

# SUPPLEMENT

## Pipeline output

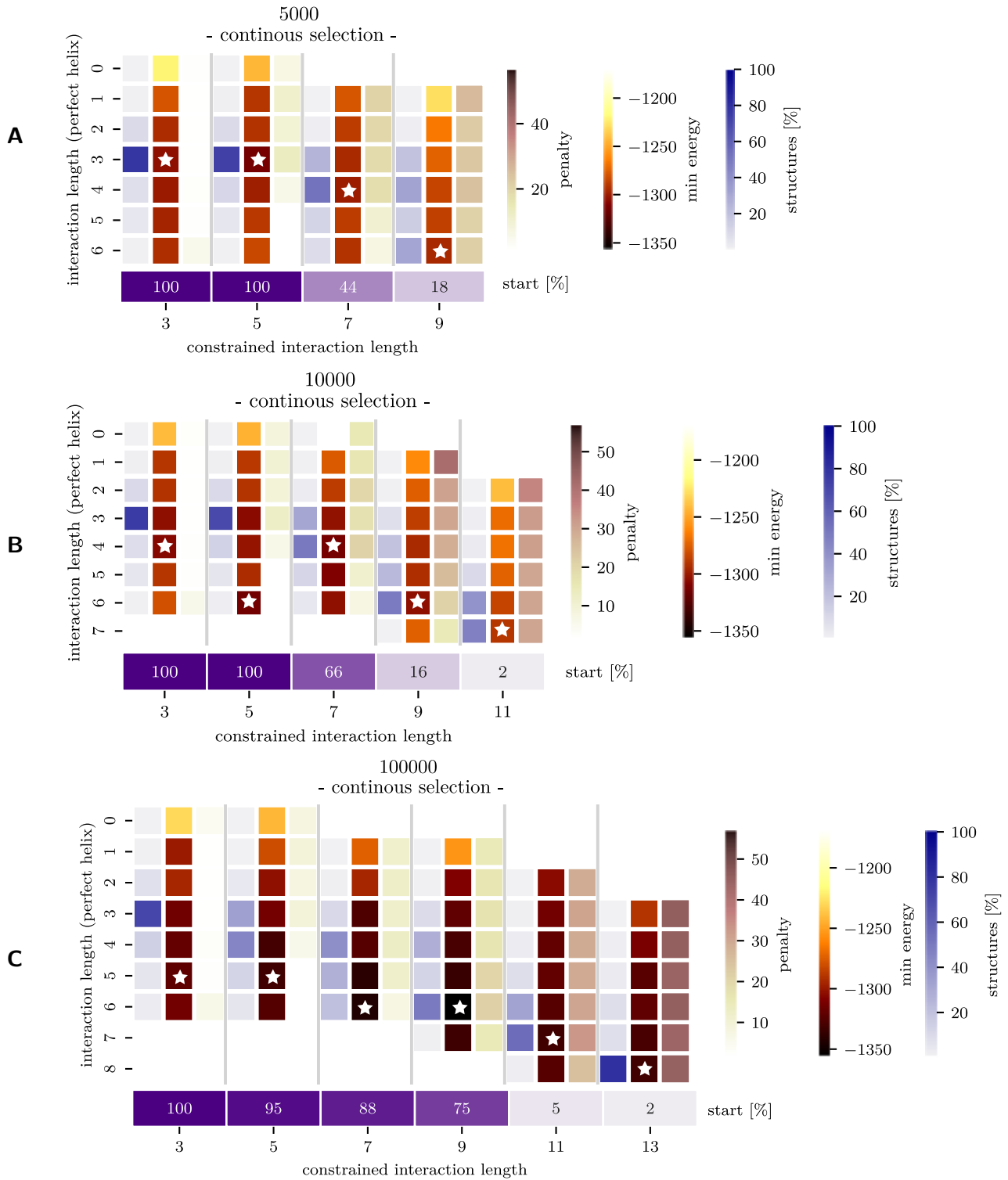
---

- $n_{\text{step}}$
- dotbracket representation
- SimRNA based values
  - energy
  - energy plus penalty
  - temperature
- 'constancy':  $n_{\text{step}}$  until bps change in the structure
- basepairlist of
  - whole structure
  - the interaction site
  - intramolecular bps of chain-A
  - intramolecular bps of chain-B
  - difference to the start structure for this expansion step
  - difference to the constrained structure
  - difference to the constrained interaction site
  - difference to the  $n_{\text{step}}$  structure before
- intermolecular bps that do not belong to the main interaction e.g separated by intramolecular bps
- multiplets
- bps that occur neither in the start nor in the target structure
- basepair count of
  - chain-A
  - chain-B
  - interaction length (perfect helix)
  - interaction length with loops allowed
  - intermolecular bp do not belong to the main interaction
  - bps that occur neither in the start nor in the target structure
- count of bps that differ to
  - the start structure for this expansion step
  - the  $n_{\text{step}}$  structure before
  - the constrained structure
  - the constrained interaction

