



MASTER 1 <sup>PJI</sup> INFORMATIQUE

UNIVERSITÉ LILLE 1

---

# Prédiction de l'activité des peptides

*CRISTAL - Equipe BONSAI*

---

*Auteur:*  
Emilie ALLART

*Tuteurs:*  
Maude PUPIN  
Laurent NOÉ

*9 mars 2015*

## Remerciements

Je remercie Maude et Laurent pour m'avoir permis d'effectuer mon projet dans l'équipe Bonsai. Merci de m'avoir suivie tout au long du développement de celui-ci, et de me laisser aller jusqu'au bout de mon projet en me donnant un stage.

# Contents

Remerciements . . . . .	1
Introduction . . . . .	3
<b>1 Contexte</b>	<b>4</b>
1.1 NRP . . . . .	4
1.1.1 Monomère . . . . .	5
1.1.2 Structure . . . . .	5
1.1.3 Activité . . . . .	6
1.1.4 Clusters . . . . .	6
1.2 Norine . . . . .	8
1.3 Problématique . . . . .	9
1.4 Plan . . . . .	9
<b>2 Mise en oeuvre</b>	<b>10</b>
2.1 Récupération depuis Norine . . . . .	10
2.2 Création des empreintes . . . . .	10
2.3 Lancement des méthodes d'apprentissage . . . . .	12
2.4 Mesures de robustesse . . . . .	12
<b>3 Résultats</b>	<b>14</b>
3.1 Résultats obtenus . . . . .	14
3.2 Comparaison avec les résultats précédents . . . . .	16
Conclusion . . . . .	16
Glossaire . . . . .	18
Annexe . . . . .	19

## Introduction

Au cours de la première année de master informatique, nous avons la possibilité de choisir dans le cadre du module PJI un projet à effectuer dans un laboratoire de recherche. J'ai donc effectué le mien au sein de l'équipe Bonsai, une équipe orientée bioinformatique faisant partie de CRISTAL.

Bonsai est un groupe de recherche en bioinformatique affilié avec INRIA Lille - Nord Europe et le Centre de Recherche en Informatique, Signal et Automate de Lille (CRISTAL, Université Lille 1, CNRS). Leur objectif principal est de définir des modèles et des algorithmes efficaces pour l'analyse de séquence à grande échelle dans le domaine de la biologie moléculaire. Cela comprend par exemple la génomique comparative et la métagénomique .

Une branche en particulier est orientée vers les peptides non ribosomiques (ou NRPs) dirigée par Maude Pupin.

C'est pourquoi, encadrée par Maude Pupin et Laurent Noé, j'ai travaillé sur le thème de : *Rechercher les meilleurs critères pour la prediction de l'activité d'un peptide*. Ceci rejoint une étude menée auparavant par l'équipe, intitulée *A new fingerprint to predict nonribosomal peptides activity*, qui étudie la décomposition d'un NRP en monomères (sous-ensemble) pour prédire son activité.

Nous allons donc étudier les peptides de la base Norine et d'essayer de trouver de nouveaux critères pour améliorer la prédiction de l'activité d'un peptide. De plus, une automatisation du programme permettra aux personnes le souhaitant d'utiliser leur base d'apprentissage.

Le but est de pouvoir par la suite prédire l'activité d'un peptide inconnu, avec pour information sa composition en monomères et sa structure. Et d'améliorer potentiellement les résultats par rapport à la précédente recherche.

Dans ce rapport, nous allons expliquer ce qu'est un NRP et présenter Norine, puis annoncer le plan du projet. Dans une deuxième partie, nous détaillerons étape par étape le travail effectué. Enfin nous présenterons les résultats obtenus en réalisant une comparaison avec l'étude précédente, et les réponses que l'on peut en tirer.

