



Extraction de Motifs Communs dans un Ensemble de Séquences.

**Application à l'identification de sites de liaison
aux protéines dans les séquences primaires d'ADN.**

Alban MANCHERON

alban.mancheron@univ-nantes.fr

Laboratoire d'Informatique de Nantes-Atlantique (LINA)
Centre National de la Recherche Scientifique (CNRS)
Formation de Recherche en Évolution (FRE) 2729.

vendredi 29 septembre 2006

1 De la biologie à l'informatique



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

1 De la biologie à l'informatique

- Introduction à la biologie



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

1 De la biologie à l'informatique

- Introduction à la biologie
- L'algorithmique du texte

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

1 De la biologie à l'informatique

- Introduction à la biologie
- L'algorithme du texte

2 Extraction de motifs

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

1 De la biologie à l'informatique

- Introduction à la biologie
- L'algorithme du texte

2 Extraction de motifs

- STABS

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

1 De la biologie à l'informatique

- Introduction à la biologie
- L'algorithme du texte

2 Extraction de motifs

- STABS
- Fonctions de score *Block-Based*

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

1 De la biologie à l'informatique

- Introduction à la biologie
- L'algorithme du texte

2 Extraction de motifs

- STARS
- Fonctions de score *Block-Based*
- StatiSTARS

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

1 De la biologie à l'informatique

- Introduction à la biologie
- L'algorithme du texte

2 Extraction de motifs

- STARS
- Fonctions de score *Block-Based*
- StatiSTARS

3 Étude des Oracles

Soutenance de thèse

Alban MANCHERON


Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

- 
- 1 De la biologie à l'informatique**
 - Introduction à la biologie
 - L'algorithme du texte
 - 2 Extraction de motifs**
 - STARS
 - Fonctions de score *Block-Based*
 - StatiSTARS
 - 3 Étude des Oracles**
 - Présentation et Construction des Oracles

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

1 De la biologie à l'informatique

- Introduction à la biologie
- L'algorithme du texte

2 Extraction de motifs

- STARS
- Fonctions de score *Block-Based*
- StatiSTARS

3 Étude des Oracles

- Présentation et Construction des Oracles
- Langage reconnu par les Oracles

1 De la biologie à l'informatique

- Introduction à la biologie
- L'algorithme du texte

2 Extraction de motifs

- STARS
- Fonctions de score *Block-Based*
- StatiSTARS

3 Étude des Oracles

- Présentation et Construction des Oracles
- Langage reconnu par les Oracles
- Vers un nouvel oracle

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

1 De la biologie à l'informatique

- Introduction à la biologie
- L'algorithmique du texte

2 Extraction de motifs

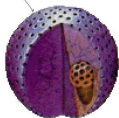
- STARS
- Fonctions de score *Block-Based*
- StatiSTARS

3 Étude des Oracles

- Présentation et Construction des Oracles
- Langage reconnu par les Oracles
- Vers un nouvel oracle

4 Conclusions & perspectives

Noyau



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

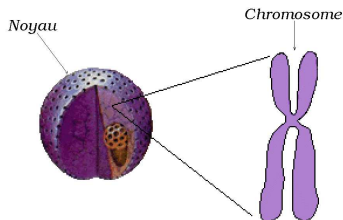
Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

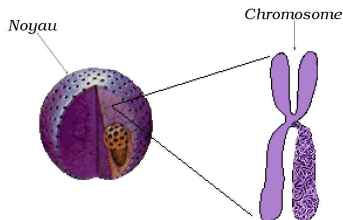
Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

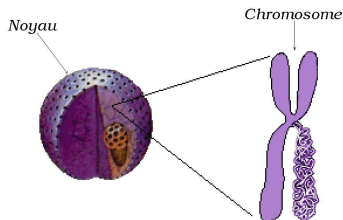
Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

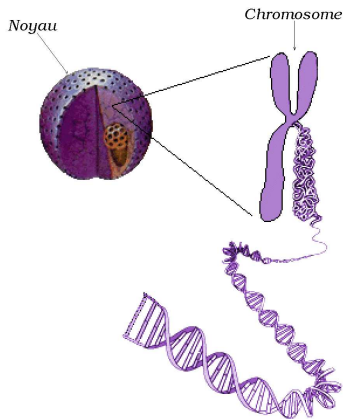
Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

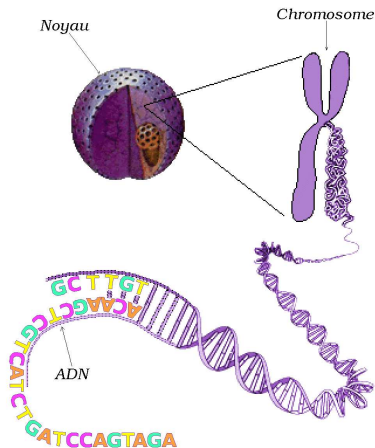
Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

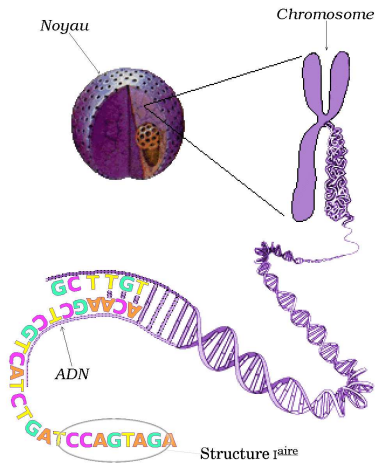
Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

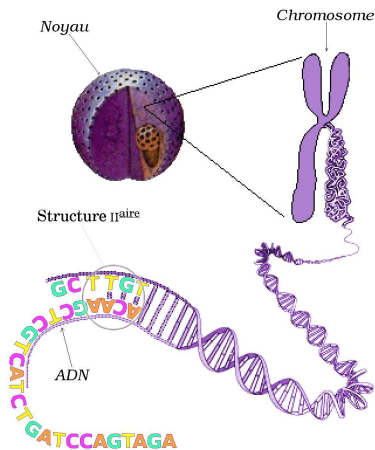
Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

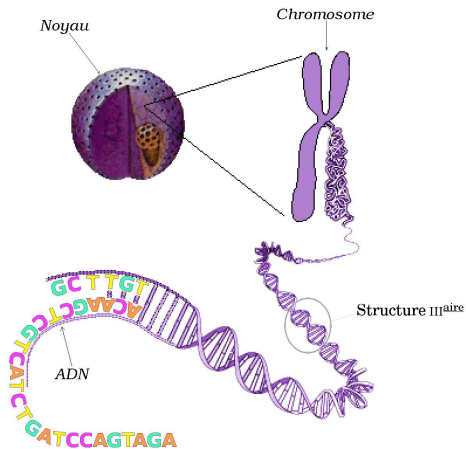
Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

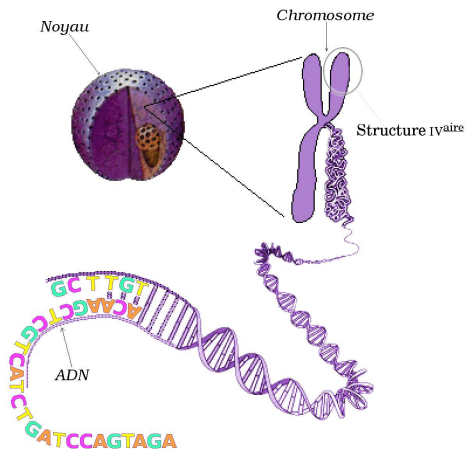
Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

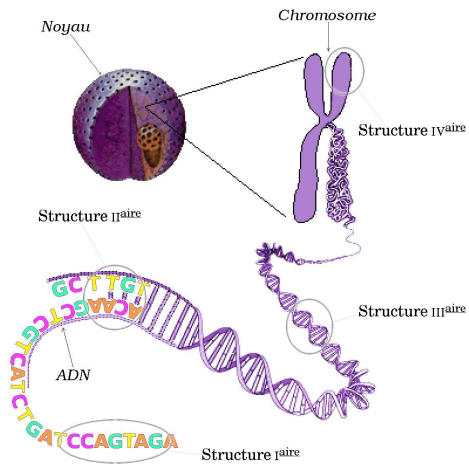
Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Codage des acides nucléiques recommandé par l'IUPAC (1970)

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Codage des acides nucléiques recommandé par l'IUPAC (1970)

Molécule(s)	Système de codage		symbolique
	trois lettres	une lettre	

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Codage des acides nucléiques recommandé par l'IUPAC (1970)

Molécule(s)	Système de codage		symbolique
	trois lettres	une lettre	
Adénine	Ade	A	<u>A</u> dénine
Cytosine	Cyt	C	<u>C</u> ytosine
Guanine	Gua	G	<u>G</u> uanine
Thymine	Thy	T	<u>T</u> hymine

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Codage des acides nucléiques recommandé par l'IUPAC (1970)

Molécule(s)	Système de codage		symbolique
	trois lettres	une lettre	
Adénine	Ade	A	<u>A</u> dénine
Cytosine	Cyt	C	<u>C</u> ytosine
Guanine	Gua	G	<u>G</u> uanine
Thymine	Thy	T	<u>T</u> hymine
Uracile	Ura	U	<u>U</u> racile

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Codage des acides nucléiques recommandé par l'IUPAC (1970)

Molécule(s)	Système de codage		symbolique
	trois lettres	une lettre	
Adénine	Ade	A	<u>A</u> dénine
Cytosine	Cyt	C	<u>C</u> ytosine
Guanine	Gua	G	<u>G</u> uanine
Thymine	Thy	T	<u>T</u> hymine
Uracile	Ura	U	<u>U</u> racile
Purine (Gua ou Ade)	Pur	R	pu <u>R</u> ine
Pyrimidine (Thy ou Cyt)	Pyr	Y	p <u>Y</u> rimidine
Amines (Ade ou Cyt)	/	M	a <u>M</u> ines
Kétines (Gua ou Thy/Ura)	/	K	<u>K</u> étines
Gua ou Cyt	/	S	interactions fortes (Strong)
Ade, Thy ou Ura	/	W	interactions faibles (Weak)
Ade, Cyt ou Thy ou Ura	/	H	tous sauf G, qui est suivi de <u>H</u>
Cyt, Gua, Thy ou Ura	/	B	tous sauf A, qui est suivi de <u>B</u>
Cyt, Gua ou Ade	/	V	tous sauf T/U, qui est suivi de <u>V</u>
Ade, Gua, Thy ou Ura	/	D	tous sauf C, qui est suivi de <u>D</u>
Ade, Cyt, Gua, Thy ou Ura	Bas	N	Toutes les bases (<u>aNy</u>)

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

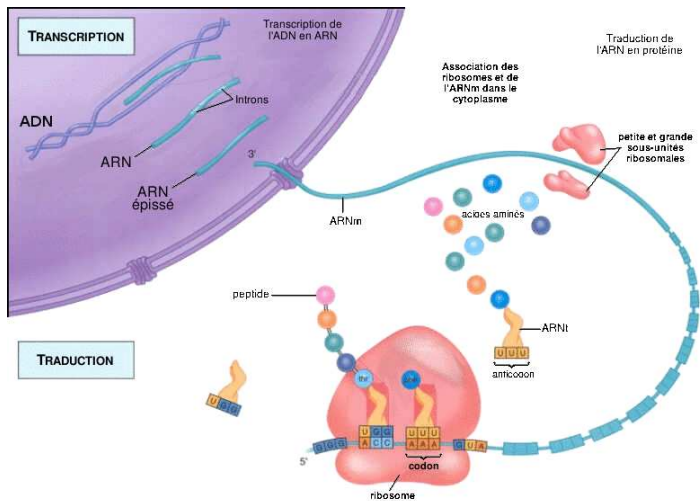
Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

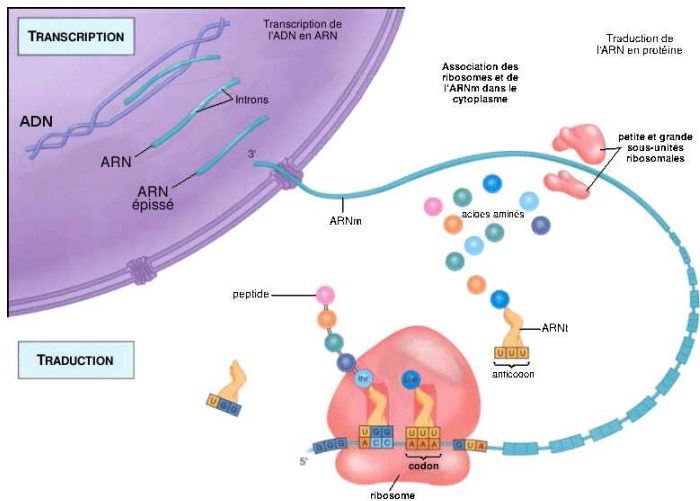
Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

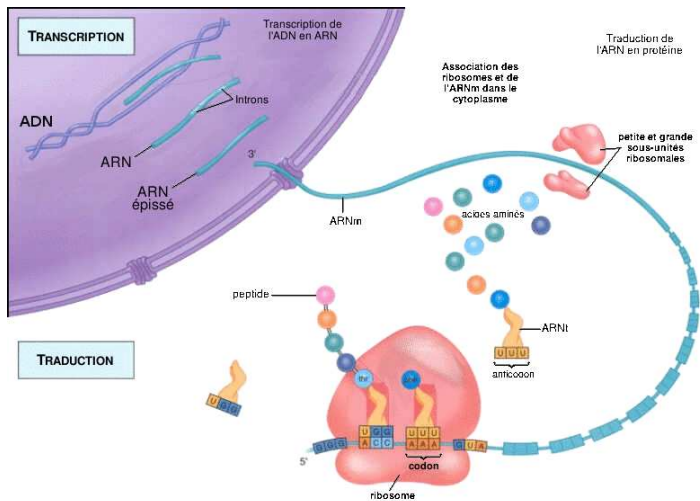
Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

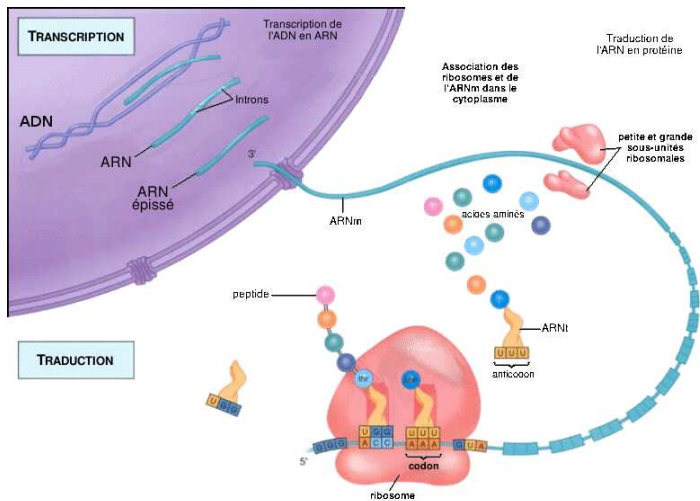
Extraction de motifs

Étude des Oracles

Conclusions & perspectives

transcription.mov

Animation disponible à l'adresse <http://www.wellesley.edu/Chemistry/chem227/>
[contact : preisberg@wellesley.edu]



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

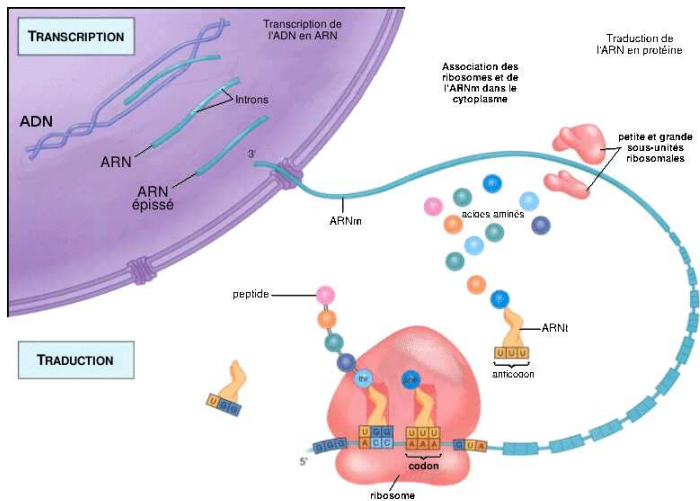
Extraction de motifs

Étude des Oracles

Conclusions & perspectives

prtsynth.mov

Animation disponible à l'adresse <http://www.wellesley.edu/Chemistry/chem227/>
[contact : preisberg@wellesley.edu]



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

- Synthèse protéique
 - sites de régulation ;
 - séquences primaires d'ADN.
- Problèmes de régulation
- Connaissance des sites de régulation
- Caractéristiques des sites de régulation

À la découverte des sites de régulation

Le problème de l'extraction de motifs communs dans les séquences primaires d'ADN ...

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

- Synthèse protéique
 - sites de régulation ;
 - séquences primaires d'ADN.
- Problèmes de régulation
- Connaissance des sites de régulation
- Caractéristiques des sites de régulation

À la découverte des sites de régulation

Le problème de l'extraction de motifs communs dans les séquences primaires d'ADN ...

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

- Synthèse protéique
 - sites de régulation ;
 - séquences primaires d'ADN.
- Problèmes de régulation
 - pathologies...
- Connaissance des sites de régulation
- Caractéristiques des sites de régulation

À la découverte des sites de régulation

Le problème de l'extraction de motifs communs dans les séquences primaires d'ADN ...

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

- Synthèse protéique
 - sites de régulation ;
 - séquences primaires d'ADN.
- Problèmes de régulation
 - pathologies.
- Connaissance des sites de régulation
 - extraction des sites de régulation
- Caractéristiques des sites de régulation

À la découverte des sites de régulation

Le problème de l'extraction de motifs communs dans les séquences primaires d'ADN ...

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

- Synthèse protéique
 - sites de régulation ;
 - séquences primaires d'ADN.
- Problèmes de régulation
 - pathologies.
- Connaissance des sites de régulation
 - action sur la synthèse protéique ;
 - traitement/soins.
- Caractéristiques des sites de régulation

À la découverte des sites de régulation

Le problème de l'extraction de motifs communs dans les séquences primaires d'ADN ...

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

- Synthèse protéique
 - sites de régulation ;
 - séquences primaires d'ADN.
- Problèmes de régulation
 - pathologies.
- Connaissance des sites de régulation
 - action sur la synthèse protéique ;
 - traitement/soins.
- Caractéristiques des sites de régulation

À la découverte des sites de régulation

Le problème de l'extraction de motifs communs dans les séquences primaires d'ADN ...

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

- Synthèse protéique
 - sites de régulation ;
 - séquences primaires d'ADN.
- Problèmes de régulation
 - pathologies.
- Connaissance des sites de régulation
 - action sur la synthèse protéique ;
 - traitement/soins.
- Caractéristiques des sites de régulation

À la découverte des sites de régulation

Le problème de l'extraction de motifs communs dans les séquences primaires d'ADN ...

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

- Synthèse protéique
 - sites de régulation ;
 - séquences primaires d'ADN.
- Problèmes de régulation
 - pathologies.
- Connaissance des sites de régulation
 - action sur la synthèse protéique ;
 - traitement/soins.
- Caractéristiques des sites de régulation
 - fixation de protéines régulatrices sur l'ADN ;
 - motifs à l'échelle de la séquence primaire d'ADN ;
 - motifs à l'échelle de la 3D (secondaire).

À la découverte des sites de régulation

Le problème de l'extraction de motifs communs dans les séquences primaires d'ADN ...

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

- Synthèse protéique
 - sites de régulation ;
 - séquences primaires d'ADN.
- Problèmes de régulation
 - pathologies.
- Connaissance des sites de régulation
 - action sur la synthèse protéique ;
 - traitement/soins.
- Caractéristiques des sites de régulation
 - fixation de protéines régulatrices sur l'ADN ;
 - similarités au niveau de la séquence primaire d'ADN ;
 - petite taille (environ 10-30 nucléotides).

À la découverte des sites de régulation

Le problème de l'extraction de motifs communs dans les séquences primaires d'ADN ...

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

- Synthèse protéique
 - sites de régulation ;
 - séquences primaires d'ADN.
- Problèmes de régulation
 - pathologies.
- Connaissance des sites de régulation
 - action sur la synthèse protéique ;
 - traitement/soins.
- Caractéristiques des sites de régulation
 - fixation de protéines régulatrices sur l'ADN ;
 - similarités au niveau de la séquence primaire d'ADN ;
 - petite taille (environ 10-30 nucléotides).

À la découverte les sites de régulation

Le problème de l'extraction de motifs communs dans les séquences primaires d'ADN ...

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

- Synthèse protéique
 - sites de régulation ;
 - séquences primaires d'ADN.
- Problèmes de régulation
 - pathologies.
- Connaissance des sites de régulation
 - action sur la synthèse protéique ;
 - traitement/soins.
- Caractéristiques des sites de régulation
 - fixation de protéines régulatrices sur l'ADN ;
 - similarités au niveau de la séquence primaire d'ADN ;
 - petite taille (environ 10-30 nucléotides).

À la découverte les sites de régulation

Le problème de l'extraction de motifs communs dans les séquences primaires d'ADN ...

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

- Synthèse protéique
 - sites de régulation ;
 - séquences primaires d'ADN.
- Problèmes de régulation
 - pathologies.
- Connaissance des sites de régulation
 - action sur la synthèse protéique ;
 - traitement/soins.
- Caractéristiques des sites de régulation
 - fixation de protéines régulatrices sur l'ADN ;
 - similarités au niveau de la séquence primaire d'ADN ;
 - petite taille (environ 10-30 nucléotides).

À la découverte des sites de régulation

Le problème de l'extraction de motifs communs dans les séquences primaires d'ADN ...

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

- Synthèse protéique
 - sites de régulation ;
 - séquences primaires d'ADN.
- Problèmes de régulation
 - pathologies.
- Connaissance des sites de régulation
 - action sur la synthèse protéique ;
 - traitement/soins.
- Caractéristiques des sites de régulation
 - fixation de protéines régulatrices sur l'ADN ;
 - similarités au niveau de la séquence primaire d'ADN ;
 - petite taille (environ 10-30 nucléotides).

À la découverte des sites de régulation

Le problème de l'extraction de motifs communs dans les séquences primaires d'ADN ...

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

- Synthèse protéique
 - sites de régulation ;
 - séquences primaires d'ADN.
- Problèmes de régulation
 - pathologies.
- Connaissance des sites de régulation
 - action sur la synthèse protéique ;
 - traitement/soins.
- Caractéristiques des sites de régulation
 - fixation de protéines régulatrices sur l'ADN ;
 - similarités au niveau de la séquence primaire d'ADN ;
 - petite taille (environ 10-30 nucléotides).

À la découverte des sites de régulation

Le problème de l'extraction de motifs communs dans les séquences primaires d'ADN ...

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

- Synthèse protéique
 - sites de régulation ;
 - séquences primaires d'ADN.
- Problèmes de régulation
 - pathologies.
- Connaissance des sites de régulation
 - action sur la synthèse protéique ;
 - traitement/soins.
- Caractéristiques des sites de régulation
 - fixation de protéines régulatrices sur l'ADN ;
 - similarités au niveau de la séquence primaire d'ADN ;
 - petite taille (environ 10-30 nucléotides).

À la découverte des sites de régulation

Le problème de l'extraction de motifs communs dans les séquences primaires d'ADN ...

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

- Synthèse protéique
 - sites de régulation ;
 - séquences primaires d'ADN.
- Problèmes de régulation
 - pathologies.
- Connaissance des sites de régulation
 - action sur la synthèse protéique ;
 - traitement/soins.
- Caractéristiques des sites de régulation
 - fixation de protéines régulatrices sur l'ADN ;
 - similarités au niveau de la séquence primaire d'ADN ;
 - petite taille (environ 10-30 nucléotides).

À la découverte des sites de régulation

Le problème de l'extraction de motifs communs dans les séquences primaires d'ADN ...

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

- Synthèse protéique
 - sites de régulation ;
 - séquences primaires d'ADN.
- Problèmes de régulation
 - pathologies.
- Connaissance des sites de régulation
 - action sur la synthèse protéique ;
 - traitement/soins.
- Caractéristiques des sites de régulation
 - fixation de protéines régulatrices sur l'ADN ;
 - similarités au niveau de la séquence primaire d'ADN ;
 - petite taille (environ 10-30 nucléotides).

À la découverte des sites de régulation

Le problème de l'extraction de motifs communs dans les séquences primaires d'ADN ...

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

- Synthèse protéique
 - sites de régulation ;
 - séquences primaires d'ADN.
- Problèmes de régulation
 - pathologies.
- Connaissance des sites de régulation
 - action sur la synthèse protéique ;
 - traitement/soins.
- Caractéristiques des sites de régulation
 - fixation de protéines régulatrices sur l'ADN ;
 - similarités au niveau de la séquence primaire d'ADN ;
 - petite taille (environ 10-30 nucléotides).

À la découverte les sites de régulation

Le problème de l'extraction de motifs communs dans les séquences primaires d'ADN ...

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

- Synthèse protéique
 - sites de régulation ;
 - séquences primaires d'ADN.
- Problèmes de régulation
 - pathologies.
- Connaissance des sites de régulation
 - action sur la synthèse protéique ;
 - traitement/soins.
- Caractéristiques des sites de régulation
 - fixation de protéines régulatrices sur l'ADN ;
 - similarités au niveau de la séquence primaire d'ADN ;
 - petite taille (environ 10-30 nucléotides).

À la découverte les sites de régulation

Le problème de l'extraction de motifs communs dans les séquences primaires d'ADN ...

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

- Synthèse protéique
 - sites de régulation ;
 - séquences primaires d'ADN.
- Problèmes de régulation
 - pathologies.
- Connaissance des sites de régulation
 - action sur la synthèse protéique ;
 - traitement/soins.
- Caractéristiques des sites de régulation
 - fixation de protéines régulatrices sur l'ADN ;
 - similarités au niveau de la séquence primaire d'ADN ;
 - petite taille (environ 10-30 nucléotides).

À la découverte les sites de régulation

Le problème de l'extraction de motifs communs dans les séquences primaires d'ADN ...

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Un mot s est une séquence finie $s = s_1 s_2 \cdots s_n$ de symboles pris dans un ensemble noté Σ .

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Un mot s est une séquence finie $s = s_1 s_2 \cdots s_n$ de symboles pris dans un ensemble noté Σ .

⇒ Lettres, Alphabet, Longueur.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Un mot s est une séquence finie $s = s_1 s_2 \cdots s_n$ de symboles pris dans un ensemble noté Σ .

⇒ Lettres, Alphabet, Longueur.

Exemple (Facteurs, préfixes et suffixes)

Soit le mot CAGCTGACTAGCTACGACT.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Un mot s est une séquence finie $s = s_1 s_2 \cdots s_n$ de symboles pris dans un ensemble noté Σ .

⇒ Lettres, Alphabet, Longueur.

