Relevant Cycles in Biopolymers and Random Graphs

Petra M. Gleiss [a] and Peter F. Stadler [a,b]

[a] *Institute for Theoretical Chemistry and Molecular Structural Biology*
University of Vienna, Währingerstrasse 17, A-1090 Vienna, Austria
`{pmg,studla}@tbi.univie.ac.at`

[b] *The Santa Fe Institute, 1399 Hyde Park Rd, Santa Fe NM 87501, USA*
`stadler@santafe.edu`

Abstract

Short cycles are an important characteristic of molecular graphs in organic chemistry as well as in structural biology. Minimum cycle bases are of particular interest, despite the fact that they are usually not unique. Hence, one sometimes resorts to the set relevant cycles, defined as the union of all minimum cycles bases. Here we introduce the set of essential cycles as the intersection of a graph's minimum cycle bases and provide an algorithm for their computation. Furthermore, we extend previous bounds on the length of minimal cycles bases to certain book-embeddable graphs.

Key words: Minimal Cycle Basis, Relevant Cycles, Essential Cycles, Biopolymer Graphs.

Subject Classification: 05C38, 05C85.

---

1 Introduction

Organic carbon compounds, such as the example shown in Figure 1, may exhibit elaborate polycyclic structures. Biopolymers, such as RNA, DNA, or proteins form well-defined three dimensional structures which are of utmost importance for their biological function. The most salient features of these structures are captured by their *contact graphs* which have the atoms of small molecules or the monomers of a biopolymer as their vertices, and edges that connect spatially adjacent objects. While this simplification of the 3D shape obviously neglects a wealth of structural details, it encapsulates the type of structural information that can be obtained by a variety of experimental and computational methods.

Fourth Slovene International Conference in Graph Theory
Bled, June 28 – July 8 1999

25 June 1999

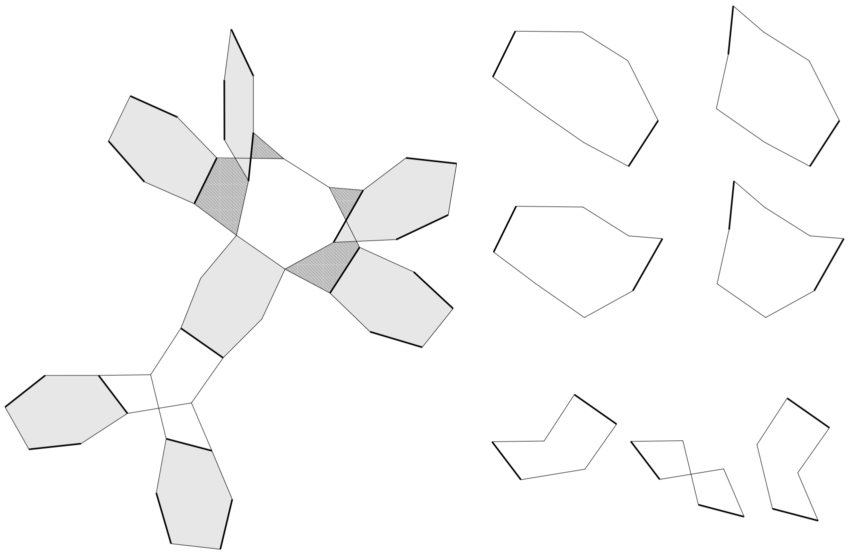

Fig. 1. Compound 8 from [19] (aromatic “double bonds” are indicated by thick lines). The nine essential cycles are marked with gray shades. There are two groups of relevant non-essential cycles: four 8-rings and three 6-rings. A minimal cycle basis contains two of the three 6-rings and one of the four 8-rings.

Biopolymers share a number of common features distinguishing them from other classes of the molecular contact graphs. In particular, they have spanning path T corresponding to the covalent backbone. The remaining non-covalent bonds $B = E \setminus T$ then determine the “fold” or three-dimensional structure of the molecule. Nucleic acids, both RNA and DNA, often form a special type of contact structures known as *secondary structures*. These graphs are outer-planar and sub-cubic, i.e., the maximal vertex degree is 3.

The description of cyclic structures is an important part of graph theory as exemplified e.g. by the book [31]. Naturally, short cycles are particularly useful for this purpose. Minimal cycles bases (MCB) are of particular practical interest because they encapsulate the entire cycle space in concise manner. In section 3 we extend previous work [22] on the length of minimal cycle bases to more general graphs.

The MCB of a secondary structure graph is unique [22]. Its special role for the standard energy model of RNA folding is described for instance in [10, 32]. In recent years, however, there has been increasing evidence that so-called pseudo-knots play an important role, see e.g. [11]. These structural elements violate outer-planarity and — in the simplest case — lead to the *bisecondary structures* introduced in [26]. These graphs can be embedded into a book with 2 pages in such a way that the spine is formed by the spanning path T. The MCB is not unique for most graphs, including most non-trivial bisecondary structures.

Phillipe Vismara [30] considers the set $\mathcal{R}$ of *relevant cycles*, defined as the union of all minimal cycle bases, as a natural extension. Relevant cycles, and the related set of shortest cycles (section 5) have a variety of applications in science and engineering, among them in structural analysis [20] and in chemical structure storage and retrieval systems [8]. In section 4 we introduce the set of *essential cycles* as the intersection of all minimum cycle basis, see also Figure 1, and we present an algorithm for its efficient computation.