# Chapter 1

## Contexte

### 1.1 NRP

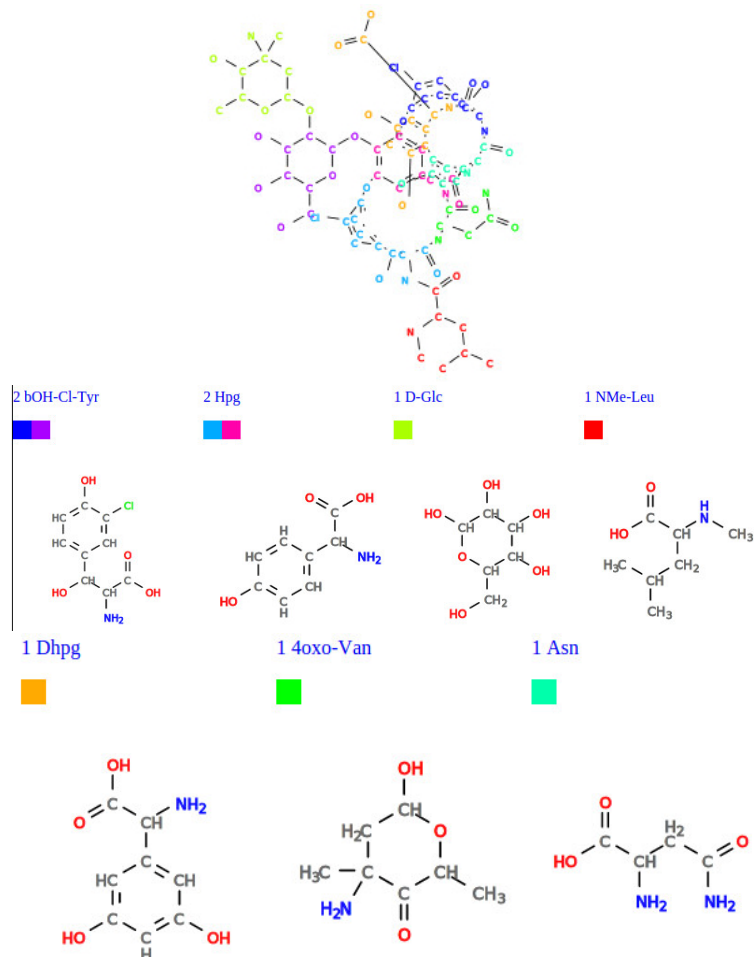


Figure 1.1: Décomposition de la vancomycin en monomères

... Dans la figure 1.1 page 4, on peut voir .... !!!

Les bactéries et les champignons comportent des peptides synthétisés par une voie indépendante du ribosome. Ils ne suivent pas la voie classique allant de la transcription de l'ADN, à la traduction de l'ARN en peptide sur le ribosome. Ils utilisent une voie alternative utilisant des NRPSs (ou non-ribosomal peptide synthetase) qui sont des complexes multi-enzymatiques organisés en module. Les NRPs sont des composés chimiques synthétisés par de grandes enzymes qui regroupent des acides aminés mais aussi leur(s) dérivés ainsi que d'autres composés (des lipides ou des glucides par exemple). Ces composés sont associés par des liaisons peptidiques et non peptidiques.

Les NRPs servent déjà à fournir des médicaments mais leur capacité à fournir de nouveaux médicaments et produits pharmacologiques n'est pas encore assez exploitée. Il faut donc continuer à en découvrir de nouveaux et en apprendre d'avantage sur leurs activités.

### 1.1.1 Monomère

Pour mieux comprendre les NRPs, une décomposition en monomères a été mise en place. Comme dit auparavant, un NRP est un assemblage d'acides aminés mais aussi de lipides et glucides, que nous appelons monomères (les unités de base) . Un NRP a la spécificité d'être de petite taille (2 à 50 monomères) et d'être fort diversifié car il existe plus de 500 monomères comptant les 21 acides aminés et tous leur(s) dérivés. Ainsi, on peut décrire un NRP par l'ensemble des monomères qui le composent et en tirer des informations.

