

Supplementary material of “IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming”

Kengo Sato^{1,*} Yuki Kato^{2,*} Michiaki Hamada¹ Tatsuya Akutsu³
Kiyoshi Asai^{1,4}

S1 Prediction of pseudoknot-free structures with IPknot

We evaluated the performance of predicting secondary structures without pseudoknots using our method and the other existing methods. The dataset was obtained from the RNA STRAND database [1]. We selected 141 *pseudoknot-free* RNA sequences whose length is between 140 nt and 500 nt and has at most 85 % identity to the other sequences. Since both of IPknot and CentroidFold employ the same posterior probability distribution, the McCaskill model [3] with Boltzmann likelihood parameters [2], accuracy of these two methods is comparable. However, IPknot attempts to predict additional base pairs that are pseudoknotted to base pairs at lower levels of the decomposition. In particular, the iterative refinement algorithm might emphasize base-pairing probabilities for pseudoknotted base pairs. Such nature of IPknot makes the number of false positive base pairs increase.

S2 Accuracy of common secondary structure prediction depends on the quality of alignments

To investigate the correlation between the quality of multiple alignments and the accuracy of common secondary structure prediction, we prepared various quality of multiple alignments produced by: ClustalW, ProbConsRNA, CentroidAlign and manual curation (reference). Figure S2 clearly reveals that the accuracy of common secondary structure prediction deeply depends on the quality of multiple alignments, that is, the higher alignment quality we can have, the more accurate prediction we can perform.

Figure 7 in the main paper and Fig. S3 show that IPknot can make robust predictions compared with *hxmatch*. This results from the averaged base-pairing probabilities that *on average* produce the optimal structure when it is mapped to each individual sequence in the given alignment.

*The authors wish it to be known that in their opinion the first two authors should be regarded as joint First Authors.

¹Graduate School of Frontier Sciences, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8561, Japan.

²Graduate School of Information Science, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan.

³Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan.

⁴Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), 2-41-6, Aomi, Koto-ku, Tokyo 135-0064, Japan.

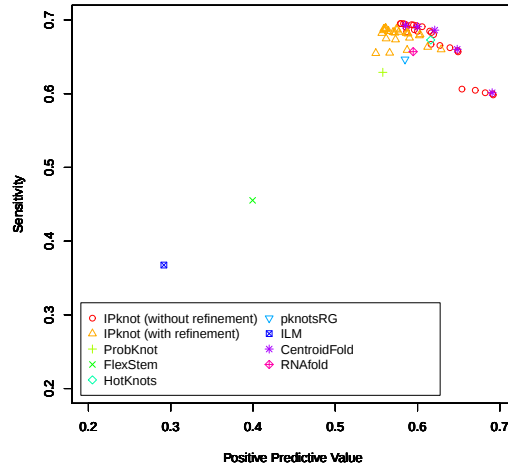


Figure S1: The PPV–Sensitivity plots of the experiments on the pkfree141 dataset.

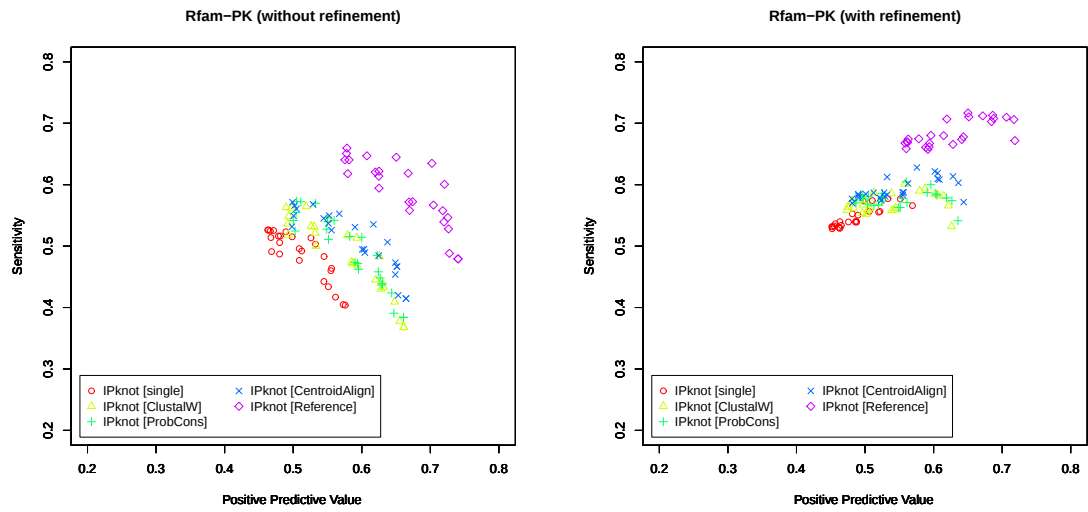


Figure S2: The PPV–Sensitivity plots for various quality of alignments.

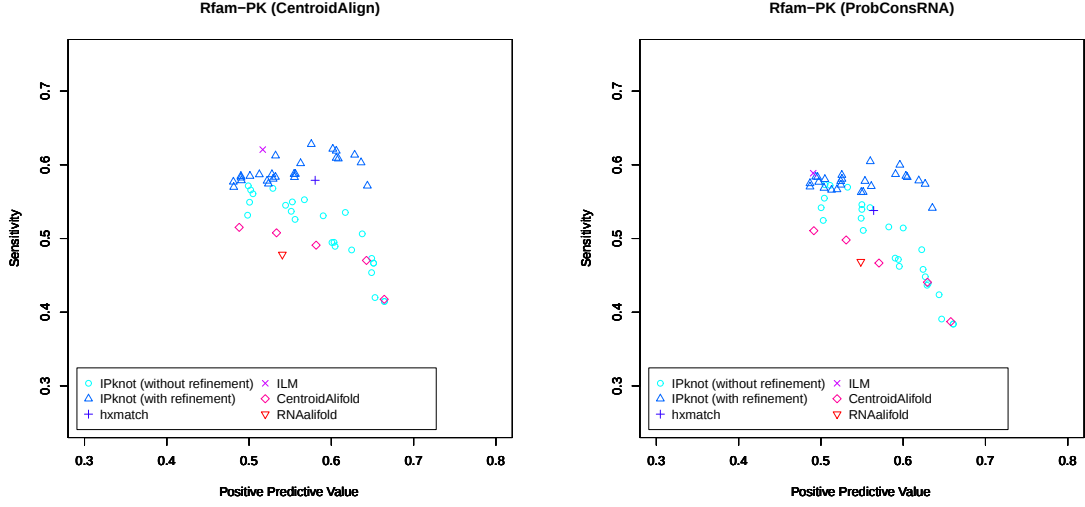


Figure S3: The PPV–Sensitivity plots of the experiments on the Rfam-PK dataset.

