

Supplementary Data for Inverse Folding with RNA-As-Graphs Produces a Large Pool of Candidate Sequences with Target Topologies

Swati Jain^{†1}, Yunwen Tao^{†1}, and Tamar Schlick^{1,2,3,*}

¹Department of Chemistry, New York University, New York, NY, USA

²Courant Institute of Mathematical Sciences, New York University, New
York, NY, USA

³NYU-ECNU Center for Computational Chemistry at NYU Shanghai,
Shanghai, China

*Corresponding author: schlick@nyu.edu

[†]These authors contributed equally

Mutation Optimization after GA (Sequence 13 for 7_4 topology)

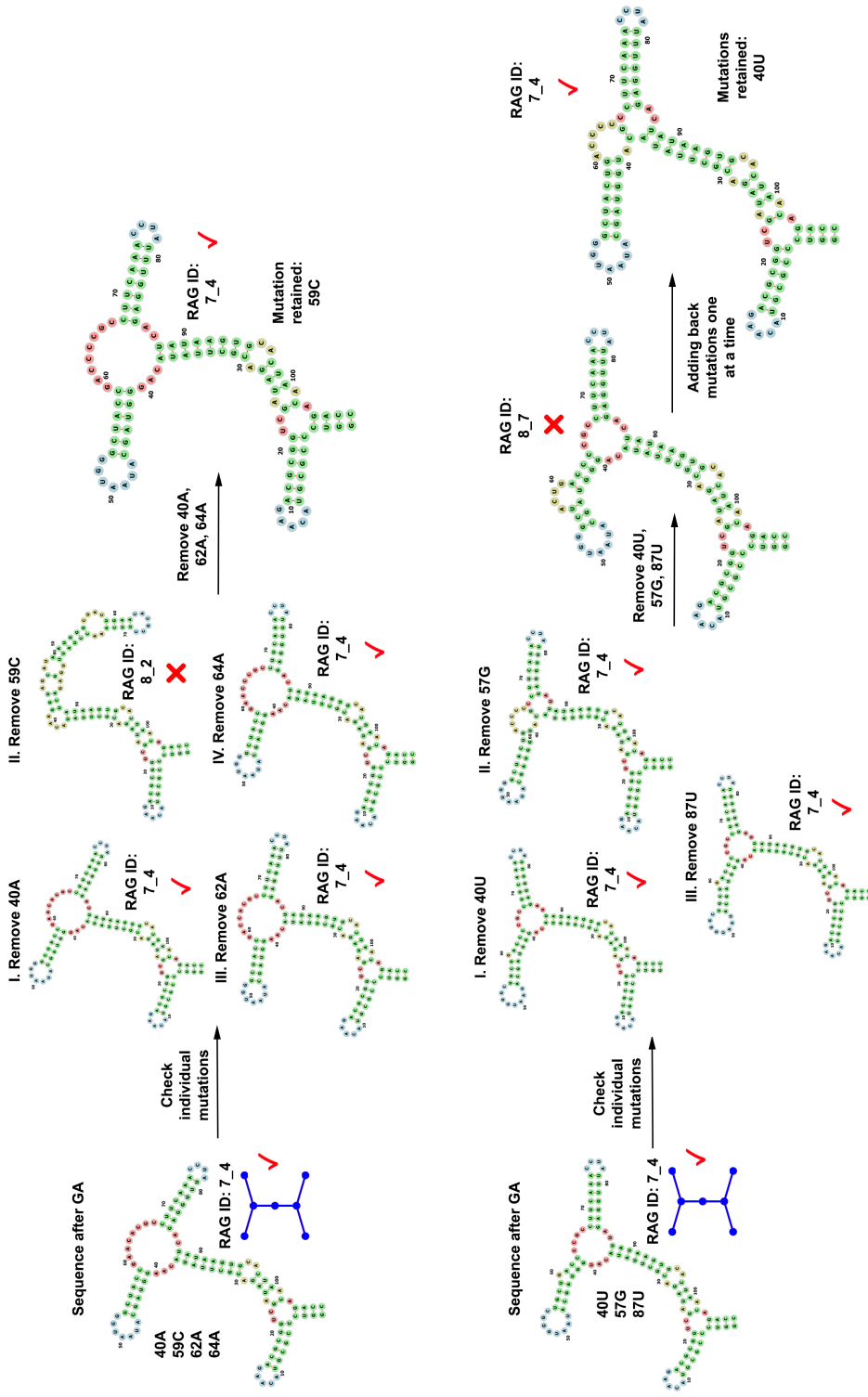


Figure S1: Mutation optimization procedure applied to two sequences generated by GA for starting sequence 13 for the 7_4 topology (see Figure 4 of the main paper) to retain only essential mutations. Only NUPACK predicted 2D structures and their corresponding tree graph topologies are shown (RNAfold predicted 2D structures for shown sequences have the correct 7_4 topology).

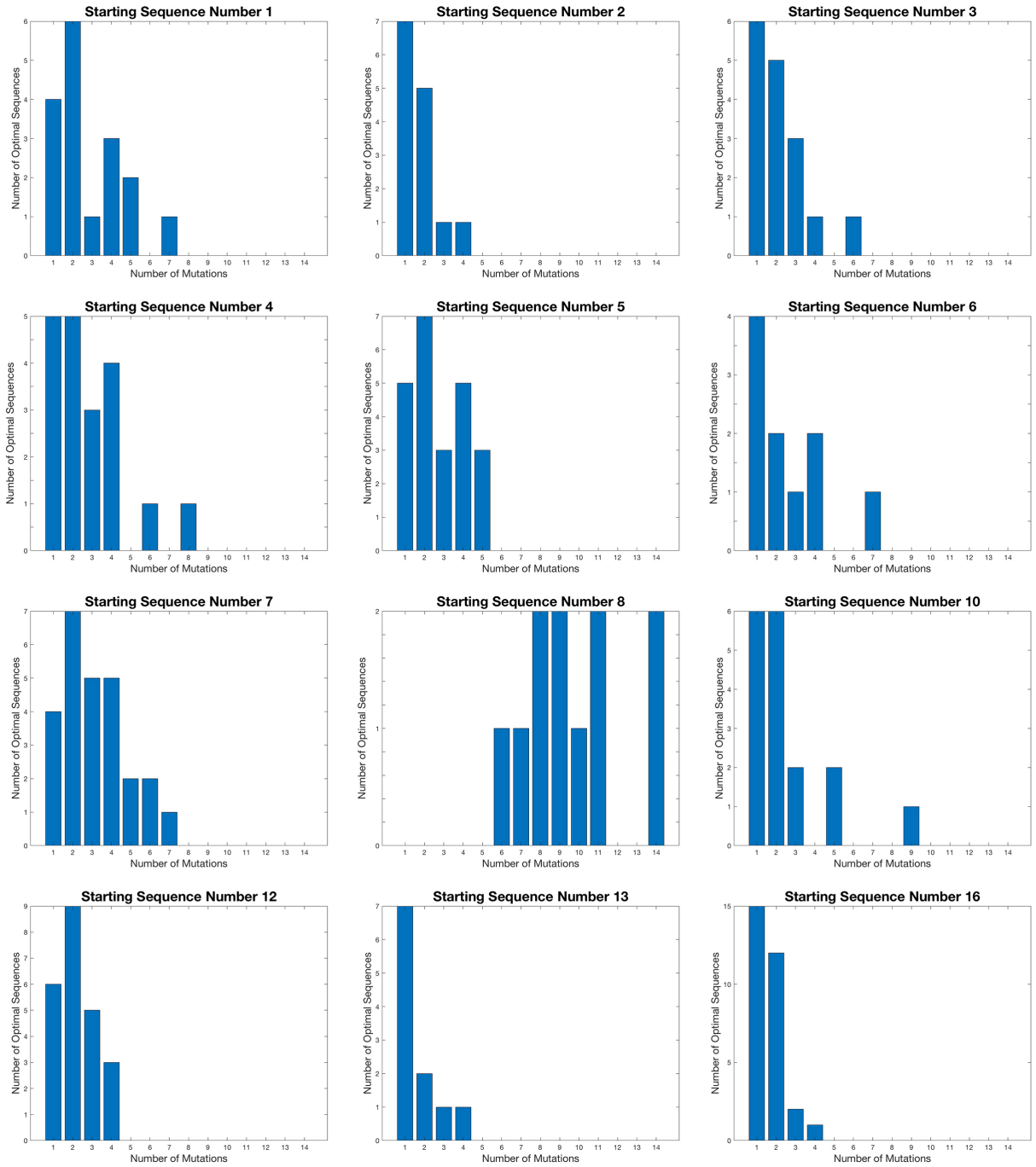


Figure S2: Distribution for the number of mutations for unique sequences generated by RAG-IF for 22 successful sequences (out of 27 starting sequences) for the 7_4 target topology (cont on next page).

