Supplementary material: CoBold, a method for identifying different functional classes of transient RNA structure features that can impact RNA structure formation *in vivo*

Adrián López Martín[1], Mohamed Mounir[1], and Irmtraud
M. Meyer [*1,2]

[1]Berlin Institute for Medical Systems Biology, Max Delbrück
Center for Molecular Medicine in the Helmholtz Association,
Hannoversche Str. 28, 10115 Berlin, Germany,
[2]Freie Universität Berlin, Department of Biology, Chemistry and
Pharmacy, Institute of Chemistry and Biochemistry, Thielallee 63,
14195 Berlin, Germany

August 11, 2020

---

[*]To whom correspondence should be addressed. Tel: +49 30 9406 3292; Fax: +49 30 9406 3291; Email: irmtraud.meyer@cantab.net

1 Additional figures

The main article, "CoBold, a method for identifying different functional classes of transient RNA structure features that can impact RNA structure formation *in vivo*", discusses the proposition of a new computational tool. Through the entire discussion of the method, a major part of the study consists of the analysis of the information that the method allows us to extract from different RNA systems. Those systems have the particular characteristic that they can adopt different conformations depending on the environmental properties. The systems discussed are the following ones:

- HDV ribozyme [Figure S1]:

  The Hepatitis Delta Virus (HDV) is a single-stranded (ss) RNA virus that aggravates symptoms of human hepatitis. This particular ribozyme regulates the self-cleaving process required for viral replication. This is achieved via two different structural configurations that either allow or prevent replication depending on the environmental conditions. In the active state, the region on the sequence where the self-cleaving happens is exposed, so HDV can undergo the process and replicate. During poor conditions for replication, however, the self-cleaving region is sequestered by a hairpin, thereby preventing replication. Notwithstanding, if optimal replication conditions are achieved once more, the 5' region of the sequestering hairpin is itself sequestered, so the self-cleaving region is exposed again and the replication can occur. The HDV ribozyme sequence has a length of 152 nt. In this particular case, we employed the alignment RF00094 from Rfam [1], with M28267.1 635-775 as reference sequence.

- 5'UTR of Leviviridae Levivirus [Figure S2]:

  The Levivirus is a ssRNA bacteriophage, *i.e.* a virus that kills certain bacteria. Its 5' UTR controls the translation of its proteins by regulating the binding of the ribosome and the replicase, which compete for the same region. Without the need for more proteins, the ribosomal binding site is sequestered by an upstream complementary sequence and trapped in a helix. If protein translation is needed, however, this complementary sequence is paired a sequence further, thereby releasing the sequence and enabling ribosome binding. This particular sequence has a length of 158 nt. For this analysis we used a custom alignment, chosing GQ153927.1 1-132 as reference sequence.

- SAM riboswitch [Figure S3]:

  This particular riboswitch is a structure widely shared among bacteria which allows for a switch in gene expression depending on the presence of a certain coenzyme (in this case, S-adenosylmethionine (SAM)). With SAM binding, the riboswitch adopts an RNA structure that stops transcription. This structure comprises two helices, one called anti-anti-terminator and one called terminator (in 5' to 3' order). Without SAM binding, this

RNA structure is partly re-arranged. In this rearrangement, the two anti-anti-terminator and terminator helices are destroyed and converted into a single new helix that pairs the 3' arm of the former anti-anti-terminator helix and the 5' arm of the former terminator helix into a new anti-terminator helix. This anti-terminator helix allows for transcription to proceed. The SAM riboswitch has a length of 215 nt. In this particular study we employed a custom alignment, with AL009126.3 1258276-1258464 as reference sequence.

- Tryphtophan (Trp) operon leader [Figure S4]:

  This structure regulates the transcription of the trp operon in *E. coli*. Similarly to the above examples, it senses if the optimal conditions for transcription are given or not. In this case, however, the region is responsive to the transcription speed itself. While transcription is being conducted by the ribosome as its nominal speed, a terminator hairpin forms which stops transcription. If conditions are not optimal for transcription and the ribosome stalls, an anti-terminator structure emerges that sequesters the terminator which in turns allows the ribosome to transcribe the whole sequence, instead of just a partial one. The sequence of the trp operon leader has a length of 152 nt. For the purpose of this analysis we used Rfam alignment RF00513 with AE005174.2 2263095-2263188 as reference.

- ZTP riboswitch [Figure S5]:

  This structure switch regulates transcription in bacteria via the presence or absence of the purine biosynthetic intermediate 5-aminoimidazole-4-carbozamide riboside 5'-triphosphate (ZTP) molecule. Here, we focus on the structure in *Clostridium beijerinckii pfl*. This riboswitch's structure was recently studied [2] and its cofolding path was extracted by analysing the SHAPE values of its different substranscripts. We wanted to check if our proposed method was able to recover that particular structure that was extracted experimentally with a to of time and effort. The ZTP riboswitch we use has a length of 172 nt. In this study we employed an improved version of the RF01750 alignment with CP000721.1 1211931-1212029 as reference sequence.

Due to the different conformations and elements to discuss, all the results could not be presented in the main article due to length restrictions, as the inclusion of all those figure and additional discussion would have extended the length of the main article by a considerable amount. Thus, those additional elements regarding the HDV ribozyme, the SAM riboswitch and the tryphtophan operon leader are being included in the following supplementary material. All the following figures were plot using Rchie/R4RNA [3].

2 SAM riboswitch

When analysing the results given by the comparative part of COBOLD for the SAM riboswitch for its two, bound and unbound, conformations [Figure S7, Figure S8] we can observe the evolutionary conserved clashing helices that can potentially appear during the transcription process. In this particular case we see how, for the unbound conformation, the clashing helices that are more prevalent are beneficial (3'-TRANS, 5'-CIS) or neutral (3'-MID, 5'-MID); some of those beneficial helices disappear when considering only the results with the lowest p-value [Figure S9, Figure S10]. However, for the bound conformation the observed predictions are more difficult to explain, as we cans see how many different helix types appear, some negatives ones. Even though the negatives cases disappear when considering the lowest p-value plots, where mostly neutral types (5'-MID, 3'-MID) appear; the appearance of those negative helices could seem puzzling. However, if we consider that those negative clashing helices usually spawn regions that are mutually exclusive for both conformation, they can be seen as turning points for the structure, that could appear under the right circumstances to impede the formation of the bound structure, allowing only the unbound structure to emerge.

When analysing the results given by the non-comparative part of COBOLD for the SAM riboswitch for its two, bound and unbound, conformations [Figure S11, Figure S12], we can observe the different clashing helices that can appear during transcription and interact with any of the reference helices, while also looking only the ones that are supported by evolutionary conservation. In this particular case, for the bound conformation we observe some beneficial clashing helices appearing (5'-CIS) that are mostly sequestering structures that only appear in that particular conformation. These helices are actually the most prevalent when considering only the ones that are also well evolutionary conserved. Those particular clashing helices can be seen as elements that can help the riboswitch to transition between conformation given the right environmental conditions. For the unbound conformation only the complete set of predictions of the non-comparative part of COBOLD are shown, as there was no intersection with the comparative part. The ones that can be observed are mostly of the neutral (5'-MID) of beneficial (3'-TRANS) type. Some of the disruptive (5'-TRANS, 3'-CIS) type can also be observed, they are once more affecting mutually exclusive structures, however, again can be interpreted as control helices that allow the formation of one of another structure in terms of the conditions promoting or not the appearance of that particular clashing helix.

3 Tryphtophan operon leader

In the case of the terminator and antiterminator conformations of the Tryphtophan operon leader, when focusing on the results given by the comparative part of COBOLD [Figure S13, Figure S14], we can observe a similar picture of the one appearing for the SAM riboswitch. Most of the clashing helices appearing

are of the beneficial or neutral type. There are some of the disruptive type, but those tend to disappear when considering the best conserved ones [Figure S15, Figure S16]. The only disruptive clashing helices that are well evolutionary conserved are the ones that sequester some of the helices that belong to structures that are mutually exclusive between the terminator and antiterminator conformation, as the can act as control elements that allow the formation of the structure according to the particular environmental conditions during the transcription process

The results of the non-comparative part of CoBold for the two conformation of the Tryphtophan operon leader [Figure S17, Figure S18] once more let us know the co-transcriptionallyemerging clashing helices, and in this case we see how there are an increased number of disruptive (3'-Cis, 5'-Trans) clashes. However, we also can appreciate how those clashing helices in this particular case are actually the same for both conformations (being 3'-Transfor the terminator conformation and 5'-Transfor the antiterminator one), which shows in an even clearer way how these specific helices acts are control elements that allow certain exclusive structures for different conformations. These control helices will capture some of the basepairs during transcription and thus, only the best suited conformation acording to the current environmental characteristics will arise at the end.

