Benasque RNA 2018

Capturing the Landscapes of RNA 2D Structures by Decomposition of Partition Functions

Kiyoshi Asai

Dept. Comp. Bio. and Medical Sciences, University of Tokyo
Artificial Intelligence Research Center（AIRC）, AIST

This talk in mainly depends on

1. Ryota Mori et al., Efficient calculation of exact probability distributions of integer features on RNA secondary structures. BMC genomics 2014,15 Suppl 10: 12.
2. Taichi Hagio et al., Capturing alternative secondary structures of RNA by decomposition of base-pairing probabilities. *BMC bioinformatics* 2018, 19(Suppl 1):38

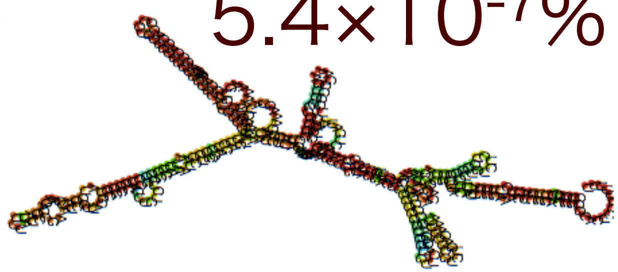
ANY 2D structure has very small probability

$$P(\sigma \mid x) = \frac{1}{Z(x)} \exp \frac{-E(\sigma, x)}{kT}$$

x : sequence

σ : 2D structure

$5.4 \times 10^{-7}\%$

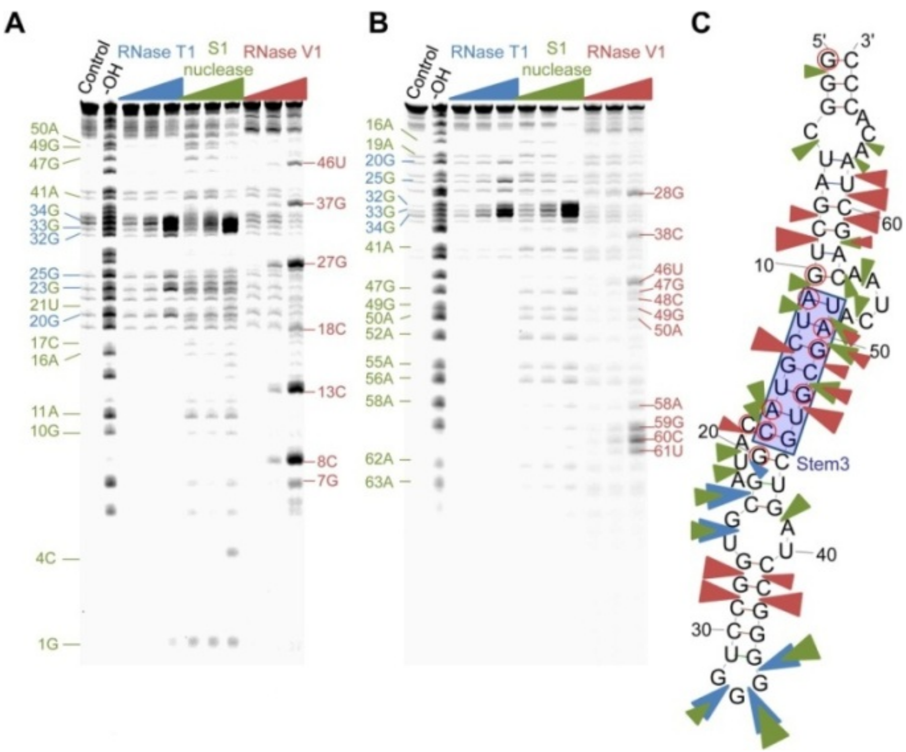


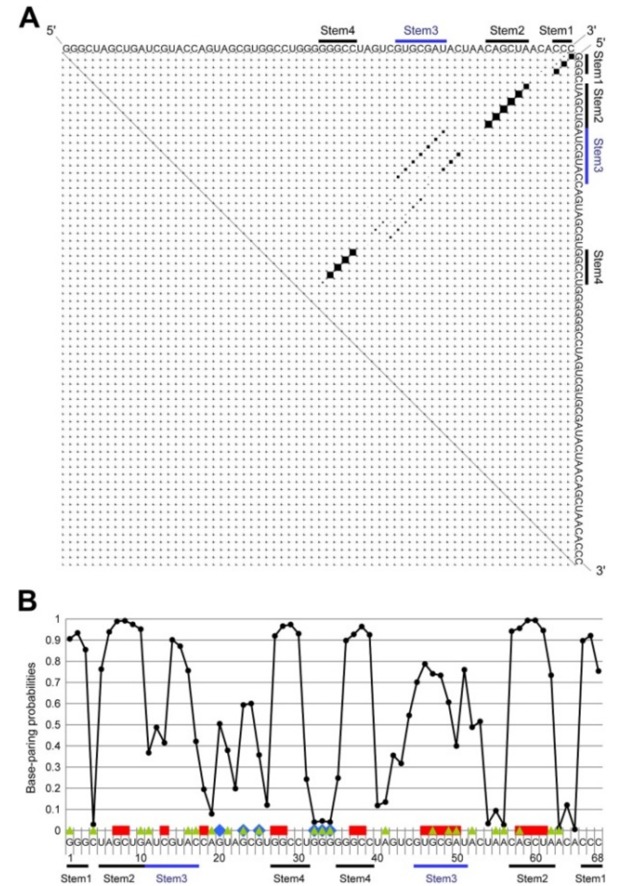
# of possible 2D. structures

$$\approx \frac{(2|x|)!}{(|x|+1)!|x|!}$$

Base Pairing Probabilities representing alternative strcutures

$$P^{(bp)}(i,j) \equiv P((i,j) \in \sigma^* \mid x) = \sum_{\sigma \mid (i,j) \in \sigma} P(\sigma \mid x)$$

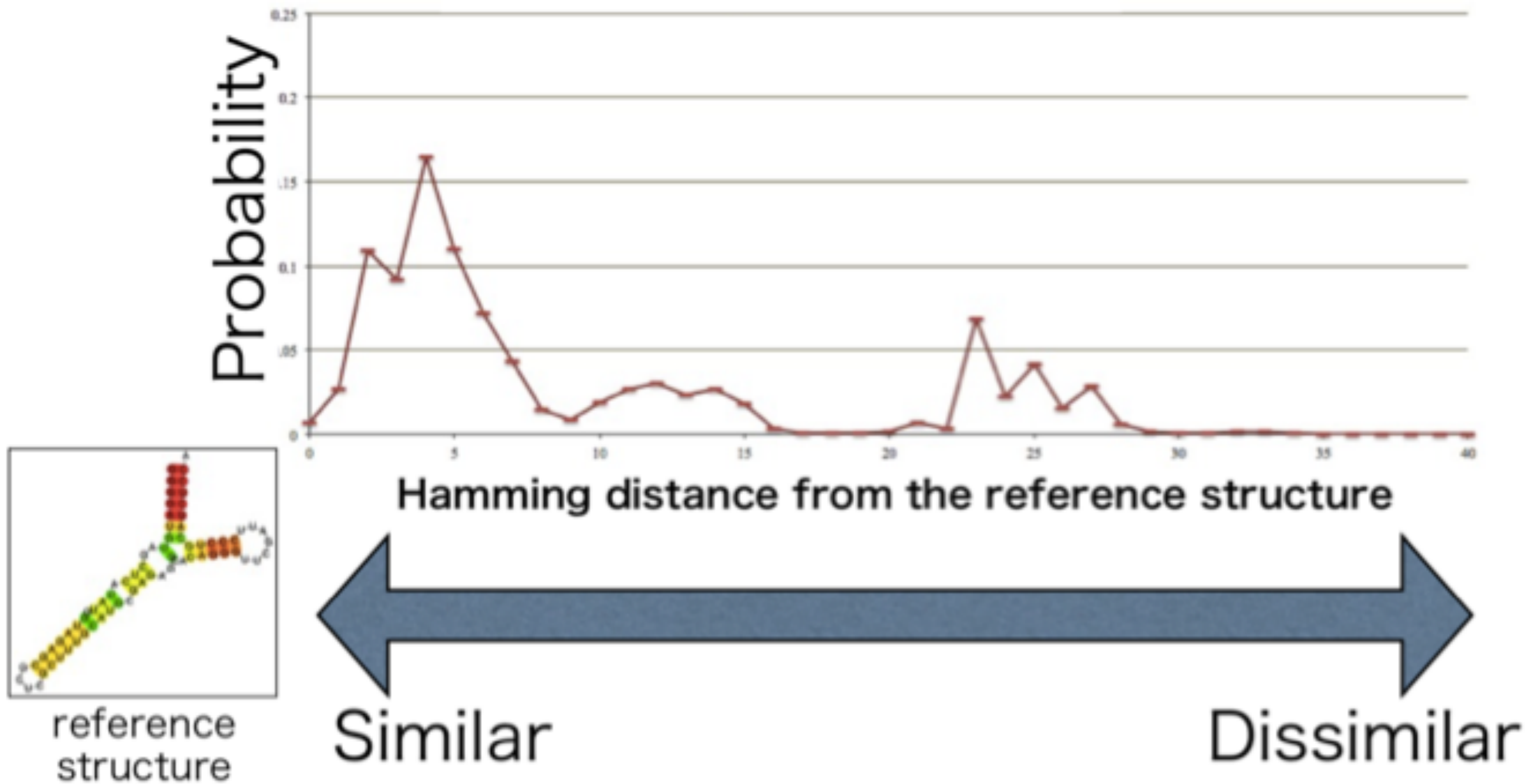




Adachi H et al., Biochimie, 93(7):1081-8(2011).

1D
landscape

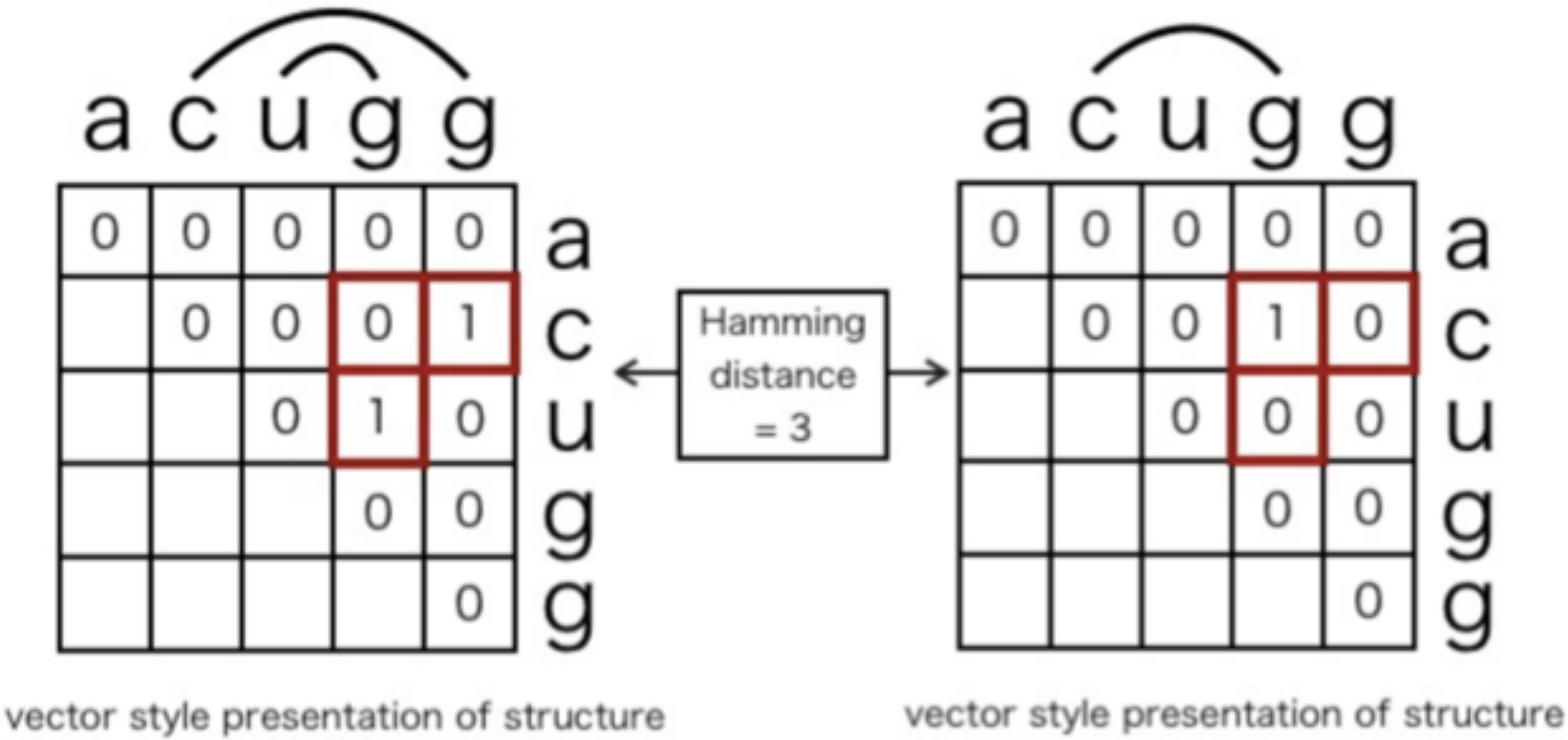
Probability on Hamming distance from a reference structure



RNAbor: Freyhult E et al., Bioinformatics. 2007; 23(16):2054–62.