Voici la décomposition monomérique de la vancomycine : Asn,bOH-Cl-Tyr,NMe-Leu,Leu,Hpg,D-Glc, Van,bOH-Cl-Tyr,Dhpg,Hpg

### 1.1.2 Structure

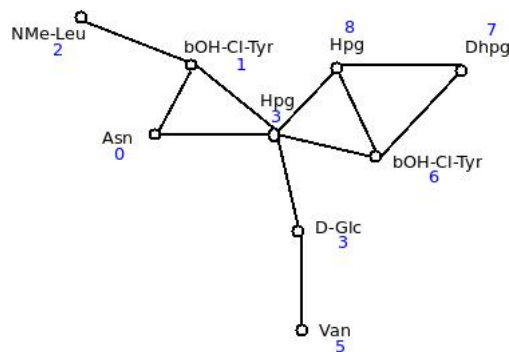


Figure 1.2: Représentation graphique de la vancomycine

0	1	2	3	4	5	6	7	8
Asn	bOH-Cl-Tyr	NMe-Leu	Hpg	D-Glc	Van	bOH-Cl-Tyr	Dhpg	Hpg
@1,3	@0,2,3	@1	@1,0,4,6,8	@3,5	@4	@3,7,8	@6,8	@3,7,6

Table 1.1: Représentation des liens entre les monomères composant la vancomycine

Il est plus aisé de représenter un NRP sous la forme d'un graphe avec pour noeuds les monomères qui le constituent, et pour liens les liaisons qui les relient. Le NRP ne contient pas seulement des liaisons peptidiques, mais également des liaisons non-peptidiques. Cela engendre donc l'apparition possible de structures cycliques (partielles ou non) et de ramifications sur la structure primaire de la molécule.

Grâce à la structure du NRP (graphe), on peut décrire les liens entre les monomères et analyser l'effet qu'ils ont sur l'activité du peptide. C'est pourquoi, dans le projet, l'arité de chaque monomère, c'est-à-dire le nombre de liaisons que possède un monomère, est étudié.

On peut décrire cette structure par une description linéaire qui liste les monomères dans un ordre, permettant d'identifier un monomère par sa position, et à la suite, à l'aide de symboles '@', se trouvent les liens qui existent entre eux

Voici la description linéaire de la vancomycine : Asn,bOH-Cl-Tyr,NMe-Leu,Hpg,D-Glc, Van,bOH-Cl-Tyr,Dhpg,Hpg @1,3 @0,2,3 @1 @1,0,4,6,8 @3,5 @4 @3,7,8 @6,8 @3,7,6

### 1.1.3 Activité

En effet, les NRPs sont une mine d'or pour les biologistes, ils ont un large domaine d'activité au niveau biologique et pharmacologique. Ils peuvent, par exemple, avoir comme activité:

- antibiotique : lutte contre les bactéries *ex* : *ACV* (précurseur de la *penicilline*)
- anticancéreux : lutte contre le cancer *ex* : *actinomycine D*
- toxine : tue les cellules *ex* : *callipeltin D*
- sidérophore : agit comme un aimant avec les molécules de fer *ex* : *amphibactin I*
- inhibiteur de la protéase : lutte contre les virus *ex* : *cyanostatin B*

Un NRP peut posséder plusieurs activités à la fois, mais nous ne considérons, dans ce travail, que ceux qui ne possèdent qu'une seule activité.

### 1.1.4 Clusters

Bien souvent les monomères partagent des propriétés physico-chimiques, soit parce qu'ils ont une structure similaire, soit parce qu'ils dérivent tous d'un même composant auquel s'est ajouté un groupement (groupement acetyl, methyl, etc...), ou qui a changé de conformité. Cela nous permet de les ranger dans des clusters (ou des familles).

Ainsi, nous examinons ces clusters et regardons s'ils peuvent nous aider à améliorer la prédiction de l'activité d'un peptide.

