Supplement to Accompany:

Accelerated RNA Secondary Structure Design Using Pre-Selected Sequences for Helices and Loops

Stanislav Bellaousov,[1] Mohammad Kayedkhordeh,[1] Raymond J. Peterson,[2] and David H. Mathews[1,3,*]

[1]Department of Biochemistry & Biophysics and Center for RNA Biology, University of Rochester Medical Center, 601 Elmwood Ave, Box 712, Rochester, New York 14642, United States

[2]Celadon Laboratories, Inc., Suite 521, 6525 Belcrest Road, Hyattsville, MD 20782, United States

[3]Department of Biostatistics and Computational Biology, University of Rochester Medical Center, 601 Elmwood Ave, Box 712, Rochester, New York 14642, United States

*Address correspondence to David_Mathews@urmc.rochester.edu

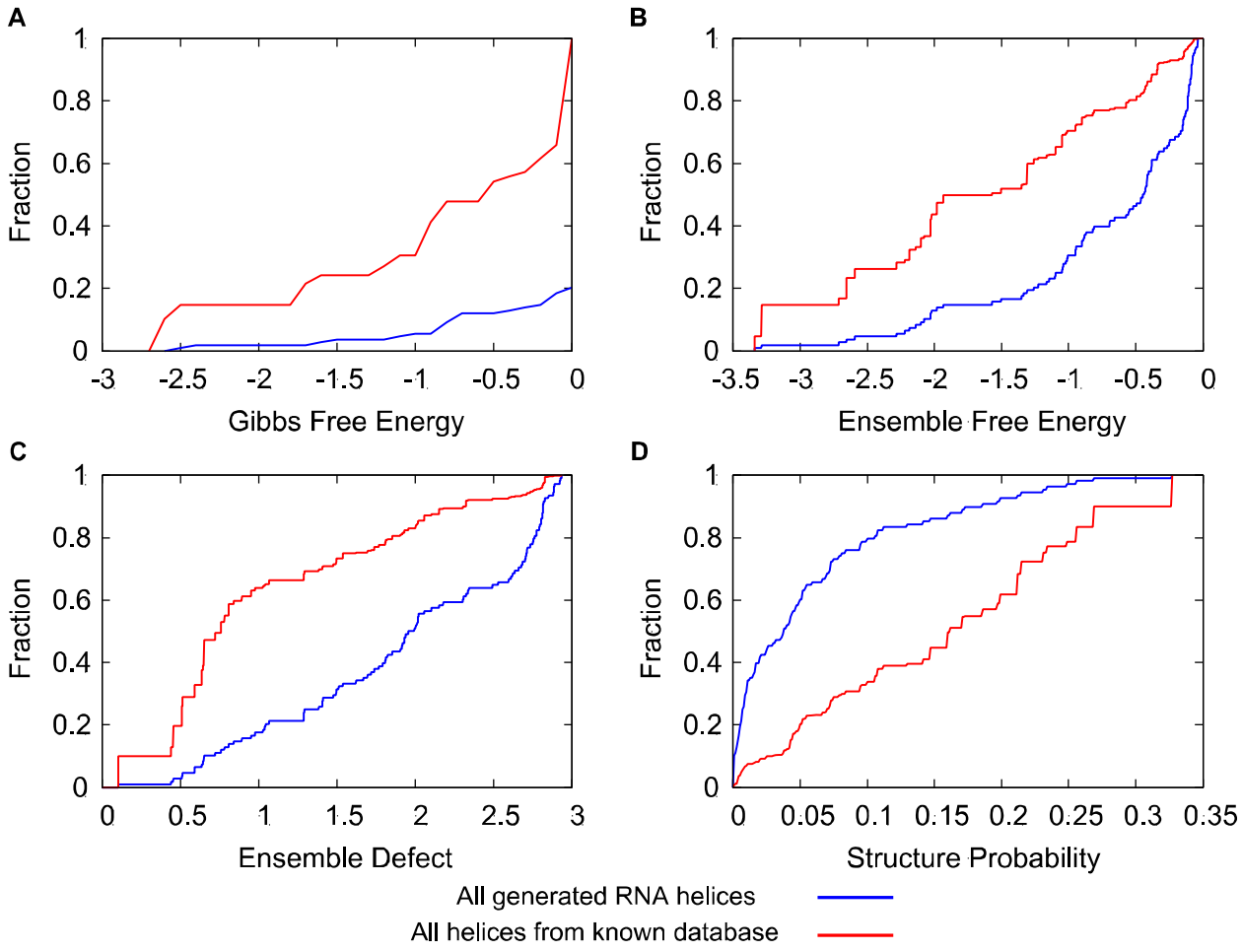

Figure S1. Trends in natural sequences for helices of length 3. Distributions of folding free energy change, ensemble folding free energy change, ensemble defect, and structure probability for three base pair helices. Cumulative distribution plots are provided in red for unique sequences observed in the database of RNA structures and in blue for all possible helices. Panel A is Gibbs free energy change, panel B is ensemble Gibbs free energy change, panel C is ensemble defect, and panel D is the probability of helix formation.

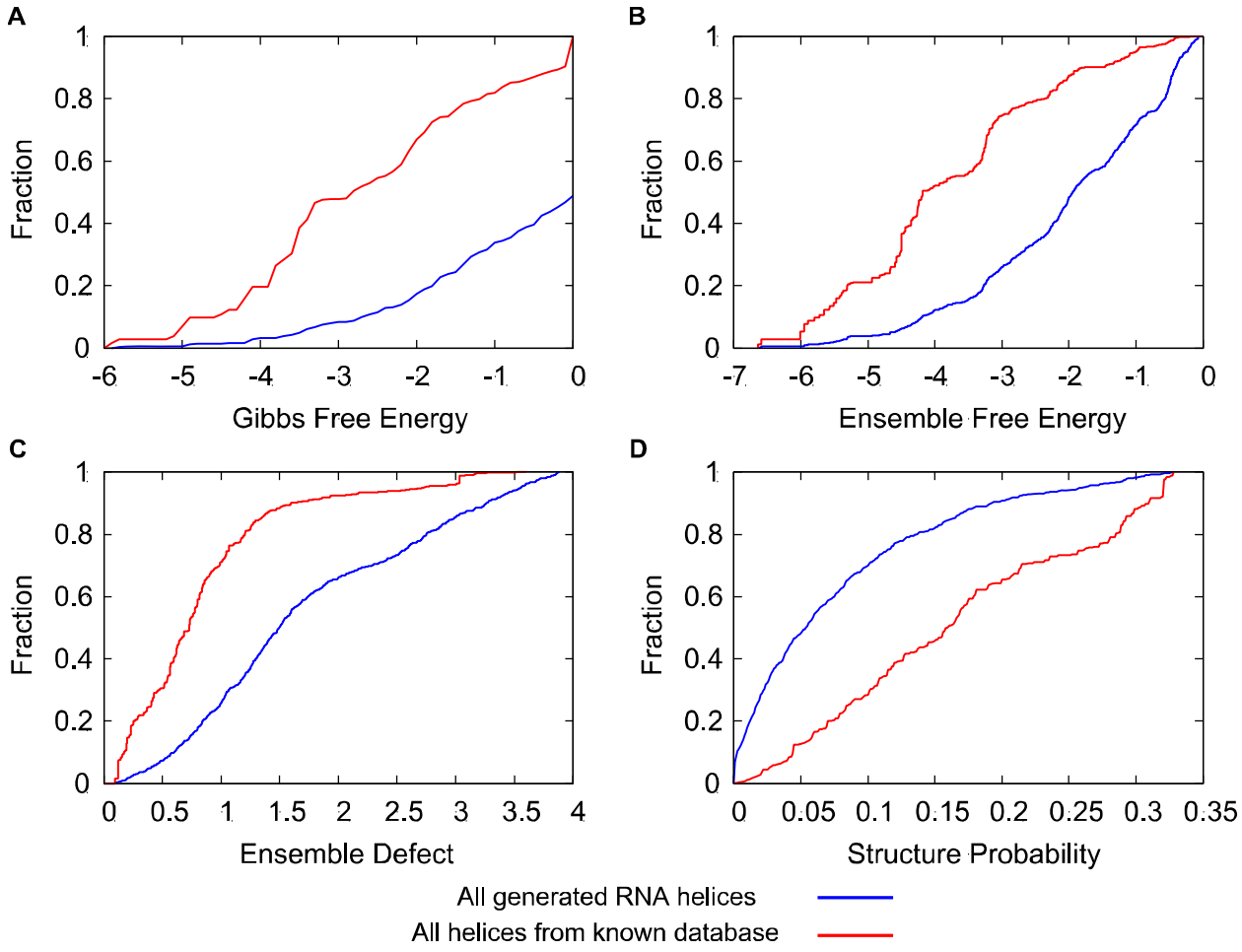

Figure S2. Trends in natural sequences for helices of length four. Distributions of folding free energy change, ensemble folding free energy change, ensemble defect, and structure probability for four base pair helices. Cumulative distribution plots are provided in red for unique sequences observed in the database of RNA structures and in blue for all possible helices. Panel A is Gibbs free energy change, panel B is ensemble Gibbs free energy change, panel C is ensemble defect, and panel D is the probability of helix formation.