RintD: Mori R et al., BMC genomics 2014, 15 Suppl 10: 12

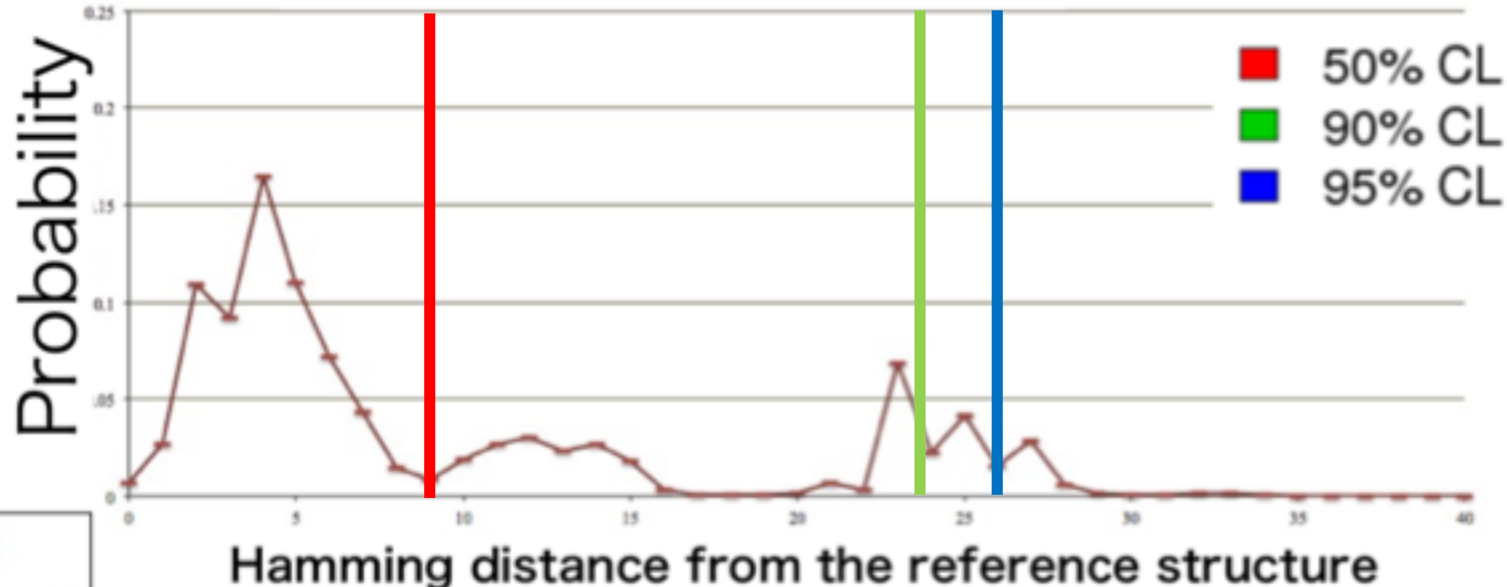
Definition of Hamming distance of 2D structures

