



# BBC – Projet- 2

Test pour détecter la  
présence de la bactérie E. coli  
dans l'eau

Bouyiatotis Stéphane - Marques Alexandre -  
Walter Simon



# BBC – Projet 2 : Test pour détecter la présence de la bactérie E. coli dans l'eau

---

## Introduction

---

Le but de ce projet est d'utiliser les outils bio-informatiques pour détecter la présence d'une bactérie, plus précisément le E. coli ou Escherichia coli, dans l'eau. La protéine cible nous est fournie. C'est l'OmpF Porin contenue dans la membrane de la bactérie E. Coli.

## Le E. Coli

---

Dans cette partie du document nous décrirons un peu plus le E. Coli. En connaissant mieux ce qu'est cette bactérie nous seront plus à même de la détecter.

### Qu'est-ce qu'une bactérie ?

Une bactérie est un organisme vivant microscopique et procaryote, le plus souvent unicellulaire ne comportant pas de noyaux et présent dans tous les milieux. Elles sont parfois pluricellulaires (généralement filamenteuses), la plupart des espèces bactériennes ne vivant pas individuellement en suspension, mais en communautés complexes adhérant à des surfaces au sein d'un gel muqueux.

Les bactéries les plus grosses mesurent plus de 2  $\mu\text{m}$  et, jusqu'au début du XXI<sup>e</sup> siècle, les spécialistes considéraient que les plus petites mesuraient 0,2  $\mu\text{m}$ , mais il existe des ultramicrobactéries.

Les bactéries sont ubiquitaires et sont présentes dans tous les types de biotopes rencontrés sur Terre. Elles peuvent être isolées du sol, des eaux douces, marines ou saumâtres, de l'air, des profondeurs océaniques, des déchets radioactifs, de la croûte terrestre, sur la peau et dans l'intestin des animaux ou des humains. Les bactéries ont une importance considérable dans les cycles biogéochimiques comme le cycle du carbone et la fixation de l'azote de l'atmosphère.<sup>1</sup>

### Qu'est-ce que le E. Coli ?

Escherichia coli, également appelée colibacille et abrégée en E. coli, est une bactérie intestinale des mammifères, très commune chez l'être humain. En effet, elle compose environ 80 % de notre flore intestinale. Découverte en 1885 par Theodor Escherich, dans des selles de chèvres, c'est un coliforme fécal. Cependant, certaines souches d'E. Coli peuvent être pathogènes, entraînant alors des gastro-entérites, infections urinaires, méningites ou sepsis.<sup>2</sup>

---

1 Source : <https://fr.wikipedia.org/wiki/Bact%C3%A9rie>

2 Source : [https://fr.wikipedia.org/wiki/Escherichia\\_coli](https://fr.wikipedia.org/wiki/Escherichia_coli)

## Pourquoi le rechercher ?

Nous savons que quelques formes du E. Coli peuvent provoquer des maladies et dans certain cas il peut être pathogène provoquant des problèmes intestinaux, mais aussi mortelles. De plus en 2011 une épidémie de E. Coli, plus précisément Escherichia coli O104:H4 du type Escherichia coli entérohémorragique (ECEH), avait touché en partie l'Europe, appelé aussi crise du concombre, car on croyait que ça venait de concombre provenant d'Espagne alors que cela provenait de graines germées de haricots. C'est le 18 juin 2011 que la souche est retrouvée dans un ruisseau près de Francfort et à proximité d'une station d'épuration.

Suite à ceci la recherche de présence de E. Coli dans l'eau permettra d'éviter que ce genre de cas se reproduisent.

## Manière de détecter la présence du E. Coli

Nous allons vous présenter deux façons de détecter la bactérie E. Coli dans l'eau :