4 HDV ribozyme

If we study the results of the comparative part of CoBold for the HDV ribozyme conformations (active, alternative 1 and alternative 2) [Figure S19, Figure S20, Figure S21], we can observe how the clashing helices appearing are mostly neutral (5'-Mid, 3'-Mid) . However, one particular clashing helix peaks our attention, as it the only negative one while also being shared between the three different conformations and potentially clashing with structural elements that are specific to each conformation. This helix, that also appears when considering only the lowest p-value clashes [Figure S22, Figure S23 is one of the already mentioned control clashing helices that might serve as a selection point during transcription, allowing one of other conformation to arise in terms of the cell conditions.

This particular helix, however, is not observed when considering the predictions of the non-comparative part of CoBold [Figure S24, Figure S25], where most of the helices that can be observed are only of the neutral (5'-Mid, 3'-Mid) type, and only a neutral (5'-Mid) helix is present in the intersection between the non-comparative and comparative part. This helix is however present in both conformations and also clashes with an structure that, in this case, is present in both of them. That element could act as safety device, (similar to a 5'-Cishelix), sequestering some of the nucleotides of the 5' region of a helix to ensure the formation of that particular element by forbidding any other possible pairing and let the reference helix emerge by breaking when the 3' region of the reference helix is being transcribed.

References

[1] Kalvari,I., Argasinska,J., Quinones-Olvera,N., Nawrocki,E.P., Rivas,E., Eddy,S.R., Bateman,A., Finn,R.D., and Petrov,A.I. (2017) Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Research,* **46**(D1), D335D342.

[2] Strobel,E.J., Cheng,L., Berman,K.E., Carlson,P.D., and Lucks,J.B. (2019) A ligand-gated strand displacement mechanism for ZTP riboswitch transcription control. *Nature Chemical Biology,* **15**(11), 1067–1076.

[3] Lai,D., Proctor,J.R., Zhu,J.Y.A., and Meyer,I.M. (2012) R- chie : a web server and R package for visualizing RNA secondary structures. *Nucleic Acids Research,* **40**(12), e95–e95.

[4] Zhu,J.Y.A. and Meyer,I.M. (2015) Four RNA families with functional transient structures. *RNA Biology,* **12**(1), 5–20.

5   FIGURE LEGENDS
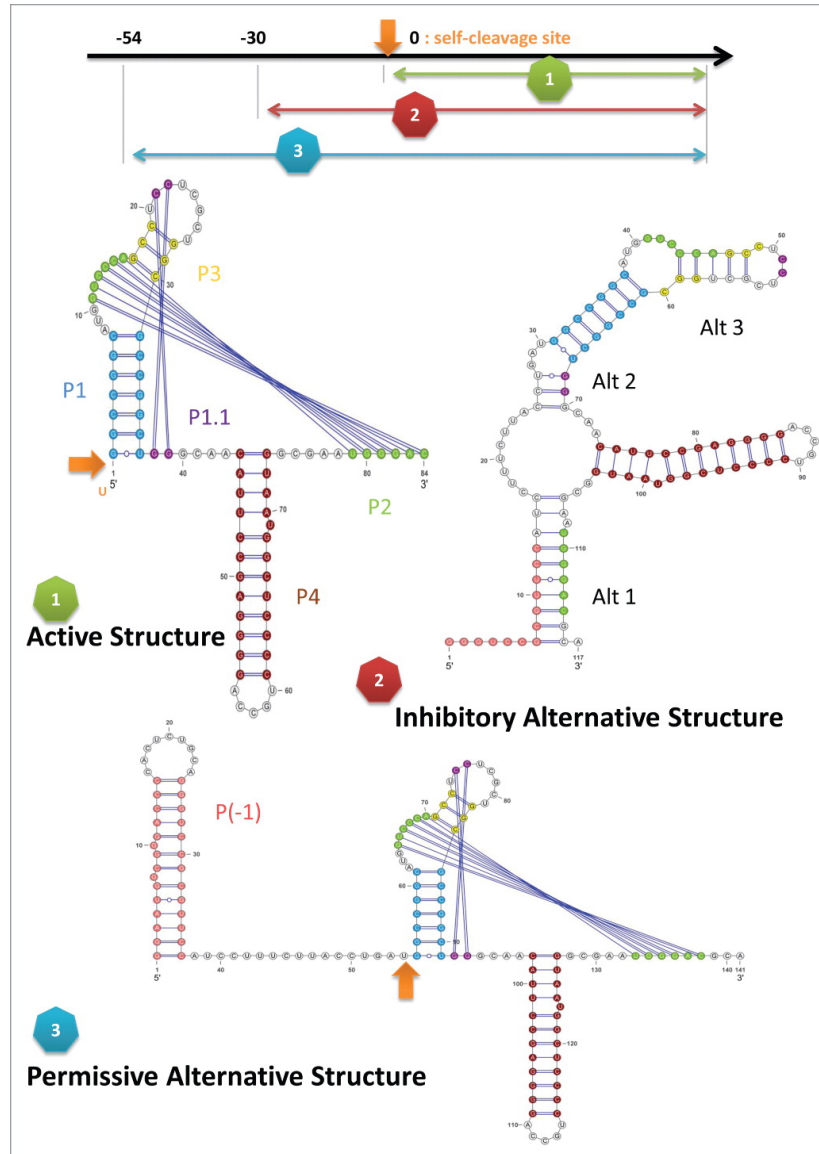
Figure S1: Picture showing the three different conformations the HDV rybozime can adopt: the active structure, the inhibitory structure and the permissive alternative structure. The self-cleaving site is indicated by the orange arrow, being exposed in the active and permissive structure and sequestered in the inhibitory one. Picture extracted from [4]
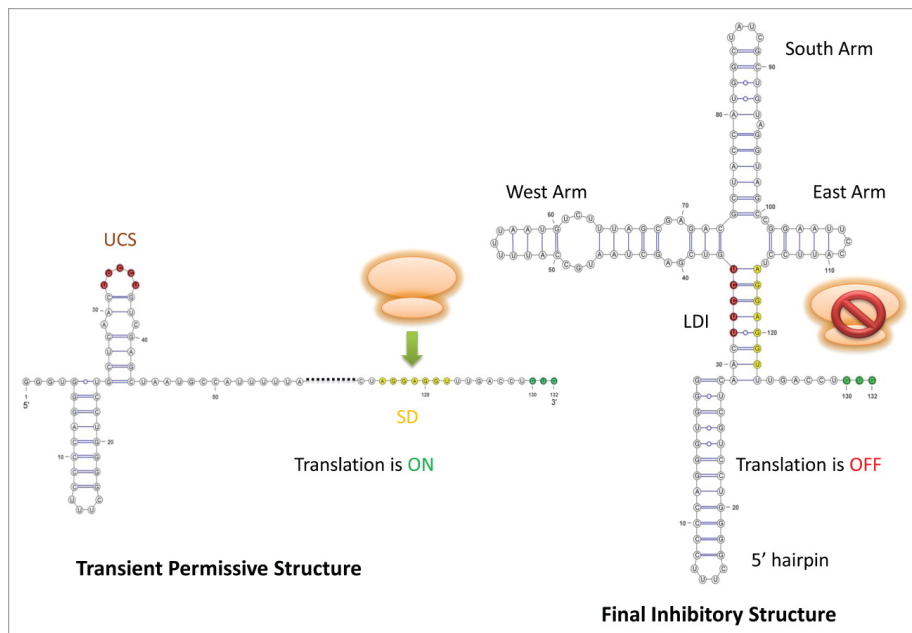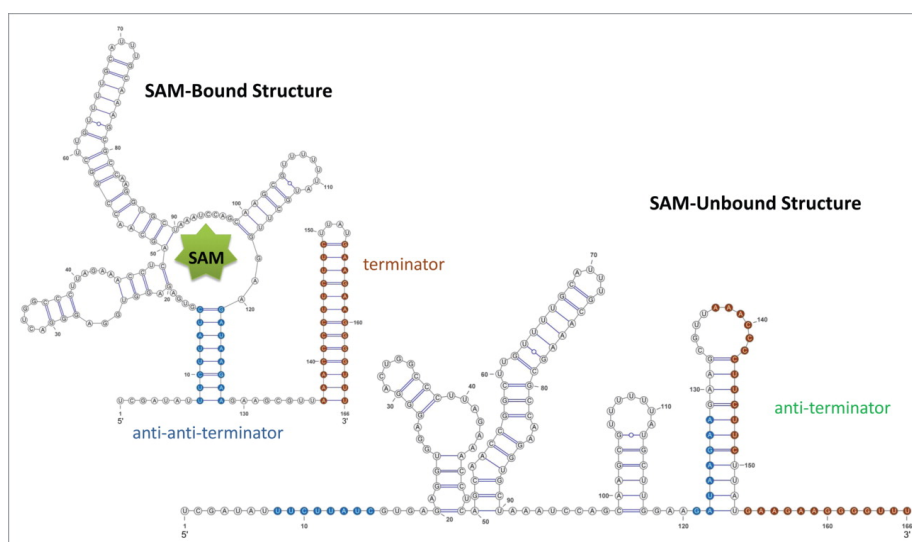
Figure S2: Picture showing the two different conformations the $5'UTR$ of the Leviviridae Levivirus can adopt: the inhibitory structure and the permissive alternative structure. The ribosome binding site (SD) site is indicated by the yellow nucleotides, being exposed in the permissive structure and sequestered in the inhibitory one. Picture extracted from [4]