Voici une sous-partie des clusters que nous avons utilisée :

**Chromophores non siderophore**

ChrD;ChrA;ChrAct

**Siderophores**

ChrI;ChrP;OH-Asp;D-OH-Asp;OH-His;Ac-OH-Orn;D-Fo-OH-Orn;D-OH-Orn;D-Ac-OH-Orn;D-OH-cOrn;OH-Orn;NAc-Fo-OH-Orn;Fo-OH-Orn;OH-cOrn

**Peptaibols (antibiotiques)**

NAc-Dpr;Ac-Ser;Ac-Ival;Ac-Val;Ac-Trp;Ac-Phe;Ac-Aib;NAc-Leu;Ivalol;Leuol;Valol;Ileol;Pheol;Trpol;Se  
Leuol;Aib;Ival;4OH-Pro;Et-Nva

**Hpg et derives (antibiotiques)**

Hpg;D-OH-dHpg;Cl2-Hpg;NMe-Hpg;Cl-Hpg;OH-dHpg;D-Hpg

**Sucres (antibiotiques)**

2OMe-Rha;Ara;Aco;D-Gal;D-Ara;Ere;Glc;Oli;D-Glc;bD-Gal;U4oxo-Van;Rha;D-Man;4oxo-Van;Act;Lyx;Ria;Van



Bonsai  
Software

Norine

home general search structure search monomers help

## vancomycin

### Peptide

- Norine ID:** NOR00681
- Family:** vancomycin
- Synonym(s):** K258, vancin, vancled
- Activity:** antibiotic
- Class:** glycopeptide
- Formula:** C<sub>66</sub>H<sub>75</sub>ClN<sub>9</sub>O<sub>24</sub>
- Molecular weight:** 1449.2537 g/mol
- Comment:** Vancomycin is a glycopeptide antibiotic used in the prophylaxis and treatment of infections caused by Gram-positive bacteria. It has traditionally been reserved as a drug of "last resort", used only after treatment with other antibiotics had failed, although the emergence of vancomycin-resistant organisms means that it is increasingly being displaced from this role by linezolid and the carbapenems. Two bonds of the central Hpg are due to oxidative ring closure reactions.
- Entry information:**
  - status: curated
  - last modification date: 2007-02-12
  - Norine team (ProBioGEM (UPRES EA 1026 USTL), France, LIFL (UMR CNRS 8022 / Univ. Lille1), France)
  - [View all entry history](#)

### Structure

- Type:** other
- Number of monomers:** 9
- Monomeric composition :**

1	2	3	4	5	6	7	8	9
Asn	bOH-Cl-Tyr	NMe-Leu	Hpg	D-Glc	Van	bOH-Cl-Tyr	Dhpg	Hpg
- Graph representation:** Asn,bOH-Cl-Tyr,NMe-Leu,Hpg,D-Glc,Van,bOH-Cl-Tyr,Dhpg,Hpg @1,2 @0,2,2 @1 @1,0,4,6,8 @1,5 @4 @3,7,8 @6,8 @3,7,8
- Visualization**

### Organisms

- Nocardia orientalis**
- taxonomy:** cellular organisms, Bacteria, Actinobacteria, Actinobacteria (class), Actinobacteridae, Actinomycetales, Pseudonocardiaceae, Pseudonocardiaceae, Amycolatopsis
- gram:** positive
- synonyms:** Amycolatopsis orientalis, Streptomyces orientalis, Amycolatopsis orientalis orientalis
- taxid:** 91398

### References

- Vancomycin assembly: nature's way**  
Walsh CT, Hubbard BK, Angewandte Chemie, 2003, Feb 17,42(7):720-65.  
[PubMed: 12596194](#)

### Links

No comment available on this peptide [Add a comment](#)

8

(cyclique, linéaire, double cyclique, ...), qui donne la structure monomérique des peptides. On peut également trouver la description des monomères.

### 1.3 Problématique

Les NRPS, bien que fort intéressants, sont difficile à produire et à analyser. On ne peut donc pas utiliser les méthodes habituelle d'analyse de forme 3D ou de leur réaction physico-chimique. C'est pourquoi le but de l'étude précédente est de donner des modèles de prediction informatique permettant de prédire l'activité d'un peptide. Ces modèles se basaient uniquement sur l'empreinte monomérique du peptide, c'est-à-dire un comptage de chaque monomère pour voir lesquels sont présents et en déterminer d'après l'apprentissage quel 'schéma' ils suivent.