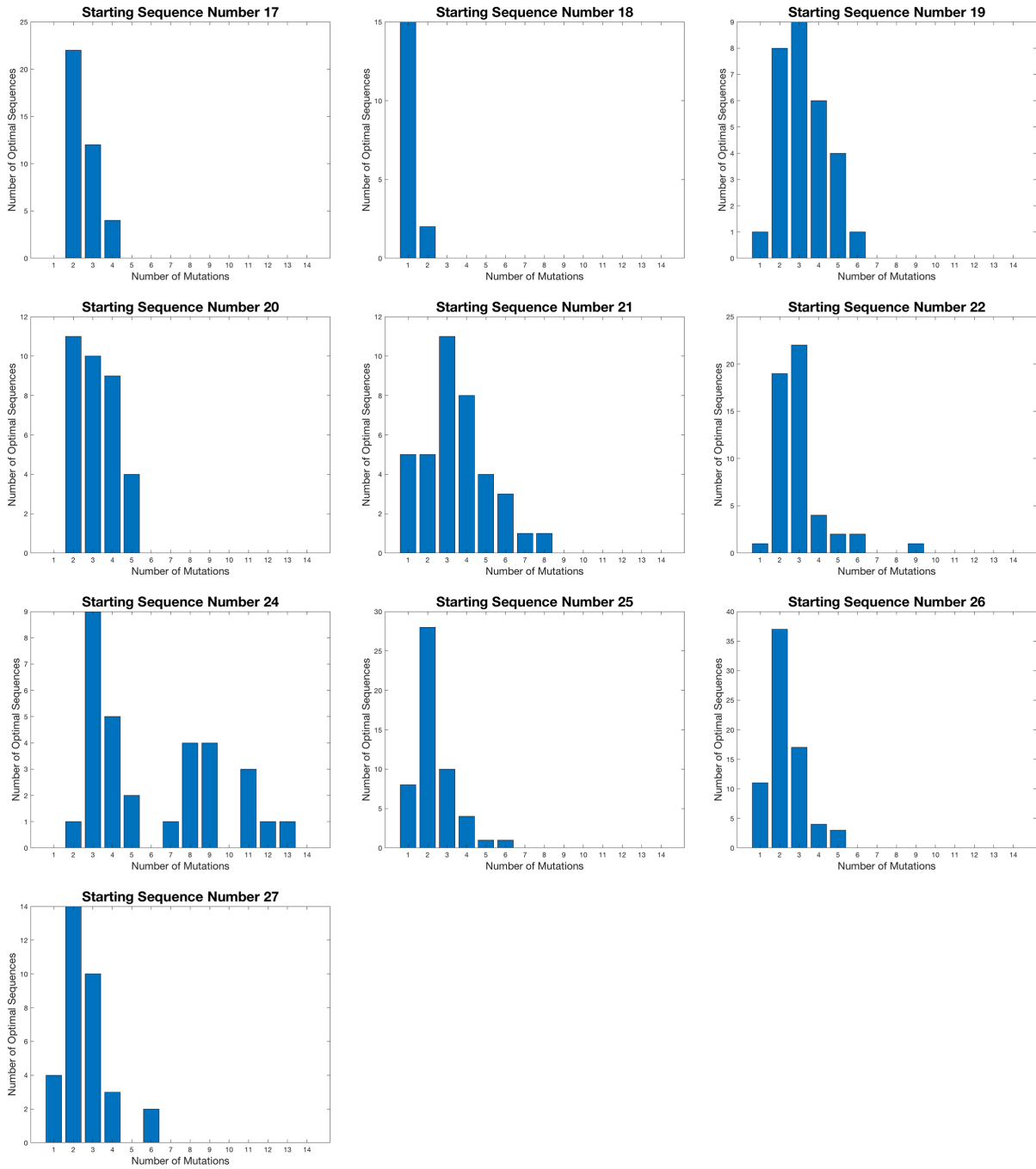


Figure S2: Distribution for the number of mutations for unique sequences generated by RAG-IF for 22 successful sequences (out of 27 starting sequences) for the 7_4 target topology (cont from previous page).

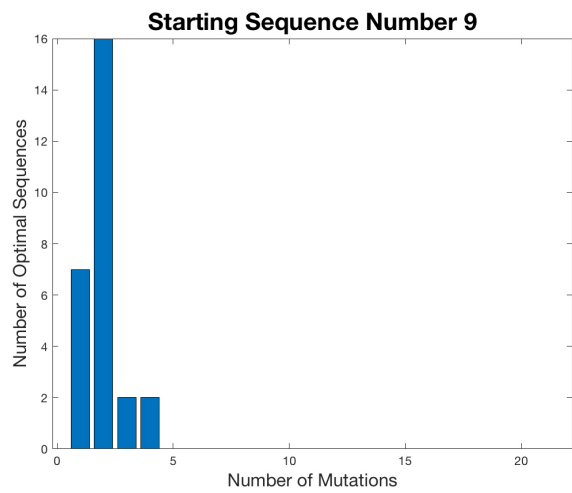
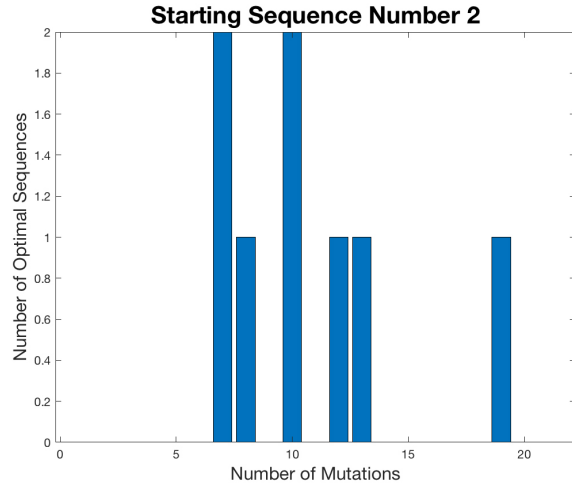
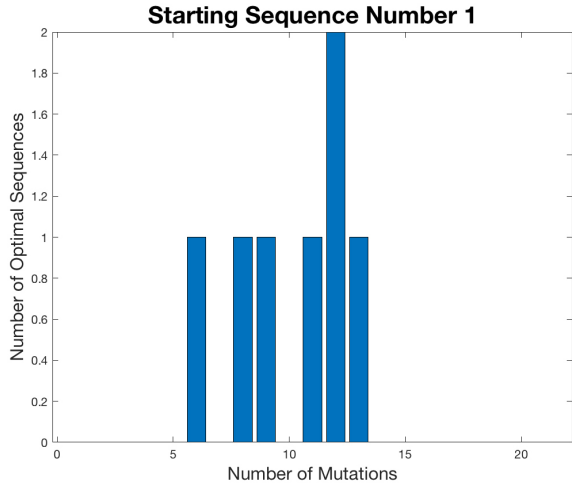


Figure S3: Distribution for the number of mutations for unique sequences generated by RAG-IF for 5 successful sequences (out of 9 starting sequences) for the 8_4 target topology.

