Irreducible Surfaces

Elena Rivas[1] **and Sean R. Eddy**[2]

Department of Genetics,

Washington University, St. Louis, MO 63130 USA

Supplementary material for

"A dynamic programming algorithm for
RNA structure prediction including pseudoknots"
by Elena Rivas and Sean R. Eddy.

[1]Email: elena@genetics.wustl.edu.
[2]Email: eddy@genetics.wustl.edu.

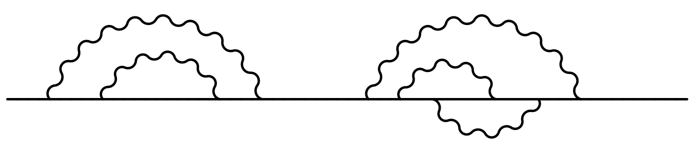
In this supplementary material we are give a more precise mathematical definition and some properties of Irreducible Surfaces (ISs).

Irreducible Surfaces are of importance because they constitute the basic elements of RNA folding: hairpin loops, stems, bulges, internal loops, and multiloops. ISs can also be pseudoknotted. In this regard, ISs are an extension of the "k–loops" introduced by Sankoff (1985) which are inherently nested.

Our algorithm is an expansion in ISs, that is, it is an attempt to express a given full structure in terms of its basic constituents, much in the same way as Schwinger-Dyson diagrams do for quantum field theory. In this supplementary material we introduce some relevant concepts about the topology of the graphs we are using.

Definition 1 (Graph) A graph is any continuous fragment of RNA that folds into an arbitrary non-trivial secondary structure.

Example of a graph,



(1)

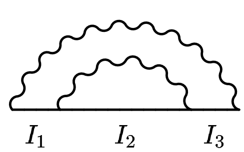
Definition 2 (Irreducible Graph) A graph is irreducible when the RNA backbone cannot be disrupted, without disrupting any secondary interactions. The above graph is reducible to two irreducible components.

Definition 3 (Interval) We define an interval, I, of an irreducible graph (IG) as set of points in between two secondary interactions, including the paired vertices.

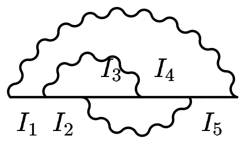
Definition 4 (Order of an IG) If a given IG has γ intervals, the graph is said to be of order γ, IG^γ.

Example, the graph in (1) contains two irreducible graphs,

$IG^3 =$



has 3 intervals (2)

$IG^5 =$ 

has 5 intervals (3)

Definition 5 (Surface) A surface of order γ, S^γ, is a collection of intervals (possibly disjoint) such that moving from interval to interval through a secondary interaction it closes in itself.

$S^\gamma(I_1, I_2, ..., I_\gamma)$ indicates that there is a closed path from $I_1 \to I_2 \to ... \to I_\gamma \to I_1$ via secondary interactions at the end of each interval. Secondary interactions may be covered more than once in order to complete the surface. When an interval has to be covered in the $3'$-end to $5'$-end, that is indicated by the transposition symbol I^T.

Examples, the graph in (2) includes surfaces $S^1(I_2)$, $S^2(I_1, I_3)$, and $S^3(I_1, I_2^T, I_3)$. The graph in (3) includes surfaces $S^2(I_2, I_4^T)$, $S^3(I_1, I_3^T, I_5)$, and $S^5(I_1, I_4, I_3, I_2, I_5)$.

Property 1 An IG of order γ contains always at least one surface which has order γ.

Theorem 1 For a collection of intervals to define a surface, the exterior edges have to H-bond to each other.

Proof:

Consider a collection of intervals with 2γ edges. The number of interior edges has to be even in order to have a surface (*i.e.* a closing path) since no more than one H-bond can go to an edge (we do not allow interaction of more than two nucleotides at once). Let the number of interior edges be $2n$. Then

$$\#\text{order of graph} = \gamma + 2 * n; \quad \#\text{H-bonds} = \gamma + n. \tag{4}$$

In order to have a surface, $(\gamma + 2 * n) - (\gamma + n) = n$ H-bonds have to be covered twice. Therefore $2n$ intervals have connect to each other, which means that the interior edges bond to each other, so the exterior edges also have to bind to each other in order to define a surface.

Definition 6 (Irreducible Surface) An irreducible surface of order γ (IS^γ) is a surface S^γ such that if one of the secondary interactions is interrupted there is not any other surface that covers the same space $I_1 \cup I_2 \cup ... \cup I_\gamma$.

Examples, surfaces $S^1(I_2)$ and $S^2(I_1, I_3)$ for the graph in (2), and surfaces $S^2(I_2, I_4)$ and $S^3(I_1, I_3, I_5)$ for the graph in (3) are irreducible.

Theorem 2 A surface is irreducible if and only if the intervals that constitute the surface are disjoint.

Proof:

$\Leftarrow$ If the intervals are disjoint, no H-bond is shared by any two intervals, therefore if an H-bond is broken, the surface becomes an open path.

$\Rightarrow$ Suppose a surface S^γ in which two intervals I_i, I_j are not disjoint. Then, intervals I_i and I_j share an H-bond. Because the surface closed in itself by definition, there is another pair of non-disjoint intervals I'_i and I'_j that share the other end of the H-bond. Now break that common interaction, then $S^{\gamma-2}(I_1, ..., I_i \cup I_j, ..., I'_i \cup I'_j, ..., I_\gamma)$ is a surface.

Corollary 1 A surface is irreducible if and only if no H-bond has to be crossed more than one time in traversing the surface.

Proof:

The intervals are disjoint if and only if no H-bond has to be crossed more than once.

Property 2 A surface of order γ always contains at least one irreducible surface IS^β, such that $\beta \leq \gamma$.

Proof:

Suppose that S^γ has $\beta \leq \gamma$ H-bonds $\Rightarrow \exists \beta$ disjoint intervals that form a surface $\Rightarrow$ there is an $IS^\beta \subset S^\gamma$. This IS^β is the one of higher order that a IG^γ can have.

In summary, our algorithm is a recursion for generic graphs (the wx, vx, vhx, whx,... matrices) in terms of irreducible surfaces. Properties 1 and 2 assure us that for a given generic graph there is a recursion possible in terms of ISs, and that the recursion has a finite number of terms.