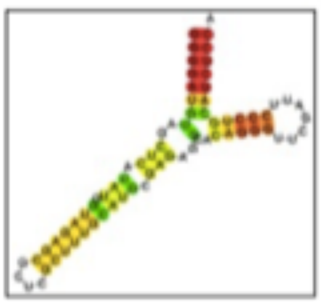


RintD: Mori R et al., BMC genomics 2014, 15 Suppl 10: 12

1D landscape

Credibility limit of a 2D structure

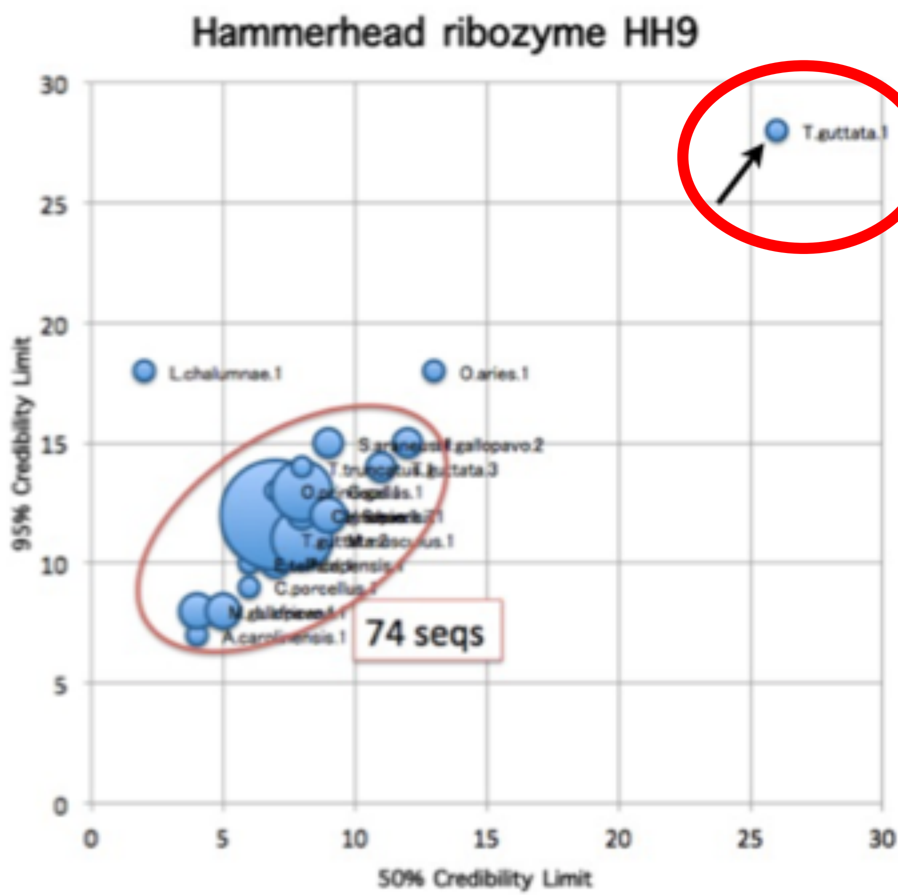


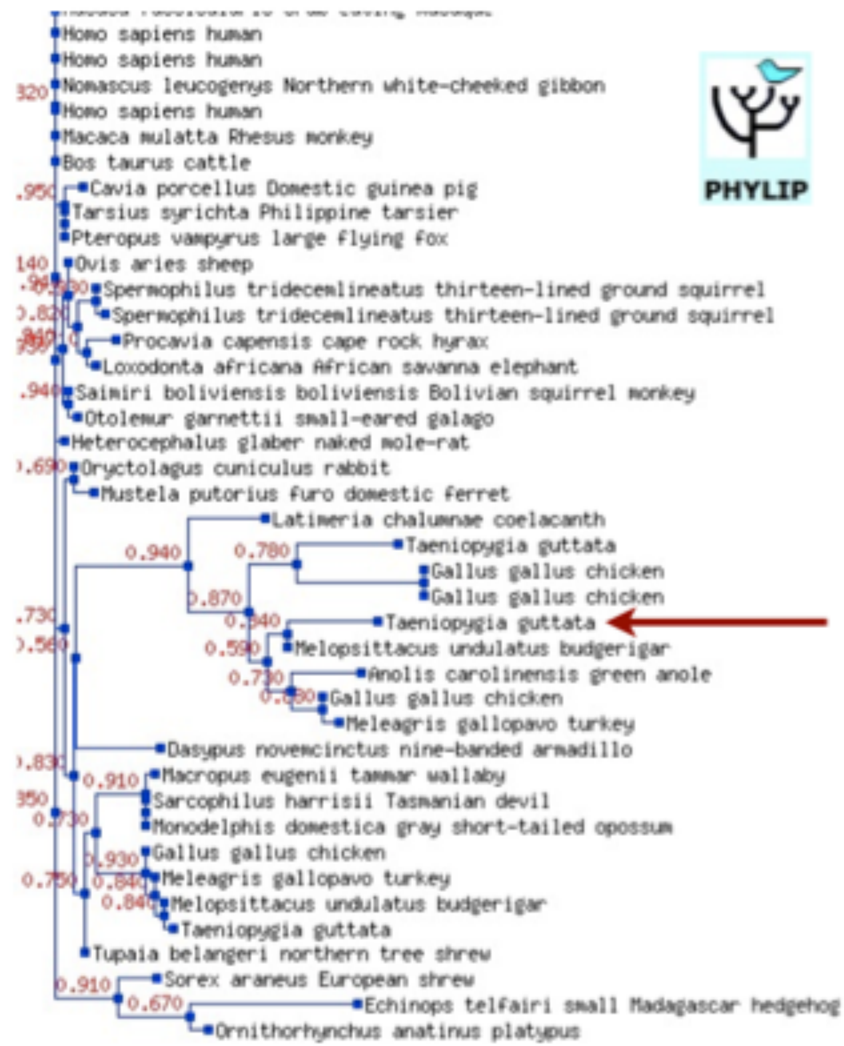


reference structure

RintD: Mori R et al., BMC genomics 2014, 15 Suppl 10: 12

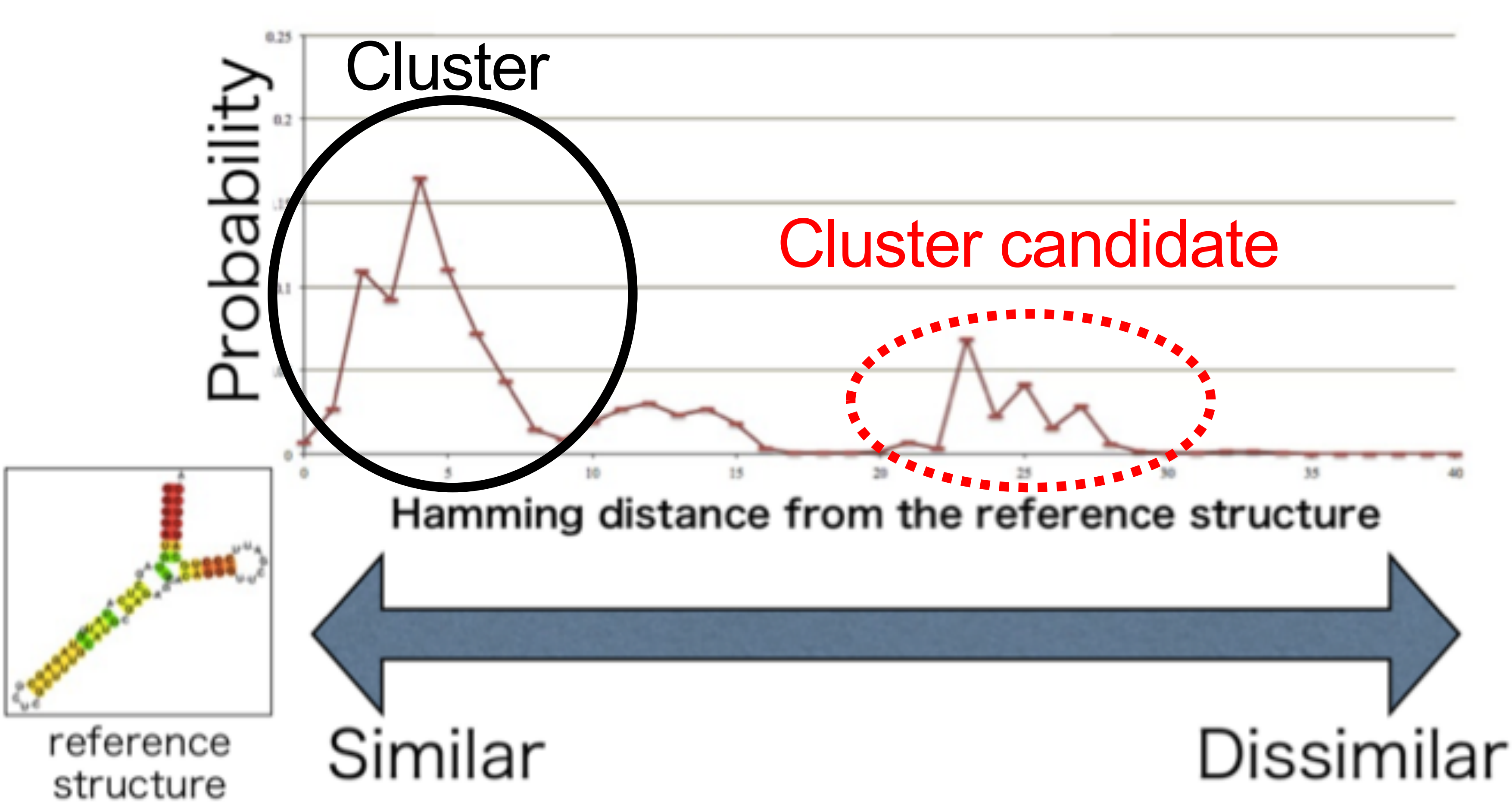
Credibility limits found a wrong classification





1D landscape

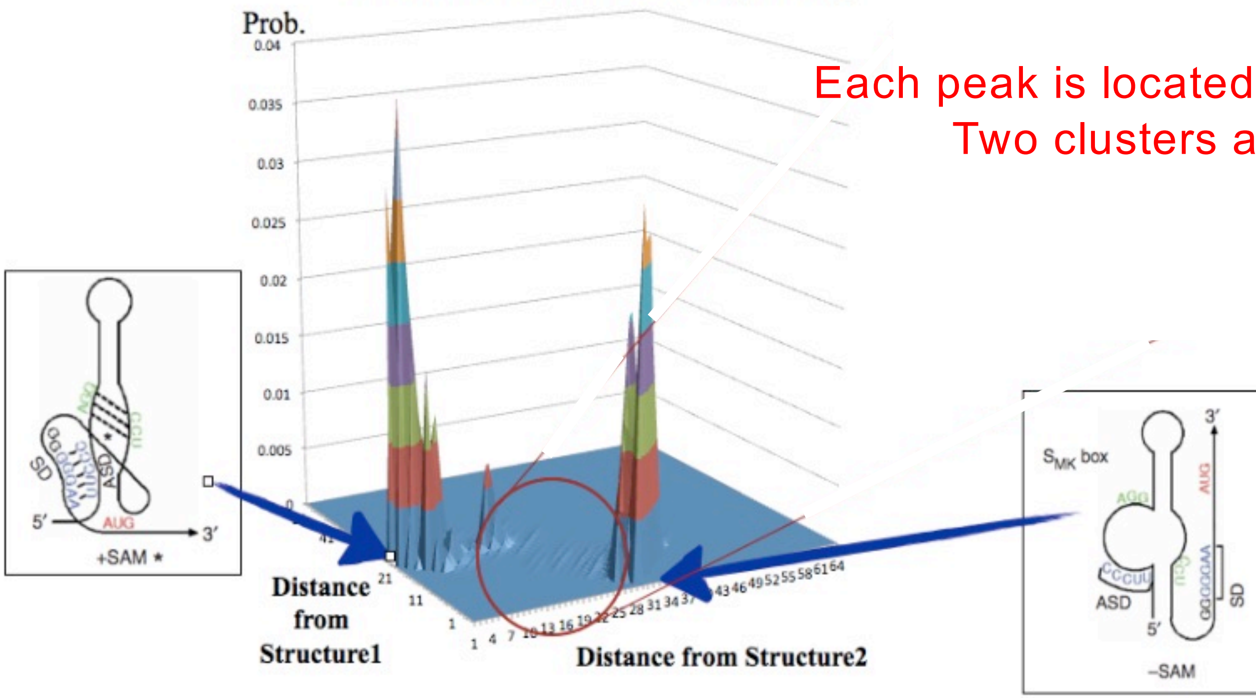
Probability over Hamming distance from a reference structure



RintD: Mori R et al., BMC genomics 2014, 15 Suppl 10: 12

2D landscape

Probability over Hamming distances from two reference structures



Each peak is located near x or y axe. Two clusters are found.

RNA2Dfold: Vienna RNA Package 2.0. Algorithms Mol Biol. 2011; 6:26

RintD: Mori R et al., BMC genomics 2014, 15 Suppl 10: 12

Dynamic Programming for Partition Function

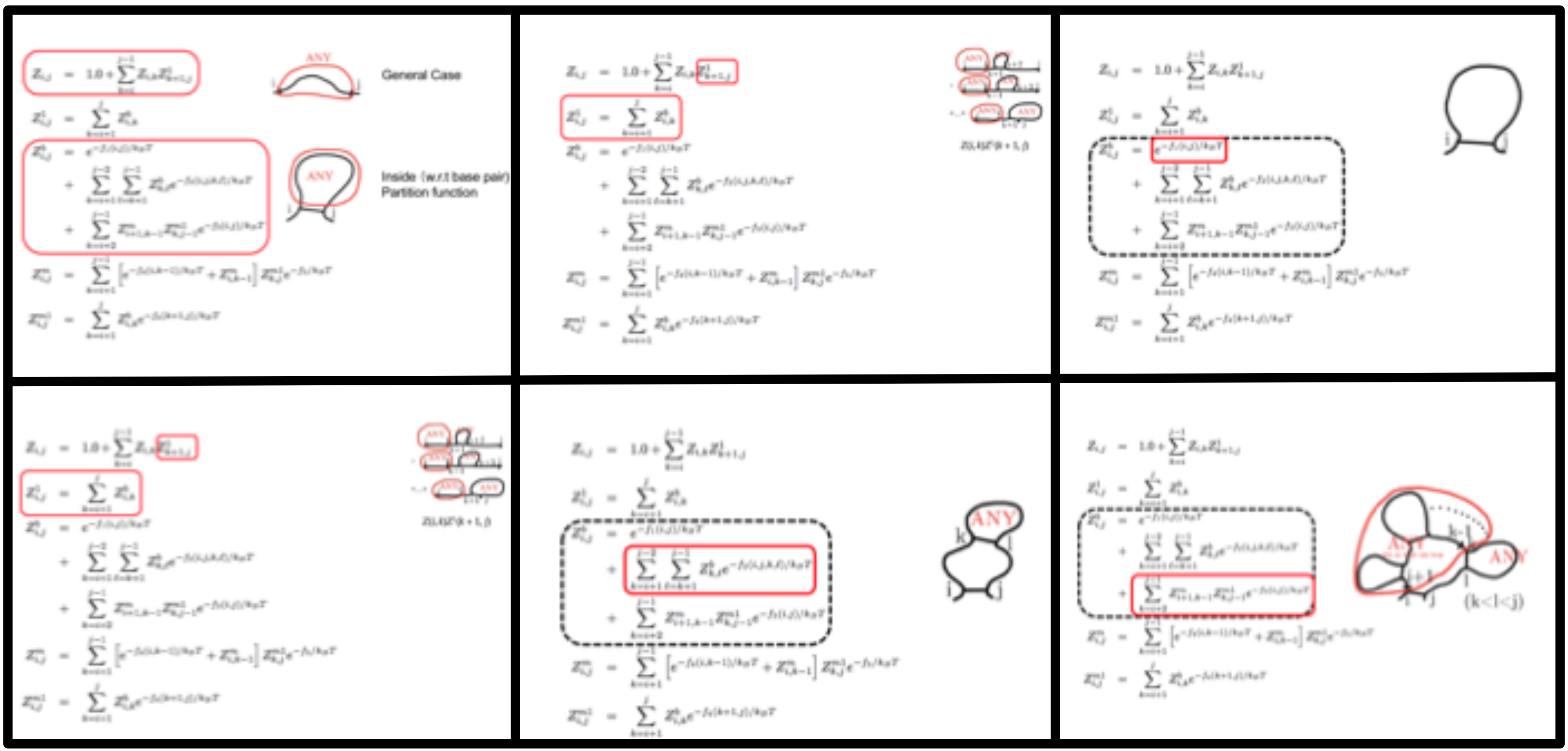
McCaskill (1990), Biopolymers. 1990, 29 (6-7): 1105-19

$$
\begin{aligned}
Z_{i,j} &= 1.0 + \sum_{k=i}^{j-1} Z_{i,k} Z^{1}_{k+1,j} \\
Z^{1}_{i,j} &= \sum_{k=i+1}^{j} Z^{b}_{i,k} \\
Z^{b}_{i,j} &= e^{-f_1(i,j)/k_B T} \\
&+ \sum_{k=i+1}^{j-2} \sum_{\ell=k+1}^{j-1} Z^{b}_{k,\ell} e^{-f_2(i,j,k,\ell)/k_B T} \\
&+ \sum_{k=i+2}^{j-1} Z^{m}_{i+1,k-1} Z^{m1}_{k,j-1} e^{-f_3(i,j)/k_B T} \\
Z^{m}_{i,j} &= \sum_{k=i+1}^{j-1} \left[e^{-f_4(i,k-1)/k_B T} + Z^{m}_{i,k-1} \right] Z^{m1}_{k,j} e^{-f_5/k_B T} \\
Z^{m1}_{i,j} &= \sum_{k=i+1}^{j} Z^{b}_{i,k} e^{-f_4(k+1,j)/k_B T}
\end{aligned}
$$

Dynamic Programming for Partition Function

McCaskill (1990), Biopolymers. 1990, 29 (6-7): 1105-19

McCaskill Algorithm corresponds to INSIDE Algorithm of SCFG



DP on polynomials for partition function over Hamming distance

$$Z_{i,j} = x^{g_1^Z(i,j)} + \sum_{h=i}^{j-1} Z_{i,h} Z^1_{h+1,j} x^{g_2^Z(i,j,h)}$$

$$Z^1_{i,j} = \sum_{h=i+1}^{j} Z^b_{i,h} x^{g_3^Z(i,j,h)}$$

$$Z^b_{i,j} = e^{-f_1(i,j)/kT} x^{g_4^Z(i,j)} + \sum_{h=i+1}^{j-2} \sum_{\ell=h}^{j-1} Z^b_{h,\ell} e^{-f_2(i,h,\ell,j)/kT} x^{g_5^Z(i,h,\ell,j)} + \sum_{h=i+2}^{j-1} Z^m_{i+1,h-1} Z^{m1}_{h,j-1} e^{-f_3(i,j)/kT} x^{g_6^Z(i,j,h)}$$