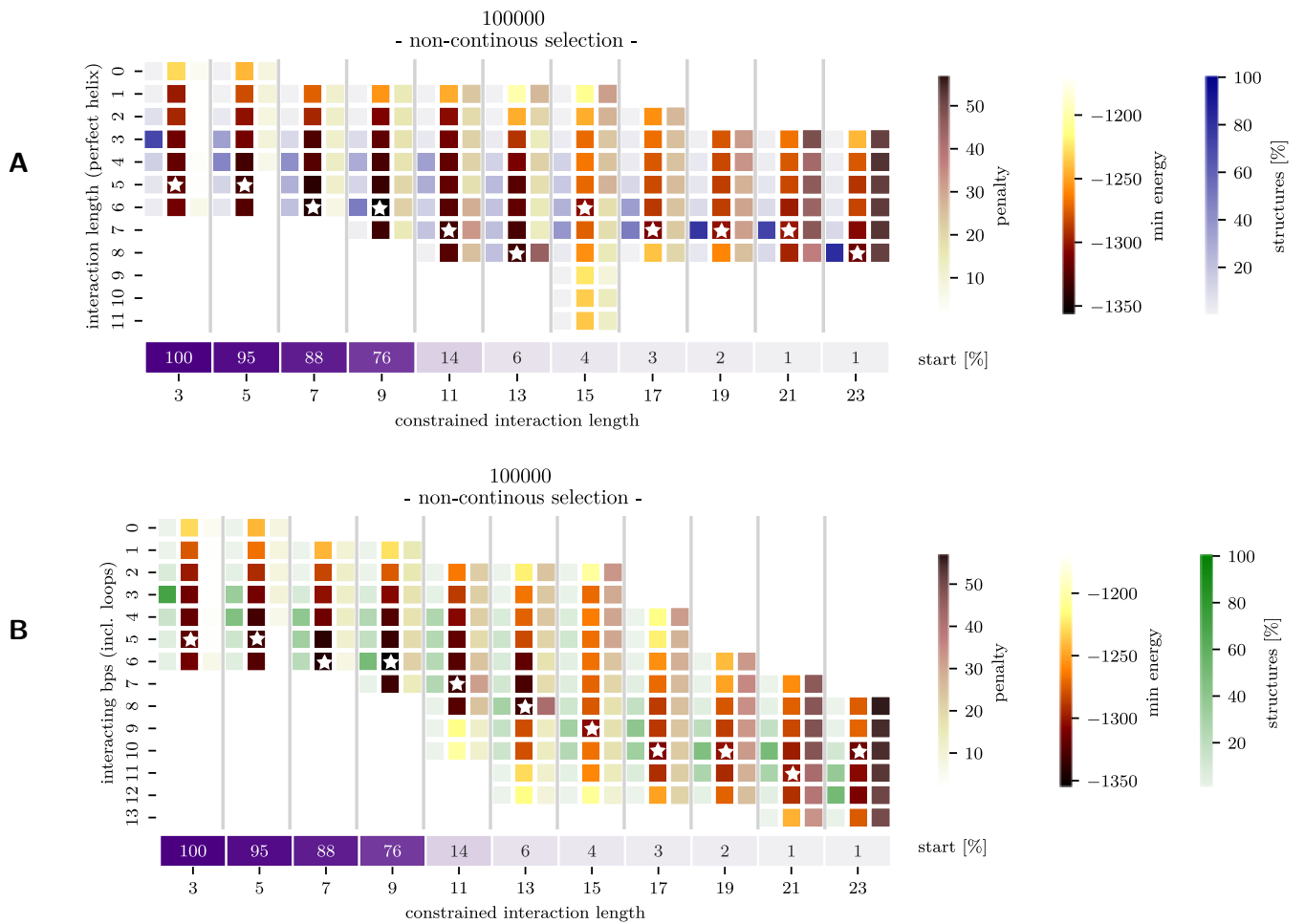
---

**List 1:** The main pipeline output is in the form of csv files for each  $n_{\text{sim}}$ , containing for each sampled 3D structure information as listed above, in particular about the formed base pairs. Furthermore, for each  $n_{\text{run}}$  with all its  $n_{\text{sim}}$ , two summaries are offered, one with the frequency of occurring dotbracket structures and the other with all occurring bps. After each interaction expansion step, the `interaction-csv` file contains all the structures that can be considered for further expansion. The best one is selected for next expansion step and also translated into a full atom PDB.