One of the oldest results on cycle bases [27, 34, 35] relates minimal cycles bases and the shortest cycles containing a given edge. This connection is briefly reviewed in section 5. Finally, we present a few computational results concerning the expected size of the sets of relevant and essential cycles, respectively.

2 Notation

Let Γ be a finite, loop-free, undirected, simple graph with vertex set V and edge set E. A graph Γ' with vertex set $V' \subseteq V$ and edge set $E' \subseteq E$ is a *subgraph* of Γ. A subgraph is *induced* if, for any two vertices $x, y \in V'$ we have $\{x, y\} \in E'$ if and only if $\{x, y\} \in E$. We write $\Gamma' < \Gamma$.

For the purpose of this paper we shall say that a graph Γ is a *biopolymer graph* if it has a spanning path. If Γ is Hamiltonian, i.e., if there is a spanning path T and an edge $e \in B$ such that $H = T \cup \{e\}$ is a Hamiltonian cycle we shall speak about circular biopolymer graphs. Just like "linear" biopolymers, such structures frequently occur in nature.

A *generalized* cycle in Γ is the edge set C of a subgraph $\Gamma' < \Gamma$ in which every vertex has an even degree. A *cycle* in Γ is the edge set of a minimal (equivalently: connected) subgraph of Γ which has only vertices of degree 2. We shall write $V[C]$ for the vertex set of Γ', i.e., for the set of vertices that are connected by the edges in C.

The set $\mathcal{E}$ of all subsets of E forms an m-dimensional vector space over GF(2) with vector addition $X \oplus Y := (X \cup Y) \setminus (X \cap Y)$ and scalar multiplication $1 \cdot X = X$, $0 \cdot X = \emptyset$ for all $X, Y \in \mathcal{E}$. The set $\mathcal{C}$ of all generalized cycles forms a subspace of $(\mathcal{E}, \oplus, \cdot)$ which is called the *cycle space* of Γ. A basis $\mathcal{B}$ of the cycle space $\mathcal{C}$ is called a *cycle basis* of Γ [4]. The dimension of the cycle space is the *cyclomatic number* or *first Betti number* $\nu(\Gamma) = |E| - |V| + 1$.

The length $|C|$ of a cycle C is the number of its edges. Two quantities associated with a cycle basis $\mathcal{B}$ are of particular interest: its length $\ell(\mathcal{B})$ and the size of its largest cycle $c(\mathcal{B})$ defined as

$$\ell(\mathcal{B}) = \sum_{C\in\mathcal{B}} |C| \qquad\qquad c(\mathcal{B}) = \max_{C\in\mathcal{B}} |C| \tag{1}$$

$\ell(\mathcal{B})$ is minimal if and only if $c(\mathcal{B})$ is minimal [5, Thm.4]. Such a basis is called a *minimal cycle basis* of Γ; its length will denote by $\ell(\Gamma)$. Let $u_{\mathcal{B}}(l)$ denote the number of cycles of length l in a basis $\mathcal{B}$. Suppose $\mathcal{M}$ and $\mathcal{M}'$ are to minimal cycle bases. Then $u_{\mathcal{M}}(l) = u_{\mathcal{M}'}(l)$ for all l [5, Thm.3].

A p-book $\mathfrak{B}$ is a set of p distinct half-planes (the *pages* of the book) that share a common boundary line ℓ, called the *spine* of the book. An embedding of a graph Γ into a book $\mathfrak{B}$ consists of an ordering of the vertices along the spine of the book together with an assignment of each edge to a page of the book, in which edges assigned to the same page do not cross. If Γ has a spanning path T and the vertices are arranged along the spine in their order of occurrence along T, we shall say for simplicity that T is the spine of the book embedding.

3 An Upper Bounds on $\ell(\Gamma)$

A sharp upper bound on $\ell(\Gamma)$ is proved in [16, Thm.6]:

$$\ell(\Gamma) \le \ell(K_{|V|}) = 3(|V|-1)(|V|-2)/2 \tag{2}$$

For 2-connected outerplanar and planar graphs we have $\ell(\Gamma) \le 3|V| - 6$ and $\ell(\Gamma) \le 6|V| - 15$, respectively [22, Thm.11]. The global upper bound $\ell(\Gamma) \le \nu(\Gamma) + \kappa(\mathsf{T}(\Gamma))$, where $\kappa(\mathsf{T}(\Gamma))$ is the connectivity of the tree graph of Γ, is derived in [23].

The behavior of $\ell(\Gamma)$ under most graph operations is hard to predict. The deletion of a single edge, for instance, can drastically increase $\ell(\Gamma)$, see Figure 2.

In the case of biopolymer graphs we can construct a decomposition into outerplanar graphs which will enable us to derive an upper bound on $\ell(\Gamma)$.

Definition 1 *Let* $\Gamma = (V, E)$ *be a graph with spanning path* T*. Consider a partition* $\{B_1, B_2, \dots, B_\beta\}$ *of* $B = E \setminus T$ *such that* $\Gamma_k = (V, B_k \cup T)$ *is outerplanar. We call the subgraph* Γ_k *of* Γ *an* outerplanar constituent *and write* $\Gamma = \Gamma_1 \vee \Gamma_2 \vee \cdots \vee \Gamma_\beta$*.*

Note that $\Gamma = \bigvee_{k=1}^{\beta} \Gamma_k$ is embeddable in a β-book $\mathfrak{B}$ with spine T. The bisecondary structure graphs introduced in [26] are exactly those that have at most two outerplanar constituents. Equivalently, they are characterized as subgraph of planar Hamiltonian graphs [1].

