

**E.N.S.S.I.B.**  
ECOLE NATIONALE SUPERIEURE  
DES SCIENCES DE L'INFORMATION  
ET DES BIBLIOTHEQUES

**UNIVERSITE  
CLAUDE BERNARD  
LYON I**

**D.E.S.S. en INFORMATIQUE DOCUMENTAIRE**

**Rapport de Stage**

**UTILISATION DES RESSOURCES D'INTERNET DANS LE  
DEVELOPPEMENT D'UNE INFRASTRUCTURE DE  
RECHERCHE MEDICALE**

BLADIER Agnès

Sous la direction de

Florin Grigorescu  
Chercheur INSERM  
Laboratoire d'Endocrinologie Moléculaire  
Montpellier

1996

1996  
ID ST  
3

**E.N.S.S.I.B.**  
**ECOLE NATIONALE SUPERIEURE**  
**DES SCIENCES DE L'INFORMATION**  
**ET DES BIBLIOTHEQUES**

**UNIVERSITE**  
**CLAUDE BERNARD**  
**LYON I**

**D.E.S.S. en INFORMATIQUE DOCUMENTAIRE**

**Rapport de Stage**

**UTILISATION DES RESSOURCES D'INTERNET DANS LE**  
**DEVELOPPEMENT D'UNE INFRASTRUCTURE DE**  
**RECHERCHE MEDICALE**

BLADIER Agnès



Sous la direction de

Florin Grigorescu  
Chercheur INSERM  
Laboratoire d'Endocrinologie Moléculaire  
Montpellier

1996  
ID ST  
3

1996

## **RESUME**

La recherche en génétique requiert des informations, nombreuses et variées. Afin d'améliorer la collecte de renseignements, souvent disponibles sur Internet, une liste de sites Web d'intérêt a été constituée et un manuel d'utilisateurs, destiné à faciliter l'utilisation de ceux-ci, réalisé. Puis, Afin de promouvoir les activités, actuelles ou futures, du laboratoire d'Endocrinologie Moléculaire, un site Web a été créé.

## **DESCRIPTEURS**

Internet

HTML

Manuel d'utilisateurs

Génétique

## **ABSTRACT**

Genetic researches require diverse and numerous informations. To improve the gathering of data, often available on Internet, a Web site list was constituted and a manuel, meant to facilitate their use, realised. Then, to promote the Molecular Endocrinology Laboratory's, actual or future, activities, a Web site was created.

## **KEYWORDS**

Internet

HTML

User's Manuel

Genetics

**Je remercie toutes les personnes grâce auxquelles ce stage et les réalisations décrites dans ce mémoires ont été possibles.**

## Query GenBank Database

Enter your query in the boxes below. A document matches a box if all words within it appear in one field of the document. The documents matching a box are combined with boxes following it using the selected operator to its right. The AND and BUT NOT operators limit retrieval and the OR operator expands it.

Checking the "fr" (Field Restriction) box limits the match to the fields checked in the table below.

Additional help is [here](#).

r	<input type="text"/>	<input type="radio"/> <i>and</i> <input type="radio"/> <i>or</i> <input type="radio"/> <i>but</i>
not	<input type="text"/>	<input type="radio"/> <i>and</i> <input type="radio"/> <i>or</i> <input type="radio"/> <i>but</i>
r	<input type="text"/>	<input type="radio"/> <i>and</i> <input type="radio"/> <i>or</i> <input type="radio"/> <i>but</i>
not	<input type="text"/>	<input type="radio"/> <i>and</i> <input type="radio"/> <i>or</i> <input type="radio"/> <i>but</i>

Return up to  documents.\*\*

\*\* Additional documents can be retrieved using the **(More)** button from the document list. \*\*

Select Database:  *GenBank*     *GenBank Updates*     *Both*

Field Restriction: (Applied only on query lines where "fr" is checked)

- Locus*     *Definition*     *Accession No.*     *NID*     *Keywords*
- Source*     *Reference*     *Comment*     *Features*

Rev. 06/27/96

# GENBANK

Genbank est produite par le NIH (National Institutes of Health ) basé à Bethesda (Maryland). Les informations proviennent de journaux mais aussi directement des soumissions d'auteurs .

Chaque entrée présente les informations suivantes :

- Description concise de la séquence
- Nom scientifique
- Taxonomie de la source
- Caractéristiques de la séquence
- Références bibliographiques (liens avec Medline).