Exemple (Facteurs, préfixes et suffixes)

Soit le mot **CAGCT**GACTAGCTACGACT.
CAGCT est un préfixe

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Un mot s est une séquence finie $s = s_1 s_2 \cdots s_n$ de symboles pris dans un ensemble noté Σ .

⇒ Lettres, Alphabet, Longueur.

Exemple (Facteurs, préfixes et suffixes)

Soit le mot CAGCTGACTAGCT**ACGACT**.

ACGACT est un suffixe

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Un mot s est une séquence finie $s = s_1 s_2 \cdots s_n$ de symboles pris dans un ensemble noté Σ .

⇒ Lettres, Alphabet, Longueur.

Exemple (Facteurs, préfixes et suffixes)

Soit le mot CAGCTG**ACTAG**CTACGACT.

ACTAG est un facteur

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Deux problématiques :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Deux problématiques :

La recherche de motifs

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Deux problématiques :

La recherche de motifs

L'extraction de motifs

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Deux problématiques :

La recherche de motifs

- Motif(s) connu(s) *a priori* ;

L'extraction de motifs

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Deux problématiques :

La recherche de motifs

- Motif(s) connu(s) *a priori* ;

L'extraction de motifs

- Motif(s) **in**connu(s) *a priori* ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Deux problématiques :

La recherche de motifs

- Motif(s) connu(s) *a priori* ;
- Motif(s) avec ou sans erreurs ;

L'extraction de motifs

- Motif(s) inconnu(s) *a priori* ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Deux problématiques :

La recherche de motifs

- Motif(s) connu(s) *a priori* ;
- Motif(s) avec ou sans erreurs ;

L'extraction de motifs

- Motif(s) inconnu(s) *a priori* ;
- Motif(s) avec ou sans erreurs ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Deux problématiques :

La recherche de motifs

- Motif(s) connu(s) *a priori* ;
- Motif(s) avec ou sans erreurs ;
- Solutions généralement d'ordre polynomial.

L'extraction de motifs

- Motif(s) inconnu(s) *a priori* ;
- Motif(s) avec ou sans erreurs ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Deux problématiques :

La recherche de motifs

- Motif(s) connu(s) *a priori* ;
- Motif(s) avec ou sans erreurs ;
- Solutions généralement d'ordre polynomial.

L'extraction de motifs

- Motif(s) inconnu(s) *a priori* ;
- Motif(s) avec ou sans erreurs ;
- Solutions généralement d'ordre exponentiel.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Deux problématiques :

La recherche de motifs

- Motif(s) connu(s) *a priori* ;
- Motif(s) avec ou sans erreurs ;
- Solutions généralement d'ordre polynomial.

L'extraction de motifs

- Motif(s) inconnu(s) *a priori* ;
- Motif(s) avec ou sans erreurs ;
- Solutions généralement d'ordre exponentiel.

Dans les deux cas :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Deux problématiques :

La recherche de motifs

- Motif(s) connu(s) *a priori* ;
- Motif(s) avec ou sans erreurs ;
- Solutions généralement d'ordre polynomial.

L'extraction de motifs

- Motif(s) inconnu(s) *a priori* ;
- Motif(s) avec ou sans erreurs ;
- Solutions généralement d'ordre exponentiel.

Dans les deux cas :

Besoin de caractériser la notion de ressemblance entre motifs

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Deux problématiques :

La recherche de motifs

- Motif(s) connu(s) *a priori* ;
- Motif(s) avec ou sans erreurs ;
- Solutions généralement d'ordre polynomial.

L'extraction de motifs

- Motif(s) inconnu(s) *a priori* ;
- Motif(s) avec ou sans erreurs ;
- Solutions généralement d'ordre exponentiel.

Dans les deux cas :

Besoin de caractériser la notion de ressemblance entre motifs

⇒ la similarité...

- Critère subjectif;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

- Critère subjectif ;
- Risque de confusion

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

- Critère subjectif ;
- Risque de confusion
- Un exemple :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

- Critère subjectif ;
- Risque de confusion
- Un exemple :
 - *deux mammifères se ressemblent plus qu'un mammifère et un poisson.*

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

- Critère subjectif ;
- Risque de confusion
- Un exemple :
 - *deux mammifères se ressemblent plus qu'un mammifère et un poisson.*
 - *Quid du dauphin ???*

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

- Critère subjectif ;
- Risque de confusion
- Un exemple :
 - *deux mammifères se ressemblent plus qu'un mammifère et un poisson.*
 - *Quid du dauphin ???*
- Problème de formalisation

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

- Critère subjectif ;
- Risque de confusion
- Un exemple :
 - *deux mammifères se ressemblent plus qu'un mammifère et un poisson.*
 - *Quid du dauphin ???*
- Problème de formalisation
- Problème d'interprétation

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

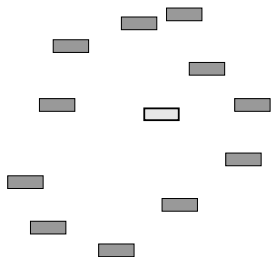
Extraction de motifs

Étude des Oracles

Conclusions & perspectives

L'algorithmique du texte

Similarité par rapport à un modèle externe vs. similarité entre motifs.



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

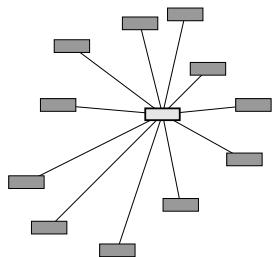
Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

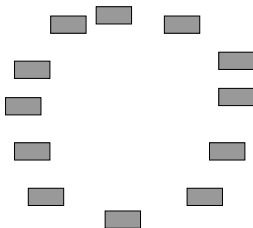
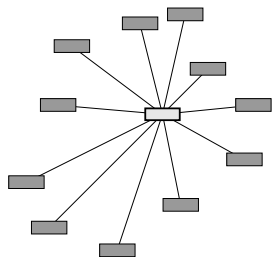
Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

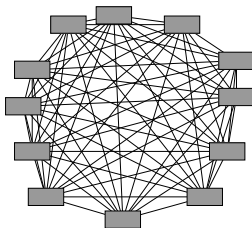
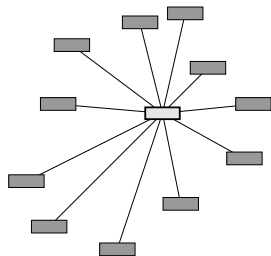
Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Extraction de motifs communs à un ensemble de séquences :EM.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de motifs ($m_1 \in s_1, \dots, m_t \in s_t$) telles que m_1, \dots, m_t soient similaires.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Extraction de motifs communs à un ensemble de séquences :EM.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de motifs ($m_1 \in s_1, \dots, m_t \in s_t$) telles que m_1, \dots, m_t soient similaires.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Extraction de motifs communs à un ensemble de séquences :EM.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de motifs ($m_1 \in s_1, \dots, m_t \in s_t$) telles que m_1, \dots, m_t soient similaires.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Extraction de motifs communs à un ensemble de séquences :EM.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de motifs ($m_1 \in s_1, \dots, m_t \in s_t$) telles que m_1, \dots, m_t soient similaires.

Extraction sous contrainte de quorum de motifs communs dans un ensemble de séquences : EM_q.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), un quorum q ($0 \leq q \leq 100\%$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de $R \geq \lceil qt \rceil$ motifs ($m_1 \in s'_1, \dots, m_R \in s'_R$) telles que $\{s'_1, \dots, s'_R\} \subseteq \mathcal{S}$ ($s'_i \neq s'_j$) et que m_1, \dots, m_R soient similaires.

Extraction de motifs communs à un ensemble de séquences :EM.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de motifs ($m_1 \in s_1, \dots, m_t \in s_t$) telles que m_1, \dots, m_t soient similaires.

Extraction sous contrainte de quorum de motifs communs dans un ensemble de séquences : EM_q.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), un quorum q ($0 \leq q \leq 100\%$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de $R \geq \lceil qt \rceil$ motifs ($m_1 \in s'_1, \dots, m_R \in s'_R$) telles que $\{s'_1, \dots, s'_R\} \subseteq \mathcal{S}$ ($s'_i \neq s'_j$) et que m_1, \dots, m_R soient similaires.

Extraction de motifs communs à un ensemble de séquences :EM.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de motifs ($m_1 \in s_1, \dots, m_t \in s_t$) telles que m_1, \dots, m_t soient similaires.

Extraction sous contrainte de quorum de motifs communs dans un ensemble de séquences : EM_q.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), un quorum q ($0 \leq q \leq 100\%$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de $R \geq \lceil qt \rceil$ motifs ($m_1 \in s'_1, \dots, m_R \in s'_R$) telles que $\{s'_1, \dots, s'_R\} \subseteq \mathcal{S}$ ($s'_i \neq s'_j$) et que m_1, \dots, m_R soient similaires.

		Algorithmes	
À base			

• Ces méthodes sont :

• Les méthodes proposées doivent offrir :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

		Algorithmes	
		Déterministes	
À base			

• Ces méthodes sont :

• Les méthodes proposées doivent offrir :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

		Algorithmes	
		Déterministes	Probabilistes
À base			

• Ces méthodes sont :

• Les méthodes proposées doivent offrir :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

		Algorithmes	
		Déterministes	Probabilistes
À base	de fouille		

• Ces méthodes sont :

• Les méthodes proposées doivent offrir :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

		Algorithmes	
		Déterministes	Probabilistes
À base	d'index		
	de fouille		

• Ces méthodes sont :

• Les méthodes proposées doivent offrir :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

		Algorithmes	
		Déterministes	Probabilistes
À base	d'index	MOTIF [SMITH & al., 1990]	
	de fouille		

• Ces méthodes sont :

• Les méthodes proposées doivent offrir :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

		Algorithmes	
		Déterministes	Probabilistes
À base	d'index	MOTIF [SMITH & al., 1990] PRATT [JONASSEN, 1997]	
	de fouille		

• Ces méthodes sont :

• Les méthodes proposées doivent offrir :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

		Algorithmes	
		Déterministes	Probabilistes
À base	d'index	MOTIF [SMITH & al., 1990] PRATT [JONASSEN, 1997] SMILE [MARSAN & al., 2001]	
	de fouille		

• Ces méthodes sont :

• Les méthodes proposées doivent offrir :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

		Algorithmes	
		Déterministes	Probabilistes
À base	d'index	MOTIF [SMITH & al., 1990] PRATT [JONASSEN, 1997] SMILE [MARSAN & al., 2001]	DISCOVER [WANG & al., 1994]
	de fouille		

• Ces méthodes sont :

• Les méthodes proposées doivent offrir :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

		Algorithmes	
		Déterministes	Probabilistes
À base	d'index	MOTIF [SMITH & al., 1990] PRATT [JONASSEN, 1997] SMILE [MARSAN & al., 2001]	DISCOVER [WANG & al., 1994]
	de fouille	MEME [BAILEY & al., 1994]	

• Ces méthodes sont :

• Les méthodes proposées doivent offrir :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

		Algorithmes	
		Déterministes	Probabilistes
À base	d'index	MOTIF [SMITH & al., 1990] PRATT [JONASSEN, 1997] SMILE [MARSAN & al., 2001]	DISCOVER [WANG & al., 1994]
	de fouille	MEME [BAILEY & al., 1994] SPLASH [CALIFANO, 2000]	

• Ces méthodes sont :

• Les méthodes proposées doivent offrir :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

		Algorithmes	
		Déterministes	Probabilistes
À base	d'index	MOTIF [SMITH & al., 1990] PRATT [JONASSEN, 1997] SMILE [MARSAN & al., 2001]	DISCOVER [WANG & al., 1994]
	de fouille	MEME [BAILEY & al., 1994] SPLASH [CALIFANO, 2000] TEIRESIAS [RIGOUTSOS & al., 1998]	

• Ces méthodes sont :

• Les méthodes proposées doivent offrir :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

		Algorithmes	
		Déterministes	Probabilistes
À base	d'index	MOTIF [SMITH & al., 1990] PRATT [JONASSEN, 1997] SMILE [MARSAN & al., 2001]	DISCOVER [WANG & al., 1994]
	de fouille	MEME [BAILEY & al., 1994] SPLASH [CALIFANO, 2000] TEIRESIAS [RIGOUTSOS & al., 1998] WINNOWER [PEVZNER & al., 2000]	

• Ces méthodes sont :

• Les méthodes proposées doivent offrir :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

		Algorithmes	
		Déterministes	Probabilistes
À base	d'index	MOTIF [SMITH & al., 1990] PRATT [JONASSEN, 1997] SMILE [MARSAN & al., 2001]	DISCOVER [WANG & al., 1994]
	de fouille	MEME [BAILEY & al., 1994] SPLASH [CALIFANO, 2000] TEIRESIAS [RIGOUTSOS & al., 1998] WINNOWER [PEVZNER & al., 2000]	Meta-MEME [GRUNDY & al., 1997]

• Ces méthodes sont :

• Les méthodes proposées doivent offrir :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

		Algorithmes	
		Déterministes	Probabilistes
À base	d'index	MOTIF [SMITH & al., 1990] PRATT [JONASSEN, 1997] SMILE [MARSAN & al., 2001]	DISCOVER [WANG & al., 1994]
	de fouille	MEME [BAILEY & al., 1994] SPLASH [CALIFANO, 2000] TEIRESIAS [RIGOUTSOS & al., 1998] WINNOWER [PEVZNER & al., 2000]	Meta-MEME [GRUNDY & al., 1997] GIBBS [LAWRENCE & al., 1993]

• Ces méthodes sont :

• Les méthodes proposées doivent offrir :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

		Algorithmes	
		Déterministes	Probabilistes
À base	d'index	MOTIF [SMITH & al., 1990] PRATT [JONASSEN, 1997] SMILE [MARSAN & al., 2001]	DISCOVER [WANG & al., 1994]
	de fouille	MEME [BAILEY & al., 1994] SPLASH [CALIFANO, 2000] TEIRESIAS [RIGOUTSOS & al., 1998] WINNOWER [PEVZNER & al., 2000]	Meta-MEME [GRUNDY & al., 1997] GIBBS [LAWRENCE & al., 1993] PROJECTION [BUHLER & al., 2001]

- Ces méthodes sont :

- Les méthodes proposées doivent offrir :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

		Algorithmes	
		Déterministes	Probabilistes
À base	d'index	MOTIF [SMITH & al., 1990] PRATT [JONASSEN, 1997] SMILE [MARSAN & al., 2001]	DISCOVER [WANG & al., 1994]
	de fouille	MEME [BAILEY & al., 1994] SPLASH [CALIFANO, 2000] TEIRESIAS [RIGOUTSOS & al., 1998] WINNOWER [PEVZNER & al., 2000]	Meta-MEME [GRUNDY & al., 1997] GIBBS [LAWRENCE & al., 1993] PROJECTION [BUHLER & al., 2001] CONSENSUS [HERTZ & al., 1995]

- Ces méthodes sont :
 - soit trop spécifiques, soit inefficaces ;
 - et souvent difficiles à paramétrer ;
- Les méthodes proposées doivent offrir :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

		Algorithmes	
		Déterministes	Probabilistes
À base	d'index	MOTIF [SMITH & al., 1990] PRATT [JONASSEN, 1997] SMILE [MARSAN & al., 2001]	DISCOVER [WANG & al., 1994]
	de fouille	MEME [BAILEY & al., 1994] SPLASH [CALIFANO, 2000] TEIRESIAS [RIGOUTSOS & al., 1998] WINNOWER [PEVZNER & al., 2000]	Meta-MEME [GRUNDY & al., 1997] GIBBS [LAWRENCE & al., 1993] PROJECTION [BUHLER & al., 2001] CONSENSUS [HERTZ & al., 1995]

- Ces méthodes sont :
 - soit trop spécifiques, soit inefficaces ;
 - souvent difficiles à paramétrer :

- Les méthodes proposées doivent offrir :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

		Algorithmes	
		Déterministes	Probabilistes
À base	d'index	MOTIF [SMITH & al., 1990] PRATT [JONASSEN, 1997] SMILE [MARSAN & al., 2001]	DISCOVER [WANG & al., 1994]
	de fouille	MEME [BAILEY & al., 1994] SPLASH [CALIFANO, 2000] TEIRESIAS [RIGOUTSOS & al., 1998] WINNOWER [PEVZNER & al., 2000]	Meta-MEME [GRUNDY & al., 1997] GIBBS [LAWRENCE & al., 1993] PROJECTION [BUHLER & al., 2001] CONSENSUS [HERTZ & al., 1995]

- Ces méthodes sont :
 - soit trop spécifiques, soit inefficaces ;
 - souvent difficiles à paramétrer :

- nombre de paramètres trop important ;
 - paramètres difficiles à régler ;

- Les méthodes proposées doivent offrir :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

		Algorithmes	
		Déterministes	Probabilistes
À base	d'index	MOTIF [SMITH & al., 1990] PRATT [JONASSEN, 1997] SMILE [MARSAN & al., 2001]	DISCOVER [WANG & al., 1994]
	de fouille	MEME [BAILEY & al., 1994] SPLASH [CALIFANO, 2000] TEIRESIAS [RIGOUTSOS & al., 1998] WINNOWER [PEVZNER & al., 2000]	Meta-MEME [GRUNDY & al., 1997] GIBBS [LAWRENCE & al., 1993] PROJECTION [BUHLER & al., 2001] CONSENSUS [HERTZ & al., 1995]

- Ces méthodes sont :
 - soit trop spécifiques, soit inefficaces ;
 - souvent difficiles à paramétrer :
 - nombre de paramètres trop important ;
 - manque d'intuition.
- Les méthodes proposées doivent offrir :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

		Algorithmes	
		Déterministes	Probabilistes
À base	d'index	MOTIF [SMITH & al., 1990] PRATT [JONASSEN, 1997] SMILE [MARSAN & al., 2001]	DISCOVER [WANG & al., 1994]
	de fouille	MEME [BAILEY & al., 1994] SPLASH [CALIFANO, 2000] TEIRESIAS [RIGOUTSOS & al., 1998] WINNOWER [PEVZNER & al., 2000]	Meta-MEME [GRUNDY & al., 1997] GIBBS [LAWRENCE & al., 1993] PROJECTION [BUHLER & al., 2001] CONSENSUS [HERTZ & al., 1995]

- Ces méthodes sont :
 - soit trop spécifiques, soit inefficaces ;
 - souvent difficiles à paramétrer :
 - nombre de paramètres trop important ;
 - manque d'intuition.
- Les méthodes proposées doivent offrir :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

		Algorithmes	
		Déterministes	Probabilistes
À base	d'index	MOTIF [SMITH & al., 1990] PRATT [JONASSEN, 1997] SMILE [MARSAN & al., 2001]	DISCOVER [WANG & al., 1994]
	de fouille	MEME [BAILEY & al., 1994] SPLASH [CALIFANO, 2000] TEIRESIAS [RIGOUTSOS & al., 1998] WINNOWER [PEVZNER & al., 2000]	Meta-MEME [GRUNDY & al., 1997] GIBBS [LAWRENCE & al., 1993] PROJECTION [BUHLER & al., 2001] CONSENSUS [HERTZ & al., 1995]

- Ces méthodes sont :
 - soit trop spécifiques, soit inefficaces ;
 - souvent difficiles à paramétrer :
 - nombre de paramètres trop important ;
 - manque d'intuition.
- Les méthodes proposées doivent offrir :
 - un compromis acceptable entre généricité et efficacité ;
 - une paramétrisation simple, rapide et intuitive ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

		Algorithmes	
		Déterministes	Probabilistes
À base	d'index	MOTIF [SMITH & al., 1990] PRATT [JONASSEN, 1997] SMILE [MARSAN & al., 2001]	DISCOVER [WANG & al., 1994]
	de fouille	MEME [BAILEY & al., 1994] SPLASH [CALIFANO, 2000] TEIRESIAS [RIGOUTSOS & al., 1998] WINNOWER [PEVZNER & al., 2000]	Meta-MEME [GRUNDY & al., 1997] GIBBS [LAWRENCE & al., 1993] PROJECTION [BUHLER & al., 2001] CONSENSUS [HERTZ & al., 1995]

- Ces méthodes sont :
 - soit trop spécifiques, soit inefficaces ;
 - souvent difficiles à paramétrer :
 - nombre de paramètres trop important ;
 - manque d'intuition.
- Les méthodes proposées doivent offrir :
 - un compromis acceptable entre généricité et efficacité ;
 - une paramétrisation simple, rapide et intuitive ;
 - une complexité temporelle et spatiale polynomiale.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

		Algorithmes	
		Déterministes	Probabilistes
À base	d'index	MOTIF [SMITH & al., 1990] PRATT [JONASSEN, 1997] SMILE [MARSAN & al., 2001]	DISCOVER [WANG & al., 1994]
	de fouille	MEME [BAILEY & al., 1994] SPLASH [CALIFANO, 2000] TEIRESIAS [RIGOUTSOS & al., 1998] WINNOWER [PEVZNER & al., 2000]	Meta-MEME [GRUNDY & al., 1997] GIBBS [LAWRENCE & al., 1993] PROJECTION [BUHLER & al., 2001] CONSENSUS [HERTZ & al., 1995]

- Ces méthodes sont :
 - soit trop spécifiques, soit inefficaces ;
 - souvent difficiles à paramétrer :
 - nombre de paramètres trop important ;
 - manque d'intuition.
- Les méthodes proposées doivent offrir :
 - un compromis acceptable entre généralité et efficacité ;
 - une paramétrisation simple, rapide et intuitive ;
 - une complexité temporelle et spatiale polynomiale.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

		Algorithmes	
		Déterministes	Probabilistes
À base	d'index	MOTIF [SMITH & al., 1990] PRATT [JONASSEN, 1997] SMILE [MARSAN & al., 2001]	DISCOVER [WANG & al., 1994]
	de fouille	MEME [BAILEY & al., 1994] SPLASH [CALIFANO, 2000] TEIRESIAS [RIGOUTSOS & al., 1998] WINNOWER [PEVZNER & al., 2000]	Meta-MEME [GRUNDY & al., 1997] GIBBS [LAWRENCE & al., 1993] PROJECTION [BUHLER & al., 2001] CONSENSUS [HERTZ & al., 1995]

- Ces méthodes sont :
 - soit trop spécifiques, soit inefficaces ;
 - souvent difficiles à paramétrer :
 - nombre de paramètres trop important ;
 - manque d'intuition.
- Les méthodes proposées doivent offrir :
 - un compromis acceptable entre généricité et efficacité ;
 - une paramétrisation simple, rapide et intuitive ;
 - une complexité temporelle et spatiale polynomiale.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

		Algorithmes	
		Déterministes	Probabilistes
À base	d'index	MOTIF [SMITH & al., 1990] PRATT [JONASSEN, 1997] SMILE [MARSAN & al., 2001]	DISCOVER [WANG & al., 1994]
	de fouille	MEME [BAILEY & al., 1994] SPLASH [CALIFANO, 2000] TEIRESIAS [RIGOUTSOS & al., 1998] WINNOWER [PEVZNER & al., 2000]	Meta-MEME [GRUNDY & al., 1997] GIBBS [LAWRENCE & al., 1993] PROJECTION [BUHLER & al., 2001] CONSENSUS [HERTZ & al., 1995]

- Ces méthodes sont :
 - soit trop spécifiques, soit inefficaces ;
 - souvent difficiles à paramétrer :
 - nombre de paramètres trop important ;
 - manque d'intuition.
- Les méthodes proposées doivent offrir :
 - un compromis acceptable entre généralité et efficacité ;
 - une paramétrisation simple, rapide et intuitive ;
 - une complexité temporelle et spatiale polynomiale.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Introduction à la biologie

L'algorithmique du texte

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Extraction de motifs communs à un ensemble de séquences.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*, 1 \leq i \leq t$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de motifs ($m_1 \in s_1, \dots, m_t \in s_t$) telles que m_1, \dots, m_t soient similaires.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Stars

Fonctions de score Block-Based

StatiStars

Étude des Oracles

Conclusions & perspectives

Extraction de motifs communs à un ensemble de séquences.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de motifs ($m_1 \in s_1, \dots, m_t \in s_t$) telles que m_1, \dots, m_t soient similaires.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Extraction de motifs communs à un ensemble de séquences.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de motifs ($m_1 \in s_1, \dots, m_t \in s_t$) telles que m_1, \dots, m_t soient similaires.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Extraction de motifs communs à un ensemble de séquences.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de motifs ($m_1 \in s_1, \dots, m_t \in s_t$) telles que m_1, \dots, m_t soient similaires.

- **Principe :**

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Extraction de motifs communs à un ensemble de séquences.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de motifs ($m_1 \in s_1, \dots, m_t \in s_t$) telles que m_1, \dots, m_t soient similaires.

- **Principe :**

- choisir une séquence comme référence ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Extraction de motifs communs à un ensemble de séquences.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de motifs ($m_1 \in s_1, \dots, m_t \in s_t$) telles que m_1, \dots, m_t soient similaires.

• Principe :

- choisir une séquence comme référence ;
- pour chaque séquence de l'ensemble, chercher les **parties conservées** par rapport à la séquence de référence ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Extraction de motifs communs à un ensemble de séquences.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de motifs ($m_1 \in s_1, \dots, m_t \in s_t$) telles que m_1, \dots, m_t soient similaires.

• Principe :

- choisir une séquence comme référence ;
- pour chaque séquence de l'ensemble, chercher les **parties conservées** par rapport à la séquence de référence ;
- ne garder que ces **parties conservées** (les candidats) pour la prochaine comparaison.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Extraction de motifs communs à un ensemble de séquences.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de motifs ($m_1 \in s_1, \dots, m_t \in s_t$) telles que m_1, \dots, m_t soient similaires.

- Principe :
- Problèmes posés :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Extraction de motifs communs à un ensemble de séquences.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de motifs ($m_1 \in s_1, \dots, m_t \in s_t$) telles que m_1, \dots, m_t soient similaires.

- **Principe :**
- **Problèmes posés :**
 - choix de la séquence de référence ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Extraction de motifs communs à un ensemble de séquences.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de motifs ($m_1 \in s_1, \dots, m_t \in s_t$) telles que m_1, \dots, m_t soient similaires.

- **Principe :**
- **Problèmes posés :**
 - choix de la séquence de référence ;
 - ordre de traitement des séquences ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Extraction de motifs communs à un ensemble de séquences.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de motifs ($m_1 \in s_1, \dots, m_t \in s_t$) telles que m_1, \dots, m_t soient similaires.

- **Principe :**
- **Problèmes posés :**
 - choix de la séquence de référence ;
 - ordre de traitement des séquences ;
 - définir la notion de similarité.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Extraction de motifs communs à un ensemble de séquences.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de motifs ($m_1 \in s_1, \dots, m_t \in s_t$) telles que m_1, \dots, m_t soient similaires.

- **Principe :**
- **Problèmes posés :**
- **Complexité :**

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Extraction de motifs communs à un ensemble de séquences.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de motifs ($m_1 \in s_1, \dots, m_t \in s_t$) telles que m_1, \dots, m_t soient similaires.