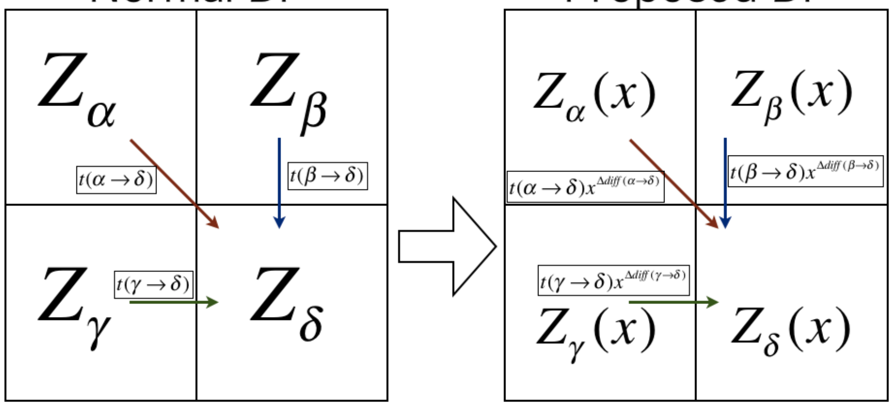
$$Z^m_{i,j} = \sum_{h=i}^{j-1} \left[\begin{array}{l} e^{-f_4(i,h-1)/kT} x^{g_7^Z(i,j,h)} \\ + Z^m_{i,h-1} x^{g_8^Z(i,j,h)} \end{array} \right] Z^{m1}_{h,j} e^{-f_5/kT}$$

$$Z^{m1}_{i,j} = \sum_{h=i+1}^{j} Z^b_{i,h} e^{-f_4(h+1,j)/kT} x^{g_3^Z(i,j,h)}$$

Mori et al, BMC genomics, 2014

Calculating partition function over Hamming distance

Normal DP　　　　Proposed DP



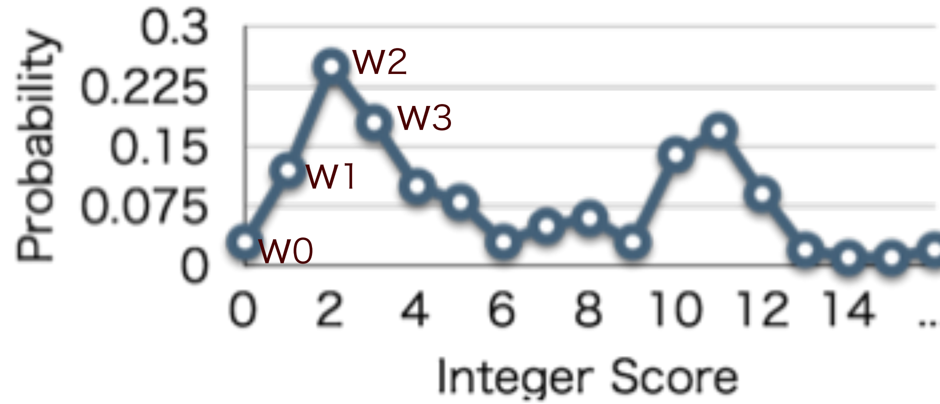
DP matrix　　　　DP matrix

$$Z_\delta = Z_\alpha \cdot t(\alpha \to \delta) + Z_\beta \cdot t(\beta \to \delta) + Z_\gamma \cdot t(\gamma \to \delta)$$

$Z_\bullet : scalar$

$$Z_\delta(x) = Z_\alpha(x) \cdot t(\alpha \to \delta)x^{\Delta diff(\alpha \to \delta)} + Z_\beta(x) \cdot t(\beta \to \delta)x^{\Delta diff(\beta \to \delta)} + Z_\gamma(x) \cdot t(\gamma \to \delta)x^{\Delta diff(\gamma \to \delta)}$$

$Z_\bullet(x) : polynomial$



$$Z_n = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + \cdots$$

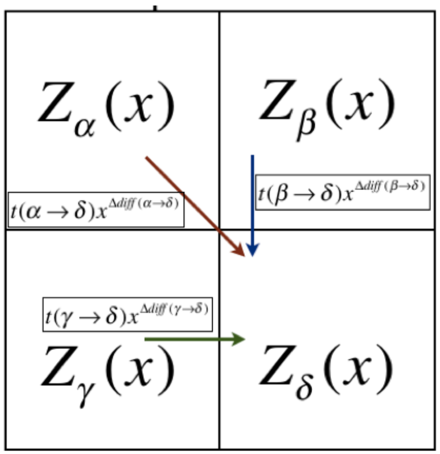
Newberg LA, Lawrence CE.. J Comput Biol. 2009; 16(1):1–18.

Calculating partition function over Hamming distance

$$s(\alpha \rightarrow \delta) = 1 \qquad Z_\alpha(x) = w_{\alpha 0} + w_{\alpha 1}x + w_{\alpha 2}x^2 + w_{\alpha 3}x^3 + \cdots$$

$$s(\beta \rightarrow \delta) = 0 \qquad Z_\beta(x) = w_{\beta 0} + w_{\beta 1}x + w_{\beta 2}x^2 + w_{\beta 3}x^3 + \cdots$$

$$s(\gamma \rightarrow \delta) = 2 \qquad Z_\gamma(x) = w_{\gamma 0} + w_{\gamma 1}x + w_{\gamma 2}x^2 + w_{\gamma 3}x^3 + \cdots$$



$$Z_\delta(x) = Z_\alpha(x)t(\alpha \rightarrow \delta)x^1 + Z_\beta(x)t(\beta \rightarrow \delta)x^0 + Z_\gamma(x)t(\gamma \rightarrow \delta)x^2$$

$$w_{\alpha 0}t(\alpha \rightarrow \delta)x + w_{\alpha 1}t(\alpha \rightarrow \delta)x^2 + w_{\alpha 2}t(\alpha \rightarrow \delta)x^3 + w_{\alpha 3}t(\alpha \rightarrow \delta)x^4 + \cdots$$

$$w_{\beta 0}t(\beta \rightarrow \delta) + w_{\beta 1}t(\beta \rightarrow \delta)x + w_{\beta 2}t(\beta \rightarrow \delta)x^2 + w_{\beta 3}t(\beta \rightarrow \delta)x^3 + \cdots$$

$$+)\quad w_{\gamma 0}t(\gamma \rightarrow \delta)x^2 + w_{\gamma 1}t(\gamma \rightarrow \delta)x^3 + w_{\gamma 2}t(\gamma \rightarrow \delta)x^4 + \cdots$$

$$Z_\delta(x) = w_{\beta 0}t(\beta \rightarrow \delta) + \begin{pmatrix} w_{\alpha 0}t(\alpha \rightarrow \delta) \\ + \\ w_{\beta 1}t(\beta \rightarrow \delta) \end{pmatrix} x + \begin{pmatrix} w_{\alpha 1}t(\alpha \rightarrow \delta) \\ + \\ w_{\beta 2}t(\beta \rightarrow \delta) \\ + \\ w_{\gamma 0}t(\gamma \rightarrow \delta) \end{pmatrix} x^2 + \cdots$$

Newberg LA, Lawrence CE.. J Comput Biol. 2009; 16(1):1–18.

DP on polynomials for partition function over Hamming distance

$$Z_{i,j} = x^{g_1^Z(i,j)} + \sum_{h=i}^{j-1} Z_{i,h} Z^1_{h+1,j} x^{g_2^Z(i,j,h)}$$

$$Z^1_{i,j} = \sum_{h=i+1}^{j} Z^b_{i,h} x^{g_3^Z(i,j,h)}$$

$$Z^b_{i,j} = e^{-f_1(i,j)/kT} x^{g_4^Z(i,j)} + \sum_{h=i+1}^{j-2} \sum_{\ell=h}^{j-1} Z^b_{h,\ell} e^{-f_2(i,h,\ell,j)/kT} x^{g_5^Z(i,h,\ell,j)} + \sum_{h=i+2}^{j-1} Z^m_{i+1,h-1} Z^{m1}_{h,j-1} e^{-f_3(i,j)/kT} x^{g_6^Z(i,j,h)}$$

$$Z^m_{i,j} = \sum_{h=i}^{j-1} \left[e^{-f_4(i,h-1)/kT} x^{g_7^Z(i,j,h)} + Z^m_{i,h-1} x^{g_8^Z(i,j,h)} \right] Z^{m1}_{h,j} e^{-f_5/kT}$$

$$Z^{m1}_{i,j} = \sum_{h=i+1}^{j} Z^b_{i,h} e^{-f_4(h+1,j)/kT} x^{g_3^Z(i,j,h)}$$

$$g_0(i,j,h) = \sum_{p=i}^{h} \sum_{q=h+1}^{j} \sigma_{p,q},$$

$$g_0^Z(i,j) = \sum_{p=i}^{j-1} \sum_{q=p+1}^{j} \sigma_{p,q},$$

the gain functions are described as follows:

$$\begin{aligned}
g_1^Z(i,j) &= g_0^Z(i,j),\\
g_2^Z(i,j,h) &= g_0(i,j,h),\\
g_3^Z(i,j,h) &= g_0(i,j,h+1) + g_0^Z(h+1,j),\\
g_4^Z(i,j) &= g_0^Z(i,j) + 1 - 2\sigma_{p,q},\\
g_5^Z(i,j,h,\ell) &= g_0^Z(i,j) - g_0^Z(h,\ell) + 1 - 2\sigma_{p,q},\\
g_6^Z(i,j,h) &= g_0^Z(i,j) - g_0^Z(i+1,h-1) - g_0^Z(h,j-1) + 1 - 2\sigma_{p,q},\\
g_7^Z(i,j,h) &= g_0^Z(i,h-1) + g_0(i,j,h),\\
g_8^Z(i,j,h) &= g_0(i,j,h-1).
\end{aligned}$$

Mori et al, BMC genomics, 2014

DP on polynomials
for partition function over Hamming distance

$$Z_{i,j} = x^{g_1^Z(i,j)} + \sum_{h=i}^{j-1} Z_{i,h} Z^1_{h+1,j} x^{g_2^Z(i,j,h)}$$

$$Z^1_{i,j} = \sum_{h=i+1}^{j} Z^b_{i,h} x^{g_3^Z(i,j,h)}$$

$$Z^b_{i,j} = e^{-f_1(i,j)/kT} x^{g_4^Z(i,j)} + \sum_{h=i+1}^{j-2} \sum_{\ell=h}^{j-1} Z^b_{h,\ell} e^{-f_2(i,h,\ell,j)/kT} x^{g_5^Z(i,h,\ell,j)} + \sum_{h=i+2}^{j-1} Z^m_{i+1,h-1} Z^{m1}_{h,j-1} e^{-f_3(i,j)/kT} x^{g_6^Z(i,j,h)}$$

$$Z^m_{i,j} = \sum_{h=i}^{j-1} \left[e^{-f_4(i,h-1)/kT} x^{g_7^Z(i,j,h)} + Z^m_{i,h-1} x^{g_8^Z(i,j,h)} \right] Z^{m1}_{h,j} e^{-f_5/kT}$$

$$Z^{m1}_{i,j} = \sum_{h=i+1}^{j} Z^b_{i,h} e^{-f_4(h+1,j)/kT} x^{g_3^Z(i,j,h)}$$

Product of 2-dim DP matrices
$O(L^4)$

Mori et al, BMC genomics, 2014

Calculating partition function over Hamming distance

$$\Delta = d_{max} - d_{min} + 1$$

$$Z(d,\sigma) = \sum_{s=d_{min}}^{d_{max}} Z(s,\sigma)\delta_{sd}$$

$$= \sum_{s=d_{min}}^{d_{max}} Z(s,\sigma) \sum_{r=d_{min}}^{d_{max}} \frac{\exp\left[2\pi i \frac{r(s-d)}{\Delta}\right]}{\Delta}$$

$$= \frac{1}{\Delta} \sum_{r=d_{min}}^{d_{max}} \exp\left[2\pi i \frac{-rd}{\Delta}\right] \sum_{s=d_{min}}^{d_{max}} Z(s,\sigma) \exp\left[2\pi i \frac{rs}{\Delta}\right]$$

$$= \frac{1}{\Delta} \sum_{r=d_{min}}^{d_{max}} \exp\left[2\pi i \frac{-rd}{\Delta}\right] \tilde{Z}_{1,L}(\sigma).$$