Les bases dbEST et dbSTS sont des sous-produits de GENBANK (voir SRS p97)

- Entrer les mots clés
- Appliquer les opérateurs si nécessaire
- Limiter la recherche à certains champs (cliquer "fr" et sélectionner les champs)
- Sélectionner la base
- Décider du nombre maximum de réponses (1)
- Cliquer " Run Query"

NB: Voir "The GenBank DNA Sequence Database" en annexe

Source: BENSON D., BOGUSKI M., LIPMAN D.J., OSTELL J. (1994) GenBank. Nucleic Acid Research, 22 : 3441-3444.  
[http://www2.ncbi.nlm.nih.gov/genbank/query\\_form.html](http://www2.ncbi.nlm.nih.gov/genbank/query_form.html)

- 
- (1) Choix entre
- 1st 100
  - 2nd 100
  - 3rd 100
  - 4th 100
  - 5th 100
  - 6th 100
  - 7th 100
  - 8th 100
  - 9th 100
  - All

<b>INTRODUCTION GENERALE .....</b>	<b>6</b>
------------------------------------	----------

<b>PREMIERE PARTIE :LE LIEU ET SES MISSIONS .....</b>	<b>7</b>
---	----------

1. L'INSTITUT UNIVERSITAIRE DE RECHERCHE CLINIQUE (IURC) .....	8
--	---

2. LE LABORATOIRE D'ENDOCRINOLOGIE MOLECULAIRE .....	8
--	---

2.1. L'équipe .....	8
---------------------	---

2.2. La Recherche.....	8
------------------------	---

2.3. Le matériel .....	9
------------------------	---

3. LE STAGE .....	10
-------------------	----

<b>DEUXIEME PARTIE :UTILISATION D'INTERNET POUR LA RECHERCHE EN BIOLOGIE MOLECULAIRE.....</b>	<b>11</b>
---	-----------

INTRODUCTION.....	12
-------------------	----

1. PREMIERS ELEMENTS .....	12
----------------------------	----

1.1. Comment travaillent-ils ? .....	12
--------------------------------------	----

1.2. Quelles sont leurs connaissances ? .....	12
---	----

1.3. Quels sont les besoins des chercheurs ? .....	13
--	----

2. REALISATION D'UNE LISTE DE LIENS .....	13
---	----

2.1. Rechercher dans Internet.....	13
------------------------------------	----

1. Les catalogues .....	13
-------------------------	----

2. Les moteurs de recherche .....	14
-----------------------------------	----

2.2. Bases de données.....	14
----------------------------	----

1.Demandes expresses.....	14
---------------------------	----

2. Discussion avec un "expert" .....	14
--------------------------------------	----

3. Exploration à partir des moteurs de recherche.....	15
---	----

2.3. Centres de recherche.....	15
--------------------------------	----

2.4. Journaux .....	16
---------------------	----

2.5. Divers.....	17
------------------	----

2.6. Listes de discussion et groupes de news .....	17
--	----

1.Listes de discussion.....	17
-----------------------------	----

2. Groupes de news .....	17
--------------------------	----

2.7. Conclusion .....	18
-----------------------	----

3. REALISATION D'UN MANUEL D'UTILISATEURS .....	19
---	----

3.1. Que décrire ? .....	19
--------------------------	----

3.2 Comment présenter le manuel ? .....	19
---	----

3.3. Sommaire du manuel obtenu .....	20
--------------------------------------	----

3.4. Internet et son vocabulaire .....	21
--	----

3.5. Accès à Internet.....	22
----------------------------	----

3.6. Optimisation d'une interrogation grâce aux opérateurs booléens.....	24
--	----

3.7. Description des moteurs de recherche .....	25
---	----

3.8. Description des bases de données et des outils associés.....	28
---	----

1. Sites répertoriés .....	28
----------------------------	----

2. Présentation des bases de données.....	29
---	----

a. genbank.....	31
-----------------	----

b. pdb.....	31
-------------	----

3. Présentation des outils de recherche .....	36
---	----

4. RESULTATS.....	41
5. CONCLUSION.....	43
<b>TROISIEME PARTIE : CREATION D'UN SITE WEB SUR LE LABORATOIRE D'ENDOCRINOLOGIE MOLECULAIRE.....</b>	<b>44</b>
INTRODUCTION.....	45
1. INTRODUCTION A HTML.....	45
2. QUEL MATERIEL ?.....	47
2.1. <i>Quelle machine et quel logiciel utiliser ?</i> .....	47
2.2. <i>Netscape Gold 2.0.</i> .....	47
3. QUEL BUT ?.....	47
4. QUELLES INFORMATIONS ?.....	48
4.1. <i>Le laboratoire d'Endocrinologie Moléculaire</i> .....	48