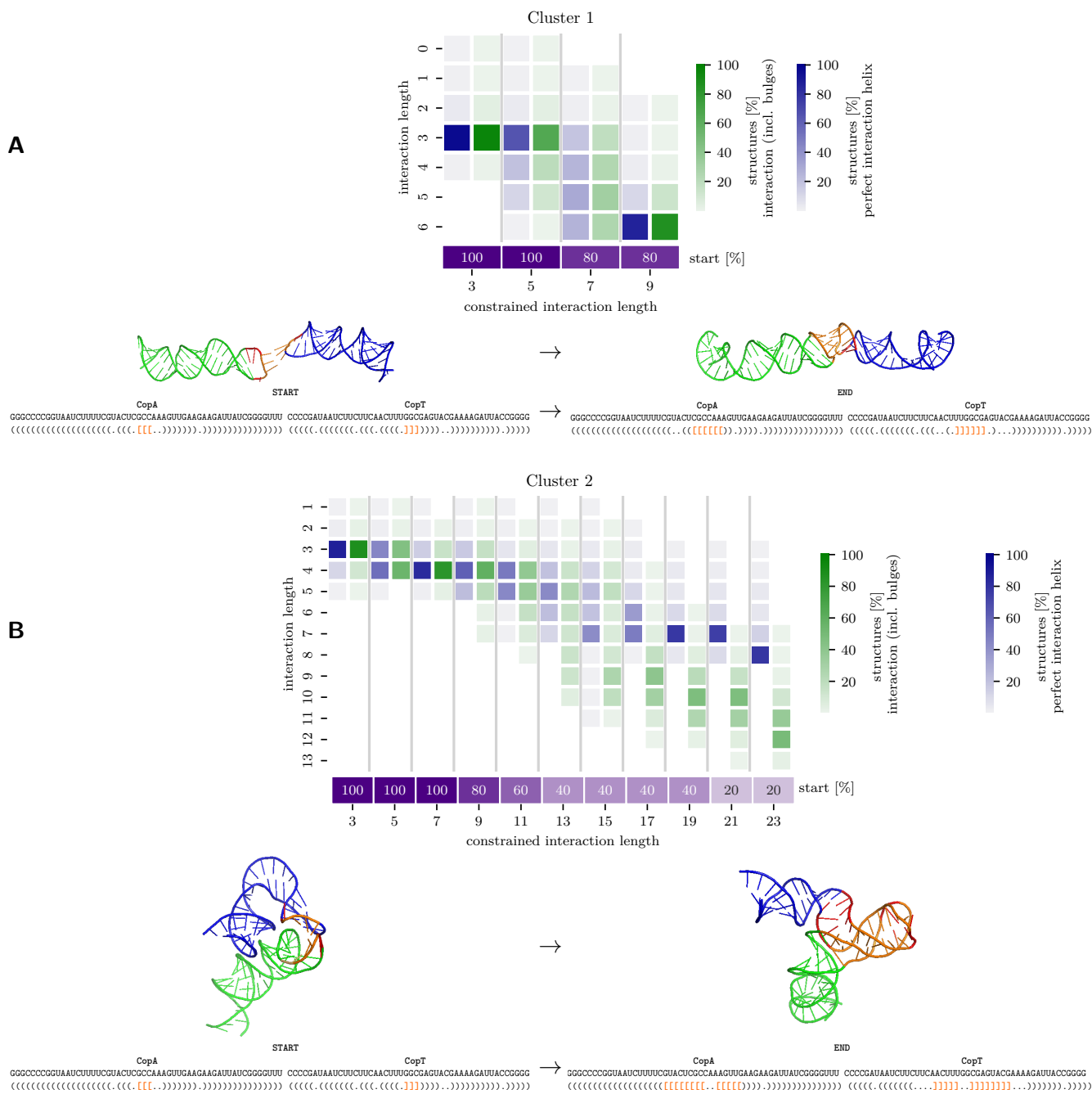




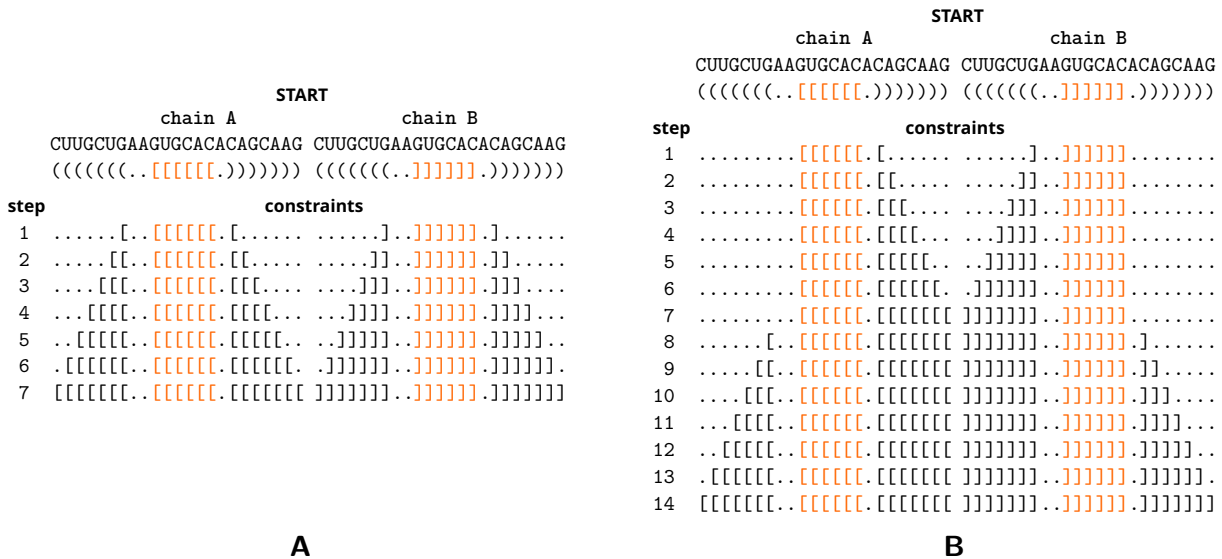
**FIGURE S2:** CopA–CopT extension of the perfect interaction helix (setting I). (A)  $n_{\text{step}} = 5000$ . The purple bar at the bottom shows the percentage of  $n_{\text{cluster}} \times n_{\text{run}} = 50$  runs that have passed the previous extension checkpoints. For each extension step with the respective constrained interaction length (x-axis) (cf. Fig. S1), statistics are shown in 3 columns separated by the grey line. The first column shows the histogram (blue) for the perfect helix interaction length (y-axis). Based on the number of interacting basepairs formed, the second column (yellow-red) shows the best SimRNA energy, with an asterisk marking the lowest in the column, and the third column (brown) shows the penalty used in this set. (B)  $n_{\text{step}} = 10000$ . (C)  $n_{\text{step}} = 100000$ .