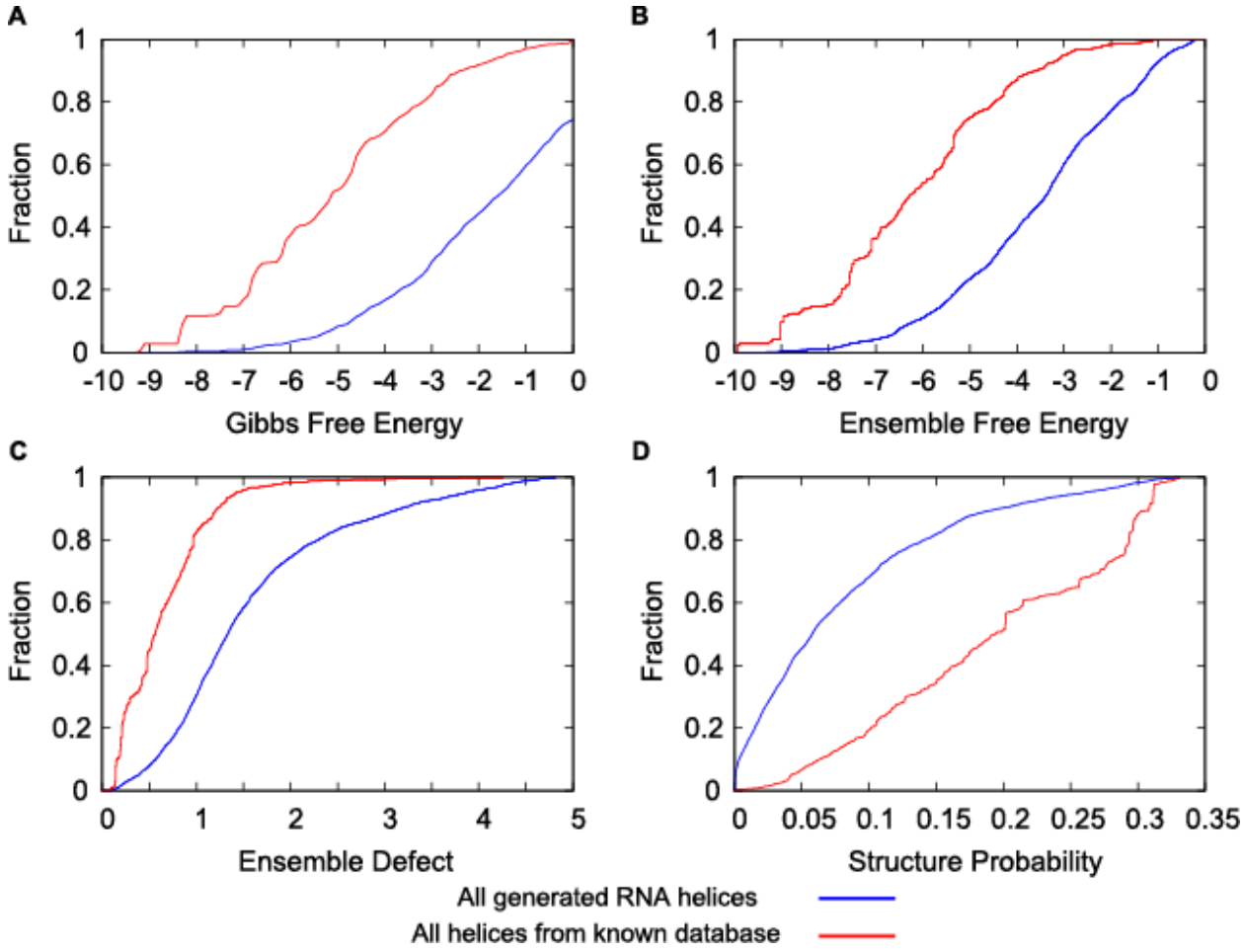

Figure S3. Trends in natural sequences for helices of length five. Distributions of folding free energy change, ensemble folding free energy change, ensemble defect, and structure probability for five base pair helices. Cumulative distribution plots are provided in red for unique sequences observed in the database of RNA structures and in blue for all possible helices. Panel A is Gibbs free energy change, panel B is ensemble Gibbs free energy change, panel C is ensemble defect, and panel D is the probability of helix formation.

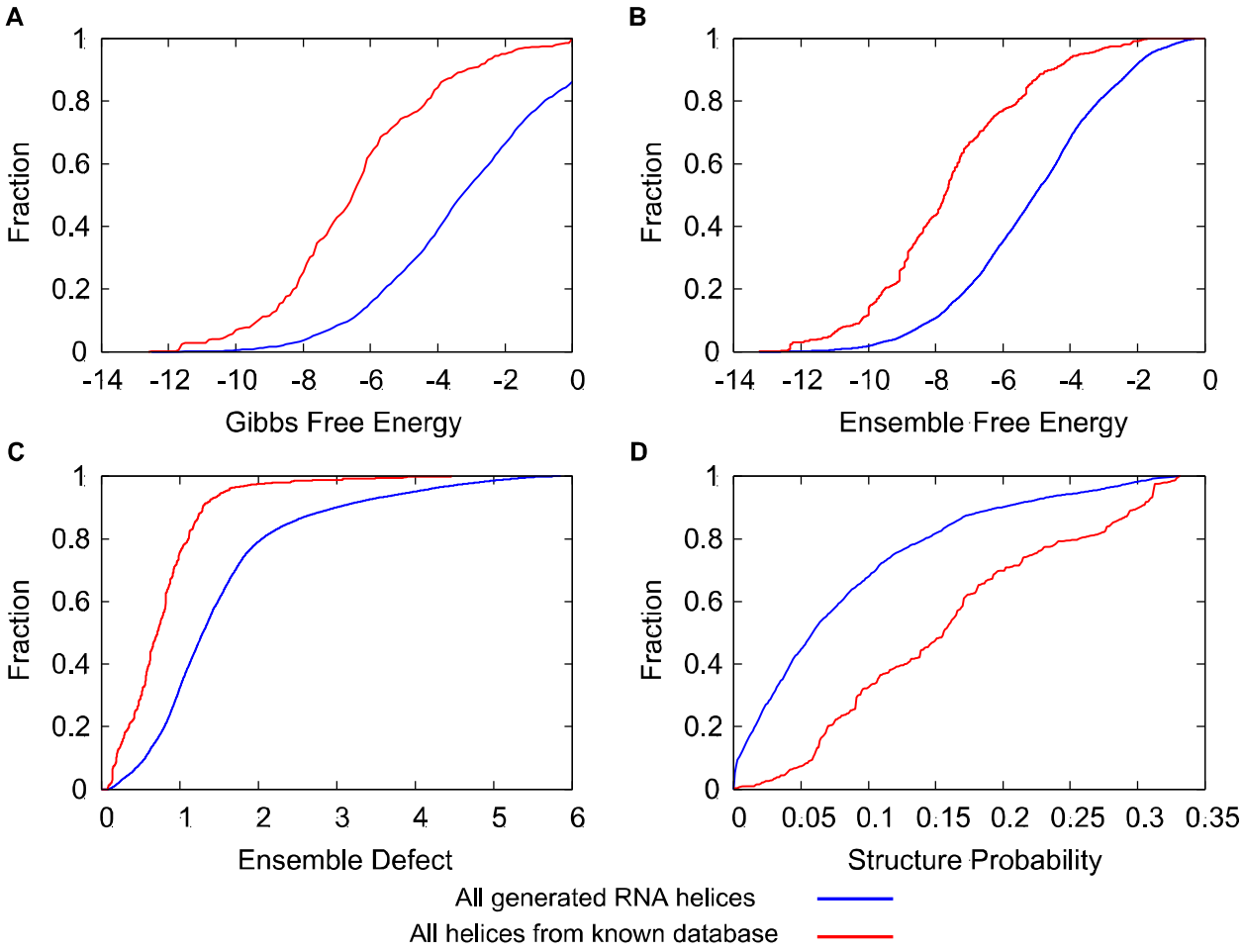

Figure S4. Trends in natural sequences for helices of length six. Distributions of folding free energy change, ensemble folding free energy change, ensemble defect, and structure probability for six base pair helices. Cumulative distribution plots are provided in red for unique sequences observed in the database of RNA structures and in blue for all possible helices. Panel A is Gibbs free energy change, panel B is ensemble Gibbs free energy change, panel C is ensemble defect, and panel D is the probability of helix formation.

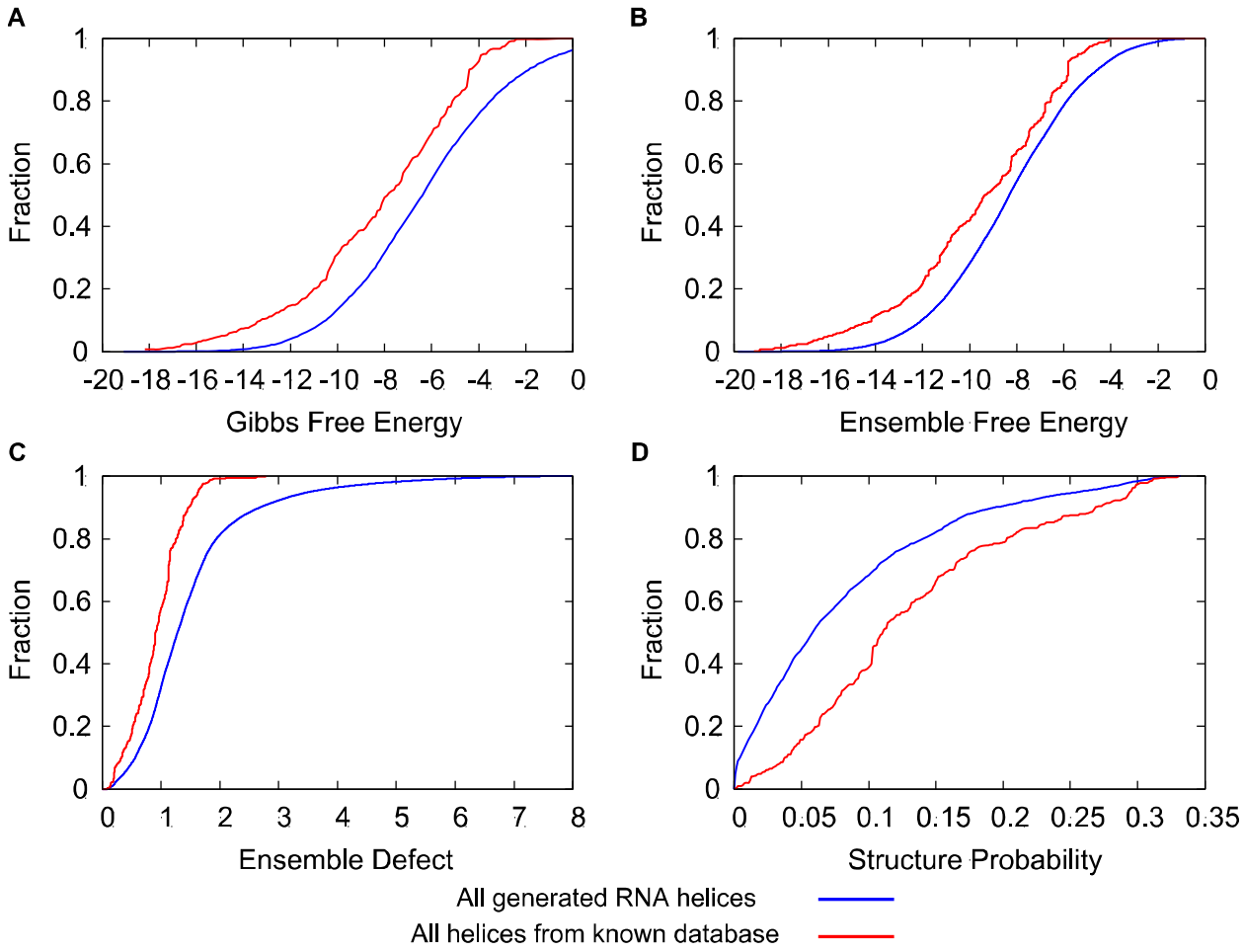

Figure S5. Trends in natural sequences for helices of length eight. Distributions of folding free energy change, ensemble folding free energy change, ensemble defect, and structure probability for eight base pair helices. Cumulative distribution plots are provided in red for unique sequences observed in the database of RNA structures and in blue for all possible helices. Panel A is Gibbs free energy change, panel B is ensemble Gibbs free energy change, panel C is ensemble defect, and panel D is the probability of helix formation.