4.2. <i>Les sites sélectionnés</i> .....	48
4.3. <i>Les données liées aux activités de recherche</i> .....	48
4.4. <i>Les activités de services du laboratoire</i> .....	48
4.5. <i>Les positions par rapport aux étudiants</i> .....	49
4.6. <i>La liste des publications</i> .....	49
5. QUEL RESULTAT ?.....	49
6. CONCLUSION.....	51
<b>CONCLUSION.....</b>	<b>52</b>
<b>BIBLIOGRAPHIE.....</b>	<b>53</b>
<b>ANNEXE.....</b>	<b>57</b>



# INTRODUCTION GENERALE

Ce rapport a été réalisé dans le cadre d'un stage au Laboratoire d'Endocrinologie Moléculaire de l'Institut Universitaire de Recherche Clinique, stage dont les réalisations ont eu pour finalité de préparer la phase de récolte des informations en vue de la création d'une base de données.

En effet, la croissance d'Internet et la multiplication des bases de données génétiques, rendent la localisation des données, spécifiques à un sujet d'étude, de plus en plus difficile. C'est la raison qui a motivé Dr Grigorescu à vouloir créer une base de données spécialisée sur les mutations du récepteur de l'insuline.

La première partie de ce mémoire est consacrée à la présentation de l'environnement dans lequel a été effectuée le stage.

La deuxième partie décrit la méthode suivie pour constituer une liste de sites Web utiles pour les chercheurs en génétique ainsi que la création d'un manuel facilitant l'utilisation de ces sites.

La création d'un site Web est le sujet de la troisième et dernière partie. Celui-ci devra permettre l'accès à la base de données, une fois celle-ci achevée; pour l'instant, seules les informations générales sur le laboratoire et quelques listes de données sont accessibles.

# **PREMIERE PARTIE :**

## **LE LIEU ET SES MISSIONS**

## **1. L'INSTITUT UNIVERSITAIRE DE RECHERCHE CLINIQUE (IURC)**

Cet Institut, ouvert en février 1996, est situé sur le campus de la Faculté de Médecine et sous l'autorité du doyen. Ayant pour vocation d'accueillir des équipes universitaires de recherche clinique reconnues par le Ministère, c'est une structure ouverte à l'ensemble de la communauté de recherche biomédicale, où sont développées des techniques de pointe en imagerie cellulaire et biologie moléculaire. Les équipes présentes à l'IURC sont les suivantes:

- Centre Régional d'Imagerie Cellulaire (CRIC)
- Laboratoire de Dermatologie Moléculaire
- Laboratoire de Force Atomique
- Laboratoire de Biologie et Biochimie des Lipides
- Laboratoire de Nutrition Humaine
- Biostatistique, Epidémiologie et Recherche Clinique
- Laboratoire d'Endocrinologie Moléculaire
- Rein et Hypertension

L'Institut est géré par un Conseil de gestion qui a pour attribution de coordonner et promouvoir les activités scientifiques de l'Institut (projets communs de développement, fonctions des postes communs d'activité, ...).

## **2. LE LABORATOIRE D'ENDOCRINOLOGIE MOLECULAIRE**

### **2.1. L'équipe**

L'équipe de l'IURC est associée au Service d'Endocrinologie de l'Hôpital Lapeyronie et forme le Groupe de Recherche sur la Pathologie Moléculaire de l'Insulino-résistance. Ce groupe est formé de neuf personnes, chercheurs et cliniciens dont la collaboration doit permettre d'associer les observations cliniques et les résultats de la recherche en génétique moléculaire recherche génétique.

### **2.2. La Recherche**

Le diabète sucré est une maladie très fréquente qui atteint plus de 2% de la population française et dont les complications vasculaires pèsent lourdement sur le coût de la santé publique. Pour ces deux raisons, la recherche en diabétologie est très développée et est plus récemment axée sur la biologie moléculaire et la génétique.

Le diabète est une maladie que tout le monde pense connaître : en effet, qui n'a pas lu ou vu une histoire dans laquelle il était question d'une personne devant s'injecter une dose d'insuline. Ce type de diabète est qualifié d'insulino-dépendant ou de type I et est dû à l'absence partielle ou totale de sécrétion d'insuline. Le diabète de type I est considéré comme une maladie auto-immune.