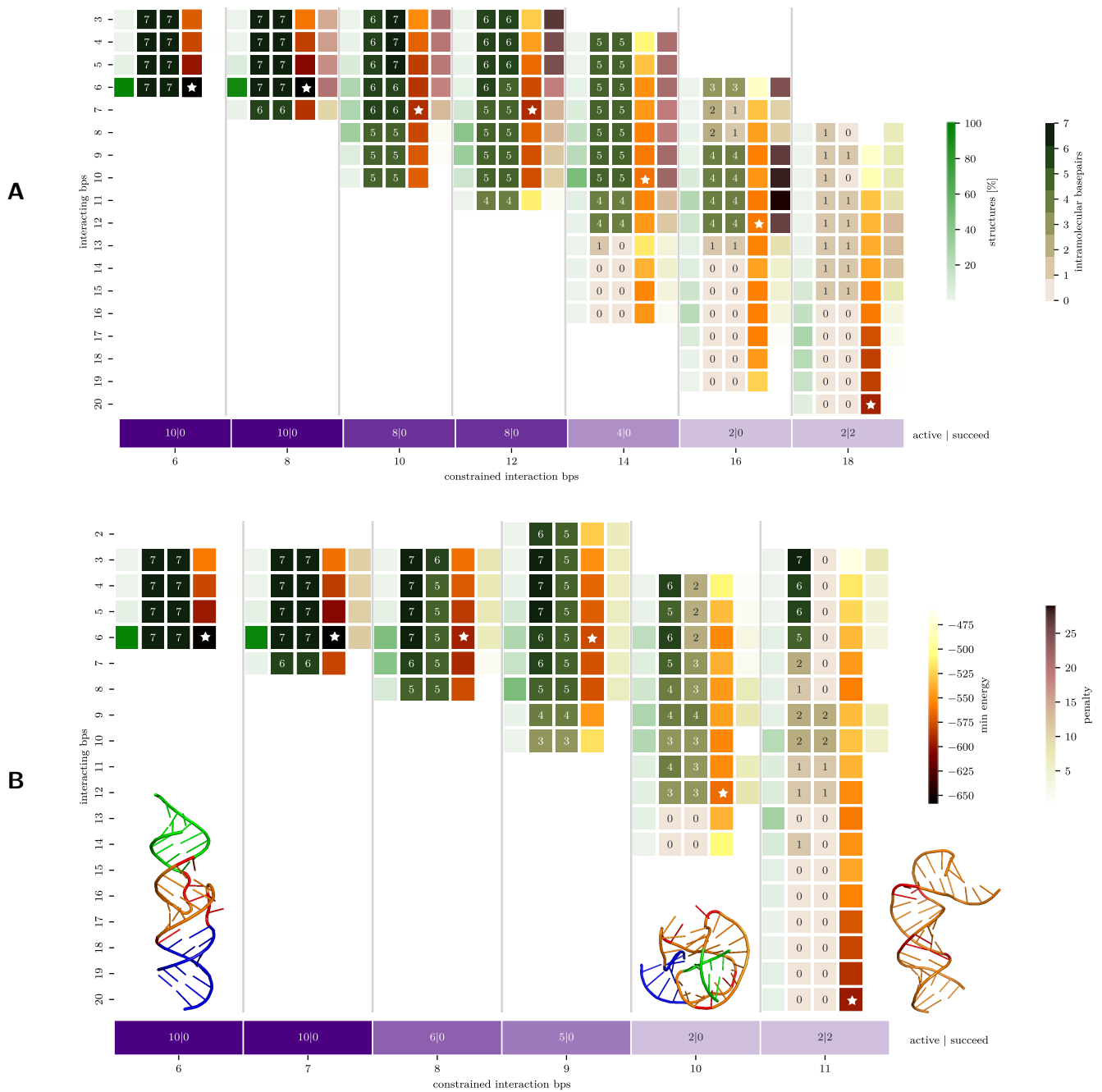
**FIGURE S3:** CopA–CopT extension of the interaction site (setting II and  $n_{\text{step}} = 100000$ ) with small interior loops or bulges allowed within the interaction site during the selection process. (A) The columns with the blue histogram represent the perfect interaction helix lengths (bps) in contrast to (B) the green histogram with the number of interacting bps with loops allowed within the interaction site. For further legend description see supplemental Figure S2



**FIGURE S4:** CopA–CopT interaction formation from two different starting conformations. Two very different starting conformations were selected from a total  $n_{\text{cluster}} = 10$  (summary results for all 10 are shown in Fig. 4B): (A) a start structure with head-to-head orientations of the two helices makes extension difficult and all runs achieve at most 6 interaction pairs. (B) a start structure with almost parallel orientation of the two helices can be extended to interaction with up to 12 pairs (interrupted by a short interior loop). The progress of the interaction expansion can be followed in the respective plots. The purple bar at the bottom shows the percentage of  $n_{\text{cluster}} \times n_{\text{run}} = 50$  runs that have passed the previous extension with the respective constrained interaction length (x-axis). The blue histograms show the length of the longest uninterrupted helix, the green histograms show the total number of intra-molecular pairs. Detail results from the individual start clusters, as well as 3D structures of the respective start and representative end structures can be found on the github page [www.github.com/irenekb/RR1-3D/tree/master/examples/from\\_2D/CopA-CopT](http://www.github.com/irenekb/RR1-3D/tree/master/examples/from_2D/CopA-CopT).