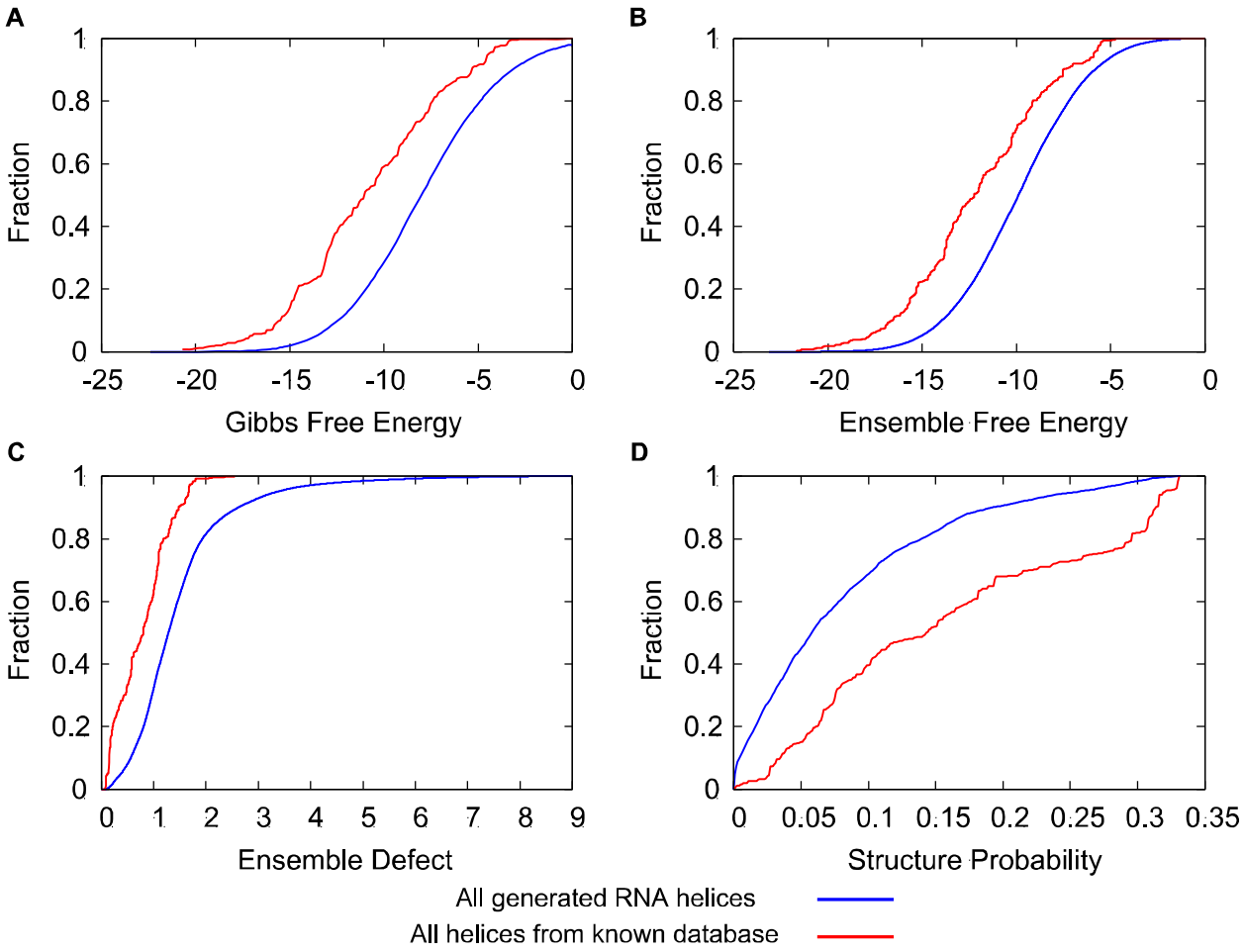

Figure S6. Trends in natural sequences for helices of length nine. Distributions of folding free energy change, ensemble folding free energy change, ensemble defect, and structure probability for nine base pair helices. Cumulative distribution plots are provided in red for unique sequences observed in the database of RNA structures and in blue for all possible helices. Panel A is Gibbs free energy change, panel B is ensemble Gibbs free energy change, panel C is ensemble defect, and panel D is the probability of helix formation.

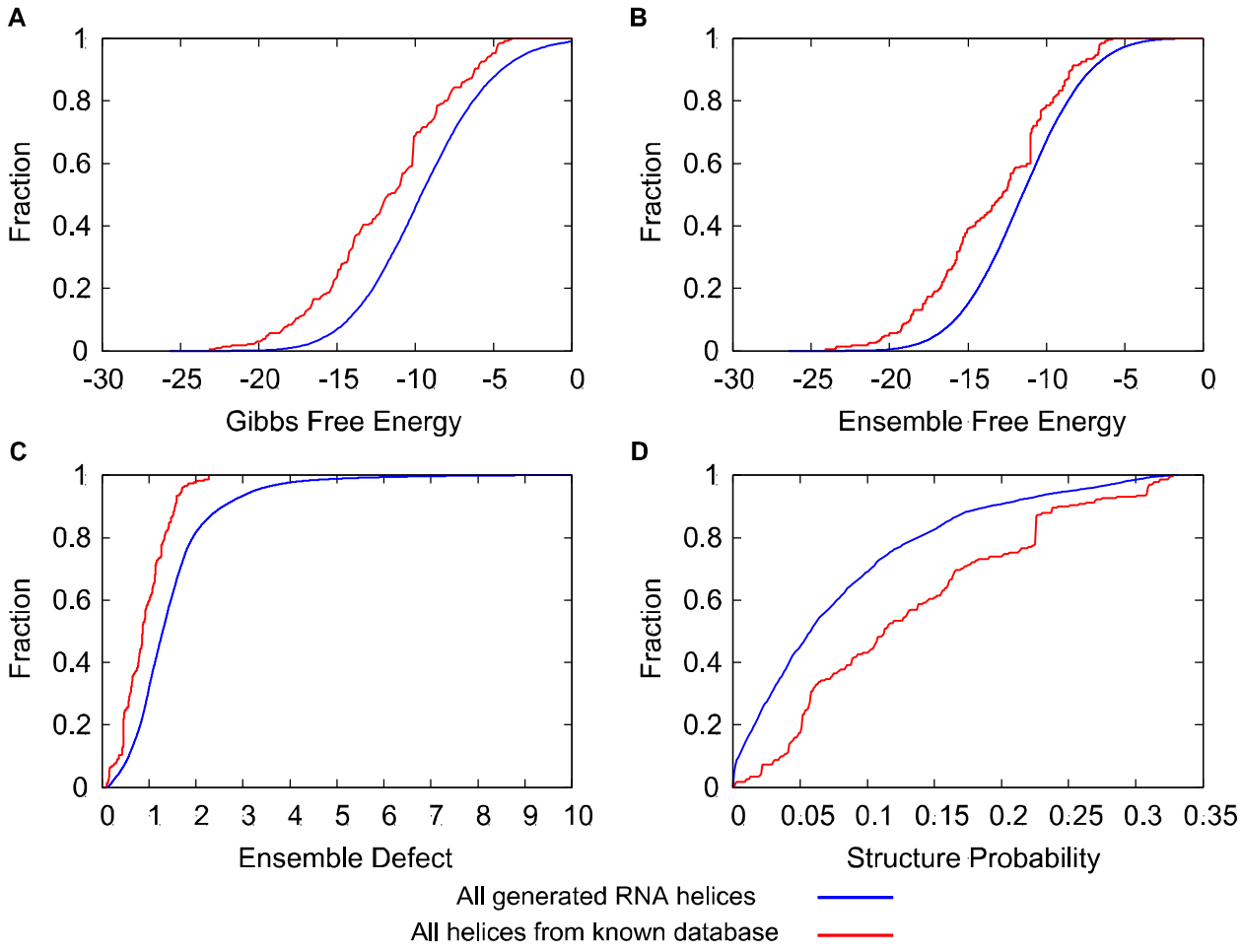

Figure S7. Trends in natural sequences for helices of length ten. Distributions of folding free energy change, ensemble folding free energy change, ensemble defect, and structure probability for ten base pair helices. Cumulative distribution plots are provided in red for unique sequences observed in the database of RNA structures and in blue for all possible helices. Panel A is Gibbs free energy change, panel B is ensemble Gibbs free energy change, panel C is ensemble defect, and panel D is the probability of helix formation.

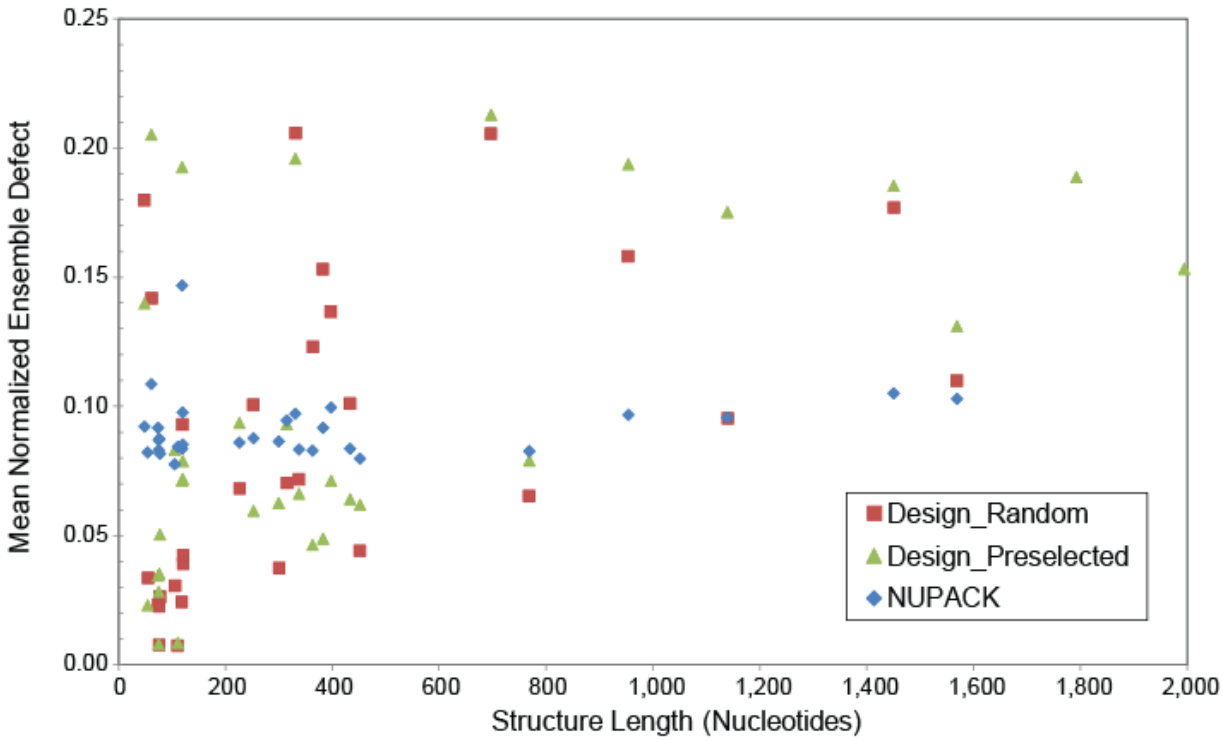

Figure S8. Mean normalized ensemble defect for long structures. Designs were made for sequences of up to 1995 nucleotides (Table S7). Mean NED is shown for ten calculations for each target structure. Design times were capped to 75 days of running time (6,480,000 seconds). Points missing for NUPACK and *Design_Random* had one or more designs that reached the maximum runtime and were terminated, so the mean could not be calculated. Designs were performed on a single core of an Opteron 2427 processor. NUPACK was run using rna99 thermodynamic parameters at 37 °C, the NED threshold was set to 0.1 so that NUPACK produced structures of similar NED as *Design_Random*, and other parameters were set to defaults.