Toutefois, le plus fréquent des diabètes, est dit diabète non insulino-dépendant (DNID) ou de type II ; dans ce cas, le défaut ne vient pas de l'insuline, qui est normalement produite, mais de son incapacité à produire son effet biologique. Il y a insulino-résistance. Dans la résistance à l'insuline, le récepteur membranaire - c'est-à-dire la protéine qui permet à l'insuline de pénétrer dans la cellule - joue un rôle primordial dans la pathogénie..

Le Laboratoire d'Endocrinologie Moléculaire s'intéresse particulièrement au déterminisme génétique et à la pathologie moléculaire de la résistance à l'insuline. En effet, personne ne sait exactement à quoi est dû le diabète de type II et les chercheurs s'emploient à rechercher et identifier des mutations génétiques du récepteur de l'insuline et de la protéine IRS-1 (Insulin Receptor Substrate 1). Pour mener à bien une telle investigation, ils étudient l'ADN génomique de personnes atteintes de maladies rares dues à une insulino-résistance sévère (ex: leprechaunisme, maladie d'Alström). Ils espèrent ainsi découvrir des marqueurs génétiques pouvant ensuite être utilisés pour le dépistage des personnes à risque.

Il importe donc aux chercheurs d'être à même de trouver les articles de confrères diabétologues mais aussi de pouvoir avoir accès aux séquences d'ADN ou de protéines, déposées dans les banques de données. Ce qui, vue l'étendue des sujets d'intérêt, doit rapporter beaucoup d'informations, mais demander beaucoup de temps et exiger une assiduité certaine.

### **2.3. Le matériel**

Jusqu'à l'année dernière, l'équipe de recherche faisait partie du CRBM (Centre de Recherche en Biochimie Macromoléculaire) et les chercheurs avaient accès au centre de documentation ainsi qu'aux nombreux abonnements (revues, CD-Rom). A l'IURC, une telle structure n'existe pas encore et il faut donc parvenir à trouver l'information seuls.

Pour cela, il a été prévu d'utiliser au maximum les possibilités d'Internet grâce à la station Indigo- IRIS R3000 Silicon Graphics nommée Spectrum (système d'exploitation IRIX 4.05) datant de 1992.

Cette machine est équipée avec la version 4.2 de NCSA Mosaic. Pourquoi ne pas avoir changé de version ou être passé à Netscape ? Tout simplement parce qu'un changement, même de version, impliquait un changement du système d'exploitation, et que cela était impossible pour de nombreuses raisons que je n'exposerai pas.

### 3. LE STAGE

Ma présence devait permettre d'avancer la réalisation d'une base de données des mutations du récepteur de l'insuline, base réalisée en langage acedb développé par Jean Thierry-Mieg et Richard Durbin au CNRS. Or, après avoir commencé à réfléchir à la structure de la base de données, il s'est avéré que pour réaliser celle-ci, il faut avoir obtenu toutes les informations concernant toutes les mutations du récepteur de l'insuline découvertes jusqu'à ce jour.

Mon travail a, ainsi, surtout porté sur Internet et son utilisation pour un laboratoire public:

- utilisation centrée sur le laboratoire: sélection des sites d'intérêt (bases de données, journaux, ...) sur Internet
- utilisation tournée vers l'extérieur: création d'un site Web présentant le laboratoire et offrant des services.

Ce travail est donc un préliminaire à la création de la base de données IRMDB: Insulin Receptor Mutation Database.

## **DEUXIEME PARTIE :**

# **UTILISATION D'INTERNET POUR LA RECHERCHE EN BIOLOGIE MOLECULAIRE**

## INTRODUCTION

Avant de chercher à utiliser Internet pour la recherche dans le laboratoire, il faut :

- comprendre la manière actuelle de travailler des chercheurs afin d'améliorer leur recherche d'information
- savoir quelles sont leurs connaissances sur Internet et l'interrogation des bases de données
- en déduire quels sont les besoins des différents chercheurs

Une fois ces éléments en main, il est important de trouver une personne qui puisse fournir les rudiments sur les bases de données utiles en biologie moléculaire ou génétique.