- **Principe :**
- **Problèmes posés :**
- **Complexité :**
 - la construction de l'arbre des solutions s'effectue en $O(p t n^2)$;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Extraction de motifs communs à un ensemble de séquences.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de motifs ($m_1 \in s_1, \dots, m_t \in s_t$) telles que m_1, \dots, m_t soient similaires.

- **Principe :**
- **Problèmes posés :**
- **Complexité :**
 - la construction de l'arbre des solutions s'effectue en $O(p t n^2)$;
 - l'algorithme complet s'exécute en moyenne en $O(p^2 t^3 n^2)$.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Définition (Fonctions *Block-Based*)

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Définition (Fonctions *Block-Based*)

Fonctions composantes de score :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Définition (Fonctions *Block-Based*)

Fonctions composantes de score :

- $f^= (k) \Rightarrow$ bloc consécutif de k « correspondances » ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Définition (Fonctions *Block-Based*)

Fonctions composantes de score :

- $f^= (k) \Rightarrow$ bloc consécutif de k « correspondances » ;
- $f^{\neq} (k) \Rightarrow$ bloc consécutif de k « non correspondances » ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Définition (Fonctions *Block-Based*)

Fonctions composantes de score :

- $f^= (k) \Rightarrow$ bloc consécutif de k « correspondances » ;
- $f^\neq (k) \Rightarrow$ bloc consécutif de k « non correspondances » ;

\Rightarrow Fonction *Block-Based* :

$$S_{\langle f^=, f^\neq \rangle} (1^{k_1} 0^{\ell_1} \dots 1^{k_n} 0^{\ell_n}) := \sum_{i=1}^n f^= (k_i) + f^\neq (\ell_i)$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Définition (Fonctions *Block-Based*)

Fonctions composantes de score :

- $f^= (k) \Rightarrow$ bloc consécutif de k « correspondances » ;
- $f^\neq (k) \Rightarrow$ bloc consécutif de k « non correspondances » ;

\Rightarrow Fonction *Block-Based* :

$$S_{\langle f^=, f^\neq \rangle} (1^{k_1} 0^{\ell_1} \dots 1^{k_n} 0^{\ell_n}) := \sum_{i=1}^n f^= (k_i) + f^\neq (\ell_i)$$

Exemple

$$\begin{array}{rcccccccccccc} w_1 & = & P & C & V & E & P & I & U & Y & G & F \\ w_2 & = & U & C & V & R & T & I & U & P & G & S \\ A(w_1, w_2) & = & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \end{array}$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Définition (Fonctions *Block-Based*)

Fonctions composantes de score :

- $f^= (k) \Rightarrow$ bloc consécutif de k « correspondances » ;
- $f^\neq (k) \Rightarrow$ bloc consécutif de k « non correspondances » ;

\Rightarrow Fonction *Block-Based* :

$$S_{\langle f^=, f^\neq \rangle} (1^{k_1} 0^{\ell_1} \dots 1^{k_n} 0^{\ell_n}) := \sum_{i=1}^n f^= (k_i) + f^\neq (\ell_i)$$

Exemple

$$\begin{array}{rcccccccccc} w_1 & = & P & C & V & E & P & I & U & Y & G & F \\ w_2 & = & U & C & V & R & T & I & U & P & G & S \\ A(w_1, w_2) & = & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \end{array}$$

$$f^= (k) = 0, f^\neq (k) = k$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Définition (Fonctions *Block-Based*)

Fonctions composantes de score :

- $f^= (k) \Rightarrow$ bloc consécutif de k « correspondances » ;
- $f^\neq (k) \Rightarrow$ bloc consécutif de k « non correspondances » ;

\Rightarrow Fonction *Block-Based* :

$$S_{\langle f^=, f^\neq \rangle} (1^{k_1} 0^{\ell_1} \dots 1^{k_n} 0^{\ell_n}) := \sum_{i=1}^n f^= (k_i) + f^\neq (\ell_i)$$

Exemple

$$\begin{array}{rcccccccccc} w_1 & = & P & C & V & E & P & I & U & Y & G & F \\ w_2 & = & U & C & V & R & T & I & U & P & G & S \\ A(w_1, w_2) & = & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \end{array}$$

$$f^= (k) = 0, f^\neq (k) = k$$

$$\begin{aligned} S_{\langle f^=, f^\neq \rangle} (0110011010) &= f^= (0) + f^\neq (1) + f^= (2) + f^\neq (2) \\ &\quad + f^= (2) + f^\neq (1) + f^= (1) + f^\neq (1) \\ &= 0 + 1 + 0 + 2 + 0 + 1 + 0 + 1 = 5. \end{aligned}$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

$$\Sigma = \{a, b, c\}$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

$$\Sigma = \{a, b, c\} \quad \left[\begin{array}{l} s_1 = ababc \\ s_2 = babac \\ s_3 = ccbab \end{array} \right.$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

$$\Sigma = \{a, b, c\} \quad \left[\begin{array}{l} s_1 = ababc \\ s_2 = babac \\ s_3 = ccbab \end{array} \right. \quad \left[\begin{array}{l} e = 1 \\ p = 3 \\ f^=(k) = k \\ f^\neq(k) = -k \end{array} \right.$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

$$\Sigma = \{a, b, c\} \quad \begin{cases} s_1 = ababc \\ s_2 = babac \\ s_3 = ccbab \end{cases} \quad \begin{cases} e = 1 \\ p = 3 \\ f^=(k) = k \\ f^\neq(k) = -k \end{cases}$$

[ababc]
●

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

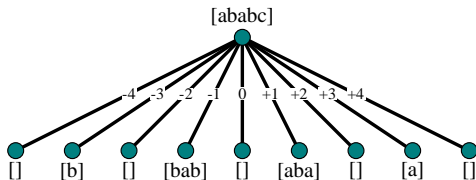
Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

$$\Sigma = \{a, b, c\} \quad \begin{cases} s_1 = ababc \\ s_2 = babac \\ s_3 = ccbab \end{cases} \quad \begin{cases} e = 1 \\ p = 3 \\ f^=(k) = k \\ f^{\neq}(k) = -k \end{cases}$$



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

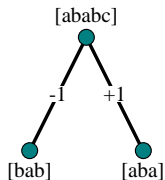
Étude des Oracles

Conclusions & perspectives

$$\Sigma = \{a, b, c\}$$

$$\begin{cases} s_1 = ababc \\ s_2 = babac \\ s_3 = ccbab \end{cases}$$

$$\begin{cases} e = 1 \\ p = 3 \\ f^=(k) = k \\ f^{\neq}(k) = -k \end{cases}$$



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

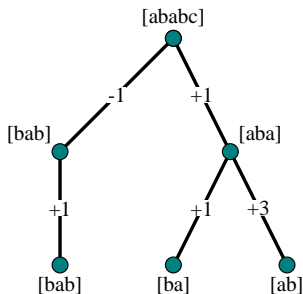
StatiSTARS

Étude des Oracles

Conclusions & perspectives

$$\Sigma = \{a, b, c\}$$

$$\begin{cases} s_1 = ababc \\ s_2 = babac \\ s_3 = ccbab \end{cases} \quad \begin{cases} e = 1 \\ p = 3 \\ f^=(k) = k \\ f^{\neq}(k) = -k \end{cases}$$



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- Jeux d'essais :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- Jeux d'essais :
 - Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- Jeux d'essais :
 - Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
 - Caractéristiques :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- Jeux d'essais :

- Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
- Caractéristiques :
 - 1 000 ensembles de séquences ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- Jeux d'essais :

- Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
- Caractéristiques :
 - 1 000 ensembles de séquences ;
 - chaque ensemble contient au moins 10 séquences ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- Jeux d'essais :

- Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
- Caractéristiques :
 - 1 000 ensembles de séquences ;
 - chaque ensemble contient au moins 10 séquences ;
 - chaque séquence est de longueur > 100 ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- Jeux d'essais :

- Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
- Caractéristiques :
 - 1 000 ensembles de séquences ;
 - chaque ensemble contient au moins 10 séquences ;
 - chaque séquence est de longueur > 100 ;
 - chaque séquence a 3 motifs aléatoirement altérés, insérés à des positions aléatoires.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- Jeux d'essais :

- Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
- Caractéristiques :
 - 1 000 ensembles de séquences ;
 - chaque ensemble contient au moins 10 séquences ;
 - chaque séquence est de longueur > 100 ;
 - chaque séquence a 3 motifs aléatoirement altérés, insérés à des positions aléatoires.
- Critères d'évaluation : indicateurs sur la fiabilité des tests statistiques.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

• Jeux d'essais :

- Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
- Caractéristiques :
 - 1 000 ensembles de séquences ;
 - chaque ensemble contient au moins 10 séquences ;
 - chaque séquence est de longueur > 100 ;
 - chaque séquence a 3 motifs aléatoirement altérés, insérés à des positions aléatoires.
- Critères d'évaluation : indicateurs sur la fiabilité des tests statistiques.

Principe des tests statistiques :

	Motif présent	Motif absent
Motif détecté		
Motif non détecté		

• Jeux d'essais :

- Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
- Caractéristiques :
 - 1 000 ensembles de séquences ;
 - chaque ensemble contient au moins 10 séquences ;
 - chaque séquence est de longueur > 100 ;
 - chaque séquence a 3 motifs aléatoirement altérés, insérés à des positions aléatoires.
- Critères d'évaluation : indicateurs sur la fiabilité des tests statistiques.

Principe des tests statistiques :

	Motif présent	Motif absent
Motif détecté	Vrai Positifs (VP)	
Motif non détecté		

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

• Jeux d'essais :

- Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
- Caractéristiques :
 - 1 000 ensembles de séquences ;
 - chaque ensemble contient au moins 10 séquences ;
 - chaque séquence est de longueur > 100 ;
 - chaque séquence a 3 motifs aléatoirement altérés, insérés à des positions aléatoires.
- Critères d'évaluation : indicateurs sur la fiabilité des tests statistiques.

Principe des tests statistiques :

	Motif présent	Motif absent
Motif détecté	Vrai Positifs (VP)	Faux Positifs (FP)
Motif non détecté		

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

• Jeux d'essais :

- Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
- Caractéristiques :
 - 1 000 ensembles de séquences ;
 - chaque ensemble contient au moins 10 séquences ;
 - chaque séquence est de longueur > 100 ;
 - chaque séquence a 3 motifs aléatoirement altérés, insérés à des positions aléatoires.
- Critères d'évaluation : indicateurs sur la fiabilité des tests statistiques.

Principe des tests statistiques :

	Motif présent	Motif absent
Motif détecté	Vrai Positifs (VP)	Faux Positifs (FP)
Motif non détecté	Faux Négatifs (FN)	

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

• Jeux d'essais :

- Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
- Caractéristiques :
 - 1 000 ensembles de séquences ;
 - chaque ensemble contient au moins 10 séquences ;
 - chaque séquence est de longueur > 100 ;
 - chaque séquence a 3 motifs aléatoirement altérés, insérés à des positions aléatoires.
- Critères d'évaluation : indicateurs sur la fiabilité des tests statistiques.

Principe des tests statistiques :

	Motif présent	Motif absent
Motif détecté	Vrai Positifs (VP)	Faux Positifs (FP)
Motif non détecté	Faux Négatifs (FN)	Vrai Négatifs (VN)

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

• Jeux d'essais :

- Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
- Caractéristiques :
 - 1 000 ensembles de séquences ;
 - chaque ensemble contient au moins 10 séquences ;
 - chaque séquence est de longueur > 100 ;
 - chaque séquence a 3 motifs aléatoirement altérés, insérés à des positions aléatoires.
- Critères d'évaluation : indicateurs sur la fiabilité des tests statistiques.

Principe des tests statistiques :

	Motif présent	Motif absent
Motif détecté	Vrai Positifs (VP)	Faux Positifs (FP)
Motif non détecté	Faux Négatifs (FN)	Vrai Négatifs (VN)

$$\text{Sensibilité} = \frac{VP}{VP + FN}$$

$$\text{Spécificité} = \frac{VN}{VN + FP}$$

$$VPP = \frac{VP}{VP + FP}$$

$$VPN = \frac{VN}{VN + FN}$$

• Jeux d'essais :

- Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
- Caractéristiques :
 - 1 000 ensembles de séquences ;
 - chaque ensemble contient au moins 10 séquences ;
 - chaque séquence est de longueur > 100 ;
 - chaque séquence a 3 motifs aléatoirement altérés, insérés à des positions aléatoires.
- Critères d'évaluation : indicateurs sur la fiabilité des tests statistiques.

Principe des tests statistiques :

	Motif présent	Motif absent
Motif détecté	Vrai Positifs (VP)	Faux Positifs (FP)
Motif non détecté	Faux Négatifs (FN)	Vrai Négatifs (VN)

$$\text{Sensibilité} = \frac{VP}{VP + FN}$$

$$\text{Spécificité} = \frac{VN}{VN + FP}$$

$$VPP = \frac{VP}{VP + FP}$$

$$VPN = \frac{VN}{VN + FN}$$

• Jeux d'essais :

- Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
- Caractéristiques :
 - 1 000 ensembles de séquences ;
 - chaque ensemble contient au moins 10 séquences ;
 - chaque séquence est de longueur > 100 ;
 - chaque séquence a 3 motifs aléatoirement altérés, insérés à des positions aléatoires.
- Critères d'évaluation : indicateurs sur la fiabilité des tests statistiques.

Principe des tests statistiques :

	Motif présent	Motif absent
Motif détecté	Vrai Positifs (VP)	Faux Positifs (FP)
Motif non détecté	Faux Négatifs (FN)	Vrai Négatifs (VN)

$$\text{Sensibilité} = \frac{VP}{VP + FN}$$

$$\text{Spécificité} = \frac{VN}{VN + FP}$$

$$VPP = \frac{VP}{VP + FP}$$

$$VPN = \frac{VN}{VN + FN}$$

• Jeux d'essais :

- Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
- Caractéristiques :
 - 1 000 ensembles de séquences ;
 - chaque ensemble contient au moins 10 séquences ;
 - chaque séquence est de longueur > 100 ;
 - chaque séquence a 3 motifs aléatoirement altérés, insérés à des positions aléatoires.
- Critères d'évaluation : indicateurs sur la fiabilité des tests statistiques.

Principe des tests statistiques :

	Motif présent	Motif absent
Motif détecté	Vrai Positifs (VP)	Faux Positifs (FP)
Motif non détecté	Faux Négatifs (FN)	Vrai Négatifs (VN)

$$\text{Sensibilité} = \frac{VP}{VP + FN}$$

$$\text{Spécificité} = \frac{VN}{VN + FP}$$

$$VPP = \frac{VP}{VP + FP} = 0,95$$

$$VPN = \frac{VN}{VN + FN}$$

• Jeux d'essais :

- Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
- Caractéristiques :
 - 1 000 ensembles de séquences ;
 - chaque ensemble contient au moins 10 séquences ;
 - chaque séquence est de longueur > 100 ;
 - chaque séquence a 3 motifs aléatoirement altérés, insérés à des positions aléatoires.
- Critères d'évaluation : indicateurs sur la fiabilité des tests statistiques.

Principe des tests statistiques :

	Motif présent	Motif absent
Motif détecté	Vrai Positifs (VP)	Faux Positifs (FP)
Motif non détecté	Faux Négatifs (FN)	Vrai Négatifs (VN)

$$\text{Sensibilité} = \frac{VP}{VP + FN} = 0,91$$

$$VPP = \frac{VP}{VP + FP} = 0,95$$

$$\text{Spécificité} = \frac{VN}{VN + FP}$$

$$VPN = \frac{VN}{VN + FN}$$

- Jeux d'essais :
 - Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- Jeux d'essais :
 - Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :
 - Sélection de 5 ensembles parmi les 55 disponibles selon :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- Jeux d'essais :
 - Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :
 - Sélection de 5 ensembles parmi les 55 disponibles selon :
 - le nombre de séquences ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- Jeux d'essais :
 - Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :
 - Sélection de 5 ensembles parmi les 55 disponibles selon :
 - le nombre de séquences ;
 - la taille des séquences.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

• Jeux d'essais :

- Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :

n°	Nb. de séq.	Longueur [min, max]	Protéine
1	9	[53, 806]	ArgR
2	8	[53, 806]	ArgR (site en tandem)
3	16	[100, 384]	LexA
4	18	[299, 5795]	PurR
5	8	[251, 2594]	TyrR

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

• Jeux d'essais :

- Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :

n°	Nb. de séq.	Longueur [min, max]	Protéine
1	9	[53, 806]	ArgR
2	8	[53, 806]	ArgR (site en tandem)
3	16	[100, 384]	LexA
4	18	[299, 5795]	PurR
5	8	[251, 2594]	TyrR

- Analyse des résultats pour le jeu 1 :

Résultat de STARS vs. « Sequence Logo » proposé :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

• Jeux d'essais :

- Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :

n°	Nb. de séq.	Longueur [min, max]	Protéine
1	9	[53, 806]	ArgR
2	8	[53, 806]	ArgR (site en tandem)
3	16	[100, 384]	LexA
4	18	[299, 5795]	PurR
5	8	[251, 2594]	TyrR

- Analyse des résultats pour le jeu 1 :

Résultat de STARS vs. « Sequence Logo » proposé :

TGAATxAxxATxCAxxTxxxxTGxxT

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

• Jeux d'essais :

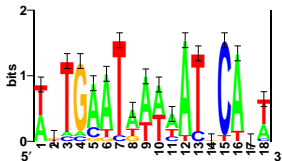
- Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :

n°	Nb. de séq.	Longueur [min, max]	Protéine
1	9	[53, 806]	ArgR
2	8	[53, 806]	ArgR (site en tandem)
3	16	[100, 384]	LexA
4	18	[299, 5795]	PurR
5	8	[251, 2594]	TyrR

- Analyse des résultats pour le jeu 1 :

Résultat de STABS vs. « Sequence Logo » proposé :

TGAATxAxxATxCAxxTxxxxxTGxxT



• Jeux d'essais :

- Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :

n°	Nb. de séq.	Longueur [min, max]	Protéine
1	9	[53, 806]	ArgR
2	8	[53, 806]	ArgR (site en tandem)
3	16	[100, 384]	LexA
4	18	[299, 5795]	PurR
5	8	[251, 2594]	TyrR

- Analyse des résultats pour le jeu 2 :

Résultat de STARS vs. « Sequence Logo » proposé :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

• Jeux d'essais :

- Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :

n°	Nb. de séq.	Longueur [min, max]	Protéine
1	9	[53, 806]	ArgR
2	8	[53, 806]	ArgR (site en tandem)
3	16	[100, 384]	LexA
4	18	[299, 5795]	PurR
5	8	[251, 2594]	TyrR

- Analyse des résultats pour le jeu 2 :

Résultat de STARS vs. « Sequence Logo » proposé :

TGAATxAxxATxCAxxTAxxxTGxxTxxxAATxCA

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

• Jeux d'essais :

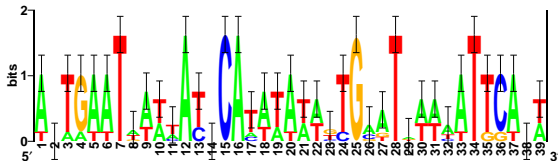
- Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :

n°	Nb. de séq.	Longueur [min, max]	Protéine
1	9	[53, 806]	ArgR
2	8	[53, 806]	ArgR (site en tandem)
3	16	[100, 384]	LexA
4	18	[299, 5795]	PurR
5	8	[251, 2594]	TyrR

- Analyse des résultats pour le jeu 2 :

Résultat de STABS vs. « Sequence Logo » proposé :

TGAATxAxxATxCAxxxTAxxxxTGxxTxxxxAATxCA



• Jeux d'essais :

- Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :

n°	Nb. de séq.	Longueur [min, max]	Protéine
1	9	[53, 806]	ArgR
2	8	[53, 806]	ArgR (site en tandem)
3	16	[100, 384]	LexA
4	18	[299, 5795]	PurR
5	8	[251, 2594]	TyrR

- Analyse des résultats pour le jeu 3 :

Résultat de STARS vs. « Sequence Logo » proposé :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

• Jeux d'essais :

- Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :

n°	Nb. de séq.	Longueur [min, max]	Protéine
1	9	[53, 806]	ArgR
2	8	[53, 806]	ArgR (site en tandem)
3	16	[100, 384]	LexA
4	18	[299, 5795]	PurR
5	8	[251, 2594]	TyrR

- Analyse des résultats pour le jeu 3 :

Résultat de STABS vs. « Sequence Logo » proposé :

ACTGTATATxxAxxCAG

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STABS

Fonctions de score Block-Based

StatiSTABS

Étude des Oracles

Conclusions & perspectives

• Jeux d'essais :

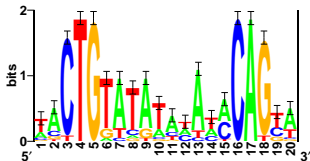
- Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :

n°	Nb. de seq.	Longueur [min, max]	Protéine
1	9	[53, 806]	ArgR
2	8	[53, 806]	ArgR (site en tandem)
3	16	[100, 384]	LexA
4	18	[299, 5795]	PurR
5	8	[251, 2594]	TyrR

- Analyse des résultats pour le jeu 3 :

Résultat de STABS vs. « Sequence Logo » proposé :

ACTGTATATxxAxxCAG



• Jeux d'essais :

- Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :

n°	Nb. de séq.	Longueur [min, max]	Protéine
1	9	[53, 806]	ArgR
2	8	[53, 806]	ArgR (site en tandem)
3	16	[100, 384]	LexA
4	18	[299, 5795]	PurR
5	8	[251, 2594]	TyrR

- Analyse des résultats pour le jeu 4 :

Résultat de STARS vs. « Sequence Logo » proposé :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

• Jeux d'essais :

- Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :

n°	Nb. de séq.	Longueur [min, max]	Protéine
1	9	[53, 806]	ArgR
2	8	[53, 806]	ArgR (site en tandem)
3	16	[100, 384]	LexA
4	18	[299, 5795]	PurR
5	8	[251, 2594]	TyrR

- Analyse des résultats pour le jeu 4 :

Résultat de STARS vs. « Sequence Logo » proposé :

GCxAxCGTTTTTC

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

• Jeux d'essais :

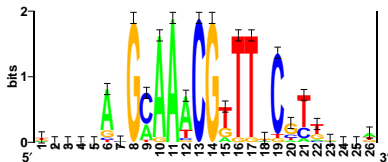
- Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :

n°	Nb. de seq.	Longueur [min, max]	Protéine
1	9	[53, 806]	ArgR
2	8	[53, 806]	ArgR (site en tandem)
3	16	[100, 384]	LexA
4	18	[299, 5795]	PurR
5	8	[251, 2594]	TyrR

- Analyse des résultats pour le jeu 4 :

Résultat de STABS vs. « Sequence Logo » proposé :

GCxAxCGTTTTTC



• Jeux d'essais :

- Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :

n°	Nb. de séq.	Longueur [min, max]	Protéine
1	9	[53, 806]	ArgR
2	8	[53, 806]	ArgR (site en tandem)
3	16	[100, 384]	LexA
4	18	[299, 5795]	PurR
5	8	[251, 2594]	TyrR

- Analyse des résultats pour le jeu 5 :

Résultat de STARS vs. « Sequence Logo » proposé :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

• Jeux d'essais :

- Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :

n°	Nb. de séq.	Longueur [min, max]	Protéine
1	9	[53, 806]	ArgR
2	8	[53, 806]	ArgR (site en tandem)
3	16	[100, 384]	LexA
4	18	[299, 5795]	PurR
5	8	[251, 2594]	TyrR

- Analyse des résultats pour le jeu 5 :

Résultat de STARS vs. « Sequence Logo » proposé :

TGTAAAxxAAxxTxTAC

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

• Jeux d'essais :

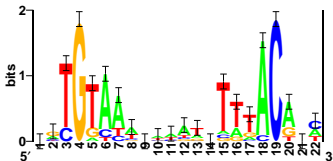
- Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :

n°	Nb. de séq.	Longueur [min, max]	Protéine
1	9	[53, 806]	ArgR
2	8	[53, 806]	ArgR (site en tandem)
3	16	[100, 384]	LexA
4	18	[299, 5795]	PurR
5	8	[251, 2594]	TyrR

- Analyse des résultats pour le jeu 5 :

Résultat de STABS vs. « Sequence Logo » proposé :

TGTA^AA^xx^AA^xx^Tx^TA^C



• Jeux d'essais :

- Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :

n°	Nb. de séq.	Longueur [min, max]	Protéine
1	9	[53, 806]	ArgR
2	8	[53, 806]	ArgR (site en tandem)
3	16	[100, 384]	LexA
4	18	[299, 5795]	PurR
5	8	[251, 2594]	TyrR

- Analyse des résultats de STARS satisfaisante.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

• Jeux d'essais :

- Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :

n°	Nb. de séq.	Longueur [min, max]	Protéine
1	9	[53, 806]	ArgR
2	8	[53, 806]	ArgR (site en tandem)
3	16	[100, 384]	LexA
4	18	[299, 5795]	PurR
5	8	[251, 2594]	TyrR

- Analyse des résultats de STARS satisfaisante.
- Analyse des résultats des autres méthodes (PRATT, TEIRESIAS, MEME, GIBBS, WINNOWER, SMILE, CONSENSUS)

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

• Jeux d'essais :

- Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :

n°	Nb. de séq.	Longueur [min, max]	Protéine
1	9	[53, 806]	ArgR
2	8	[53, 806]	ArgR (site en tandem)
3	16	[100, 384]	LexA
4	18	[299, 5795]	PurR
5	8	[251, 2594]	TyrR

- Analyse des résultats de STARS satisfaisante.
- Analyse des résultats des autres méthodes (PRATT, TEIRESIAS, MEME, GIBBS, WINNOWER, SMILE, CONSENSUS)

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Théorème (Moyenne des fonctions *Block-Based*)

*La moyenne de la variable aléatoire S_n associée aux scores obtenus par les fonctions *Block-Based* est de la forme $\alpha n + \beta + o(1)$.*

Théorème (Variance des fonctions *Block-Based*)

*La variance de la variable aléatoire S_n associée aux scores obtenus par les fonctions *Block-Based* est de la forme $\gamma n + o(n)$.*

Théorème (Moyenne des fonctions *Block-Based*)

*La moyenne de la variable aléatoire S_n associée aux scores obtenus par les fonctions *Block-Based* est de la forme $\alpha n + \beta + o(1)$.*

Théorème (Variance des fonctions *Block-Based*)

*La variance de la variable aléatoire S_n associée aux scores obtenus par les fonctions *Block-Based* est de la forme $\gamma n + o(n)$.*

Théorème (Moyenne des fonctions *Block-Based*)

*La moyenne de la variable aléatoire S_n associée aux scores obtenus par les fonctions *Block-Based* est de la forme $\alpha n + \beta + o(1)$.*

Théorème (Variance des fonctions *Block-Based*)

*La variance de la variable aléatoire S_n associée aux scores obtenus par les fonctions *Block-Based* est de la forme $\gamma n + o(n)$.*

Théorème (Moyenne des fonctions *Block-Based*)

*La moyenne de la variable aléatoire S_n associée aux scores obtenus par les fonctions *Block-Based* est de la forme $\alpha n + \beta + o(1)$.*