Figure S3: Picture showing the two different conformations the SAM-responsive riboswitch can adopt: the bound structure and the unbound structure. The terminator, antiterminator and antiantiterminator hairpins are indicated in brown, green and blue respectively. It can be seen how the terminator is sequestered in the unbound state and free at the bound one. Picture extracted from [4]
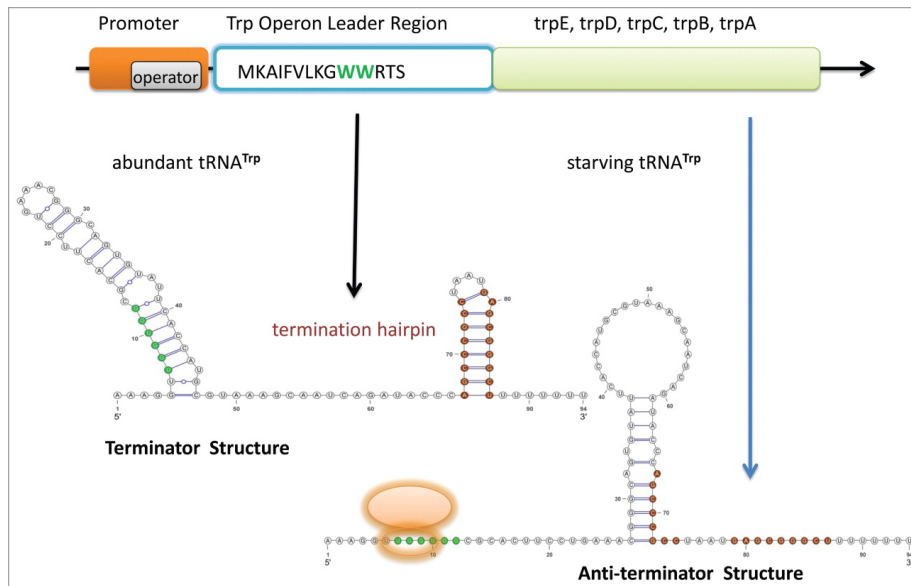
Figure S4: Picture showing the two different conformations the tryphtophan operon leader can adopt: the terminator structure and the antiterminator structure. The ribosome binding site and antiterminator hairpin are indicated in green and brown respectively. It can be seen how the binding site is sequestered in the terminator state and free and accessible at the antiterminator one, where the terminator is sequestered by the antiterminator. Picture extracted from [4]
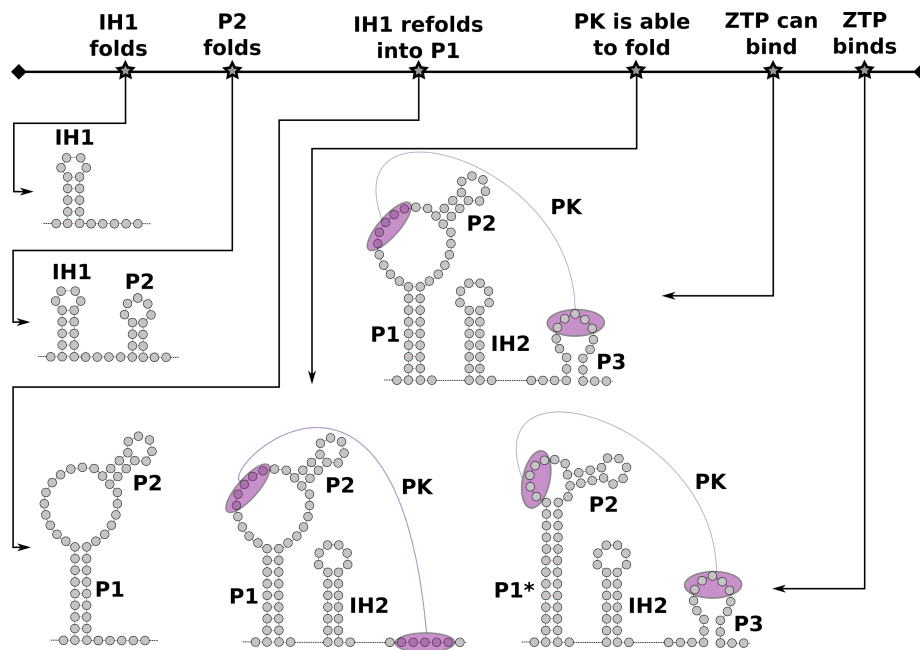
Figure S5: Picture showing the cofolding path as well as the different conformations for the ZTP riboswitch for the ZTP-bound and ZTP-unbound states. At the top a time-line can be observed with the relevant structural steps happening during transcription while, at the bottom we show the secondary structure of the different transient forms of the ZTP riboswitch, labelling the most relevant helices involved. Figure adapted from [2]
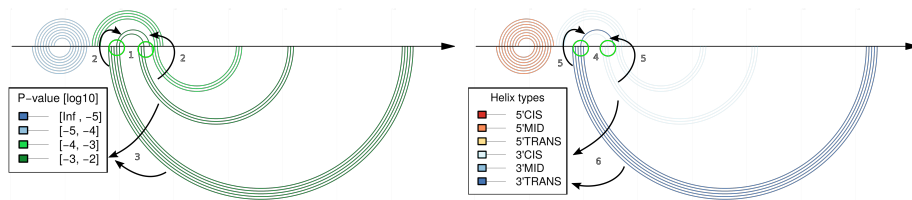
Figure S6: **Explanatory figure showing the clashing predictions with the minimum p-value of the permissive conformation of the Leviridae Levivirus by the comparative part of** CoBold**:** The horizontal line represents the input MSA. Each arc or semi-circle corresponds to a single base-pair connecting the corresponding two sequence positions. **Arcs on top** represent the reference RNA structure, in this case the permissive conformation of the Levivirus. **Arcs below** the horizontal line depict the transient helices predicted by the comparative mode of CoBold. For each shown transient helix, it is easy to identify the corresponding helix or helices it is clashing with (in this case indicated by the arrows) as these share base-paired positions along the horizontal line, see the feature highlighted by the circles and arrows (**1–2,4–5**). The **left part** of the figure shows the predicted helices and the reference helix/helices they are clashing with in the known structure coloured by statistical significance (*i.e.* p-value) (**5**) whereas the **right part** of the figure shows them coloured by functional class (**6**). In this case, the most prominent feature is thus the predicted transient helix comprising four base-pairs of high statistical significance (see dark green colour and legend, as indicated by the arrow on the bottom left) and of beneficial type (*i.e.* 3'-Trans) (see dark blue colour and legend on the bottom right) which may aid in the formation of the 3' helix of the reference RNA structure conformation.
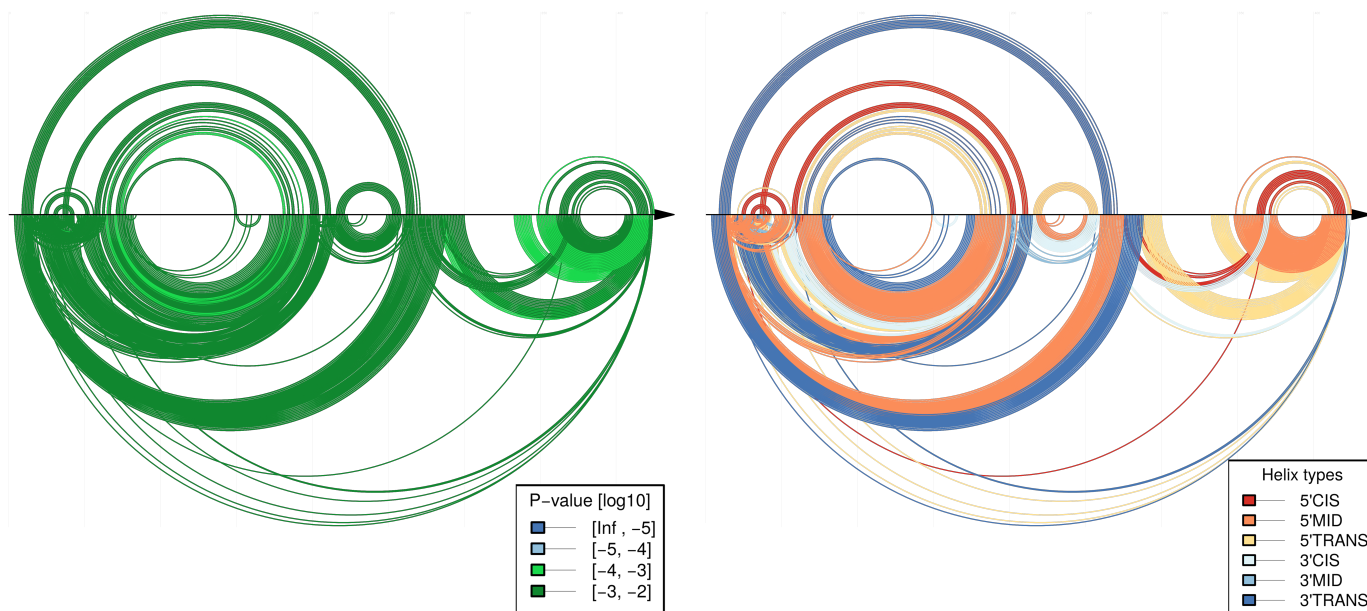
Figure S7: **Clashing predictions of the bound conformation of the SAM Riboswitch by the comparative part of** CoBold**:** In the top part of both figures the reference structure corresponding to the *bound conformation of the SAM Riboswitch* is shown. In the bottom part of both figures, only the clashing helices with the comparative part of CoBold and directly competing with one of the helices of the reference structure are shown. Each arc between two positions represent a basepair between those two nucleotides. Consecutive arcs represent helices spawning a consecutive number of basepairs. Each competing helix can be coloured in terms of two different schemes: p-values and helix type. When coloured in terms of helix type, each competing helices at the bottom and the corresponding reference helix being clashed are coloured in terms of their relative position. When coloured in terms of p-value, both the competing and reference competed helix are coloured in terms of the p-value of the clashing helix. Reference helices not being clashed with are coloured in black. The clashes appearing in the picture on the left are coloured in terms of their p-value. The clashes appearing in the picture on the right are coloured in terms of their helix types instead.
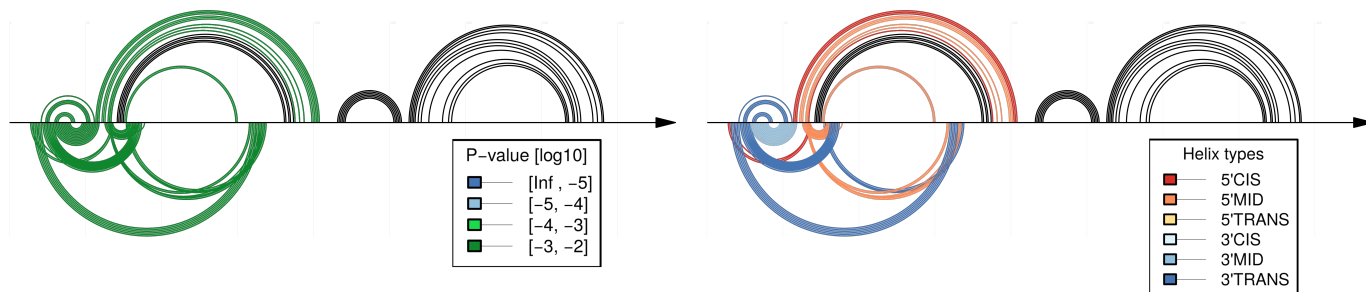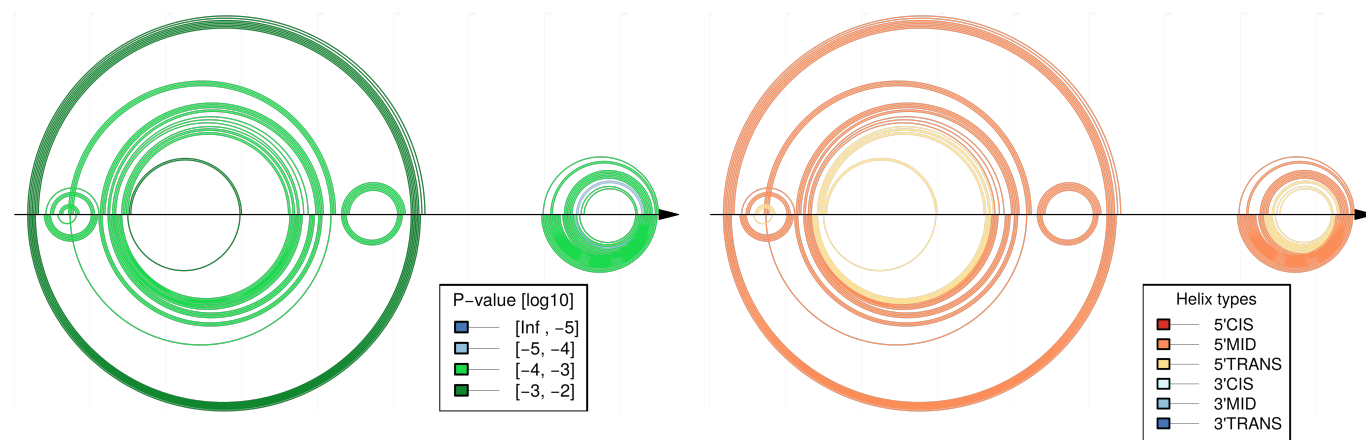
Figure S8: **Clashing predictions of the unbound conformation of the SAM Riboswitch by the comparative part of** CoBold:For a more detailed general description of the figure see Fig. 1. In the top part of both figures the reference structure corresponding to the *unbound conformation of the SAM Riboswitch* is shown. In the bottom part of both figures, only the clashing helices with the comparative part of CoBold and directly competing with one of the helices of the reference structure are shown. The clashes appearing in the picture on the left are coloured in terms of their p-value. The clashes appearing in the picture on the right are coloured in terms of their helix types instead.
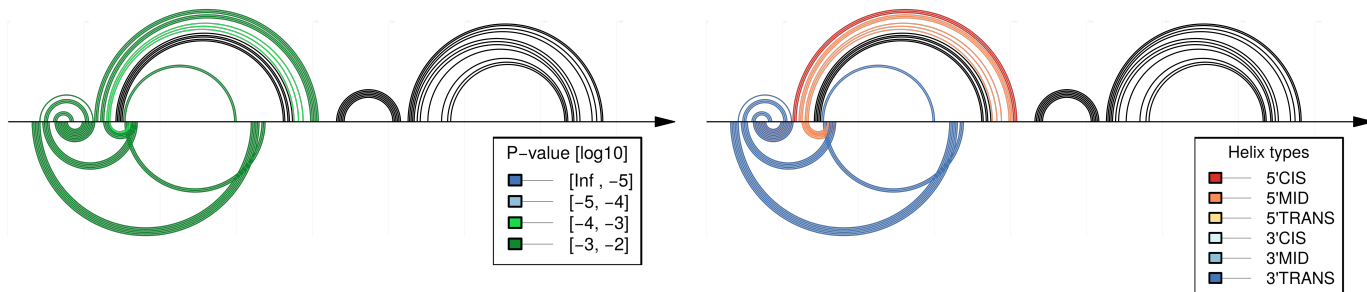

