Lecture 6: RNA-RNA interaction

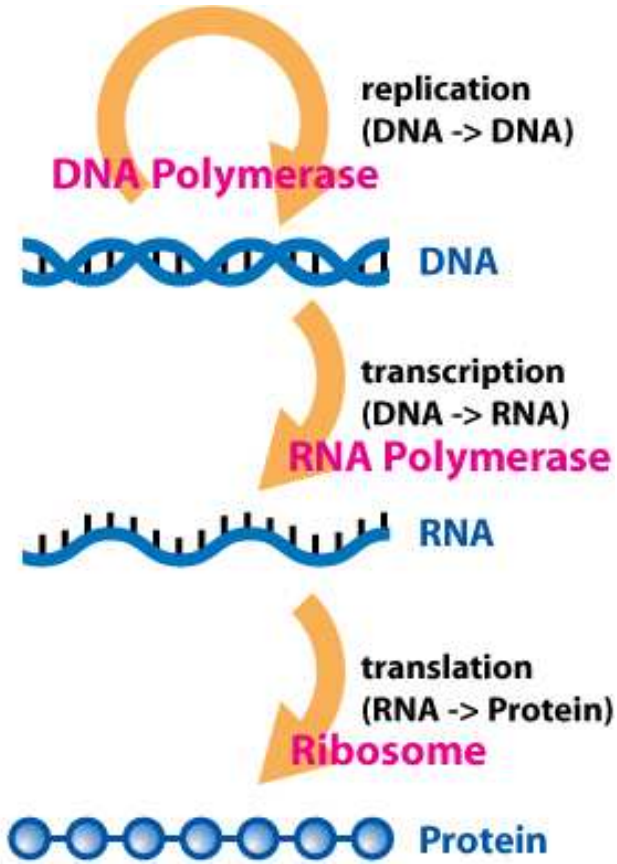
Hamidreza Chitsaz

Colorado State University
chitsaz@cs.colostate.edu

Spring 2020
February 18,20, 2020

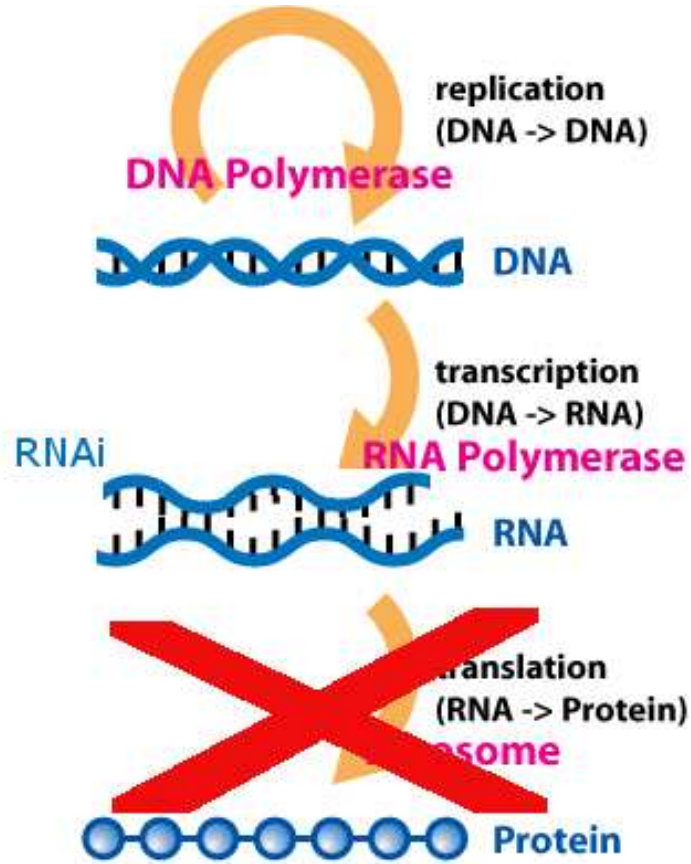
Central dogma

DNA → RNA → Protein



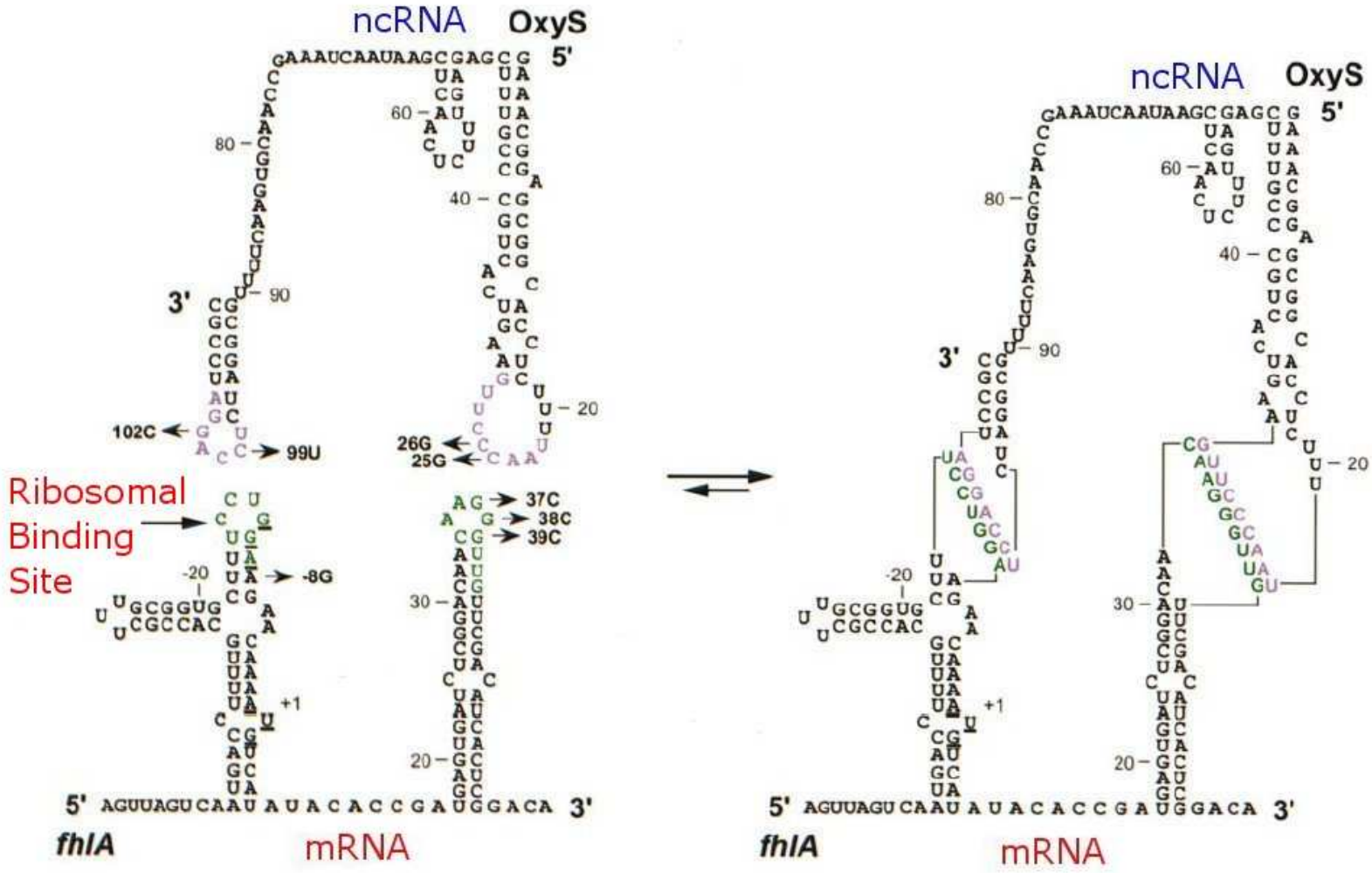
Motivation

Post-transcriptional regulation of gene expression



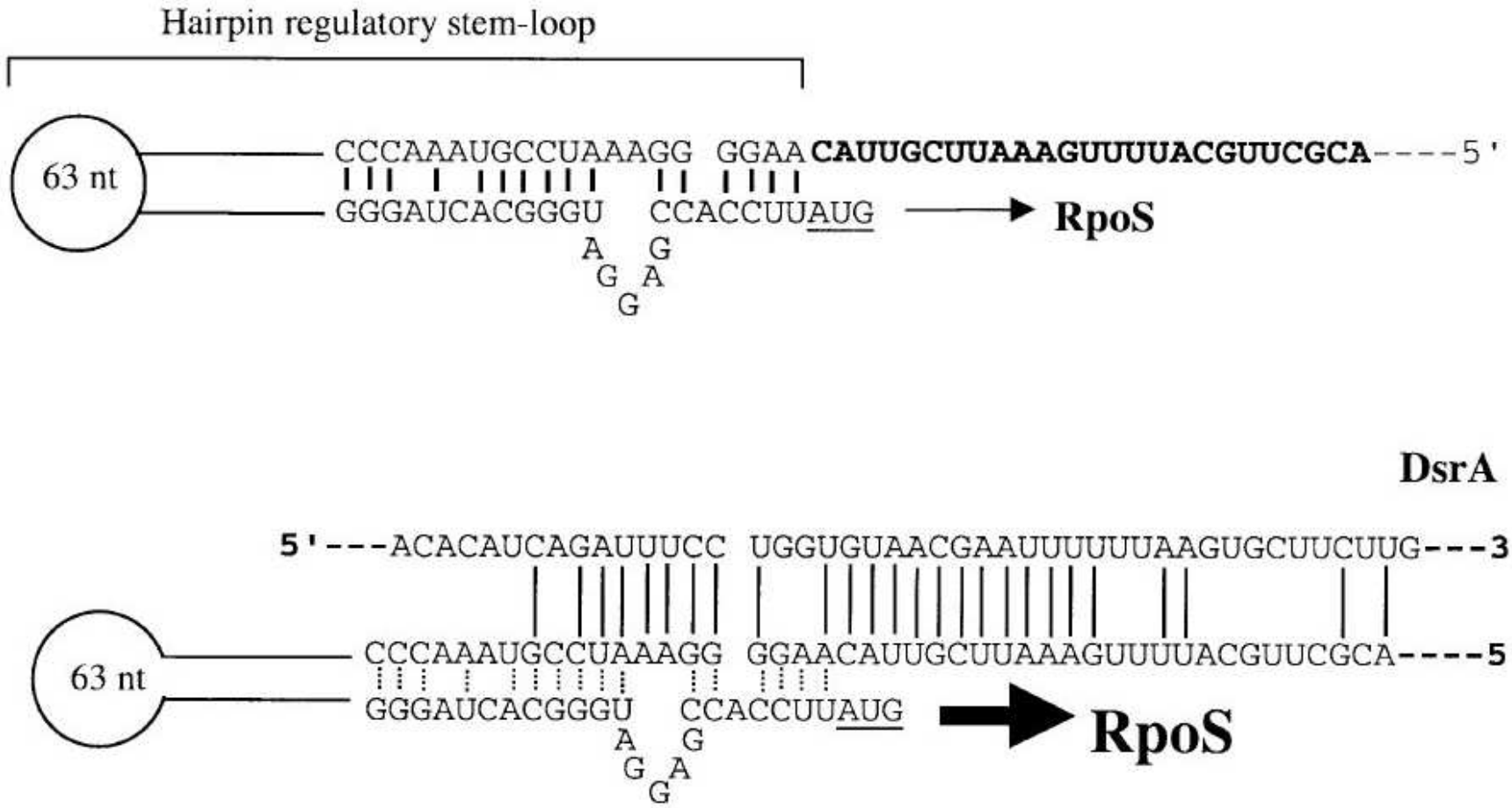
Regulatory RNA

Repression example (Argaman and Altuvia, J. Mol. Biol. 2000)



Regulatory RNA

Activation example (Repoila, Majdalani, and Gottesman, Mol. Microbiol. 2003)



Background

RNA-RNA MFE structure prediction

- Avoid intramolecular base pairing
  RNAhybrid (Rehmsmeier *et al.* 2004), **RNAduplex** (Bernhart *et al.* 2006), **UNAFold** (Markham *et al.* 2008)
  No internal structure
- Concatenate input sequences as a single strand; no pseudoknots
  PairFold (Andronescu *et al.* 2005), **RNAcofold** (Bernhart *et al.* 2006)
  No kissing hairpins
- Predict binding sites
  RNAup (Mückstein *et al.* 2008), **intaRNA** (Busch *et al.* 2008)
  Just one binding site not complete structure
- Concatenate input sequences; consider special pseudoknots
  NUPACK (Dirks *et al.* 2003,2007)

  Still no kissing hairpins!

Background

RNA-RNA MFE structure prediction

- Avoid intramolecular base pairing
  RNAhybrid (Rehmsmeier *et al.* 2004), **RNAduplex** (Bernhart *et al.* 2006), **UNAFold** (Markham *et al.* 2008)
  No internal structure
- Concatenate input sequences as a single strand; no pseudoknots
  PairFold (Andronescu *et al.* 2005), **RNAcofold** (Bernhart *et al.* 2006)
  No kissing hairpins
- Predict binding sites
  RNAup (Mückstein *et al.* 2008), **intaRNA** (Busch *et al.* 2008)
  Just one binding site not complete structure
- Concatenate input sequences; consider special pseudoknots
  NUPACK (Dirks *et al.* 2003,2007)

  Still no kissing hairpins!

Background

RNA-RNA MFE structure prediction

- Avoid intramolecular base pairing
  RNAhybrid (Rehmsmeier *et al.* 2004), **RNAduplex** (Bernhart *et al.* 2006), **UNAFold** (Markham *et al.* 2008)
  No internal structure
- Concatenate input sequences as a single strand; no pseudoknots
  PairFold (Andronescu *et al.* 2005), **RNAcofold** (Bernhart *et al.* 2006)
  No kissing hairpins
- Predict binding sites
  RNAup (Mückstein *et al.* 2008), **intaRNA** (Busch *et al.* 2008)
  Just one binding site not complete structure
- Concatenate input sequences; consider special pseudoknots
  NUPACK (Dirks *et al.* 2003,2007)

  Still no kissing hairpins!

Background

RNA-RNA MFE structure prediction

- Avoid intramolecular base pairing
  RNAhybrid (Rehmsmeier *et al.* 2004), **RNAduplex** (Bernhart *et al.* 2006), **UNAFold** (Markham *et al.* 2008)
  No internal structure
- Concatenate input sequences as a single strand; no pseudoknots
  PairFold (Andronescu *et al.* 2005), **RNAcofold** (Bernhart *et al.* 2006)
  No kissing hairpins
- Predict binding sites
  RNAup (Mückstein *et al.* 2008), **intaRNA** (Busch *et al.* 2008)
  Just one binding site not complete structure
- Concatenate input sequences; consider special pseudoknots
  NUPACK (Dirks *et al.* 2003,2007)

  Still no kissing hairpins!