S3 The number of iterations of the iterative refinement algorithm

Figure S4 shows difference in prediction accuracy when changing the number of iterations of the iterative refinement algorithm for base-pairing probabilities. Although applying the algorithm once significantly improved the prediction accuracy, applying the algorithm twice might make no obvious difference.

S4 The number of decomposed levels of pseudoknots

This supplement provides experimental results on the RS-pkfree141 (pseudoknot-free) dataset and the RS-pk399 (pseudoknotted) dataset for the maximum decomposed levels $m = 1, 2, 3$. Note that CentroidFold is equivalent to IPknot with the maximum decomposed level $m = 1$. It can be observed that the selection of a conflicting level between predicted structures and correct structures causes the degradation of the accuracy compared with the best results. For example, for the pseudoknot-free dataset, IPknot with the maximum decomposed level $m = 3$ and the iterative refinement algorithm cut down the accuracy compared with CentroidFold because of increasing false positive base pairs. These results indicate that the appropriate number of decomposed levels should be selected, although correct structures might be unknown. This is a drawback of our method.

S5 Comparison with a variant of IPknot

We can consider a variant of IPknot that does not use the decomposition of pseudoknotted secondary structures. In such a case, the expected gain function (2) in the main paper is simply employed. The IP problem is formulated as follows:

$$\begin{aligned} &\text{maximize} && \sum_{i < j \text{ s.t. } p_{ij} > \theta} p_{ij} y_{ij} \end{aligned} \quad (\text{S1})$$

$$\begin{aligned} &\text{subject to} && \sum_{j=1}^{i-1} y_{ji} + \sum_{j=i+1}^n y_{ij} \leq 1 \quad (1 \leq \forall i \leq n), \end{aligned} \quad (\text{S2})$$

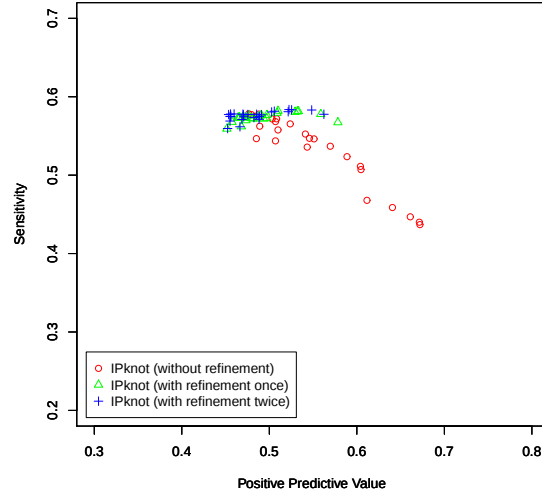


Figure S4: The PPV–Sensitivity plots of the experiments on the RS-pk388 dataset, ranging in the number of iterations of the iterative refinement algorithm from 0 to 2.

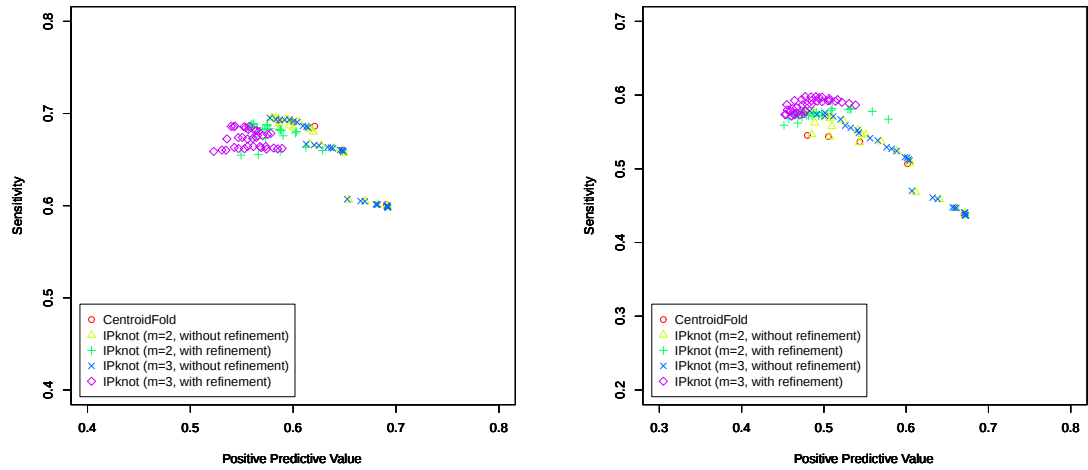


Figure S5: The PPV–Sensitivity plots comparing the number of decomposed levels. (Left) the experiment on the RS-pkfree141 dataset. (Right) the experiment on the RS-pk388 dataset.

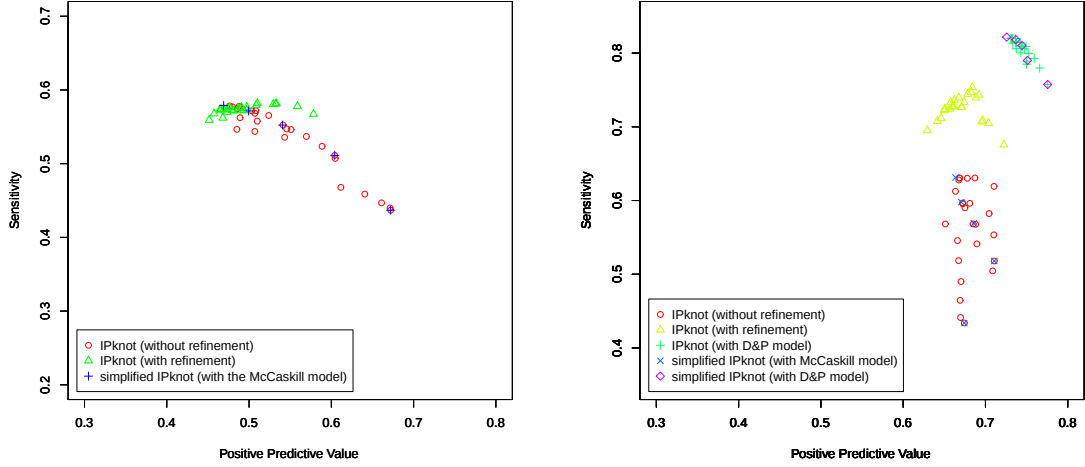


Figure S6: The PPV–Sensitivity plots comparing the original IPknot and its variant. (Left) the experiment on the RS-pk388 dataset. (Right) the experiment on the pk168 dataset.