Figure S9: **Clashing predictions with the minimum p-value of the bound conformation of the SAM Riboswitch by the comparative part of** CoBold: For a more detailed general description of the figure see Fig. 1. In the top part of both figures the reference structure corresponding to the *bound conformation of the SAM Riboswitch* is shown. In the bottom part, only the clashing helices with the minimum p-value predicted with the comparative part of CoBold and directly competing with one of the helices of the reference structure are shown. The clashes appearing in the picture on the left are coloured in terms of their p-value. The clashes appearing in the picture on the right are coloured in terms of their helix types instead.
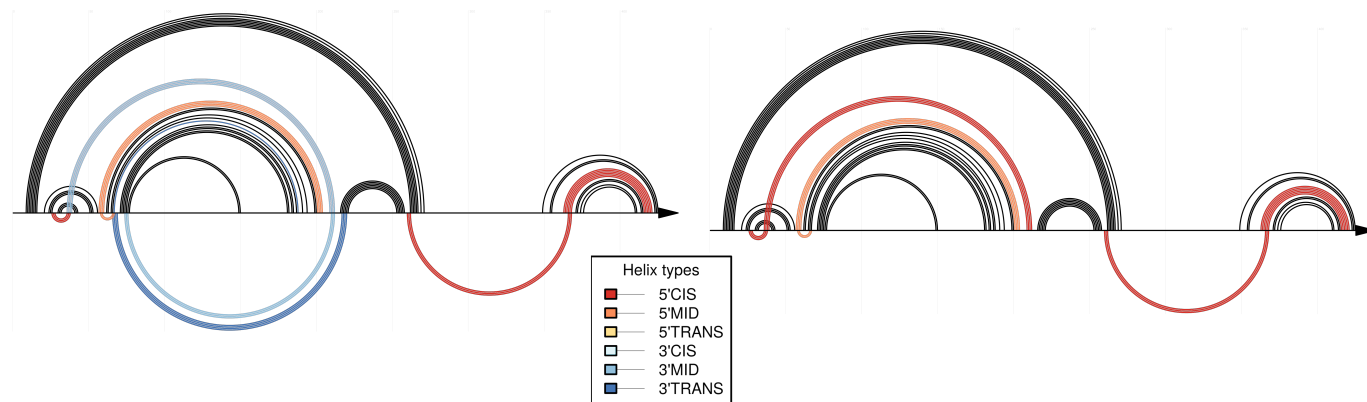
14