Background (continued)

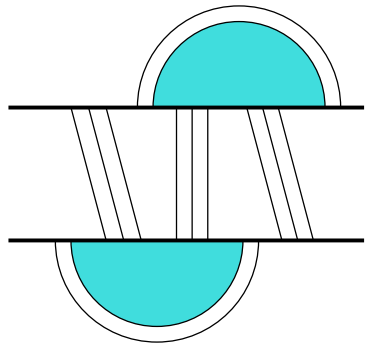
RNA-RNA MFE structure prediction

Consider inter- and intramolecular base pairing

IRIS (Pervouchine 2004), **inteRNA** (Alkan *et al.* 2005), **Grammatical Approach** (Kato *et al.* 2009)

Voilà, now we are talking business.

The problem is NP-Hard (Alkan *et al.* 2005); no surprise as pseudoknots are NP-Hard. Exclude *zigzags* and crossing interactions to lift the curse of complexity and obtain an exact $O(n^6)$-time $O(n^4)$-space DP algorithm (albeit for simple base-pair counting).



First order zigzag. A general zigzag involves an arbitrary number of kissing hairpins.

Ahhh...but MFE is often wrong!

Question: how about

1. computing base pairing probabilities,
2. sampling from the Boltzmann ensemble of interaction structures, clustering, centroids, etc.,
3. and computing equilibrium concentrations and melting temperature for RNA-RNA compounds?

Answer: the key enabling technology is the **partition function**. All of the above can be computed from the partition function.

Ahhh...but MFE is often wrong!

Question: how about

1. computing base pairing probabilities,
2. sampling from the Boltzmann ensemble of interaction structures, clustering, centroids, etc.,
3. and computing equilibrium concentrations and melting temperature for RNA-RNA compounds?

Answer: the key enabling technology is the **partition function**. All of the above can be computed from the partition function.

Ahhh...but MFE is often wrong!

Question: how about

1. computing base pairing probabilities,
2. sampling from the Boltzmann ensemble of interaction structures, clustering, centroids, etc.,
3. and computing equilibrium concentrations and melting temperature for RNA-RNA compounds?

Answer: the key enabling technology is the **partition function**. All of the above can be computed from the partition function.

Ahhh...but MFE is often wrong!

Question: how about

1. computing base pairing probabilities,
2. sampling from the Boltzmann ensemble of interaction structures, clustering, centroids, etc.,
3. and computing equilibrium concentrations and melting temperature for RNA-RNA compounds?

Answer: the key enabling technology is the **partition function**. All of the above can be computed from the partition function.

Partition function

$$Q(T) = \sum_{f \in F} e^{-G_f/RT},$$

$F =$ All permissible interaction structures,

$$p(f) \propto e^{-G_f/RT},$$

and Q is the normalizing factor. Also other thermodynamic quantities can be derived from Q.

Partition function

$$Q(T) = \sum_{f \in F} e^{-G_f/RT},$$

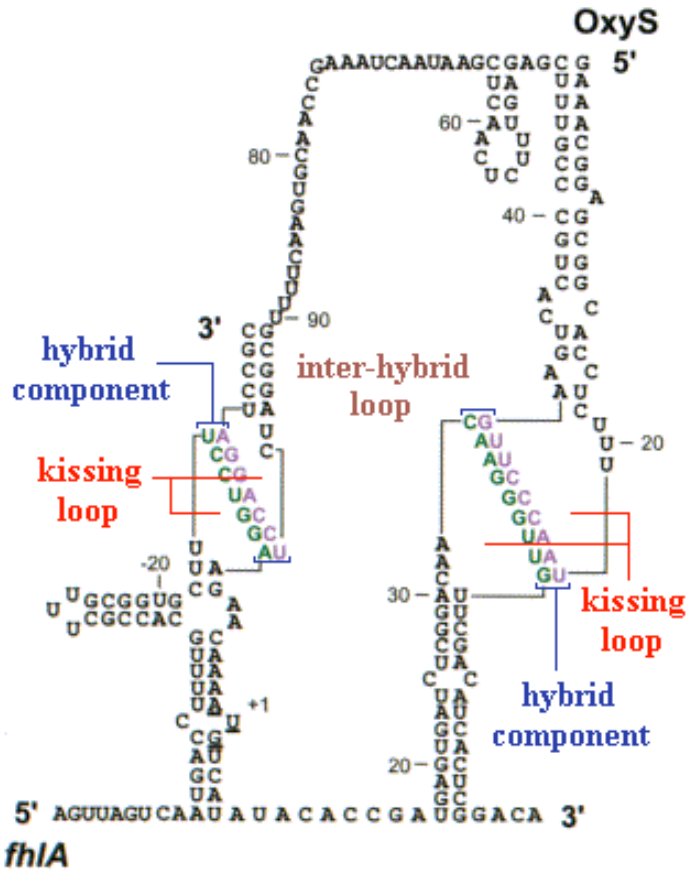
$$F = \text{All permissible interaction structures},$$

$$p(f) \propto e^{-G_f/RT},$$

and Q is the normalizing factor. Also other thermodynamic quantities can be derived from Q.

Our extension of the Turner model

Chitsaz *et al.*, Bioinformatics 25(12): i365-i373

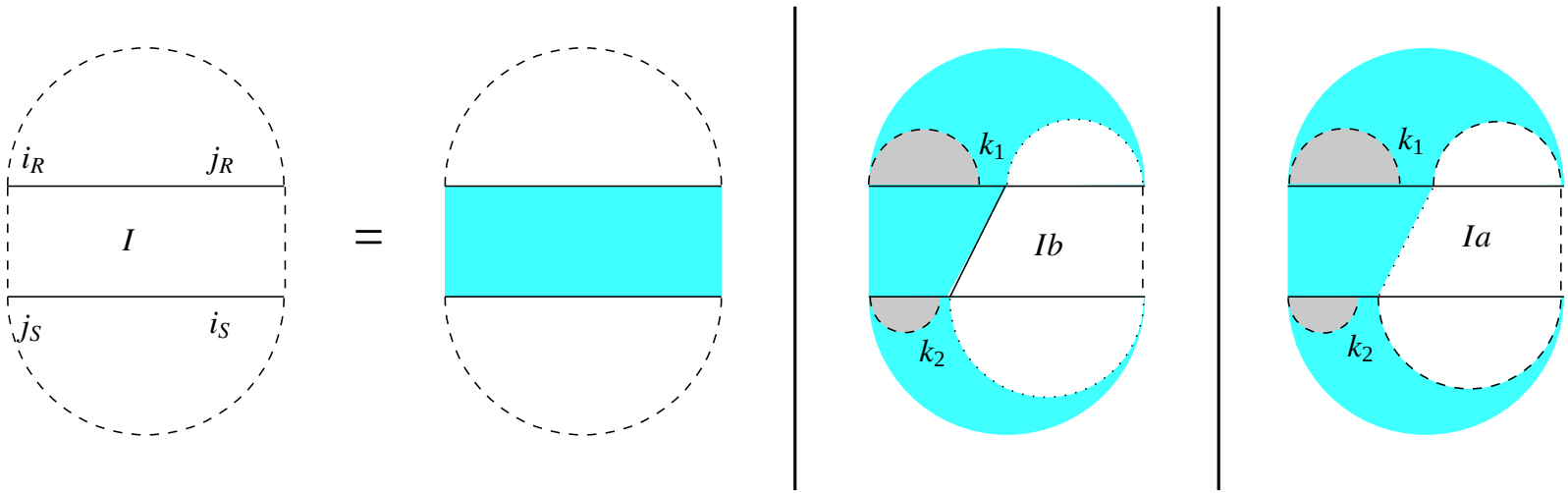


Hybrid component: as if intramolecular, with penalties.

Kissing loop: like multibranch loop.

Partition function for two strands

straight vertical line: intermolecular bond
solid: a base pair
dotted: not a base pair
dashed: either of those two

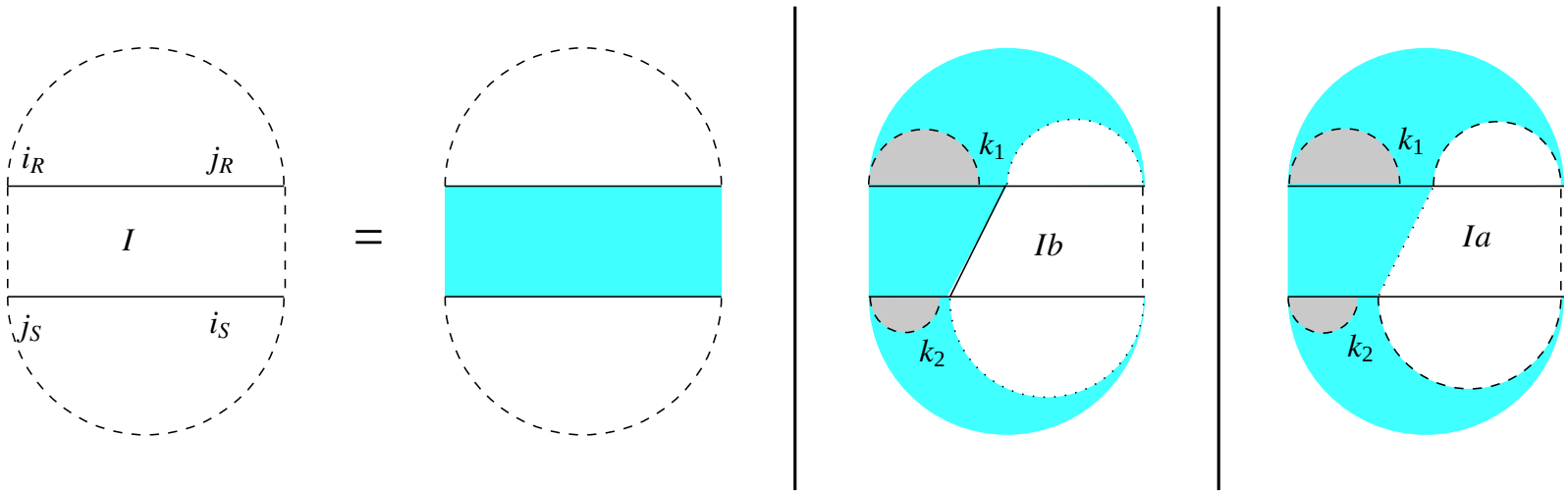


$$Q^I_{i_R,j_R,i_S,j_S} = Q_{i_R,j_R} Q_{i_S,j_S} + \sum_{\substack{i_R \le k_1 < j_R \\ i_S < k_2 \le j_S}} Q_{i_R,k_1-1} Q_{k_2+1,j_S} Q^{Ib}_{k_1,j_R,i_S,k_2} +$$

$$\sum_{\substack{i_R \le k_1 < j_R \\ i_S < k_2 \le j_S}} Q_{i_R,k_1-1} Q_{k_2+1,j_S} Q^{Ia}_{k_1,j_R,i_S,k_2}.$$

Partition function for two strands

straight vertical line: intermolecular bond
solid: a base pair
dotted: not a base pair
dashed: either of those two

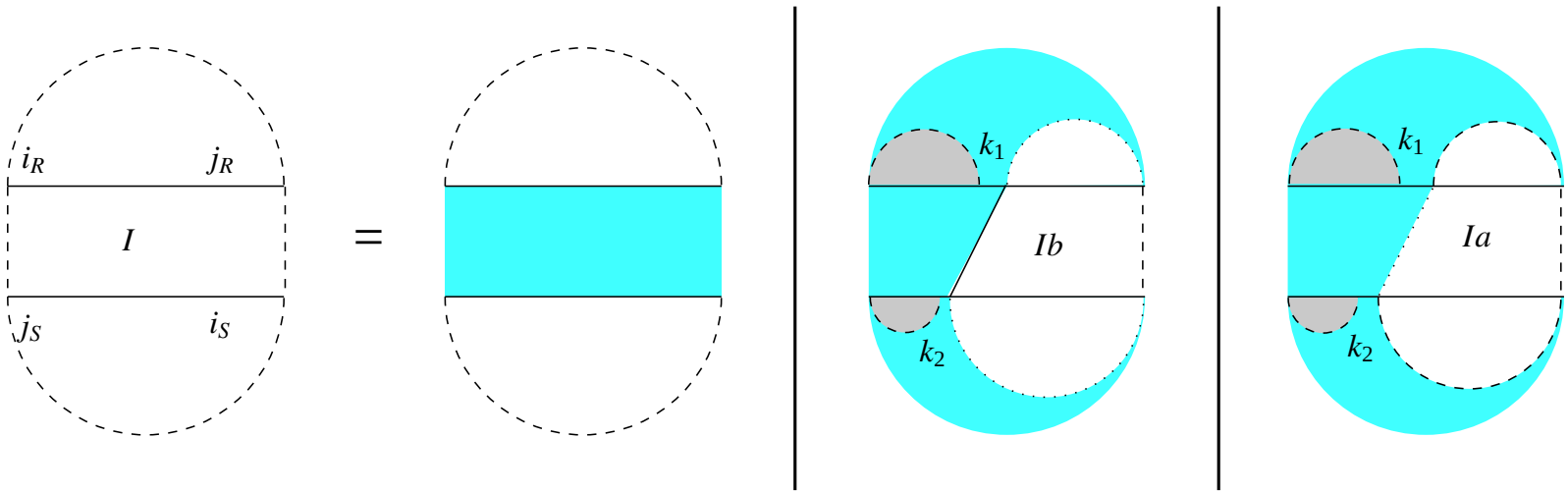


$$Q^{I}_{i_R,j_R,i_S,j_S} = Q_{i_R,j_R} Q_{i_S,j_S} + \sum_{\substack{i_R \le k_1 < j_R \\ i_S < k_2 \le j_S}} Q_{i_R,k_1-1} Q_{k_2+1,j_S} Q^{Ib}_{k_1,j_R,i_S,k_2} +$$

$$\sum_{\substack{i_R \le k_1 < j_R \\ i_S < k_2 \le j_S}} Q_{i_R,k_1-1} Q_{k_2+1,j_S} Q^{Ia}_{k_1,j_R,i_S,k_2}.$$