where $\theta = 1/(\gamma + 1)$. We call this variant the *simplified* IPknot. We can employ two kinds of base-pairing probability matrices: one is calculated by the probability distribution *with* pseudoknots such as the D&P model; the other *without* pseudoknots such as the McCaskill model. The former performs an exact estimation in the sense of γ -centroid estimator, but spends enormous time and space. The latter runs fast, but may sacrifice the accuracy. Figure S6 shows comparisons between the original IPknot and simplified IPknot, indicating that the base-pairing probability produced by the iterative refinement algorithm is more accurate and efficient than the simplified IPknot. The iterative refinement algorithm is based on the decomposition of pseudoknotted secondary structures into a set of pseudoknot-free structures, and the mixture of the distribution of pseudoknot-free structures improves the accuracy. On the other hand, the IP problem (S1)–(S2) does not provide the decomposition unlike the original IPknot. Therefore, the iterative refinement algorithm cannot be directly applied to the simplified IPknot.

S6 Capability of representing arbitrary pseudoknots by the decomposition

We can construct a graph from any given pseudoknotted structure $y \in \mathcal{S}(x)$. Each vertex in the graph corresponds to a base pair such that $y_{ij} = 1$. Two vertices are connected by an edge if and only if the corresponding base pairs y_{ij} and y_{kl} are pseudoknotted, that is, $i < k < j < l$ or $k < i < l < j$. Thus, the decomposition of y into m pseudoknot-free substructures $(y^{(1)}, \dots, y^{(m)})$ (i.e. m -partite structure) is equivalent to the m -coloring problem. It is obvious that the chromatic number m is at most the number of vertices in the graph, which is less than or equal to $\lfloor |x|/2 \rfloor$ in this case. Several algorithms for finding the chromatic number of any graph have been studied. For a graph $G = (V, E)$ with vertices $V = \{v_1, \dots, v_n\}$, the Welsh–Powell algorithm [4] can find a coloring that uses $\max_{1 \leq i \leq n} \min\{d(v_i) + 1, i\}$ colors by a greedy algorithm, where $d(v_i)$ is the number of edges connected to the vertex v_i , called the degree of the vertex v_i . In other words, the chromatic number of the graph is at most one more than the graph’s maximum degree. This means that the number of decompositions of $y \in \mathcal{S}(x)$ is at most one more than the maximum number of pseudoknotted base pairs for each base pair. Therefore, for any RNA secondary structure y , there exists $m \leq \lfloor |x|/2 \rfloor$

such that y can be decomposed into m pseudoknot-free substructures.

References

- [1] Andronescu,M., Bereg,V., Hoos,H.H. and Condon,A. (2008) RNA STRAND: the RNA secondary structure and statistical analysis database. *BMC Bioinform.*, **9**, 340.
- [2] Andronescu,M., Condon,A., Hoos,H.H., Mathews,D.H. and Murphy,K.P. (2010b) Computational approaches for RNA energy parameter estimation. *RNA*, **16**, 2304–2318.
- [3] McCaskill,J.S. (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, **29**, 1105–1119.
- [4] Welsh,D.J.A. and Powell,M.B. (1967) An upper bound for the chromatic number of a graph and its application to timetabling problems. *The Computer Journal*, **10**, 85–86.

A Experimental results in detail

In Appendix, the detailed tables of experimental results performed in this work are shown. We evaluated the prediction accuracy through Matthews Correlation Coefficient (MCC) as well as PPV and sensitivity. MCC is a balanced accuracy measure between PPV and sensitivity, defined as:

$$MCC = \frac{TP \cdot TN - FN \cdot FP}{\sqrt{(TP + FN)(TN + FP)(TP + FP)(TN + FN)}},$$

where TP is the number of base pairs appearing in both the reference and predicted structures, FP is the number of base pairs that appear in the predicted structure but not in the reference, TN is the number of base pairs that appear in neither structure, and FN is the number of base pairs that appear in the reference but not in the predicted structure.

Table S1: IPknot without refinement on the RS-pk388 dataset.

$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC	$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC
1	1	0.672	0.437	0.537	8	1	0.507	0.544	0.523
1	2	0.671	0.440	0.538	8	2	0.510	0.558	0.531
1	4	0.661	0.447	0.538	8	4	0.507	0.568	0.534
1	8	0.641	0.459	0.537	8	8	0.503	0.572	0.534
1	16	0.612	0.468	0.530	8	16	0.488	0.578	0.528
2	1	0.605	0.507	0.551	16	1	0.485	0.547	0.512
2	2	0.604	0.511	0.552	16	2	0.489	0.562	0.522
2	4	0.589	0.524	0.552	16	4	0.486	0.575	0.526
2	8	0.570	0.537	0.550	16	8	0.480	0.577	0.524
2	16	0.551	0.546	0.546	16	16	0.476	0.578	0.522
4	1	0.543	0.536	0.537					
4	2	0.546	0.547	0.544					
4	4	0.541	0.552	0.544					
4	8	0.524	0.565	0.542					
4	16	0.508	0.572	0.536					

Table S2: IPknot with one-time refinement on the RS-pk388 dataset.