### Méthode de type présence/absence (P/A)

Un volume de 100 ml d'échantillon d'eau est mélangé dans une bouteille stérile avec un milieu de culture comme le Colilert®, puis incubé pendant 24 h à 35 °C. Le milieu de culture contient de l'ONPG (ortho-nitrophényl-β-D-galactopyranoside) et du MUG (4-méthyl-umbelliféryl-β-D-glucuronide). Lorsque des coliformes totaux sont présents dans l'échantillon, l'ONPG est utilisé par l'enzyme β-D-galactosidase, une enzyme spécifique aux coliformes totaux. L'utilisation de l'ONPG provoque l'apparition d'une coloration jaune dans le milieu de culture. Lorsque E. coli est présent dans l'échantillon, le MUG est utilisé par l'enzyme β-D-glucuronidase, une enzyme spécifique à cette bactérie, ce qui induit une fluorescence bleue dans le milieu de culture, lorsqu'éclairé par un rayonnement ultraviolet. D'autres bactéries peuvent aussi utiliser l'ONPG ou le MUG, mais le milieu de culture spécifique contient habituellement des inhibiteurs qui empêchent leur croissance (CEAEQ, 2015a). Il est possible que des laboratoires accrédités utilisent des milieux différents, mais basés sur les mêmes principes.<sup>3</sup>

### Méthode par filtration sur membrane (FM)

Cette méthode permet de détecter et de quantifier simultanément la bactérie E. coli et les coliformes totaux sur une même gélose (CEAEQ, 2015b). Un volume d'eau (habituellement 100 ml) est filtré sur une membrane filtrante qui est déposée sur la gélose MI (à titre d'exemple). Cette gélose est ensuite incubée pendant 24 heures à 35 °C. Au terme de la période d'incubation, les bactéries E. coli retenues sur la membrane filtrante forment des colonies qui apparaissent bleues en lumière visible et les autres coliformes totaux forment des colonies fluorescentes sous un éclairage UV. La formation des couleurs qui permettent de distinguer spécifiquement les colonies résultent de l'utilisation de substrats enzymatiques apparentés à ceux employés dans la méthode de type présence/absence décrite plus haut (CEAEQ, 2015b). Certains laboratoires accrédités emploient d'autres milieux de culture gélosés qui sont basés sur les mêmes principes de fonctionnement que le milieu MI.<sup>4</sup>

---

3 Source : <https://www.inspq.qc.ca/eau-potable/e-coli>

4 Source : <https://www.inspq.qc.ca/eau-potable/e-coli>

## Le E. Coli : informations supplémentaires

Nous avons vue de manière générale qu'est-ce que le E. Coli ainsi qu'une partie de son histoire. Nous avons aussi vue qu'il pouvait prendre certaine forme potentiellement dangereuse pour l'homme, comme Escherichia coli O104:H4 ou bien Escherichia coli O157:H7. Dans cette partie nous allons rajouter des informations sur le E. Coli.

- C'est une bactérie de la famille des Enterobacteriaceae.
- Elle fermente le glucose
- Le patrimoine génétique de la souche E. coli de laboratoire non pathogène a été entièrement séquencé en 1997. Son génome comprend 4,6 millions de paires de bases codant environ 4 200 protéines
- Le E. Coli est connue pour avoir une faible stabilité, c'est à dire ça capacité à changer sans apport extérieur.
  - Ça stabilité est meilleur si elle est en groupe (3 ou plus) où le produit forme des corps d'inclusion insolubles
  - Corps d'inclusions insoluble : Quand un groupe de bactérie se réunie autour pur une surexpression d'une protéiné et s'accumule en un agrégat insoluble.
- Sensible aux bactériophages comme les phages T4 et lambda.

Lorsque l'on compare le E. Coli à ses compères pathogènes en moyenne seulement 40% de leurs gènes sont commun. Ce qui montre que le E. Coli possède un grand potentiel évolutif et de versatilité.

## Méthodologie

---

Dans cette partie nous allons expliquer la méthodologie que nous allons mettre en place pour la bonne résolution du projet.

Nous savons déjà quelle protéine est notre cible, c'est OmpF Porin. C'est une protéine situé sur la membrane extérieure du E. Coli et qui permet le passage passif de petite molécule aux travers. Avec cela nous allons voir qu'elles sont les différents épitopes qui réagissent à cette protéine, du moins à des parties de cette protéine.

### Protéine OmpF plus en détail :

- <https://www.ncbi.nlm.nih.gov/protein/1100855735>
- <https://www.uniprot.org/uniprot/P02931>

## Les outils de découvertes

Nous savons qu'il y a deux types d'épitopes, les T et les B, donc nous chercherons des outils pour les deux cas pour avoir un maximum de résultat.