DP on complex numbers

$$x = \exp\left[2\pi i \frac{r}{\Delta}\right]$$

Calculating partition function over Hamming distance

$$
\begin{aligned}
Z(d,\sigma) &= \sum_{s=d_{\min}}^{d_{\max}} Z(s,\sigma)\delta_{sd} \\
&= \sum_{s=d_{\min}}^{d_{\max}} Z(s,\sigma) \sum_{r=d_{\min}}^{d_{\max}} \frac{\exp\left[2\pi i \frac{r(s-d)}{\Delta}\right]}{\Delta} \\
&= \frac{1}{\Delta} \sum_{r=d_{\min}}^{d_{\max}} \exp\left[2\pi i \frac{-rd}{\Delta}\right] \sum_{s=d_{\min}}^{d_{\max}} Z(s,\sigma) \exp\left[2\pi i \frac{rs}{\Delta}\right] \\
&= \frac{1}{\Delta} \boxed{\sum_{r=d_{\min}}^{d_{\max}} \exp\left[2\pi i \frac{-rd}{\Delta}\right] \tilde{Z}_{1,L}(\sigma).}
\end{aligned}
$$

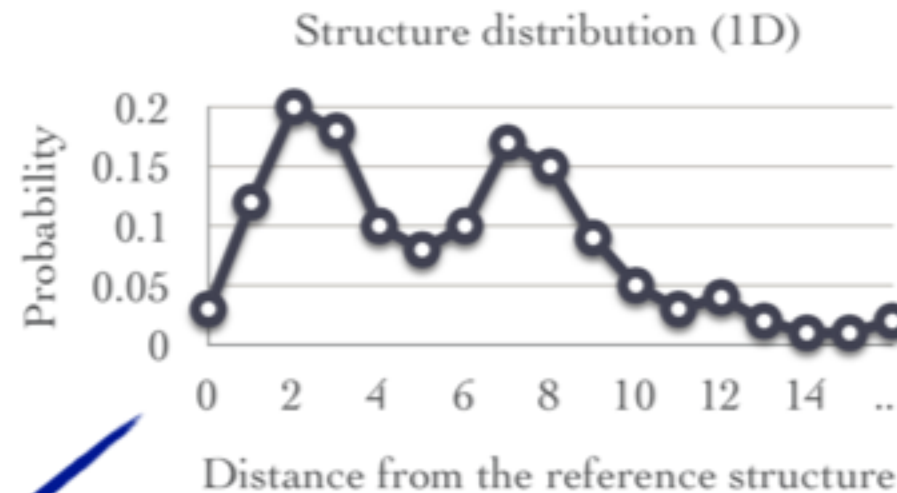
DFT

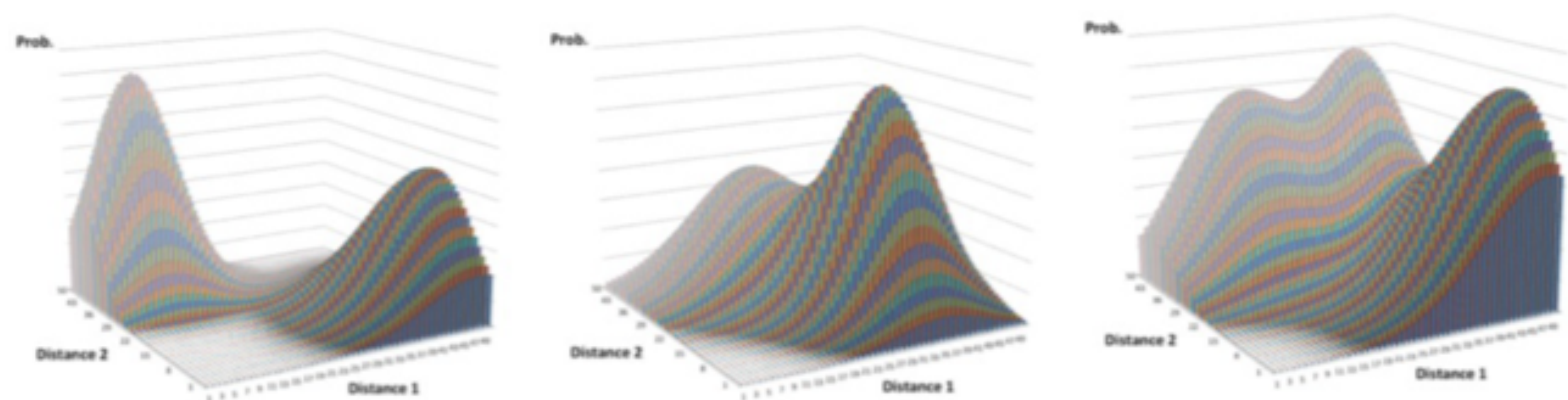
Hamming distances from two ref. structures

$$p_l = \exp\left(2\pi i \frac{l}{d_{p\max}+1}\right), q_m = \exp\left(2\pi i \frac{m}{d_{q\max}+1}\right)$$

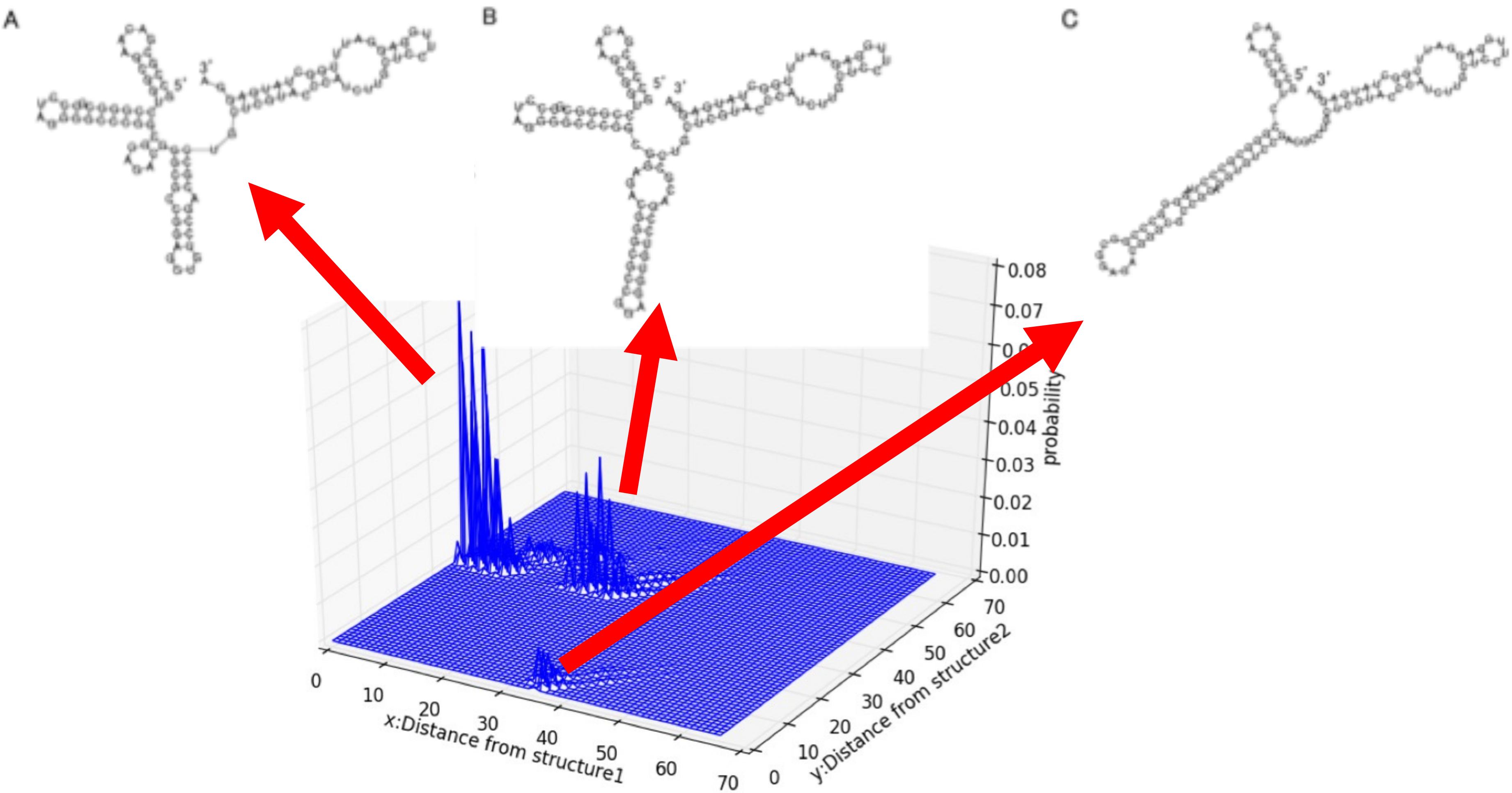
$$Z_{l,m}(p_l, q_m) = \sum_{d_p=0}^{d_{p\max}} \sum_{d_q=0}^{d_{q\max}} w_{d_p,d_q} p_l^{d_p} q_m^{d_q}$$

$$Z_{l,m} \xrightarrow{DFT} \sum_{d_q=0}^{d_{q\max}} w_{d_p,d_q} q_m^{d_q} \xrightarrow{DFT} w_{d_p,d_q}$$





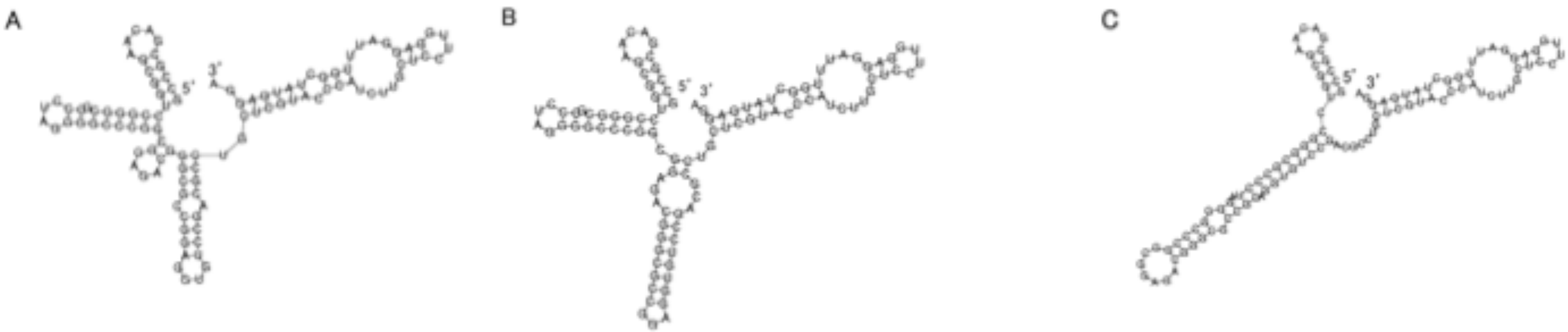
Representative 2D structures of the clusters



RNAborMEA: Clote P et al., BMC Bioinformatics. 2012; 13 Suppl 5:6.