$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC	$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC
1	1	0.578	0.567	0.570	8	1	0.468	0.562	0.510
1	2	0.559	0.578	0.566	8	2	0.473	0.570	0.517
1	4	0.533	0.582	0.554	8	4	0.479	0.573	0.521
1	8	0.510	0.582	0.542	8	8	0.481	0.572	0.522
1	16	0.487	0.575	0.527	8	16	0.469	0.574	0.516
2	1	0.532	0.581	0.554	16	1	0.452	0.559	0.500
2	2	0.529	0.581	0.552	16	2	0.457	0.568	0.507
2	4	0.509	0.579	0.540	16	4	0.464	0.573	0.513
2	8	0.491	0.577	0.529	16	8	0.465	0.573	0.514
2	16	0.476	0.576	0.521	16	16	0.466	0.573	0.514
4	1	0.494	0.572	0.529					
4	2	0.497	0.577	0.533					
4	4	0.497	0.572	0.531					
4	8	0.486	0.572	0.525					
4	16	0.473	0.573	0.518					

Table S3: IPknot with two-time refinement on the RS-pk388 dataset.

$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC	$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC
1	1	0.562	0.578	0.567	8	1	0.466	0.562	0.509
1	2	0.548	0.583	0.563	8	2	0.469	0.571	0.515
1	4	0.525	0.583	0.551	8	4	0.471	0.575	0.518
1	8	0.506	0.582	0.540	8	8	0.470	0.577	0.518
1	16	0.482	0.573	0.523	8	16	0.456	0.578	0.511
2	1	0.522	0.584	0.550	16	1	0.452	0.560	0.500
2	2	0.522	0.581	0.548	16	2	0.455	0.569	0.506
2	4	0.503	0.581	0.538	16	4	0.456	0.575	0.509
2	8	0.486	0.578	0.527	16	8	0.455	0.575	0.509
2	16	0.470	0.578	0.519	16	16	0.453	0.577	0.509
4	1	0.488	0.571	0.526					
4	2	0.491	0.577	0.530					
4	4	0.489	0.575	0.527					
4	8	0.475	0.577	0.521					
4	16	0.460	0.578	0.513					

Table S4: IPknot ($m = 3$) without refinement on the RS-pk388 dataset.

$\gamma^{(1)}$	$\gamma^{(2)}$	$\gamma^{(3)}$	PPV	Sen	MCC	$\gamma^{(1)}$	$\gamma^{(2)}$	$\gamma^{(3)}$	PPV	Sen	MCC
1	1	1	0.672	0.437	0.537	2	2	16	0.598	0.516	0.552
1	1	2	0.672	0.437	0.537	2	4	4	0.588	0.524	0.552
1	1	4	0.672	0.437	0.537	2	4	8	0.582	0.527	0.551
1	1	8	0.672	0.437	0.537	2	4	16	0.576	0.529	0.549
1	1	16	0.672	0.437	0.537	2	8	8	0.565	0.539	0.549
1	2	2	0.671	0.440	0.538	2	8	16	0.556	0.541	0.546
1	2	4	0.671	0.440	0.538	2	16	16	0.542	0.549	0.543
1	2	8	0.671	0.440	0.538	4	4	4	0.541	0.552	0.544
1	2	16	0.671	0.440	0.538	4	4	8	0.533	0.556	0.541
1	4	4	0.661	0.447	0.538	4	4	16	0.526	0.559	0.539
1	4	8	0.659	0.447	0.537	4	8	8	0.520	0.567	0.541
1	4	16	0.656	0.448	0.537	4	8	16	0.511	0.571	0.538
1	8	8	0.638	0.459	0.536	4	16	16	0.501	0.576	0.535
1	8	16	0.632	0.461	0.535	8	8	8	0.501	0.572	0.533
1	16	16	0.607	0.470	0.529	8	8	16	0.491	0.574	0.528
2	2	2	0.604	0.511	0.552	8	16	16	0.483	0.579	0.526
2	2	4	0.603	0.513	0.553	16	16	16	0.473	0.579	0.521
2	2	8	0.601	0.515	0.553						

Table S5: IPknot ($m = 3$) with refinement on the RS-pk388 dataset.

$\gamma^{(1)}$	$\gamma^{(2)}$	$\gamma^{(3)}$	PPV	Sen	MCC	$\gamma^{(1)}$	$\gamma^{(2)}$	$\gamma^{(3)}$	PPV	Sen	MCC
1	1	1	0.538	0.586	0.559	2	2	16	0.485	0.592	0.533
1	1	2	0.531	0.589	0.556	2	4	4	0.485	0.587	0.531
1	1	4	0.522	0.590	0.552	2	4	8	0.480	0.587	0.528
1	1	8	0.514	0.592	0.549	2	4	16	0.474	0.587	0.525
1	1	16	0.507	0.593	0.545	2	8	8	0.467	0.583	0.519
1	2	2	0.517	0.594	0.551	2	8	16	0.462	0.582	0.516
1	2	4	0.506	0.595	0.546	2	16	16	0.456	0.580	0.511
1	2	8	0.498	0.597	0.542	4	4	4	0.477	0.578	0.523
1	2	16	0.490	0.598	0.539	4	4	8	0.475	0.575	0.520
1	4	4	0.494	0.597	0.540	4	4	16	0.471	0.573	0.517
1	4	8	0.484	0.598	0.535	4	8	8	0.468	0.575	0.516
1	4	16	0.477	0.598	0.532	4	8	16	0.464	0.575	0.514
1	8	8	0.473	0.594	0.527	4	16	16	0.459	0.578	0.513
1	8	16	0.464	0.593	0.522	8	8	8	0.464	0.573	0.513
1	16	16	0.455	0.587	0.514	8	8	16	0.461	0.571	0.510
2	2	2	0.505	0.592	0.544	8	16	16	0.454	0.574	0.508
2	2	4	0.498	0.591	0.540	16	16	16	0.453	0.573	0.507
2	2	8	0.492	0.592	0.537						

Table S6: Comparison of IPknot with competitive methods on the RS-pk388 dataset.

	PPV	Sen	MCC
IPknot without refinement ($\gamma^{(1)} = 2, \gamma^{(2)} = 2$)	0.604	0.511	0.552
IPknot with refinement ($\gamma^{(1)} = 1, \gamma^{(2)} = 1$)	0.578	0.567	0.570
ProbKnot	0.450	0.480	0.462
FlexStem	0.383	0.416	0.396
HotKnots	0.467	0.497	0.480
pknotsRG	0.421	0.457	0.437
ILM	0.333	0.427	0.375
CentroidFold ($\gamma = 2$)	0.602	0.507	0.549
RNAfold	0.417	0.455	0.433

Table S7: IPknot without refinement on the pk168 dataset.