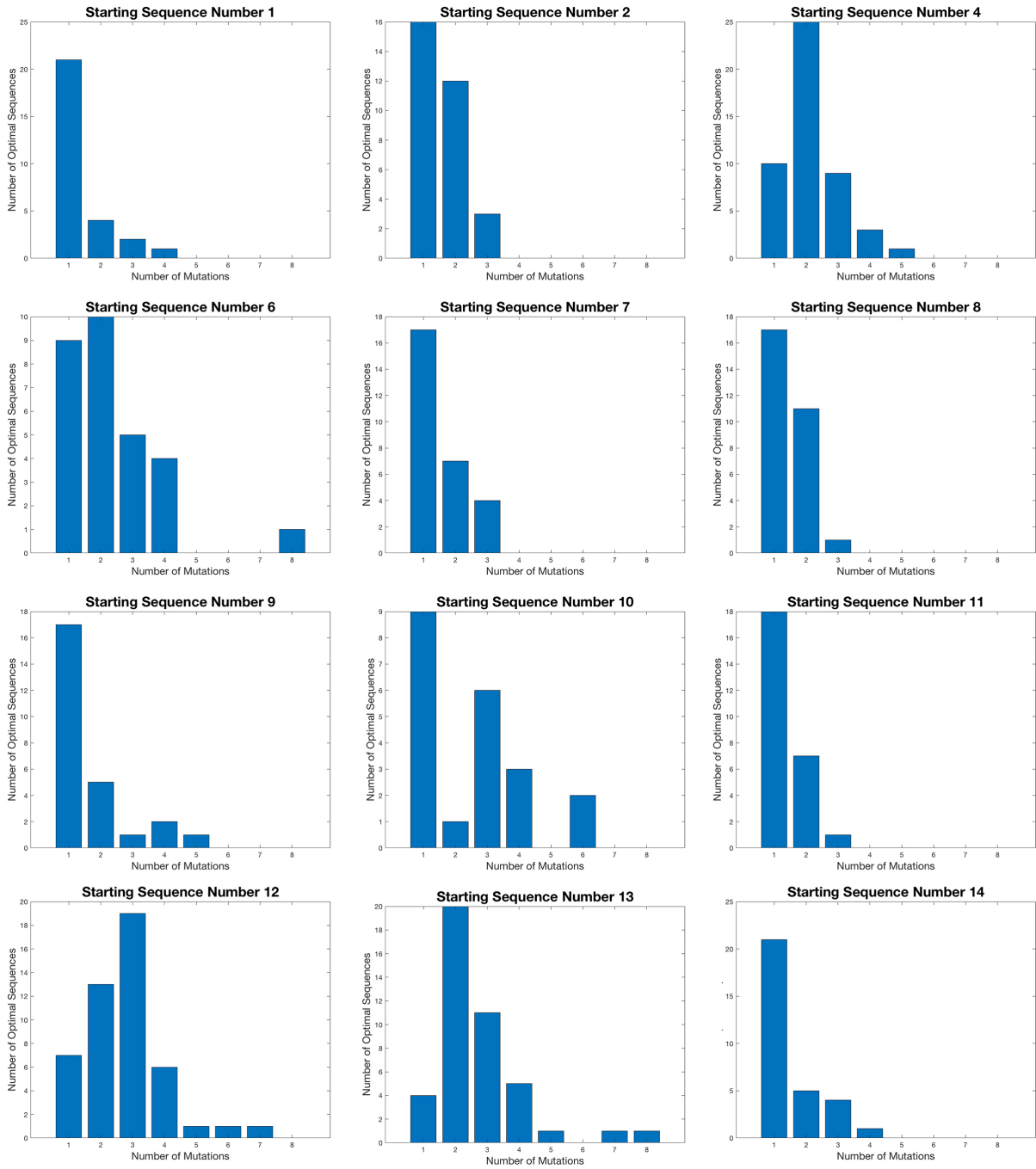


Figure S4: Distribution for the number of mutations for unique sequences generated by RAG-IF for 17 successful sequences (out of 20 starting sequences) for the 8_6 target topology (cont on next page).

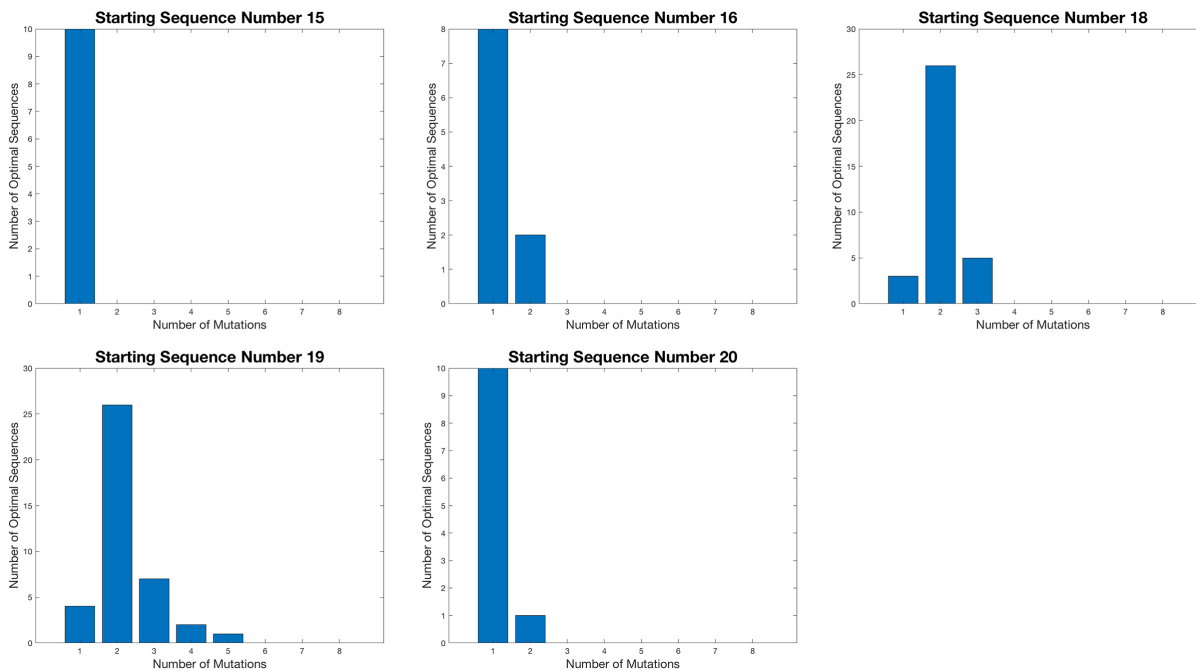


Figure S4: Distribution for the number of mutations for unique sequences generated by RAG-IF for 17 successful sequences (out of 20 starting sequences) for the 8_6 target topology (cont from previous page).

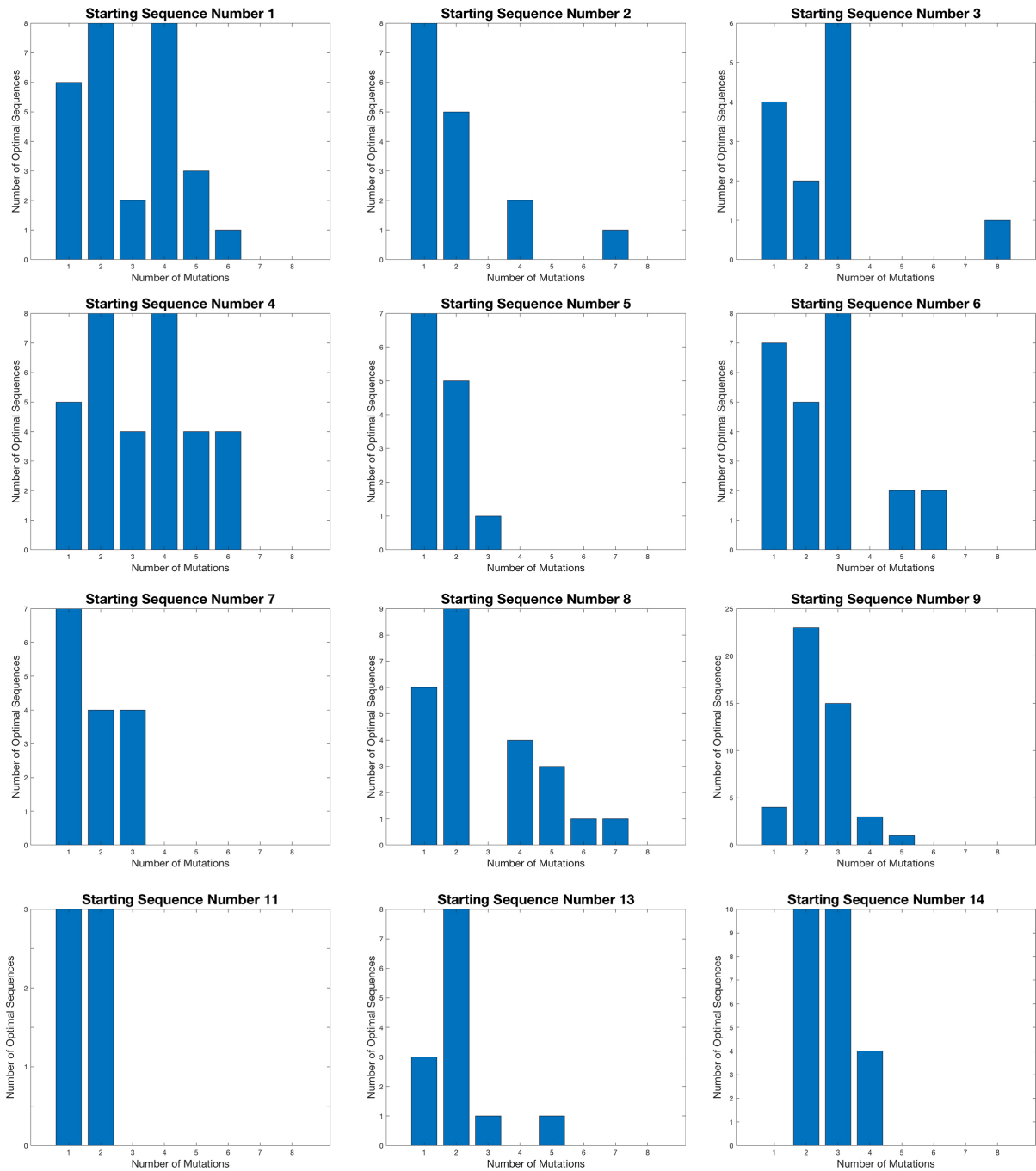


Figure S5: Distribution for the number of mutations for unique sequences generated by RAG-IF for 24 successful sequences (out of 26 starting sequences) for the 8.7 target topology (cont on next page).