Figure S10: **Clashing predictions with the minimum p-value of the unbound conformation of the SAM Riboswitch by the comparative part of** CoBold**:** For a more detailed general description of the figure see Fig. 1. In the top part of both figures the reference structure corresponding to the *unbound conformation of the SAM Riboswitch* is shown. In the bottom part, only the clashing helices with the minimum p-value predicted with the comparative part of CoBold and directly competing with one of the helices of the reference structure are shown. The clashes appearing in the picture on the left are coloured in terms of their p-value. The clashes appearing in the picture on the right are coloured in terms of their helix types instead.

Figure S11: **Clashing predictions of the bound conformation of the SAM Riboswitch by the non-comparative part of** CoBold**:** For a more detailed general description of the figure see Fig. 1.In the top part of the figure the reference structure corresponding to the *bound conformation of the SAM Riboswitch* is shown. In the bottom part of both figures, only the clashing helices predicted with the non-comparative part of CoBold and directly competing with one of the helices of the reference structure are shown. In this case the arcs appearing in bothm the left and right pictures, are coloured in terms of the kind of clashes the helices are involved in. The right part only shows the intersection between the predictions of bot the comparative and non-comparative part of CoBold.
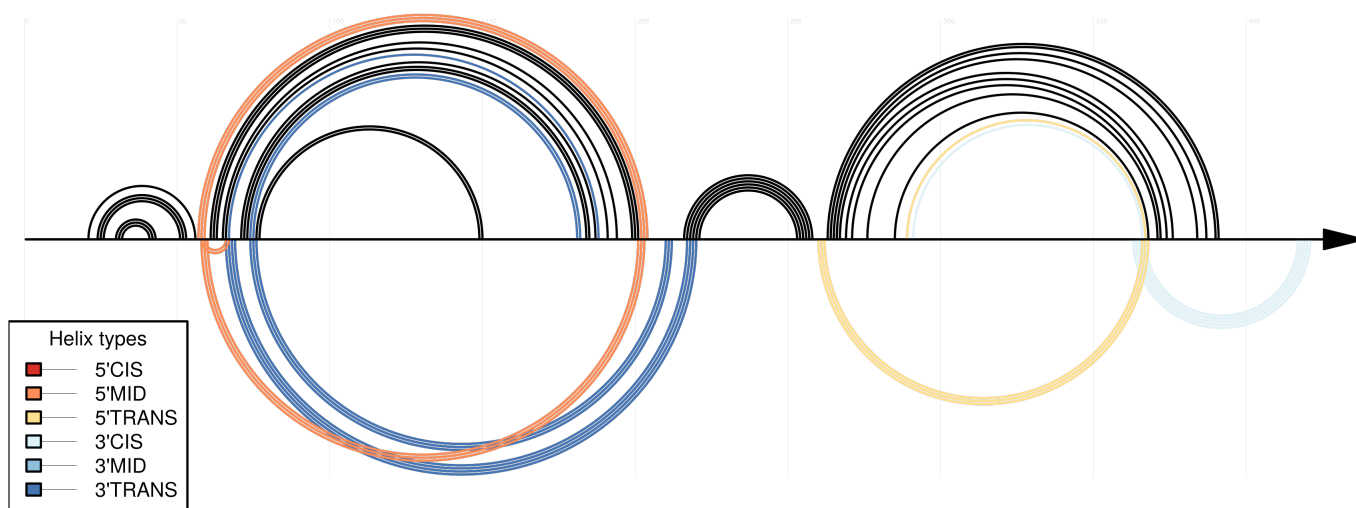
Figure S12: **Clashing predictions of the unbound conformation of the SAM Riboswitch by the non-comparative part of** CoBold**:** For a more detailed general description of the figure see Fig. 1.In the top part of the figure the reference structure corresponding to the *unbound conformation of the SAM Riboswitch* is shown. In the bottom part of both figures, only the clashing helices predicted with the non-comparative part of CoBold and directly competing with one of the helices of the reference structure are shown. In this case the arcs are coloured in terms of the kind of clashes the helices are involved in. As no common predictions appeared in this case between the comparative and non-comparative part of CoBold, only the correspondent to the non-comparative part are shown.
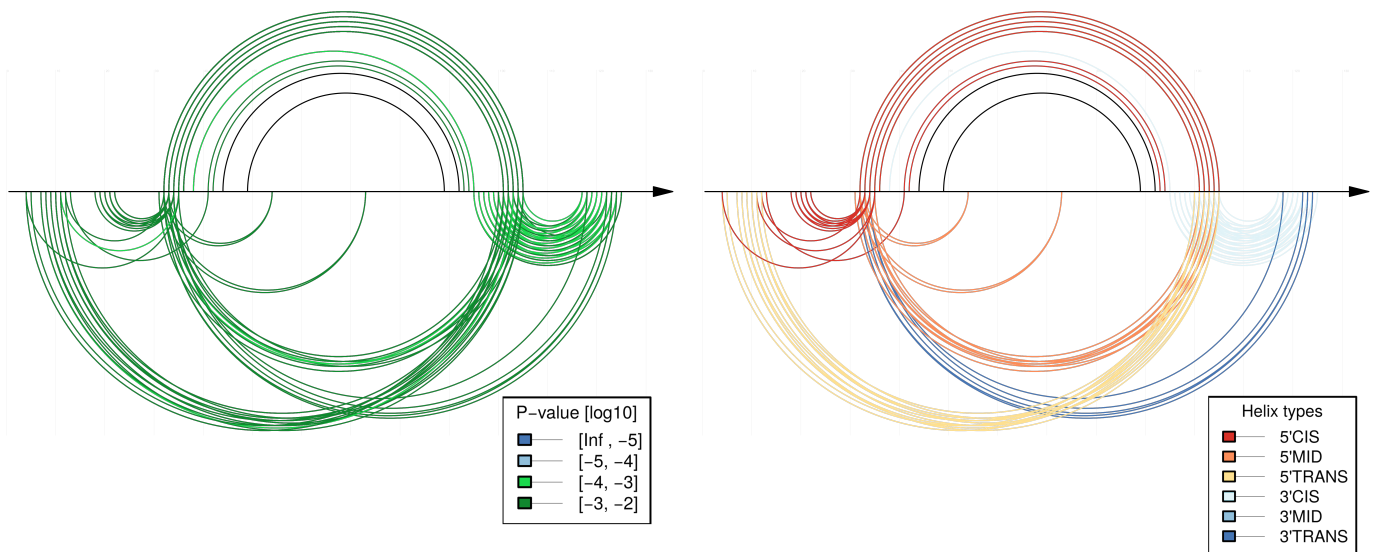
Figure S13: **Clashing predictions of the terminator conformation of the TRP operon leader by the comparative part of** CoBold**:** For a more detailed general description of the figure see Fig. 1. In the top part of both figures the reference structure corresponding to the *terminator conformation of the TRP operon leader* is shown. In the bottom part of both figures, only the clashing helices with the comparative part of CoBold and directly competing with one of the helices of the reference structure are shown. The clashes appearing in the picture on the left are coloured in terms of their p-value. The clashes appearing in the picture on the right are coloured in terms of their helix types instead.
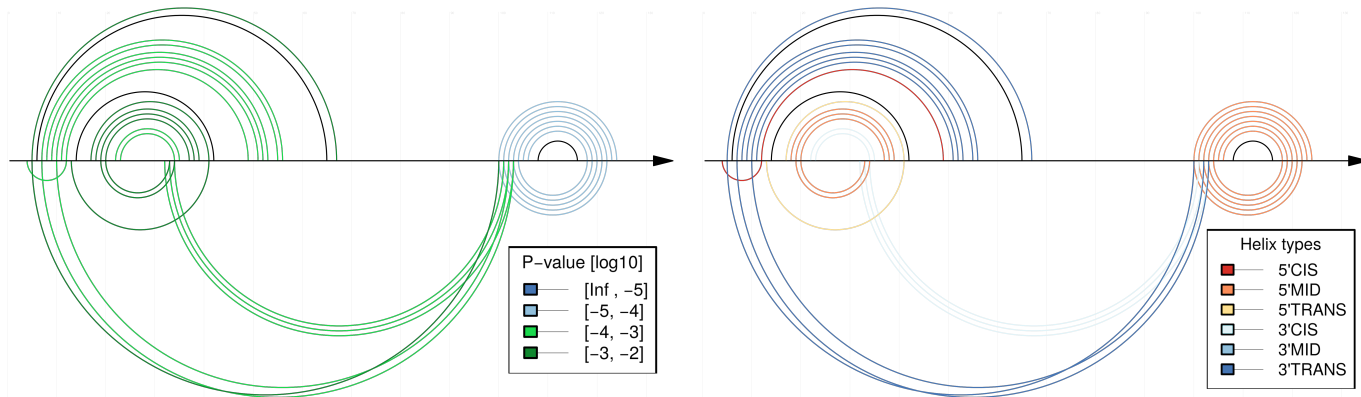
Figure S14: **Clashing predictions of the antiterminator conformation of the TRP operon leader by the comparative part of** CoBold**:** For a more detailed general description of the figure see Fig. 1. In the top part of both figures the reference structure corresponding to the *antiterminator conformation of the TRP operon leader* is shown. In the bottom part of both figures, only the clashing helices with the comparative part of CoBold and directly competing with one of the helices of the reference structure are shown. The clashes appearing in the picture on the left are coloured in terms of their p-value. The clashes appearing in the picture on the right are coloured in terms of their helix types instead.
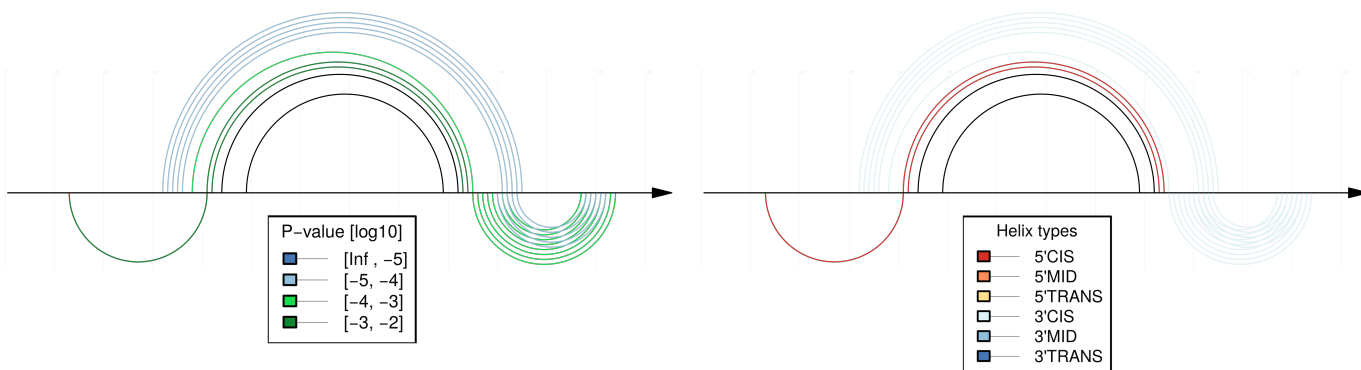
Figure S15: **Clashing predictions with the minimum p-value of the terminator conformation of the TRP operon leader by the comparative part of** CoBold**:** For a more detailed general description of the figure see Fig. 1. In the top part of both figures the reference structure corresponding to the *terminator conformation of the TRP operon leader* is shown. In the bottom part, only the clashing helices with the minimum p-value predicted with the comparative part of CoBold and directly competing with one of the helices of the reference structure are shown. The clashes appearing in the picture on the left are coloured in terms of their p-value. The clashes appearing in the picture on the right are coloured in terms of their helix types instead.

