jr., *PNAS*, 2007, **104**, 7039-7044.
72 J. A. Cruz and E. Westhof, *Cell*, 2009, **136**, 604-609.
73 S. Blouin, T. D. Craggs, D. A. Lafontaine and J. C. Penedo, in *Methods in Molecular Biology*, eds. T. Moss and B. Leblanc, Humana Press, 2009, vol. 543, ch. 28, pp. 475-502.
74 P. St.-Pierre, K. McCluskey, E. Shaw, J. C. Penedo and D. Lafontaine, *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 2014, **1839**, 1005-1019.
75 A. Haller, U. Rieder, M. Aigner, S. C. Blanchard and R. Micura, *Nature Chemical Biology*, 2011, **7**, 393-400.
76 S. H. Eschbach, P. St-Pierre, J. C. Penedo and D. A. Lafontaine, *RNA Biology*, 2012, **9**, 535-541.
77 O. Dudko, G. Hummer and A. Szabo, *PNAS*, 2008, **105**, 15755-15760.
78 A. Serganov, L. Huang and D. J. Patel, *Nature*, 2008, **455**, 1263-1268.
79 L. R. Fiegland, A. D. Garst, R. T. Batey and D. J. Nesbitt, *Biochemistry*, 2012, **51**, 9223-9233.
80 E. D. Holmstrom, J. T. Polaski, R. T. Batey and D. J. Nesbitt, *Journal of the American Chemical Society*, 2014, **136**, 16832-16843.
81 K. Neupane, H. Yu, D. A. Foster, F. Wang and M. T. Woodside, *Nucleic Acids Research*, 2011, **39**, 7677-7687.
82 P. A. Dalgarno, J. Bordello, R. Morris, P. St-Pierre, A. Dube, I. D. W. Samuel, D. A. Lafontaine and J. C. Penedo, *Nucleic Acids Research*, 2013, **41**, 4253-4265.
83 M.-K. Lee, M. Gal, L. Frydman and G. Varani, *PNAS*, 2010, **107**, 9192-9197.

View Article Online
DOI: 10.1039/C6CP08798A

84 K. C. Suddala, J. Wang, Q. Hou and N. G. Walter, *Journal of the American Chemical Society*, 2015, **137**, 14075-14083.

85 T. Moschen, C. H. Wunderlich, R. Spitzer, J. Levic, R. Micura, M. Tollinger and C. Kreutz, *Angew. Chem. Int. Ed.*, 2015, **54**, 560-563.

86 E. D. Holmstrom, J. T. Polaski, R. T. Batey and D. J. Nesbitt, *JACS*, 2014, **136**, 16832-16843.

87 C. A. Pierse and O. K. Dudko, *Biophysical Journal*, 2013, **105**.

88 O. K. Dudko, G. Hummer and A. Szabo, *Physical Review Letters*, 2006, **96**, 4.

89 B. Heppell, S. Blouin, A.-M. Dussault, J. Muhlbacher, E. Ennifar, J. C. Penedo and D. A. Lafontaine, *Nature Chemical Biology*, 2011, **7**, 384-392.