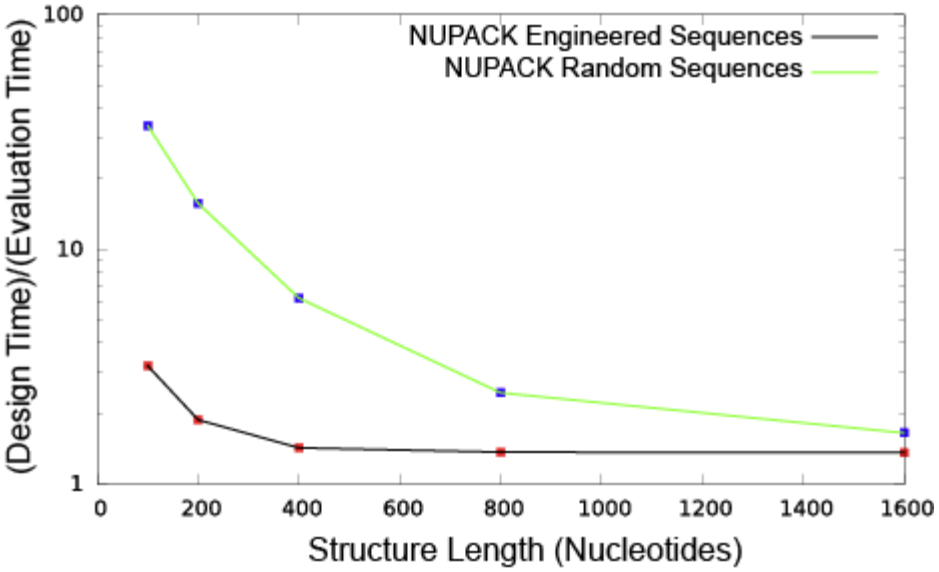

Figure S9. Asymptotic convergence as a function of length of the ratio of design cost to design evaluation on engineered structures and random structures, using the structures provided by Zadeh et al. (2011b). The thermodynamic parameters were rna1999 at 37 °C. The normalized ensemble defect threshold was set to 0.01 and other parameters were set to defaults. The mean performance for ten calculations is shown.

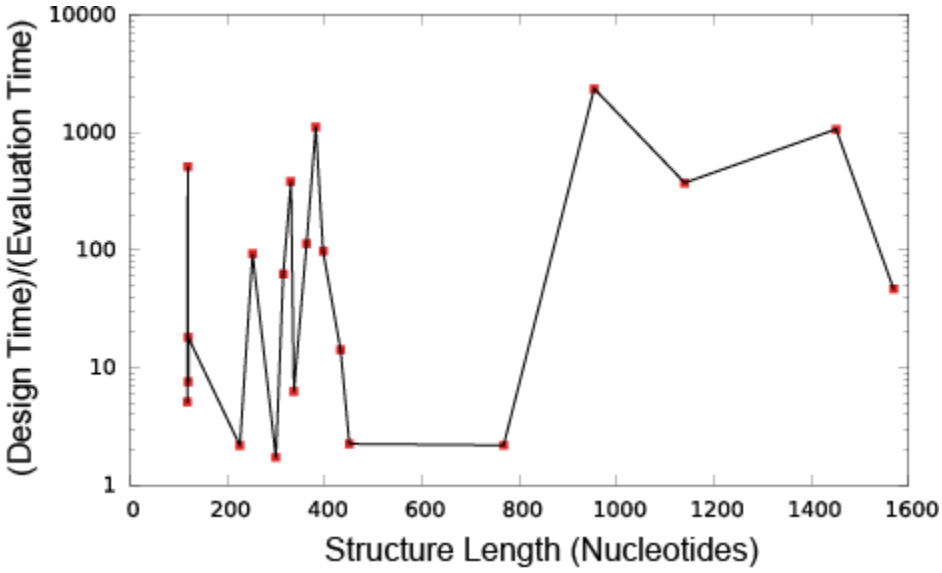

Figure S10. Asymptotic convergence as a function of length of the ratio of design cost to design evaluation on natural RNA structures for NUPACK. The thermodynamic parameters were rna1999 at 37 °C. The normalized ensemble defect threshold was set to 0.01 and other parameters were set to defaults. The mean performance for ten calculations is shown. The structures used here are those from Table S7, with the total time plotted in Figure S8.

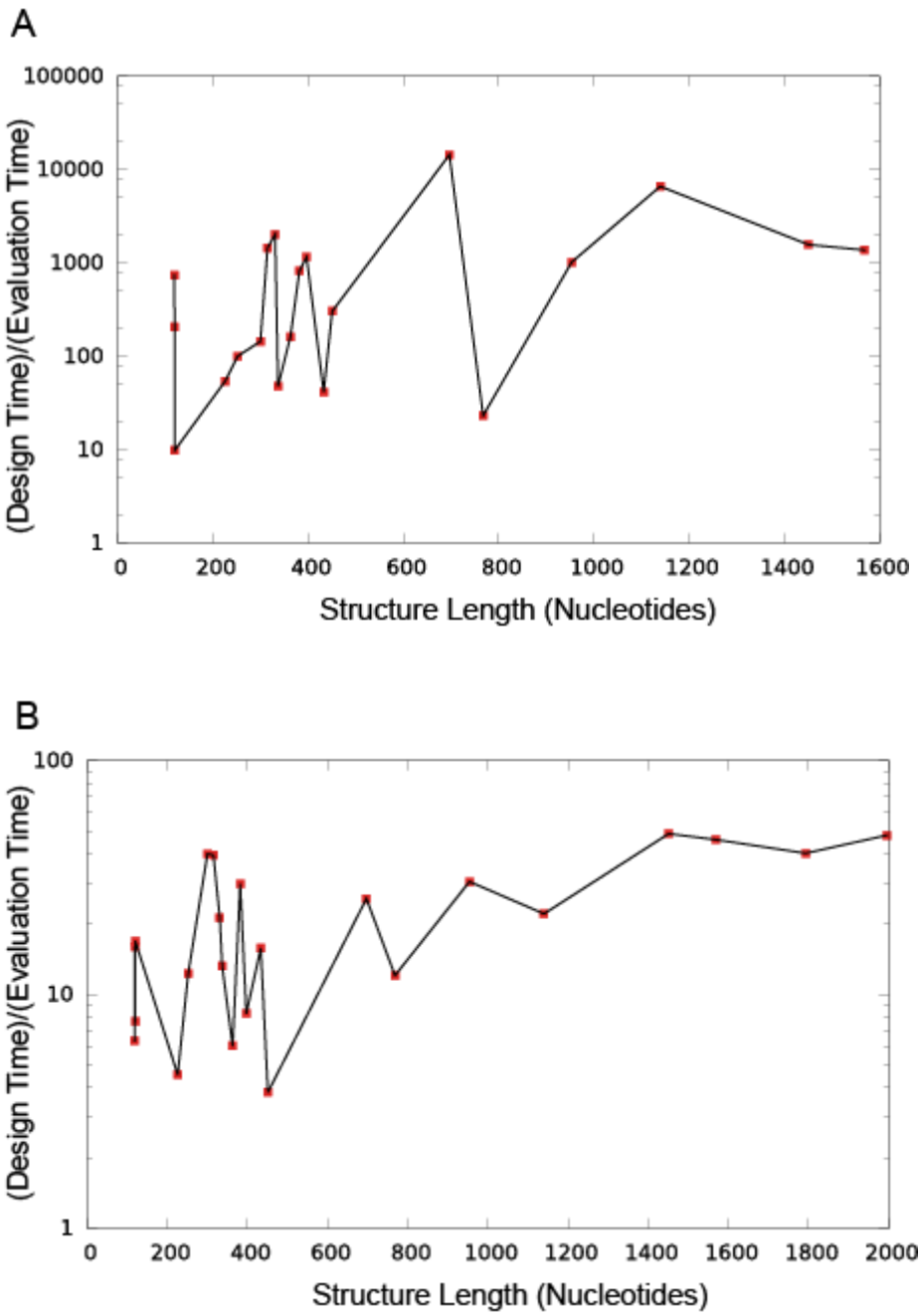

Figure S11. Asymptotic convergence as a function of length of the ratio of design cost to design evaluation on natural RNA structures for (Panel A) *Design_Random* and (Panel B) *Design_Preselected.* Parameters were set to default values. The mean performance for ten calculations is shown. The structures used here are those from Table S7, with the total time plotted in Figure S8.