Fig. 2. The l.h.s. graph has $\nu(\Gamma_1) = 3$ and $\ell(\Gamma_1) = 38$. Deleting a single edge leads to a r.h.s. graph with $\nu(\Gamma_2) = 2$ but $\ell(\Gamma_2) = 44$.

Theorem 2 *Let* $\Gamma = \bigvee_{k=1}^{\beta} \Gamma_k$. *Then:*

$$\ell(\Gamma) \leq \sum_{k=1}^{\beta} \ell(\Gamma_k) \tag{3}$$

Proof. First we observe that Γ_k is connected for $1 \leq k \leq \beta$, hence $\nu(\Gamma_k) = |T| + |B_k| - |V| + 1 = |B_k|$, while $\nu(\Gamma) = |B| + |T| - |V| + 1 = |B| = \sum_k |B_k|$, i.e., $\nu(\Gamma) = \sum_{k=1}^{n} \nu(\Gamma_k)$.

The minimal cycle bases $\mathcal{M}_k$ of the outerplanar components are Γ_k are easily constructed: they are given by the faces of the outerplanar embeddings [22]. Each of these cycles contains at least one edge in B_k and none of the edges in B_l, $l \neq k$, whence $\mathcal{M} = \bigcup_{k=1}^{\beta} \mathcal{M}_k$ is a set of independent cycles of Γ containing $\sum_k |\mathcal{M}_k| = \sum_k \nu(\Gamma_k) = \nu(\Gamma)$ cycles. In other words, $\mathcal{M}$ is a cycle basis of Γ. Equation (3) now follows from $\ell(\mathcal{M}) = \sum_k \ell(\mathcal{M}_k) = \ell(\Gamma_k)$. $\square$

In order to derive a bound in terms of $|V|$ and $|E|$ from theorem 2 we need the following technical

Lemma 3 *Let* $\Gamma = (V, E)$ *be a connected outerplanar graph. Then* $\ell(\Gamma) \leq 2|E| - |V|$. *Equality holds if and only if* Γ *is 2-connected.*

Proof. We write $\psi = 2|E| - |V|$. A connected out-planar graphs consists of 2-connected components Γ_i and trees T_j that connect the components. The trees can be further subdivided into paths P_k that connect 2-connected components and end trees, that have exactly one vertex in common with a path P_k or a 2-connected component Γ_i.

(i) Let U be an end tree with vertex of attachment z. For the graph (V', E') with $V = V \setminus (V[U] \setminus \{z\})$ and $E' = E \setminus U)$ we have $|V'| = |V| - |U|$ and $|E'| = |E| - |U|$. Hence $\psi' = \psi - |U|$, i.e., ψ strictly decreases by removing "end trees".

(ii) Suppose Γ has no end trees. Then removing a connecting path P_i increases the number of components by 1, $c' = c + 1$. Now consider a path P attached at the vertices u and v. The graph (V', E') with $V' = V \setminus (V[P] \setminus \{u, v\})$ and $E' = E \setminus U$ has $|V'| = |V| - |P| + 1$ vertices and $|E'| = |E| - |P|$ edges. Hence $\psi' = \psi - |P| - 1$, i.e., ψ strictly decreases by removing connecting paths.

After removing all end trees and connecting path, the remaining 2-connected components may be attached to each other by a common vertex. Splitting this vertex increases the number of vertices by 1 and leaves the edges unchanged. We may repeat this operation until we are left with a disjoint union of 2-connected outerplanar graphs. The final value of ψ' of course equals the sum of the ψ-values for each of the components.
The MCB of a 2-connected outerplanar graph has length ψ, see e.g. [28, 22]. Thus $\ell(\Gamma) = \psi$ if and only if Γ is a disjoint union of 2-connected outerplanar graphs, and the lemma follows. $\square$

Theorem 4 *If* $\Gamma = \bigvee_{k=1}^{\beta} \Gamma_k$, *then*

$$\ell(\Gamma) \le 2|E| + (\beta - 2)(|V| - 1) \le \beta(3|V| - 5) \tag{4}$$

The first inequality is strict for $\beta > 1$.

Proof. The MCB of a 2-connected outerplanar graph has length $2|E| - |V| = 2|T| + 2|B| - |V| = 2(|V| - 1) + 2|B| - |V| = |V| - 1 + 2|B|$. The inequality follows immediately. Since at most one of the outerplanar components Γ_k can be 2-connected, $\ell(\Gamma_k) = 2|E_k| - |V_k|$ for at most one Γ_k. Thus $\ell(\Gamma) \le \beta(|V| - 1) + 2|B|$. The corollary now follows from $|B| = |E| - |T| = |E| - (|V| - 1)$. We have $|E| = |T| + \sum_{k=1}^{\beta}(|E_k| - |T|)$, where E_k are the edge sets of the outerplanar constituents. For all outerplanar graphs holds $|E| \le 2|V| - 3$ [7]. The second inequality now follows from a short computation. $\square$

For $\beta = 2$ the bound can be further improved since Γ is planar in this case, which implies $\ell(\Gamma) \le 2|E| - g(\Gamma)$ where $g(\Gamma) \ge 3$ is the girth of Γ [22].

4 Relevant Cycles and Essential Cycles

If Γ is outerplanar, then its minimal cycle basis is unique [22]. This result can be extended to a larger class of series-parallel graphs [24]. In general, of course, we do not have uniqueness. In chemical ring perception and in the context of biopolymer structures the *union of the minimum cycle bases* $\mathcal{R}$ appears to be more suitable than using a particular minimum cycle basis [29, 30]. The set $\mathcal{R}$ consists exactly of the *relevant cycles* [25], which cannot be represented as a $\oplus$-sum of shorter cycles.