Figure S16: **Clashing predictions with the minimum p-value of the antiterminator conformation of the TRP operon leader by the comparative part of** CoBold**:** For a more detailed general description of the figure see Fig. 1. In the top part of both figures the reference structure corresponding to the *antiterminator conformation of the TRP operon leader* is shown. In the bottom part, only the clashing helices with the minimum p-value predicted with the comparative part of CoBold and directly competing with one of the helices of the reference structure are shown. The clashes appearing in the picture on the left are coloured in terms of their p-value. The clashes appearing in the picture on the right are coloured in terms of their helix types instead.
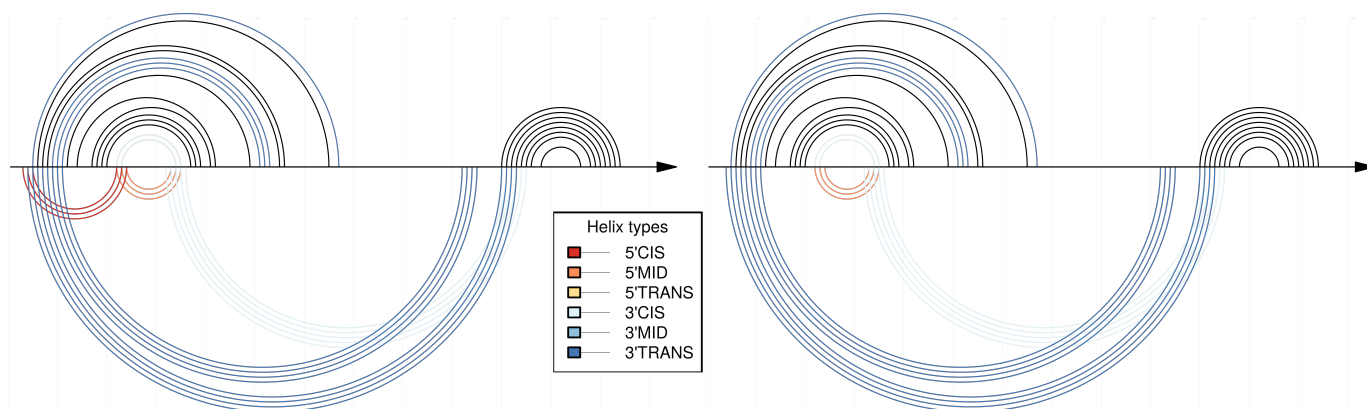
19

Figure S17: **Clashing predictions of the terminator conformation of the TRP operon leader by the non-comparative part of** CoBold**:** For a more detailed general description of the figure see Fig. 1. In the top part of the figure the reference structure corresponding to the *terminator conformation of the TRP operon leader* is shown. In the bottom part of both figures, only the clashing helices predicted with the non-comparative part of CoBold and directly competing with one of the helices of the reference structure are shown. In this case the arcs appearing in bothm the left and right pictures, are coloured in terms of the kind of clashes the helices are involved in. The right part only shows the intersection between the predictions of bot the comparative and non-comparative part of CoBold.
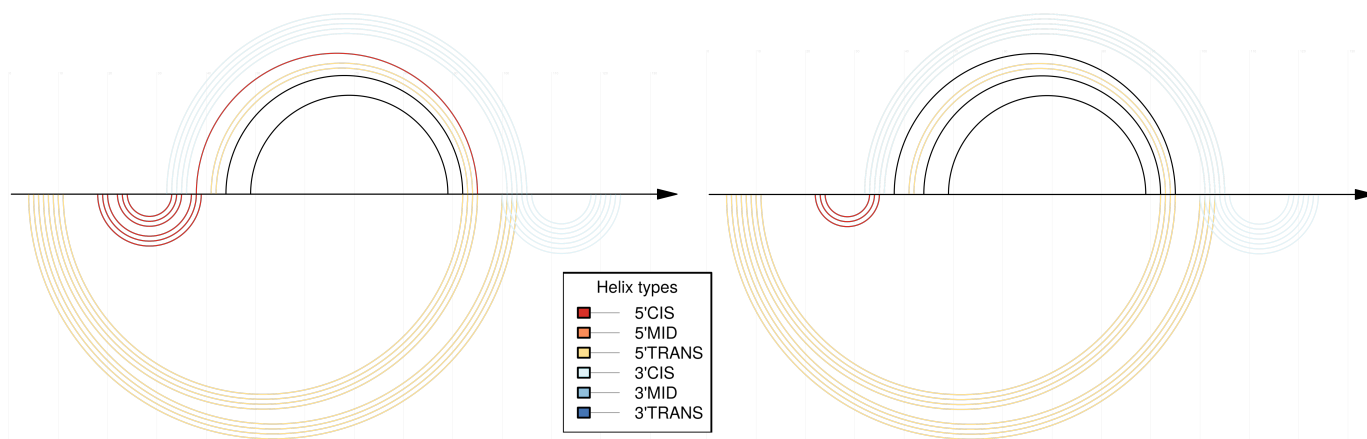