$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC	$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC
1	1	0.674	0.434	0.532	8	1	0.666	0.546	0.592
1	2	0.670	0.441	0.535	8	2	0.675	0.590	0.620
1	4	0.669	0.465	0.549	8	4	0.672	0.595	0.621
1	8	0.670	0.490	0.565	8	8	0.672	0.597	0.621
1	16	0.667	0.518	0.579	8	16	0.678	0.630	0.642
2	1	0.709	0.505	0.587	16	1	0.651	0.568	0.598
2	2	0.711	0.518	0.596	16	2	0.664	0.612	0.626
2	4	0.710	0.553	0.615	16	4	0.668	0.628	0.636
2	8	0.704	0.582	0.629	16	8	0.669	0.630	0.637
2	16	0.710	0.619	0.652	16	16	0.669	0.630	0.637
4	1	0.690	0.541	0.600					
4	2	0.688	0.568	0.614					
4	4	0.685	0.568	0.612					
4	8	0.681	0.596	0.625					
4	16	0.687	0.631	0.646					

Table S8: IPknot with refinement on the pk168 dataset.

$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC	$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC
1	1	0.722	0.676	0.690	8	1	0.646	0.712	0.671
1	2	0.704	0.705	0.696	8	2	0.657	0.724	0.682
1	4	0.688	0.739	0.706	8	4	0.661	0.727	0.686
1	8	0.678	0.744	0.704	8	8	0.660	0.727	0.685
1	16	0.668	0.740	0.696	8	16	0.658	0.733	0.688
2	1	0.695	0.707	0.692	16	1	0.629	0.695	0.654
2	2	0.697	0.708	0.694	16	2	0.642	0.707	0.667
2	4	0.692	0.743	0.710	16	4	0.650	0.722	0.678
2	8	0.683	0.747	0.706	16	8	0.651	0.724	0.679
2	16	0.684	0.753	0.711	16	16	0.650	0.724	0.679
4	1	0.674	0.733	0.696					
4	2	0.671	0.726	0.691					
4	4	0.671	0.727	0.691					
4	8	0.662	0.728	0.687					
4	16	0.662	0.736	0.692					

Table S9: IPknot with the D&P model on the pk168 dataset.

$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC	$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC
1	1	0.776	0.757	0.757	8	1	0.737	0.806	0.764
1	2	0.766	0.780	0.765	8	2	0.737	0.814	0.768
1	4	0.759	0.793	0.769	8	4	0.737	0.815	0.768
1	8	0.753	0.799	0.769	8	8	0.737	0.815	0.767
1	16	0.747	0.806	0.769	8	16	0.732	0.820	0.768
2	1	0.750	0.785	0.760	16	1	0.733	0.813	0.765
2	2	0.750	0.789	0.762	16	2	0.732	0.820	0.768
2	4	0.749	0.809	0.771	16	4	0.732	0.820	0.768
2	8	0.742	0.814	0.770	16	8	0.732	0.820	0.768
2	16	0.736	0.820	0.770	16	16	0.731	0.820	0.767
4	1	0.743	0.800	0.764					
4	2	0.744	0.809	0.769					
4	4	0.744	0.809	0.769					
4	8	0.737	0.814	0.768					
4	16	0.732	0.820	0.768					

Table S10: Comparison of IPknot with competitive methods on the pk168 dataset.

	PPV	Sen	MCC
IPknot without refinement ($\gamma^{(1)} = 2, \gamma^{(2)} = 16$)	0.710	0.619	0.652
IPknot with refinement ($\gamma^{(1)} = 2, \gamma^{(2)} = 16$)	0.684	0.753	0.711
IPknot with D&P model ($\gamma^{(1)} = 2, \gamma^{(2)} = 4$)	0.749	0.809	0.771
ProbKnot	0.648	0.601	0.613
FlexStem	0.711	0.770	0.731
HotKnots	0.727	0.690	0.696
pknotsRG	0.761	0.777	0.761
ILM	0.647	0.597	0.611
NUPACK	0.691	0.670	0.670
PKNOTS	0.730	0.735	0.722
CentroidFold ($\gamma = 2$)	0.714	0.503	0.589
RNAfold	0.658	0.507	0.567

Table S11: IPknot without refinement on the RS-pkfree141 dataset.

$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC	$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC
1	1	0.692	0.598	0.638	8	1	0.600	0.685	0.637
1	2	0.691	0.600	0.638	8	2	0.596	0.687	0.636
1	4	0.682	0.601	0.635	8	4	0.595	0.693	0.639
1	8	0.670	0.605	0.631	8	8	0.593	0.694	0.638
1	16	0.654	0.606	0.624	8	16	0.585	0.694	0.634
2	1	0.649	0.657	0.649	16	1	0.587	0.684	0.630
2	2	0.649	0.659	0.650	16	2	0.586	0.689	0.632
2	4	0.639	0.662	0.647	16	4	0.583	0.695	0.633
2	8	0.627	0.665	0.642	16	8	0.580	0.695	0.631
2	16	0.616	0.667	0.638	16	16	0.579	0.695	0.631
4	1	0.619	0.680	0.646					
4	2	0.617	0.683	0.646					
4	4	0.615	0.685	0.645					
4	8	0.606	0.691	0.643					
4	16	0.599	0.692	0.640					

Table S12: IPknot with one refinement on the RS-pkfree141 dataset.

$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC	$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC
1	1	0.629	0.660	0.641	8	1	0.570	0.683	0.621
1	2	0.612	0.663	0.634	8	2	0.572	0.683	0.621
1	4	0.587	0.659	0.619	8	4	0.574	0.688	0.625
1	8	0.566	0.655	0.605	8	8	0.575	0.687	0.625
1	16	0.549	0.655	0.596	8	16	0.562	0.688	0.618
2	1	0.602	0.679	0.636	16	1	0.557	0.682	0.612
2	2	0.603	0.680	0.637	16	2	0.559	0.686	0.616
2	4	0.590	0.676	0.628	16	4	0.561	0.689	0.618
2	8	0.573	0.673	0.617	16	8	0.560	0.688	0.617
2	16	0.562	0.675	0.612	16	16	0.559	0.688	0.616
4	1	0.586	0.682	0.629					
4	2	0.588	0.682	0.630					
4	4	0.588	0.681	0.629					
4	8	0.577	0.684	0.624					
4	16	0.565	0.685	0.618					

Table S13: IPknot ($m = 3$) without refinement on the RS-pkfree141 dataset.