Partition function for two strands

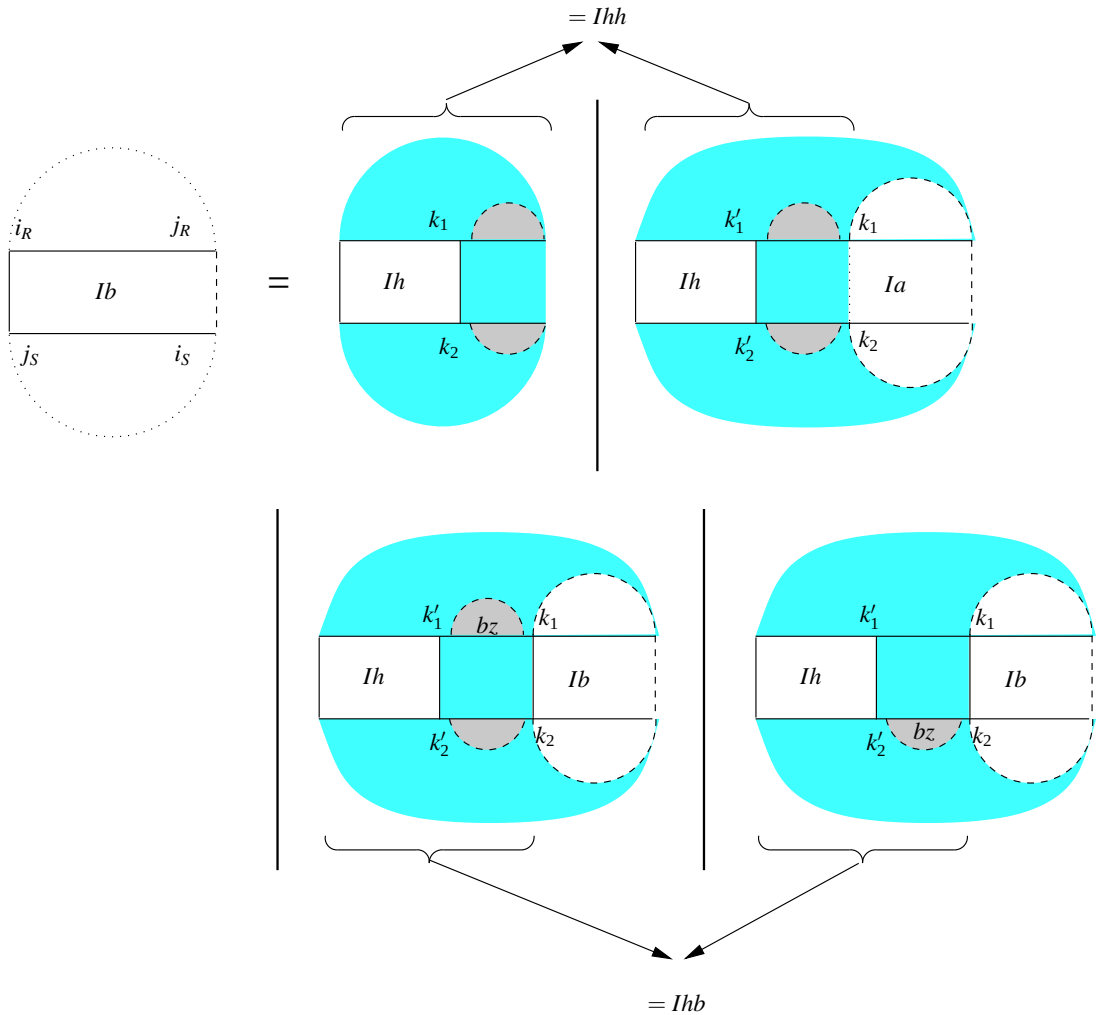
straight vertical line: intermolecular bond
solid: a base pair
dotted: not a base pair
dashed: either of those two



$$Q^{I}_{i_R,j_R,i_S,j_S} = Q_{i_R,j_R} Q_{i_S,j_S} + \sum_{\substack{i_R \le k_1 < j_R \\ i_S < k_2 \le j_S}} Q_{i_R,k_1-1} Q_{k_2+1,j_S} Q^{Ib}_{k_1,j_R,i_S,k_2} +$$

$$\sum_{\substack{i_R \le k_1 < j_R \\ i_S < k_2 \le j_S}} Q_{i_R,k_1-1} Q_{k_2+1,j_S} Q^{Ia}_{k_1,j_R,i_S,k_2}.$$

Q^{Ib}



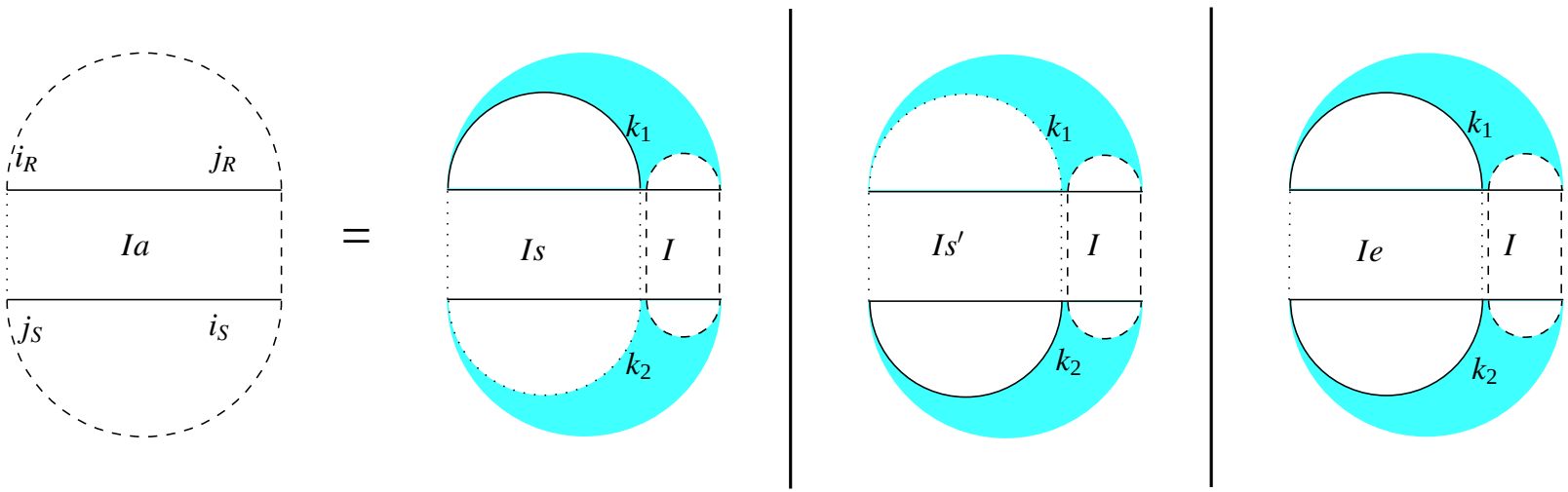
b: stands for bond

Q^{Ia}

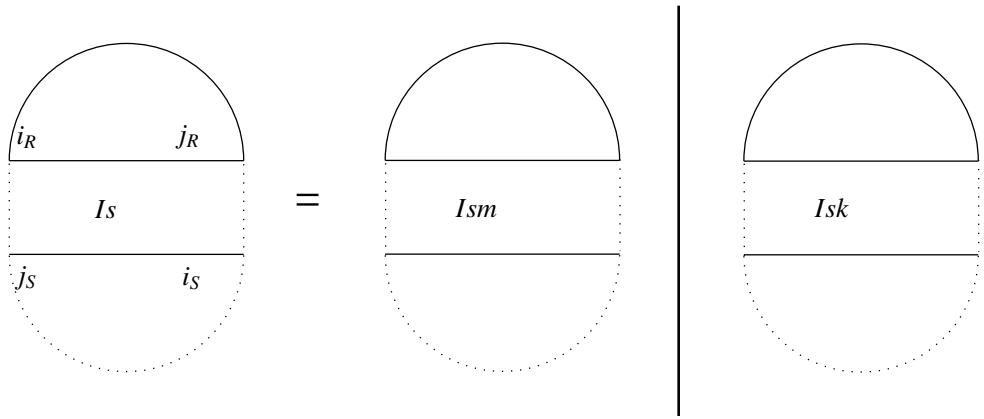
a: stands for arc

s: stands for subsume

e: stands for equivalent



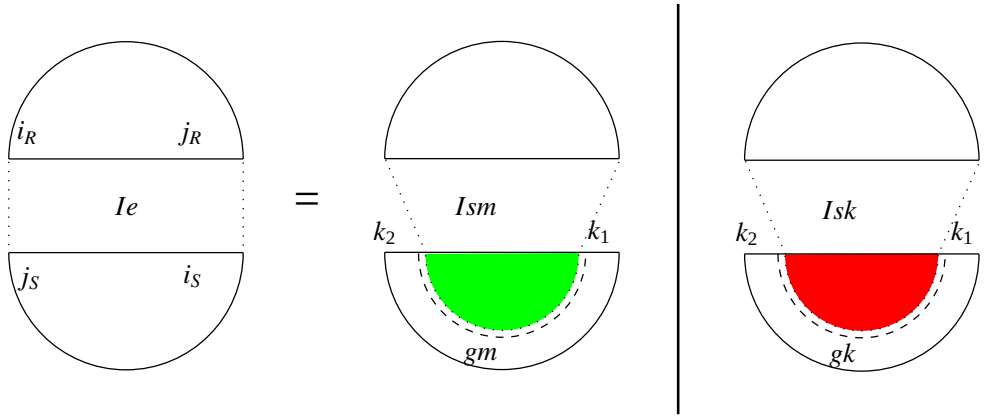
Q^{Is} and Q^{Ie}



s: stands for subsume

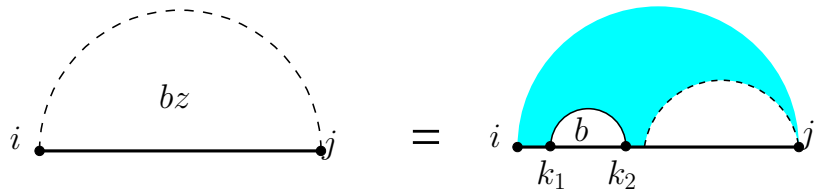
k: stands for kissing-loop

m: stands for multi-loop

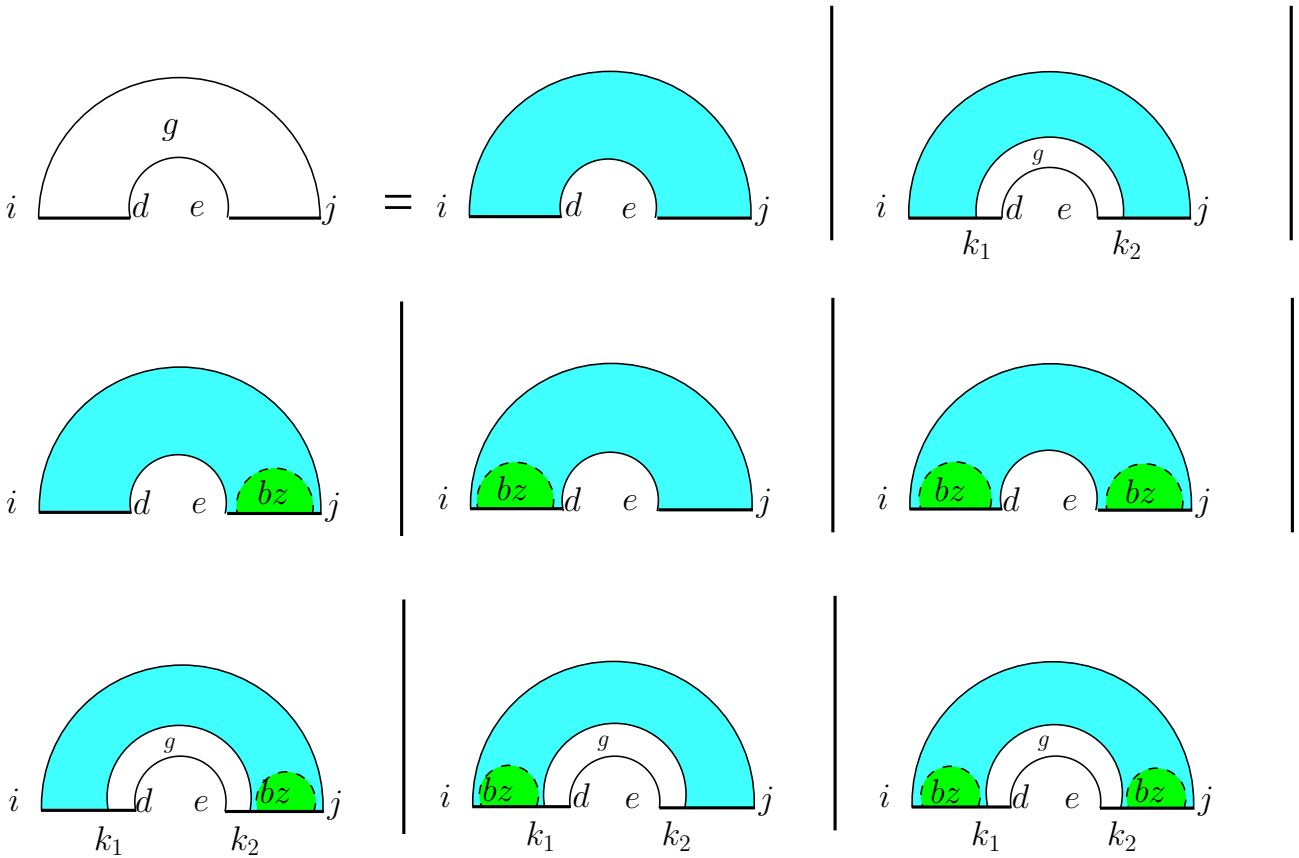


e: stands for equivalent

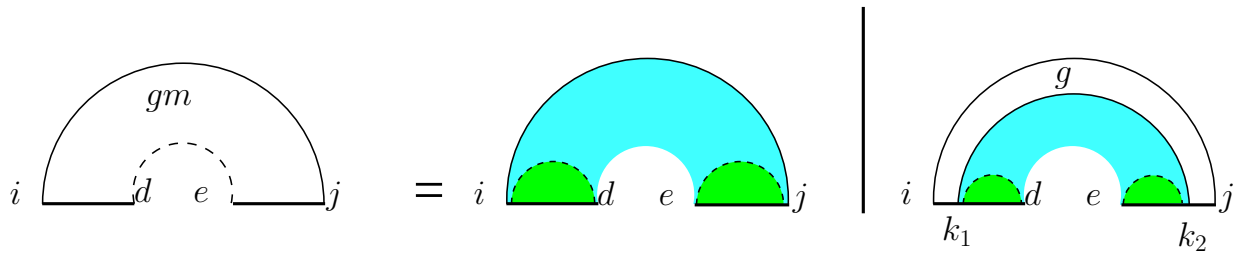
All tables



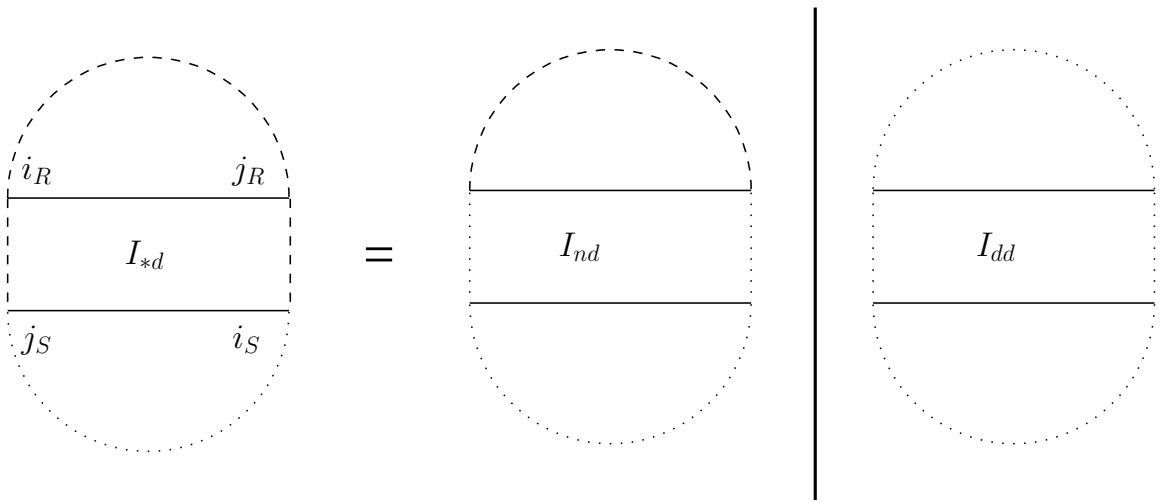
All tables


i
g
d
e
j
=
i
d
e
j
i
g
d
e
j
k1
k2
i
d
e
bz
j
i
bz
d
e
j
i
bz
d
e
bz
j
i
g
d
e
bz
j
k1
k2
i
bz
g
d
e
j
k1
k2
i
bz
g
d
e
bz
j
k1
k2