Figure S18: **Clashing predictions of the antiterminator conformation of the TRP operon leader by the non-comparative part of** CoBold**:** For a more detailed general description of the figure see Fig. 1. In the top part of the figure the reference structure corresponding to the *antiterminator conformation of the TRP operon leader* is shown. In the bottom part of both figures, only the clashing helices predicted with the non-comparative part of CoBold and directly competing with one of the helices of the reference structure are shown. In this case the arcs are coloured in terms of the kind of clashes the helices are involved in. As no common predictions appeared in this case between the comparative and non-comparative part of CoBold, only the correspondent to the non-comparative part are shown.
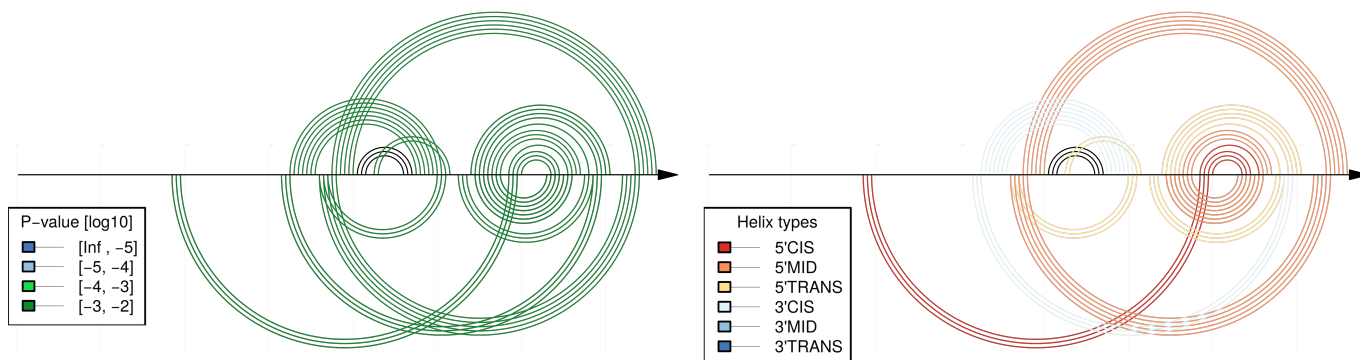
20

Figure S19: **Clashing predictions of the active conformation of the HDV riboswitch by the comparative part of** CoBold**:** For a more detailed general description of the figure see Fig. 1. In the top part of both figures the reference structure corresponding to the *active conformation of the HDV riboswitch* is shown. In the bottom part of both figures, only the clashing helices with the comparative part of CoBold and directly competing with one of the helices of the reference structure are shown. The clashes appearing in the picture on the left are coloured in terms of their p-value. The clashes appearing in the picture on the right are coloured in terms of their helix types instead.
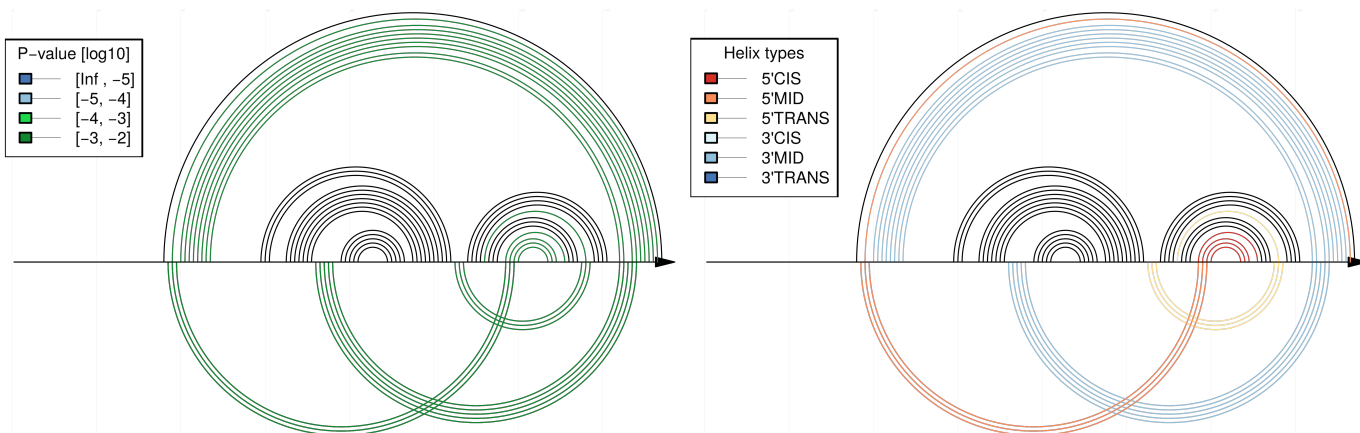

Figure S20: **Clashing predictions of the first alternative conformation of the HDV riboswitch by the comparative part of** CoBold**:** For a more detailed general description of the figure see Fig. 1. In the top part of both figures the reference structure corresponding to the *first alternative conformation of the HDV riboswitch* is shown. In the bottom part of both figures, only the clashing helices with the comparative part of CoBold and directly competing with one of the helices of the reference structure are shown. The clashes appearing in the picture on the left are coloured in terms of their p-value. The clashes appearing in the picture on the right are coloured in terms of their helix types instead.
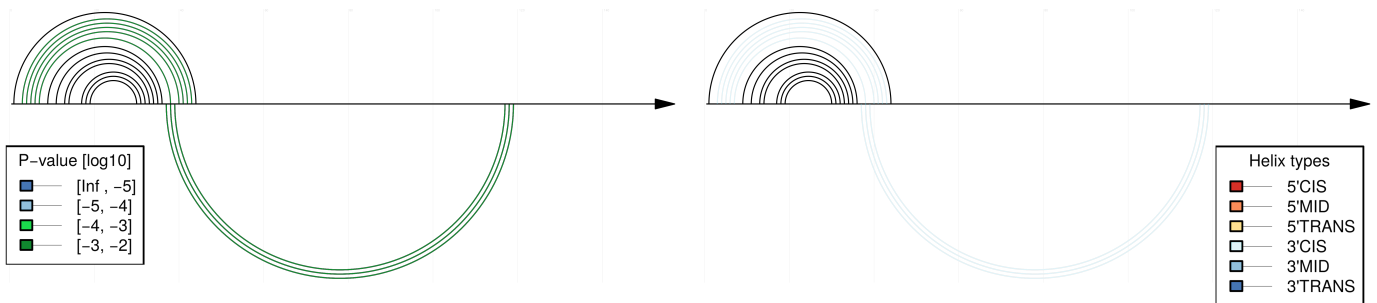
Figure S21: **Clashing predictions of the second alternative conformation of the HDV riboswitch by the comparative part of** CoBold**:** For a more detailed general description of the figure see Fig. 1. In the top part of both figures the reference structure corresponding to the *second alternative conformation of the HDV riboswitch* is shown. In the bottom part of both figures, only the clashing helices with the comparative part of CoBold and directly competing with one of the helices of the reference structure are shown. The clashes appearing in the picture on the left are coloured in terms of their p-value. The clashes appearing in the picture on the right are coloured in terms of their helix types instead.
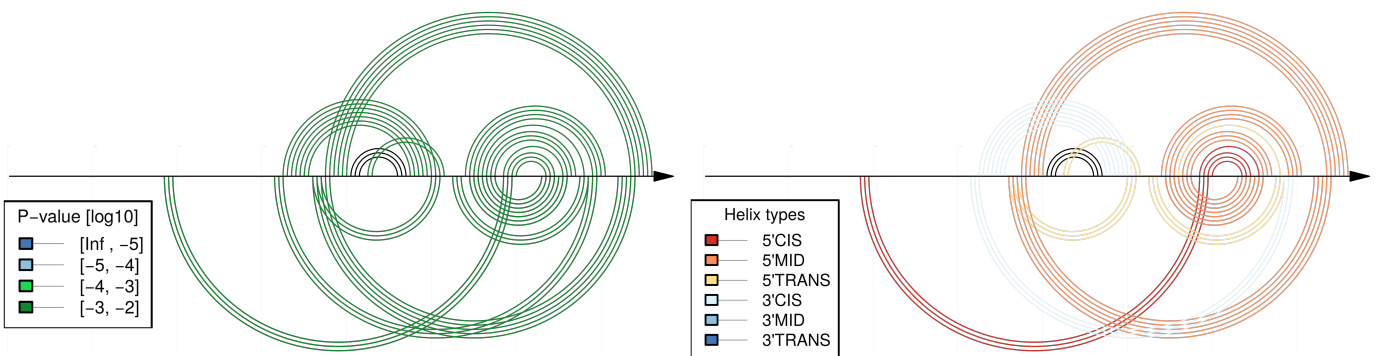