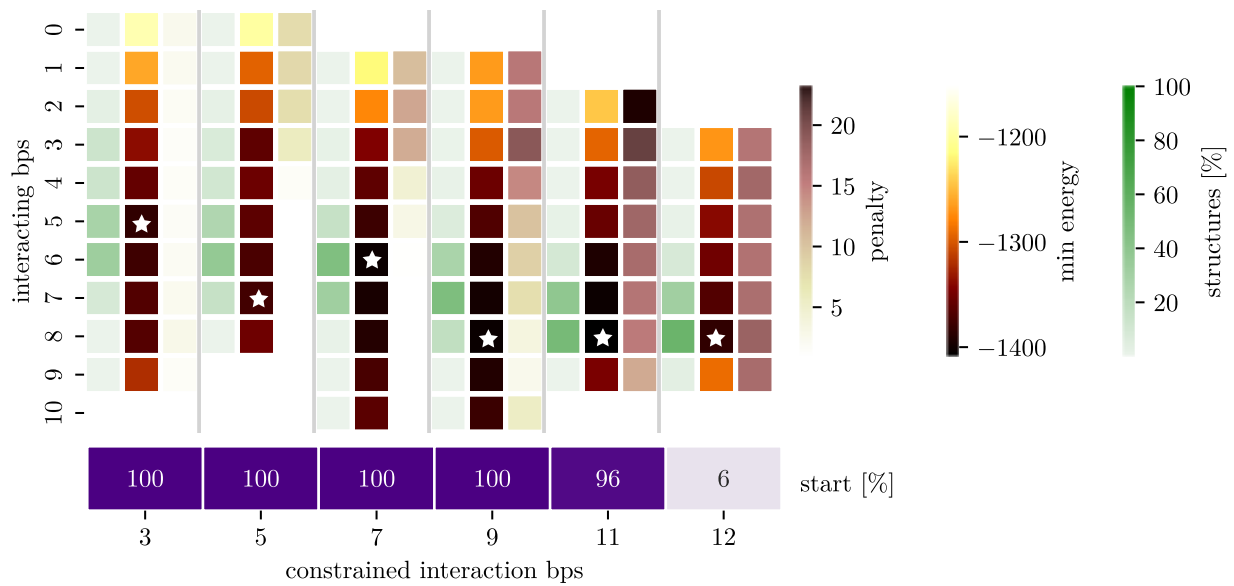
**FIGURE S5:** Sequence and dot-bracket representation of the HIV-1 DIS PDB 1ZCI, followed by the 2D pathway for the stepwise extension of the interaction site (interaction in orange). (A) Setting I with a symmetric extension (B) Setting II with an asymmetric extension.