Pour notre part, nous ajouterons d'autres critères, à savoir l'occurrence des clusters de monomères et le comptage des liens. Puis, par le biais de méthodes d'apprentissage et de validation croisée, nous chercherons à voir lequel ou lesquels de ces critères prédisent au mieux l'activité du peptide concerné.

Dans l'idéal, nous aurons de meilleurs modèles pour prédire l'activité d'un peptide. Nous comparerons les différents résultats entre eux et avec ceux de la précédente étude, pour trouver les meilleurs critères.

### 1.4 Plan

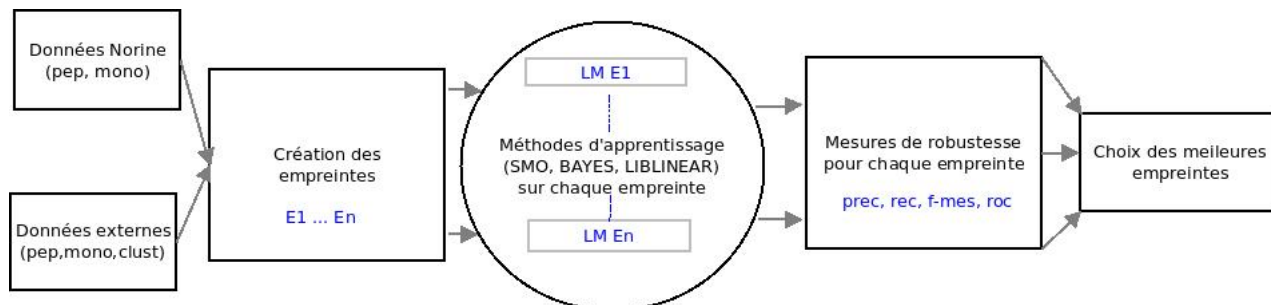


Figure 1.3: Schéma des étapes à suivre

## Chapter 2

# Mise en oeuvre

### 2.1 Récupération depuis Norine



Figure 2.1: Schéma de l'avancée

Norine met à disposition une passerelle REST, nous permettant de récupérer plus facilement les données. De cette façon, par l'intermédiaire de fichier JSON, nous avons récupéré les informations sur chaque peptide nous intéressant, ainsi que la liste de tous les monomères de la base.

Sur chaque peptide, si toutes les données sont fournies, nous conservons son identifiant (représenté par les lettres NOR suivies de 5 chiffres), sa composition monomérique, son activité et sa structure linéaire.

Pour rester cohérent avec l'étude précédente, nous avons enlevé une activité redondante nommée 'surfactant' et nous ne conservons que les peptides ayant une seule activité. Nous avons également créé un filtre qui ne garde que les peptides dont l'activité est assez représentative, le nombre de peptides possédant chacune de ces activités doit dépasser un certain seuil que nous fixons préalablement. Pour notre étude, nous avons fixé ce seuil à 20 peptides.

Une fois les données récupérées, nous les traitons selon les critères que nous souhaitons tester.

### 2.2 Création des empreintes

Une fois les données récupérées de Norine et filtrées, nous réalisons les empreintes.

#### **Empreinte en monomère**

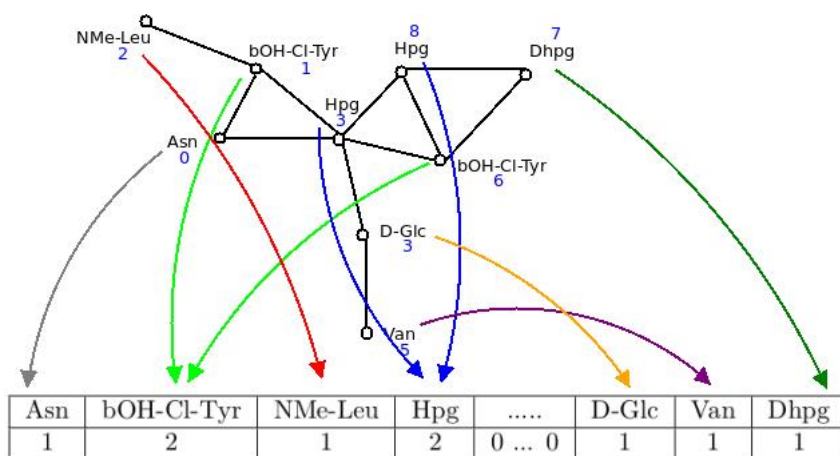
Au préalable nous avons prélevé l'ensemble des monomères de Norine. Mais