Vismara [29, 30] proposed an algorithm for computing $\mathcal{R}$ that works by first extracting so-called prototypes from a set of short cycles similar to Horton's algorithm for finding *a* MCB [16]. The computation of the prototypes requires $\mathcal{O}(|E|^3\nu(\Gamma))$ operations. The set $\mathcal{R}$ is then obtained by a backtracking procedure from the prototypes with $\mathcal{O}(|V|\,|\mathcal{R}|)$ operations. For *some* classes of graphs $|\mathcal{R}|$ grows exponentially with $|V|$, see [30] for an example. However, in

Table 1

Algorithm: extract $\mathcal{J}$ from $\mathcal{R}$.

```
Input: $\mathcal{R}$
$k = 3$; $\mathcal{B} \leftarrow \emptyset$; $\mathcal{R}_= \leftarrow \emptyset$; $\mathcal{J} \leftarrow \emptyset$;
repeat
    $C \leftarrow$ shortest cycle in $\mathcal{R}$; $\mathcal{R} \leftarrow \mathcal{R} \setminus \{C\}$.
    if $|C| > k$ or $\mathcal{R} = \emptyset$ then
        $r = |\mathcal{B}|$        /* Rank of $\{C \in \mathrm{MCB} : |C| \leq k\}$ */
        for all $C' \in \mathcal{B}_=$ do
            if $\mathrm{rank}[(\mathcal{B} \cup \mathcal{R}_=) \setminus \{C'\}] < r$ then
                $\mathcal{J} \leftarrow \mathcal{J} \cup \{C'\}$
        $k \leftarrow |C|$; $\mathcal{R}_= \leftarrow \emptyset$; $\mathcal{B}_= \leftarrow \emptyset$;
    if $\mathcal{R} = \emptyset$ then
        return $\mathcal{J}$
    $\mathcal{R}_= \leftarrow \mathcal{R}_= \cup \{C\}$;
    /* Extract an MCB */
    if $\mathcal{B} \cup \{C\}$ independent then
        $\mathcal{B} \leftarrow \mathcal{B} \cup \{C\}$; $\mathcal{B}_= \leftarrow \mathcal{B}_= \cup \{C\}$;
until
```

the final section of this contribution we report computational evidence that, typically, $|\mathcal{R}|$ is not too much larger than the minimal possible value $\nu(\Gamma)$.

Lemma 5 *If* Γ *contains* K_4 *as a subgraphs then* $\mathcal{R}$ *is dependent, i.e.,* $|\mathcal{R}| > \nu(\Gamma)$.

Proof. K_4 contains four triangles, each of which is a relevant cycle of any graph containing the K_4. From $\nu(K_4) = 3$ we conclude immediately that one of them is the $\oplus$-sum of the other three. □

Definition 6 *The* set $\mathcal{J}$ of essential cycles *is the intersection of all minimal cycles bases.*

Note that $\mathcal{J}$ can be empty. As an example consider the complete graph K_4, see Lemma 5. Similarly, $\mathcal{J} = \emptyset$ for larger complete graphs. Not surprisingly, $\mathcal{J}$ can be computed rather easily from $\mathcal{R}$.

Lemma 7 *Let* $\mathcal{B}$ *be a minimal cycle basis of* Γ, $\mathcal{B}_k = \{|C| \in \mathcal{B} : |C| < k\}$, $\mathcal{R}_k = \{|C| \in \mathcal{R} : |C| = k\}$, *and* $C \in \mathcal{R}_k$. *Then* $C \in \mathcal{J}$ *if and only if* $\mathrm{rank}[\mathcal{B}_k \cup \mathcal{R}_k \setminus \{C\}] < |\mathcal{B}_{k+1}|$.

Proof. By definition, $C \in \mathcal{J}$ if and only if $\mathcal{R} \setminus \{C\}$ does not contain a cycle basis. If $|C| = k$, it follows from the matroid properties of the cycle space that we have to consider only cycles up to length k. With $\mathcal{R}_{\leq k} = \bigcup_{j \leq k} \mathcal{R}_j$ we have $C \in \mathcal{J}$ if and only if $\mathrm{rank}[\mathcal{R}_{\leq k} \setminus \{C\}] < \mathrm{rank}[\mathcal{R}_{\leq k}]$. The lemma now follows from $\mathrm{rank}[\mathcal{R}_{\leq k}] = \mathrm{rank}[\mathcal{B}_k \cup \mathcal{R}_k]$. □

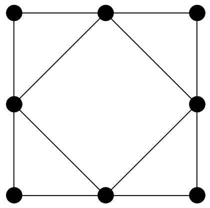

Fig. 3. An example with $\mathcal{S} \neq \mathcal{R}$ from [17]. This graph is outerplanar and hence has a unique minimal cycle basis consisting of the five cells. Since each edge is contained one of the triangles, $\mathcal{S}$ contains only the four triangles.