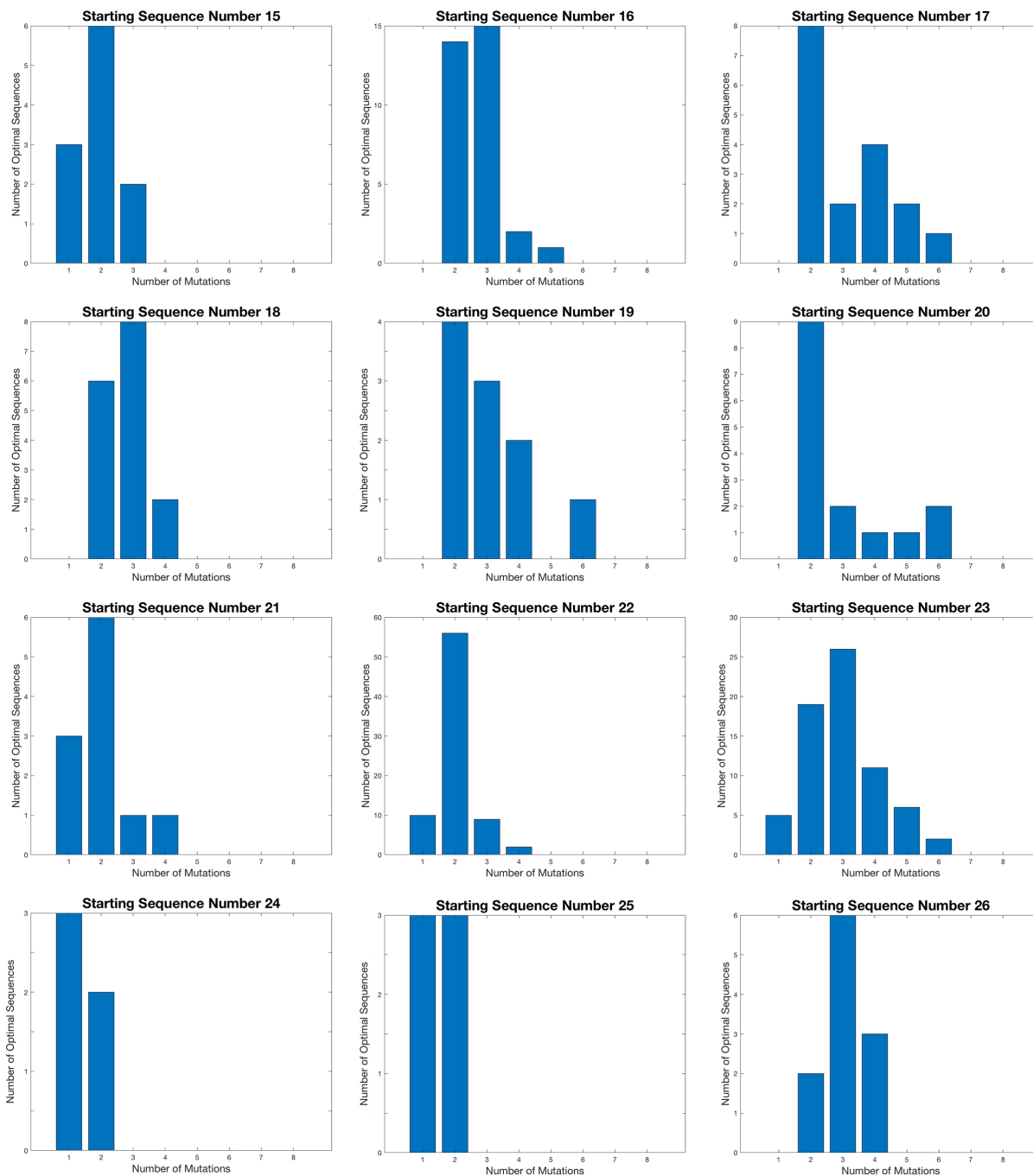


Figure S5: Distribution for the number of mutations for unique sequences generated by RAG-IF for 24 successful sequences (out of 26 starting sequences) for the 8.7 target topology (cont from previous page).

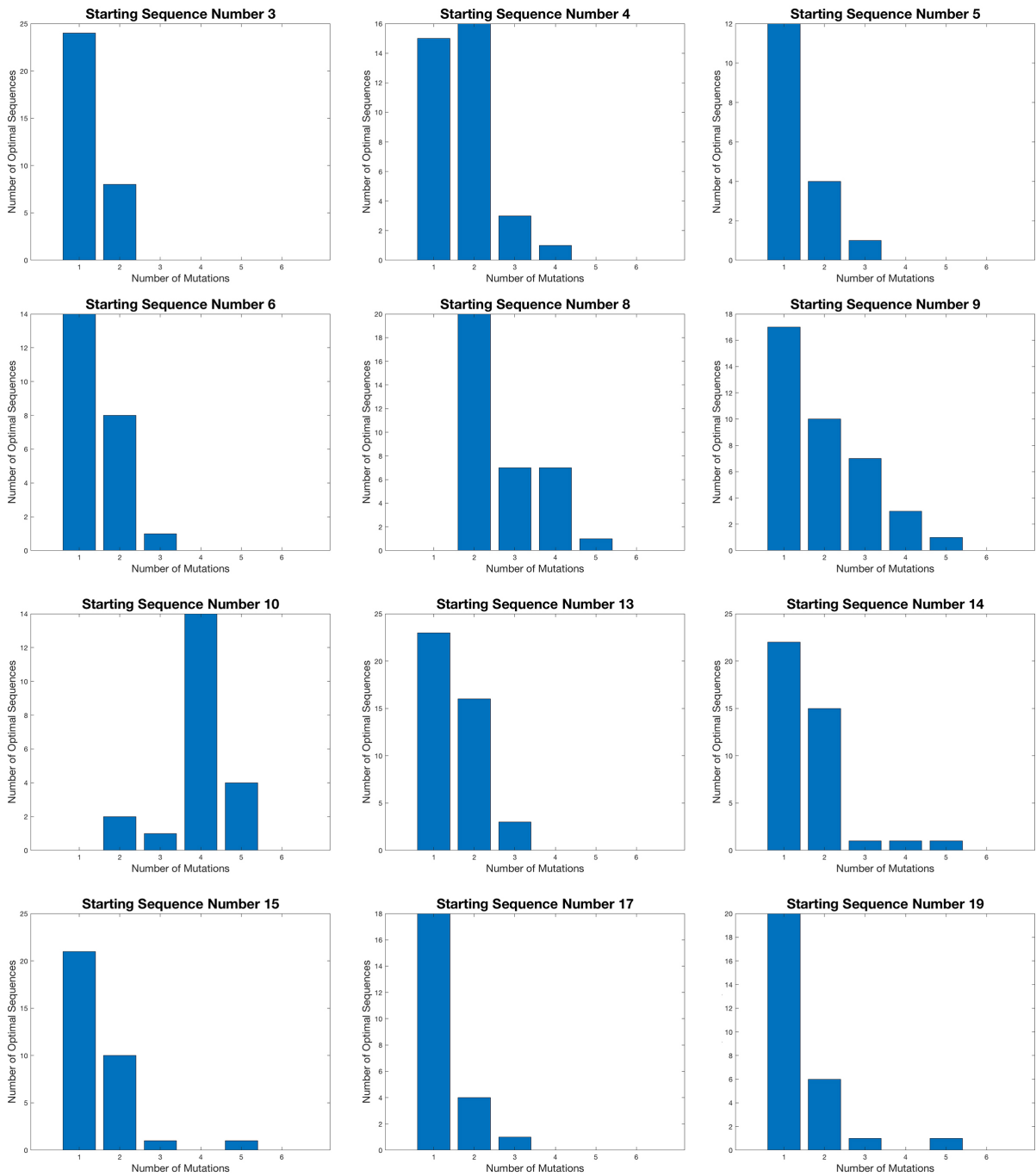


Figure S6: Distribution for the number of mutations for unique sequences generated by RAG-IF for 24 successful sequences (out of 34 starting sequences) for the 8_9 target topology (cont on next page).