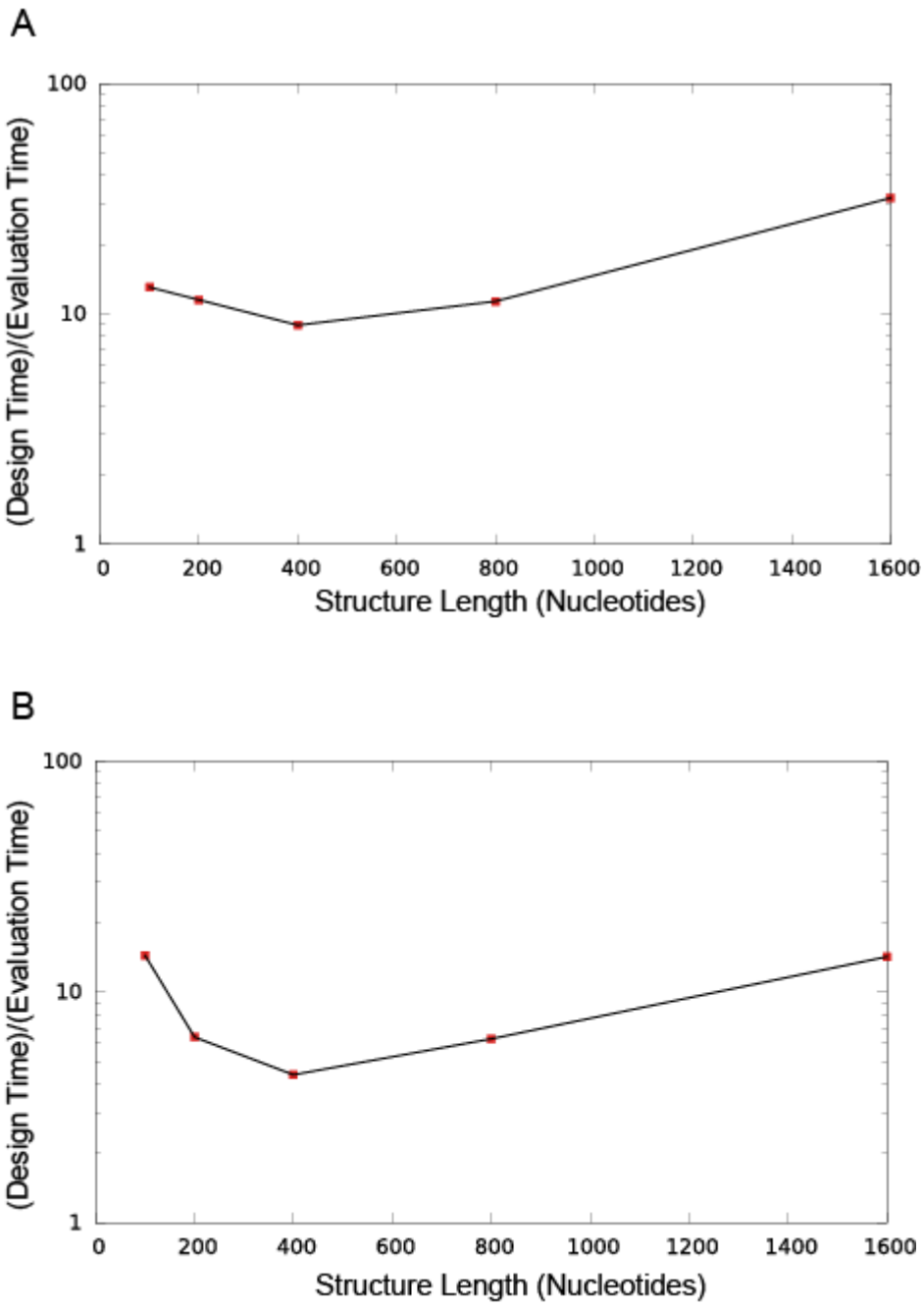

Figure S12. Asymptotic convergence as a function of length of the ratio of design cost to design evaluation on engineered RNA structures for (Panel A) ***Design_Random*** **and (Panel B)** ***Design_Preselected.*** Parameters were set to default values. The mean performance for ten calculations is shown. The structures used here are engineered structures provided by Zadeh et al. (2011b).

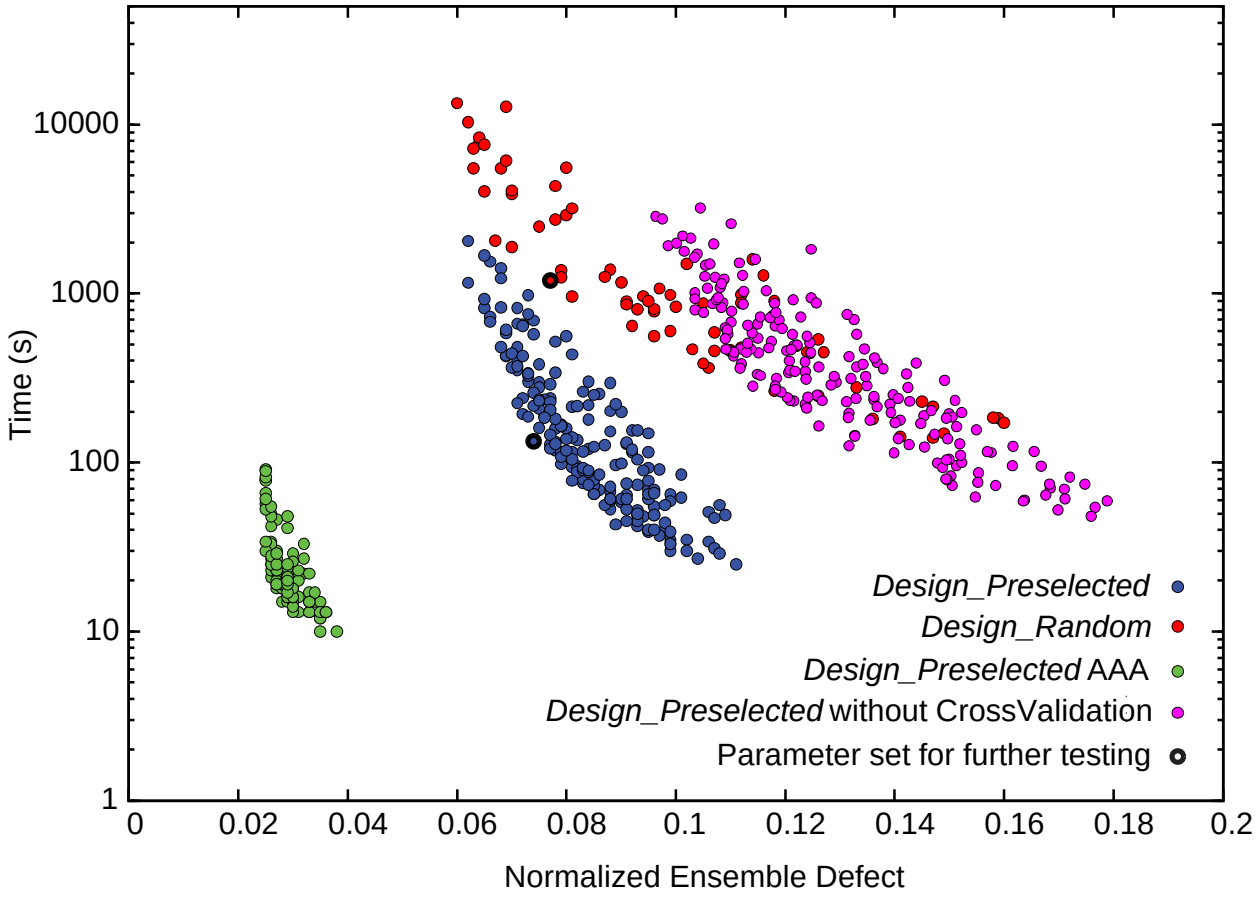

Figure S13. Algorithm performance distribution across sets of parameters. Blue and red dots represent performance of *Design* algorithm in *Preselected* and *Random* modes, respectively. Each dot is the mean for a set of parameters. Green shows the performance of the *Design* algorithm in *Preselected* mode using all Adenines in place of single stranded regions. These three results are shown in Figure 2. Purple dots show the performance of *Design* algorithm in *Preselected* mode using a database of helix and loop sequences randomly assembled from a pool of sequences remaining after the selection steps based on thermodynamics, thus not performing the cross-hybridization evaluation between the sequences. Performance is evaluated as mean time as a function of mean normalized ensemble defect.

Table S1. Randomly generated sequences for evaluation of parameter dependence.