The algorithm for computing $\mathcal{J}$ from $\mathcal{R}$ is summarized in Table 1. Its worst case complexity is determined by the $|\nu(\Gamma)|$ rank computations in step 7, which in practice can be divided into two parts. Let $\mathcal{B}_{=k} = \mathcal{B}_{k+1} \setminus \mathcal{B}_k$ denote the set cycles with length k in the MCB. For each length k it suffices to perform a Gaussian elimination on $\mathcal{B}_k \cup (\mathcal{R}_k \setminus \mathcal{B}_{=k})$ once. This step requires at most $\mathcal{O}(|\mathcal{R}|\,|\mathcal{B}|\,|E|)$ operations. The ranks can now be computed by performing Gaussian elimination on the union of the result of the first step (which has only $\mathcal{O}(|\mathcal{B}|)$ rows) and $\mathcal{B}_{=k} \setminus \{C\}$ for each $C \in \mathcal{B}_{=k}$. For each C, this can be done with no more than $\mathcal{O}(|\mathcal{B}|^2|E|)$ steps. In the worst case, hence, $\mathcal{J}$ can be obtained in $\mathcal{O}(|\mathcal{R}|\,\nu(G)^2\,|E|)$ operations.

5 Shortest Cycles

One of the oldest results concerning minimum cycle bases is the following

Proposition 8 *[27, 34, 35] Let C be a cycle in Γ. If there is an edge $e \in C$ such that C is one of the shortest cycles that contain e, then $C \in \mathcal{R}$.*

Hence the *set of shortest cycles*

$$\mathcal{S} = \{C \mid \exists e \in E : C \text{ is a shortest cycle containing } e\} \tag{5}$$

is of interest. By definition, $\mathcal{S} \subseteq \mathcal{R}$. Note that $\mathcal{S} \neq \emptyset$, since for each edge $e \in E$ there is at least one shortest cycle. On the other hand, $\mathcal{S}$ need not contain a cycle basis as the example in Figure 3 shows.

Immediately, the question arises, how well is $\mathcal{R}$ approximated by the shortest cycles $\mathcal{S}$. We briefly mention two infinite classes of graphs for which $\mathcal{R} = \mathcal{S}$.

David Hartvigsen and Russel Mardon considered the minimum cycle basis problem for graph with *perturbed edge weights* $w(e)$, $e \in E$ which are chosen such that any two distinct edge-(multi)sets have different total weights [13, 14]. In this setting the MCB becomes unique for all graphs. Translated to unweighted graphs, this means that no two minimum cycle bases contain all edges in the same number of cycles. Hence, given a minimum cycle basis $\mathcal{M}$ of Γ, there is a perturbed edge weighting such that $\mathcal{M}$ is the unique MCB of the edge-weighted version. This simple observation allows us to translate some of their results to the unweighted scenario. For instance, theorem 1.2 of [14] characterizes the perturbed graphs for which the MCB consists of entirely of shortest cycles as the planar graphs without a dual containing a double claw[1]. In the unweighted case this implies:

Corollary 9 *If* Γ *planar and none of its duals contains double claw, then* $\mathcal{S} = \mathcal{R}$.

The converse is not true. For instance, all triangles in a complete graph (which for $|V| > 4$ is not planar) are relevant, and of course they are shortest cycles.

A graph is *null-homotopic* [2, 9, 18] if it has a cycle basis consisting only of triangles. This is the case for instance for chordal graphs (in which every cycle C of length $|C| \geq 4$ contains a chord, i.e., an edge connecting two of the vertices that are non-adjacent in C.), and in particular for complete graphs K_m, $m \geq 3$. Trivially, if Γ is null-homotopic, then $\mathcal{R} = \mathcal{S}$.

Since the cycles in $\mathcal{S}$ are not independent in general, it seems natural to consider the set

$$\mathcal{U} = \{C \,|\, \exists e \in E : \, C \,\text{is the}\ \textit{unique}\ \text{shortest cycle containing}\ e\} \tag{6}$$

instead. As each cycle C in $\mathcal{U}$ is the unique shortest cycle for any perturbed edge weighting the discussion in [14] implies that C is contained all minimal cycle bases, i.e., $\mathcal{U} \subseteq \mathcal{J}$. Trivially, if $\mathcal{U}$ is a cycle basis, then the MCB is unique and $\mathcal{U} = \mathcal{J} = \mathcal{S} = \mathcal{R}$. This provides a sometimes convenient way to establish the uniqueness of the MCB, see Figure 4.

However, uniqueness of the minimum cycle basis, i.e., $\mathcal{J} = \mathcal{R}$, in general does not imply that $\mathcal{U} = \mathcal{J}$. The example in Figure 3 is outerplanar and hence has a unique MCB, but $\mathcal{U}$ contains only the four triangles. In a more restricted setting, however, which includes secondary structure graphs, we have

Lemma 10 *Let* Γ *be sub-cubic 2-connected outerplanar graph. Then* $\mathcal{U}$ *is the minimal cycle basis.*

Proof. Γ has a unique Hamiltonian H cycle which forms the boundary of the planar embedding [28]. The minimal cycle basis is given by the cells of planar embedding [22]. For each edge $e \in H$ there is unique shortest cycle, namely the cell in which it is contained. Since the vertex degree is at most 3, each

---

[1] A *double claw* with ends x and y is a subgraph that consisting of three internally node disjoint paths from x to y.