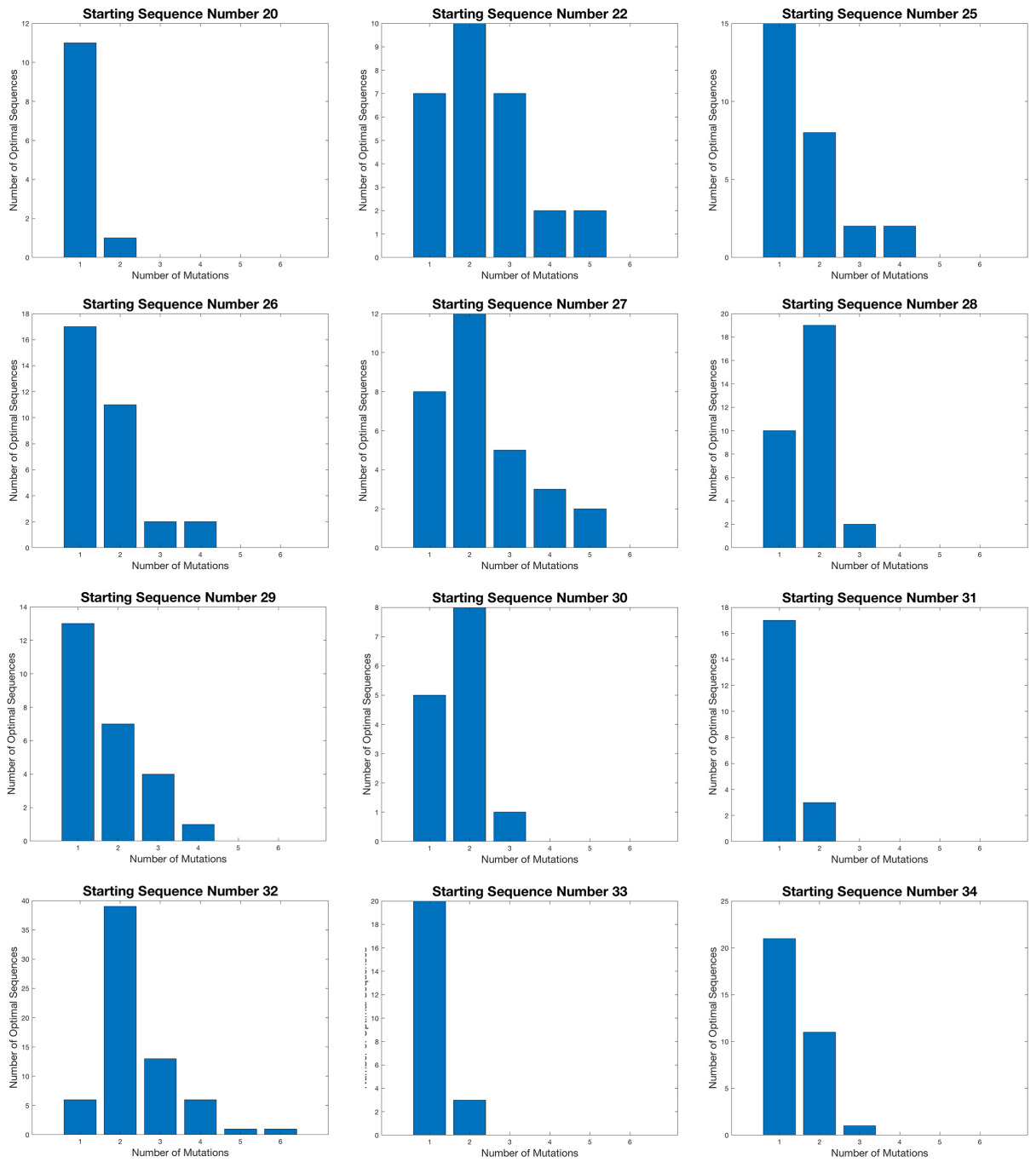


Figure S6: Distribution for the number of mutations for unique sequences generated by RAG-IF for 24 successful sequences (out of 34 starting sequences) for the 8_9 target topology (cont from previous page).

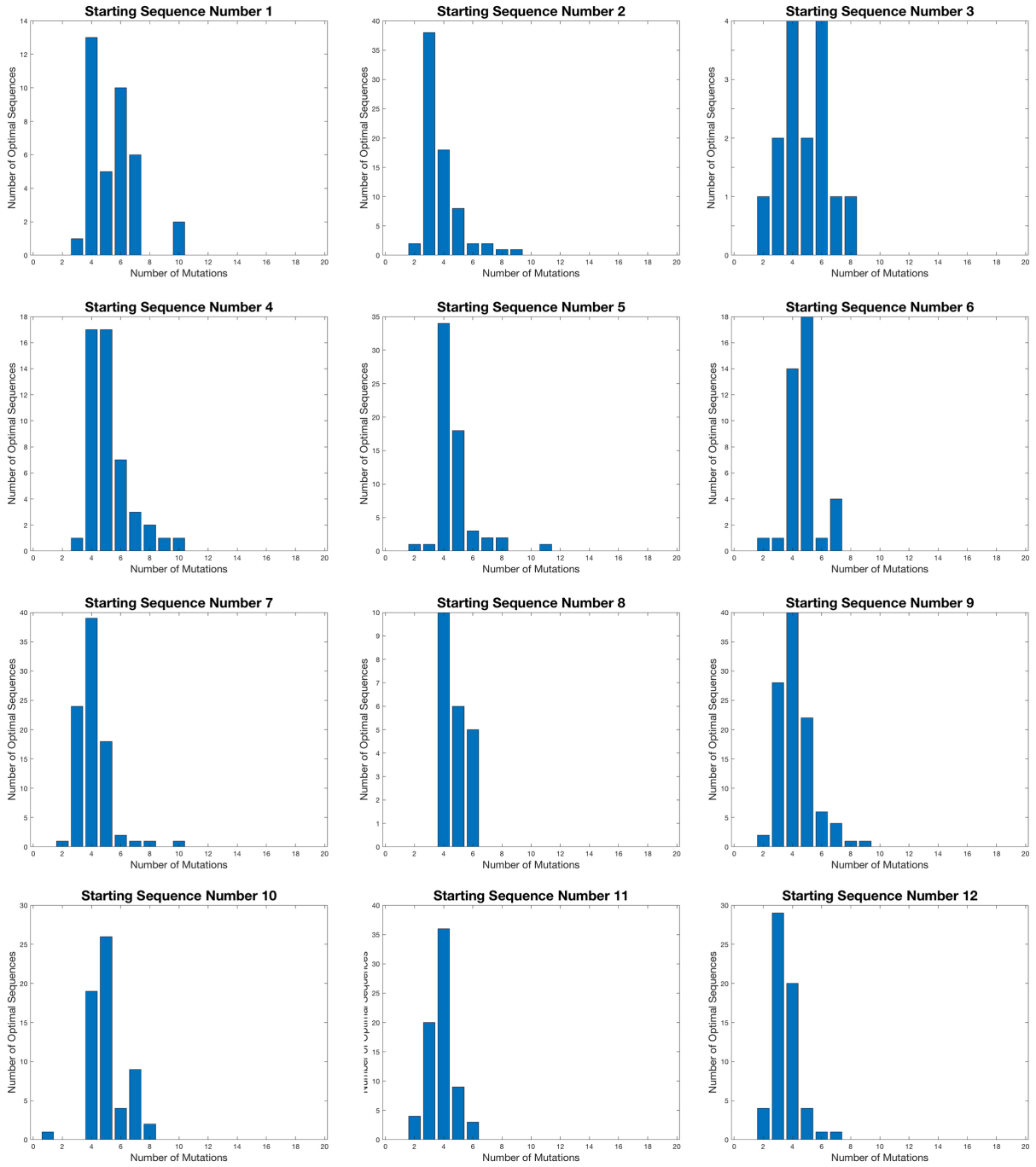


Figure S7: Distribution for the number of mutations for unique sequences generated by RAG-IF for 68 successful sequences (out of 71 starting sequences) for the 8_12 target topology (cont on next page).