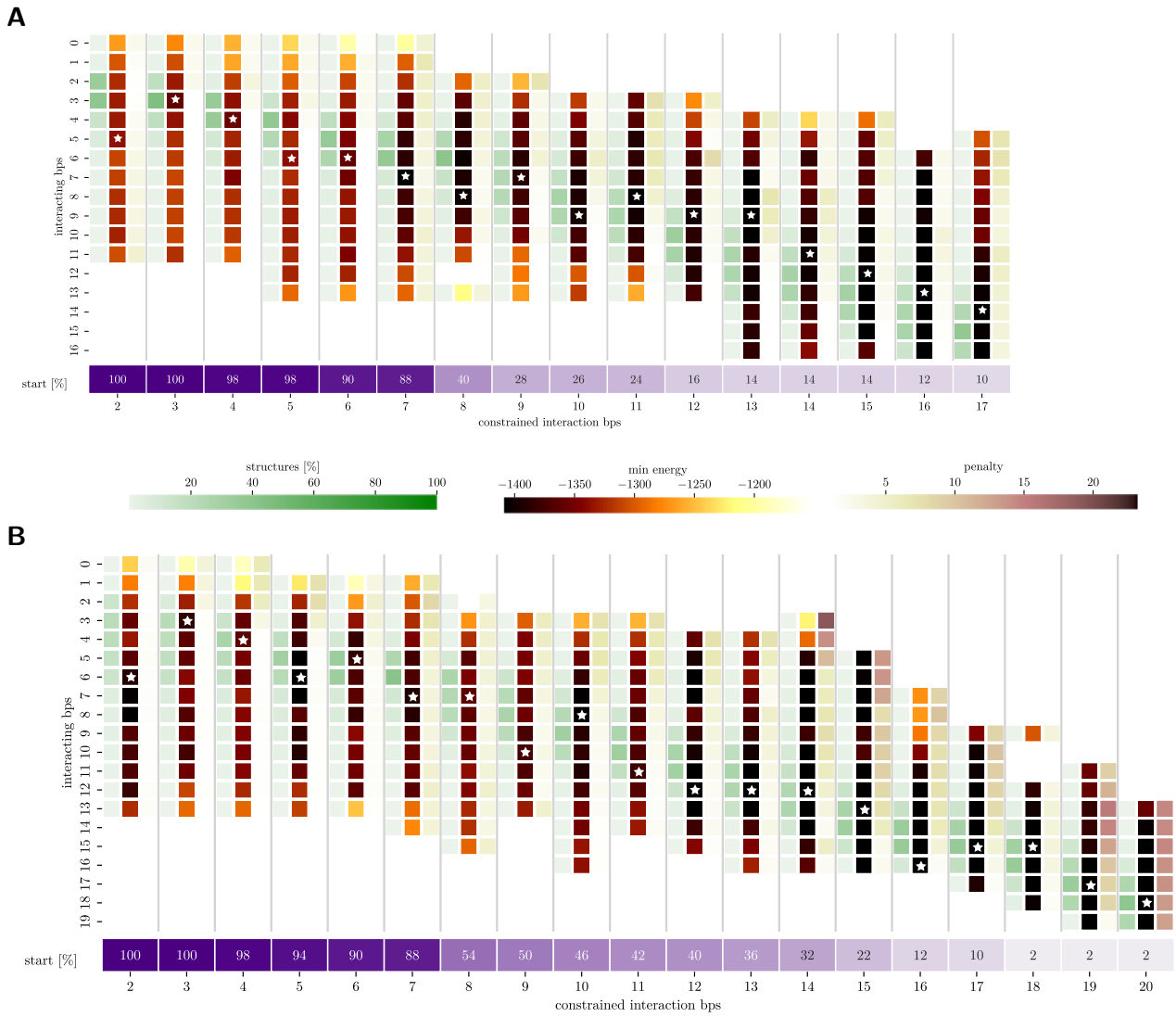
**FIGURE S6:** HIV-1 DIS interaction extension with  $n_{\text{sim}} = 10$  simulations and  $n_{\text{step}} = 1\text{million}$ . (A) Setting I with a symmetric extension. The purple frame at the bottom shows the number of runs still active and the number of runs successfully reaching the full duplex in the corresponding extension step along the given 2D path (cf. Fig. S5). We report statistics for structures at each extension step, where structures are further grouped by the number of interaction base pairs (incl. loops) formed (y-axis). For each extension step, corresponding to a number of constrained interaction base pairs (x-axis), statistics are reported in five columns (separated by the grey line) using different color gradients. The first shows the percentage of structures in the set; the second and third the average intra-molecular helix length of chain A and chain B; the fourth the best SimRNA energy, with an asterisk marking the lowest in the column; and the third the penalty used for that set. (B) Setting II with an asymmetric extension. The three 3D structures shown here represent the PDB starting structure for this simulation path (6 interaction bps), the structure with the best energy (marked with asterisks) at an interaction constraint of 10 bps (reaching 12 interaction bps), and a constraint of 12 bps (reaching the full duplex of 20 bps). The interaction is colored in orange, the loops in red, and the two chains in blue and green. PyMOL session files allowing to interactively explore the three 3D structures can be found on the github page [www.github.com/irenekb/RRI-3D/tree/master/examples/from\\_pdb/HIV](http://www.github.com/irenekb/RRI-3D/tree/master/examples/from_pdb/HIV).