Théorème (Variance des fonctions *Block-Based*)

*La variance de la variable aléatoire S_n associée aux scores obtenus par les fonctions *Block-Based* est de la forme $\gamma n + o(n)$.*

	Moyenne	Loi
Source sans mémoire		
<i>Bacillus Subtilis</i>		
Différents Génomes		

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

StARS

Fonctions de score Block-Based

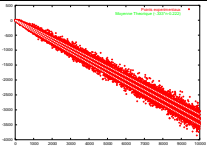
StatiStARS

Étude des Oracles

Conclusions & perspectives

Fonctions de score *Block-Based*

Distribution Gaussienne des fonctions

	Moyenne	Loi
Source sans mémoire		
<i>Bacillus Subtilis</i>		
Différents Génomes		

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

StARS

Fonctions de score Block-Based

StatiStARS

Étude des Oracles

Conclusions & perspectives

Fonctions de score *Block-Based*

Distribution Gaussienne des fonctions

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

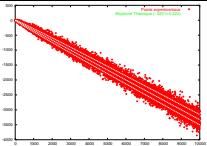
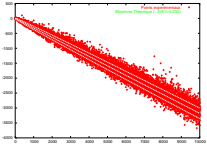
StARS

Fonctions de score *Block-Based*

StatiStARS

Étude des Oracles

Conclusions & perspectives

	Moyenne	Loi
Source sans mémoire		
<i>Bacillus Subtilis</i>		
Différents Génomes		

Fonctions de score *Block-Based*

Distribution Gaussienne des fonctions

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

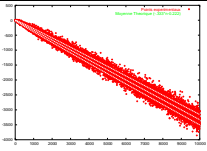
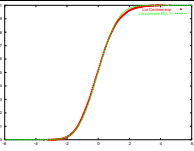
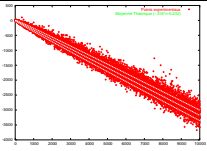
StARS

Fonctions de score Block-Based

StatiStARS

Étude des Oracles

Conclusions & perspectives

	Moyenne	Loi
Source sans mémoire		
<i>Bacillus Subtilis</i>		
Différents Génomes		

Fonctions de score *Block-Based*

Distribution Gaussienne des fonctions

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

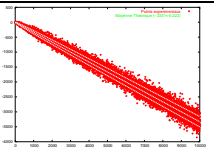
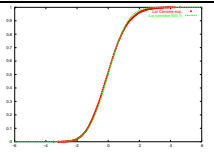
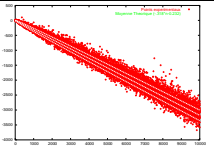
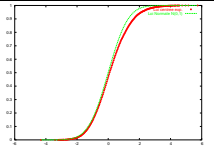
StARS

Fonctions de score *Block-Based*

StatiStARS

Étude des Oracles

Conclusions & perspectives

	Moyenne	Loi
Source sans mémoire		
<i>Bacillus Subtilis</i>		
Différents Génomes		

Fonctions de score *Block-Based*

Distribution Gaussienne des fonctions

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

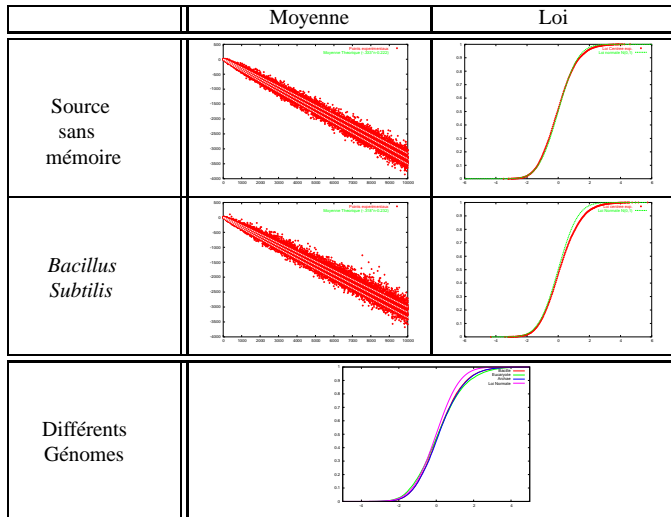
StARS

Fonctions de score *Block-Based*

StatiStARS

Étude des Oracles

Conclusions & perspectives



Extraction sous contrainte de quorum de motifs communs dans un ensemble de séquences.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), un quorum q ($0 \leq q \leq 100\%$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de $R \geq \lceil qt \rceil$ motifs ($m_1 \in s'_1, \dots, m_R \in s'_R$) telles que $\{s'_1, \dots, s'_R\} \subseteq \mathcal{S}$ ($s'_i \neq s'_j$) et que m_1, \dots, m_R soient similaires.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

StARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Extraction sous contrainte de quorum de motifs communs dans un ensemble de séquences.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), un quorum q ($0 \leq q \leq 100\%$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de $R \geq \lceil qt \rceil$ motifs ($m_1 \in s'_1, \dots, m_R \in s'_R$) telles que $\{s'_1, \dots, s'_R\} \subseteq \mathcal{S}$ ($s'_i \neq s'_j$) et que m_1, \dots, m_R soient similaires.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

StARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Extraction sous contrainte de quorum de motifs communs dans un ensemble de séquences.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), un quorum q ($0 \leq q \leq 100\%$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de $R \geq \lceil qt \rceil$ motifs $(m_1 \in s'_1, \dots, m_R \in s'_R)$ telles que $\{s'_1, \dots, s'_R\} \subseteq \mathcal{S}$ ($s'_i \neq s'_j$) et que m_1, \dots, m_R soient similaires.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Extraction sous contrainte de quorum de motifs communs dans un ensemble de séquences.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), un quorum q ($0 \leq q \leq 100\%$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de $R \geq \lceil qt \rceil$ motifs ($m_1 \in s'_1, \dots, m_R \in s'_R$) telles que $\{s'_1, \dots, s'_R\} \subseteq \mathcal{S}$ ($s'_i \neq s'_j$) et que m_1, \dots, m_R soient similaires.

• Principe :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Extraction sous contrainte de quorum de motifs communs dans un ensemble de séquences.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), un quorum q ($0 \leq q \leq 100\%$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de $R \geq \lceil qt \rceil$ motifs ($m_1 \in s'_1, \dots, m_R \in s'_R$) telles que $\{s'_1, \dots, s'_R\} \subseteq \mathcal{S}$ ($s'_i \neq s'_j$) et que m_1, \dots, m_R soient similaires.

• Principe :

- choisir une séquence comme référence ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Extraction sous contrainte de quorum de motifs communs dans un ensemble de séquences.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), un quorum q ($0 \leq q \leq 100\%$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de $R \geq \lceil qt \rceil$ motifs $(m_1 \in s'_1, \dots, m_R \in s'_R)$ telles que $\{s'_1, \dots, s'_R\} \subseteq \mathcal{S}$ ($s'_i \neq s'_j$) et que m_1, \dots, m_R soient similaires.

• Principe :

- choisir une séquence comme référence ;
- pour chaque séquence de l'ensemble, chercher les **parties conservées** par rapport à la séquence de référence ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Extraction sous contrainte de quorum de motifs communs dans un ensemble de séquences.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), un quorum q ($0 \leq q \leq 100\%$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de $R \geq \lceil qt \rceil$ motifs $(m_1 \in s'_1, \dots, m_R \in s'_R)$ telles que $\{s'_1, \dots, s'_R\} \subseteq \mathcal{S}$ ($s'_i \neq s'_j$) et que m_1, \dots, m_R soient similaires.

• Principe :

- choisir une séquence comme référence ;
- pour chaque séquence de l'ensemble, chercher les **parties conservées** par rapport à la séquence de référence ;
- évaluer ces parties conservées d'après les **propriétés statistiques** des fonctions *Block-Based*.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Extraction sous contrainte de quorum de motifs communs dans un ensemble de séquences.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), un quorum q ($0 \leq q \leq 100\%$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de $R \geq \lceil qt \rceil$ motifs $(m_1 \in s'_1, \dots, m_R \in s'_R)$ telles que $\{s'_1, \dots, s'_R\} \subseteq \mathcal{S}$ ($s'_i \neq s'_j$) et que m_1, \dots, m_R soient similaires.

- Principe :
- Problèmes posés :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Extraction sous contrainte de quorum de motifs communs dans un ensemble de séquences.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), un quorum q ($0 \leq q \leq 100\%$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de $R \geq \lceil qt \rceil$ motifs $(m_1 \in s'_1, \dots, m_R \in s'_R)$ telles que $\{s'_1, \dots, s'_R\} \subseteq \mathcal{S}$ ($s'_i \neq s'_j$) et que m_1, \dots, m_R soient similaires.

- **Principe :**
- **Problèmes posés :**
 - choix de la séquence de référence ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Extraction sous contrainte de quorum de motifs communs dans un ensemble de séquences.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), un quorum q ($0 \leq q \leq 100\%$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de $R \geq \lceil qt \rceil$ motifs $(m_1 \in s'_1, \dots, m_R \in s'_R)$ telles que $\{s'_1, \dots, s'_R\} \subseteq \mathcal{S}$ ($s'_i \neq s'_j$) et que m_1, \dots, m_R soient similaires.

- **Principe :**
- **Problèmes posés :**
 - choix de la séquence de référence ;
 - ordre de traitement des séquences ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

StARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Extraction sous contrainte de quorum de motifs communs dans un ensemble de séquences.

Données : un ensemble de séquences $\mathcal{S} = \{s_1, \dots, s_t\}$ ($s_i \in \Sigma^*$, $1 \leq i \leq t$), un quorum q ($0 \leq q \leq 100\%$), ainsi qu'une notion de similarité globale.

Problème : trouver toutes les collections de $R \geq \lceil qt \rceil$ motifs $(m_1 \in s'_1, \dots, m_R \in s'_R)$ telles que $\{s'_1, \dots, s'_R\} \subseteq \mathcal{S}$ ($s'_i \neq s'_j$) et que m_1, \dots, m_R soient similaires.

- **Principe :**
- **Problèmes posés :**
 - choix de la séquence de référence ;
 - ordre de traitement des séquences ;
 - intégration de la contrainte de quorum.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

StARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

⇒ Gestion du quorum au niveau des positions de la séquence de référence.

s_ℓ 

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

⇒ Gestion du quorum au niveau des positions de la séquence de référence.



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

⇒ Gestion du quorum au niveau des positions de la séquence de référence.



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

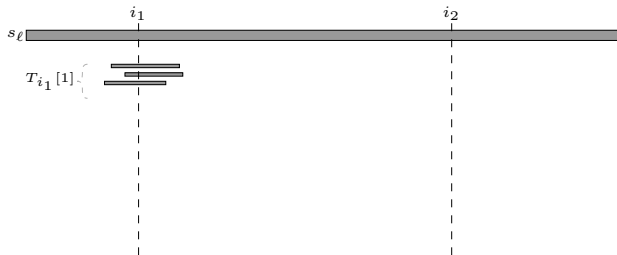
Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

⇒ Gestion du quorum au niveau des positions de la séquence de référence.



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

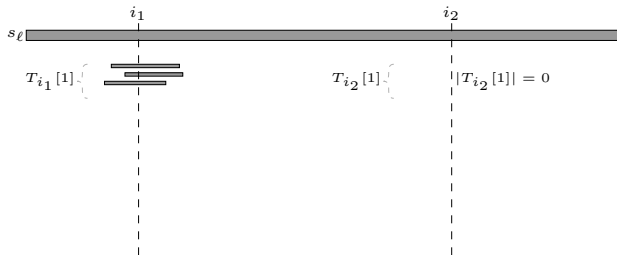
Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

⇒ Gestion du quorum au niveau des positions de la séquence de référence.



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

StARS

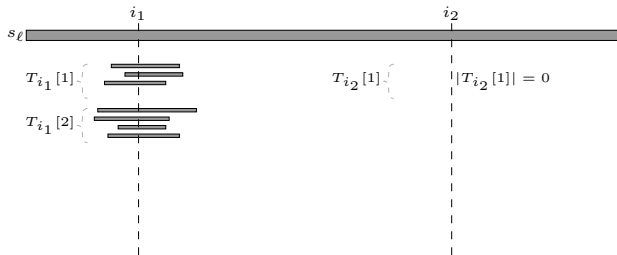
Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

⇒ Gestion du quorum au niveau des positions de la séquence de référence.



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

StARS

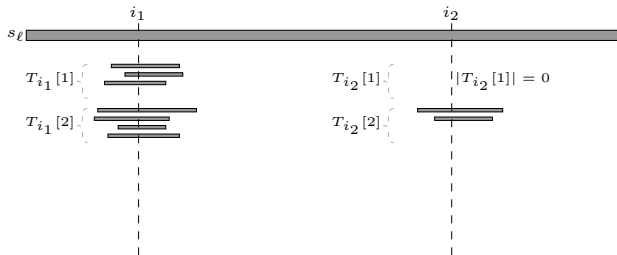
Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

⇒ Gestion du quorum au niveau des positions de la séquence de référence.



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

StARS

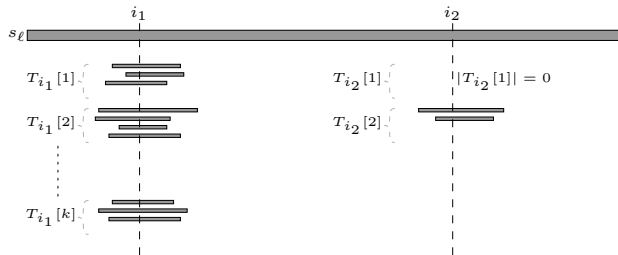
Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

⇒ Gestion du quorum au niveau des positions de la séquence de référence.



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

StARS

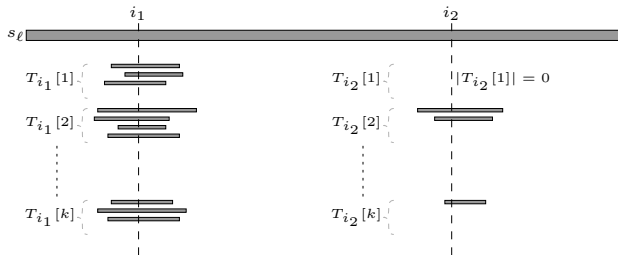
Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

⇒ Gestion du quorum au niveau des positions de la séquence de référence.



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

StARS

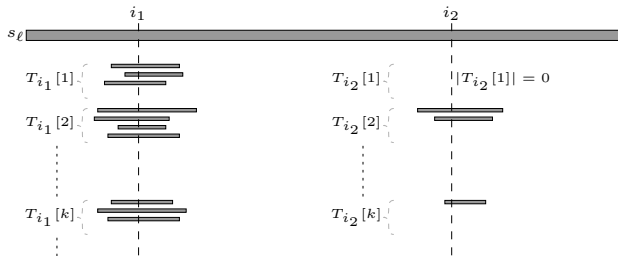
Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

⇒ Gestion du quorum au niveau des positions de la séquence de référence.



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

StARS

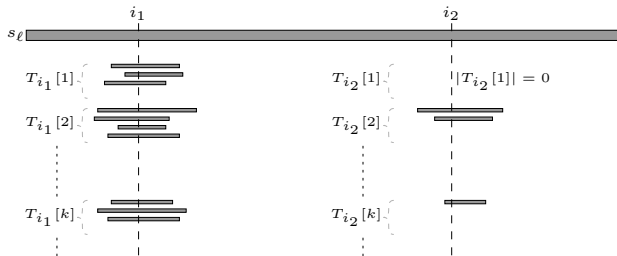
Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

⇒ Gestion du quorum au niveau des positions de la séquence de référence.



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

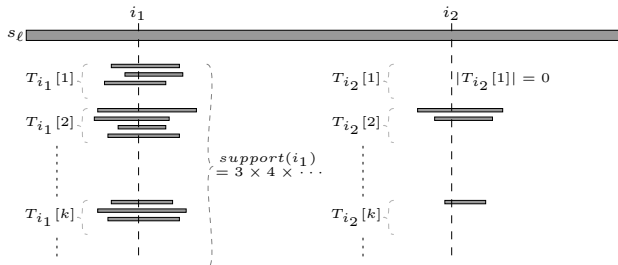
Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

⇒ Gestion du quorum au niveau des positions de la séquence de référence.



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

StARS

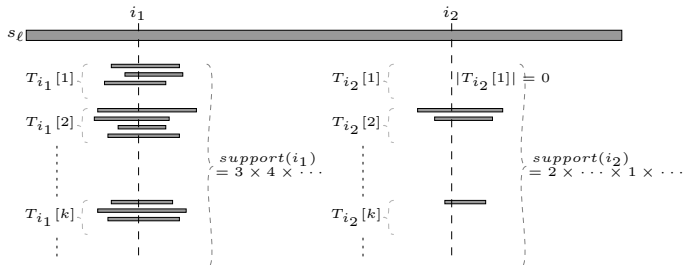
Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

⇒ Gestion du quorum au niveau des positions de la séquence de référence.



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

StARS

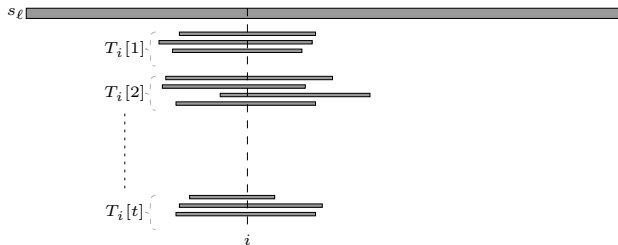
Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

⇒ Gestion du quorum au niveau des positions de la séquence de référence.



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

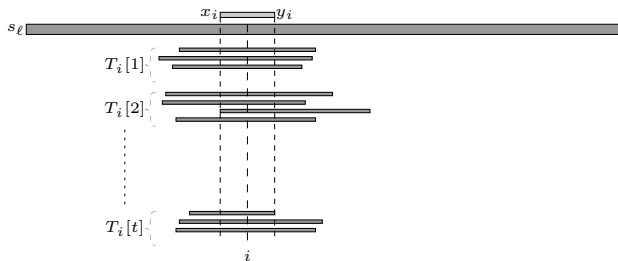
Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

⇒ Gestion du quorum au niveau des positions de la séquence de référence.



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

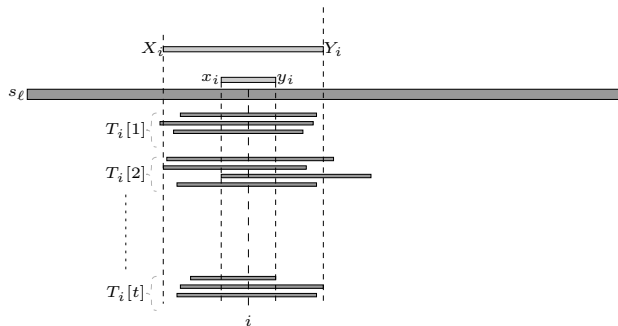
Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

⇒ Gestion du quorum au niveau des positions de la séquence de référence.



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

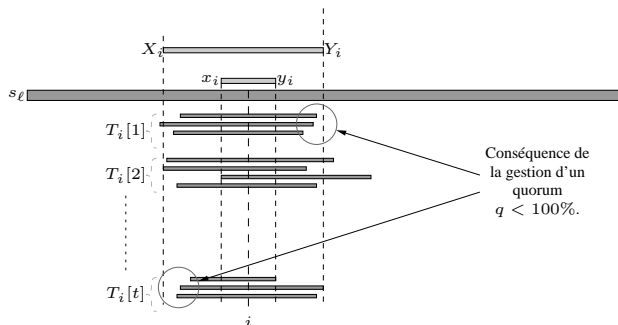
Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

⇒ Gestion du quorum au niveau des positions de la séquence de référence.



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

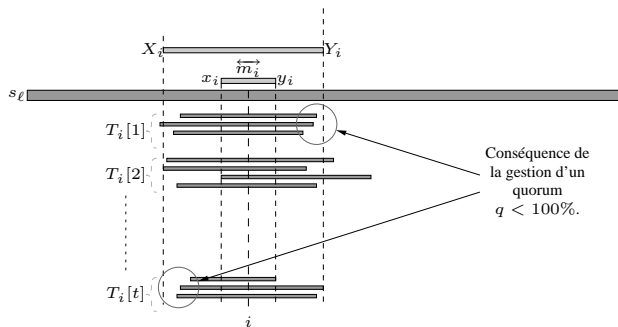
Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

⇒ Gestion du quorum au niveau des positions de la séquence de référence.



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

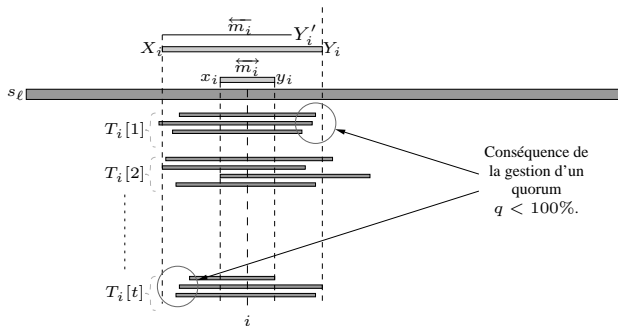
Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

⇒ Gestion du quorum au niveau des positions de la séquence de référence.



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

StARS

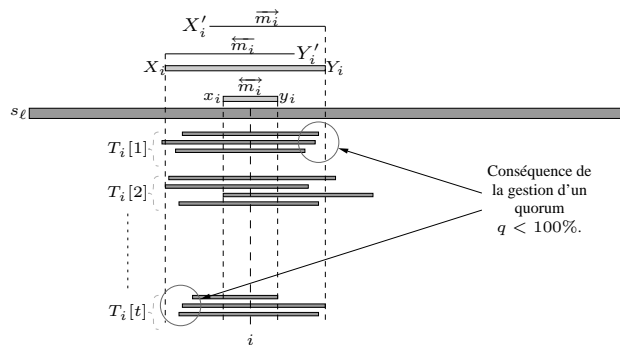
Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

⇒ Gestion du quorum au niveau des positions de la séquence de référence.



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- **Algorithme complet :**

- **Complexité :**

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- **Algorithme complet :**
 - Pour une position i , seuls les motifs maximaux au moins aussi grands que les motifs caractéristiques gauches et droits respectant la contrainte de quorum sont étudiés :

- **Complexité :**

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- **Algorithme complet :**

- Pour une position i , seuls les motifs maximaux au moins aussi grands que les motifs caractéristiques gauches et droits respectant la contrainte de quorum sont étudiés :
- L'algorithme s'arrête lorsque p solutions ont été calculées.

- **Complexité :**

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- **Algorithme complet :**

- Pour une position i , seuls les motifs maximaux au moins aussi grands que les motifs caractéristiques gauches et droits respectant la contrainte de quorum sont étudiés :
- L'algorithme s'arrête lorsque p solutions ont été calculées.

- **Complexité :**

- L'algorithme s'exécute en $O(p t^2 (\log q t + n^2 \log n))$.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- Jeux d'essais :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- Jeux d'essais :
 - Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- Jeux d'essais :

- Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
- Caractéristiques :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- Jeux d'essais :

- Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
- Caractéristiques :
 - 100 ensembles de séquences ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- Jeux d'essais :

- Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
- Caractéristiques :
 - 100 ensembles de séquences ;
 - chaque ensemble contient au moins 10 séquences ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- Jeux d'essais :

- Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
- Caractéristiques :
 - 100 ensembles de séquences ;
 - chaque ensemble contient au moins 10 séquences ;
 - chaque séquence est de longueur > 100 ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- Jeux d'essais :

- Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
- Caractéristiques :
 - 100 ensembles de séquences ;
 - chaque ensemble contient au moins 10 séquences ;
 - chaque séquence est de longueur > 100 ;
 - au moins $\lceil q t \rceil$ séquences contiennent chacun des 3 motifs aléatoirement altérés, insérés à des positions aléatoires.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

StARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- Jeux d'essais :

- Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
- Caractéristiques :
 - 100 ensembles de séquences ;
 - chaque ensemble contient au moins 10 séquences ;
 - chaque séquence est de longueur > 100 ;
 - au moins $\lceil q t \rceil$ séquences contiennent chacun des 3 motifs aléatoirement altérés, insérés à des positions aléatoires.

- Résultats en moyenne :

Contrainte	VPP	Sensibilité
70%	0,83	
80%		
90%		
100%		

- Jeux d'essais :

- Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
- Caractéristiques :
 - 100 ensembles de séquences ;
 - chaque ensemble contient au moins 10 séquences ;
 - chaque séquence est de longueur > 100 ;
 - au moins $\lceil q t \rceil$ séquences contiennent chacun des 3 motifs aléatoirement altérés, insérés à des positions aléatoires.

- Résultats en moyenne :

Contrainte	VPP	Sensibilité
70%	0,83	
80%	0,92	
90%		
100%		

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

StARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- Jeux d'essais :

- Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
- Caractéristiques :
 - 100 ensembles de séquences ;
 - chaque ensemble contient au moins 10 séquences ;
 - chaque séquence est de longueur > 100 ;
 - au moins $\lceil q t \rceil$ séquences contiennent chacun des 3 motifs aléatoirement altérés, insérés à des positions aléatoires.

- Résultats en moyenne :

Contrainte	VPP	Sensibilité
70%	0,83	
80%	0,92	
90%	0,96	
100%		

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- Jeux d'essais :

- Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
- Caractéristiques :
 - 100 ensembles de séquences ;
 - chaque ensemble contient au moins 10 séquences ;
 - chaque séquence est de longueur > 100 ;
 - au moins $\lceil q t \rceil$ séquences contiennent chacun des 3 motifs aléatoirement altérés, insérés à des positions aléatoires.

- Résultats en moyenne :

Contrainte	VPP	Sensibilité
70%	0,83	
80%	0,92	
90%	0,96	
100%	0,99	

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

StARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- Jeux d'essais :

- Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
- Caractéristiques :
 - 100 ensembles de séquences ;
 - chaque ensemble contient au moins 10 séquences ;
 - chaque séquence est de longueur > 100 ;
 - au moins $\lceil q t \rceil$ séquences contiennent chacun des 3 motifs aléatoirement altérés, insérés à des positions aléatoires.

- Résultats en moyenne :

Contrainte	VPP	Sensibilité
70%	0,83	0,87
80%	0,92	
90%	0,96	
100%	0,99	

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

StARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- Jeux d'essais :

- Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
- Caractéristiques :
 - 100 ensembles de séquences ;
 - chaque ensemble contient au moins 10 séquences ;
 - chaque séquence est de longueur > 100 ;
 - au moins $\lceil q t \rceil$ séquences contiennent chacun des 3 motifs aléatoirement altérés, insérés à des positions aléatoires.

- Résultats en moyenne :

Contrainte	VPP	Sensibilité
70%	0,83	0,87
80%	0,92	0,88
90%	0,96	
100%	0,99	

- Jeux d'essais :

- Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
- Caractéristiques :
 - 100 ensembles de séquences ;
 - chaque ensemble contient au moins 10 séquences ;
 - chaque séquence est de longueur > 100 ;
 - au moins $\lceil q t \rceil$ séquences contiennent chacun des 3 motifs aléatoirement altérés, insérés à des positions aléatoires.