Pour découvrir les épitopes qui réagissent à l'OmpF Porin nous allons utiliser divers outils qui nous sont fournis qui vont être listé ci-dessous. Comme nous avons un certain nombre d'outils nous ne les utiliserons pas tous.

### Prédiction d'épitopes linéaires (B cellules) <sup>5</sup>:

- **Bepipred** : Utilise un modèle de Markov caché et l'échelle d'hydrophilie de Parker.
- **Bepipred 2.0** : Utilise une forêt d'arbres décisionnels
- **Bcepred** <sup>6</sup>: Utilise une combinaison de propriété physico-chimique : hydrophilie, souplesse/mobilité, accessibilité, polarité, surface exposée, « coude » et antigénicité.
- **COBEpro** : Utilise une machine à vecteurs de support
- **Lbtope** <sup>7</sup>: Utilise une machine à vecteurs de support, des propriétés physico-chimiques et une échelle d'antigénicité sur des paires d'acides-aminés.
- **ABCpred (par peptides)** <sup>8</sup>: Utilise un Réseau de neurones récurrents

**Deux façons de faire** : par résidus (chaque résidu de protéines reçoit un score d'antigénicité) et par peptides (la prédiction est limitée à un nombre précis de protéines).

Profile de propriétés physico-chimiques des acides-aminés d'une protéine : <https://web.expasy.org/protscale/>

### Prédiction d'épitopes linéaires (T cellules) :

- MHC Class-I Binding Peptide Prediction, nHLAPred<sup>9</sup>
- MHC Class-II Binding Peptide Prediction, ProPred<sup>10</sup>

5 Choix méthodes de prédiction des B épitopes linéaires inspiré de : <https://www.biorxiv.org/content/10.1101/833418v1.full.pdf>  
 6 [https://webs.iitd.edu.in/raghava/bcepred/bcepred\\_submission.html](https://webs.iitd.edu.in/raghava/bcepred/bcepred_submission.html)  
 7 <https://webs.iitd.edu.in/raghava/lbtope/protein.php>  
 8 [https://webs.iitd.edu.in/raghava/abcpred/ABC\\_submission.html](https://webs.iitd.edu.in/raghava/abcpred/ABC_submission.html)  
 9 <http://crdd.osdd.net/raghava/nhlapred/>  
 10 <http://crdd.osdd.net/raghava/propred/>

## La sélection

Pour la sélection de nos épitopes les plus favorables pour détecter le E. Coli dans l'eau nous allons dans un premier temps chercher des épitopes avec chacun de nos outils, en ignorant volontairement les chaînes très petites. Nous sélectionnerons donc des grandes chaînes d'acides aminés possédant un bon score avec chacun de ses outils, puis nous les compareront pour voir les répétitions et ceux-ci seront sélectionnés formant notre liste d'épitopes potentiels. Pour tous les départager nous regarderons plus en détail cette chaîne, son score générale, et en rajoutant sa moyenne d'hydrophilie. Nous regarderons aussi si cette épitope ne peut pas être confondue avec d'autre corps étranger.

Pour notre recherche et notre classification nous nous contenterons d'utiliser des outils de prédiction d'épitope linéaire pour les B cellules car les épitopes B sont des parties d'antigène présents sous la forme entière de l'antigène et susceptibles d'être reconnus par un paratope de type anticorps <sup>11</sup>et c'est aussi là que nous avons le plus d'outils.

---

<sup>11</sup> <https://fr.wikipedia.org/wiki/%C3%89pitope>

## Réalisation

Dans cette partie nous allons décrire comment nous avons cherché et listé les épitopes

### Recherche des épitopes

Pour trouver les épitopes nous avons choisi 3 outils Bepipred<sup>12</sup>, Bepipred 2.0<sup>13</sup> et ABCpred<sup>14</sup>. Lors de nos différents tests ce sont les outils qui se sont avérés être les plus pratiques à utiliser. Nous avons ensuite trié les chaînes trop petites, car si elles sont inférieures à 10 elles ont plus de chance d'être trouvées sur d'autres protéines.

Chaque outil nous a donné un bon nombre de chaînes potentielles de toute taille. De ce fait nous avons dû les trier et nous sommes arrivés à un total de 39 chaînes restantes tous outils confondus. Après le tri des chaînes courtes, nous avons regardé les chaînes qui se retrouvent dans les différents outils pour faire notre liste d'épitopes potentiels.