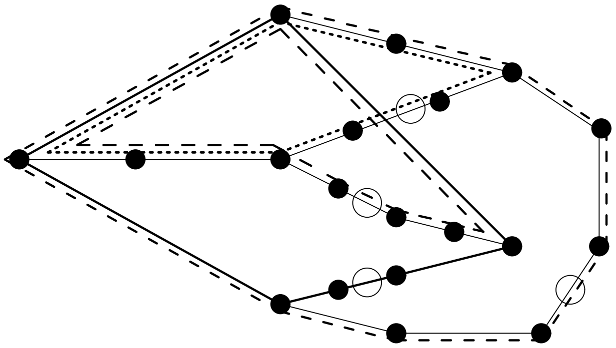

Fig. 4. Γ is a subdivision of $K_{3,3}$, hence $\nu(\Gamma) = 4$. Since $\mathcal{U}$ contains the four marked cycles, Γ has a unique minimal cycle basis. For each of these cycles, an edge for which the cycle is the unique shortest one, is indicated by a circle.

cycle $|C|$ must contain at least one boundary edge e, i.e., $\mathcal{U}$ is the collection of all cells, and hence the MCB. $\square$

Remark. Biopolymer graphs of nucleic acid, be it secondary structures, bisecondary structures, or even more elaborate models, do not contain triangles. Furthermore, the only class of quadrangles is formed by so-called base-pairing stacks, along which edges from T and B alternate. It is easy to verify that each quadrangle is the unique shortest cycle for each of the two backbone edges $e, e' \in T$. Thus $\mathcal{U}$ contains all base-pairing stacks which correspond to the stabilizing structural elements.

A cycle basis $\mathcal{B}$ is called *fundamental* [15, 33] if there exists an ordering of its cycles such that $C_j \setminus (C_1 \cup C_2 \cup \cdots \cup C_{j-1}) \neq \emptyset$ for $2 \leq j \leq \nu(G)$. If Γ contains a spanning tree T such that each C_j is the unique cycle in $T \cup \{e_j\}$ for some edge $e_j \notin T$, $\mathcal{B}$ is called *strictly fundamental* [21]. For a thorough discussion of fundamental cycle bases and related concepts see [15, 12]. The problem of finding a strictly fundamental cycle basis of minimal length is NP complete [6]. An example of graph with triangular (and hence minimal) cycle basis that is not even fundamental can be found in [3].

Lemma 11 *Let Γ be a planar graph with a unique minimal cycle basis. Then $\mathcal{U} \neq \emptyset$.*

Proof. By [22, Cor.13], any MCB of a planar graph is fundamental and hence there is an edge e that is contained in exactly one cycle C. Since all shortest cycles containing e are contained in $\mathcal{R}$ by Prop. 8, we conclude from the uniqueness of the MCB that C is the unique shortest cycle containing e, i.e., $C \in \mathcal{U}$. $\square$

The lemma immediately generalizes to graphs with a unique MCB that is fundamental. We *conjecture* that uniqueness of the MCB implies $\mathcal{U} \neq \emptyset$.

6 Computational Results

Three classes of labelled random graph on n and m vertices are of particular interest in the context of biopolymers: The unconstrained random graphs $\mathfrak{G}_{n,m}$, the random graphs $\mathfrak{H}_{n,m}$ obtained from the cycle $(1,2,\dots,n)$ by inserting additional $m-n$ chords, and a class $\mathfrak{T}_{n,m}$ of connected random graphs obtained from a random spanning tree by inserting additional $m-n+1$ edges.

Let Δ denote the number of triangles. Introducing the random variables X_{ijk}, $i,j,k \in V$, such that $X_{ijk}=1$ if (i,j,k) is a triangle and $X_{ijk}=0$ otherwise, we see that $\mathbb{E}[\Delta]=\sum_{i<j<k}\mathbb{E}[X_{ijk}]$. In an unconstrained random graph with edge drawing probability $p=2m/(n(n-1))$ we have of course $\mathbb{E}[X_{ijk}]=p^3$. In the models $\mathfrak{H}_{n,m}$ and $\mathfrak{T}_{n,m}$ the expected values depend on whether or not two or three of the vertices (i,j,k) are adjacent along the prescribed Hamiltonian cycle or spanning tree, respectively. It is easy to see, however, that this affects only $\mathcal{O}(n^2)$ of the $\binom{n}{3}$ possible triangles.

All triangles being relevant, we expect that a minimum cycle basis will consist almost exclusively of triangles when $\mathbb{E}[\Delta] \gg \nu(\Gamma)$. Making use of the average vertex degree $d=2|E|/|V|=p(n-1)$ this condition becomes

$$\frac{n(n-1)(n-2)}{6}\frac{d^3}{(n-1)^3} \gg \frac{(d-2)n}{2}+1 \tag{7}$$

For large n, equ.(7) simplifies to $d^3/6 \gg d/(2n)$, i.e., $d \gg \sqrt{3n}$ for all three