Number	Length (nt)	Sequence
1	64	AGCUUAACGAGGGAAUGACGUUAGAAGUCACUGACUGUGUUAGAUGGUGCUGGGUCUACAUGGG
2	92	GGUAUAGUAAUUUAUACGACUCAUUGCCUCACAUUAAUAAAUGCUCACCCCCAAAUGCGUCACCCUUAGCGAGCCGUCGUGAGUCUUUAAUA
3	104	GGGUUCUUGAGAUAGUUUCAGAACUUAUGCUUACUUGCCGCUUACAAGUUCCUUCUCAACGAACCAGAAAGGAAUGGAGCGCCAGUUGAGCCUAGAG UAAGCGA
4	118	GCAACGUGAGCCGCCAUGUAUUAGAGUCGGGGUUCGCCUGUCUAUUUAAGUAUCAGGGGACCUUACAAAUUCCUAGAUUCCAGAGAAUGCCUACAA ACAGCCUAGCCAUUAGCGAUC
5	113	AGGCAUAUGCGCCUCAAUAUGAACCCCACGACCUAUGAUCGGUCUACAACAGGACUUGAAACGGACACAAACGACAUCUGCAGCUUUCCUGCGAGUU CAAAAUUCGACAGGCC
6	256	ACUCCACAAGAAGUGGCCAAUCCACGGCUUCUCUAGCGAUUGCCAAACUUCCGUCCCGUCCCAUACUGGCCGUUGAAGCAAACUUUAUUUGCACCCGU AAACUGUGCGAGCACCAUUUAUCUUCUCCUCCAAGUAGGAGUAAACACGCCAGGAUAAUGCGACGGUCUUCCAACCGCUUCUCGUUGGCAGACGCGU AACCACCAGAAAGACUGAAGUGUACCGCAGUCUAUGACCUUAGAACCCCUCAUUUAUCGGG
7	360	CGAACUGCGGUGUUGGUCCGUCCGGGUGUGUGGUGCAACCUCCAUGUCAAAACUUUCUUGCCGCCGUCCUCACACGGCUACGGAACAGCUCGAGCUCU AAGAUACUGAGCGUUGCCUAACGGAGUCCUGUACUGUGGGCGCAAAGGAUAUGCUCCUCAAUGUAGGGUCGAAUACUGGCUGGGUAGACCCGCACAG UAUCUCUACGAUCGCAGACAGCUCUUCUAUAAAUCACAUUUGUCGAAUAGACUGAAGCAUUUCCGCAUAGAGACCCAUGAGCUGCGCCCUGGGGGCC UGCUACCGGGAAGCCAUUCUCCGUCAUUCUAGAUGGCCUUCCAUCGUCGGUACAACGGAAGGGCCCGC
8	361	GAGCUACAGGUCGCUCUUUCUUUUUCAGCUUCUCAGACAAAUUUCCGGAAGUACCCUUGCGGGCCGCCAGAACGGGCAAGCGGGGAGGAAGGGGACA GUAAGACCCCUCAAGAUUAUCCGUGAACGCUGUCCGGGAUCACAGGUUAUUACUUGGCACGAGCAGAGACUAUGCAAUAUGCAGUUAAGCACCUAGU UUGGGCGAAGGUCGACUCUGAUGGUCGAACGCUACUGAAUCUGUGGGACGGGCACUGAAGCCCACUCUGUCGGGCCAUUAACUGGCCCAUCAAAGGU UUUUAACGCCAGAGUUUUAUCGCUGUCGCCAGUCAUUGAAGACAAGUUAGGGUCGUUUGACUUUCCUGGU
9	367	CAACGCAUAUUCUAAUGCAAUCAAUAGAAGUAAAUUUGCGGGUACUUGUAUCUCGGCAGAAUUGAAUCACCGGGUAAUGCUUUUAAAGAACCCCGGU CACGGCGUUCGGACCCGAAUUGACCUGUGGCCGAUUAUUUAGCUCACUAGAAGGUUGAGAAGUAGCUUCGCCAUCCUAGUAGAGGAUUUAGGCUUUC GCCCGCGGGUAAAAUUAAAAUGCAACAAACUCCGUGGCCGGAUCUACCGGCUCCAUGUGGCGCUAGUAGUUAUACACUCGAAGUCUGAAUUUCGUGA UCUAAAGUUCGCAAUUUUACGUGCCUAGUUUCGCUCAUUAAGAAUCCCUAAGGCCUCAUGCCUACCAAUAUUUGUC
10	368	GGGAUUAUCCUCGAGCGAUUCUCGGCAGAAUUCUCAAGGCUAUCUACAAUCCGGAUGGCUACCAACCGCCCCAACGGGCUCGUUAAGUUGCAUUGCG CAGAGCACGGGAUAGACGGACCGCGGACCUAGAAAUCAGUGAGACACAGUAUGGGUGUGCACGAGCUCCAAAGUUUCUUAUAUCUCGCAUAUUAGAA GAGCAUAUAACAGAUACAGACAAAUUGCAGCGCGACGGGUAGACUACAUCCAAUGUUGCACAGCUCGCACUUAUUGUUCUAUCACGGCACAUCGGCU AUAAAUAGAUGUGAGGAUAACAUUGUGUGUGCCCGGGCUGGCUACAUGGUUUCGAUUAAACCUCUCGCGACAGUCGC
11	397	CAUCGUGAAAAUCCAAUUCGCUUAUCCCACAAGAUAGGUUAUGCUUCAAGUAGCCCGAGGAACCUGAACCAAUUGGCAGUACACAGAUUCUCUGAGG ACUACAGGGGUUGCACCAUUCCGCUGACCCUUGAGGUCGAAAGAUUGUACGUUCCCGUUUCCAAACUGGCACAACGAGGAAAUGGAGACUCAUACU UACGAGGGUUCACCGACCUUUGGUAGCAUUGUGAAAACGUGCGUGUACGGUAACCACACACUUUCGUCGACCAUUCUGGUACAGAUCGAUCUCUCCG UUCGUGAUGUGUCUCCCGUUGAUGGGCAGGUAUCCCCGAUAGGCAAUCAAGUUGGCGGAUGGGCUUCUUAUUGACC
12	431	UAAAGACGAUCGAGAGAUUGAUAGUCUGGUACUUCCGUACCCCUAAUAGCAGUGGACCAAUCUCUAGUCAAUGACCCAAACGGGGGCGGCACGAAAC UACUAGUACCUUUUUCUCACUGAAUCCAGACCUAACGUACGCCUACGUUGCCAGUAACUUUGCACGACUUCGGACAGUUACGCCCUCCCCAUCGCUUA GUGGCGAUGCCAAUUCCGCCUUAUUUAAGCAAAGCGGUUGCGCUAUCGGCGCCUUCAGUCGAGGUUCUAUAGGCUAUGACUCAGUGCAGUAGUUAAC GGCACGCGAGUUCUCGUAGGAAGAUCGCUUGAGAAGUCUCACUGGCCAGCGCUCAUGCAUGCUAGGUUCAAUACCGGACGUAUAGUAGUUUUUUGAA GGCGGAGACAGAUAAUAGAUUGCA

Table S2. Randomly generated sequences for testing performance on one parameter set.

Number	Length (nt)	Sequence
1	48	AUGAGUGACGUAAUGUCUUUUAUCGGAGUUUUUGUCCCCGUAGAACUC
2	54	UUGCACAAAGGUACUGCAGCCUAAGACGGUUCGGGUUGUACCUCUGUAGCCUAC
3	74	AAGAUCGUCAAGAUCCCUGUAAUAAUAACAACGAACUCUCUCGAAAUGGGAACGAUCCAGACGUCACGGCUGAG
4	61	AUACGGAAUAACAAGGCGAGGUGUAGCCCCCGACUGUAGGAGUAGGGUAUCGGACUGGAUU
5	75	CCGAUCCGCCGGGAUCCCUCAGGUCGGUAAUGCCGCUUUAACGGCGUGUGAUAGGGAACCCAUGGCUCAGCAAUG
6	75	CCGGCACCCCACCGAAUACGUCUCGACUUGUUUUACUCCAUUUAUAUUUCCAGGUAUCAGGCAGAACCCUCCCUA
7	76	CUGCAUUGAAAUAACUCGCGUGCAUUUAGGGGGAGUGAGUUUACUUUCAUGCAGUAGUCCAUGGAAUACCCUGUCG
8	77	GGUACUAAGUCCCGGCUAAUAGUAUAACUCCAAGGAGUUCCAGACGCGGUUAUGACCACUGACUCCACCCCUAGUGU
9	105	GGGUGCGAGAUGUUCUCGAACGAGUACCCCCACGUACGCCGUACGUUUGUCACAUAGUGAGGGUUAGGGGAACCUCGACAU GCCCGUCUAGGGGUGGACCAGAAG
10	110	UUCUCAUGUUUCGUACAUACUUAAUGCCAACUUGUCGGUCCGGACUCCGCUGAUGUGUCGUGGGACUGUGACAGCGCGUAU GGAGAUAUAGGGAAGCACCGCGUUUCCCA
11	118	AAGGAUAAUUGAAUCACCACAUUGGCCUUGGCCGGUUAACCUGAUGUGACAUAACUACGACGGGAUGCUCUCAACUGCGGU GCGUAAGUGACAUUCUCGGGAAACAGGGCAAGGAGGU
12	119	GCUUGAGCCCUCAUUCCCUCCGUUUUGGCAUACGUAGGCAUACUUCACGUGUCCUAACGGCCGUACUGUAGGAACACCCUA UUCAGCAUUGACAUCGGUCGGAUGCUUUAGUUCCUUGU
13	120	CAAUUUCAUAUCGAACUGGGUUUACCAUCCCUUGACCGGUUGCCGUGGAAUCAGAUCCUCCAGUUGUCACGUCCGCGGGGC GAAUGGUACGGACUCGAGUGGAGUUACCAACGAAAGACU
14	120	CGUUGUCCGUCGGCCAUUGAUGAAAACAUUCGUCUCGACCACUCUAGUCAUCUGCCGAGUUAAUUUGGACAGUGGGGGCGA GCUAUGUUAAGUGUCAACCUUUUUGCGCCGCGCCGAACA
15	226	CGACAGCAUAACAGUUGCAGUGCUCGCCGGGUAGCACGAAGAGGUCCCUGAGACCCAGGGAACCUGCACGAUGGGCAUUUC UCCAGGACAGCCAGAGAAUGAUAUCGUACGGAUUACGCGGUAUUGUGGUCAAACUCUGGAAAAUUAACCUACUUACUGAG AAGACCGAUAAUAAGUCUCGGCUCAAGUGGGGUAUAGCCCCCCCGUAUGCCAGCCUCAGUGAGCC
16	252	UUAGUUGGAUUCACGUGACUGUCGUGUUUCCCAGCUCUUACAUGAAGAUCCUAGCUUCGAUUUGACGCAAUCUACUAUAGG UGUAAGGGUCCACUCGUUGCUCGUCACAACUAGUUAUAGACCCGGGGGCUGACUCAGCUGUAUCUUAAAGUCGUGACAAUG CUUCCAAAGAAUAGGAGGGAGGACCCCUGUGUAUGGACAUCGUUCGCGGGGUAAAGUCAAACUAUAGCUCCGGCGCGUGCA GUAGGGACG
17	300	UCCCGGGCAAGCACGUCUUGCUGAUCCAGCGCAACGAAGACGACUUAGGUGGGGAGCCCGUGGAAGUGGUUGCUAGUACCU CUCACCCCCCACAGAUUGUGGCUGUAAAUGGCCGGUUUUUGGACAUACCAGCUUGAGGUAACUGAAGGAAGUUCGACCGGU CCAAGUUUAAAUUCCGAGAUAGUUGCCCUAAUCAAAUAGCGGGUCUGAGAACUAUACCCGACGGGUCUCCCUGGCGUCCCU UUUAGCGGAUCUUUAUGCGCAAUGAUGGGCUUCCGGGAUUCAAUGUCGUAUUGGCCA
18	315	GCGCUGAGUAUAACCCAUGAGGCUAGCGAUGUGUUCCGUGCAGUUCGAAUCACGUUGCCAAUCUCCGUCUCUGGUUCACCU AGCAGUGGUACAUCAGUUCUAUGCGGACUAGUUUGCCUAACACACGGCCGGCCACACCUUCCCUAGCCAUCUCGUCCAAUG ACAAUCGGUCCAUUUAGCACUACAACGAGCUCAUCUUUCCUAGCAGUUCCCAGUGCAUUUCUGUGCUUUCAAAGUGGACCU CCGCUGCCAGAAAGUUUCCAAAUAGCGGCAUGGCAUACAAUAAACUAGGUUCACGCUUCUAUUCUUUAACAA
19	331	ACUGCUCUUUUCCCUAUAAAACGAGGUUGCGGGCAUCGUACGCUUAUGAUGACAAUUAGCCCAUGAGUGCUAUAUUCGCCG AGUAGGUGACUCCGACCUAGUAGUUCACGGUGAGUGAGCCUAGAUGGAGGGCGAACCCGAUGGUACCCAGGUUAAGUGGCA AGGCGUGUCGGUCCAGUCGCUGAUGGGAAUACCGGACAUGUGGAUGGGUCAGUACCUACUUCACUGCAUAGGGAGGGAACC AUACCCACGUAUACGGGGCUCUCCCCUAGACUACACGCCUACUUCCGUUACGCAAUGUUAUAUACUCUCUGCAGACUGAGUC CAGCUAA
20	337	UGCAAUCUAAAGGUUCGUGAGAGUAGUUUGUGAAUAAAAACAGUACAGAUGUUAGUUCGGUGCAGAAGAAGCCACCCCUG ACCACUUAAGUAAAAGAACACUGGAUUCAGUCUACAUACAGAAGAACAAGACCGUCCGUGAGUUUAAGACGUCGCCGGCUA CA
21	366	UUUAAACCCGUUUUCUGAGAGUUGGCAUAAGCGAGCCACUGUAGCCACCGGUGGGUGCGGGGCCGAUGGGCAGGUGAAGGA CUUGAGUUGUCGCGAGAACCUCGUCCUCCAGACCCGCACAUGAACACGGACAUGGAUCAACUGCUAUUAACACAACACCC GACCGGUGGAGUGCUAGCUUCUAAGCAAGUUCUGUGGACAUGCAGGACCUAAUUCAUCGCUCUCUUCGGACGGAAAAUAAA UUUCUGUAGGAUACUCAUAACCGUGCAGUAGCGCGUAUUCUGAAGCACAACUUUCAGCACGUUGCUAUGUCUGAUGCACCG CGUAGGAUGGGCAUUCAUCCCCGGACUGGGCUCGUAACC