- Résultats en moyenne :

Contrainte	VPP	Sensibilité
70%	0,83	0,87
80%	0,92	0,88
90%	0,96	0,92
100%	0,99	

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

StARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- Jeux d'essais :

- Séquences générées aléatoirement dans lesquelles ont été insérés des motifs.
- Caractéristiques :
 - 100 ensembles de séquences ;
 - chaque ensemble contient au moins 10 séquences ;
 - chaque séquence est de longueur > 100 ;
 - au moins $\lceil q t \rceil$ séquences contiennent chacun des 3 motifs aléatoirement altérés, insérés à des positions aléatoires.

- Résultats en moyenne :

Contrainte	VPP	Sensibilité
70%	0,83	0,87
80%	0,92	0,88
90%	0,96	0,92
100%	0,99	0,90

- Jeux d'essais :
 - Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :

n°	Nb. de séq.	Longueur [min, max]	Protéine
1	9	[53, 806]	ArgR
2	8	[53, 806]	ArgR (site en tandem)
3	16	[100, 384]	LexA
4	18	[299, 5795]	PurR
5	8	[251, 2594]	TyrR

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

StARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- Jeux d'essais :

- Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :

n°	Nb. de séq.	Longueur [min, max]	Protéine
1	9	[53, 806]	ArgR
2	8	[53, 806]	ArgR (site en tandem)
3	16	[100, 384]	LexA
4	18	[299, 5795]	PurR
5	8	[251, 2594]	TyrR

- Analyse des résultats :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

StARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- Jeux d'essais :

- Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :

n°	Nb. de séq.	Longueur [min, max]	Protéine
1	9	[53, 806]	ArgR
2	8	[53, 806]	ArgR (site en tandem)
3	16	[100, 384]	LexA
4	18	[299, 5795]	PurR
5	8	[251, 2594]	TyrR

- Analyse des résultats :

- résultats plus fins que ceux de STARS (déjà plus performant que PRATT, TEIRESIAS, MEME, WINNOWER et SMILE) ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- Jeux d'essais :

- Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :

n°	Nb. de séq.	Longueur [min, max]	Protéine
1	9	[53, 806]	ArgR
2	8	[53, 806]	ArgR (site en tandem)
3	16	[100, 384]	LexA
4	18	[299, 5795]	PurR
5	8	[251, 2594]	TyrR

- Analyse des résultats :

- résultats plus fins que ceux de STARS (déjà plus performant que PRATT, TEIRESIAS, MEME, WINNOWER et SMILE) ;
- meilleure complexité que STARS ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- Jeux d'essais :

- Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :

n°	Nb. de séq.	Longueur [min, max]	Protéine
1	9	[53, 806]	ArgR
2	8	[53, 806]	ArgR (site en tandem)
3	16	[100, 384]	LexA
4	18	[299, 5795]	PurR
5	8	[251, 2594]	TyrR

- Analyse des résultats :

- résultats plus fins que ceux de STARS (déjà plus performant que PRATT, TEIRESIAS, MEME, WINNOWER et SMILE) ;
- meilleure complexité que STARS ;
- gestion de la contrainte de quorum ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

• Jeux d'essais :

- Séquences biologiques (*Escherichia coli*), contenant des sites de liaison aux protéines, validés expérimentalement :

n°	Nb. de séq.	Longueur [min, max]	Protéine
1	9	[53, 806]	ArgR
2	8	[53, 806]	ArgR (site en tandem)
3	16	[100, 384]	LexA
4	18	[299, 5795]	PurR
5	8	[251, 2594]	TyrR

• Analyse des résultats :

- résultats plus fins que ceux de STARS (déjà plus performant que PRATT, TEIRESIAS, MEME, WINNOWER et SMILE) ;
- meilleure complexité que STARS ;
- gestion de la contrainte de quorum ;
- indication statistique sur la pertinence des solutions.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

- **Origine :** Structure introduite en 1999 par Cyril ALLAUZEN, Maxime CROCHEMORE et Mathieu RAFFINOT.
- **Objectif :**

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

*Présentation et Construction des
Oracles*

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- **Origine :** Structure introduite en 1999 par Cyril ALLAUZEN, Maxime CROCHEMORE et Mathieu RAFFINOT.
- **Objectif :**

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

*Présentation et Construction des
Oracles*

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- **Origine** : Structure introduite en 1999 par Cyril ALLAUZEN, Maxime CROCHEMORE et Mathieu RAFFINOT.

- **Objectif** : construire un automate pour un mot s donné qui soit :

- acyclique ;

- minimal ;

- sans transitions inutiles (transitions qui ne sont jamais prises) ;

- déterministe ;

- complet ;

- ...

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- **Origine** : Structure introduite en 1999 par Cyril ALLAUZEN, Maxime CROCHEMORE et Mathieu RAFFINOT.
- **Objectif** : construire un automate pour un mot s donné qui soit :
 - acyclique ;
 - déterministe ;
 - qui reconnaisse au moins les facteurs/suffixes de s ;
 - minimal en nombre d'états ;
 - linéaire en nombre de transitions.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- **Origine** : Structure introduite en 1999 par Cyril ALLAUZEN, Maxime CROCHEMORE et Mathieu RAFFINOT.
- **Objectif** : construire un automate pour un mot s donné qui soit :
 - acyclique ;
 - déterministe ;
 - qui reconnaisse au moins les facteurs/suffixes de s ;
 - minimal en nombre d'états ;
 - linéaire en nombre de transitions.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- **Origine** : Structure introduite en 1999 par Cyril ALLAUZEN, Maxime CROCHEMORE et Mathieu RAFFINOT.
- **Objectif** : construire un automate pour un mot s donné qui soit :
 - acyclique ;
 - déterministe ;
 - qui reconnaisse au moins les facteurs/suffixes de s ;
 - minimal en nombre d'états ;
 - linéaire en nombre de transitions.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- **Origine** : Structure introduite en 1999 par Cyril ALLAUZEN, Maxime CROCHEMORE et Mathieu RAFFINOT.
- **Objectif** : construire un automate pour un mot s donné qui soit :
 - acyclique ;
 - déterministe ;
 - qui reconnaisse au moins les facteurs/suffixes de s ;
 - minimal en nombre d'états ;
 - linéaire en nombre de transitions.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- **Origine** : Structure introduite en 1999 par Cyril ALLAUZEN, Maxime CROCHEMORE et Mathieu RAFFINOT.
- **Objectif** : construire un automate pour un mot s donné qui soit :
 - acyclique ;
 - déterministe ;
 - qui reconnaisse au moins les facteurs/suffixes de s ;
 - minimal en nombre d'états ;
 - linéaire en nombre de transitions.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- **Origine** : Structure introduite en 1999 par Cyril ALLAUZEN, Maxime CROCHEMORE et Mathieu RAFFINOT.
- **Objectif** : construire un automate pour un mot s donné qui soit :
 - acyclique ;
 - déterministe ;
 - qui reconnaisse au moins les facteurs/suffixes de s ;
 - minimal en nombre d'états ;
 - linéaire en nombre de transitions.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- **Origine** : Structure introduite en 1999 par Cyril ALLAUZEN, Maxime CROCHEMORE et Mathieu RAFFINOT.
- **Objectif** : construire un automate pour un mot s donné qui soit :
 - acyclique ;
 - déterministe ;
 - qui reconnaisse au moins les facteurs/suffixes de s ;
 - minimal en nombre d'états ;
 - linéaire en nombre de transitions.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

*Présentation et Construction des
Oracles*

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

• Intérêts principaux :

- facilité d'implémentation/d'utilisation ;
- complexités temporelle et spatiale.

• Utilisation :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

*Présentation et Construction des
Oracles*

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

• Intérêts principaux :

- facilité d'implémentation/d'utilisation ;
- complexités temporelle et spatiale.

• Utilisation :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

*Présentation et Construction des
Oracles*

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

• Intérêts principaux :

- facilité d'implémentation/d'utilisation ;
- complexités temporelle et spatiale.

• Utilisation :

- [Multi-][Turbo-]BOM (Backward Oracle Matching) : recherche de motifs dans une ou plusieurs séquences [ALLAUZEN & al., 2001] ;
- COMPOR : compression de textes [LEFFEVRE & al., 2002] ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

• Intérêts principaux :

- facilité d'implémentation/d'utilisation ;
- complexités temporelle et spatiale.

• Utilisation :

- [Multi-][Turbo-]BOM (Backward Oracle Matching) : recherche de motifs dans une ou plusieurs séquences [ALLAUZEN & *al.*, 2001] ;
- COMPROR : compression de textes [LEFEBVRE & *al.*, 2002] ;
- FORREPEATS : détection de motifs répétés dans de grandes séquences [LEFEBVRE & *al.*, 2003] ;
- *OpenMusic* : improvisation musicale assistée par ordinateur [ASSAYAG & *al.*, 2004] ;
- POPITAM : identification de protéines à partir de spectrométries de masses [HERNANDEZ & *al.*, 2003].

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

• Intérêts principaux :

- facilité d'implémentation/d'utilisation ;
- complexités temporelle et spatiale.

• Utilisation :

- [Multi-][Turbo-]BOM (Backward Oracle Matching) : recherche de motifs dans une ou plusieurs séquences [ALLAUZEN & *al.*, 2001] ;
- COMPROR : compression de textes [LEFEBVRE & *al.*, 2002] ;
- FORREPEATS : détection de motifs répétés dans de grandes séquences [LEFEBVRE & *al.*, 2003] ;
- *OpenMusic* : improvisation musicale assistée par ordinateur [ASSAYAG & *al.*, 2004] ;
- POPITAM : identification de protéines à partir de spectrométries de masses [HERNANDEZ & *al.*, 2003].

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

• Intérêts principaux :

- facilité d'implémentation/d'utilisation ;
- complexités temporelle et spatiale.

• Utilisation :

- [Multi-][Turbo-]BOM (Backward Oracle Matching) : recherche de motifs dans une ou plusieurs séquences [ALLAUZEN & *al.*, 2001] ;
- COMPROR : compression de textes [LEFEBVRE & *al.*, 2002] ;
- FORREPEATS : détection de motifs répétés dans de grandes séquences [LEFEBVRE & *al.*, 2003] ;
- *OpenMusic* : improvisation musicale assistée par ordinateur [ASSAYAG & *al.*, 2004] ;
- POPITAM : identification de protéines à partir de spectrométries de masses [HERNANDEZ & *al.*, 2003].

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

• Intérêts principaux :

- facilité d'implémentation/d'utilisation ;
- complexités temporelle et spatiale.

• Utilisation :

- [Multi-][Turbo-]BOM (Backward Oracle Matching) : recherche de motifs dans une ou plusieurs séquences [ALLAUZEN & *al.*, 2001] ;
- COMPROR : compression de textes [LEFEBVRE & *al.*, 2002] ;
- FORREPEATS : détection de motifs répétés dans de grandes séquences [LEFEBVRE & *al.*, 2003] ;
- *OpenMusic* : improvisation musicale assistée par ordinateur [ASSAYAG & *al.*, 2004] ;
- POPITAM : identification de protéines à partir de spectrométries de masses [HERNANDEZ & *al.*, 2003].

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

• Intérêts principaux :

- facilité d'implémentation/d'utilisation ;
- complexités temporelle et spatiale.

• Utilisation :

- [Multi-][Turbo-]BOM (Backward Oracle Matching) : recherche de motifs dans une ou plusieurs séquences [ALLAUZEN & *al.*, 2001] ;
- COMPROR : compression de textes [LEFEBVRE & *al.*, 2002] ;
- FORREPEATS : détection de motifs répétés dans de grandes séquences [LEFEBVRE & *al.*, 2003] ;
- *OpenMusic* : improvisation musicale assistée par ordinateur [ASSAYAG & *al.*, 2004] ;
- POPITAM : identification de protéines à partir de spectrométries de masses [HERNANDEZ & *al.*, 2003].

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

• Intérêts principaux :

- facilité d'implémentation/d'utilisation ;
- complexités temporelle et spatiale.

• Utilisation :

- [Multi-][Turbo-]BOM (Backward Oracle Matching) : recherche de motifs dans une ou plusieurs séquences [ALLAUZEN & *al.*, 2001] ;
- COMPROR : compression de textes [LEFEBVRE & *al.*, 2002] ;
- FORREPEATS : détection de motifs répétés dans de grandes séquences [LEFEBVRE & *al.*, 2003] ;
- *OpenMusic* : improvisation musicale assistée par ordinateur [ASSAYAG & *al.*, 2004] ;
- POPITAM : identification de protéines à partir de spectrométries de masses [HERNANDEZ & *al.*, 2003].

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes du mot *gaccattctc*



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

*Présentation et Construction des
Oracles*

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes du mot *gaccattctc*

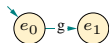


$$i + 1 = 1$$

$$s[i + 1] = g$$

$$S_\varepsilon(0) = -1$$

Oracle des Suffixes du mot *gaccattctc*



$$i + 1 = 1$$

$$s[i + 1] = g$$

$$S_\varepsilon(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

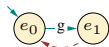
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes du mot *gaccattctc*



$$i + 1 = 2$$

$$s[i + 1] = a$$

$$S_g(1) = 0$$

$$S_g(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

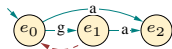
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes du mot *gaccatttc*



$$i + 1 = 2$$

$$s[i + 1] = a$$

$$S_g(1) = 0$$

$$S_g(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

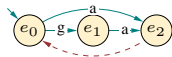
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes du mot *gaccattctc*



$$i + 1 = 3$$

$$s[i + 1] = c$$

$$S_{ga}(2) = 0$$

$$S_{ga}(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

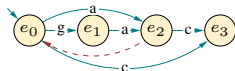
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes du mot *gaccattctc*



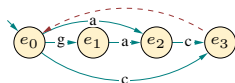
$$i + 1 = 3$$

$$s[i + 1] = c$$

$$S_{ga}(2) = 0$$

$$S_{ga}(0) = -1$$

Oracle des Suffixes du mot *gaccattctc*



$$i + 1 = 4$$

$$s[i + 1] = c$$

$$S_{gac}(3) = 0$$

$$S_{gac}(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

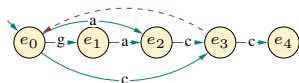
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes du mot *gaccattctc*



$$i + 1 = 4$$

$$s[i + 1] = c$$

$$S_{gac}(3) = 0$$

$$S_{gac}(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

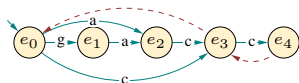
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes du mot *gaccattctc*



$$i + 1 = 5$$

$$s[i + 1] = a$$

$$S_{gacc}(4) = 3$$

$$S_{gacc}(3) = 0$$

$$S_{gacc}(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

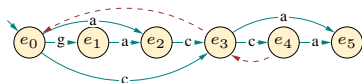
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes du mot *gaccattctc*



$$i + 1 = 5$$

$$s[i + 1] = a$$

$$S_{gacc}(4) = 3$$

$$S_{gacc}(3) = 0$$

$$S_{gacc}(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

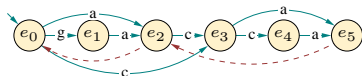
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes du mot *gaccattctc*



$$i + 1 = 6$$

$$s[i + 1] = t$$

$$S_{gacca}(5) = 2$$

$$S_{gacca}(2) = 0$$

$$S_{gacca}(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

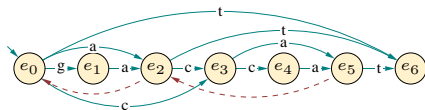
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes du mot *gaccattctc*



$$i + 1 = 6$$

$$s[i + 1] = t$$

$$S_{gacca}(5) = 2$$

$$S_{gacca}(2) = 0$$

$$S_{gacca}(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

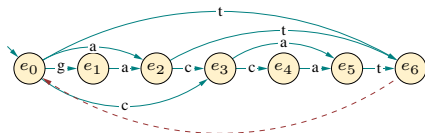
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes du mot *gaccattctc*



$$i + 1 = 7$$

$$s[i + 1] = t$$

$$S_{gaccat}(6) = 0$$

$$S_{gaccat}(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

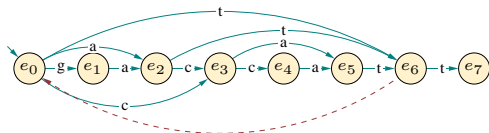
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes du mot *gaccattctc*



$$i + 1 = 7$$

$$s[i + 1] = t$$

$$S_{gac\text{c}at}(6) = 0$$

$$S_{gac\text{c}at}(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

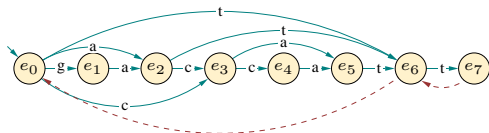
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes du mot *gaccattctc*



$$i + 1 = 8$$

$$s[i + 1] = c$$

$$S_{gaccatt}(7) = 6$$

$$S_{gaccatt}(6) = 0$$

$$S_{gaccatt}(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

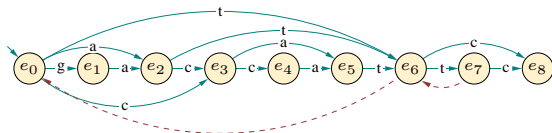
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes du mot *gaccattctc*



$$i + 1 = 8$$

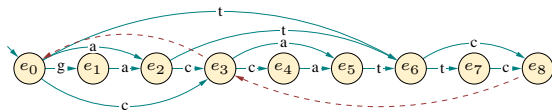
$$s[i + 1] = c$$

$$S_{gaccatt}(7) = 6$$

$$S_{gaccatt}(6) = 0$$

$$S_{gaccatt}(0) = -1$$

Oracle des Suffixes du mot *gaccatttc*



$$i + 1 = 9$$

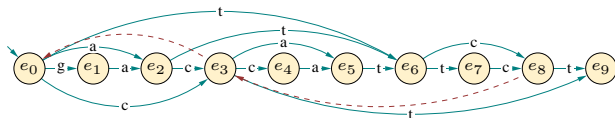
$$s[i + 1] = t$$

$$S_{gaccatttc}(8) = 3$$

$$S_{gaccatttc}(3) = 0$$

$$S_{gaccatttc}(0) = -1$$

Oracle des Suffixes du mot *gaccatttc*



$$i + 1 = 9$$

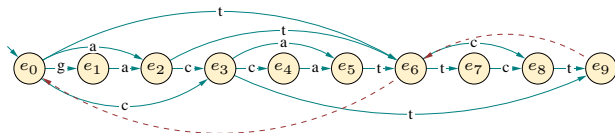
$$s[i + 1] = t$$

$$S_{gaccatttc}(8) = 3$$

$$S_{gaccatttc}(3) = 0$$

$$S_{gaccatttc}(0) = -1$$

Oracle des Suffixes du mot *gaccattctc*



$$i + 1 = 10$$

$$s[i + 1] = c$$

$$S_{gaccattctc}(9) = 6$$

$$S_{gaccattctc}(6) = 0$$

$$S_{gaccattctc}(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

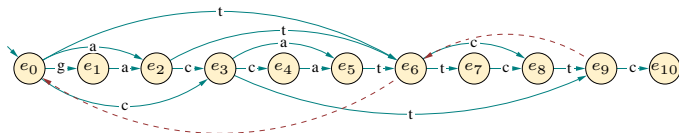
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes du mot *gaccattctc*



$$i + 1 = 10$$

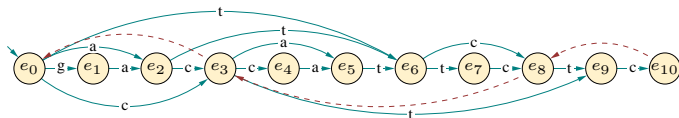
$$s[i + 1] = c$$

$$S_{gaccattctc}(9) = 6$$

$$S_{gaccattctc}(6) = 0$$

$$S_{gaccattctc}(0) = -1$$

Oracle des Suffixes du mot *gaccattctc*



$$S_{gaccattctc}(10) = 8$$

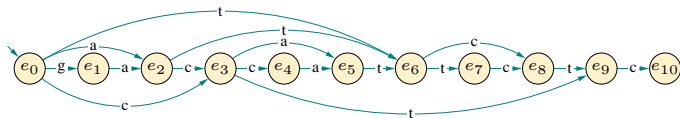
$$S_{gaccattctc}(8) = 3$$

$$S_{gaccattctc}(3) = 0$$

Oracles

Facteurs

Suffixes

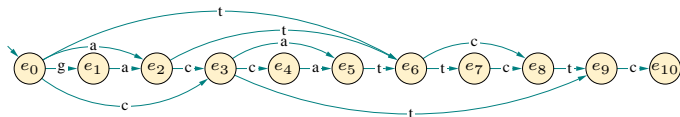


Oracles

Facteurs

Suffixes

Tous les états sont finaux

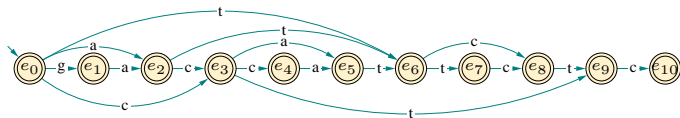


Oracles

Facteurs

Suffixes

Tous les états sont finaux



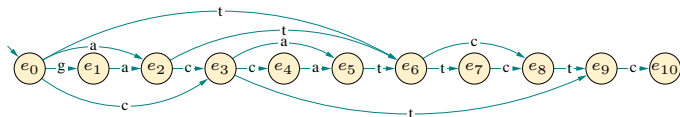
Oracles

Facteurs

Suffixes

Tous les états sont finaux

Seuls les états e_i tels qu'il existe un chemin étiquetant un suffixe de s sont terminaux



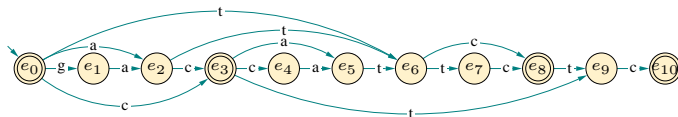
Oracles

Facteurs

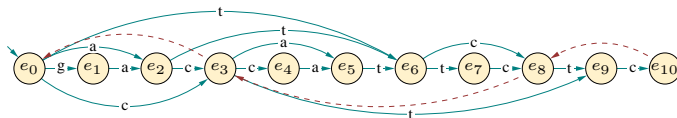
Suffixes

Tous les états sont finaux

Seuls les états e_i tels qu'il existe un chemin étiquetant un suffixe de s sont terminaux



Oracle des Suffixes du mot *gaccattctc*

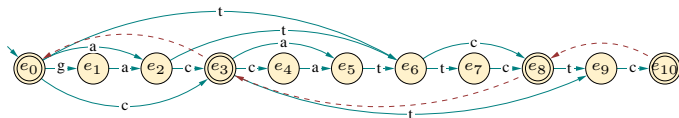


$$S_{gaccattctc}(10) = 8$$

$$S_{gaccattctc}(8) = 3$$

$$S_{gaccattctc}(3) = 0$$

Oracle des Suffixes du mot *gaccatttc*



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

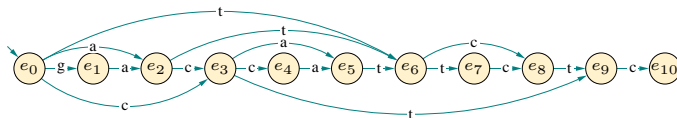
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Facteurs du mot *gaccattctc*



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

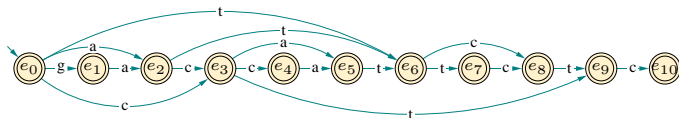
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Facteurs du mot *gaccattctc*



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- L'Oracle des Facteurs/Suffixes de *s* peut reconnaître des mots qui ne sont pas facteurs/suffixes de *s*.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

*Présentation et Construction des
Oracles*

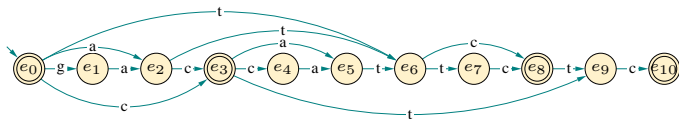
Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- L'Oracle des Facteurs/Suffixes de *s* peut reconnaître des mots qui ne sont pas facteurs/suffixes de *s*.

$atc \notin \text{Fact}(gaccattctc)$
 $\notin \text{Suff}(gaccattctc)$



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des
Oracles

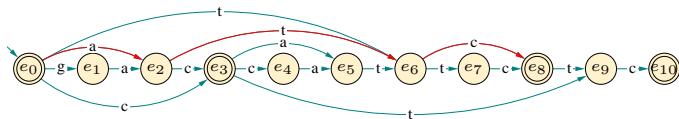
Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- L'Oracle des Facteurs/Suffixes de *s* peut reconnaître des mots qui ne sont pas facteurs/suffixes de *s*.

$atc \notin \text{Fact}(gaccattctc)$
 $\notin \text{Suff}(gaccattctc)$



- L'Oracle des Facteurs/Suffixes de s peut reconnaître des mots qui ne sont pas facteurs/suffixes de s .
- L'Oracle des Facteurs/Suffixes n'est pas minimal en nombre de mots reconnus dans l'ensemble des automates homogènes à $n + 1$ états reconnaissant au moins les facteurs/suffixes de s .

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- L'Oracle des Facteurs/Suffixes de s peut reconnaître des mots qui ne sont pas facteurs/suffixes de s .
- L'Oracle des Facteurs/Suffixes n'est pas minimal en nombre de mots reconnus dans l'ensemble des automates homogènes à $n + 1$ états reconnaissant au moins les facteurs/suffixes de s .

⇒ Besoin de caractériser le langage reconnu par les Oracles !

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- L'Oracle des Facteurs/Suffixes de s peut reconnaître des mots qui ne sont pas facteurs/suffixes de s .
- L'Oracle des Facteurs/Suffixes n'est pas minimal en nombre de mots reconnus dans l'ensemble des automates homogènes à $n + 1$ états reconnaissant au moins les facteurs/suffixes de s .

⇒ Besoin de caractériser le langage reconnu par les Oracles !

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- L'Oracle des Facteurs/Suffixes de s peut reconnaître des mots qui ne sont pas facteurs/suffixes de s .
- L'Oracle des Facteurs/Suffixes n'est pas minimal en nombre de mots reconnus dans l'ensemble des automates homogènes à $n + 1$ états reconnaissant au moins les facteurs/suffixes de s .

⇒ Besoin de caractériser le langage reconnu par les Oracles !

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- L'Oracle des Facteurs/Suffixes de s peut reconnaître des mots qui ne sont pas facteurs/suffixes de s .
- L'Oracle des Facteurs/Suffixes n'est pas minimal en nombre de mots reconnus dans l'ensemble des automates homogènes à $n + 1$ états reconnaissant au moins les facteurs/suffixes de s .

⇒ Besoin de caractériser le langage reconnu par les Oracles !