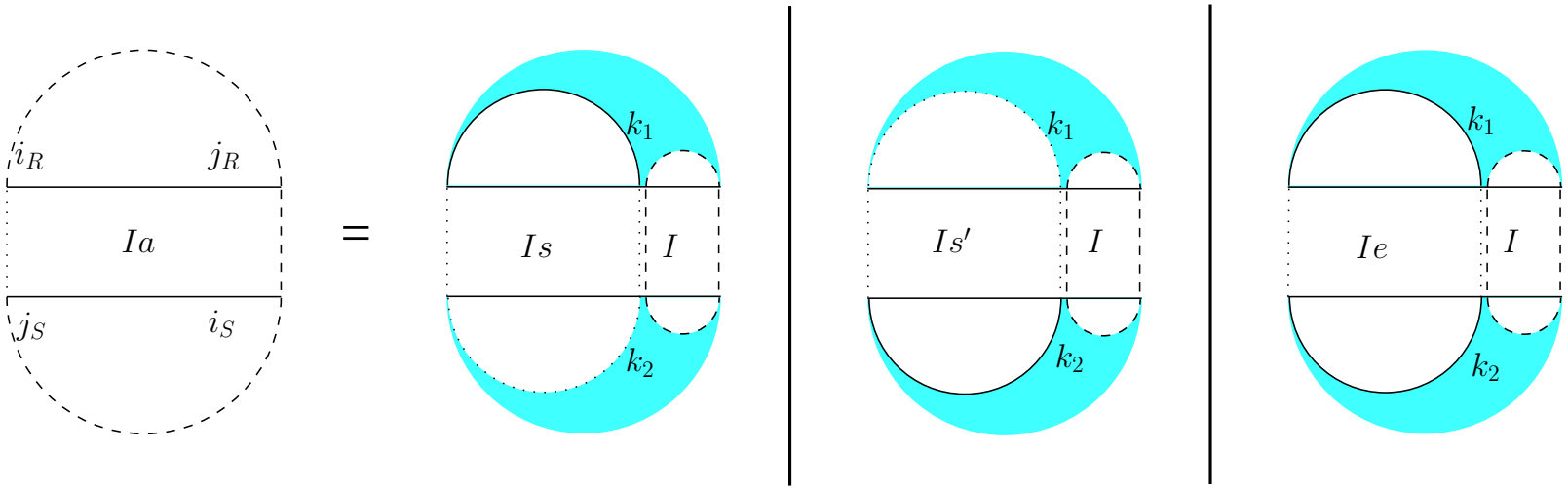
All tables



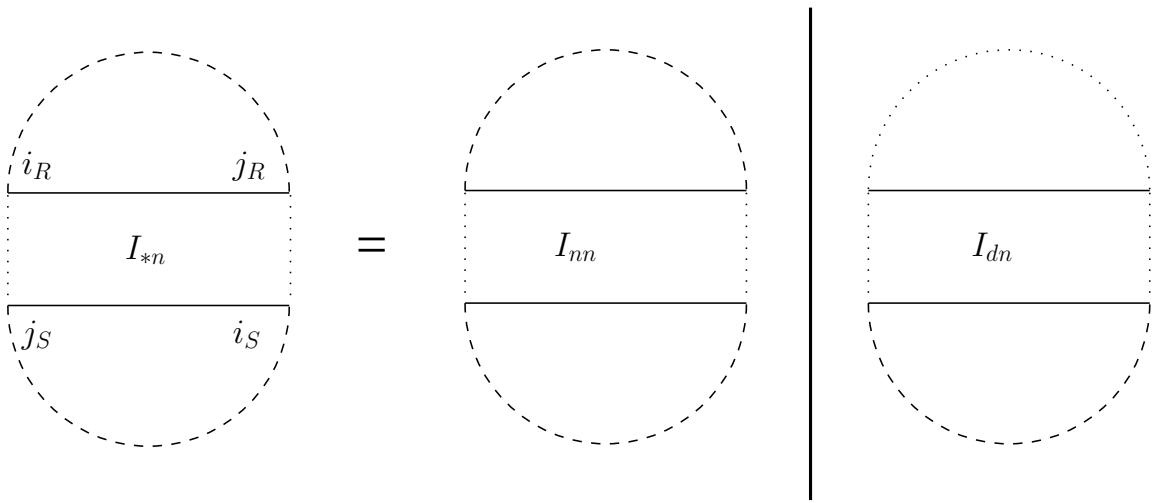
All tables



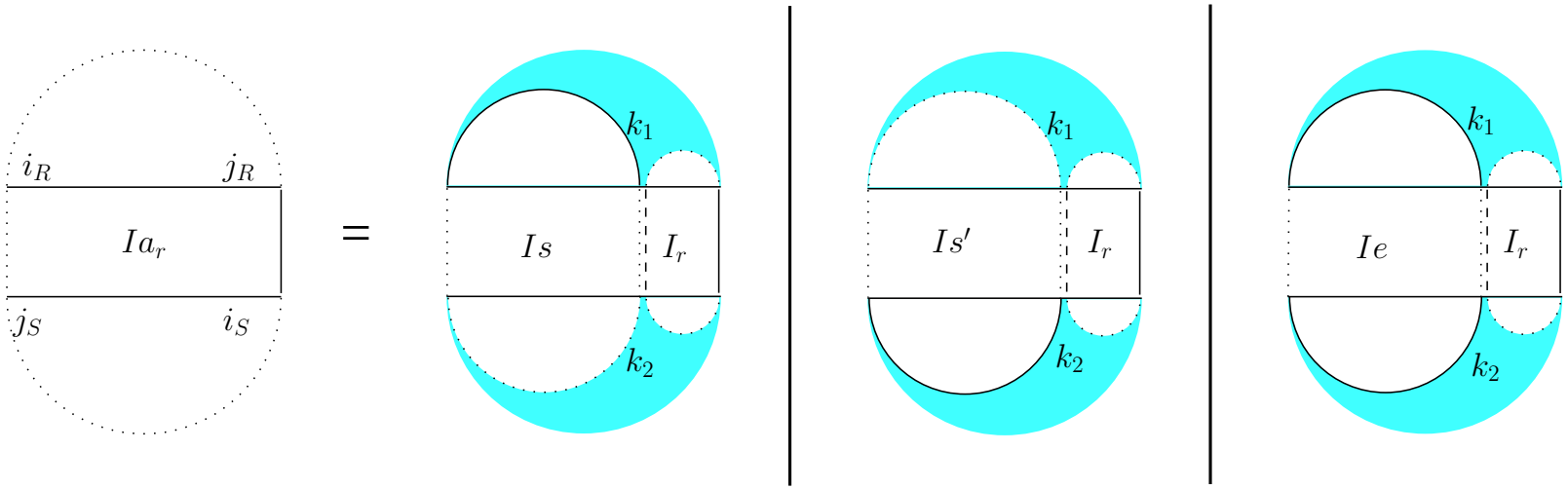
All tables



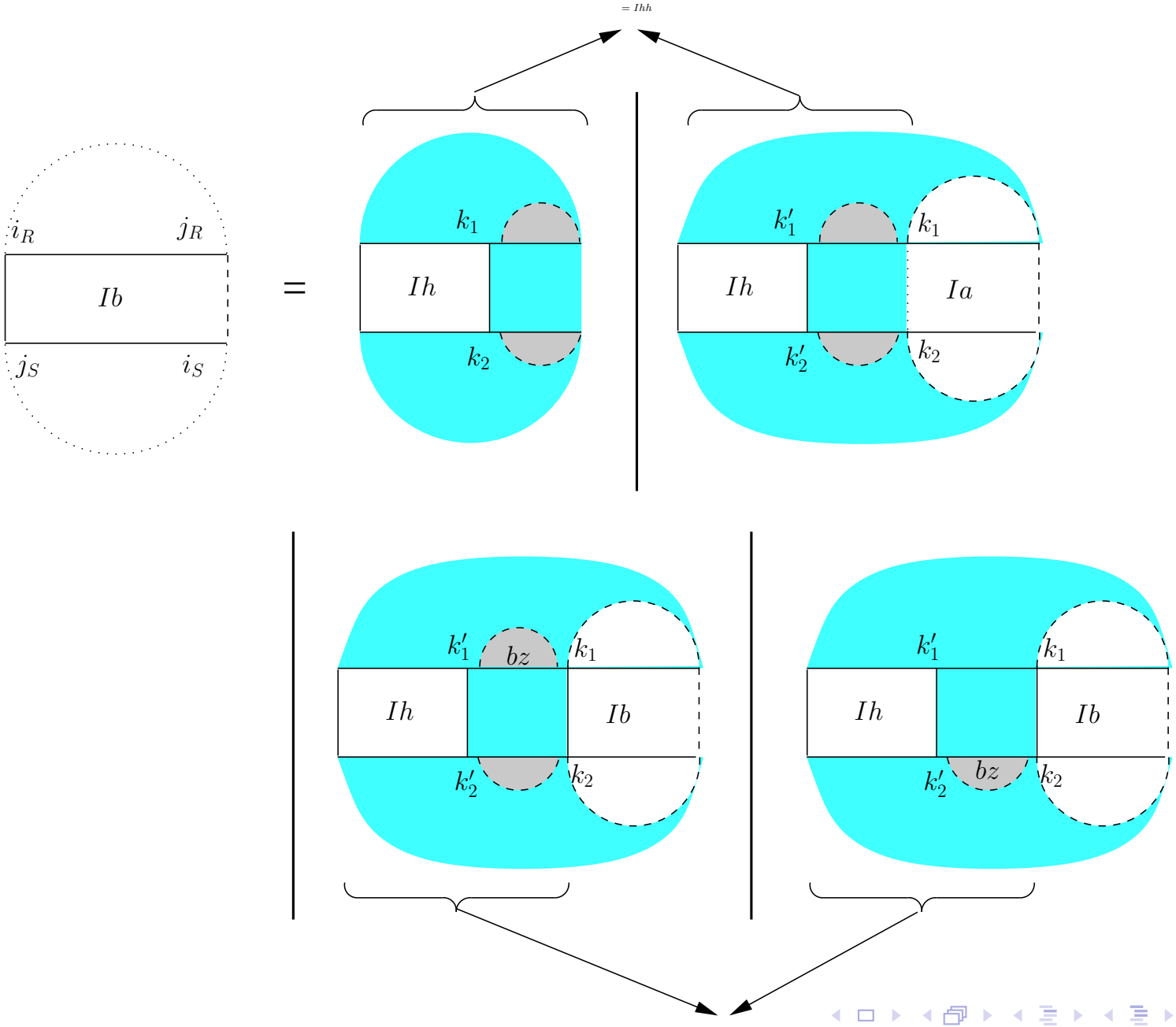
All tables



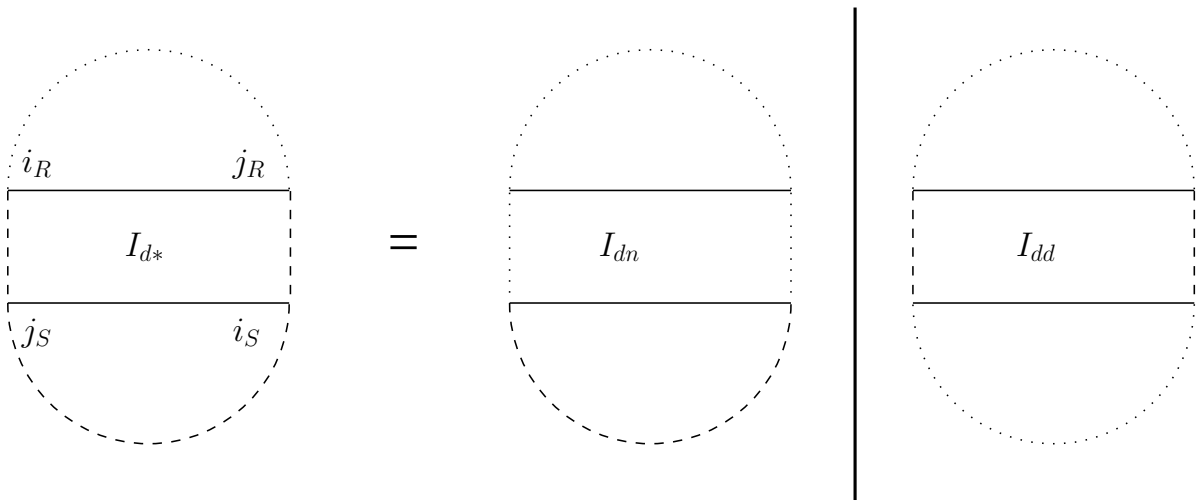
All tables



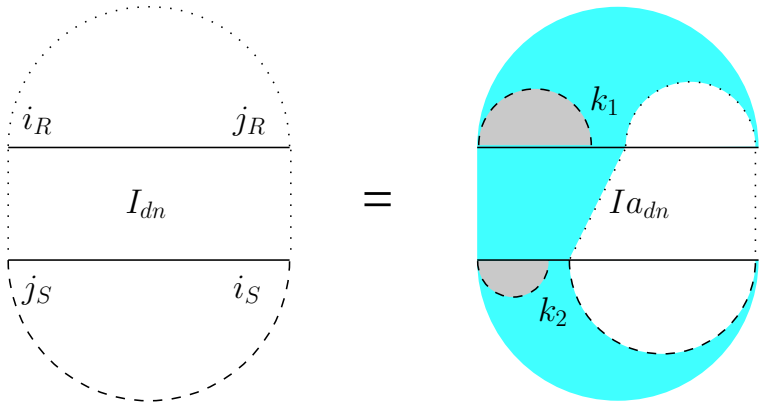
All tables


$= Ihh$
i_R
j_R
Ib
j_S
i_S
=
k_1
Ih
k_2
k_1'
k_1
Ih
Ia
k_2'
k_2
k_1'
bz
k_1
Ih
Ib
k_2'
k_2
k_1'
k_1
Ih
Ib
k_2'
bz
k_2