22	382	UGCAACAGGCACUCGUGCGCGAGCGGUCUAUAGAUUAUUCAUAUAGUCUCAUCGCUUCUUGAGGACUCAGUAUAUUCGGGG GCCCACUACCUCUGGGGGAUAUCCCCAACGUAGUGGUCUGAGAUACGAUCUUCCACUUGAAAUGUAGAGACAACAGGUAAC UCAAACGGGGGGUUCCCACUGUUGGCUCUUAAGAGCGAGAGACCUACCAAAGAAGAACUUGGAACUGUCCCAGCAGAGGAA AAAUCUCCGGACUUGACAAUCUCGGUUAAUGCCAUACUGUUCUCCGAACAGACAAGAGGUGACGGUUGCCGGAUAACGGUU AUACACUAGGUGCGAUACCCUGGCGAAUCGUUUUGCCCAGGGCGCGCGAUAUAACCUG
23	397	AAGCAGCUGGCAUGCCCUACCGCAUUGACUUUGUAGGCUGAUGUUACUAGGACUUCCGUCAUGACACGUCAUGAUUUGGCC UCUAUUUUUACCAGUGAGACACUAGUUCAAUUCAUGGCACCCGCGCGAGGUGCGGUGCGGACCAAGGCCUAGUUGUUGGAA UCAACAACCCACCUUUCUGGAUGCGAAUACUCCGAACUCGGGUGUCGCGACAUGACCCCCUCGUAUCGCUGUUCCUCGGAA UCAAGGGCUUCUAAUUACAGCGUGAUGGCCGAUCAAUUAGUGACCCACAGGAGGAACACGAAUUGCGGAAUUAUUCGCUGA UCUAACAUUUUGCCAAGAAUCCUGCGGGUUUAUAUUAUAAGCCCCCGUUCGACGAGAACUCGGCGUUACCCCA
24	433	UUCUUCUUCAAUAGGCAUCCAUUCGCGUCCCACCUGGUAGAGUCGUUAGAAGAAGGUGGCCAUCGCUACACGCUGACAUUU UAGAAUGGCACUUAUCCGUACGCCAAUAGCCCUAGUGGUAACUUCCACUGCCCCUUAAUCUCAGAUAAAUGAACGCAUCAG CCCGGUUUCGGCUGGAUGCUGAUUUUAAGCAGGGCGCGGCUUGCCUGAGCCCGGAGCAUUCCGCGGAUCAAUGCGGGUUGC UGGCAAGAUUAGCUGCCACCGUUCUUAUGUGAUCUGUAACCGGGCGGCCACGCACACUUCAUCAAGUCUCACCCCUACGAG UAUAGUUCCUAGAAGCCAAGAAUACUCUCCCGCUCAUUAGGCCCCGUCCGACAGCUCGUAGCCCCGUGCCCGCAAGAUAAG UUACUGGAUAGACGCUUCAGCUAGUAUA
25	451	CCUGGGUGCGCGCGUGCGAGUAGAGAAGAAUAUCCGUUCUUUGGUGGAAGGUAGGACAGGUCAAAUCCGGCGAAUAGAUU GAAAUCGAUGCUAUAUUACCAGUGCUCAGUAAUAGCCACGCGAGGGUGCUUAGAUGUUAUCGCUACAUAGAAAAUUUCGC GUUCACAACGUGCUGCGUCAGAAAGUGUAGUUUACCGCAGGAGGCGCGCAAUGUUUAAGCAGGGCGGGCUCCUUCGAAGUU GGGAAAAGGUCAUACAGACCUUGAGGAGUUCGGACCGACUAAAGUGCUGCGGCGUUGAACACGACAGGAGCUAUGCAGCAG UAUUCGCAACGAGUACUUGAUUGACAACUCGAGCGUCUAGCUUGGCAGUGAUGGAUGGGGUUAGGAACUAGAGUACUUAU CUCGCAUCGCUAUAGCCGAAGACGGCCUAAUCGCAUGGAGUUGACAGGG

Table S3. Performance comparison between two modes for natural sequences, using median NED and median time.

Names	Length	*Design_Preselected*		*Design_Random*	
		Median NED	**Median Time (s)**	**Median NED**	**Median Time (s)**
Schistosoma haematobium[1]	48	0.12±0.025	1±0	0.19±0.026	14±3
Peach latent mosaic viroid[2]	54	0.02±0.005	1±1	0.03±0.018	2±1
synthetic construct[3]	61	0.19±0.020	0±0	0.14±0.000	71±4
Saccharomyces cerevisiae[4]	74	0.03±0.009	1±0	0.02±0.004	2±1
Gallus gallus domesticus[5]	75	0.02±0.011	0±0	0.03±0.013	3±1
Galago senegalensis[6]	75	0.00±0.001	0±0	0.01±0.002	5±1
Nicotiana rustica[7]	76	0.04±0.013	0±0	0.02±0.009	3±1
Mycoplasma mycoides[8]	77	0.05±0.003	0±0	0.03±0.001	8±5
Thermus thermophilus[8]	105	0.09±0.006	4±0	0.03±0.001	294±85
Homo sapiens[6]	110	0.00±0.001	0±0	0.01±0.001	28±4
Stilbum vulgare[9]	118	0.07±0.007	2±0	0.02±0.002	39±16
Avocado sunblotch viroid[1]	119	0.17±0.009	5±3	0.10±0.049	302±14
Methanococcus vannielii[9]	120	0.07±0.004	2±0	0.04±0.010	56±23
Streptomyces griseus[9]	120	0.07±0.011	5±2	0.03±0.009	62±25
Homo sapiens[3]	226	0.09±0.011	8±2	0.07±0.001	91±24
Anabaena[10]	252	0.06±0.002	29±20	0.12±0.061	217±59
Homo sapiens A[8]	300	0.06±0.004	148±74	0.04±0.003	497±323
Sulfolobus acidocaldarius[11]	315	0.09±0.005	198±32	0.06±0.014	9223±1179
Homo sapiens[12]	331	0.18±0.033	160±13	0.19±0.033	9896±7412
Escherichia coli[13]	337	0.06±0.014	85±27	0.06±0.031	316±62
Escherichia coli[14]	363	0.04±0.007	51±15	0.11±0.014	1362±369
Streptomyces aureofaciens[14]	382	0.05±0.011	294±158	0.15±0.041	8047±1818
Mus spretus[15]	397	0.07±0.005	50±22	0.13±0.028	10860±2349
Tetrahymena thermophila[10]	433	0.06±0.011	255±94	0.09±0.026	528±142
Oryctolagus cuniculus[15]	451	0.06±0.003	49±9	0.04±0.007	4122±2322
Natural Structure Average	**201**	**0.07**	**54**	**0.07**	**1842**

All calculations were run on a cluster with 24 dual processors, six core Opteron 2427 nodes. Error estimates are median absolute deviations. [1]hammerhead ribozyme type I; [2] hammerhead ribozyme type III; [3] hairpin ribozyme; [4] $tRNA^{ACG}$; [5] $tRNA^{BCA}$, where "B" represents 2'-O-methylcytidine; [6] Y RNA; [7] $tRNA^{GPA}$, where "P" represents pseudouridine; [8] SRP RNA; [9] 5S RNA; [10] group I intron; [11] RNase P; [12] 7SK RNA; [13] RNase E 5' UTR; [14] tmRNA; [15] telomerase RNA.

Table S4. Performance comparison between two modes for random sequences.

Names	Length	*Design_Preselected*		*Design_Random*	
		Median NED	**Median Time (s)**	**Median NED**	**Median Time (s)**
Sequence 1	48	0.11±0.021	0±0	0.04±0.011	5±1
Sequence 2	54	0.12±0.026	0±0	0.01±0.000	1±1
Sequence 3	61	0.10±0.055	1±0	0.21±0.091	11±2
Sequence 4	74	0.05±0.010	2±1	0.03±0.002	11±7
Sequence 5	75	0.04±0.005	1±0	0.05±0.018	9±4
Sequence 6	75	0.08±0.020	1±1	0.01±0.001	5±3
Sequence 7	76	0.04±0.012	1±0	0.02±0.003	1±0
Sequence 8	77	0.05±0.008	2±1	0.02±0.005	5±3
Sequence 9	105	0.06±0.006	13±6	0.04±0.011	53±28
Sequence 10	110	0.13±0.033	1±0	0.01±0.001	12±4
Sequence 11	118	0.07±0.006	14±6	0.03±0.003	55±20
Sequence 12	119	0.08±0.012	2±1	0.02±0.007	23±5
Sequence 13	120	0.06±0.009	27±11	0.03±0.013	25±16
Sequence 14	120	0.08±0.012	1±0	0.03±0.004	30±10
Sequence 15	226	0.07±0.009	19±10	0.01±0.002	106±44
Sequence 16	252	0.05±0.006	15±5	0.03±0.003	67±31
Sequence 17	300	0.08±0.011	34±6	0.03±0.007	306±105
Sequence 18	315	0.06±0.011	368±142	0.03±0.010	629±309
Sequence 19	331	0.07±0.012	142±80	0.01±0.002	537±176
Sequence 20	337	0.11±0.010	257±103	0.05±0.036	2346±589
Sequence 21	363	0.09±0.013	357±268	0.03±0.001	1494±665
Sequence 22	382	0.05±0.003	57±28	0.03±0.007	238±141
Sequence 23	397	0.05±0.006	438±197	0.04±0.006	789±271
Sequence 24	433	0.06±0.004	518±310	0.03±0.002	5435±1584
Sequence 25	451	0.06±0.002	236±131	0.03±0.001	1295±571
Random Structure Average	**201**	**0.07**	**100**	**0.03**	**539**

All calculations were run on a cluster with 24 dual processors, six core Opteron 2427 nodes.

Table S5. NUPACK Performance comparison for natural sequences.