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- L'Oracle des Facteurs/Suffixes de s peut reconnaître des mots qui ne sont pas facteurs/suffixes de s .
- L'Oracle des Facteurs/Suffixes n'est pas minimal en nombre de mots reconnus dans l'ensemble des automates homogènes à $n + 1$ états reconnaissant au moins les facteurs/suffixes de s .

⇒ Besoin de caractériser le langage reconnu par les Oracles !

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Définition (Facteurs Remarquables)

Étant donné un mot $s \in \Sigma^*$ et son Oracle, l'ensemble des facteurs remarquables de s , noté \mathcal{F}_s , est défini par :

$$\mathcal{F}_s = \{ \min(e_i) \mid 1 \leq i < |s| \wedge (\#_{out}(e_i) > 1 \vee \min(e_i) \in Suff(s)) \},$$

où $\#_{out}(e_i)$ désigne le degré sortant de l'état e_i .

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

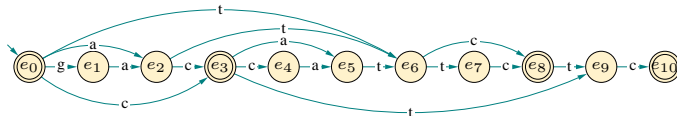
Conclusions & perspectives

Définition (Facteurs Remarquables)

Étant donné un mot $s \in \Sigma^*$ et son Oracle, l'ensemble des facteurs remarquables de s , noté \mathcal{F}_s , est défini par :

$$\mathcal{F}_s = \{ \min(e_i) \mid 1 \leq i < |s| \wedge (\#_{out}(e_i) > 1 \vee \min(e_i) \in Suff(s)) \},$$

où $\#_{out}(e_i)$ désigne le degré sortant de l'état e_i .

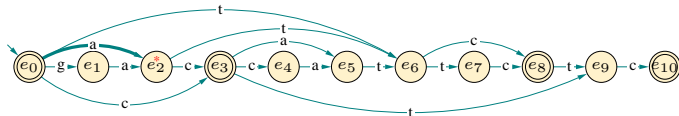


Définition (Facteurs Remarquables)

Étant donné un mot $s \in \Sigma^*$ et son Oracle, l'ensemble des facteurs remarquables de s , noté \mathcal{F}_s , est défini par :

$$\mathcal{F}_s = \{ \min(e_i) \mid 1 \leq i < |s| \wedge (\#_{out}(e_i) > 1 \vee \min(e_i) \in Suff(s)) \},$$

où $\#_{out}(e_i)$ désigne le degré sortant de l'état e_i .

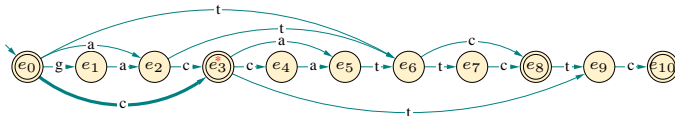


Définition (Facteurs Remarquables)

Étant donné un mot $s \in \Sigma^*$ et son Oracle, l'ensemble des facteurs remarquables de s , noté \mathcal{F}_s , est défini par :

$$\mathcal{F}_s = \{ \min(e_i) \mid 1 \leq i < |s| \wedge (\#_{out}(e_i) > 1 \vee \min(e_i) \in Suff(s)) \},$$

où $\#_{out}(e_i)$ désigne le degré sortant de l'état e_i .

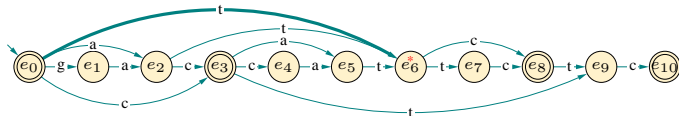


Définition (Facteurs Remarquables)

Étant donné un mot $s \in \Sigma^*$ et son Oracle, l'ensemble des facteurs remarquables de s , noté \mathcal{F}_s , est défini par :

$$\mathcal{F}_s = \{ \min(e_i) \mid 1 \leq i < |s| \wedge (\#_{out}(e_i) > 1 \vee \min(e_i) \in Suff(s)) \},$$

où $\#_{out}(e_i)$ désigne le degré sortant de l'état e_i .

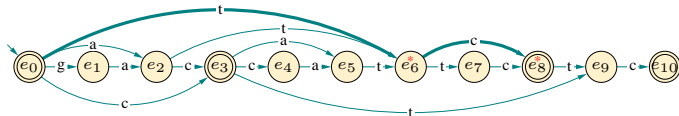


Définition (Facteurs Remarquables)

Étant donné un mot $s \in \Sigma^*$ et son Oracle, l'ensemble des facteurs remarquables de s , noté \mathcal{F}_s , est défini par :

$$\mathcal{F}_s = \{ \min(e_i) \mid 1 \leq i < |s| \wedge (\#_{out}(e_i) > 1 \vee \min(e_i) \in Suff(s)) \},$$

où $\#_{out}(e_i)$ désigne le degré sortant de l'état e_i .

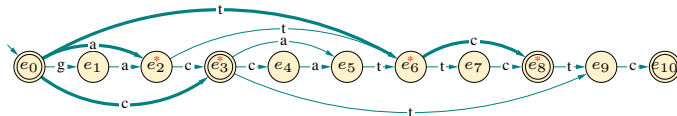


Définition (Facteurs Remarquables)

Étant donné un mot $s \in \Sigma^*$ et son Oracle, l'ensemble des facteurs remarquables de s , noté \mathcal{F}_s , est défini par :

$$\mathcal{F}_s = \{ \min(e_i) \mid 1 \leq i < |s| \wedge (\#_{out}(e_i) > 1 \vee \min(e_i) \in Suff(s)) \},$$

où $\#_{out}(e_i)$ désigne le degré sortant de l'état e_i .



$$\mathcal{F}_{gaccattctc} = \{a, c, t, tc\}$$

Exemple (contraction)

$$\mathcal{F}_{gaccattctc} = \{a, c, t, tc\}.$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

*Présentation et Construction des
Oracles*

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Exemple (contraction)

$$\mathcal{F}_{gaccattctc} = \{a, c, t, tc\}.$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

*Présentation et Construction des
Oracles*

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Exemple (contraction)

$\mathcal{F}_{gaccattctc} = \{a, c, t, tc\}$.

La paire (3, 8) est une contraction

$$\Rightarrow gaccattctc = gactc.$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

*Présentation et Construction des
Oracles*

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Exemple (Ensemble de contraction)

$$\mathcal{F}_{gaccattctc} = \{a, c, t, tc\} \Rightarrow \left\{ \begin{array}{l} (2, 5), (3, 4), \\ (3, 8), (3, 10), \\ (6, 7), (6, 9), \\ (7, 9) \end{array} \right\}$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

*Présentation et Construction des
Oracles*

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Exemple (Ensemble de contraction)

$$\mathcal{F}_{gaccattctc} = \{a, c, t, tc\} \Rightarrow \left\{ \begin{array}{l} (2, 5), (3, 4), \\ (3, 8), (3, 10), \\ (6, 7), (6, 9), \\ (7, 9) \end{array} \right\}$$

Exemple (Calcul des ensembles de contractions)

$$\left\{ \begin{array}{l} \{\} \\ \{(2, 5)\} \\ \{(3, 4)\} \\ \{(3, 8)\} \\ \{(3, 10)\} \\ \{(6, 7)\} \\ \{(6, 9)\} \\ \{(7, 9)\} \\ \{(2, 5), (6, 7)\} \\ \{(2, 5), (6, 9)\} \\ \{(2, 5), (7, 9)\} \\ \{(3, 4), (6, 7)\} \\ \{(3, 4), (6, 9)\} \\ \{(3, 4), (7, 9)\} \end{array} \right\}$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

*Présentation et Construction des
Oracles*

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Exemple (Ensemble de contraction)

$$\mathcal{F}_{gaccattctc} = \{a, c, t, tc\} \Rightarrow \left\{ \begin{array}{l} (2, 5), (3, 4), \\ (3, 8), (3, 10), \\ (6, 7), (6, 9), \\ (7, 9) \end{array} \right\}$$

Exemple (Calcul des ensembles de contractions)

{	{}	⇒	gaccattctc
	{(2, 5)}	⇒	gaccattctc
	{(3, 4)}	⇒	gaccattctc
	{(3, 8)}	⇒	gaccattctc
	{(3, 10)}	⇒	gaccattctc
	{(6, 7)}	⇒	gaccattctc
	{(6, 9)}	⇒	gaccattctc
	{(7, 9)}	⇒	gaccattctc
	{(2, 5), (6, 7)}	⇒	gaccattctc
	{(2, 5), (6, 9)}	⇒	gaccattctc
	{(2, 5), (7, 9)}	⇒	gaccattctc
	{(3, 4), (6, 7)}	⇒	gaccattctc
	{(3, 4), (6, 9)}	⇒	gaccattctc
	{(3, 4), (7, 9)}	⇒	gaccattctc

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des
Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Exemple (Ensemble de contraction)

$$\mathcal{F}_{gaccattctc} = \{a, c, t, tc\} \Rightarrow \left\{ \begin{array}{l} (2, 5), (3, 4), \\ (3, 8), (3, 10), \\ (6, 7), (6, 9), \\ (7, 9) \end{array} \right\}$$

Exemple (Calcul des ensembles de contractions)

{ }	\Rightarrow	<i>gaccattctc</i>	=	<i>gaccattctc</i>
{ (2, 5) }	\Rightarrow	<i>gaccattctc</i>	=	<i>gattctc</i>
{ (3, 4) }	\Rightarrow	<i>gaccattctc</i>	=	<i>gacattctc</i>
{ (3, 8) }	\Rightarrow	<i>gaccattctc</i>	=	<i>gactc</i>
{ (3, 10) }	\Rightarrow	<i>gaccattctc</i>	=	<i>gac</i>
{ (6, 7) }	\Rightarrow	<i>gaccattctc</i>	=	<i>gaccatctc</i>
{ (6, 9) }	\Rightarrow	<i>gaccattctc</i>	=	<i>gaccatc</i>
{ (7, 9) }	\Rightarrow	<i>gaccattctc</i>	=	<i>gaccattc</i>
{ (2, 5), (6, 7) }	\Rightarrow	<i>gaccattctc</i>	=	<i>gatctc</i>
{ (2, 5), (6, 9) }	\Rightarrow	<i>gaccattctc</i>	=	<i>gatc</i>
{ (2, 5), (7, 9) }	\Rightarrow	<i>gaccattctc</i>	=	<i>gattc</i>
{ (3, 4), (6, 7) }	\Rightarrow	<i>gaccattctc</i>	=	<i>gacatctc</i>
{ (3, 4), (6, 9) }	\Rightarrow	<i>gaccattctc</i>	=	<i>gacatc</i>
{ (3, 4), (7, 9) }	\Rightarrow	<i>gaccattctc</i>	=	<i>gacattc</i>

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Exemple (Fermeture)

La fermeture du mot *gaccattctc* est :

Exemple (Calcul des ensembles de contractions)

{ }	⇒	<i>gaccattctc</i>	=	<i>gaccattctc</i>
{ (2, 5) }	⇒	<i>gaccattctc</i>	=	<i>gattctc</i>
{ (3, 4) }	⇒	<i>gaccattctc</i>	=	<i>gacattctc</i>
{ (3, 8) }	⇒	<i>gaccattctc</i>	=	<i>gactc</i>
{ (3, 10) }	⇒	<i>gaccattctc</i>	=	<i>gac</i>
{ (6, 7) }	⇒	<i>gaccattctc</i>	=	<i>gaccatctc</i>
{ (6, 9) }	⇒	<i>gaccattctc</i>	=	<i>gaccatc</i>
{ (7, 9) }	⇒	<i>gaccattctc</i>	=	<i>gaccattc</i>
{ (2, 5), (6, 7) }	⇒	<i>gaccattctc</i>	=	<i>gatctc</i>
{ (2, 5), (6, 9) }	⇒	<i>gaccattctc</i>	=	<i>gatc</i>
{ (2, 5), (7, 9) }	⇒	<i>gaccattctc</i>	=	<i>gattc</i>
{ (3, 4), (6, 7) }	⇒	<i>gaccattctc</i>	=	<i>gacatctc</i>
{ (3, 4), (6, 9) }	⇒	<i>gaccattctc</i>	=	<i>gacatc</i>
{ (3, 4), (7, 9) }	⇒	<i>gaccattctc</i>	=	<i>gacattc</i>

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des
Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Exemple (Fermeture)

La fermeture du mot *gaccattctc* est :

$$\mathcal{E}(gaccattctc) = \left\{ \begin{array}{l} gac, gacatc, gacatctc, gacattc, gaccattctc, \\ gaccatc, gaccatctc, gaccattc, gaccattctc, \\ gactc, gatc, gatctc, gattc, gattctc \end{array} \right\}.$$

Exemple (Calcul des ensembles de contractions)

{ }	⇒	<i>gaccattctc</i>	=	<i>gaccattctc</i>
{ (2, 5) }	⇒	<i>gaccattctc</i>	=	<i>gattctc</i>
{ (3, 4) }	⇒	<i>gaccattctc</i>	=	<i>gacattctc</i>
{ (3, 8) }	⇒	<i>gaccattctc</i>	=	<i>gactc</i>
{ (3, 10) }	⇒	<i>gaccattctc</i>	=	<i>gac</i>
{ (6, 7) }	⇒	<i>gaccattctc</i>	=	<i>gaccatctc</i>
{ (6, 9) }	⇒	<i>gaccattctc</i>	=	<i>gaccatc</i>
{ (7, 9) }	⇒	<i>gaccattctc</i>	=	<i>gaccattc</i>
{ (2, 5), (6, 7) }	⇒	<i>gaccattctc</i>	=	<i>gatctc</i>
{ (2, 5), (6, 9) }	⇒	<i>gaccattctc</i>	=	<i>gatc</i>
{ (2, 5), (7, 9) }	⇒	<i>gaccattctc</i>	=	<i>gattc</i>
{ (3, 4), (6, 7) }	⇒	<i>gaccattctc</i>	=	<i>gacatctc</i>
{ (3, 4), (6, 9) }	⇒	<i>gaccattctc</i>	=	<i>gacatc</i>
{ (3, 4), (7, 9) }	⇒	<i>gaccattctc</i>	=	<i>gacattc</i>

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Théorème (Langage reconnu par les Oracles des Suffixes)

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

*Présentation et Construction des
Oracles*

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Théorème (Langage reconnu par les Oracles des Suffixes)

L'Oracle des Suffixes de s reconnaît tous (et seulement) les suffixes des mots de $\mathcal{E}(s)$.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

*Présentation et Construction des
Oracles*

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Théorème (Langage reconnu par les Oracles des Suffixes)

L'Oracle des Suffixes de s reconnaît tous (et seulement) les suffixes des mots de $\mathcal{E}(s)$.

Exemple

Pour *gaccattctc*,

$$\mathcal{E}(gaccattctc) = \left\{ \begin{array}{l} gac, gacatc, gacatctc, gacattc, gacattctc, \\ gaccatc, gaccatctc, gaccattc, gaccattctc, \\ gactc, gatc, gatctc, gattc, gattctc \end{array} \right\}.$$

Il y a 94 mots reconnus par l'Oracle des Facteurs de *gaccattctc*, dont ses 49 facteurs.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Théorème (Langage reconnu par les Oracles des Suffixes)

L'Oracle des Suffixes de s reconnaît tous (et seulement) les suffixes des mots de $\mathcal{E}(s)$.

Exemple

Pour *gaccattctc*,

$$\mathcal{E}(gaccattctc) = \left\{ \begin{array}{l} gac, gacatc, gacatctc, gacattc, gacattctc, \\ gaccatc, gaccatctc, gaccattc, gaccattctc, \\ gactc, gatc, gatctc, gattc, gattctc \end{array} \right\}.$$

Il y a 94 mots reconnus par l'Oracle des Facteurs de *gaccattctc*, dont ses 49 facteurs.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Théorème (Langage reconnu par les Oracles des Suffixes)

L'Oracle des Suffixes de s reconnaît tous (et seulement) les suffixes des mots de $\mathcal{E}(s)$.

Exemple

Pour *gaccattctc*,

$$\mathcal{E}(gaccattctc) = \left\{ \begin{array}{l} gac, gacatc, gacatctc, gacattc, gacattctc, \\ gaccatc, gaccatctc, gaccattc, gaccattctc, \\ gactc, gatc, gatctc, gattc, gattctc \end{array} \right\}.$$

Il y a 43 mots reconnus par l'Oracle des Suffixes de *gaccattctc*, dont ses 11 suffixes.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Théorème (Langage reconnu par les Oracles des Suffixes)

L'Oracle des Suffixes de s reconnaît tous (et seulement) les suffixes des mots de $\mathcal{E}(s)$.

Exemple

Pour *gaccattctc*,

$$\mathcal{E}(gaccattctc) = \left\{ \begin{array}{l} gac, gacatc, gacatctc, gacattc, gaccattctc, \\ gaccatc, gaccatctc, gaccattc, gaccattctc, \\ gactc, gatc, gatctc, gattc, gattctc \end{array} \right\}.$$

Il y a 94 mots reconnus par l'Oracle des Facteurs de *gaccattctc*, dont ses 49 facteurs.

Théorème (Langage reconnu par les Oracles des Facteurs)

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Théorème (Langage reconnu par les Oracles des Suffixes)

L'Oracle des Suffixes de s reconnaît tous (et seulement) les suffixes des mots de $\mathcal{E}(s)$.

Exemple

Pour *gaccattctc*,

$$\mathcal{E}(gaccattctc) = \left\{ \begin{array}{l} gac, gacatc, gacatctc, gacattc, gaccattctc, \\ gaccatc, gaccatctc, gaccattc, gaccattctc, \\ gactc, gatc, gatctc, gattc, gattctc \end{array} \right\}.$$

Il y a 94 mots reconnus par l'Oracle des Facteurs de *gaccattctc*, dont ses 49 facteurs.

Théorème (Langage reconnu par les Oracles des Facteurs)

L'Oracle des Facteurs de s reconnaît tous (et seulement) les facteurs des mots de $\mathcal{E}(s)$.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Théorème (Langage reconnu par les Oracles des Suffixes)

L'Oracle des Suffixes de s reconnaît tous (et seulement) les suffixes des mots de $\mathcal{E}(s)$.

Exemple

Pour *gaccattctc*,

$$\mathcal{E}(gaccattctc) = \left\{ \begin{array}{l} gac, gacatc, gacatctc, gacattc, gaccattctc, \\ gaccatc, gaccatctc, gaccattc, gaccattctc, \\ gactc, gatc, gatctc, gattc, gattctc \end{array} \right\}.$$

Il y a 94 mots reconnus par l'Oracle des Facteurs de *gaccattctc*, dont ses 49 facteurs.

Théorème (Langage reconnu par les Oracles des Facteurs)

L'Oracle des Facteurs de s reconnaît tous (et seulement) les facteurs des mots de $\mathcal{E}(s)$.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- Un mot rejeté par l'Oracle des Facteurs/Suffixes d'une séquence s n'est pas facteur/suffixe de s ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

*Présentation et Construction des
Oracles*

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- Un mot rejeté par l'Oracle des Facteurs/Suffixes d'une séquence s n'est pas facteur/suffixe de s ;
- Un mot accepté par l'Oracle des Facteurs/Suffixes d'une séquence s n'est pas toujours facteur/suffixe de s .

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- Un mot rejeté par l'Oracle des Facteurs/Suffixes d'une séquence s n'est pas facteur/suffixe de s ;
- Un mot accepté par l'Oracle des Facteurs/Suffixes d'une séquence s n'est pas toujours facteur/suffixe de s .
 - Notion de Vrai/Faux Positifs ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- Un mot rejeté par l'Oracle des Facteurs/Suffixes d'une séquence s n'est pas facteur/suffixe de s ;
- Un mot accepté par l'Oracle des Facteurs/Suffixes d'une séquence s n'est pas toujours facteur/suffixe de s .
 - Notion de Vrai/Faux Positifs ;
- Nombre de « Faux Positifs » peut-être exponentiel par rapport à $|s|$;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- Un mot rejeté par l'Oracle des Facteurs/Suffixes d'une séquence s n'est pas facteur/suffixe de s ;
- Un mot accepté par l'Oracle des Facteurs/Suffixes d'une séquence s n'est pas toujours facteur/suffixe de s .
 - Notion de Vrai/Faux Positifs ;
- Nombre de « Faux Positifs » peut-être exponentiel par rapport à $|s|$;
 - Faux Positifs engendrés par « contractions » (\Rightarrow arcs externes).

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- **Objectif :**

- gardes sur les transitions externes des Oracles.

- **Calcul :**

- mise à jour à partir de l'Oracle de $s[1..i]$.

- **Utilisation :**

- filtrer le passage par une transition externe selon la garde présente.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- **Objectif :**
 - gardes sur les transitions externes des Oracles.
- **Calcul :**
 - mise à jour à partir de l'Oracle de $s[1..i]$.
- **Utilisation :**
 - filtrer le passage par une transition externe selon la garde présente.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- **Objectif :**
 - gardes sur les transitions externes des Oracles.
- **Calcul :**
 - mise à jour à partir de l'Oracle de $s[1..i]$.
- **Utilisation :**
 - filtrer le passage par une transition externe selon la garde présente.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

*Présentation et Construction des
Oracles*

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- **Objectif :**
 - gardes sur les transitions externes des Oracles.
- **Calcul :**
 - mise à jour à partir de l'Oracle de $s[1..i]$.
- **Utilisation :**
 - filtrer le passage par une transition externe selon la garde présente.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

*Présentation et Construction des
Oracles*

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes à transitions gardées du mot *gaccattctc*



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

*Présentation et Construction des
Oracles*

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes à transitions gardées du mot *gaccattctc*



$$i + 1 = 1 \quad s[i + 1] = g$$

$$S_\varepsilon(0) = -1$$

$$L_\varepsilon(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

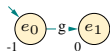
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes à transitions gardées du mot *gaccattctc*



$$i + 1 = 1 \quad s[i + 1] = g$$

$$S_\varepsilon(0) = -1$$

$$L_\varepsilon(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

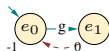
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes à transitions gardées du mot *gaccattctc*



$$i + 1 = 2 \quad s[i + 1] = a$$

$$S_g(1) = 0$$

$$L_g(1) = 0$$

$$S_g(0) = -1$$

$$L_g(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes à transitions gardées du mot *gaccattctc*



$$i + 1 = 2 \quad s[i + 1] = a$$

$$S_g(1) = 0$$

$$L_g(1) = 0$$

$$S_g(0) = -1$$

$$L_g(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes à transitions gardées du mot *gaccattctc*



$$i + 1 = 3 \quad s[i + 1] = c$$

$$S_{ga}(2) = 0$$

$$L_{ga}(2) = 0$$

$$S_{ga}(0) = -1$$

$$L_{ga}(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

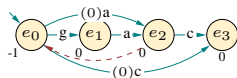
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes à transitions gardées du mot *gaccattctc*



$$i + 1 = 3 \quad s[i + 1] = c$$

$$S_{ga}(2) = 0$$

$$L_{ga}(2) = 0$$

$$S_{ga}(0) = -1$$

$$L_{ga}(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

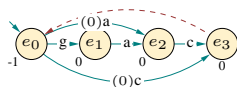
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes à transitions gardées du mot *gaccattctc*



$$i + 1 = 4 \quad s[i + 1] = c$$

$$S_{gac}(3) = 0$$

$$S_{gac}(0) = -1$$

$$L_{gac}(0) = 0$$

$$L_{gac}(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

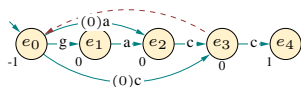
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes à transitions gardées du mot *gaccattctc*



$$i + 1 = 4 \quad s[i + 1] = c$$

$$S_{gac}(3) = 0$$

$$S_{gac}(0) = -1$$

$$L_{gac}(0) = 0$$

$$L_{gac}(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

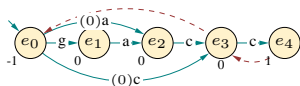
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes à transitions gardées du mot *gaccattctc*



$$i + 1 = 5 \quad s[i + 1] = a$$

$$S_{gacc}(4) = 3$$

$$L_{gacc}(4) = 1$$

$$S_{gacc}(3) = 0$$

$$L_{gacc}(3) = 0$$

$$S_{gacc}(0) = -1$$

$$L_{gacc}(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

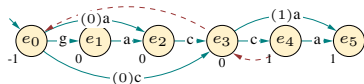
Présentation et Construction des
Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes à transitions gardées du mot *gaccattctc*



$$i + 1 = 5 \quad s[i + 1] = a$$

$$S_{gacc}(4) = 3$$

$$L_{gacc}(4) = 1$$

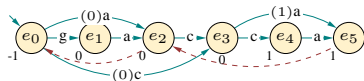
$$S_{gacc}(3) = 0$$

$$L_{gacc}(3) = 0$$

$$S_{gacc}(0) = -1$$

$$L_{gacc}(0) = -1$$

Oracle des Suffixes à transitions gardées du mot *gaccattctc*



$$i + 1 = 6 \quad s[i + 1] = t$$

$$S_{gacca}(5) = 2$$

$$L_{gacca}(5) = 1$$

$$S_{gacca}(2) = 0$$

$$L_{gacca}(2) = 0$$

$$S_{gacca}(0) = -1$$

$$L_{gacca}(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

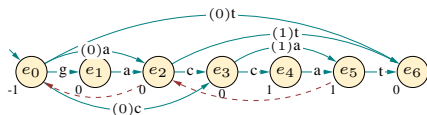
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes à transitions gardées du mot *gaccattctc*



$$i + 1 = 6 \quad s[i + 1] = t$$

$$S_{gacca}(5) = 2$$

$$L_{gacca}(5) = 1$$

$$S_{gacca}(2) = 0$$

$$L_{gacca}(2) = 0$$

$$S_{gacca}(0) = -1$$

$$L_{gacca}(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

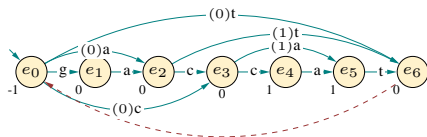
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes à transitions gardées du mot *gaccattctc*



$$i + 1 = 7 \quad s[i + 1] = t$$

$$S_{gacccat}(6) = 0$$

$$L_{gacccat}(6) = 0$$

$$S_{gacccat}(0) = -1$$

$$L_{gacccat}(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

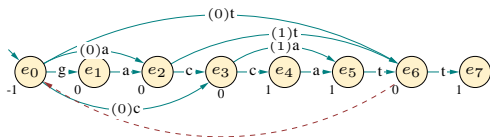
Présentation et Construction des
Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes à transitions gardées du mot *gaccattctc*