Après cela, nous avons obtenu 6 épitopes présents dans nos 3 outils et un présent dans seulement deux outils (Bepipred et ABCpred). Nous avons aussi pris le score de chacune de ces chaînes sur chaque outil puis nous en avons fait une moyenne que nous utiliserons plus tard pour notre classification.

Chaîne	Début	Fin	Taille	Bepipred	Bepred 2	ABCpred	Moyenne	Prés.
<b>KGNGENSYGGNGDM</b>	47	60	14	0.85	0.6	0.8	0.75	3
<b>QGNNSEGADAQTGNK</b>	88	102	15	0.88	0.62	0.75	0.75	3
<b>RDTARRSNGDGV</b>	185	196	12	0.86	0.63	0.78	0.76	3
<b>TNLQEAQPLGNGKKA E</b>	219	234	16	0.78	0.64	0.75	0.72	3
<b>ETRNPITNKFTNT</b>	255	268	14	0.73	0.62	0.82	0.72	3
<b>DSDNKLGVGSDDT</b>	341	352	12	0.71	0.63	0.84	0.73	3
<b>AYTKSKAKDVEGI</b>	296	308	13	0.7	0.56	0.78	0.68	2

Sur le tableau ci-dessus se trouve un tableau de tous les épitopes sélectionnés avec leurs débuts dans la chaîne, leurs fins, leurs tailles, leurs scores suivant l'outil, calculés comme étant la moyenne des scores des acides aminés de la chaîne pour Bepipred 1 et 2 et comme la moyenne des scores des chaînes trouvées à ces positions dans ABCpred, leurs scores moyens, et leurs présences. Pour l'épitope présent dans que deux outils nous avons quand même réussi à obtenir le score sur Bepipred 2 mais nous l'avons marqué en jaune et pris cela en compte dans le calcul du score. Nous utiliserons le score moyen pour la classification.

12 <http://www.cbs.dtu.dk/services/BepiPred-1.0/>

13 <http://www.cbs.dtu.dk/services/BepiPred/instructions.php>

14 <http://crdd.osdd.net/raghava/abcpred/>

## Classification

Maintenant que nous avons notre liste d'épitopes nous allons pouvoir faire la classification. Pour cela nous allons rajouter le score d'hydrophilie de chacune de ces chaînes et leurs taux de conservation.

Pour trouver le score de l'hydrophilie de nos chaînes nous utilisons Bcepred<sup>15</sup>. Nous entrons la chaîne dans l'outil et il nous donne le score d'hydrophilie pour chaque acide aminé après nous faisons sa moyenne. Voir tableau ci-dessous :

Chaîne	Hydrophilie
<b>KGNGENSYGGNGDM</b>	2.022
<b>QGNNSEGADAQTGNK</b>	2.253
<b>RDTARRSNGDGVGGS</b>	1.757
<b>TNLQEAQPLGNGKKAE</b>	1.299
<b>ETRNPITNKFTNT</b>	1.140
<b>DSDNKLGVGSDDT</b>	1.804
<b>AYTKSKAKDVEGI</b>	1.367

Pour savoir ensuite si la chaîne peut être confondu avec un autre organisme nous réalisons un blast via blast ncbi<sup>16</sup> et nous calculons le pourcentage de chance que cette chaîne provienne bien du E. Coli

Chaîne	Pourcentage
<b>KGNGENSYGGNGDM</b>	99.274%
<b>QGNNSEGADAQTGNK</b>	99.270%
<b>RDTARRSNGDGVGGS</b>	99.272%
<b>TNLQEAQPLGNGKKAE</b>	97.645%
<b>ETRNPITNKFTNT</b>	99.027%
<b>DSDNKLGVGSDDT</b>	96.333%
<b>AYTKSKAKDVEGI</b>	98.775%

Maintenant que nous avons tous nos scores nous allons maintenant pouvoir créer notre liste d'épitope.