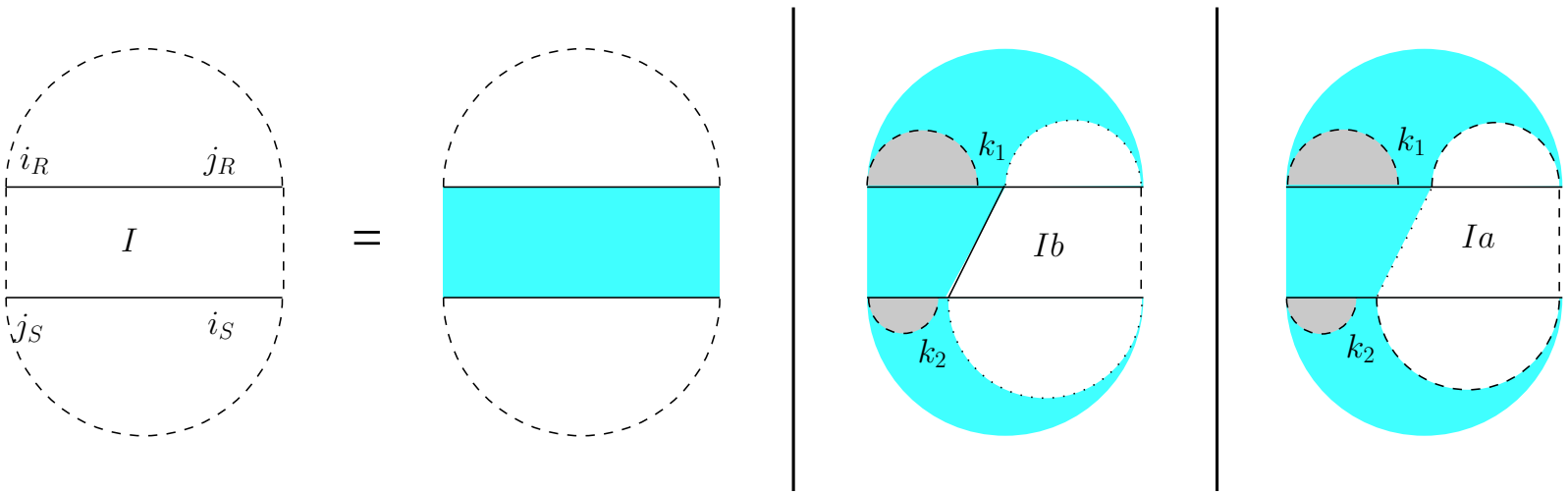
All tables



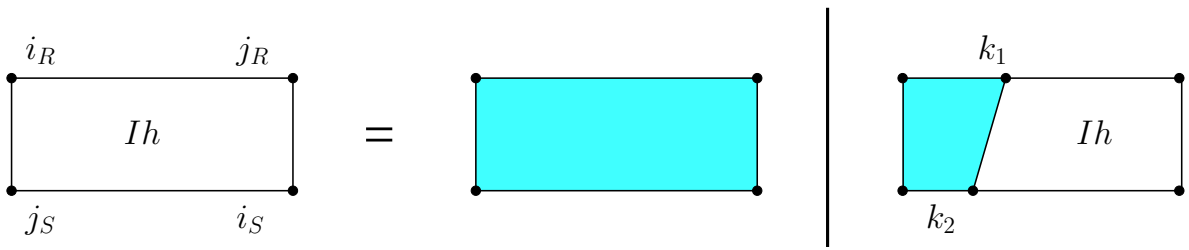
All tables



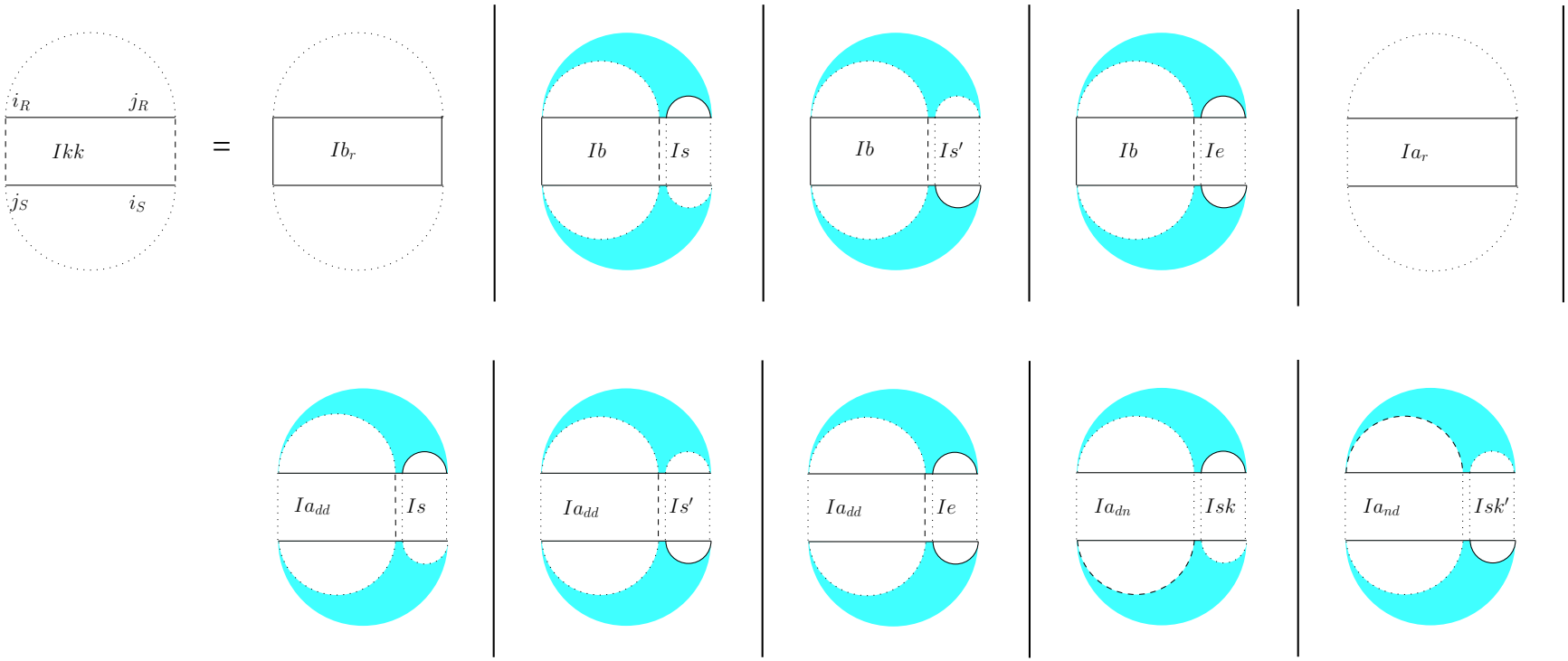
All tables



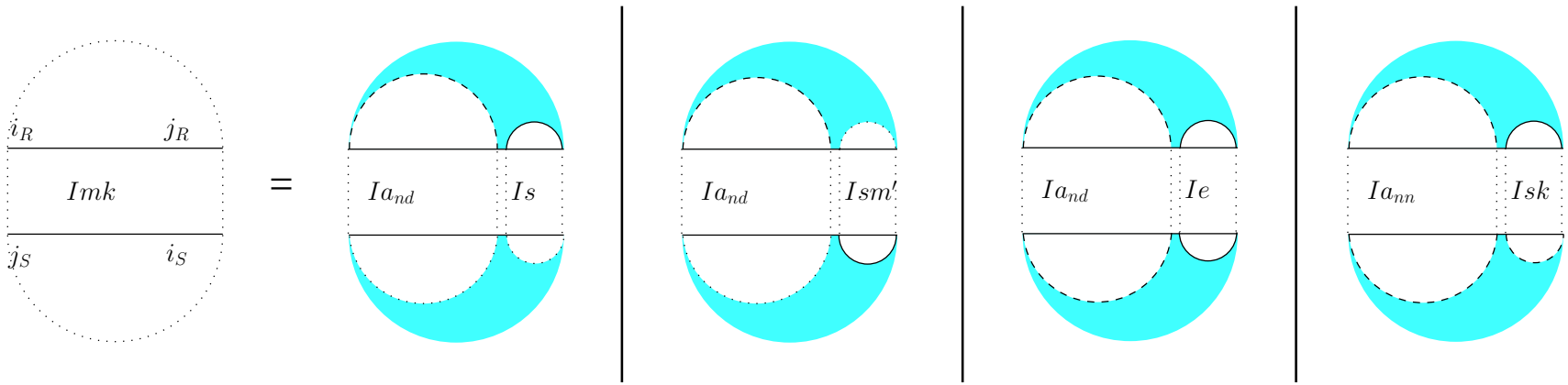
All tables



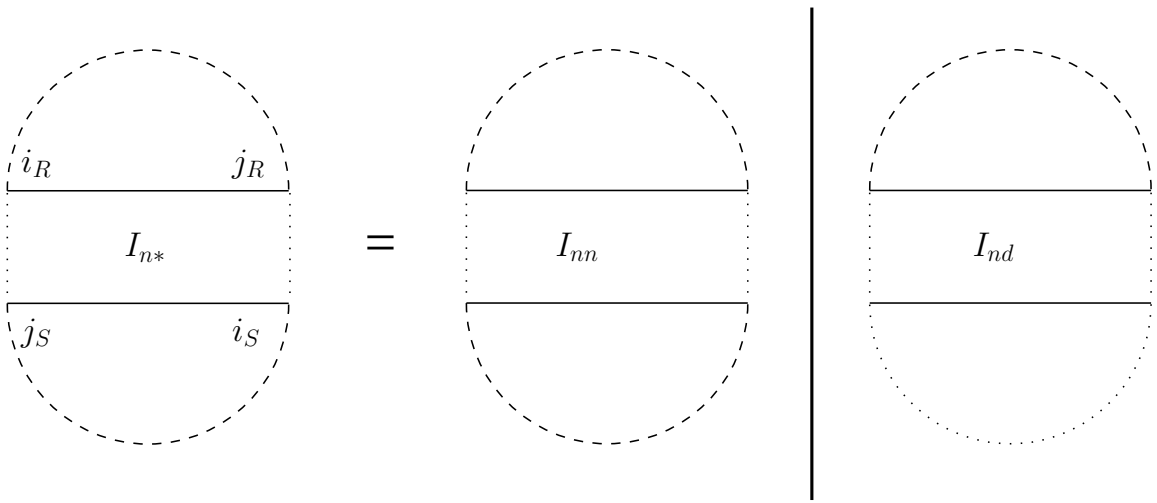
All tables


i_R
j_R
Ikk
j_S
i_S
=
Ib_r
Ib
Is
Ib
Is'
Ib
Ie
Ia_r
Ia_{dd}
Is
Ia_{dd}
Is'
Ia_{dd}
Ie
Ia_{dn}
Isk
Ia_{nd}
Isk'