Figure 2.2: Schéma de l'avancée

il est possible pour une personne souhaitant utiliser ses données de donner au programme un fichier contenant les monomères qu'il a répertoriés.

A chaque peptide, nous comptons le nombre d'occurrences de chaque monomère dans sa composition et en faisons ainsi une empreinte.



Asn	bOH-Cl-Tyr	NMe-Leu	Hpg	.....	D-Glc	Van	Dhpg
1	2	1	2	0 ... 0	1	1	1

Table 2.1: Représentation des liens entre les monomères composant la vancomycin

### Empreinte en cluster

De même que pour les monomères, nous devons au préalable charger un fichier répertoriant les clusters de monomères avec leur nom et les monomères qui les composent.

C'est le même principe qu'auparavant, nous faisons l'empreinte mais cette fois nous notons le nombre d'occurrences d'un cluster en comptant combien de fois les éléments de celui-ci apparaissent dans ce peptide.

### Empreinte en lien

Comme dit précédemment, les NRPs sont représentés par des graphes de monomères. On peut donc intégrer la notion d'arité, qui représente le nombre de lien que possède un sommet, donc pour notre problème, cela représente le nombre de monomères auquel est lié un monomère. Pour chaque arité, allant de 1 à 5, nous comptons le nombre de sommets ayant cette arité, et l'ajoutons à notre empreinte.

Selon les critères que nous voulons tester, nous ajoutons à la liste pour chaque peptide leur(s) empreinte(s). Nous pouvons ainsi tester les critères, chacun pris séparément ou combinés : par exemple les liens et les monomères ensemble ou alors les clusters seuls.

## 2.3 Lancement des méthodes d'apprentissage

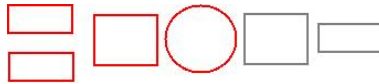


Figure 2.3: Schéma de l'avancée

Pour traiter les données, nous utilisons 3 méthodes d'apprentissage : le NaïveBayes, le SMO et le LibLinear.

Pour cela, nous avons utilisé des librairies python et une extension pour LibLinear qui n'est pas implémentée de base.

Le Naïve Bayes repose sur le théorème de Bayes, il possède une phase d'entraînement et une phase de test. Puis l'on fait de l'apprentissage probabiliste sur l'ensemble des peptides, ainsi lorsque l'on teste à quelle activité appartient un nouveau peptide, on conserve la classe qui a la plus grande probabilité.

Le SMO est une méthode d'apprentissage supervisé qui construit, au fur et à mesure de la lecture des données, une fonction objective via une descente de gradient. LibLinear reste dans la même idée.

**Validation croisée** Nous divisons les peptides en 10 groupes, nous faisons un apprentissage des empreintes sur 9 groupes et testons sur le dernier groupe si l'activité que l'on trouve correspond bien. Nous faisons cela pour chacun des groupes et obtenons ainsi des mesures de qualités.

## 2.4 Mesures de robustesse



Figure 2.4: Schéma de l'avancée

Nous obtenons ainsi un ensemble de mesures de robustesse qu'il faut comparer pour voir quel critère ou quel combinaison de critère est la plus efficace pour obtenir des informations sur un peptide. Pour notre étude, nous utilisons la précision, la sensibilité, la F-measure et le ROC.