RintW: Taichi Hagio et al., *BMC bioinformatics* 2018, 19(Suppl 1):38

Representative 2D structures of the clusters



Exact calculation

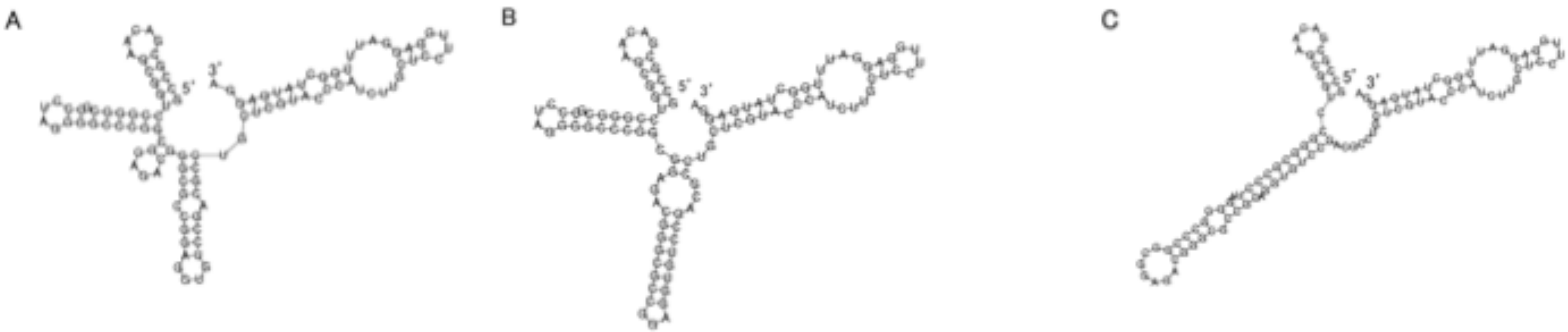
It looked easy by using Integer Programming.
Exact DP should be possible, suggested by Risa Kawaguchi in private discussion during Benasque RNA 2015

Based on sampling

RNAborMEA: Clote P et al., BMC Bioinformatics. 2012; 13 Suppl 5:6.

RintW: Taichi Hagio et al., *BMC bioinformatics* 2018, 19(Suppl 1):38

Representative 2D structures of the clusters



Based on Exact DP calculation of base-paring probabilities

Based on sampling

RNAborMEA: Clote P et al., BMC Bioinformatics. 2012; 13 Suppl 5:6.

RintW: Taichi Hagio et al., *BMC bioinformatics* 2018, 19(Suppl 1):38

Posterior decoding for Maximum Expected Accuracy

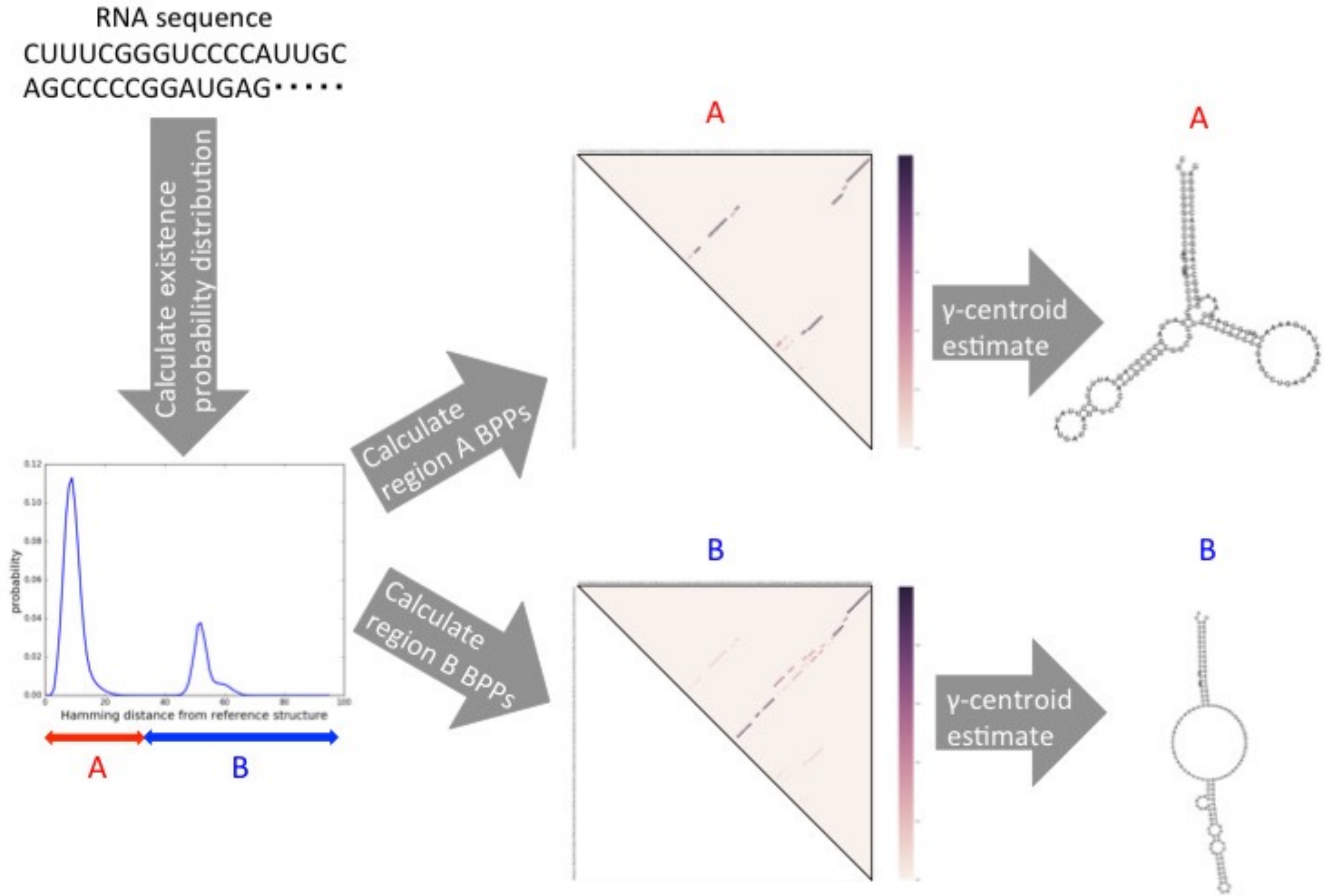
γ-centroid estimator

$$M_{i,j} = \max \begin{cases} M_{i+1,j-1} + (\gamma+1)P^{(bp)}_{i,j} - 1 \\ M_{i-1,k} \\ M_{i,k-1} \\ \max_k \left[M_{i,k} + M_{k+1,j} \right] \end{cases}$$

CentroidFold

Hamada et al. *Bioinformatics 25(4), 2009*

Overview of our method



McCaskill algorithm of BPPs

$$P_{hl} = \frac{Q_{1,h-1} Q^b_{hl} Q_{l+1,N}}{Q_{1N}}$$

$$+ \sum_{\substack{i,j \\ i<h<l<j}} P_{ij} \frac{Q^b_{hl}}{Q^b_{ij}} e^{-[F_2(i,j,h,l)/kT]}$$

$$+ \sum_{\substack{i,j \\ i<h<l<j}} P_{ij} \frac{Q^b_{hl}}{Q^b_{ij}} e^{-[(a+b)/kT]}$$

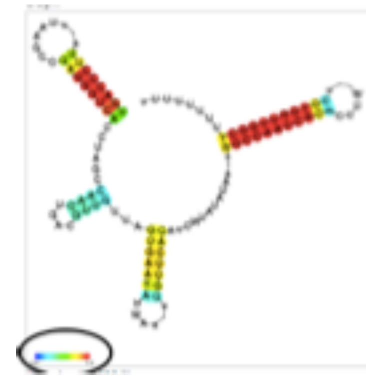
$$\times \left(e^{-[(h-i-1)c/kT]} Q^m_{l+1,j-1} + Q^m_{i+1,h-1} \right.$$

$$\left. \times e^{-[(j-l-1)c/kT]} + Q^m_{i+1,h-1} Q^m_{l+1,j-1} \right)$$

McCaskill（1990）, Biopolymers. 1990, 29 (6-7): 1105-19

Base Pairing Probability（BPP）

$$P^{(bp)}(i,j) \equiv P((i,j) \in \sigma^* \mid x) = \sum_{\sigma \mid (i,j) \in \sigma} P(\sigma \mid x)$$

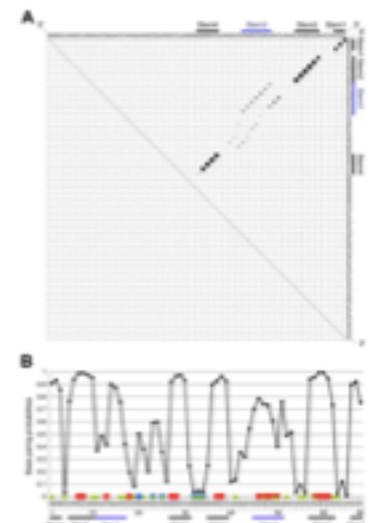


$$= \frac{Z^b_{ij} W^b_{ij}}{Z_{1,L}}$$

$Z_{1,L}$: Partition function of the whole sequence

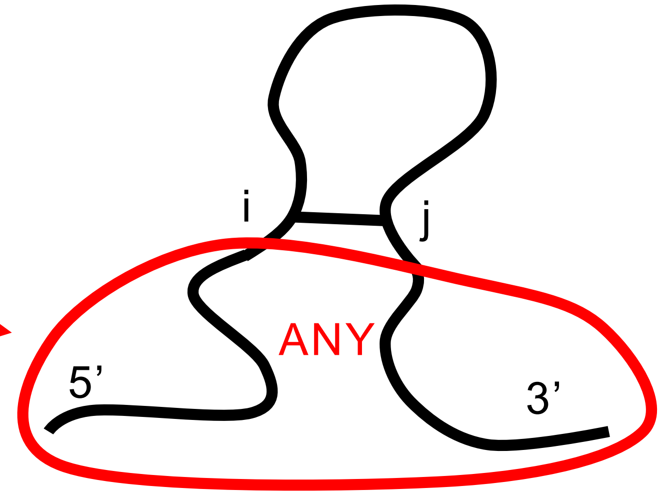
Z^b_{ij} : Inside partition function（w.r.t. (i,j) base-pair）

W^b_{ij} : Outside partition function（w.r.t. (i,j) base-pair）



Base Pairing Probability（BPP）

$$= \frac{Z^b_{ij} W^b_{ij}}{Z_{1,L}}$$



$Z_{1,L}$: Partition function of the whole sequence

Z^b_{ij} : Inside partition function（w.r.t. (i,j) base-pair）

W^b_{ij} : Outside partition function（w.r.t. (i,j) base-pair）

Dynamic Programming for Partition Function

McCaskill（1990）, Biopolymers. 1990, 29 (6-7)

$$Z_{i,j} = 1.0 + \sum_{k=i}^{j-1} Z_{i,k} Z^1_{k+1,j}$$

General Case

$$Z^1_{i,j} = \sum_{k=i+1}^{j} Z^b_{i,k}$$

$$\begin{aligned} Z^b_{i,j} &= e^{-f_1(i,j)/k_BT} \\ &+ \sum_{k=i+1}^{j-2} \sum_{\ell=k+1}^{j-1} Z^b_{k,\ell} e^{-f_2(i,j,k,\ell)/k_BT} \\ &+ \sum_{k=i+2}^{j-1} Z^m_{i+1,k-1} Z^{m1}_{k,j-1} e^{-f_3(i,j)/k_BT} \end{aligned}$$

Inside（w.r.t base pair）
Partition function

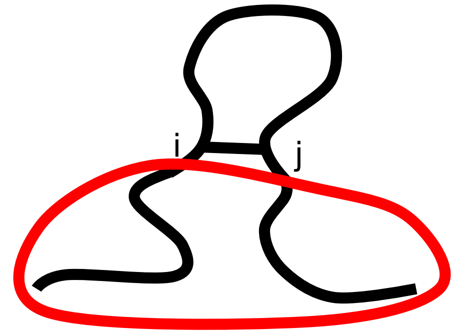
$$Z^m_{i,j} = \sum_{k=i+1}^{j-1} \left[e^{-f_4(i,k-1)/k_BT} + Z^m_{i,k-1} \right] Z^{m1}_{k,j} e^{-f_5/k_BT}$$

$$Z^{m1}_{i,j} = \sum_{k=i+1}^{j} Z^b_{i,k} e^{-f_4(k+1,j)/k_BT}$$

McCaskill Algorithm corresponds to INSIDE Algorithm of SCFG

DP algorithm of outside partition function

$$W^b_{i,j} = Z_{1,i-1} Z_{j+1,L} x^{g_1^{\mathrm{W}}(i,j)} + \sum_{\substack{k<i \\ \ell>j}} W^b_{k,\ell} e^{-f_2(k,i,j,\ell)/k_B T} x^{g_2^{\mathrm{W}}(k,i,j,\ell)} + \sum_{\substack{k<i \\ \ell>j}} W^b_{k,\ell} e^{-f_3/k_B T} \left[\begin{array}{l} Z^m_{k+1,i-1} e^{-f_4(j,\ell)/k_B T} x^{g_3^{\mathrm{W}}(k,i,j,\ell)} \\ + Z^m_{j+1,\ell-1} e^{-f_4(k,i)/k_B T} x^{g_4^{\mathrm{W}}(k,i,j,\ell)} \\ + Z^m_{k+1,j-1} Z^m_{j+1,\ell-1} x^{g_5^{\mathrm{W}}(k,i,j,\ell)} \end{array} \right.$$



This algorithm corresponds to outside algorithm of SCFG.
The DP includes inside partition function in the formulae.