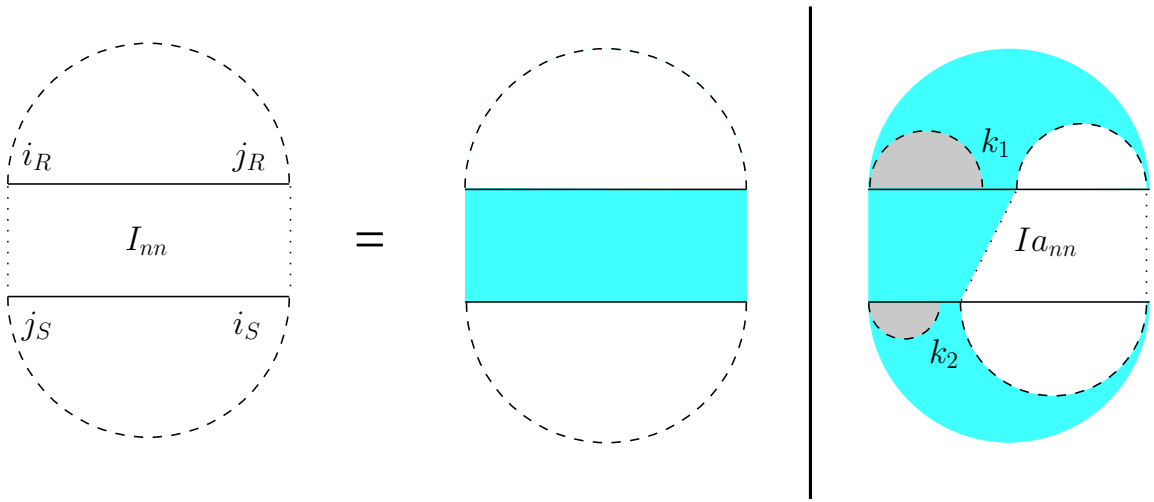
All tables



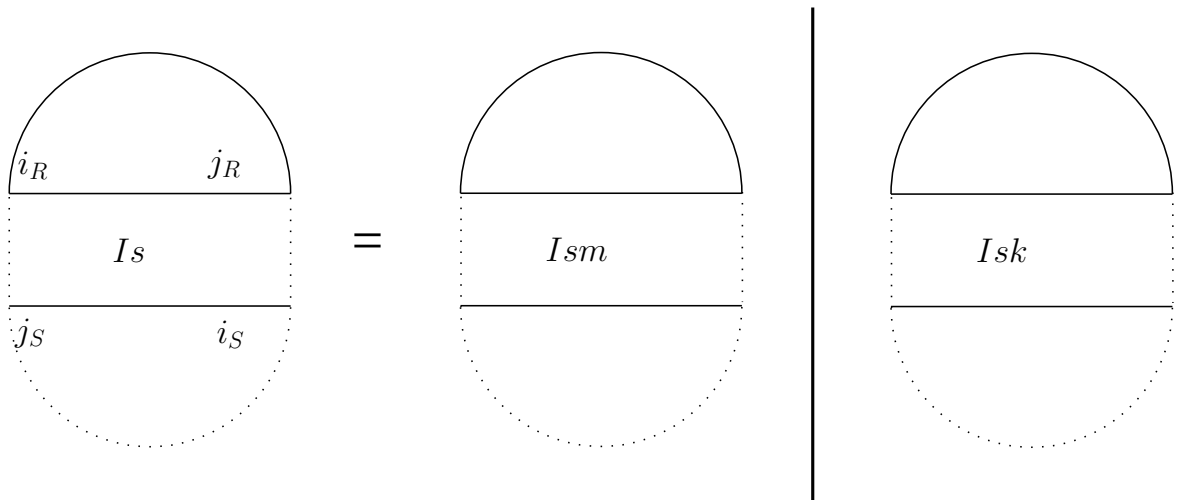
All tables



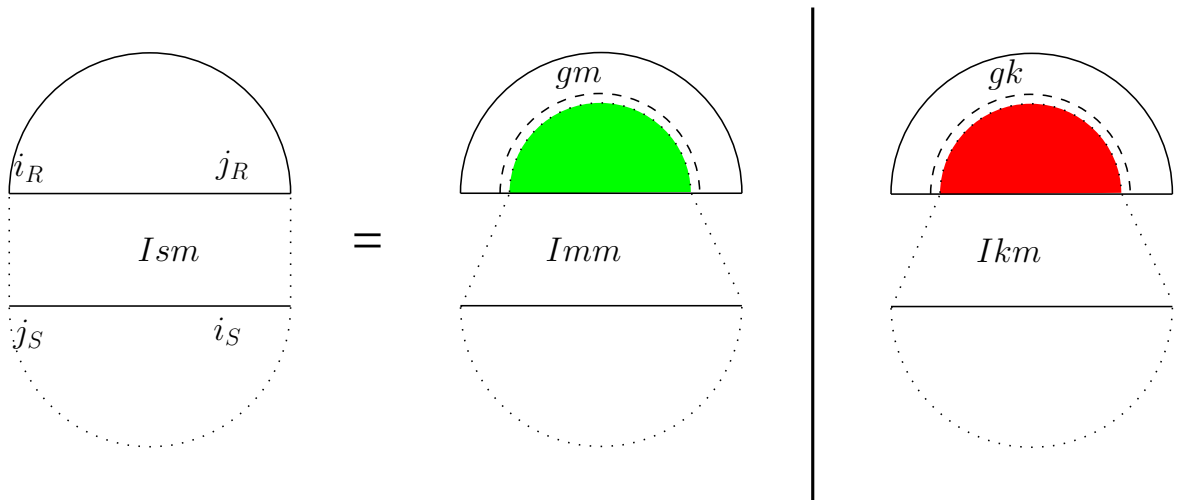
All tables



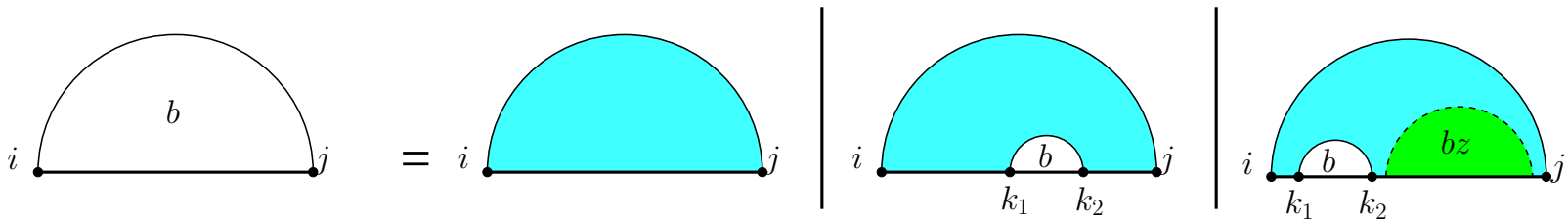
All tables



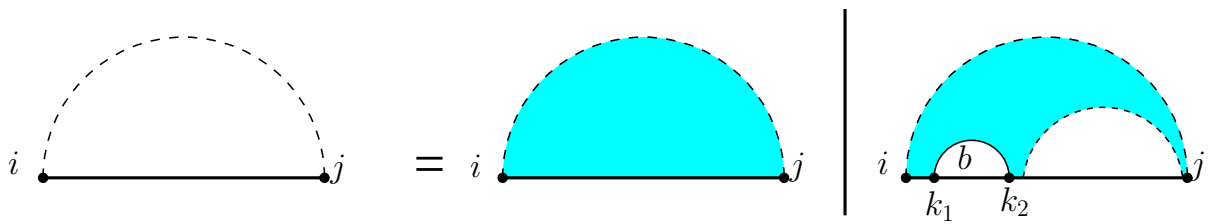
All tables



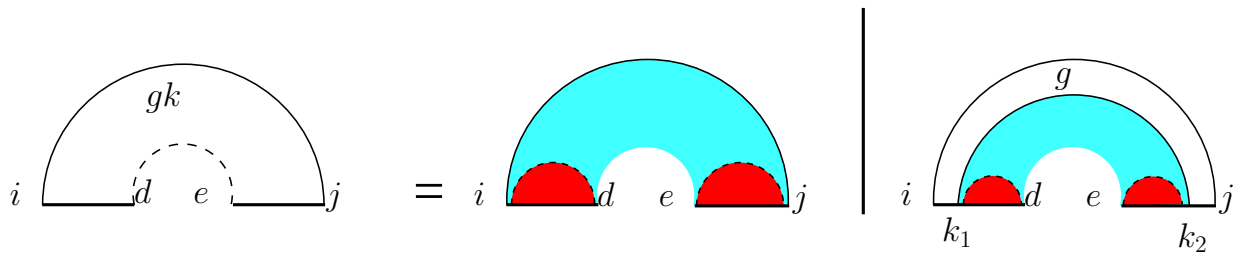
All tables



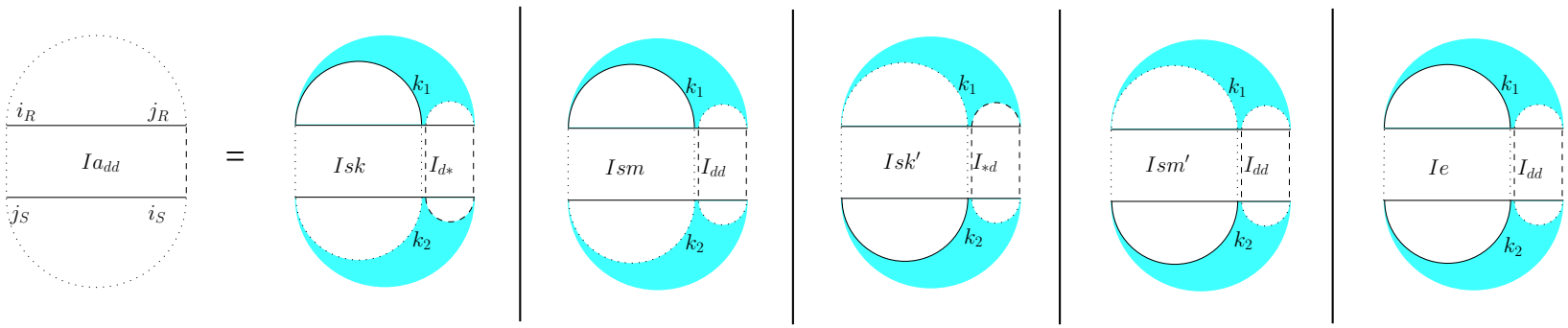
All tables



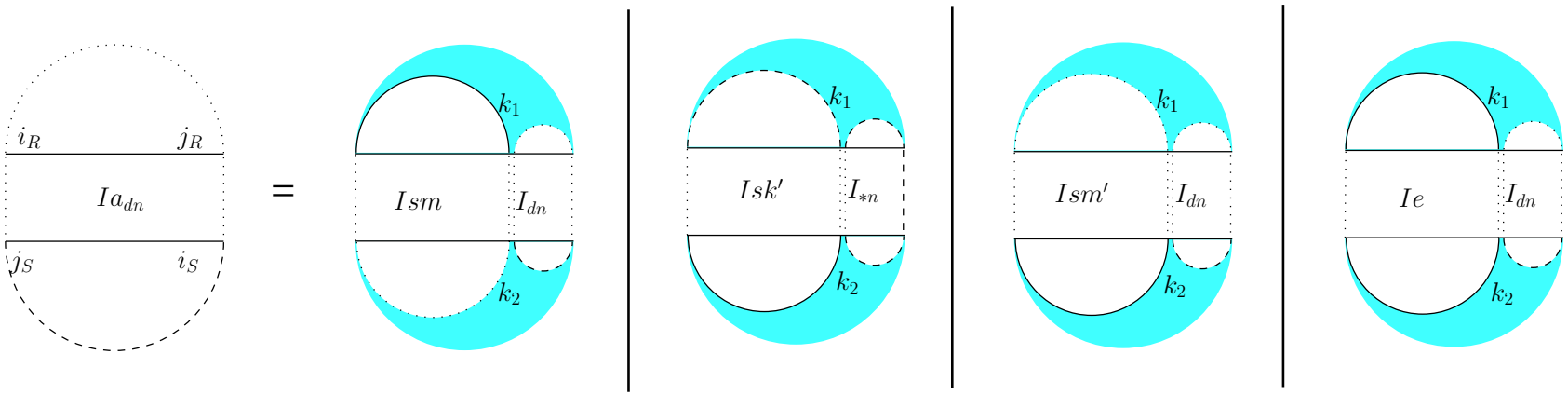
All tables



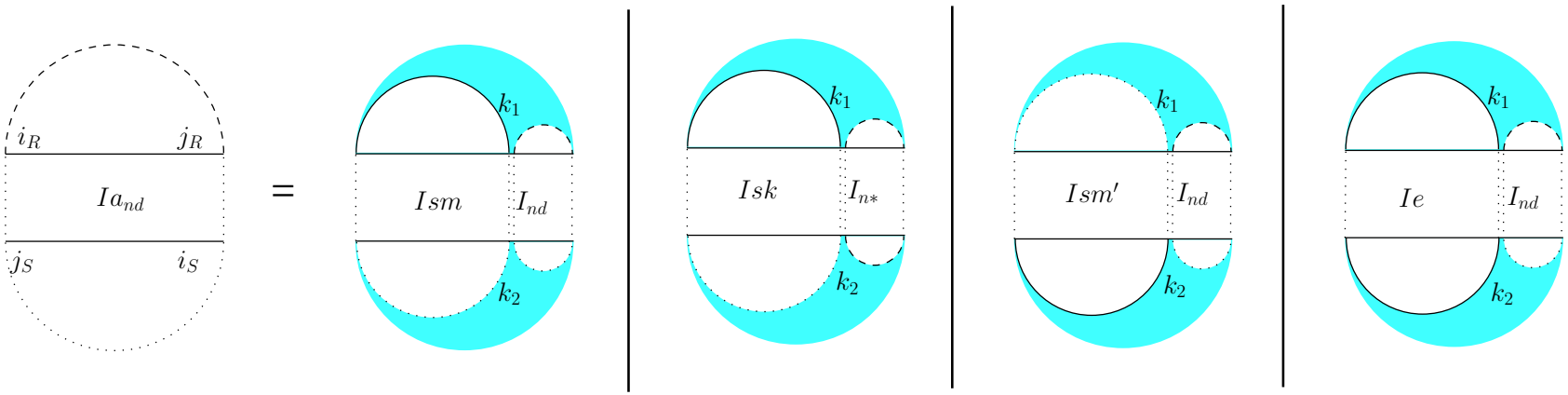
All tables


i_R
j_R
Ia_{dd}
j_S
i_S
=
k_1
Isk
I_{d*}
k_2
k_1
Ism
I_{dd}
k_2
k_1
Isk'
I_{*d}
k_2
k_1
Ism'
I_{dd}
k_2
k_1
Ie
I_{dd}
k_2

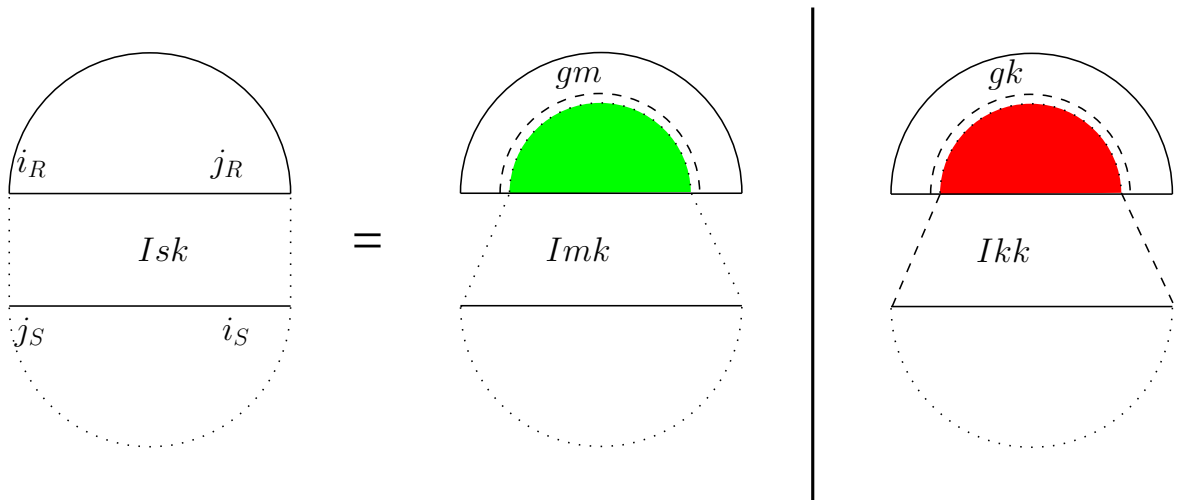
All tables


i_R
j_R
Ia_{dn}
j_S
i_S
=
k_1
Ism
I_{dn}
k_2
k_1
Isk'
I_{*n}
k_2
k_1
Ism'
I_{dn}
k_2
k_1
Ie
I_{dn}
k_2

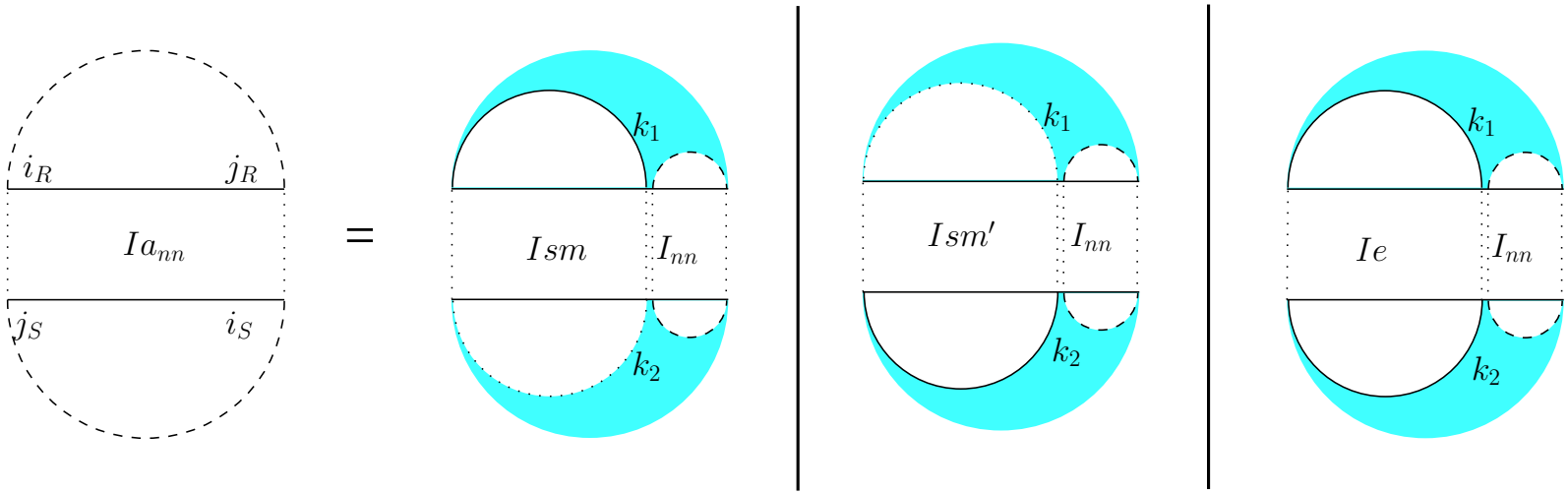
All tables


i_R
j_R
Ia_{nd}
j_S
i_S
=
k_1
Ism
I_{nd}
k_2
k_1
Isk
I_{n*}
k_2
k_1
Ism'
I_{nd}
k_2
k_1
Ie
I_{nd}
k_2