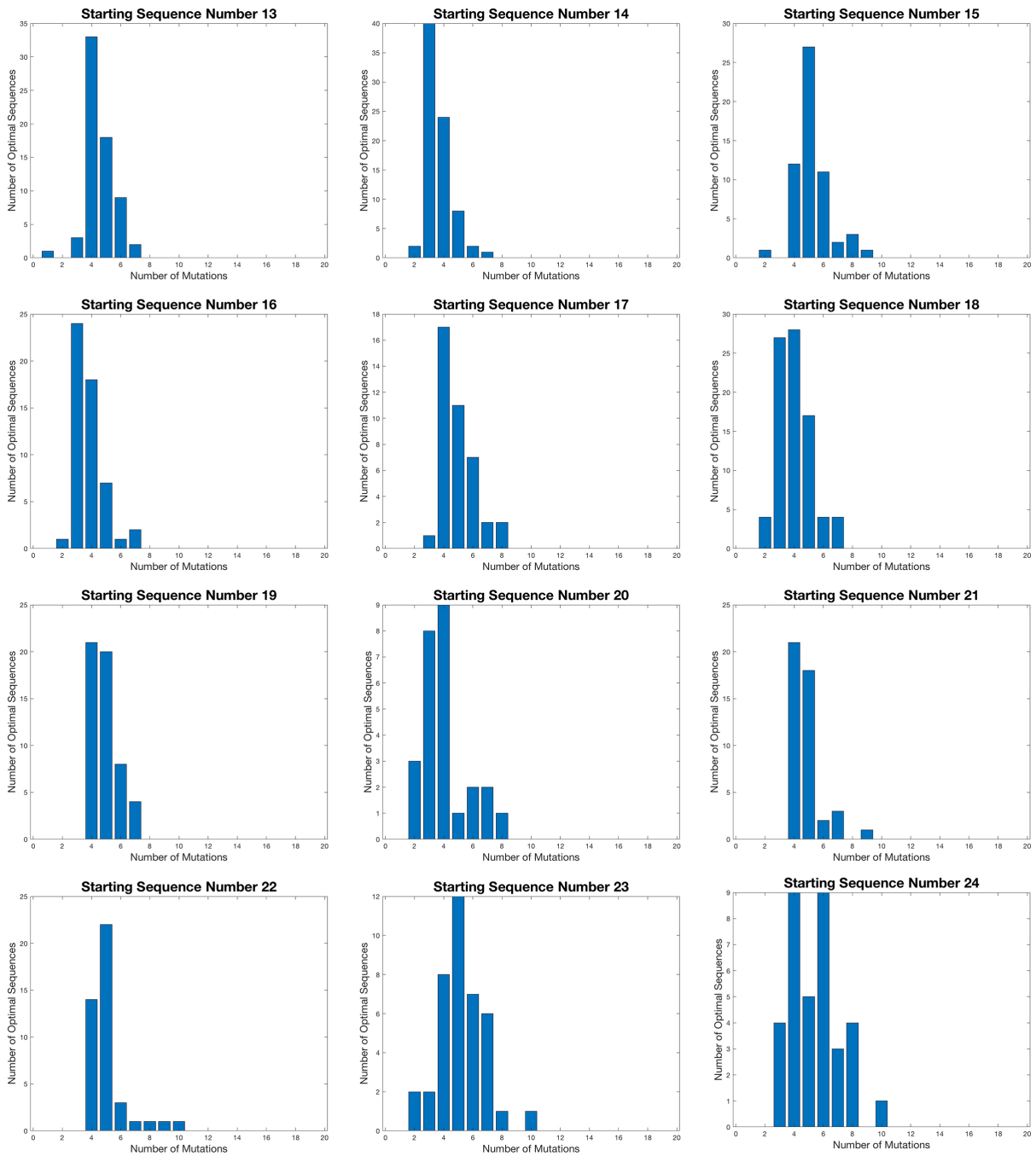


Figure S7: Distribution for the number of mutations for unique sequences generated by RAG-IF for 68 successful sequences (out of 71 starting sequences) for the 8_12 target topology (cont on next page).

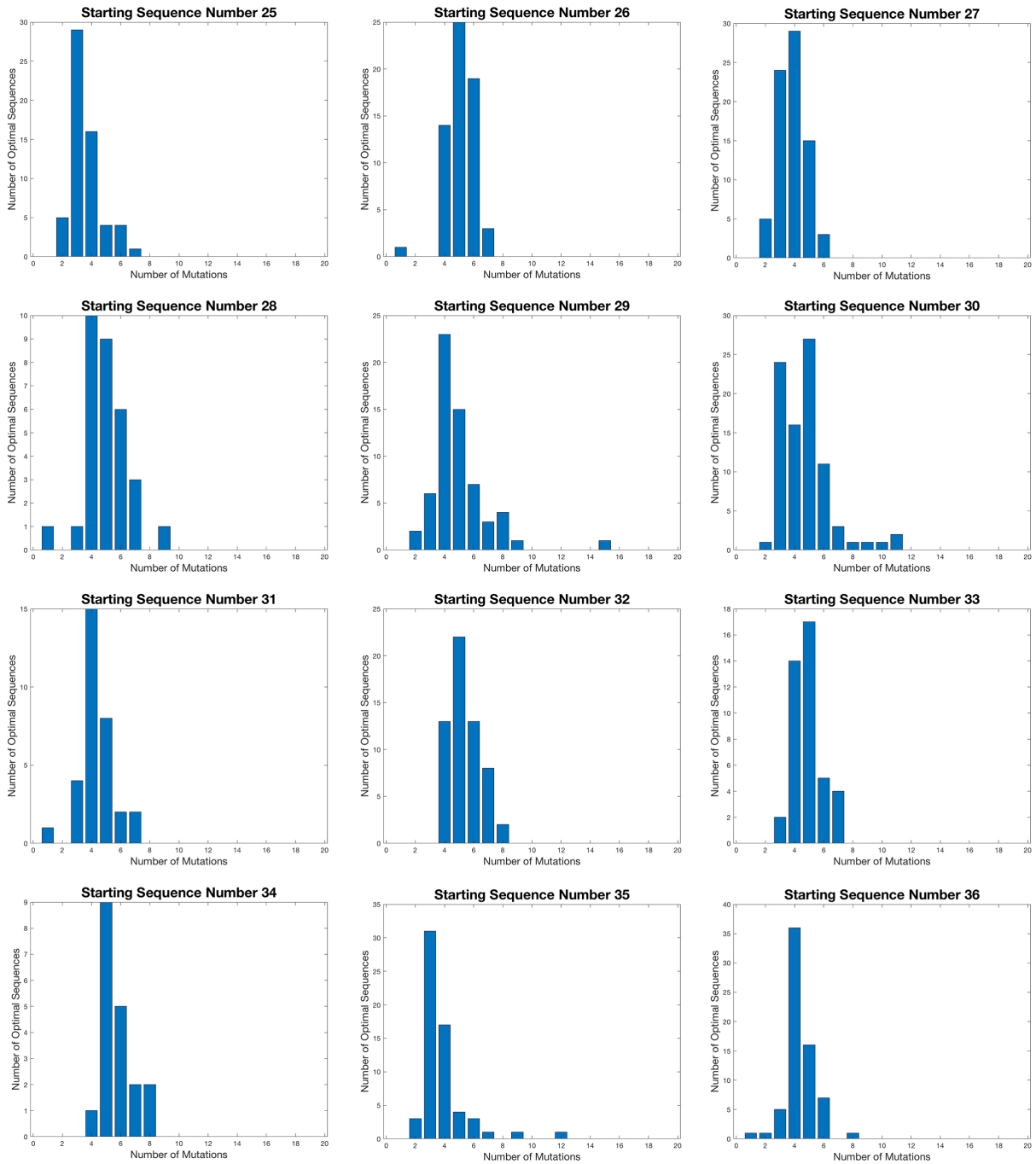


Figure S7: Distribution for the number of mutations for unique sequences generated by RAG-IF for 68 successful sequences (out of 71 starting sequences) for the 8_12 target topology (cont on next page).