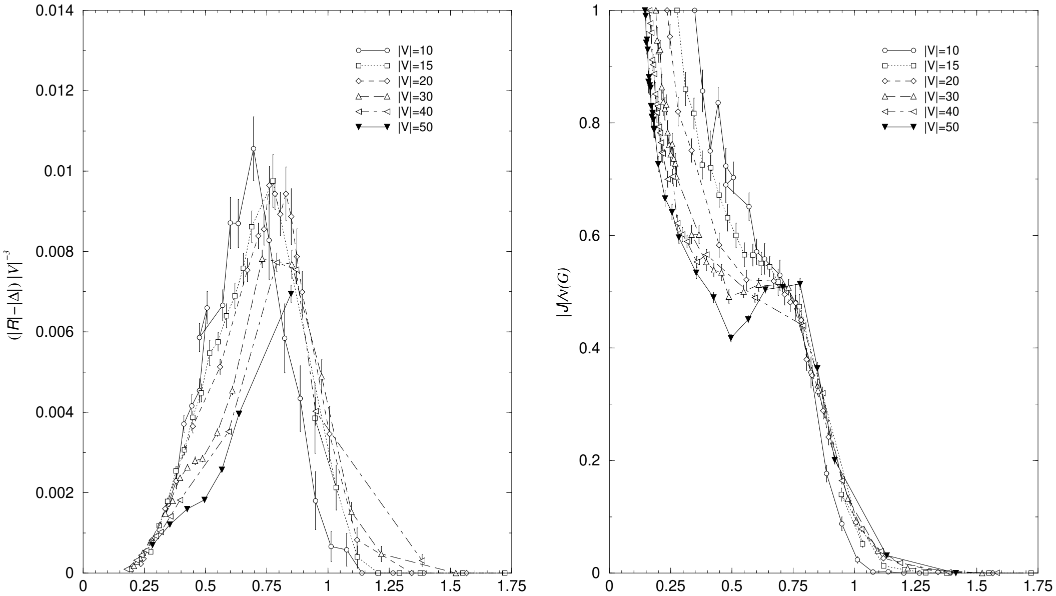

Fig. 5. Relevant and essential cycles in Hamiltonian random graphs $\mathfrak{H}_{n,m}$. L.h.s.: number of relevant *non-triangles*. R.h.s.: fraction of essential cycles in a MCB.

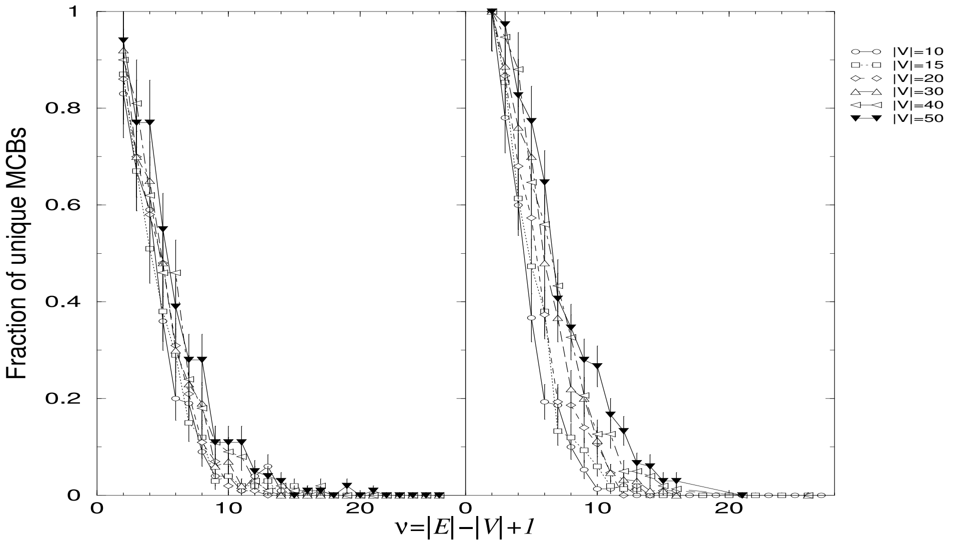

Fig. 6. Fraction of unique minimal cycle bases for the models $\mathfrak{T}_{n,m}$ (l.h.s.) and $\mathfrak{H}_{n,m}$ (r.h.s.), respectively. The transition point depends – if at all – very weakly on $|V|$, certainly not stronger that $\sqrt{|V|}$.

models. In this regime, we expect $|\mathcal{R}| \sim d^3/6$. Computational results indicate that all three random graph models indeed follow this estimate very closely.

Since the distribution of triangles is rather easily understood, our numerical work focuses on the regime $|E| \leq |V|^{3/2}$, where most relevant cycles are large, $|C| \geq 4$. Biopolymer graphs belong to this class since their degree d is bounded by a constant, i.e., $|E| \sim |V|$.

We find that the number $|\mathcal{R}'|$ of relevant non-triangles is particularly large for $|E| \approx |V|^{3/2}$, apparently scaling as $|\mathcal{R}'| \sim |V|^3$, l.h.s. of Figure 5. For $|E| \ll |V|^{3/2}$ we find that a large fraction of the cycles in a MCB are essential. Their fraction sharply drops to virtually 0 close around $|E| \approx |V|^{3/2}$. The behavior of both $|\mathcal{R}|$ and $|\mathcal{J}|$ scales in a simple way with V and E as pointed out by the r.h.s. of Figure 5.

The fraction of graphs that have a unique MCB is remarkably small. Virtually all random graphs with larger cyclomatic number $\nu(\Gamma)$ have redundant MCBs, irrespective of the number $|V|$ of vertices, see Figure 6.

Acknowledgements

Stimulating discussions with Christof Flamm and Josef Leydold as well as computational assistance by Jürgen Gleiss are gratefully acknowledged. This work was supported in part by the Austrian *Fonds zur Förderung der Wissenschaftlichen Forschung* Proj. 12591-INF, the *Jubiläumsfond der Österreich-*

ischen Nationalbank Proj. 6792, and by the *European Commission* within the framework of the Biotechnology Program (BIO-4-98-0189).