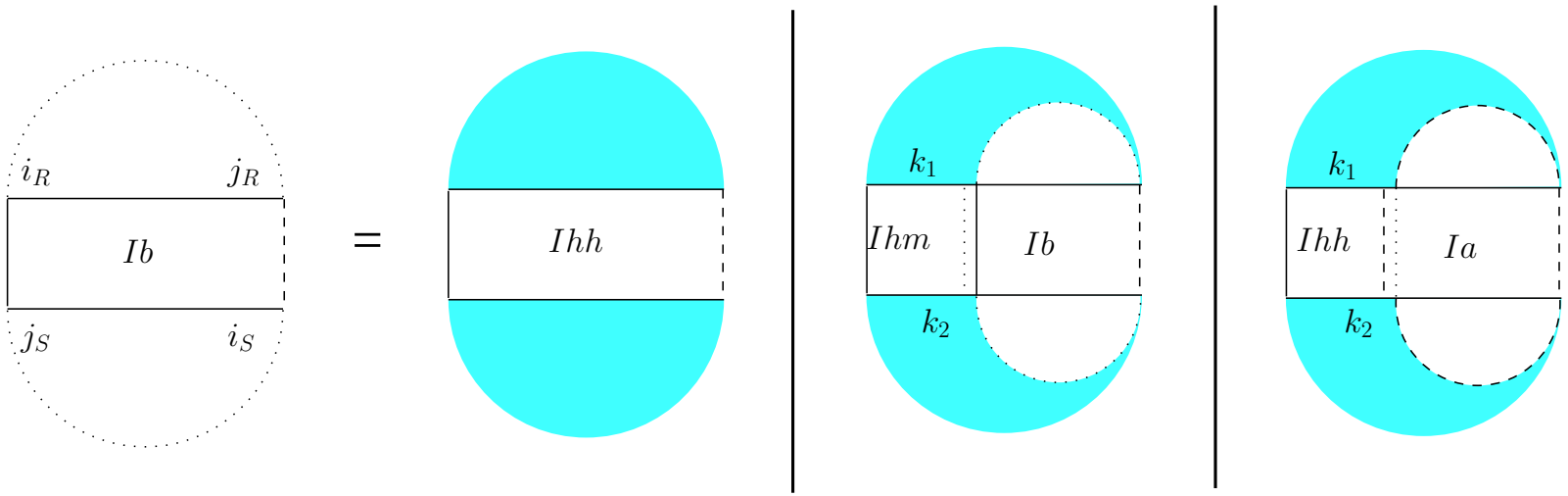
All tables



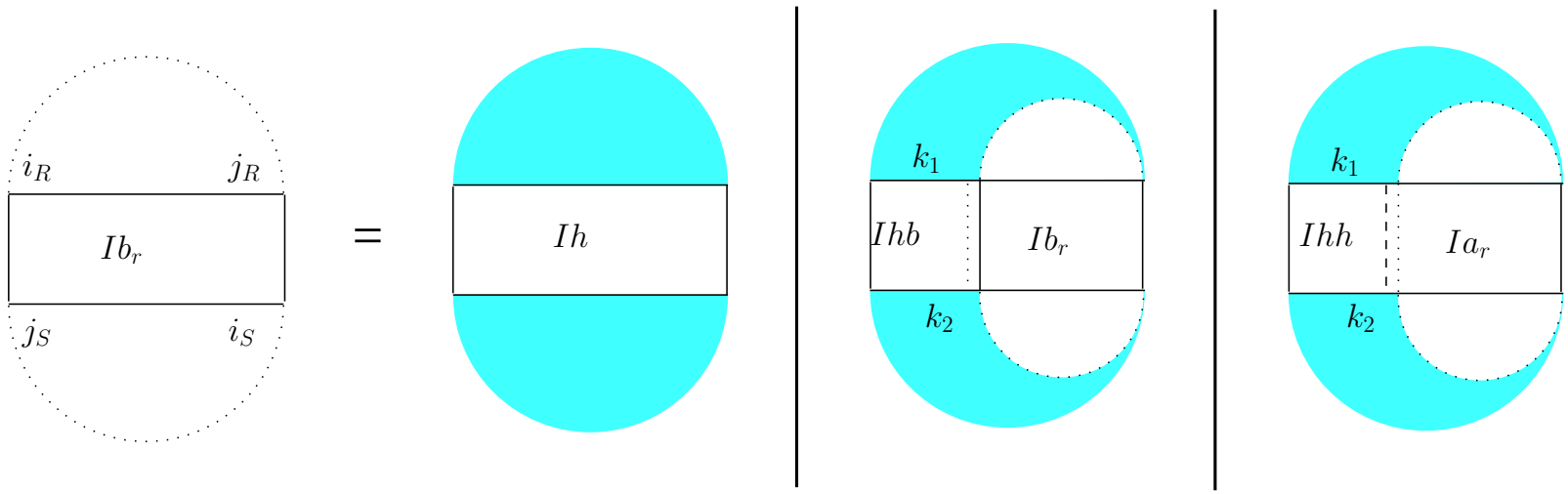
All tables



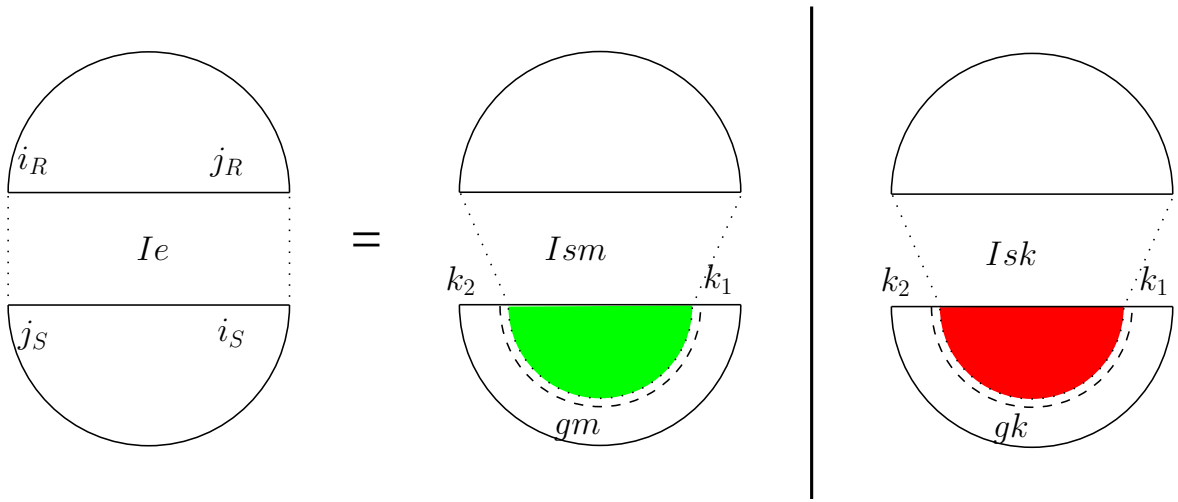
All tables



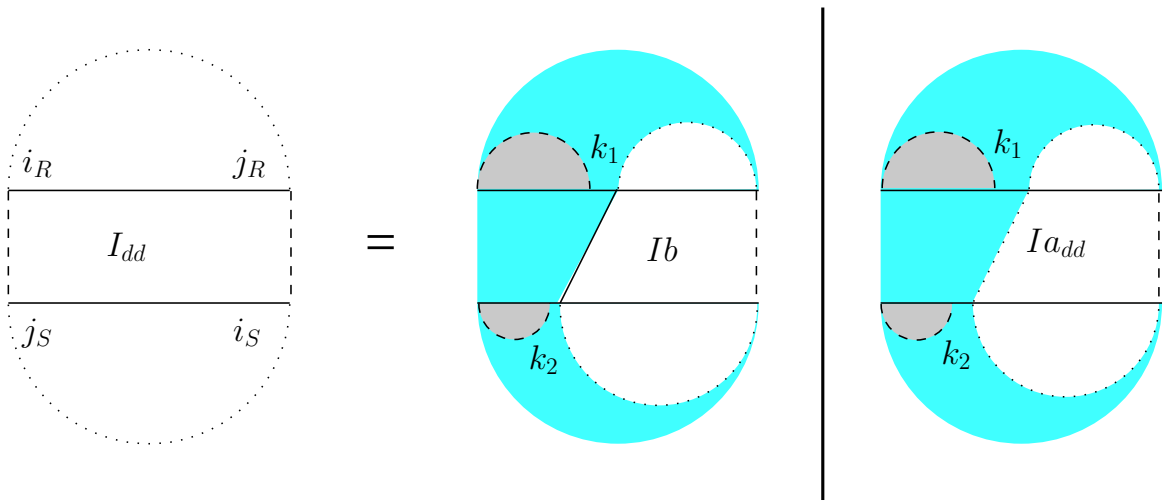
All tables



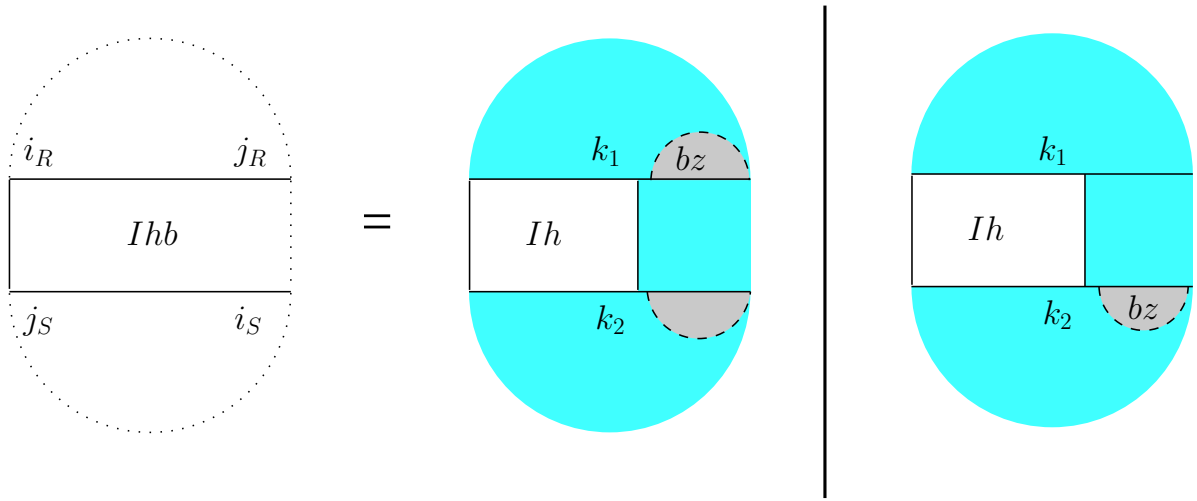
All tables



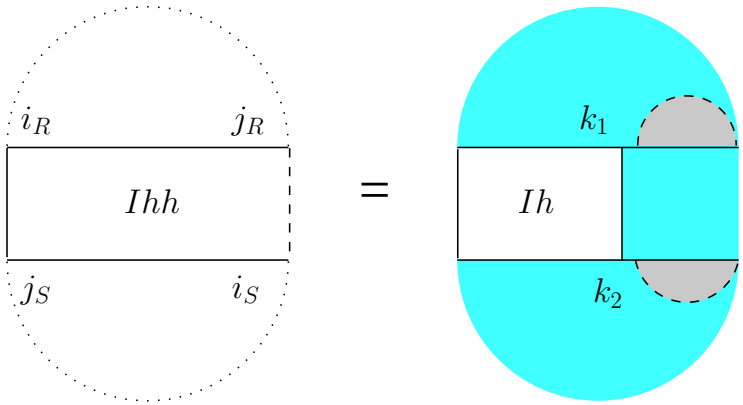
All tables



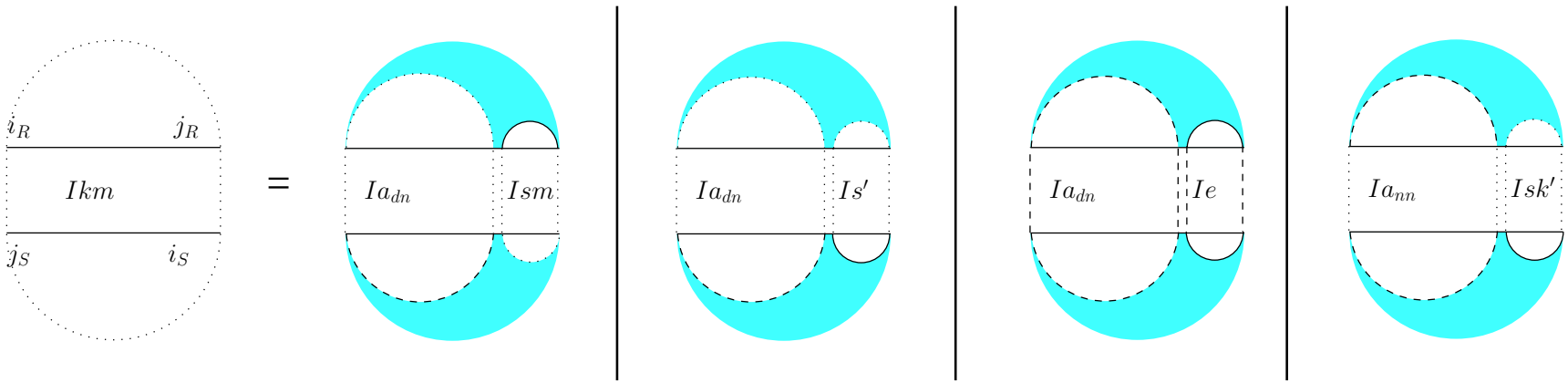
All tables



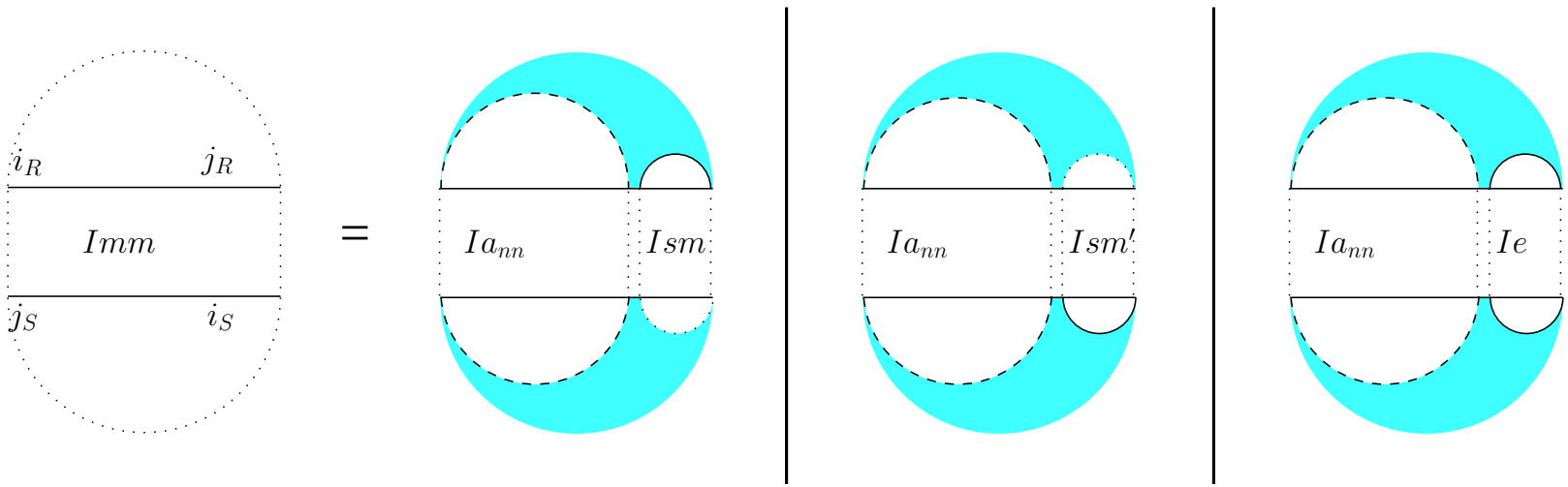
All tables



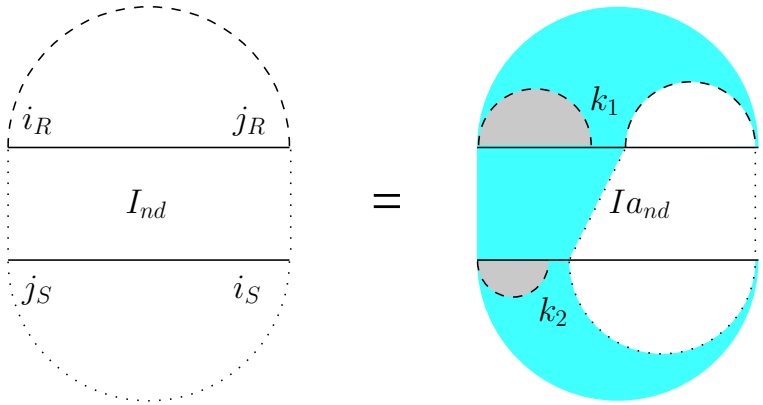
All tables



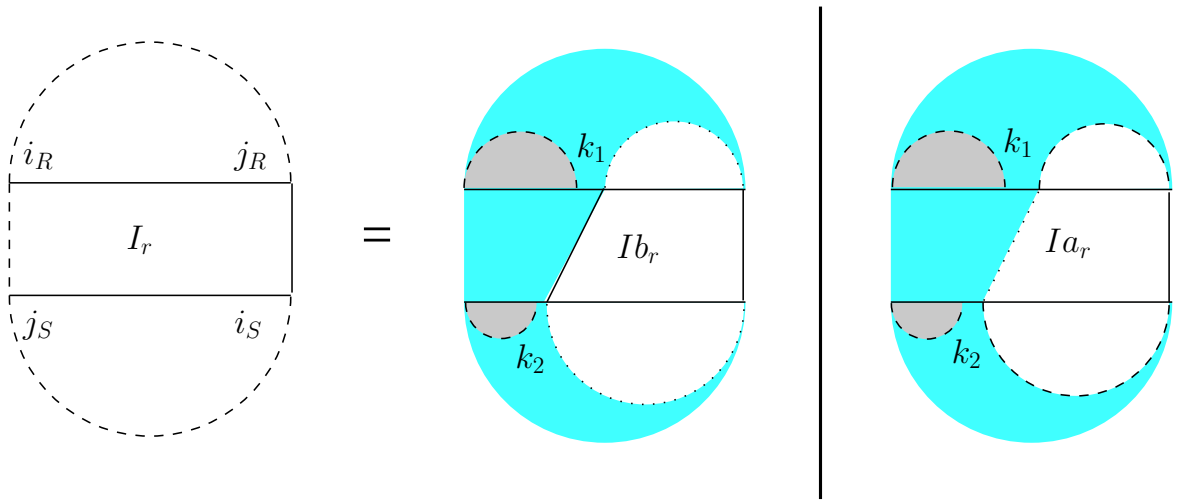
All tables



All tables



All tables



Equilibrium concentrations

For two RNAs $\mathbf{R}$ and $\mathbf{S}$

Assume five types of chemical compounds: **R**, **S**, **RR**, **SS**, **RS**.
Solve

$$K_{\mathbf{R}} = \frac{Q^I_{\mathbf{RR}}}{Q^2_{\mathbf{R}}} = \frac{N_{\mathbf{RR}}}{N^2_{\mathbf{R}}},$$

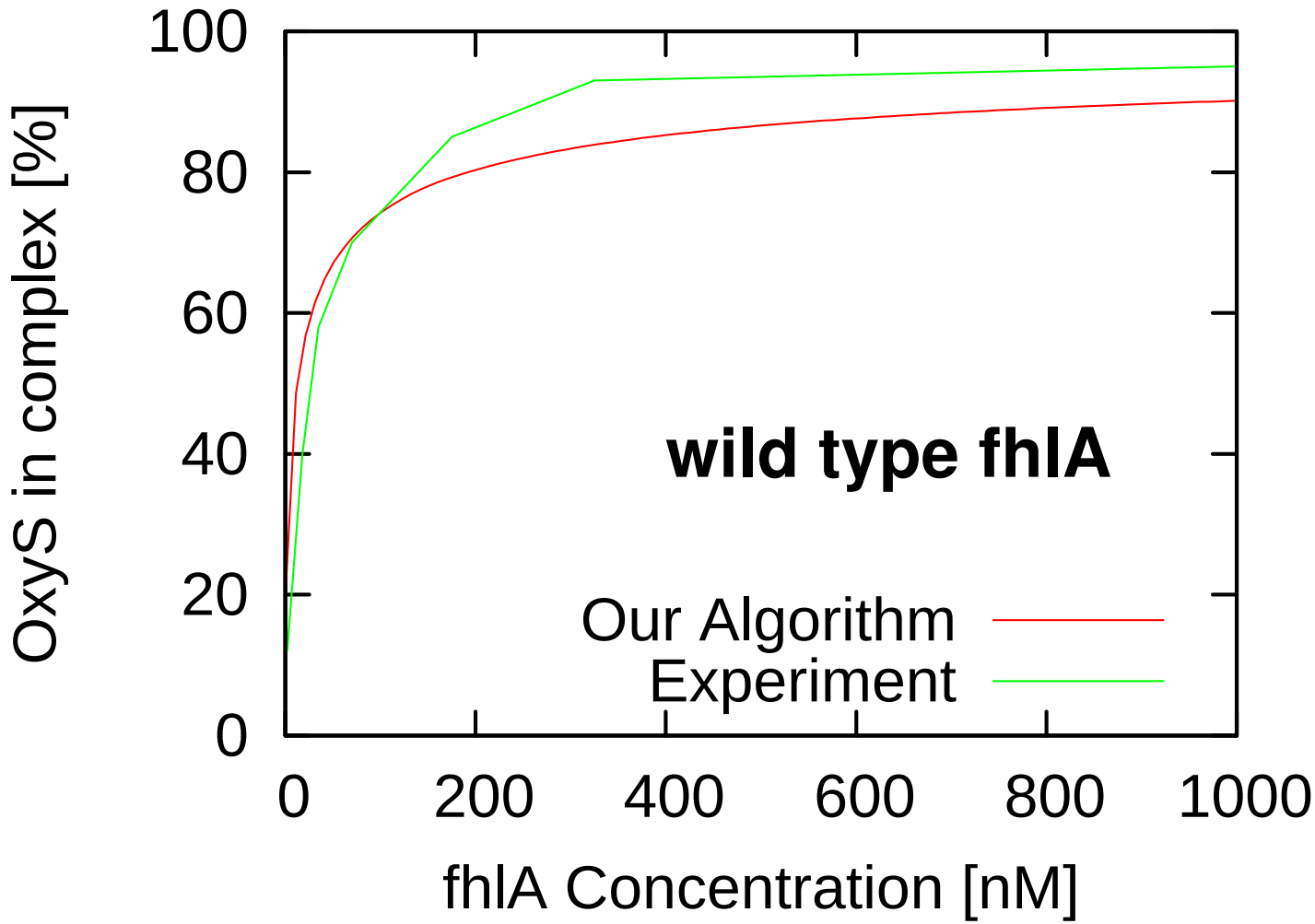
$$K_{\mathbf{S}} = \frac{Q^I_{\mathbf{SS}}}{Q^2_{\mathbf{S}}} = \frac{N_{\mathbf{SS}}}{N^2_{\mathbf{S}}},$$

$$K_{\mathbf{RS}} = \frac{Q^I_{\mathbf{RS}}}{Q_{\mathbf{R}} Q_{\mathbf{S}}} = \frac{N_{\mathbf{RS}}}{N_{\mathbf{R}} N_{\mathbf{S}}},$$