Direct DP for BPP is presented in McCaskill's original paper,
but here we present explicit formula for outside partition function.

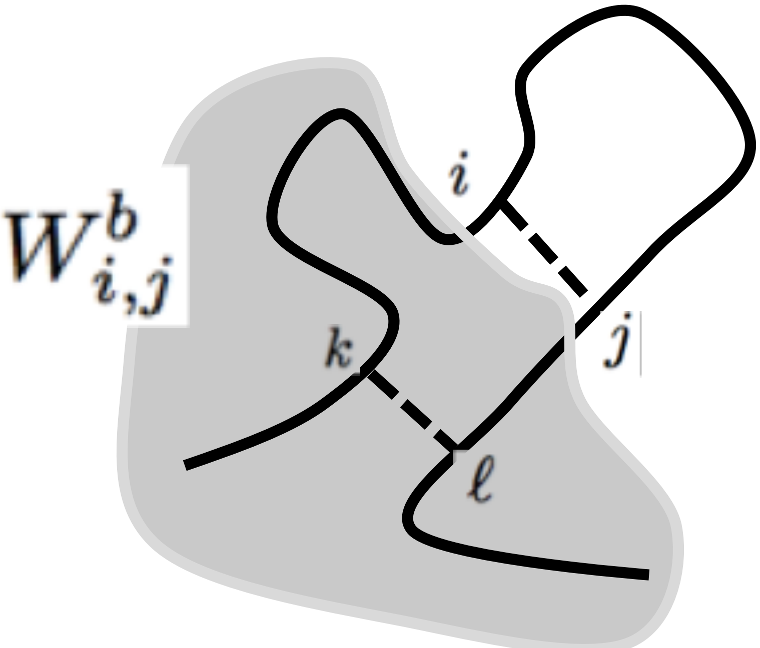
DP on polynomials for outside partition function over Hamming distance

$$
\begin{aligned}
W^b_{i,j} &= Z_{1,i-1} Z_{j+1,L} x^{g_1^{\mathrm{W}}(i,j)} \\
&+ \sum_{\substack{k<i \\ \ell>j}} W^b_{k,\ell} e^{-f_2(k,i,j,\ell)/k_B T} x^{g_2^{\mathrm{W}}(k,i,j,\ell)} \\
&+ \sum_{\substack{k<i \\ \ell>j}} W^b_{k,\ell} e^{-f_3/k_B T} \left[\begin{array}{l} Z^m_{k+1,i-1} e^{-f_4(j,\ell)/k_B T} x^{g_3^{\mathrm{W}}(k,i,j,\ell)} \\ + Z^m_{j+1,\ell-1} e^{-f_4(k,i)/k_B T} x^{g_4^{\mathrm{W}}(k,i,j,\ell)} \\ + Z^m_{k+1,j-1} Z^m_{j+1,\ell-1} x^{g_5^{\mathrm{W}}(k,i,j,\ell)} \end{array} \right.
\end{aligned}
$$

DP on polynomials for outside partition function over Hamming distance

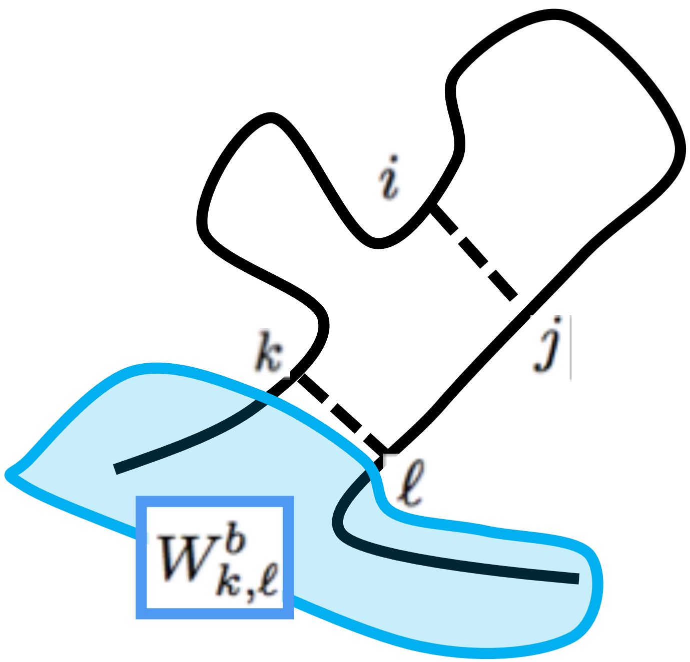
$$
\begin{aligned}
W^b_{i,j} &= Z_{1,i-1} Z_{j+1,L} x^{g_1^{\mathrm{W}}(i,j)} \\
&+ \sum_{\substack{k<i \\ \ell>j}} W^b_{k,\ell} e^{-f_2(k,i,j,\ell)/k_B T} x^{g_2^{\mathrm{W}}(k,i,j,\ell)} \\
&+ \sum_{\substack{k<i \\ \ell>j}} W^b_{k,\ell} e^{-f_3/k_B T} \Big[Z^m_{k+1,i-1} e^{-f_4(j,\ell)/k_B T} x^{g_3^{\mathrm{W}}(k,i,j,\ell)} \\
&\quad + Z^m_{j+1,\ell-1} e^{-f_4(k,i)/k_B T} x^{g_4^{\mathrm{W}}(k,i,j,\ell)} \\
&\quad + Z^m_{k+1,j-1} Z^m_{j+1,\ell-1} x^{g_5^{\mathrm{W}}(k,i,j,\ell)} \Big]
\end{aligned}
$$

DP on polynomials for outside partition function over Hamming distance



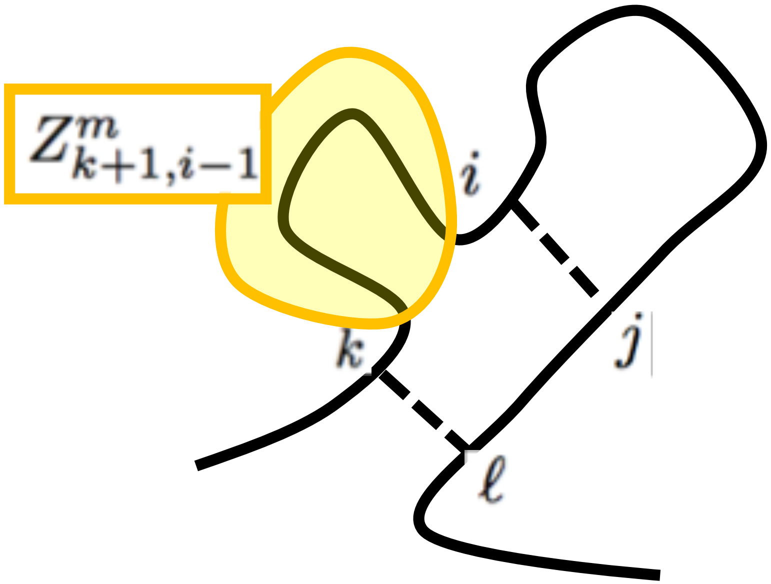
$$+ \sum_{\substack{k<i \\ \ell>j}} W^b_{k,\ell} e^{-f_3/k_B T} \Big[\; Z^m_{k+1,i-1} e^{-f_4(j,\ell)/k_B T} x^{g^W_3(k,i,j,\ell)}$$

DP on polynomials for outside partition function over Hamming distance



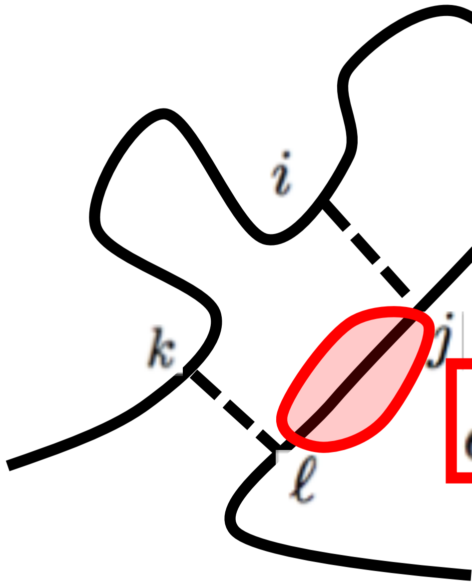
$$+ \sum_{\substack{k<i\\ \ell>j}} W^b_{k,\ell} e^{-f_3/k_BT} \left[Z^m_{k+1,i-1} e^{-f_4(j,\ell)/k_BT} x^{g_3^{\mathrm{W}}(k,i,j,\ell)} \right.$$

DP on polynomials for outside partition function over Hamming distance



$$+ \sum_{\substack{k<i \\ \ell>j}} W^b_{k,\ell} e^{-f_3/k_BT} \left[Z^m_{k+1,i-1} e^{-f_4(j,\ell)/k_BT} x^{g_3^{W}(k,i,j,\ell)} \right.$$

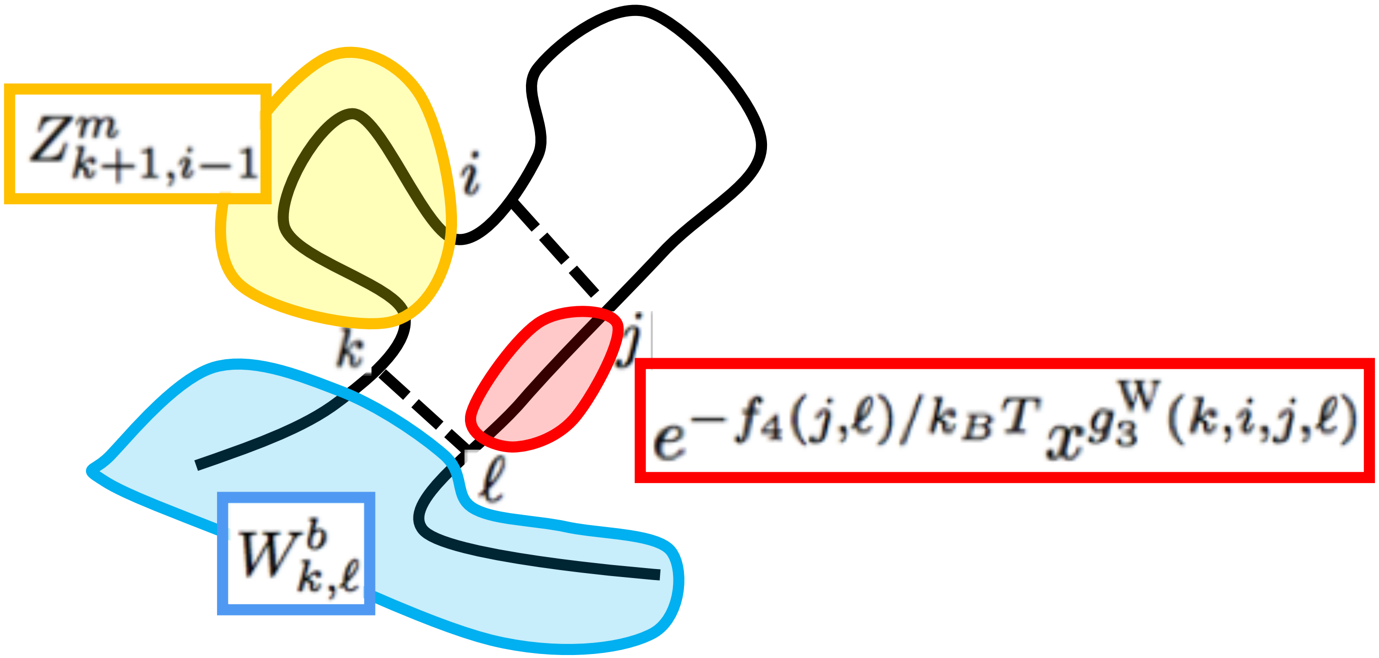
DP on polynomials for outside partition function over Hamming distance