## 1. PREMIERS ELEMENTS

### 1.1. Comment travaillent-ils ?

Le Laboratoire d'Endocrinologie Moléculaire reçoit les revues suivantes : Diabetes, Diabetologia et Biotechniques, mais ses quelques revues ne couvrent qu'une infime partie des articles publiés ayant trait aux sujets de recherche du laboratoire et la lecture d'autres journaux n'était pas systématique. De ce fait, une interrogation de Medline (sur CD-Rom), faite de façon très irrégulière, avec parfois des intervalles de 3-4 mois, permettait de savoir quels articles étaient primordiaux. Or, ceci avait lieu au CRBM, mais à l'IURC, plus de CD-Rom Medline !! Il faut donc faire autrement.

Il y a quelques années (1993), la banque de données protéiques de Brookhaven (PDB) avait été téléchargée sur la station pour faciliter l'utilisation du programme "Homology". Depuis les interrogations nécessaires se faisait sur cette version. Or, vu l'explosion de la recherche en biologie moléculaire et en génétique, la PDB est mise à jour tous les mois et un grand nombre de bases de données spécialisées se développent

### 1.2. Quelles sont leurs connaissances ?

« Internet, ça existe, les américains l'utilisent beaucoup et il faudrait s'y mettre. ». Voilà la réponse caricaturée que j'aurais pu obtenir. Plus sérieusement, ils en devinaient l'utilité sans savoir comment parvenir à trouver quoi que ce soit, surtout par manque de temps et de premiers conseils.

Pour l'interrogation des bases de données, certains pratiquaient l'interrogation simple et d'autres ne savaient pas trop comment faire.

Quant à l'utilisation de la messagerie pour communiquer rapidement et sans contrainte de décalage horaire, il semble que son utilité ait été méconnue par manque d'information.

### 1.3. Quels sont les besoins des chercheurs ?

Il leur faut se tenir au courant des découvertes de leurs confrères et donc savoir où et comment en être informés. Pour cela, il faut avoir accès aux différentes bases de données de séquences d'ADN, d'ARN et protéiques mais aussi aux sommaires de certains journaux.

Il pourrait aussi être intéressant de s'abonner à une ou des listes de discussion en biologie moléculaire ou en génétique.

**Pour réaliser tout cela, une liste de sites Web intéressants a été constituée.**

## 2. REALISATION D'UNE LISTE DE LIENS

Cette liste de liens a été constituée au fur et à mesure des explorations puis classée - ce qui fut relativement pénible avec Mosaic 4.2 car cette version ne permet pas l'organisation à posteriori des sites sélectionnés !!! Les différentes classes obtenues sont présentées ci-dessous. La liste complète, avec URLs, se trouve en annexe.

### 2.1. Rechercher dans Internet

Comment chercher dans Internet si l'on ne dispose pas de Netscape et de Netsearch ? Tout simplement en ayant retenu l'adresse d'Infoseek : <http://www.infoseek.com>. Une fois celle-ci mémorisée, il ne restait plus qu'à chercher.

Il existe en fait deux manières de chercher dans Internet, par les catalogues ou par les moteurs de recherche.

#### 1. Les catalogues

Ceux-ci ont été trouvés grâce à Infoseek et les mots-clés "internet catalogue" et à la liste donnée par IMS (Iron Mtn Systems): <http://www.ims.mariposa.ca.us/search/>

Toutefois, ces catalogues comme Galaxy (<http://www.einet.net/>) ou Whole Internet Catalogue (<http://www.gnn.com/wic/index.html>) n'ont pas été retenus, car leur méthode de recherche passe avant tout par la sélection de catégories et, comme le sujet concerne à la fois la médecine et la biologie moléculaire, quelle catégorie choisir ?

Seul Yahoo a été sélectionné après la lecture d'un article de Winship (1995)<sup>1</sup> qui comparait les performances de plusieurs outils de recherche dont Yahoo, Webcrawler et Lycos. Il en ressort que même si un outil donne plus de résultats, il n'est peut-être pas meilleur pour autant. En effet, Webcrawler et Lycos donnent sur un sujet précis un nombre plus important de réponses par rapport à Yahoo (cf. tableau ci-dessous) mais celles-ci sont moins pertinentes et parfois redondantes. En effet, les sous divisions d'un même document sont considérées comme des sites à part entière.

---

<sup>1</sup> WINSHIP I.R. (1995) World Wide Web searching tools - an evaluation. VINE, 99:375-376.