$$i + 1 = 7 \quad s[i + 1] = t$$

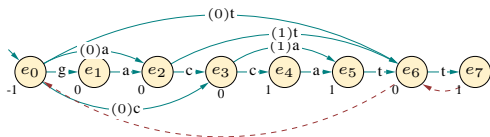
$$S_{gaccat}(6) = 0$$

$$L_{gaccat}(6) = 0$$

$$S_{gaccat}(0) = -1$$

$$L_{gaccat}(0) = -1$$

Oracle des Suffixes à transitions gardées du mot *gaccattctc*



$$i + 1 = 8 \quad s[i + 1] = c$$

$$S_{gaccatt}(7) = 6$$

$$L_{gaccatt}(7) = 1$$

$$S_{gaccatt}(6) = 0$$

$$L_{gaccatt}(6) = 0$$

$$S_{gaccatt}(0) = -1$$

$$L_{gaccatt}(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

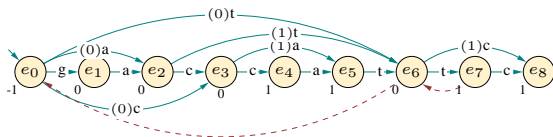
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes à transitions gardées du mot *gaccattctc*



$$i + 1 = 8 \quad s[i + 1] = c$$

$$S_{gaccatt}(7) = 6$$

$$L_{gaccatt}(7) = 1$$

$$S_{gaccatt}(6) = 0$$

$$L_{gaccatt}(6) = 0$$

$$S_{gaccatt}(0) = -1$$

$$L_{gaccatt}(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

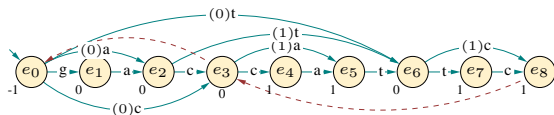
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes à transitions gardées du mot *gaccattctc*



$$i + 1 = 9 \quad s[i + 1] = t$$

$$S_{gaccattc}(8) = 3$$

$$L_{gaccattc}(8) = 1$$

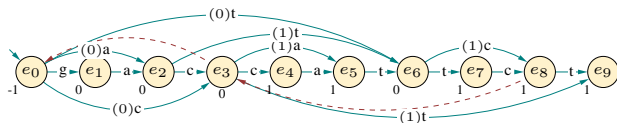
$$S_{gaccattc}(3) = 0$$

$$L_{gaccattc}(3) = 0$$

$$S_{gaccattc}(0) = -1$$

$$L_{gaccattc}(0) = -1$$

Oracle des Suffixes à transitions gardées du mot *gaccattctc*



$$i + 1 = 9 \quad s[i + 1] = t$$

$$S_{gaccattc}(8) = 3$$

$$L_{gaccattc}(8) = 1$$

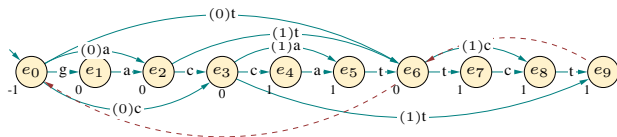
$$S_{gaccattc}(3) = 0$$

$$L_{gaccattc}(3) = 0$$

$$S_{gaccattc}(0) = -1$$

$$L_{gaccattc}(0) = -1$$

Oracle des Suffixes à transitions gardées du mot *gaccattctc*



$$i + 1 = 10 \quad s[i + 1] = c$$

$$S_{gaccattct}(9) = 3$$

$$L_{gaccattct}(9) = 1$$

$$S_{gaccattct}(3) = 0$$

$$L_{gaccattct}(3) = 0$$

$$S_{gaccattct}(0) = -1$$

$$L_{gaccattct}(0) = -1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

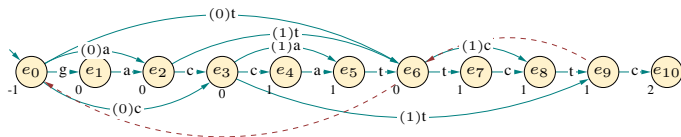
Présentation et Construction des
Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes à transitions gardées du mot *gaccattctc*



$$i + 1 = 10 \quad s[i + 1] = c$$

$$S_{gaccattct}(9) = 3$$

$$L_{gaccattct}(9) = 1$$

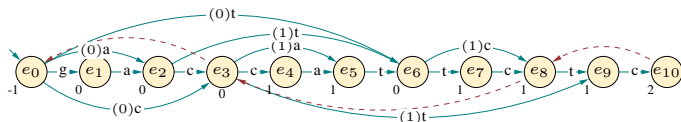
$$S_{gaccattct}(3) = 0$$

$$L_{gaccattct}(3) = 0$$

$$S_{gaccattct}(0) = -1$$

$$L_{gaccattct}(0) = -1$$

Oracle des Suffixes à transitions gardées du mot *gaccattctc*



$$\begin{aligned} S_{gaccattctc}(10) &= 8 & L_{gaccattctc}(10) &= 2 \\ S_{gaccattctc}(8) &= 3 & L_{gaccattctc}(8) &= 1 \\ S_{gaccattctc}(3) &= 0 & L_{gaccattctc}(3) &= 0 \end{aligned}$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

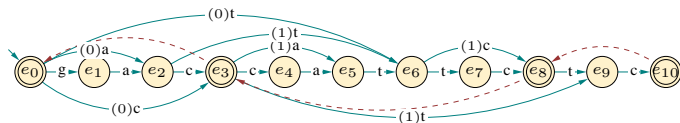
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes à transitions gardées du mot *gaccattctc*



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

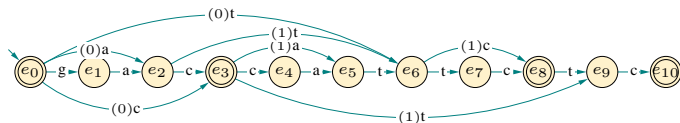
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes à transitions gardées du mot *gaccattctc*



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

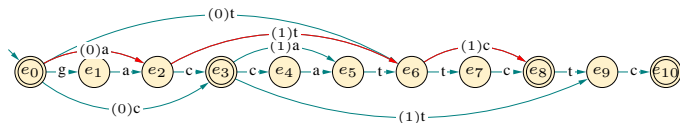
Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes à transitions gardées du mot *gaccattctc*

$atc \notin \text{Suff}(gaccattctc)$



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

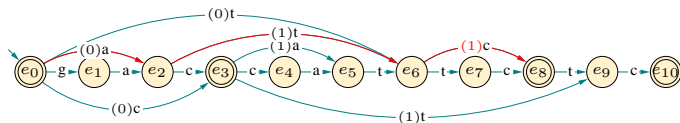
Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Suffixes à transitions gardées du mot *gaccattctc*

$atc \notin \text{Suff}(gaccattctc)$



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

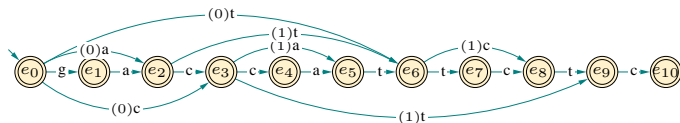
Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracle des Facteurs à transitions gardées du mot *gaccattctc*



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- Plusieurs paramètres :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

*Présentation et Construction des
Oracles*

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- Plusieurs paramètres :

- Plusieurs critères :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

*Présentation et Construction des
Oracles*

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- Plusieurs paramètres :
 - Taille de l'alphabet ($|\Sigma|$);
- Plusieurs critères :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

*Présentation et Construction des
Oracles*

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- Plusieurs paramètres :
 - Taille de l'alphabet ($|\Sigma|$);
 - Taille des séquences ($|s| = n$)...
- Plusieurs critères :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

*Présentation et Construction des
Oracles*

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- Plusieurs paramètres :
 - Taille de l'alphabet ($|\Sigma|$);
 - Taille des séquences ($|s| = n$)...
- Plusieurs critères :
 - Nombre de mots à reconnaître ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

*Présentation et Construction des
Oracles*

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- Plusieurs paramètres :
 - Taille de l'alphabet ($|\Sigma|$);
 - Taille des séquences ($|s| = n$)...
- Plusieurs critères :
 - Nombre de mots à reconnaître ;
 - Nombre de mots reconnus ;

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

*Présentation et Construction des
Oracles*

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- Plusieurs paramètres :

- Taille de l'alphabet ($|\Sigma|$);
- Taille des séquences ($|s| = n$)...

- Plusieurs critères :

- Nombre de mots à reconnaître ;
- Nombre de mots reconnus ;
- Nombre de transitions externes...

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

*Présentation et Construction des
Oracles*

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- Plusieurs paramètres :

- Taille de l'alphabet ($|\Sigma|$);
- Taille des séquences ($|s| = n$)...

- Plusieurs critères :

- Nombre de mots à reconnaître ;
- Nombre de mots reconnus ;
- Nombre de transitions externes...

Principe des tests statistiques :

	Mots à accepter	Mots à rejeter
Mots acceptés	Vrai Positifs (VP)	Faux Positifs (FP)
Mots rejetés	Faux Négatifs (FN)	Vrai Négatifs (VN)

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- Plusieurs paramètres :
 - Taille de l'alphabet ($|\Sigma|$);
 - Taille des séquences ($|s| = n$)...
- Plusieurs critères :
 - Nombre de mots à reconnaître ;
 - Nombre de mots reconnus ;
 - Nombre de transitions externes...

Principe des tests statistiques :

	Mots à accepter	Mots à rejeter
Mots acceptés	Vrai Positifs (VP)	Faux Positifs (FP)
Mots rejetés	Faux Négatifs (FN)	Vrai Négatifs (VN)

Indicateurs sur la fiabilité des tests statistiques :

$$\text{Sensibilité} = \frac{VP}{VP + FN}$$

$$\text{Spécificité} = \frac{VN}{VN + FP}$$

$$VPP = \frac{VP}{VP + FP}$$

$$VPN = \frac{VN}{VN + FN}$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- Plusieurs paramètres :
 - Taille de l'alphabet ($|\Sigma|$);
 - Taille des séquences ($|s| = n$)...
- Plusieurs critères :
 - Nombre de mots à reconnaître ;
 - Nombre de mots reconnus ;
 - Nombre de transitions externes...

Principe des tests statistiques :

	Mots à accepter	Mots à rejeter
Mots acceptés	Vrai Positifs (VP)	Faux Positifs (FP)
Mots rejetés	Faux Négatifs (FN)=0	Vrai Négatifs (VN)

Indicateurs sur la fiabilité des tests statistiques :

$$\text{Sensibilité} = \frac{VP}{VP + FN}$$

$$\text{Spécificité} = \frac{VN}{VN + FP}$$

$$VPP = \frac{VP}{VP + FP}$$

$$VPN = \frac{VN}{VN + FN}$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- Plusieurs paramètres :
 - Taille de l'alphabet ($|\Sigma|$);
 - Taille des séquences ($|s| = n$)...

- Plusieurs critères :
 - Nombre de mots à reconnaître ;
 - Nombre de mots reconnus ;
 - Nombre de transitions externes...

Principe des tests statistiques :

	Mots à accepter	Mots à rejeter
Mots acceptés	Vrai Positifs (VP)	Faux Positifs (FP)
Mots rejetés	Faux Négatifs (FN)=0	Vrai Négatifs (VN)

Indicateurs sur la fiabilité des tests statistiques :

$$\text{Sensibilité} = \frac{VP}{VP + FN} = 1$$

$$\text{Spécificité} = \frac{VN}{VN + FP}$$

$$VPP = \frac{VP}{VP + FP}$$

$$VPN = \frac{VN}{VN + FN} = 1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

● Plusieurs paramètres :

- Taille de l'alphabet ($|\Sigma|$);
- Taille des séquences ($|s| = n$)...

● Plusieurs critères :

- Nombre de mots à reconnaître ;
- Nombre de mots reconnus ;
- Nombre de transitions externes...

Principe des tests statistiques :

	Mots à accepter	Mots à rejeter
Mots acceptés	Vrai Positifs (VP)	Faux Positifs (FP)
Mots rejetés	Faux Négatifs (FN)=0	Vrai Négatifs (VN)

Indicateurs sur la fiabilité des tests statistiques :

$$\text{Sensibilité} = \frac{VP}{VP + FN} = 1$$

$$\text{Spécificité} = \frac{VN}{VN + FP}$$

$$VPP = \frac{VP}{VP + FP}$$

$$VPN = \frac{VN}{VN + FN} = 1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- Plusieurs paramètres :
 - Taille de l'alphabet ($|\Sigma|$);
 - Taille des séquences ($|s| = n$)...
- Plusieurs critères :
 - Nombre de mots à reconnaître ;
 - Nombre de mots reconnus ;
 - Nombre de transitions externes...

Principe des tests statistiques :

	Mots à accepter	Mots à rejeter
Mots acceptés	Vrai Positifs (VP)	Faux Positifs (FP)
Mots rejetés	Faux Négatifs (FN)=0	Vrai Négatifs (VN)

Indicateurs sur la fiabilité des tests statistiques :

$$\text{Sensibilité} = \frac{VP}{VP + FN} = 1$$

$$\text{Spécificité} = \frac{VN}{VN + FP} \approx 1$$

$$VPP = \frac{VP}{VP + FP}$$

$$VPN = \frac{VN}{VN + FN} = 1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

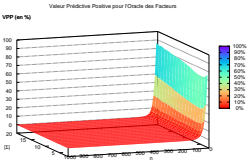
Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracles	Facteurs	Suffixes
Version Originale	<p>Valur Prédicive Positive pour l'Oracle des Facteurs</p>  <p>VPP (en %)</p> <p>100% 90% 80% 70% 60% 50% 40% 30% 20% 10% 0%</p>	
Transitions Gardées		

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

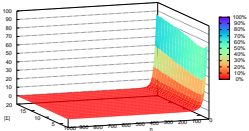
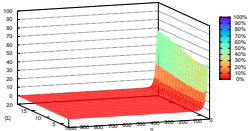
Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracles	Facteurs	Suffixes
Version Originale	<p>Valeur Prédicative Positive pour l'Oracle des Facteurs</p> 	<p>Valeur Prédicative Positive pour l'Oracle des Suffixes</p> 
Transitions Gardées		

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

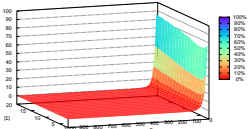
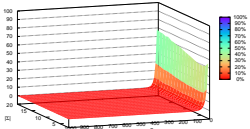
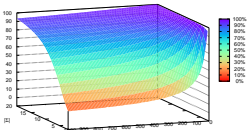
Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracles	Facteurs	Suffixes
Version Originale	<p>Valeur Prédicative Positive pour l'Oracle des Facteurs</p> 	<p>Valeur Prédicative Positive pour l'Oracle des Suffixes</p> 
Transitions Gardées	<p>Valeur Prédicative Positive pour l'Oracle des Facteurs à transitions gardées</p> 	

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

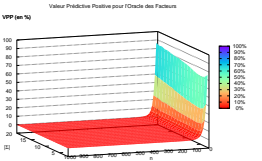
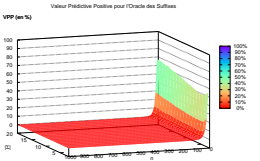
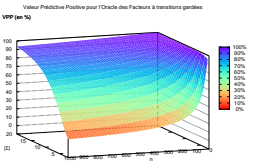
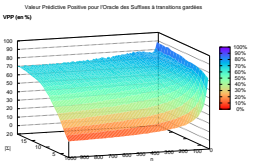
Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

Oracles	Facteurs	Suffixes
Version Originale	<p> Valeur Prédicte Positive pour l'Oracle des Facteurs</p> 	<p> Valeur Prédicte Positive pour l'Oracle des Suffixes</p> 
Transitions Gardées	<p> Valeur Prédicte Positive pour l'Oracle des Facteurs à transitions gardées</p> 	<p> Valeur Prédicte Positive pour l'Oracle des Suffixes à transitions gardées</p> 

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Présentation et Construction des Oracles

Langage reconnu par les Oracles

Vers un nouvel oracle

Conclusions & perspectives

- **Objectifs de la thèse :**

- fournir des méthodes informatiques permettant l'identification des sites de régulation de la synthèse protéique ;
⇒ répondre aux problèmes EM et EM_q.

- **Travail réalisé :**

- **Évolutions possibles :**

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

• Objectifs de la thèse :

- fournir des méthodes informatiques permettant l'identification des sites de régulation de la synthèse protéique ;
⇒ répondre aux problèmes EM et EM_q.

• Travail réalisé :

• Évolutions possibles :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

• Objectifs de la thèse :

- fournir des méthodes informatiques permettant l'identification des sites de régulation de la synthèse protéique ;
⇒ répondre aux problèmes EM et EM_q .

• Travail réalisé :

- deux algorithmes polynomiaux d'extraction ;
- formalisation et étude des fractions Block-Based ;
- implémentation et expérimentation ;
- Thèse d'habilitation en cours.

• Évolutions possibles :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

• Objectifs de la thèse :

- fournir des méthodes informatiques permettant l'identification des sites de régulation de la synthèse protéique ;
⇒ répondre aux problèmes EM et EM_q .

• Travail réalisé :

- deux algorithmes polynomiaux d'extraction ;
- formalisation et étude des fonctions *Block-Based* ;
- caractérisation du langage engendré par les Oracles ;
- Oracles à transitions gardées.

• Évolutions possibles :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

• Objectifs de la thèse :

- fournir des méthodes informatiques permettant l'identification des sites de régulation de la synthèse protéique ;
⇒ répondre aux problèmes EM et EM_q .

• Travail réalisé :

- deux algorithmes polynomiaux d'extraction ;
- formalisation et étude des fonctions *Block-Based* ;
- caractérisation du langage engendré par les Oracles ;
- Oracles à transitions gardées.

• Évolutions possibles :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

• Objectifs de la thèse :

- fournir des méthodes informatiques permettant l'identification des sites de régulation de la synthèse protéique ;
⇒ répondre aux problèmes EM et EM_q .

• Travail réalisé :

- deux algorithmes polynomiaux d'extraction ;
- formalisation et étude des fonctions *Block-Based* ;
- caractérisation du langage engendré par les Oracles ;
- Oracles à transitions gardées.

• Évolutions possibles :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

• Objectifs de la thèse :

- fournir des méthodes informatiques permettant l'identification des sites de régulation de la synthèse protéique ;
⇒ répondre aux problèmes EM et EM_q .

• Travail réalisé :

- deux algorithmes polynomiaux d'extraction ;
- formalisation et étude des fonctions *Block-Based* ;
- caractérisation du langage engendré par les Oracles ;
- Oracles à transitions gardées.

• Évolutions possibles :

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

● Objectifs de la thèse :

- fournir des méthodes informatiques permettant l'identification des sites de régulation de la synthèse protéique ;
⇒ répondre aux problèmes EM et EM_q.

● Travail réalisé :

- deux algorithmes polynomiaux d'extraction ;
- formalisation et étude des fonctions *Block-Based* ;
- caractérisation du langage engendré par les Oracles ;
- Oracles à transitions gardées.

● Évolutions possibles :

- adaptation de *SIMMS* et *StatSIMMS* à d'autres problèmes d'analyse de séquences ;
- utilisation des fonctions *Block-Based* afin de traiter les « sous-ensembles » ou le « sous-expression » du motif dans les problèmes de séquences.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

● Objectifs de la thèse :

- fournir des méthodes informatiques permettant l'identification des sites de régulation de la synthèse protéique ;
⇒ répondre aux problèmes EM et EM_q .

● Travail réalisé :

- deux algorithmes polynomiaux d'extraction ;
- formalisation et étude des fonctions *Block-Based* ;
- caractérisation du langage engendré par les Oracles ;
- Oracles à transitions gardées.

● Évolutions possibles :

- adaptation de *STARS* et *StatiSTARS* à d'autres problèmes d'analyse de séquences ;
- utilisation des fonctions *Block-Based* afin de traduire la « sur-expression » ou la « sous-expression » de motifs dans des ensembles de séquences ;
- intégration de l'Oracle à transitions gardées dans les méthodes existantes ;
- caractérisation du langage engendré par les Oracles à transitions gardées ;
- étude statistique du nombre de mots acceptés par les Oracles.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

● Objectifs de la thèse :

- fournir des méthodes informatiques permettant l'identification des sites de régulation de la synthèse protéique ;
⇒ répondre aux problèmes EM et EM_q .

● Travail réalisé :

- deux algorithmes polynomiaux d'extraction ;
- formalisation et étude des fonctions *Block-Based* ;
- caractérisation du langage engendré par les Oracles ;
- Oracles à transitions gardées.

● Évolutions possibles :

- adaptation de **STARS** et **StatiSTARS** à d'autres problèmes d'analyse de séquences ;
- utilisation des fonctions *Block-Based* afin de traduire la « sur-expression » ou la « sous-expression » de motifs dans des ensembles de séquences ;
- intégration de l'Oracle à transitions gardées dans les méthodes existantes ;
- caractérisation du langage engendré par les Oracles à transitions gardées ;
- étude statistique du nombre de mots acceptés par les Oracles.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

● Objectifs de la thèse :

- fournir des méthodes informatiques permettant l'identification des sites de régulation de la synthèse protéique ;
⇒ répondre aux problèmes EM et EM_q.

● Travail réalisé :

- deux algorithmes polynomiaux d'extraction ;
- formalisation et étude des fonctions *Block-Based* ;
- caractérisation du langage engendré par les Oracles ;
- Oracles à transitions gardées.

● Évolutions possibles :

- adaptation de **STARS** et **StatiSTARS** à d'autres problèmes d'analyse de séquences ;
- utilisation des fonctions *Block-Based* afin de traduire la « sur-expression » ou la « sous-expression » de motifs dans des ensembles de séquences ;
- intégration de l'Oracle à transitions gardées dans les méthodes existantes ;
- caractérisation du langage engendré par les Oracles à transitions gardées ;
- étude statistique du nombre de mots acceptés par les Oracles.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

● Objectifs de la thèse :

- fournir des méthodes informatiques permettant l'identification des sites de régulation de la synthèse protéique ;
⇒ répondre aux problèmes EM et EM_q.

● Travail réalisé :

- deux algorithmes polynomiaux d'extraction ;
- formalisation et étude des fonctions *Block-Based* ;
- caractérisation du langage engendré par les Oracles ;
- Oracles à transitions gardées.

● Évolutions possibles :

- adaptation de **STARS** et **StatiSTARS** à d'autres problèmes d'analyse de séquences ;
- utilisation des fonctions *Block-Based* afin de traduire la « sur-expression » ou la « sous-expression » de motifs dans des ensembles de séquences ;
- intégration de l'Oracle à transitions gardées dans les méthodes existantes ;
- caractérisation du langage engendré par les Oracles à transitions gardées ;
- étude statistique du nombre de mots acceptés par les Oracles.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

● Objectifs de la thèse :

- fournir des méthodes informatiques permettant l'identification des sites de régulation de la synthèse protéique ;
⇒ répondre aux problèmes EM et EM_q.

● Travail réalisé :

- deux algorithmes polynomiaux d'extraction ;
- formalisation et étude des fonctions *Block-Based* ;
- caractérisation du langage engendré par les Oracles ;
- Oracles à transitions gardées.

● Évolutions possibles :

- adaptation de **STARS** et **StatiSTARS** à d'autres problèmes d'analyse de séquences ;
- utilisation des fonctions *Block-Based* afin de traduire la « sur-expression » ou la « sous-expression » de motifs dans des ensembles de séquences ;
- intégration de l'Oracle à transitions gardées dans les méthodes existantes ;
- caractérisation du langage engendré par les Oracles à transitions gardées ;
- étude statistique du nombre de mots acceptés par les Oracles.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

● Objectifs de la thèse :

- fournir des méthodes informatiques permettant l'identification des sites de régulation de la synthèse protéique ;
⇒ répondre aux problèmes EM et EM_q.

● Travail réalisé :

- deux algorithmes polynomiaux d'extraction ;
- formalisation et étude des fonctions *Block-Based* ;
- caractérisation du langage engendré par les Oracles ;
- Oracles à transitions gardées.

● Évolutions possibles :

- adaptation de **STARS** et **StatiSTARS** à d'autres problèmes d'analyse de séquences ;
- utilisation des fonctions *Block-Based* afin de traduire la « sur-expression » ou la « sous-expression » de motifs dans des ensembles de séquences ;
- intégration de l'Oracle à transitions gardées dans les méthodes existantes ;
- caractérisation du langage engendré par les Oracles à transitions gardées ;
- étude statistique du nombre de mots acceptés par les Oracles.

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

That's All Folks!!!

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

Étude des Oracles

Conclusions & perspectives

Je vous remercie de votre attention...

Je vous remercie de votre attention...

...et espère que mes réponses à vos questions vous satisferont.

Séquences :

```
> S1, Lg=17. (e=1)
ACGAATCCGGTGCTTCA
> S2, Lg=17. (e=1)
TGTCAGGTTTCGCGACGA
> S3, Lg=17. (e=1)
ATACCGACCTCAGGTAT
> S4, Lg=17. (e=1)
AGTGGTATCTGGTGAGC
```

Ordre de traitement :

S2 - S4 - S3 - S1.

Fonction *Block-Based* :

$$\begin{cases} f^=(k) = k \\ f^{\neq}(k) = -k \end{cases}$$
$$\begin{aligned} p &= 3 \\ e &= 1 \end{aligned}$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Séquences :

```
> S1, Lg=17. (e=1)
ACGAATCCGGTGCTTCA
> S2, Lg=17. (e=1)
TGTCAGGTTTCGCGACGA
> S3, Lg=17. (e=1)
ATACCGACCTCAGGTAT
> S4, Lg=17. (e=1)
AGTGGTATCTGGTGAGC
```

[[1,17]]
● Score : 17

Ordre de traitement :

S2 - S4 - S3 - S1.

Fonction *Block-Based* :

$$\begin{cases} f^=(k) = k \\ f^{\neq}(k) = -k \end{cases}$$
$$p = 3$$
$$e = 1$$

Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Séquences :

```
> S1, Lg=17. (e=1)
ACGAATCCGGTGCTTCA
> S2, Lg=17. (e=1)
TGTCAGGTTTCGCGACGA
> S3, Lg=17. (e=1)
ATACCGACCTCAGGTAT
> S4, Lg=17. (e=1)
AGTGGTATCTGGTGAGC
```

Ordre de traitement :

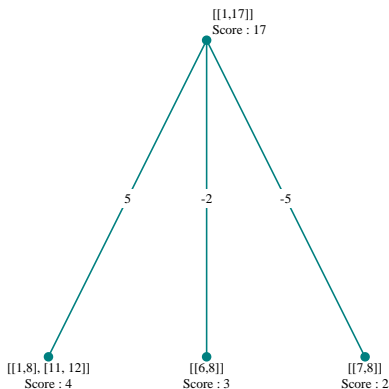
S2 - S4 - S3 - S1.