$$N_{\mathbf{RS}} = N^0_{\mathbf{R}} - 2N_{\mathbf{RR}} - N_{\mathbf{R}} = N^0_{\mathbf{S}} - 2N_{\mathbf{SS}} - N_{\mathbf{S}},$$

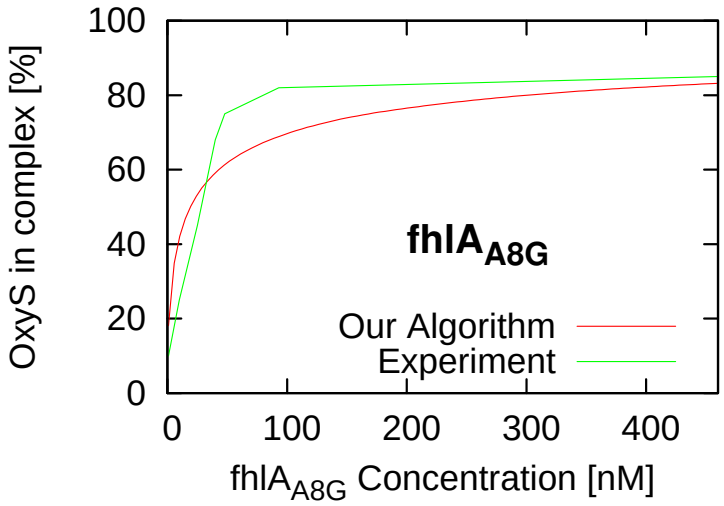
to obtain the equilibrium concentrations N. N^0 are the initial concentrations of single strands.

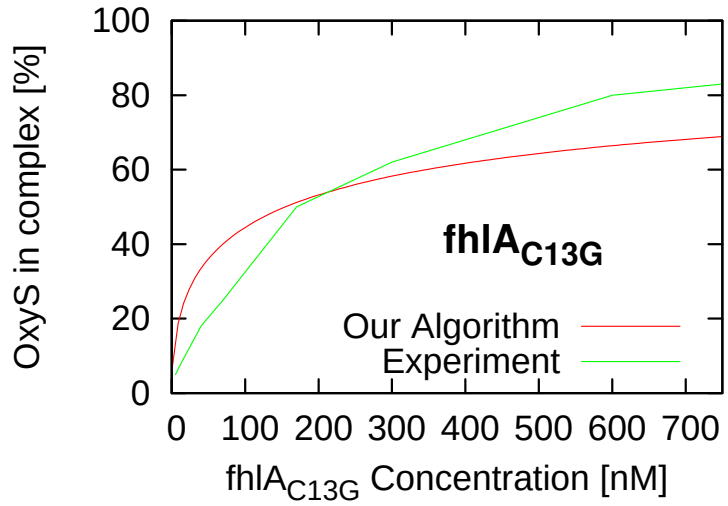
Equilibrium concentration of OxyS with wild type fhlA

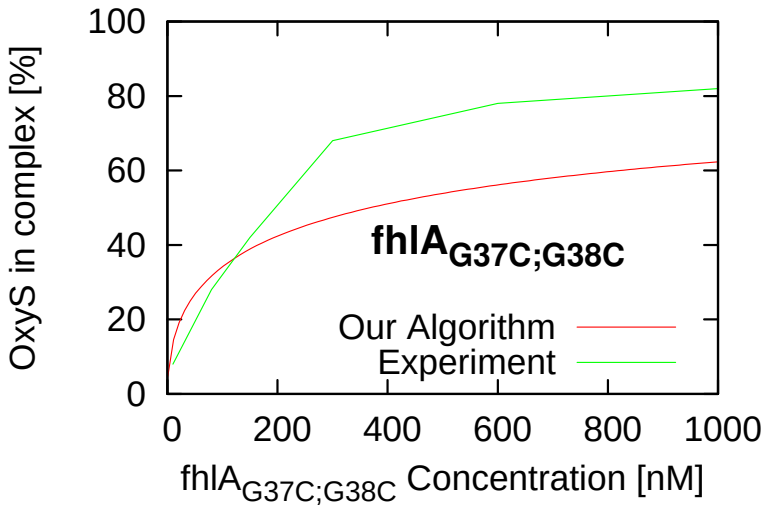


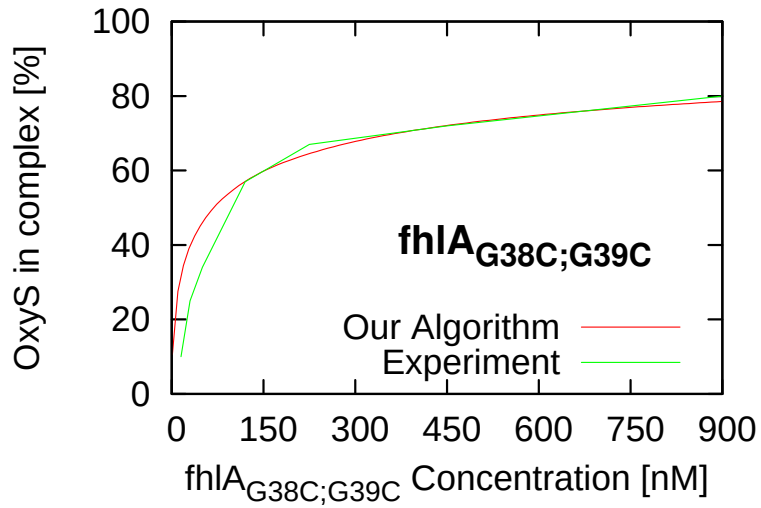
Init. [OxyS] = 2nM, [fhlA] = 0 to 1000nM

Equilibrium concentration of OxyS with fhlA mutants









Melting temperature prediction

Comparison of piRNA results over three data sets

Set	Size	Length	Avg error		
			piRNA	RNAcofold	UNAFold
I	9 short pairs	5-7nt	**1.48**°C	9.35°C	8.55°C
II	12 pairs	$\sim$ 20nt	**4.86**°C	22.97°C	9.12°C
III	62 pairs	$22-40$nt	**1.91**°C	14.34°C	26.53°C

Set	Size	Length	Spearman rank correlation		
			piRNA	RNAcofold	UNAFold
I	9 short pairs	5-7nt	**0.97**	**0.97**	0.57
II	12 pairs	$\sim$ 20nt	**0.41**	-0.03	0.1
III	62 pairs	$22-40$nt	**0.3**	-0.04	0.24