$\gamma^{(1)}$	$\gamma^{(2)}$	$\gamma^{(3)}$	PPV	Sen	MCC	$\gamma^{(1)}$	$\gamma^{(2)}$	$\gamma^{(3)}$	PPV	Sen	MCC
1	1	1	0.692	0.598	0.638	2	2	16	0.646	0.661	0.649
1	1	2	0.692	0.598	0.638	2	4	4	0.639	0.662	0.646
1	1	4	0.692	0.598	0.638	2	4	8	0.637	0.663	0.646
1	1	8	0.692	0.598	0.638	2	4	16	0.634	0.663	0.645
1	1	16	0.692	0.598	0.638	2	8	8	0.625	0.665	0.641
1	2	2	0.691	0.600	0.638	2	8	16	0.621	0.666	0.639
1	2	4	0.691	0.600	0.638	2	16	16	0.613	0.667	0.636
1	2	8	0.691	0.600	0.638	4	4	4	0.615	0.685	0.645
1	2	16	0.691	0.600	0.638	4	4	8	0.613	0.686	0.645
1	4	4	0.682	0.601	0.635	4	4	16	0.610	0.687	0.644
1	4	8	0.681	0.601	0.634	4	8	8	0.604	0.691	0.642
1	4	16	0.680	0.601	0.634	4	8	16	0.600	0.692	0.641
1	8	8	0.670	0.605	0.631	4	16	16	0.596	0.693	0.639
1	8	16	0.666	0.605	0.629	8	8	8	0.592	0.693	0.637
1	16	16	0.653	0.607	0.624	8	8	16	0.587	0.693	0.634
2	2	2	0.649	0.659	0.650	8	16	16	0.583	0.694	0.632
2	2	4	0.648	0.660	0.650	16	16	16	0.577	0.695	0.630
2	2	8	0.647	0.660	0.650						

Table S14: IPknot ($m = 3$) with refinement on the RS-pkfree141 dataset.

$\gamma^{(1)}$	$\gamma^{(2)}$	$\gamma^{(3)}$	PPV	Sen	MCC	$\gamma^{(1)}$	$\gamma^{(2)}$	$\gamma^{(3)}$	PPV	Sen	MCC
1	1	1	0.589	0.662	0.621	2	2	16	0.565	0.676	0.614
1	1	2	0.585	0.662	0.619	2	4	4	0.564	0.674	0.613
1	1	4	0.581	0.662	0.617	2	4	8	0.561	0.673	0.610
1	1	8	0.574	0.661	0.613	2	4	16	0.556	0.672	0.607
1	1	16	0.569	0.663	0.611	2	8	8	0.550	0.674	0.605
1	2	2	0.573	0.664	0.614	2	8	16	0.547	0.674	0.603
1	2	4	0.568	0.664	0.610	2	16	16	0.536	0.672	0.596
1	2	8	0.561	0.664	0.607	4	4	4	0.566	0.680	0.617
1	2	16	0.556	0.664	0.604	4	4	8	0.564	0.681	0.616
1	4	4	0.553	0.662	0.601	4	4	16	0.563	0.681	0.615
1	4	8	0.547	0.662	0.598	4	8	8	0.556	0.683	0.612
1	4	16	0.543	0.663	0.596	4	8	16	0.554	0.685	0.612
1	8	8	0.535	0.660	0.590	4	16	16	0.543	0.686	0.607
1	8	16	0.531	0.660	0.588	8	8	8	0.557	0.686	0.615
1	16	16	0.523	0.659	0.583	8	8	16	0.552	0.685	0.611
2	2	2	0.578	0.678	0.623	8	16	16	0.542	0.686	0.606
2	2	4	0.575	0.677	0.621	16	16	16	0.540	0.686	0.605
2	2	8	0.570	0.676	0.617						

Table S15: Comparison of IPknot with competitive methods on the RS-pkfree141 dataset.

	PPV	Sen	MCC
IPknot without refinement ($\gamma^{(1)} = 2, \gamma^{(2)} = 2$)	0.649	0.659	0.650
IPknot with refinement ($\gamma^{(1)} = 1, \gamma^{(2)} = 1$)	0.629	0.660	0.641
ProbKnot	0.558	0.629	0.589
FlexStem	0.400	0.455	0.423
HotKnots	0.616	0.673	0.640
pknotsRG	0.585	0.647	0.611
ILM	0.292	0.368	0.324
CentroidFold ($\gamma = 2$)	0.648	0.660	0.650
RNAfold	0.595	0.657	0.622

Table S16: IPknot without refinement on the Rfam-PK dataset (reference).

$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC	$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC
1	1	0.740	0.479	0.584	8	1	0.625	0.594	0.601
1	2	0.740	0.479	0.584	8	2	0.625	0.614	0.612
1	4	0.728	0.488	0.586	8	4	0.625	0.622	0.617
1	8	0.727	0.528	0.610	8	8	0.620	0.620	0.613
1	16	0.705	0.567	0.624	8	16	0.607	0.647	0.621
2	1	0.720	0.539	0.613	16	1	0.580	0.618	0.590
2	2	0.726	0.546	0.619	16	2	0.581	0.640	0.603
2	4	0.718	0.558	0.623	16	4	0.576	0.640	0.600
2	8	0.721	0.601	0.649	16	8	0.578	0.651	0.607
2	16	0.702	0.635	0.661	16	16	0.578	0.659	0.611
4	1	0.670	0.558	0.601					
4	2	0.674	0.572	0.611					
4	4	0.669	0.572	0.609					
4	8	0.668	0.619	0.635					
4	16	0.650	0.645	0.642					

Table S17: IPknot with refinement on the Rfam-PK dataset (reference).