**Precision**

La précision se calcule en divisant le nombre d'éléments bien classés par le nombre d'éléments en tout. Dans notre cas, elle se calcule en divisant le nombre de peptides bien classés par le nombre de peptides traités. Elle est la probabilité que les données soient bien triées, une mesure de la qualité de la classification. Donc un taux élevé est attendu dans la recherche.

$$précision = \frac{elements\text{correctement attribues à la classe } i}{nombre\text{d'elements attribues à la classe } i}$$

**Sensibilité**

La sensibilité se calcule en divisant le nombre d'éléments bien classés pour une classe donnée sur le nombre de données de cette classe. Cela représente le nombre de peptides bien annotés pour une activité par rapport au nombre de peptides rangés dans cette activité. Elle prend une valeur entre 0 et 1, plus elle tend vers 1 plus la classification est bonne.

$$sensibilite = \frac{elements\text{correctement attribues à la classe } i}{nombre\text{d'elements appartenant à la classe } i}$$

**F-measure**

La F-measure allie la précision et la sensibilité, ainsi si elle approche 1 c'est un bon résultat.

$$f - measure = \frac{2 \times precision \times sensibilite}{precision + sensibilite}$$

**AUC ou ROC**

La courbe ROC est la courbe représentant les faux-positifs (valeur rangée dans une classe alors qu'elle n'en fait pas partie) en fonction des vrai positifs (valeur bien rangée).

# Chapter 3

## Résultats



Figure 3.1: Schéma de l'avancée

### 3.1 Résultats obtenus

Faire ces tableaux pour les meilleurs résultats Puis comparatif avec graphes ?

[ht]

Activite	Naive Bayes				LibLinear				SMO			
	Prec	Rec	F	AUC	Prec	Rec	F	AUC	Prec	Rec	F	AUC
Antibiotique	0,971	0,737	0,838	0,961	0,950	0,962	0,956	0,953	0,947	0,953	0,950	0,942
Toxine	0,656	0,898	0,758	0,946	0,899	0,904	0,902	0,934	0,889	0,917	0,902	0,937
Siderophore	0,890	0,988	0,936	0,998	0,988	0,963	0,975	0,981	1	0,951	0,975	0,994
Anticancereux	0,471	0,640	0,542	0,935	0,696	0,640	0,667	0,814	0,696	0,640	0,667	0,868
Inhibiteur protease	0,870	0,909	0,889	0,996	0,952	0,909	0,930	0,954	0,952	0,909	0,930	0,975
Accuracy	81,49				93,22				92,89			

Table 3.1: Mesures de robustesse de l'apprentissage de l'étude précédente

Activite	Naive Bayes				LibLinear				SMO			
	Prec	Rec	F	AUC	Prec	Rec	F	AUC	Prec	Rec	F	AUC
Antibiotique	0,967	0,794	0,872	0,956	.	.	.	.	0,947	0,960	0,954	0,941
Toxine	0,742	0,803	0,771	0,958	.	.	.	.	0,834	0,925	0,877	0,955
Siderophore	0,719	0,967	0,825	0,995	.	.	.	.	0,967	0,967	0,967	0,990
Anticancereux	0,311	0,576	0,404	0,898	.	.	.	.	0,765	0,394	0,520	0,764
Inhibiteur protease	0,660	0,921	0,769	0,989	.	.	.	.	0,903	0,737	0,812	0,960
Accuracy	81,24				.				92,02			

Table 3.2: Mesures de robustesse de l'apprentissage sur l'empreinte monomérique

Activite	Naive Bayes				LibLinear				SMO			
	Prec	Rec	F	AUC	Prec	Rec	F	AUC	Prec	Rec	F	AUC
Antibiotique	0,970	0,794	0,873	0,959	.	.	.	.	0,953	0,965	0,959	0,949
Toxine	0,748	0,810	0,778	0,962	.	.	.	.	0,860	0,918	0,888	0,954
Siderophore	0,759	0,978	0,854	0,997	.	.	.	.	1,000	0,967	0,983	0,994
Anticancereux	0,373	0,576	0,452	0,912	.	.	.	.	0,739	0,515	0,607	0,836
Inhibiteur protease	0,507	0,921	0,654	0,994	.	.	.	.	0,914	0,842	0,877	0,979
Accuracy	81,50				.				93,16			