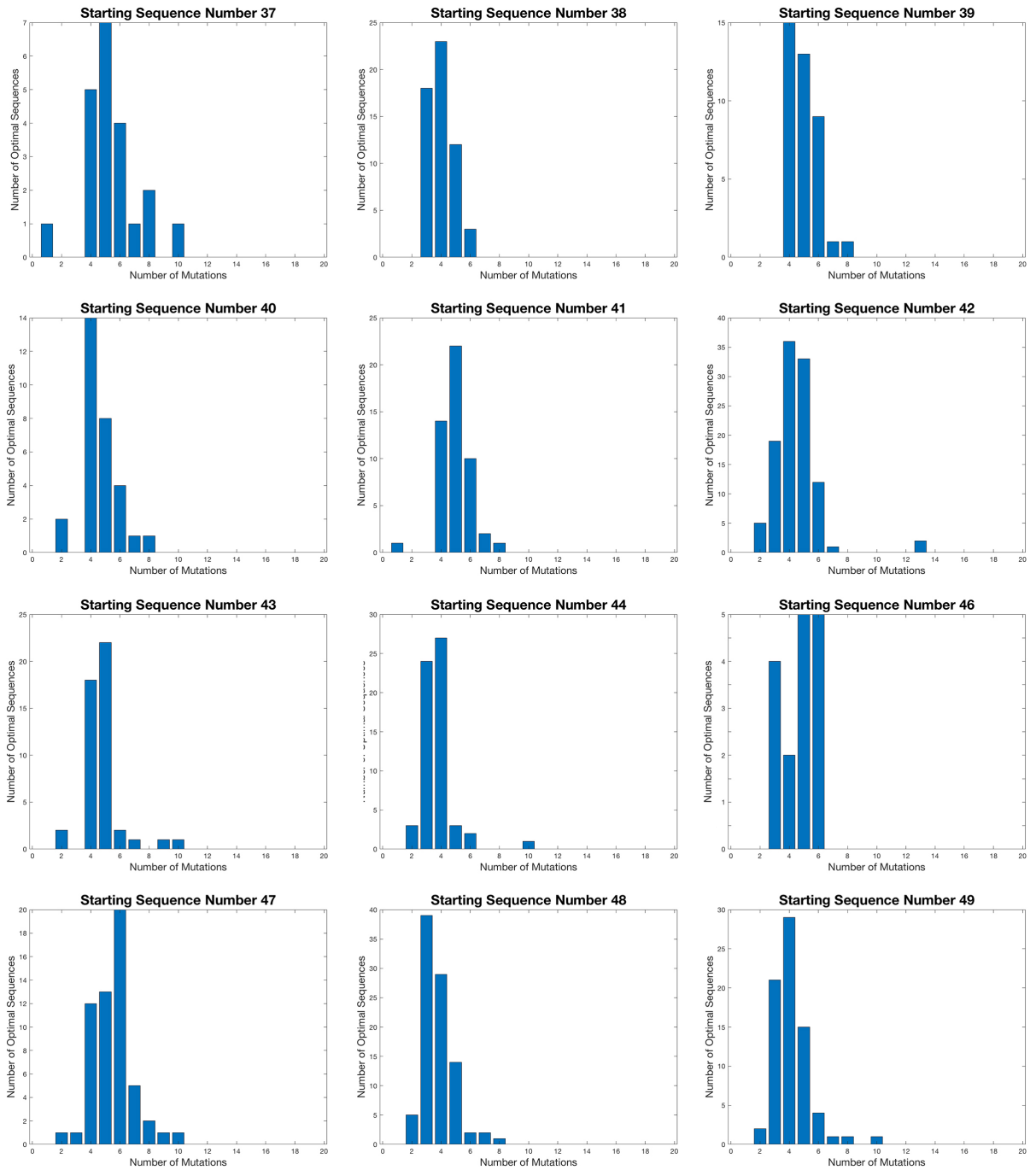


Figure S7: Distribution for the number of mutations for unique sequences generated by RAG-IF for 68 successful sequences (out of 71 starting sequences) for the 8_12 target topology (cont on next page).

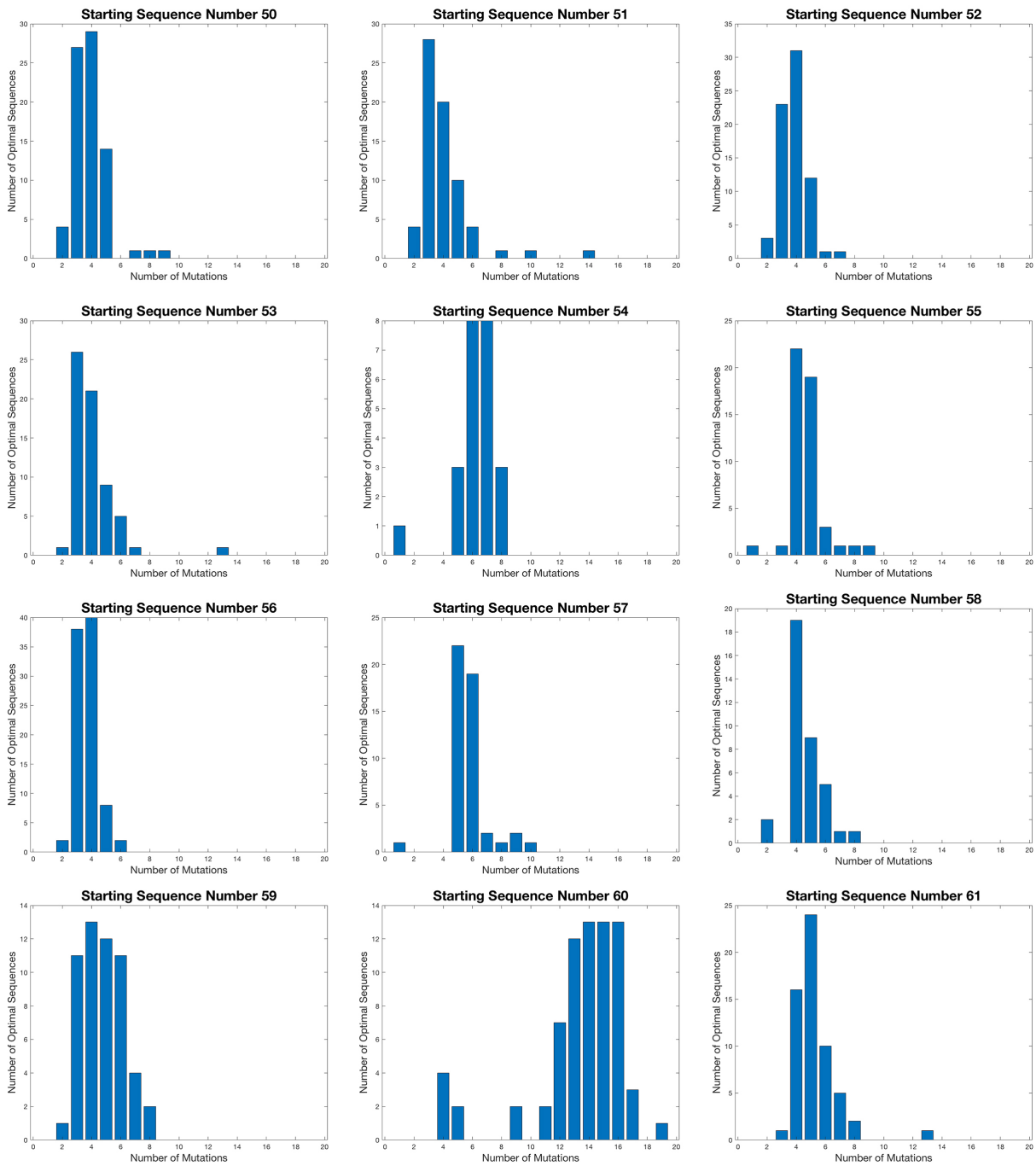


Figure S7: Distribution for the number of mutations for unique sequences generated by RAG-IF for 68 successful sequences (out of 71 starting sequences) for the 8_12 target topology (cont on next page).

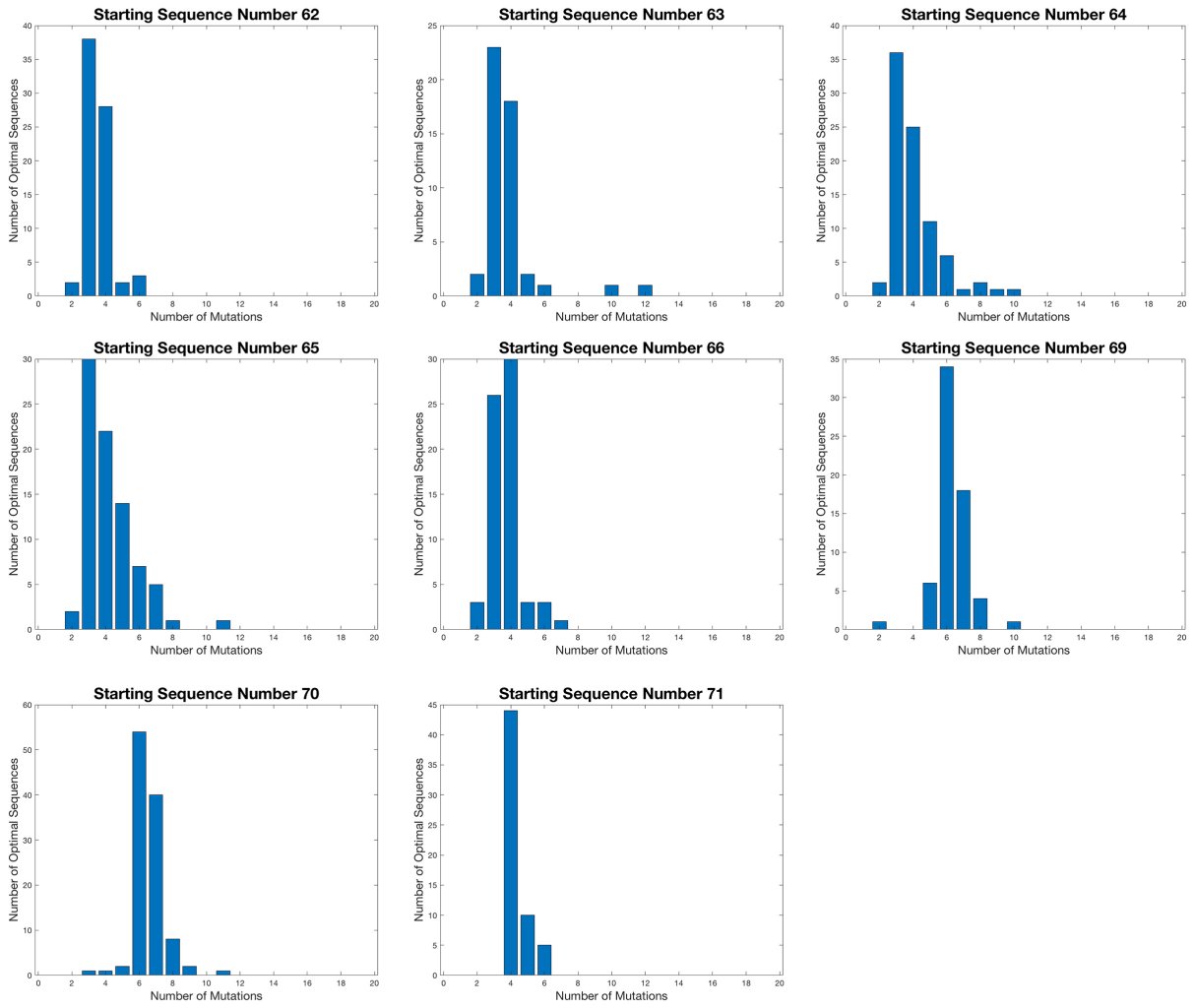


Figure S7: Distribution for the number of mutations for unique sequences generated by RAG-IF for 68 successful sequences (out of 71 starting sequences) for the 8_12 target topology (cont from previous page).