Fonction *Block-Based* :

$$\begin{cases} f^=(k) = k \\ f^{\neq}(k) = -k \end{cases}$$

$$p = 3$$

$$e = 1$$



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Séquences :

> S1, Lg=17. (e=1)
ACGAATCCGGTGCTTCA
> S2, Lg=17. (e=1)
TGTCAGGTTTCGCGACGA
> S3, Lg=17. (e=1)
ATACCGACCTCAGGTAT
> S4, Lg=17. (e=1)
AGTGGTATCTGGTGAGC

Ordre de traitement :

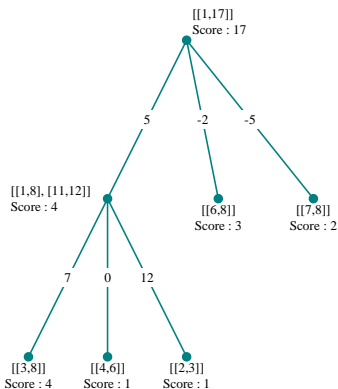
S2 - S4 - S3 - S1.

Fonction Block-Based :

$$\begin{cases} f^=(k) = k \\ f^{\neq}(k) = -k \end{cases}$$

$$p = 3$$

$$e = 1$$



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Séquences :

```
> S1, Lg=17. (e=1)
ACGAATCCGGTGCTTCA
> S2, Lg=17. (e=1)
TGTCAGGTTTCGCGACGA
> S3, Lg=17. (e=1)
ATACCGACCTCAGGTAT
> S4, Lg=17. (e=1)
AGTGGTATCTGGTGAGC
```

Ordre de traitement :

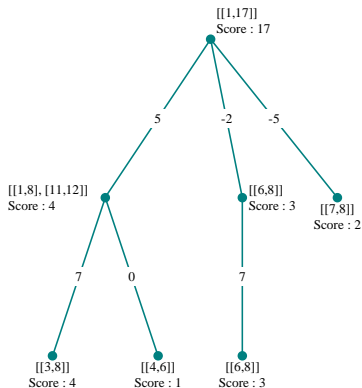
S2 - S4 - S3 - S1.

Fonction Block-Based :

$$\begin{cases} f^=(k) = k \\ f^{\neq}(k) = -k \end{cases}$$

$$p = 3$$

$$e = 1$$



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Séquences :

> S1, Lg=17. (e=1)
ACGAATCCGGTGCTTCA
> S2, Lg=17. (e=1)
TGTCAGGTTTCGCGACGA
> S3, Lg=17. (e=1)
ATACCGACCTCAGGTAT
> S4, Lg=17. (e=1)
AGTGGTATCTGGTGAGC

Ordre de traitement :

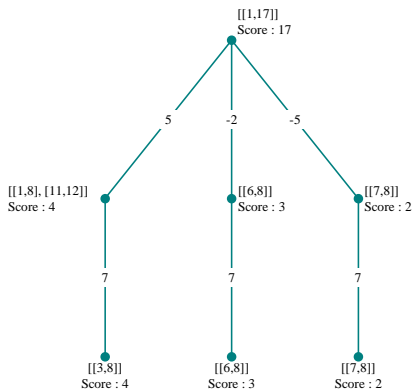
S2 - S4 - S3 - S1.

Fonction Block-Based :

$$\begin{cases} f^=(k) = k \\ f^{\neq}(k) = -k \end{cases}$$

$$p = 3$$

$$e = 1$$



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Séquences :

> S1, Lg=17. (e=1)
ACGAATCCGGTGCTTCA
> S2, Lg=17. (e=1)
TGTCAGGTTTCGCGACGA
> S3, Lg=17. (e=1)
ATACCGACCTCAGGTAT
> S4, Lg=17. (e=1)
AGTGGTATCTGGTGAGC

Ordre de traitement :

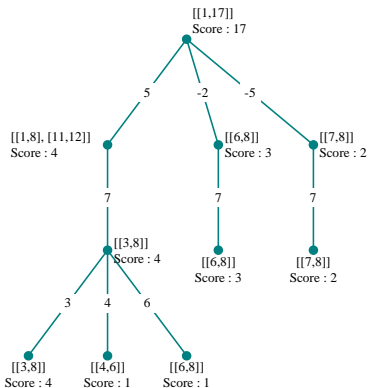
S2 - S4 - S3 - S1.

Fonction Block-Based :

$$\begin{cases} f^=(k) = k \\ f^{\neq}(k) = -k \end{cases}$$

$$p = 3$$

$$e = 1$$



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Séquences :

> S1, Lg=17. (e=1)
ACGAATCCGGTGCTTCA
> S2, Lg=17. (e=1)
TGTCAGGTTTCGCGACGA
> S3, Lg=17. (e=1)
ATACCGACCTCAGGTAT
> S4, Lg=17. (e=1)
AGTGGTATCTGGTGAGC

Ordre de traitement :

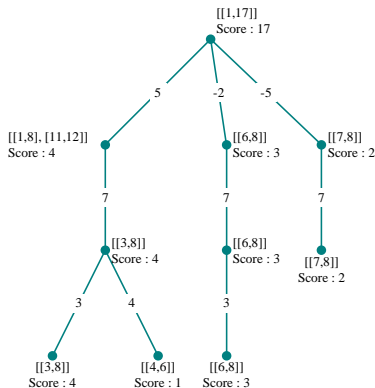
S2 - S4 - S3 - S1.

Fonction Block-Based :

$$\begin{cases} f^=(k) = k \\ f^{\neq}(k) = -k \end{cases}$$

$$p = 3$$

$$e = 1$$



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

Conclusions & perspectives

Séquences :

> S1, Lg=17. (e=1)
ACGAATCCGGTGCTTCA
> S2, Lg=17. (e=1)
TGTCAGGTTTCGCGACGA
> S3, Lg=17. (e=1)
ATACCGACCTCAGGTAT
> S4, Lg=17. (e=1)
AGTGGTATCTGGTGAGC

Ordre de traitement :

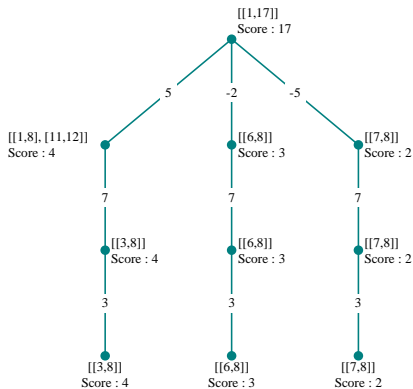
S2 - S4 - S3 - S1.

Fonction Block-Based :

$$\begin{cases} f^=(k) = k \\ f^{\neq}(k) = -k \end{cases}$$

$$p = 3$$

$$e = 1$$



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

STARS

Fonctions de score Block-Based

StatiSTARS

Étude des Oracles

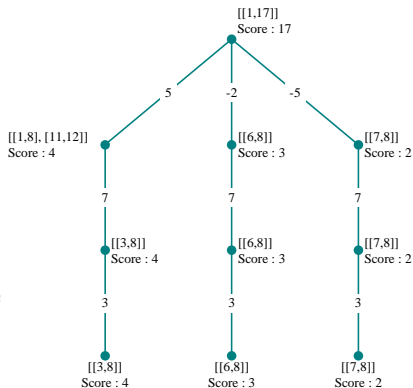
Conclusions & perspectives

Résultats :

Motif 1 : TCxGGT
 tgTCaGGTtcgcgacga
 agtggtaTCtGGTgagc
 ataccgaccTCaGGTat
 acgaaTCcGGTgcttca

Motif 2 : GGT
 tgtcaGGTtcgcgacga
 agtGGTatctggtgagc
 ataccgacctcaGGTat
 acgaaTccGGTgcttca

Motif 3 : GT
 tgtcagGTtcgcgacga
 aGTggtatctggtgagc
 ataccgacctcagGTat
 acgaaTccgGTgcttca



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

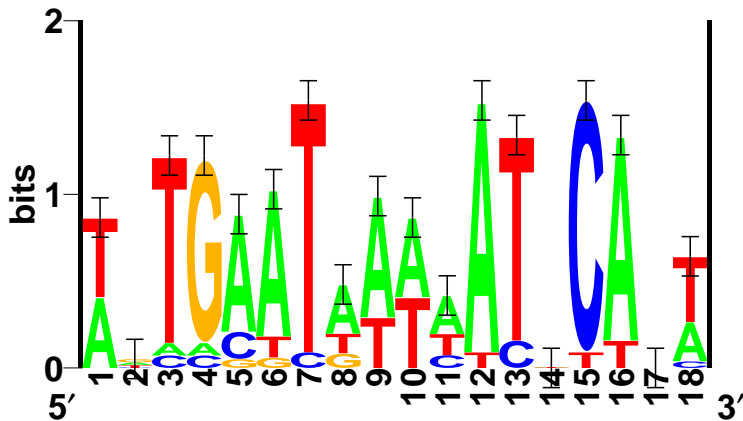
Stars

Fonctions de score Block-Based

StatiStars

Étude des Oracles

Conclusions & perspectives



Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

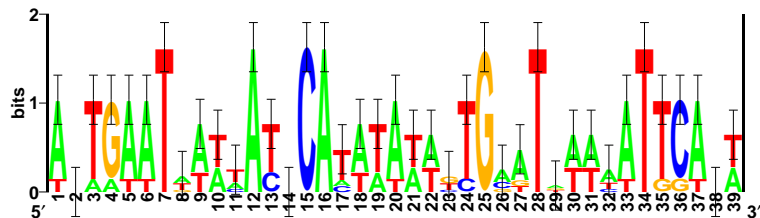
STARS

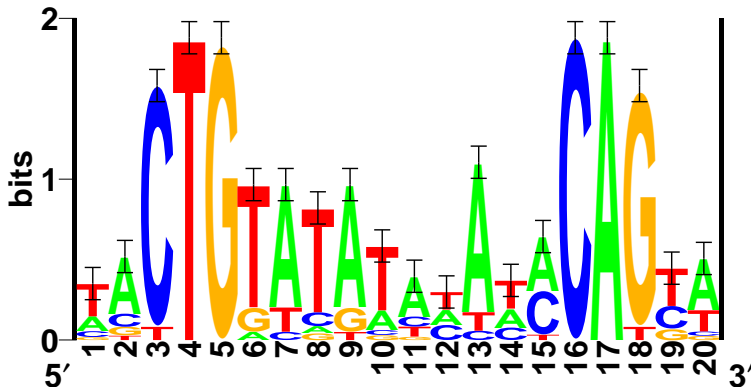
Fonctions de score Block-Based

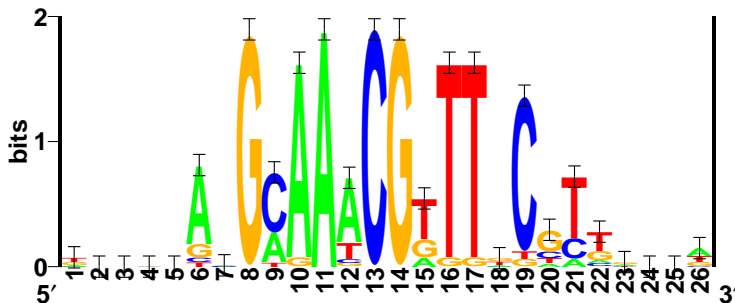
StatiSTARS

Étude des Oracles

Conclusions & perspectives







Soutenance de thèse

Alban MANCHERON

Titre & Sommaire

De la biologie à l'informatique

Extraction de motifs

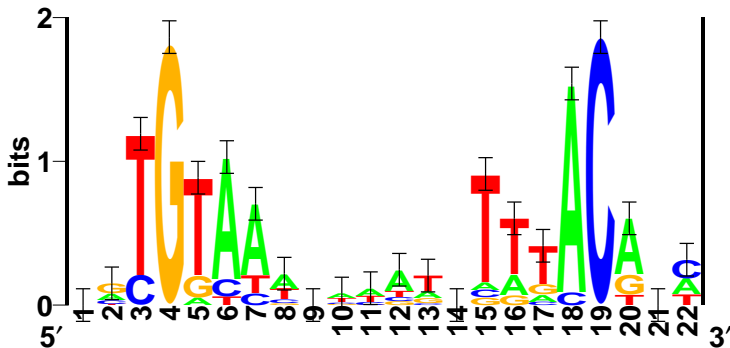
STARS

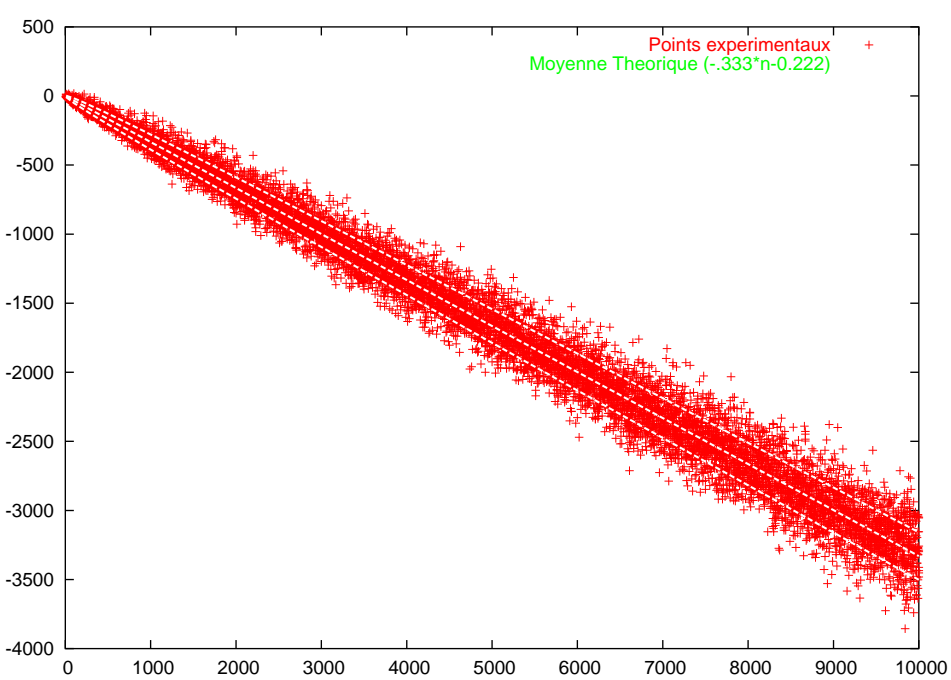
Fonctions de score Block-Based

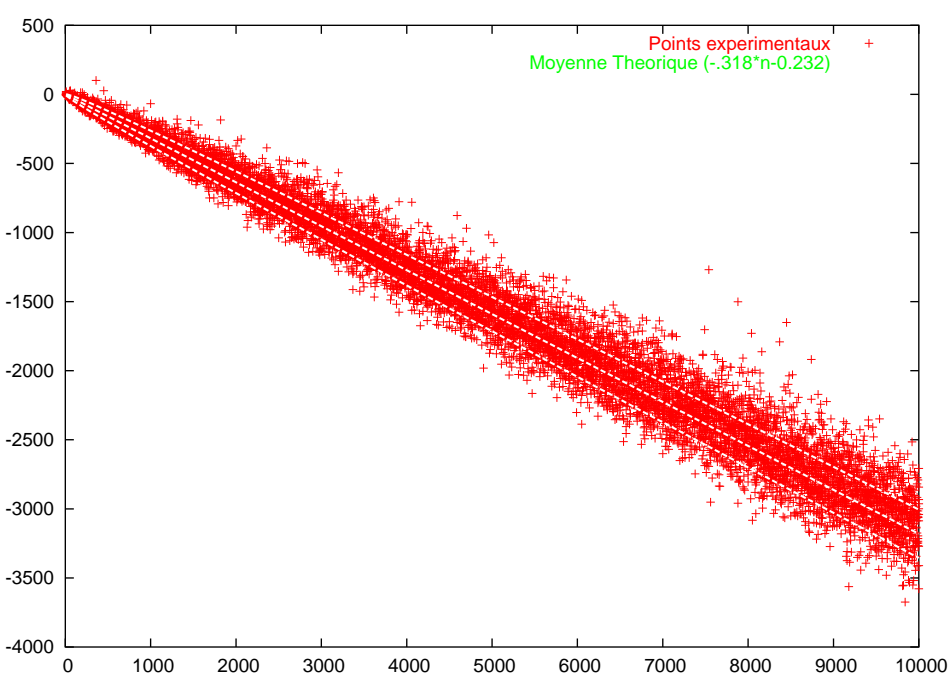
StatiSTARS

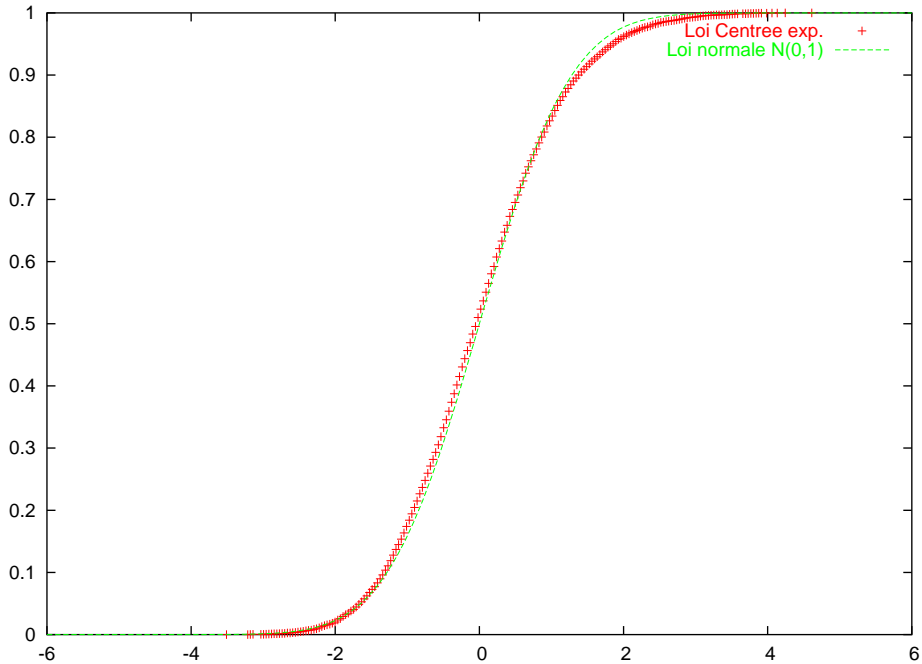
Étude des Oracles

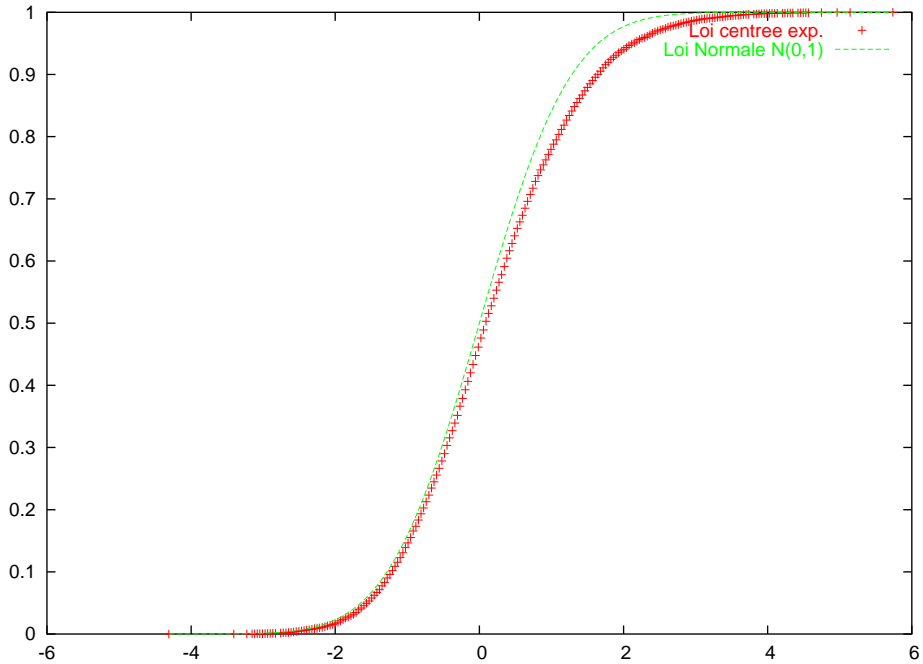
Conclusions & perspectives

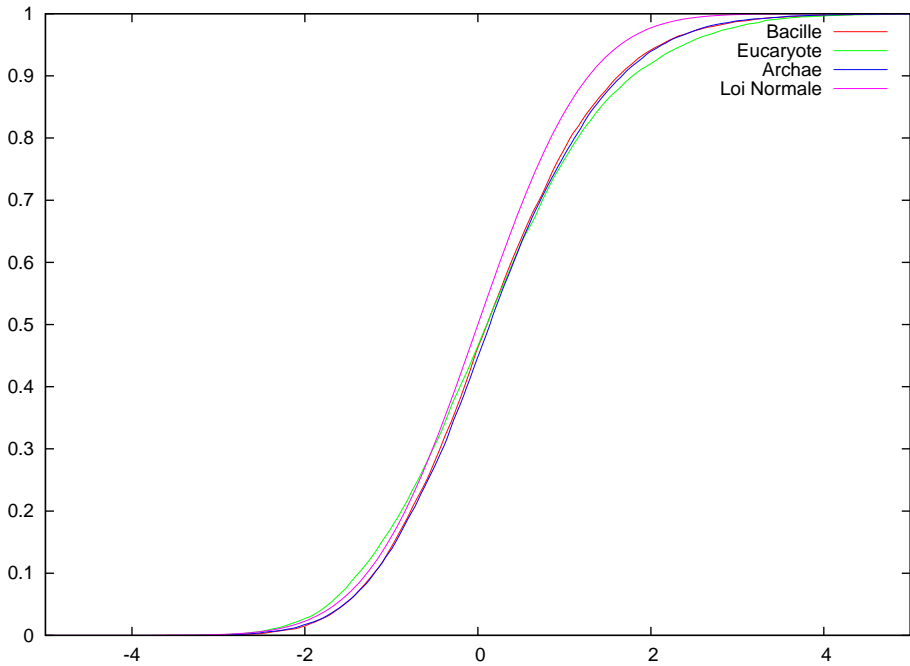






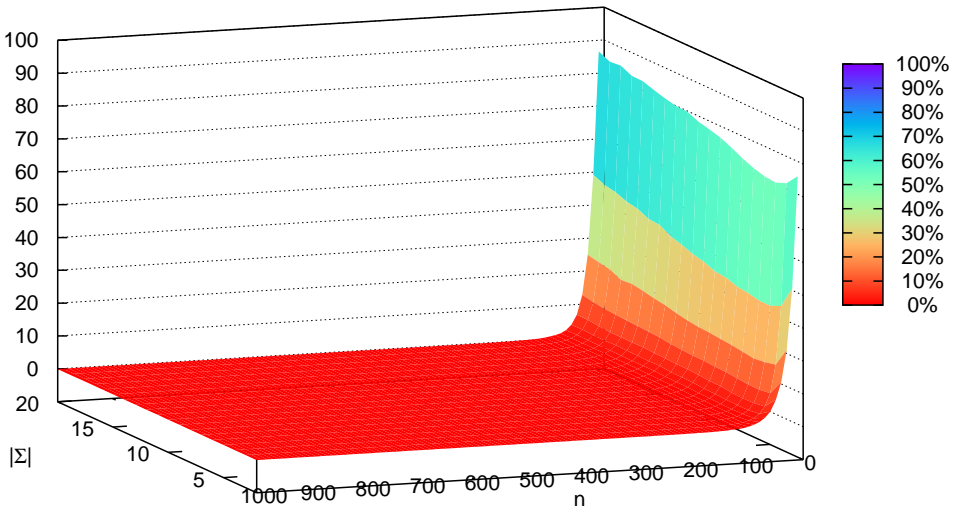






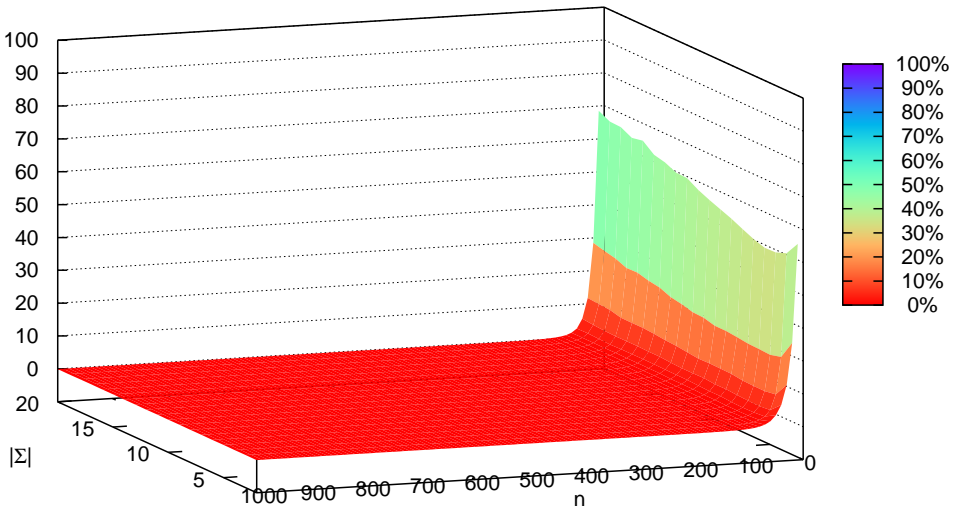
Valeur Prédicative Positive pour l'Oracle des Facteurs

VPP (en %)



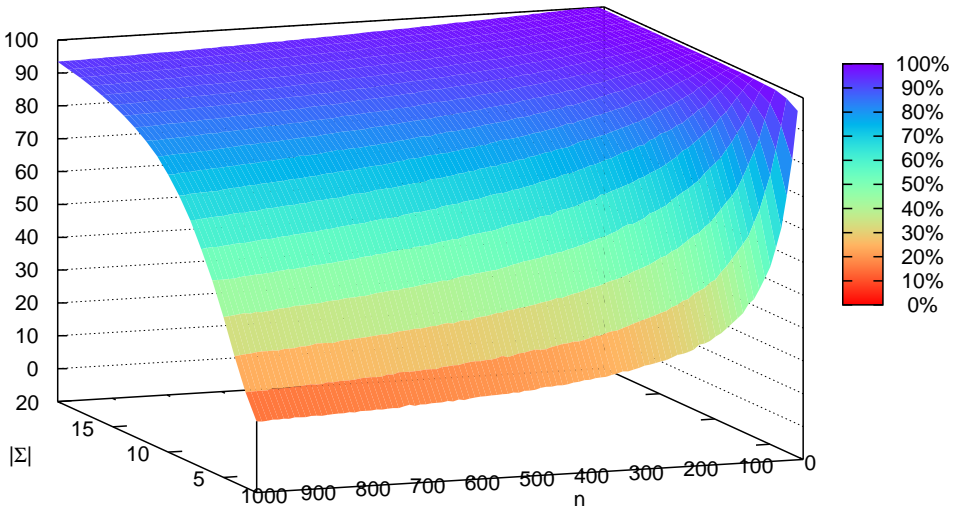
Valeur Prédicative Positive pour l'Oracle des Suffixes

VPP (en %)



Valeur Prédicative Positive pour l'Oracle des Facteurs à transitions gardées

VPP (en %)



Valeur Prédicative Positive pour l'Oracle des Suffixes à transitions gardées

VPP (en %)