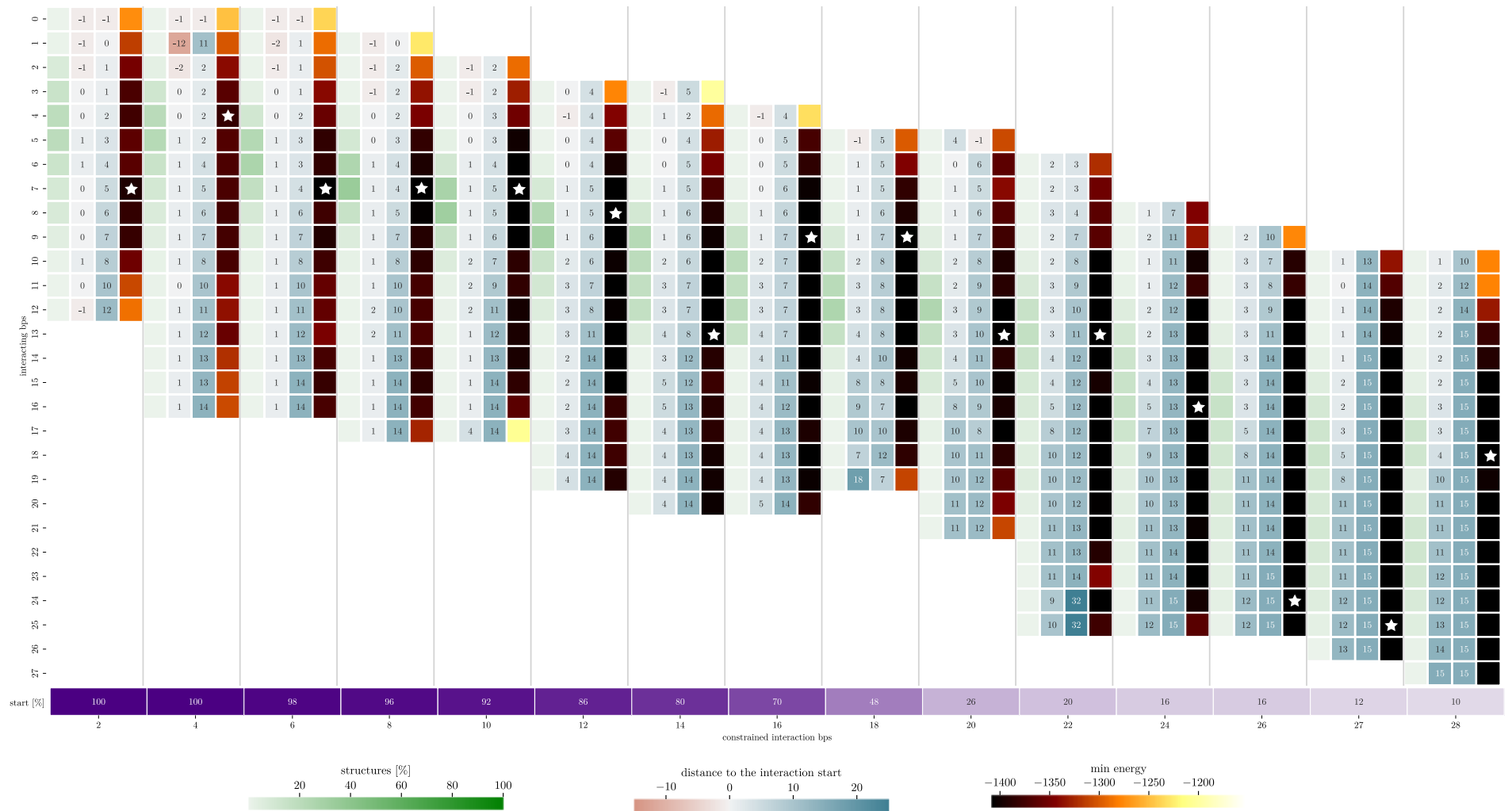




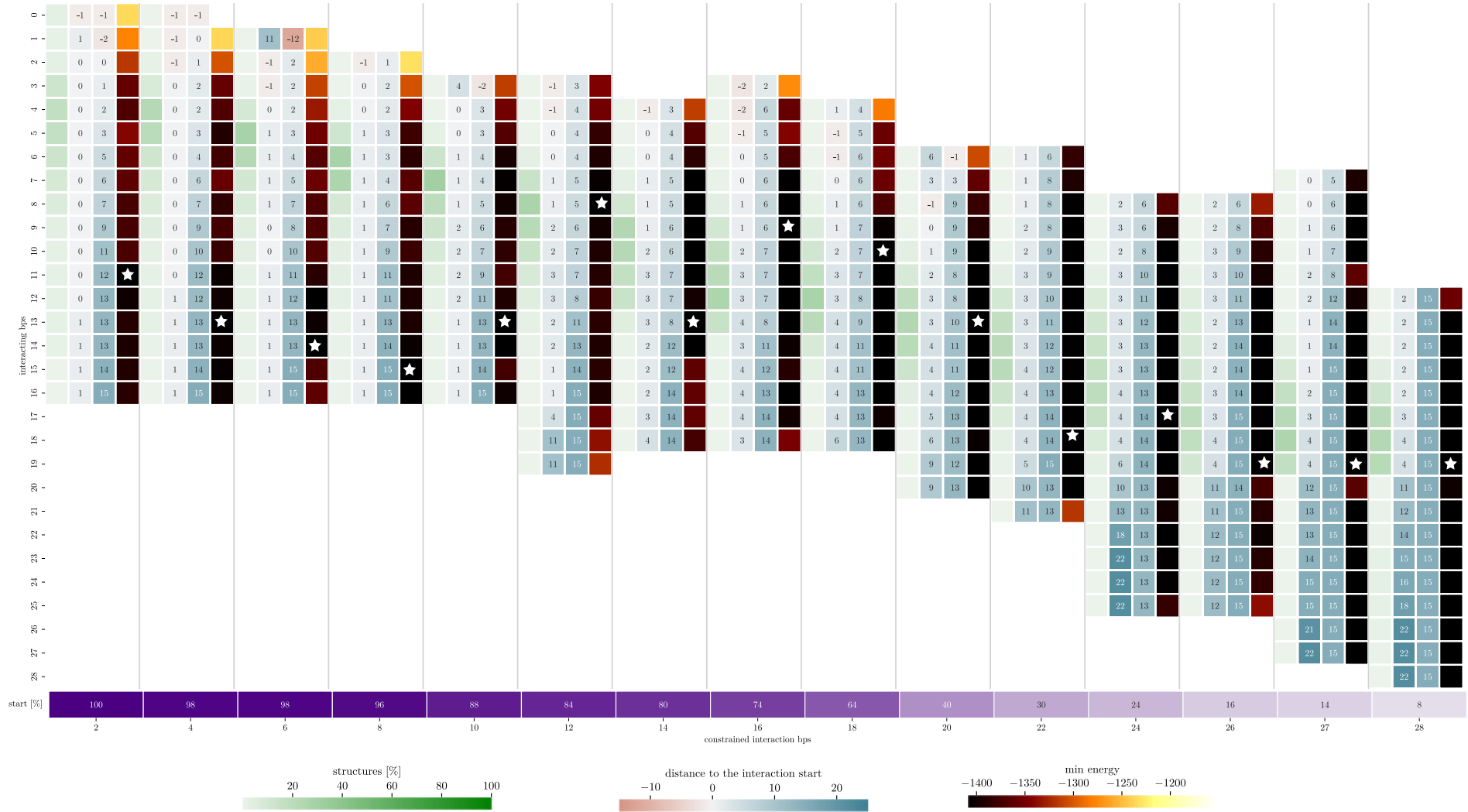
**FIGURE S11:** Interaction extension of the DsrA Fold-A-*rpoS* complex starting from SL1 (cf. Fig. S7) with  $n_{\text{step}} = 100000$ . The purple bar at the bottom shows the percentage of  $n_{\text{cluster}} \times n_{\text{run}} = 50$  structures that have passed the previous extension checkpoints at a respective constrained interaction length (x-axis). We report statistics for structures at every extension step, where structures are furthermore grouped by the number of formed interaction base pairs (y-axis). For each extension step, corresponding to a number of interaction base pairs in the constraint, statistics are reported in three columns, using different color gradients. The first one visualizes the percentage of structures in the class; the second, the best SimRNA energy (the asterisk marks the lowest in the column); and the third, the penalty used for this set.