$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC	$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC
1	1	0.719	0.672	0.689	8	1	0.592	0.657	0.618
1	2	0.717	0.706	0.707	8	2	0.589	0.661	0.617
1	4	0.706	0.710	0.704	8	4	0.594	0.667	0.624
1	8	0.672	0.712	0.687	8	8	0.594	0.662	0.621
1	16	0.651	0.710	0.675	8	16	0.578	0.675	0.619
2	1	0.684	0.703	0.689	16	1	0.560	0.659	0.601
2	2	0.688	0.708	0.693	16	2	0.562	0.669	0.606
2	4	0.686	0.713	0.695	16	4	0.558	0.668	0.604
2	8	0.650	0.717	0.677	16	8	0.561	0.670	0.606
2	16	0.619	0.707	0.656	16	16	0.563	0.674	0.609
4	1	0.641	0.674	0.652					
4	2	0.644	0.678	0.655					
4	4	0.628	0.666	0.641					
4	8	0.615	0.680	0.641					
4	16	0.596	0.680	0.630					

Table S18: Comparison of IPknot with competitive methods on the Rfam-PK dataset (reference).

	PPV	Sen	MCC
IPknot without refinement ($\gamma^{(1)} = 2, \gamma^{(2)} = 16$)	0.702	0.635	0.661
IPknot with refinement ($\gamma^{(1)} = 1, \gamma^{(2)} = 2$)	0.717	0.706	0.707
hxmatch	0.728	0.668	0.689
ILM	0.574	0.681	0.619
CentroidAlifold ($\gamma = 2$)	0.703	0.542	0.607
RNAalifold	0.609	0.547	0.568

Table S19: IPknot without refinement on the Rfam-PK dataset (CentroidAlign).

$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC	$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC
1	1	0.664	0.415	0.513	8	1	0.556	0.526	0.532
1	2	0.664	0.415	0.513	8	2	0.551	0.537	0.536
1	4	0.653	0.420	0.513	8	4	0.553	0.550	0.543
1	8	0.649	0.454	0.533	8	8	0.544	0.545	0.537
1	16	0.625	0.484	0.540	8	16	0.529	0.568	0.541
2	1	0.651	0.467	0.538	16	1	0.498	0.532	0.506
2	2	0.651	0.467	0.538	16	2	0.501	0.549	0.517
2	4	0.649	0.473	0.541	16	4	0.505	0.561	0.525
2	8	0.638	0.506	0.557	16	8	0.502	0.566	0.526
2	16	0.617	0.535	0.565	16	16	0.499	0.571	0.527
4	1	0.605	0.489	0.533					
4	2	0.603	0.495	0.536					
4	4	0.601	0.495	0.534					
4	8	0.590	0.531	0.550					
4	16	0.567	0.553	0.552					

Table S20: IPknot with refinement on the Rfam-PK dataset (CentroidAlign).

$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC	$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC
1	1	0.644	0.572	0.598	8	1	0.529	0.581	0.548
1	2	0.636	0.603	0.612	8	2	0.522	0.578	0.543
1	4	0.628	0.613	0.614	8	4	0.528	0.587	0.550
1	8	0.606	0.619	0.605	8	8	0.524	0.574	0.542
1	16	0.563	0.602	0.575	8	16	0.501	0.585	0.535
2	1	0.608	0.608	0.602	16	1	0.481	0.570	0.517
2	2	0.606	0.609	0.602	16	2	0.481	0.577	0.520
2	4	0.602	0.622	0.606	16	4	0.490	0.585	0.528
2	8	0.576	0.628	0.595	16	8	0.490	0.582	0.527
2	16	0.532	0.612	0.564	16	16	0.491	0.579	0.525
4	1	0.555	0.583	0.563					
4	2	0.556	0.587	0.566					
4	4	0.555	0.587	0.565					
4	8	0.532	0.583	0.551					
4	16	0.512	0.587	0.541					

Table S21: Comparison of IPknot with competitive methods on the Rfam-PK dataset (CentroidAlign).

	PPV	Sen	MCC
IPknot without refinement ($\gamma^{(1)} = 2, \gamma^{(2)} = 16$)	0.617	0.535	0.565
IPknot with refinement ($\gamma^{(1)} = 1, \gamma^{(2)} = 4$)	0.628	0.613	0.614
hxmatch	0.580	0.579	0.572
ILM	0.517	0.621	0.559
CentroidAlifold ($\gamma = 2$)	0.643	0.470	0.537
RNAalifold	0.540	0.478	0.499

Table S22: IPknot without refinement on the Rfam-PK dataset (ProbCons).

$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC	$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC
1	1	0.661	0.384	0.491	8	1	0.551	0.511	0.523
1	2	0.661	0.384	0.491	8	2	0.549	0.527	0.531
1	4	0.647	0.391	0.491	8	4	0.550	0.540	0.538
1	8	0.644	0.424	0.511	8	8	0.550	0.546	0.541
1	16	0.624	0.458	0.523	8	16	0.532	0.570	0.545
2	1	0.629	0.437	0.512	16	1	0.503	0.525	0.505
2	2	0.630	0.439	0.514	16	2	0.500	0.542	0.513
2	4	0.627	0.448	0.518	16	4	0.504	0.555	0.522
2	8	0.623	0.485	0.539	16	8	0.511	0.572	0.534
2	16	0.600	0.514	0.546	16	16	0.505	0.573	0.531
4	1	0.595	0.462	0.513					
4	2	0.594	0.472	0.518					
4	4	0.590	0.473	0.518					
4	8	0.582	0.516	0.540					
4	16	0.560	0.542	0.545					

Table S23: IPknot with refinement on the Rfam-PK dataset (ProbCons).

$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC	$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC
1	1	0.635	0.541	0.577	8	1	0.520	0.567	0.536
1	2	0.627	0.574	0.592	8	2	0.513	0.566	0.532
1	4	0.619	0.579	0.590	8	4	0.524	0.578	0.544
1	8	0.591	0.587	0.582	8	8	0.526	0.581	0.546
1	16	0.554	0.578	0.559	8	16	0.505	0.580	0.534
2	1	0.605	0.583	0.587	16	1	0.487	0.570	0.520
2	2	0.603	0.585	0.587	16	2	0.487	0.575	0.522
2	4	0.596	0.600	0.593	16	4	0.494	0.584	0.529
2	8	0.560	0.605	0.576	16	8	0.496	0.583	0.530
2	16	0.525	0.586	0.548	16	16	0.497	0.577	0.528
4	1	0.561	0.571	0.560					
4	2	0.551	0.563	0.551					
4	4	0.549	0.563	0.550					
4	8	0.524	0.573	0.542					
4	16	0.504	0.569	0.528					

Table S24: Comparison of IPknot with competitive methods on the Rfam-PK dataset (ProbCons).