<sup>15</sup> [https://webs.iitd.edu.in/raghava/bcepred/bcepred\\_submission.html](https://webs.iitd.edu.in/raghava/bcepred/bcepred_submission.html)

<sup>16</sup> <https://blast.ncbi.nlm.nih.gov/Blast.cgi>



Avant le calcul du score final pour classer nos épitopes nous avons normalisé notre score d'hydrophilie pour que la valeur soit entre 0 et 1 comme pour les autres.

<b>Cla s</b>	<b>Chaîne</b>	<b>Pourcenta ge</b>	<b>Score hydro</b>	<b>score moyen</b>	<b>Score final</b>
<b>1</b>	QGNNSEGADAQTGNK	99.270%	0.876	0.750	2.618
<b>2</b>	KGNGENSYGGNGDM	99.274%	0.837	0.750	2.580
<b>3</b>	RDTARRSNGDGVGGSI	99.272%	0.793	0.757	2.542
<b>4</b>	DSDNKLGVGSDDT	96.333%	0.801	0.727	2.491
<b>5</b>	TNLQEAQPLGNGKKA	97.645%	0.717	0.723	2.416
<b>6</b>	ETRNPITNKFTNT	99.027%	0.690	0.723	2.404
<b>7</b>	AYTKSKAKDVEGI	98.775%	0.728	0.680	2.396

Nous avons donc maintenant notre liste d'épitope possible pour détecter le E. Coli dans l'eau classifié du meilleur au moins bon sur un score allant de 0 à 3

## Conclusion

Dans un premier temps le projet était difficile car nous ne comprenions pas tous ainsi que sur la méthodologie à adopter pour mener à bien le travail mais après discussion et recherche nous avons pu mieux comprendre le tenant et aboutissant de ce projet.

Cela nous a permis aussi de mieux comprendre comment la bio-informatique peut aider dans le développement et la recherche médicale. Nous a fournis une meilleure connaissance de certain outils et domaine de la biologie computationnelle.

## Lexique :

---

**Antigène** : toute substance capable d'induire une réponse immunitaire

**Épitopes T (wikipedia)** : « parties de protéine [...] présentés à la surface des cellules de l'organisme dans le but d'être reconnus par le système immunitaire. Ces polypeptides sont la plupart du temps des résidus de protéines dégradées par la cellule hôte, ils sont ensuite associés à des complexes protéiques appelés CMH (pour Complexe majeur d'histocompatibilité), qui en se fixant sur la membrane plasmique de la cellule hôte permettront la présentation de l'épitope aux cellules immunitaires. »

**Peptides (futura-science)** : « Molécule comprenant au moins deux résidus d'acides aminés reliés par la liaison peptidique [...] On parle de dipeptides, de tripeptides lorsque la chaîne ne comprend que deux ou trois résidus d'acides aminés ; le terme d'oligopeptide désigne une petite chaîne d'acides aminés, un polymère ne comprenant qu'une à deux ou trois dizaines de résidus. Les chaînes d'acides aminés plus longues sont appelées des polypeptides. Le terme de protéines est réservé à ces chaînes polypeptidiques, une protéine pouvant être constituée de plusieurs chaînes réunies par des ponts disulfures [...]. »

**Paratopes (wikipedia)** : « Le paratope d'un anticorps en est la partie qui assure la fonction de recoVous aurez 15 minutes de présentation et 10 minutes de discussion, soit 25 minutes au total. Merci de bien respecter ce timing. Naissance de l'antigène. Chaque paratope reconnaît de façon spécifique une partie de l'épitope d'un antigène. C'est le site de reconnaissance de l'anticorps. »

**Cellule b, lymphocyte B (wikipedia)** : « globules blancs particuliers faisant partie des lymphocytes. Ils sont responsables de l'immunité humorale et fabriquent les immunoglobulines appelées anticorps. Pour être actifs, leurs anticorps de membrane doivent se lier directement à un antigène [...] Les plasmocytes sont donc des lymphocytes B activés et capables de produire des anticorps dirigés contre l'antigène activateur. »

**Cellule t (wikipedia)** : « Les lymphocytes T, ou cellules T, sont une catégorie de leucocytes qui jouent un grand rôle dans la réponse immunitaire secondaire. [...] Ils sont responsables de l'immunité cellulaire : les cellules infectées par un virus par exemple, ou les cellules cancéreuses reconnues comme étrangères à l'organisme [...] sont détruites par un mécanisme complexe. »