Table 3.3: Mesures de robustesse de l'apprentissage sur l'empreinte de monomères-liens

Activite	Naive Bayes				LibLinear				SMO			
	Prec	Rec	F	AUC	Prec	Rec	F	AUC	Prec	Rec	F	AUC
Antibiotique	0,976	0,759	0,854	0,957	.	.	.	.	0,953	0,965	0,959	0,947
Toxine	0,677	0,857	0,757	0,957	.	.	.	.	0,834	0,925	0,877	0,949
Siderophore	0,798	0,967	0,874	0,995	.	.	.	.	0,989	0,967	0,978	0,979
Anticancereux	0,313	0,606	0,412	0,896	.	.	.	.	0,750	0,455	0,566	0,773
Inhibiteur protease	0,625	0,921	0,745	0,988	.	.	.	.	0,903	0,737	0,812	0,964
Accuracy	80,23				.				92,52			

Table 3.4: Mesures de robustesse de l'apprentissage sur l'empreinte de monomères-clusters

Activite	Naive Bayes				LibLinear				SMO			
	Prec	Rec	F	AUC	Prec	Rec	F	AUC	Prec	Rec	F	AUC
Antibiotique	0,966	0,649	0,776	0,858	.	.	.	.	0,687	0,996	0,813	0,644
Toxine	0,417	0,769	0,541	0,831	.	.	.	.	0,400	0,014	0,026	0,584
Siderophore	0,889	0,978	0,931	0,998	.	.	.	.	1,000	0,967	0,983	0,996
Anticancereux	0,000	0,000	0,000	0,824	.	.	.	.	0,000	0,000	0,000	0,784
Inhibiteur protease	0,402	0,974	0,569	0,972	.	.	.	.	0,000	0,000	0,000	0,743
Accuracy	69,71				.				71,99			

Table 3.5: Mesures de robustesse de l'apprentissage sur l'empreinte de clusters-liens

Activite	Naive Bayes				LibLinear				SMO			
	Prec	Rec	F	AUC	Prec	Rec	F	AUC	Prec	Rec	F	AUC
Antibiotique	0,973	0,757	0,851	0,960	.	.	.	.	0,957	0,965	0,961	0,952
Toxine	0,723	0,816	0,767	0,962	.	.	.	.	0,854	0,918	0,885	0,956
Siderophore	0,784	0,967	0,866	0,997	.	.	.	.	1,000	0,967	0,983	0,996
Anticancereux	0,308	0,606	0,408	0,911	.	.	.	.	0,708	0,515	0,596	0,828
Inhibiteur protease	0,479	0,921	0,631	0,992	.	.	.	.	0,914	0,842	0,877	0,977
Accuracy	79,34				.				93,16			



Table 3.6: Mesures de robustesse de l'apprentissage sur l'empreinte de monomères-clusters-liens

Conclusion : quelles sont les meilleures combinaisons ...

## 3.2 Comparaison avec les résultats précédents

TODO tableau et/ou graphe comparatif

Conclusion : Avons-nous avancé depuis l'étude précédente? Quelles sont les meilleurs critère ou combinaison de critères pour déterminer l'activité d'un peptide ?

## Conclusion

Pour rappel, notre problématique était de chercher des meilleurs critères pour déterminer l'activité d'un peptide. A savoir que le programme est conçu pour que toutes personnes le souhaitant puisse faire l'apprentissage sur ses données.

TODO après résultats

En comparant, les résultats de l'étude effectuée sur les NRPs de Norine quelques années auparavant et les résultats que nous avons obtenus en combinant différentes connaissances sur ces NRPs, nous avons trouver des combinaisons de critères nous permettant d'apprendre avec plus de certitudes l'activité d'un NRP. Pour connaître l'activité d'un NRP avec une certitude de ... minimum, il faut combiner ....

Nous avons donc pu aller au bout de ce projet, en trouvant une combinaison meilleur, ne reste plus qu'à analyser de nouveau NRP pour, qui sait, trouver un nouvel antibiotique ...

## Glossaire

## Références

[?, ?, ?]