$$e^{-f_4(j,\ell)/k_BT} x^{g_3^{\mathrm{W}}(k,i,j,\ell)}$$

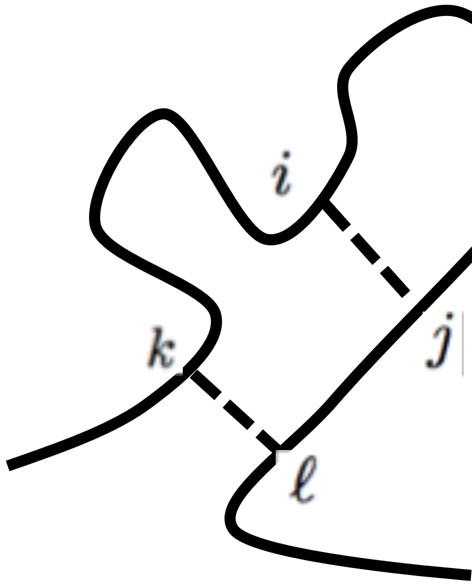
$$+ \sum_{\substack{k<i\\ \ell>j}} W^b_{k,\ell} e^{-f_3/k_BT} \Bigg[Z^m_{k+1,i-1} e^{-f_4(j,\ell)/k_BT} x^{g_3^{\mathrm{W}}(k,i,j,\ell)}$$

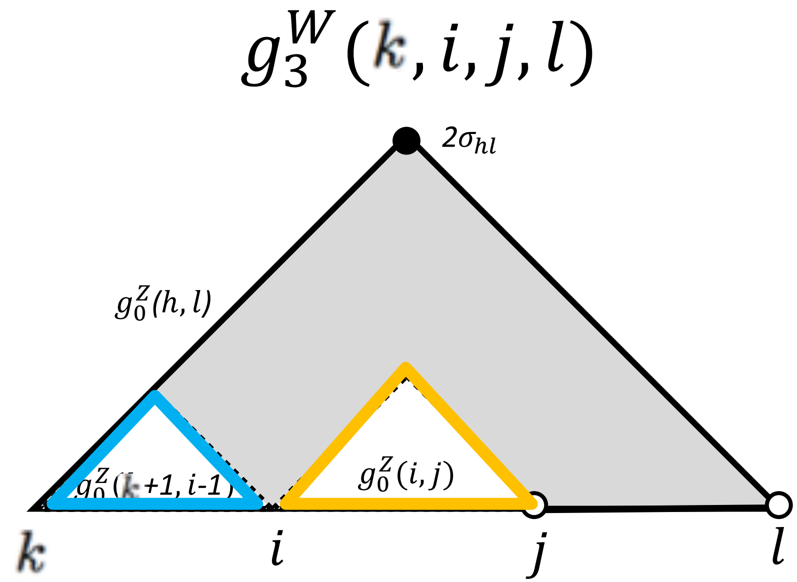
DP on polynomials for outside partition function over Hamming distance



$$+ \sum_{\substack{k < i \\ \ell > j}} W^{b}_{k,\ell} e^{-f_3/k_B T} \Big[Z^{m}_{k+1,i-1} e^{-f_4(j,\ell)/k_B T} x^{g_3^{\mathrm{W}}(k,i,j,\ell)}$$

DP on polynomials for outside partition function over Hamming distance



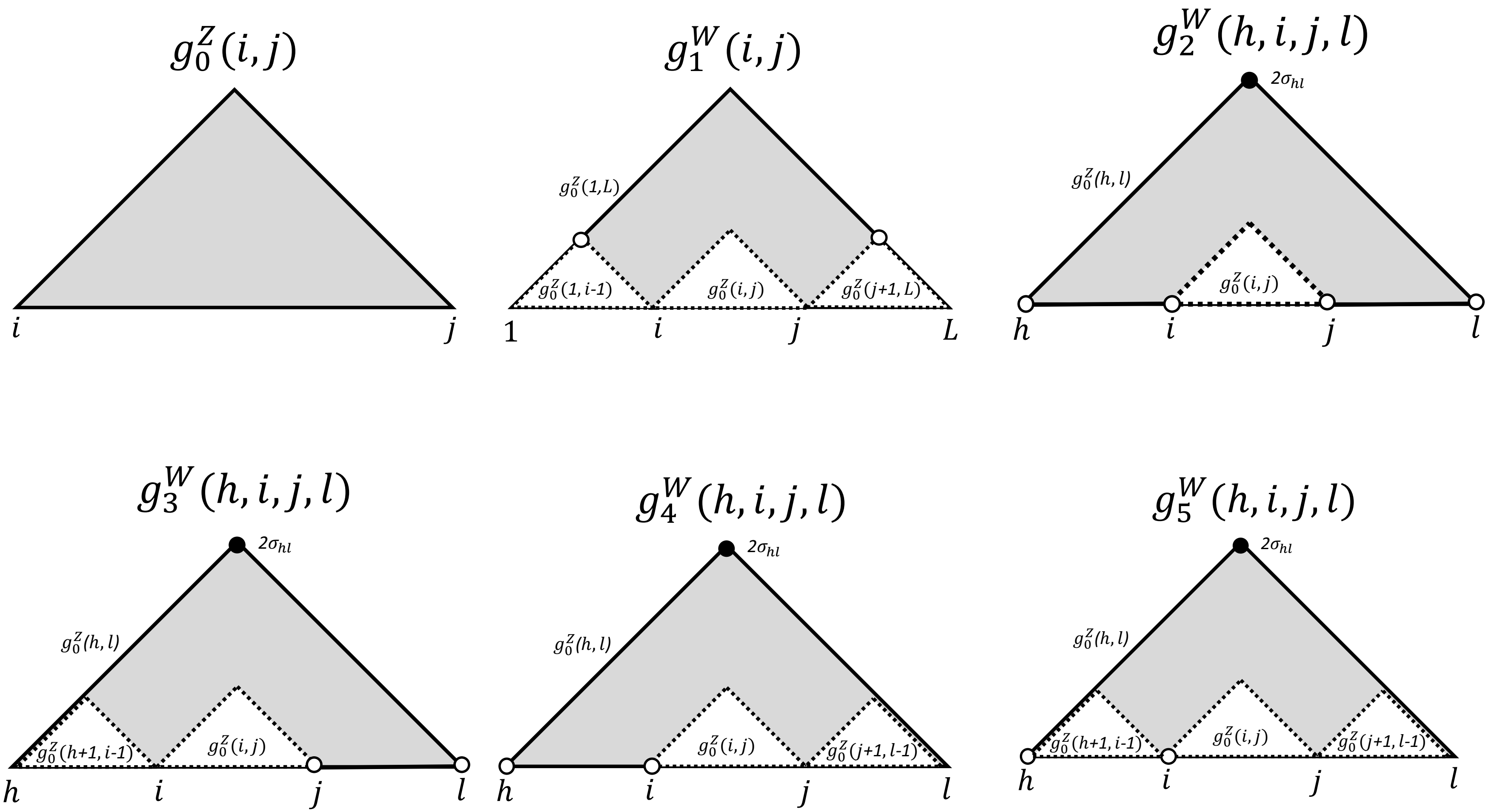


$$+ \sum_{\substack{k<i \\ \ell>j}} W^b_{k,\ell} e^{-f_3/k_BT} \Big[Z^m_{k+1,i-1} e^{-f_4(j,\ell)/k_BT} \times g_3^{\mathrm{W}}(k,i,j,\ell)$$

DP on polynomials for outside partition function over Hamming distance

$$W^b_{i,j} = Z_{1,i-1} Z_{j+1,L} x^{g_1^{\mathrm{W}}(i,j)}$$
$$+ \sum_{\substack{k<i \\ \ell>j}} W^b_{k,\ell} e^{-f_2(k,i,j,\ell)/k_B T} x^{g_2^{\mathrm{W}}(k,i,j,\ell)}$$
$$+ \sum_{\substack{k<i \\ \ell>j}} W^b_{k,\ell} e^{-f_3/k_B T} \left[\begin{array}{l} Z^m_{k+1,i-1} e^{-f_4(j,\ell)/k_B T} x^{g_3^{\mathrm{W}}(k,i,j,\ell)} \\ + Z^m_{j+1,\ell-1} e^{-f_4(k,i)/k_B T} x^{g_4^{\mathrm{W}}(k,i,j,\ell)} \\ + Z^m_{k+1,j-1} Z^m_{j+1,\ell-1} x^{g_5^{\mathrm{W}}(k,i,j,\ell)} \end{array} \right.$$

Gain functions



Base-pairing probability of specified region of Hamming distance

$$P_{ij}^{b}[r_{\min}, r_{\max}] = \frac{\sum_{d \in [r_{\min}, r_{\max}]} \sum_{t} Z_{ij}^{b}(d-t) W_{ij}^{b}(t)}{\sum_{d \in [r_{\min}, r_{\max}]} Z(d, \sigma)}$$

Calculating partition function over Hamming distance

$$\Delta = d_{max} - d_{min} + 1$$

$$\text{W}(d,\sigma) = \sum_{s=d_{min}}^{d_{max}} \text{W}(s,\sigma)\delta_{sd}$$

$$= \sum_{s=d_{min}}^{d_{max}} \text{W}(s,\sigma) \sum_{r=d_{min}}^{d_{max}} \frac{\exp\left[2\pi i \frac{r(s-d)}{\Delta}\right]}{\Delta}$$

DP on complex numbers

$$= \frac{1}{\Delta} \sum_{r=d_{min}}^{d_{max}} \exp\left[2\pi i \frac{-rd}{\Delta}\right] \sum_{s=d_{min}}^{d_{max}} \text{W}(s,\sigma) \exp\left[2\pi i \frac{rs}{\Delta}\right]$$

$$= \frac{1}{\Delta} \sum_{r=d_{min}}^{d_{max}} \exp\left[2\pi i \frac{-rd}{\Delta}\right] \tilde{\text{W}}_{1,L}(\sigma)$$

DFT

Computational complexity

	Polynomial		DFT
1D analysis			
Time :	$O(n^4)$	->	$O(n^3)$
Memory :	$O(n^2)$	->	$O(n^3)$
2D analysis			
Time :	$O(n^5)$	->	$O(n^3)$
Memory :	$O(n^2)$	->	$O(n^4)$

Concluding remarks

- Algorithm to calculate RNA base-pairing probabilities over Hamming distance has been shown.
- Implemented software, RintW, utilizes DP on polynomials and DFT for efficient calculation.
  - $O(L^4 H_{max})$ in time, $O(L^3)$ in memory
- Using RintW, alternative 2D structures of RNA are represented for RNA thermometer and Lysin Riboswitch.
- The web server is available in
  http://rtoos.cbrc.jp/

Acknowledgments

Contributers

Ryota Mori	Grad. Asai Lab.
Taichi Hagio	Unique Co. Ltd.
Shun Sakuraba	Nat. Inst. of Quantum & Radiological Sci. & Tech.
Junichi Iwakiri	Asai Lab.

Vienna RNA package (https://www.tbi.univie.ac.at/RNA/)

Collaborators in silico

Michiaki Hamada	Waseda Univ.
Hisanori Kiryu	Univ. Tokyo
Yukiteru Ono	IMSBIO Inc.
Martin Frith	AIST
Osamu Gotoh	AIST
Anish Shrestha	Asai Lab.
Zeng Chao	AIST

Collaborators experimental

Mikiko Shiomi	U-Tokyo
Tetsuro Hirose	Hokkaido Univ.
Hiroshi Abe	Hokkaido Univ..