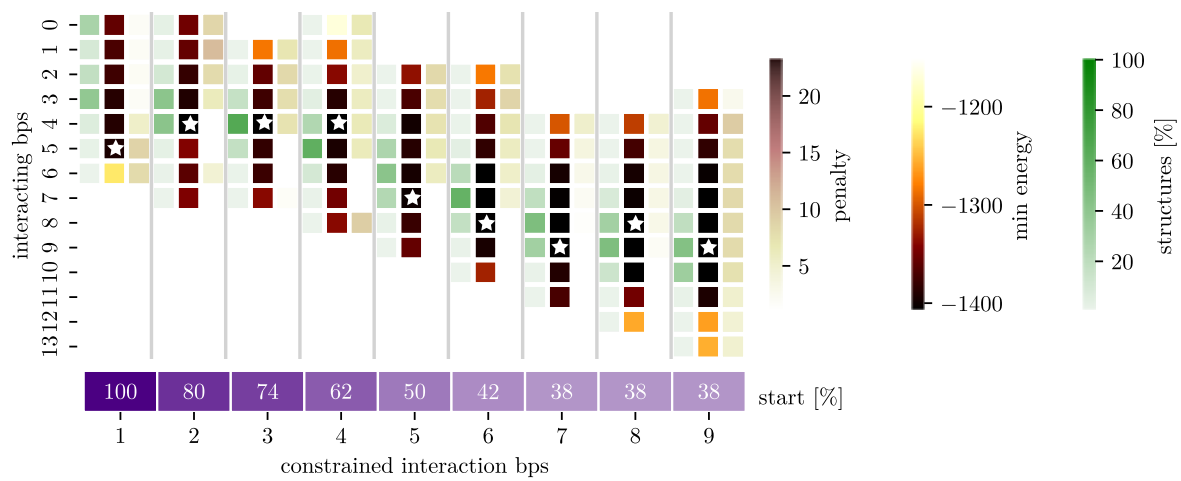
**FIGURE S12:** Interaction extension of the *DsrA-rpoS* ( $n_{\text{cluster}} \times n_{\text{run}} = 50$  and  $n_{\text{step}} = 100000$ ) starting from SL2 (cf. Fig. S9). (A) DsrA Fold-A (B) DsrA Fold-B. The legend description of these plots correspond to the one in Figure S11. Note that a constrained length of 16 bps corresponds to an extension up to the start of SL1



**FIGURE S13:** DsrA Fold-A-rpoS interaction extension ( $n_{\text{cluster}} \times n_{\text{run}} = 50$  and  $n_{\text{step}} = 100000$ ) with a two bp long interaction start site in the LR (cf. Fig. S8). The plot setup is similar to the one in Figure S11 except that the second column shows how many nucleotides the interaction extended on average towards the 5' end of DsrA (including bulges), starting from the 5' end of the interaction start site (C25) and the third one visualizes how many nucleotides the interaction extends in the 3' direction, starting from the 3' end of the DsrA interaction start site (G26).



**FIGURE S14:** Interaction extension of DsrA Fold-B-*rpoS* starting from the LR (cf. Fig. S8) with  $n_{\text{cluster}} \times n_{\text{run}} = 50$  and  $n_{\text{step}} = 100000$ . The legend description of this plot correspond to the one in Figure S13.



**FIGURE S15:** Interaction extension of the *DsrA-rpoS* complex ( $n_{\text{cluster}} \times n_{\text{run}} = 50$  and  $n_{\text{step}} = 100000$ ) with *DsrA* in the Rfam fold and starting from the LR (cf. Fig. S10). The legend description of this plot correspond to the one in Figure S11.