Figure S22: **Clashing predictions with the minimum p-value of the active conformation of the HDV riboswitch by the comparative part of** CoBold**:** For a more detailed general description of the figure see Fig. 1. In the top part of both figures the reference structure corresponding to the *active conformation of the HDV riboswitch* is shown. In the bottom part, only the clashing helices with the minimum p-value predicted with the comparative part of CoBold and directly competing with one of the helices of the reference structure are shown. The clashes appearing in the picture on the left are coloured in terms of their p-value. The clashes appearing in the picture on the right are coloured in terms of their helix types instead.
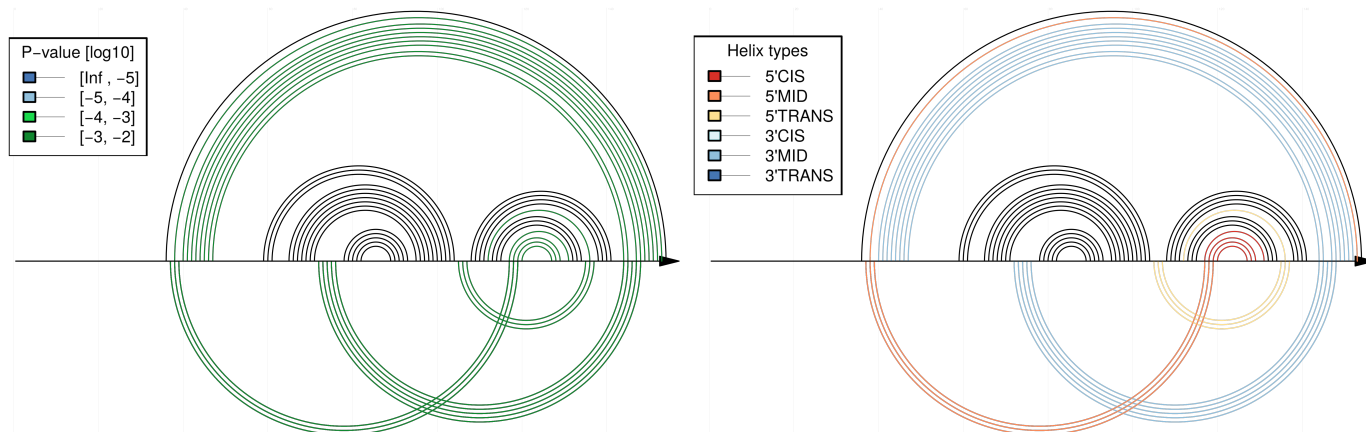
Figure S23: **Clashing predictions with the minimum p-value of the first alternative conformation of the HDV riboswitch by the comparative part of** CoBold**:** For a more detailed general description of the figure see Fig. 1. In the top part of both figures the reference structure corresponding to the *first alternative conformation of the HDV riboswitch* is shown. In the bottom part, only the clashing helices with the minimum p-value predicted with the comparative part of CoBold and directly competing with one of the helices of the reference structure are shown. The clashes appearing in the picture on the left are coloured in terms of their p-value. The clashes appearing in the picture on the right are coloured in terms of their helix types instead.
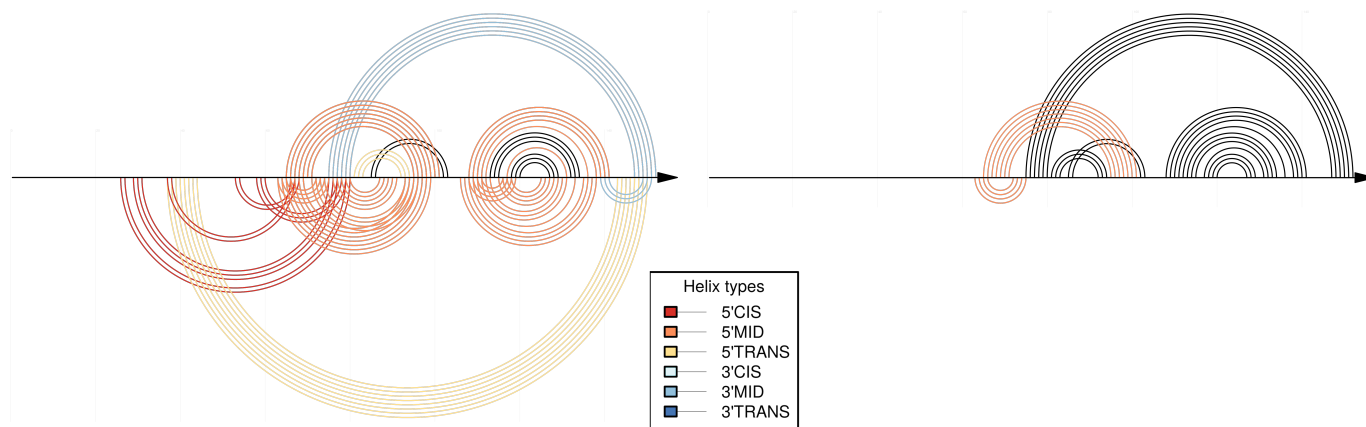

Figure S24: **Clashing predictions of the active conformation of the HDV riboswitch by the non-comparative part of** CoBold**:** For a more detailed general description of the figure see Fig. 1. In the top part of the figure the reference structure corresponding to the *active conformation of the HDV riboswitch* is shown. In the bottom part of both figures, only the clashing helices predicted with the non-comparative part of CoBold and directly competing with one of the helices of the reference structure are shown. In this case the arcs appearing in bothm the left and right pictures, are coloured in terms of the kind of clashes the helices are involved in. The right part only shows the intersection between the predictions of bot the comparative and non-comparative part of CoBold.
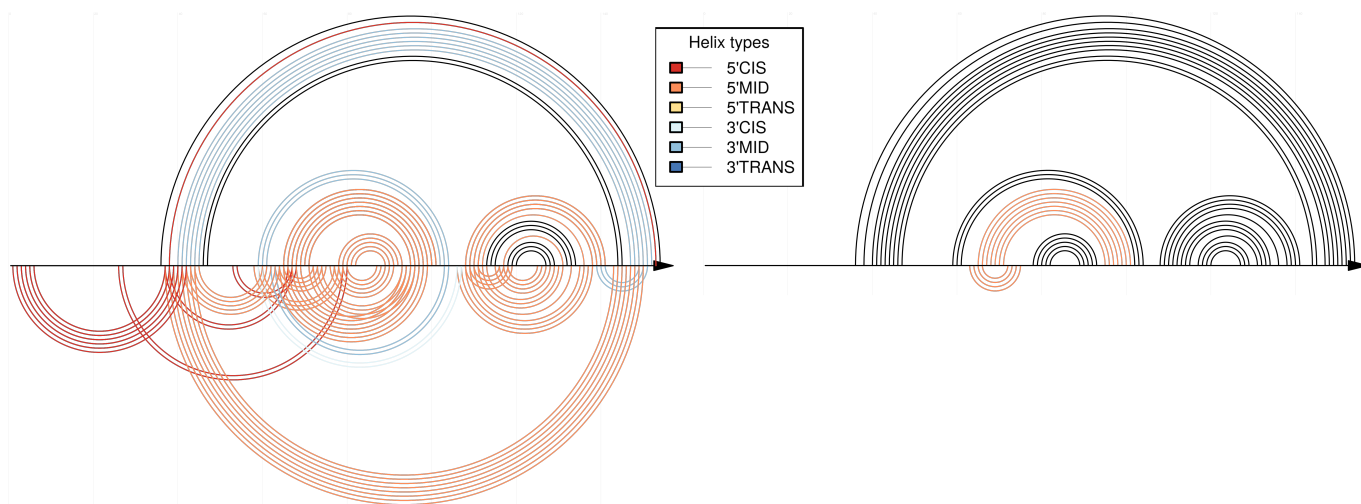
23

Figure S25: **Clashing predictions of the first alternative conformation of the HDV riboswitch by the non-comparative part of** CoBold**:** For a more detailed general description of the figure see Fig. 1. In the top part of the figure the reference structure corresponding to the *first alternative conformation of the HDV riboswitch* is shown. In the bottom part of both figures, only the clashing helices predicted with the non-comparative part of CoBold and directly competing with one of the helices of the reference structure are shown. In this case the arcs are coloured in terms of the kind of clashes the helices are involved in. As no common predictions appeared in this case between the comparative and non-comparative part of CoBold, only the correspondent to the non-comparative part are shown.