	PPV	Sen	MCC
IPknot without refinement ($\gamma^{(1)} = 2, \gamma^{(2)} = 16$)	0.600	0.514	0.546
IPknot with refinement ($\gamma^{(1)} = 2, \gamma^{(2)} = 4$)	0.596	0.600	0.593
hxmatch	0.564	0.538	0.543
ILM	0.491	0.589	0.530
CentroidAlifold ($\gamma = 2$)	0.630	0.441	0.513
RNAalifold	0.549	0.469	0.498

Table S25: IPknot without refinement on the Rfam-PK dataset (ClustalW).

$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC	$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC
1	1	0.661	0.368	0.479	8	1	0.533	0.500	0.508
1	2	0.661	0.368	0.479	8	2	0.532	0.521	0.520
1	4	0.655	0.377	0.484	8	4	0.531	0.532	0.524
1	8	0.648	0.409	0.503	8	8	0.526	0.532	0.522
1	16	0.620	0.445	0.515	8	16	0.519	0.565	0.535
2	1	0.628	0.430	0.508	16	1	0.493	0.518	0.497
2	2	0.631	0.433	0.512	16	2	0.492	0.536	0.506
2	4	0.629	0.442	0.515	16	4	0.493	0.548	0.513
2	8	0.626	0.483	0.539	16	8	0.497	0.562	0.522
2	16	0.593	0.513	0.543	16	16	0.490	0.563	0.518
4	1	0.593	0.468	0.516					
4	2	0.587	0.472	0.516					
4	4	0.585	0.474	0.516					
4	8	0.579	0.517	0.538					
4	16	0.551	0.544	0.540					

Table S26: IPknot with refinement on the Rfam-PK dataset (ClustalW).

$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC	$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC
1	1	0.626	0.532	0.568	8	1	0.502	0.552	0.520
1	2	0.622	0.566	0.585	8	2	0.496	0.553	0.517
1	4	0.615	0.581	0.590	8	4	0.507	0.562	0.527
1	8	0.588	0.594	0.583	8	8	0.504	0.559	0.524
1	16	0.538	0.586	0.554	8	16	0.499	0.575	0.529
2	1	0.606	0.584	0.588	16	1	0.474	0.558	0.508
2	2	0.602	0.584	0.586	16	2	0.475	0.564	0.510
2	4	0.580	0.590	0.579	16	4	0.481	0.572	0.517
2	8	0.558	0.601	0.573	16	8	0.481	0.573	0.518
2	16	0.513	0.587	0.542	16	16	0.484	0.562	0.513
4	1	0.545	0.562	0.547					
4	2	0.542	0.559	0.545					
4	4	0.539	0.558	0.542					
4	8	0.521	0.571	0.540					
4	16	0.497	0.567	0.524					

Table S27: Comparison of IPknot with competitive methods on the Rfam-PK dataset (ClustalW).

	PPV	Sen	MCC
IPknot without refinement ($\gamma^{(1)} = 2, \gamma^{(2)} = 16$)	0.593	0.513	0.543
IPknot with refinement ($\gamma^{(1)} = 1, \gamma^{(2)} = 4$)	0.615	0.581	0.590
hxmatch	0.530	0.490	0.501
ILM	0.466	0.571	0.509
CentroidAlifold ($\gamma = 2$)	0.624	0.434	0.510
RNAalifold	0.530	0.444	0.475

Table S28: IPknot without refinement on the Rfam-PK dataset (single).

$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC	$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC
1	1	0.576	0.404	0.469	8	1	0.480	0.487	0.474
1	2	0.573	0.405	0.469	8	2	0.480	0.506	0.485
1	4	0.562	0.417	0.473	8	4	0.481	0.517	0.490
1	8	0.551	0.434	0.479	8	8	0.478	0.516	0.489
1	16	0.545	0.442	0.481	8	16	0.471	0.525	0.490
2	1	0.555	0.460	0.494	16	1	0.468	0.491	0.470
2	2	0.556	0.464	0.496	16	2	0.467	0.514	0.482
2	4	0.545	0.483	0.503	16	4	0.467	0.525	0.487
2	8	0.533	0.503	0.509	16	8	0.464	0.526	0.486
2	16	0.526	0.513	0.511	16	16	0.463	0.526	0.485
4	1	0.509	0.477	0.482					
4	2	0.512	0.492	0.492					
4	4	0.509	0.496	0.493					
4	8	0.498	0.515	0.498					
4	16	0.489	0.523	0.498					

Table S29: IPknot with refinement on the Rfam-PK dataset (single).

$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC	$\gamma^{(1)}$	$\gamma^{(2)}$	PPV	Sen	MCC
1	1	0.569	0.566	0.560	8	1	0.463	0.529	0.487
1	2	0.551	0.577	0.556	8	2	0.462	0.531	0.487
1	4	0.533	0.577	0.547	8	4	0.465	0.534	0.490
1	8	0.511	0.574	0.533	8	8	0.463	0.529	0.487
1	16	0.497	0.574	0.526	8	16	0.455	0.533	0.484
2	1	0.522	0.556	0.532	16	1	0.452	0.528	0.481
2	2	0.520	0.555	0.530	16	2	0.451	0.532	0.482
2	4	0.506	0.556	0.523	16	4	0.457	0.536	0.487
2	8	0.490	0.550	0.511	16	8	0.453	0.531	0.483
2	16	0.481	0.553	0.508	16	16	0.452	0.532	0.482
4	1	0.486	0.539	0.505					
4	2	0.487	0.541	0.506					
4	4	0.488	0.539	0.506					
4	8	0.476	0.539	0.499					
4	16	0.463	0.540	0.492					

Table S30: Comparison of IPknot with competitive methods on the Rfam-PK dataset (single).

	PPV	Sen	MCC
IPknot without refinement ($\gamma^{(1)} = 2, \gamma^{(2)} = 16$)	0.526	0.513	0.511
IPknot with refinement ($\gamma^{(1)} = 1, \gamma^{(2)} = 1$)	0.569	0.566	0.560
ProbKnot	0.457	0.471	0.456
FlexStem	0.519	0.594	0.548
pknotsRG	0.534	0.561	0.541
ILM	0.484	0.583	0.524