References

- [1] F. Bernhart and P. C. Kainen. The book thickeness of a graph. *J. Comb. Theor. B*, 27:320–331, 1979.
- [2] J. A. Bondy. Trigraphs. *Discr. Math.*, 75:69–79, 1989.
- [3] C. Champetier. On the null-homotopy of graphs. *Discr. Math.*, 64:97–98, 1987.
- [4] W.-K. Chen. On vector spaces associated with a graph. *SIAM J. Appl. Math.*, 20:525–529, 1971.
- [5] D. M. Chickering, D. Geiger, and D. Heckerman. On finding a cycle basis with a shortest maximal cycle. *Inform. Processing Let.*, 54:55–58, 1994.
- [6] N. Deo, G. M. Prabhu, and M. S. Krishnamoorty. Algorithms for generating fundamental cycles in a graph. *ACM Trans. Math. Software*, 8:26–42, 1982.
- [7] G. A. Dirac. In abstrakten Graphen vorhandene vollständige 4-Graphen und ihre Unterteilungen. *Math. Nachr.*, 22:61–85, 1960.
- [8] G. M. Downs, V. J. Gillet, J. D. Holliday, and M. F. Lynch. Review of ring perception algorithms for chemical graphs. *J. Chem. Inf. Comput. Sci.*, 29:172–187, 1989.
- [9] P. Duchet, M. Las Vergnas, and H. Meyniel. Connected cutsets of a graph and a triangle basis of the cycle space. *Discr. Math.*, 62:145–154, 1986.
- [10] S. M. Freier, R. Kierzek, J. A. Jaeger, N. Sugimoto, M. H. Caruthers, T. Neilson, and D. H. Turner. Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl. Acad. Sci. (USA)*, 83:9373–9377, 1986.
- [11] A. P. Gultyaev, F. H. van Batenburg, and C. W. Pleij. An approximation of loop free energy values of RNA H-pseudoknots. *RNA*, 5:609–617, 1999.
- [12] D. Hartvigsen and R. Mardon. Cycle bases from orderings and coverings. *Discr. Math.*, 94:81–94, 1991.
- [13] D. Hartvigsen and R. Mardon. The prism-free planar graphs and their cycle bases. *J. Graph Theory*, 15:431–441, 1991.
- [14] D. Hartvigsen and R. Mardon. When do short cycles generate the cycle space. *J. Comb. Theory, Ser. B*, 57:88–99, 1993.
- [15] D. Hartvigsen and E. Zemel. Is every cycle basis fundamental? *J. Graph Theory*, 13:117–137, 1989.
- [16] J. D. Horton. A polynomial-time algorithm to find the shortest cycle basis of a graph. *SIAM J. Comput.*, 16:359–366, 1987.
- [17] E. Hubicka and M. M. Sysło. Minimal bases of cycles of a graph. In M. Fiedler, editor, *Recent Advances in Graph Theory, Proc. 2nd Czechoslovak Symp. on Graph Theory*, pages 283–293. Academia, 1975.

[18] R. E. Jamison. On the null-homotopy of bridged graphs. *Europ. J. Comb.*, 8:421–428, 1987.

[19] S. Kammermeier, H. Neumann, F. Hampel, and R. Herges. Diels-Alder reactions of tetradehydrodianthracene with electron-rich dienes. *Liebigs Ann.*, 1996:1795–1800, 1996.

[20] A. Kaveh. *Structural Mechanics: Graph and Matrix Methods.* Research Studies Press, Exeter, UK, 1992.

[21] G. Kirchhoff. Über die Auflösung der Gleichungen, auf welche man bei der Untersuchungen der linearen Verteilung galvanischer Ströme geführt wird. *Poggendorf Ann. Phys. Chem.*, 72:497–508, 1847.

[22] J. Leydold and P. F. Stadler. Minimal cycle basis of outerplanar graphs. *Elec. J. Comb.*, 5:R16, 1998. See `http://www.combinatorics.org` and Santa Fe Institute Preprint 98-01-011.

[23] G. Liu. On connectivities of tree graphs. *J. Graph Theory*, 12:435–459, 1988.

[24] T. A. McKee. Induced cycle structure and outerplanarity. Preprint, Wright State Univ., Dayton OH, 1998.

[25] M. Plotkin. Mathematical basis of ring-finding algorithms in CIDS. *J. Chem. Doc.*, 11:60–63, 1971.

[26] P. F. Stadler and C. Haslinger. RNA structures with pseudo-knots: Graph-theoretical and combinatorial properties. *Bull. Math. Biol.*, 61:437–467, 1999.

[27] G. F. Stepanec. Basis systems of vector cycles with extremal properties in graphs. *Uspekhi Mat. Nauk. 2*, 19:171–175, 1964. (Russian).

[28] M. M. Sysło. Characterizations of outerplanar graphs. *Discrete Math.*, 26:47–53, 1979.

[29] P. Visamara. *Reconnaissance et représentation d'éléments stucturaux pour la description d'objects complexes. Application à l'élaboration de stratégies de synthèse en chimie organique.* PhD thesis, Université Montpellier II, France, 1995. 95-MON-2-253.

[30] P. Vismara. Union of all the minimum cycle bases of a graph. *Electronic J. Comb.*, 4:#R9 (15 pages), 1997.

[31] H.-J. Voss. *Cycles and Bridges in Graphs.* Kluwer, Dordrecht, 1991.

[32] M. S. Waterman. Secondary structure of single-stranded nucleic acids. *Adv. Math. Suppl. Studies*, 1:167–212, 1978.

[33] H. Whitney. On abstract properties of linear dependence. *Am. J. Math.*, 57:509–533, 1935.

[34] A. A. Zykov. *Theory of Finite Graphs.* Nauka, Novosibirsk, USSR, 1969. (Russian).

[35] A. A. Zykov. *Fundamentals of graph theory.* BCS Associates, Moscow, Idaho, 1990. Edited by L. Boron, C. Christenson, B. Smith.