Names	Length	*NUPACK* Mean NED	*NUPACK* Mean Time (s)	*Design_Random* Median NED	*Design_Random* Median Time (s)
Schistosoma haematobium[1]	48	0.09±0.006	2±1	0.09±0.005	1±0
Peach latent mosaic viroid[2]	54	0.08±0.013	0±0	0.08±0.009	0±0
synthetic construct[3]	61	0.11±0.001	2±2	0.11±0.001	1±0
Saccharomyces cerevisiae[4]	74	0.09±0.009	1±1	0.10±0.003	1±0
Gallus gallus domesticus[5]	75	0.09±0.010	2±2	0.09±0.004	1±1
Galago senegalensis[6]	75	0.08±0.016	2±1	0.09±0.009	2±0
Nicotiana rustica[7]	76	0.09±0.014	3±2	0.09±0.007	3±1
Mycoplasma mycoides[8]	77	0.08±0.016	2±1	0.09±0.013	2±1
Thermus thermophilus[8]	105	0.08±0.014	1±0	0.08±0.013	1±0
Homo sapiens[6]	110	0.08±0.011	15±12	0.08±0.007	12±3
Stilbum vulgare[9]	118	0.08±0.013	3±2	0.09±0.009	3±1
Avocado sunblotch viroid[1]	119	0.15±0.042	334±184	0.14±0.028	298±106
Methanococcus vannielii[9]	120	0.09±0.010	5±3	0.09±0.005	4±2
Streptomyces griseus[9]	120	0.10±0.005	12±13	0.10±0.002	9±6
Homo sapiens[3]	226	0.09±0.006	12±1	0.09±0.002	11±1
Anabaena[10]	252	0.09±0.009	708±857	0.09±0.008	427±306
Homo sapiens A[8]	300	0.09±0.007	22±4	0.08±0.003	22±2
Sulfolobus acidocaldarius[11]	315	0.09±0.016	931±614	0.10±0.004	940±468
Homo sapiens[12]	331	0.10±0.009	6668±5563	0.10±0.007	5707±4523
Escherichia coli[13]	337	0.08±0.007	119±85	0.08±0.003	103±52
Escherichia coli[14]	363	0.08±0.004	2594±2224	0.08±0.003	2474±2174
Streptomyces aureofaciens[14]	382	0.09±0.005	30073±31219	0.09±0.003	23490±20550
Mus spretus[15]	397	0.10±0.013	2936±1685	0.09±0.003	2742±1110
Tetrahymena thermophila[10]	433	0.08±0.009	601±345	0.08±0.006	552±117
Oryctolagus cuniculus[15]	451	0.08±0.007	106±47	0.08±0.004	84±11
Natural Structure Average	**201**	**0.09**	**1806**	**0.09**	**1476**

All calculations were run on a cluster with 24 dual processors, six core Opteron 2427 nodes. NUPACK calculations were run using the Turner99 energy model at 37 °C. The target NED threshold was set to 0.10 and other parameters were set to defaults. For Design_Random, the parameters were set to defaults. These parameter choices were made so that the final performance of the programs for NED would be similar to facilitate the comparison of time performance. Error estimates for means are standard deviations and error estimates for medians are median absolute deviations. [1]hammerhead ribozyme type I; [2] hammerhead ribozyme type III; [3] hairpin ribozyme; [4] $tRNA^{ACG}$; [5] $tRNA^{BCA}$, where "B" represents 2'-O-

methylcytidine; [6] Y RNA; [7] $tRNA^{GPA}$, where "P" represents pseudouridine; [8] SRP RNA; [9] 5S RNA; [10] group I intron; [11] RNase P; [12] 7SK RNA; [13] RNase E 5' UTR; [14] tmRNA; [15] telomerase RNA.

Table S6. NUPACK Performance comparison for random sequences.

		NUPACK		*Design_Random*	
Names	**Length**	**Mean NED**	**Mean Time (s)**	**Median NED**	**Median Time (s)**
Sequence 1	48	0.09±0.006	1±1	0.09±0.002	1±0
Sequence 2	54	0.08±0.019	0±0	0.09±0.009	0±0
Sequence 3	61	0.09±0.006	10±7	0.09±0.005	10±5
Sequence 4	74	0.09±0.009	3±2	0.09±0.005	3±1
Sequence 5	75	0.11±0.071	3±2	0.09±0.008	2±2
Sequence 6	75	0.07±0.014	0±0	0.07±0.009	0±0
Sequence 7	76	0.07±0.016	0±0	0.07±0.009	0±0
Sequence 8	77	0.09±0.009	8±6	0.09±0.008	6±3
Sequence 9	105	0.09±0.008	6±6	0.10±0.002	4±1
Sequence 10	110	0.06±0.012	1±0	0.06±0.007	1±0
Sequence 11	118	0.08±0.012	3±1	0.08±0.009	2±1
Sequence 12	119	0.08±0.017	4±3	0.08±0.012	2±1
Sequence 13	120	0.07±0.016	2±1	0.07±0.014	2±1
Sequence 14	120	0.07±0.012	3±2	0.07±0.011	3±0
Sequence 15	226	0.08±0.006	15±3	0.08±0.005	15±2
Sequence 16	252	0.08±0.009	15±6	0.08±0.008	12±1
Sequence 17	300	0.08±0.009	82±52	0.08±0.006	64±39
Sequence 18	315	0.09±0.007	137±62	0.09±0.005	116±18
Sequence 19	331	0.08±0.008	40±13	0.08±0.008	35±4
Sequence 20	337	0.08±0.011	143±93	0.08±0.007	138±92
Sequence 21	363	0.09±0.008	83±40	0.09±0.005	77±28
Sequence 22	382	0.08±0.010	130±83	0.07±0.006	107±42
Sequence 23	397	0.09±0.006	103±47	0.08±0.003	94±35
Sequence 24	433	0.08±0.007	75±14	0.08±0.005	73±8
Sequence 25	451	0.09±0.010	112313±1	0.08±0.004	46212±44593
Random Structure Average	**201**	**0.08**	**4527**	**0.08**	**1879**

All calculations were run on a cluster with 24 dual processors, six core Opteron 2427 nodes. NUPACK calculations were run using the Turner99 energy model at 37 °C. The target NED threshold was set to 0.10 and other parameters were set to defaults. For Design_Random, the parameters were set to defaults. These parameter choices were made so that the final performance of the programs for NED would be similar to facilitate the comparison of time performance.

Table S7. Structures used for long structure benchmarks.

Index	Length	Species	RNA family
1	48	*Schistosoma haematobium*	hammerhead ribozyme type I
2	54	Peach latent mosaic viroid	hammerhead ribozyme type III
3	61	synthetic construct	hairpin ribozyme
4	74	*Saccharomyces cerevisiae*	$tRNA^{ACG}$
5	75	*Gallus gallus domesticus*	$tRNA^{BCA}$, where "B" represents 2'-O-methylcytidine
6	75	*Galago senegalensis*	Y RNA
7	76	*Nicotiana rustica*	$tRNA^{GPA}$, where "P" represents pseudouridine
8	77	*Mycoplasma mycoides*	SRP RNA
9	105	*Thermus thermophilus*	SRP RNA
10	110	*Homo sapiens*	Y RNA
11	118	*Stilbum vulgare*	5S RNA
12	119	Avocado sunblotch viroid	hammerhead ribozyme type I
13	120	*Methanococcus vannielii*	5S RNA
14	120	*Streptomyces griseus*	5S RNA
15	226	*Homo sapiens*	hairpin ribozyme
16	252	*Anabaena*	group I intron
17	300	*Homo sapiens A*	SRP RNA
18	315	*Sulfolobus acidocaldarius*	RNase P
19	331	*Homo sapiens*	7SK RNA
20	337	*Escherichia coli*	RNase E 5' UTR
21	363	*Escherichia coli*	tmRNA
22	382	*Streptomyces aureofaciens*	tmRNA
23	397	*Mus spretus*	telomerase RNA
24	433	*Tetrahymena thermophila*	group I intron
25	451	*Oryctolagus cuniculus*	telomerase RNA
26	697	*Caenorhabditis elegans (*mitochondrial*)*	16S RNA
27	768	*Saccharomyces cerevisiae B1*	group II intron
28	954	*Homo sapiens (*mitochondrial*)*	16S RNA
29	1140	*Chlamydomonas reinhardtii (*mitochondrial*)*	16S RNA
30	1451	*Giardia intestinalis*	16S RNA
31	1569	*Aquifex pyrophilus*	16S RNA
32	1793	*Toxoplasma gondii*	16S RNA
33	1995	*Drosophila melanogaster*	16S RNA

Table S8. Statistics for the overlap of the pre-selected helix database and the known helix database (for unique helixes and for all helixes).

Helix Size (nt)	**Unique Helix Overlap**	**Unique Helix Count**	**Helixes in Assembled Database**	**Percent Unique Helix Overlap**
2	10	21	10	100.0%
3	18	105	18	100.0%
4	30	376	30	100.0%
5	50	939	60	83.3%
6	31	857	100	31.0%
7	18	1030	281	6.4%
8	4	435	799	0.5%
9	3	311	300	1.0%
10	0	168	400	0.0%
TOTAL:	**164**	**4242**	**1998**	**8.2%**

Unique Helix Overlap is the number of helices in common in the assembled database and the set of sequences with known structure. Unique Helix Count is the total number of helixes in the set of sequences with known structure. Helixes in Assembled Database is the number of helixes in the assembled database. Percent Helix Overlap is the Unique Helix Overlap, divided by the Helices in Assembled Database. Therefore, it is the percent of helices in the database that are found in nature.

Table S9. Numbers of helices in helix trimming procedure.

Helix length (bp)	Starting count	Count after step 1	Count after step 2	Count after step 3	Count after step 4
3	32	20	20	20	18
4	136	94	93	93	30
5	512	257	252	252	60
6	2080	520	503	503	100
7	8192	1556	1476	1392	281
8	32896	4952	4635	4137	799
9	131072	6526	5955	4793	300
10	524800	11929	10806	8298	400
TOTAL	**699720**	**25854**	**23740**	**19488**	**1988**

Table S10. Numbers of single-stranded sequences in loop trimming procedure.

Sequence Length (nt)	**Starting count**	**Count after step 1**	**Count after step 2**	**Count after step 3**	**Count after step 4**	**Count after step 5**
3	64	64	64	40	39	18
4	256	256	252	150	108	30
5	1024	1024	996	500	281	60
6	4096	4096	3936	1000	672	100
7	16384	14080	13380	2000	1492	150
8	65536	42176	39400	5691	4126	260
9	262144	116712	106172	13183	10045	496
10	1048576	306184	268942	27436	23237	941
TOTAL	**1398080**	**484592**	**433142**	**50000**	**40000**	**2055**