Time Comparison between RNAInverse and GA

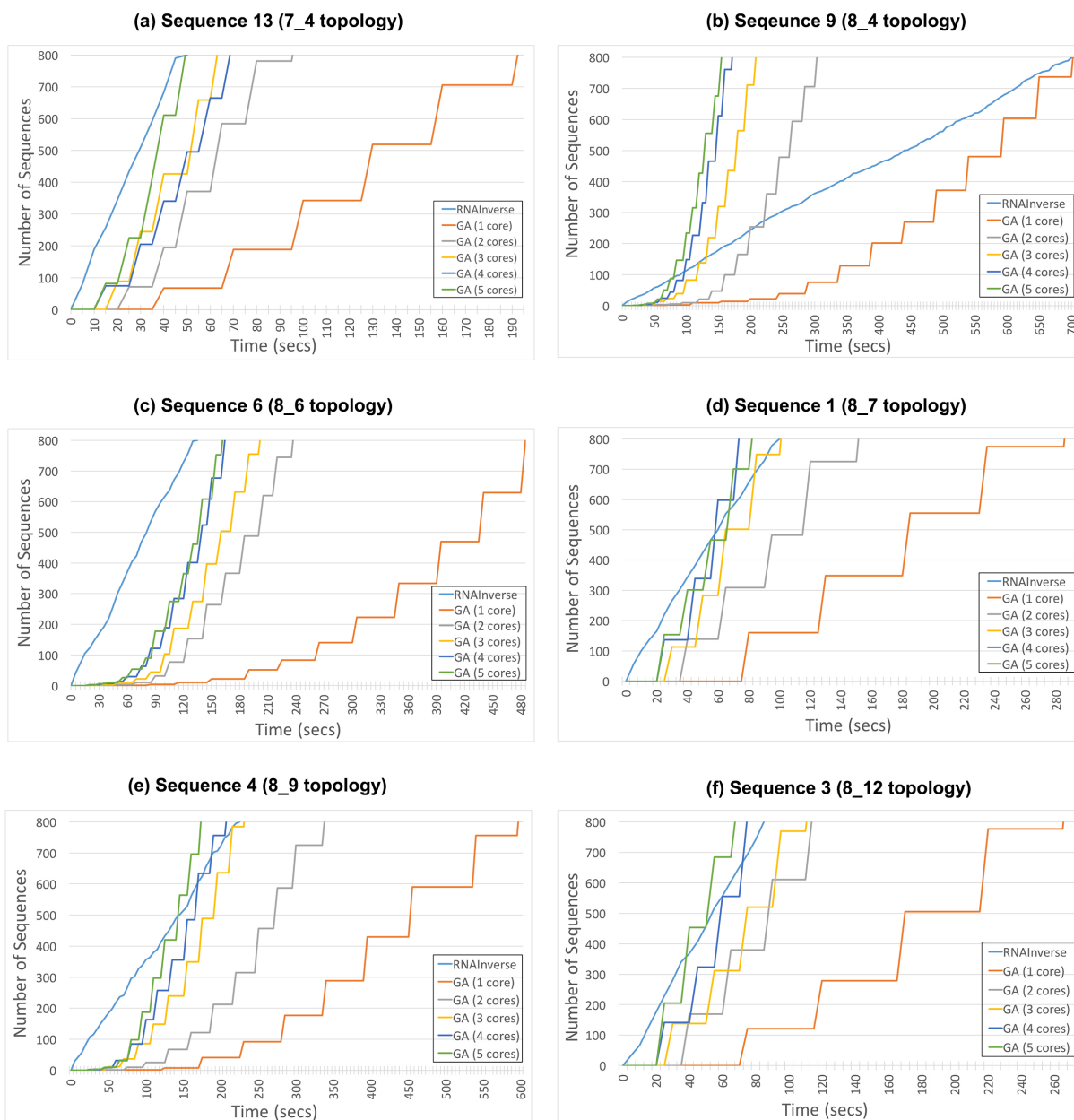


Figure S8: Run time comparison between RAG-IF GA algorithm and RNAInverse for the sample sequence for 6 target topologies shown in Figures 4-9 in the main article. For sequence 3 for the 8_12 topology, the intermediate sequence shown in Figure 9 of the main document is used. Both RNAInverse and GA were run until they generated at least 800 sequences. Note that the generated sequences by GA and RNAInverse were not screened further to check whether they fold onto the target topology.

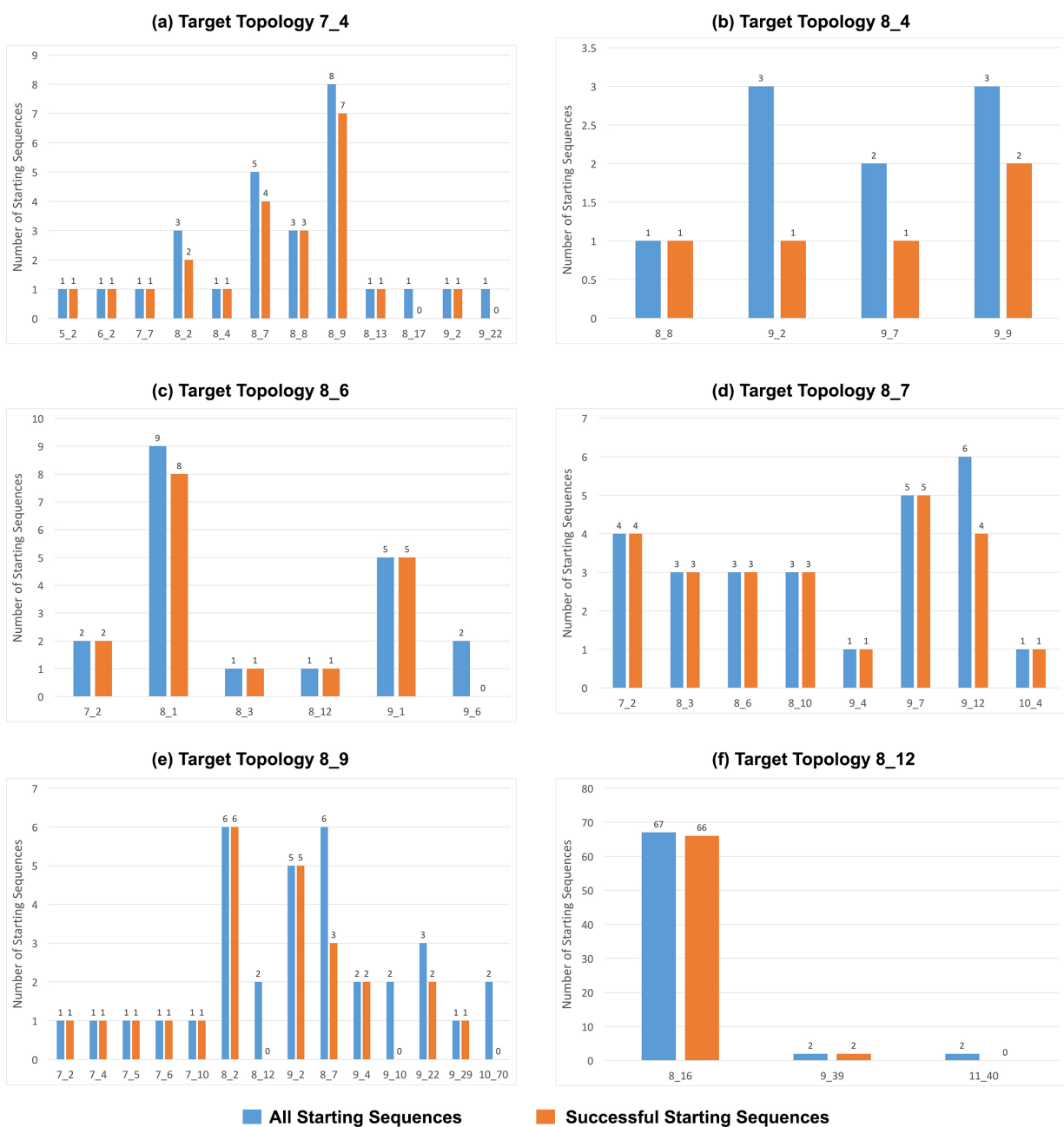


Figure S9: The NUPACK predicted topologies for the starting sequences for 6 target topologies. Number of all starting sequences are shown in blue and number of successful starting sequences (for which RAG-IF generates at least one successful sequence) are shown in orange.