**Tableau 1: Réponse de différents outils de recherche dans Internet**

	Lycos	Webcrawler	Yahoo
ebola	295	124	7
Alberta	42	42	0
Chirac	27	7	0

d'après Winship (1995)

## 2. Les moteurs de recherche

Outre Infoseek qui a permis les premières recherches (avec les mots "search engine"), les logiciels les plus connus ont été sélectionnés à partir de la même liste de l'IMS. Il s'agit de :

- Altavista: <http://www.altavista.digital.com/>
- Lycos: <http://lycos.cs.cmu.edu/>
- Webcrawler: <http://webcrawler.com>

Toutefois, deux autres listes sont à retenir:

- Links to search engines: <http://www.netour.com/intern~1.htm>
- The big search engine list: <http://www.ez2.net/search.htm>

## 2.2. Bases de données

### 1. Demandes expresses

Les seules bases de données dont il avait été question étaient Medline et la PDB (Protein Data Bank). Ces bases furent donc les deux premières de la liste, car contrairement à ce que je pensais, Medline offre la partie de sa base concernant la génétique gratuitement sur Internet (sans les Medlars toutefois). Cela ne saurait durer, certes, mais tant que c'est le cas, il faut en profiter.

### 2. Discussion avec un "expert"

Un chercheur du CRBM, Jean Derancourt, spécialiste de la chimie des protéines, a accepté de me recevoir pour me parler des bases de données qu'il utilise pour ses recherches. Il s'agit d'EMBL, GenBank, PIR et Swiss-Prot.

Une autre source de renseignements a été Sandrine Hugues, qui travaillait à recenser les bases de données sur Internet comportant des séquences génétiques d'un batracien: le Xénope. Ses renseignements, bien qu'obtenus tardivement, m'ont permis de compléter ma liste et de vérifier la pertinence de celle-ci pour un travail comme le sien..

### 3. Exploration à partir des moteurs de recherche

Les moteurs de recherche ont été interrogés avec les mots-clés suivants:

- "gene database"
- "genetic database"
- "molecular biology database"

puis les réponses obtenues "épluchées".

Les outils les plus intéressants à consulter sont les listes de sites déjà sélectionnés pour un thème donné. Les deux listes retenues sont :

- DNA Sequence and Search Ressource:  
<http://www.unl.edu/stc-95/Restools/Bitools/bitools3a.html>
- Pedro's Biomolecular Research Tools:  
[http://www.public.castate.edu/~pedro/research\\_tools.html](http://www.public.castate.edu/~pedro/research_tools.html)

Une première liste a donc été constituée sur la base des titres et sur des impressions, c'est-à-dire que les bases que l'on peut atteindre à partir de la plupart des sites des centres de recherche doivent être les plus utiles. Cette liste a ensuite été soumise aux chercheurs afin qu'ils sélectionnent les bases paraissant pouvoir réellement répondre à leurs attentes. Certaines, comme la base MPDB (Molecular Probe Database), ont été éliminées: je n'avais pas encore bien cerné le sujet. La liste restante a été améliorée en fonction des réponses obtenues avec les mots-clés "insulin, insulin receptor, ou insr" dont la liste et/ou un exemple complet étaient généralement présentés aux chercheurs accompagnés de la question Est-ce intéressant?

Deux systèmes permettant l'accès, simultané ou non, à un grand nombre de bases de données (parfois très spécialisées) ont été découverts. Il s'agit de SRS - Sequence Retrieval System et d'Entrez.

Le lien avec une des banques de données, EMBL<sup>2</sup>, a été récemment supprimé de la liste. Le site qui permet d'y accéder était saturé quelle que soit l'heure, j'ai donc considéré qu'il valait mieux passer par SRS, plus rapide.

### 2.3. Centres de recherche

Les sites des centres de recherche ont un intérêt pour les liens avec les bases de données ou les outils de manipulation des données qu'ils comportent ainsi que pour les explications diverses fournies. Ceux ci ont été sélectionnés sans réel critère, en fonction de leur renommée et des productions de bases de données.

INFOBIOGEN: <http://www.infobiogen.fr/>

CRBM - Centre de Recherche en Biochimie Macromoléculaire:

<http://www.crbm.cnrs-mop.fr/home.html>

EMBL - European Molecular Biology Laboratory: <http://www.embl-heidelberg.de/>

NCBI - National Center for Biotechnology Information: <http://www.ncbi.nlm.nih.gov/>

NIH - National Institutes of Health: <http://www.nih.gov/>

---

<sup>2</sup> EMBL : [gopher://biox.embnet.unibas.ch:13021/77/.index/xembl/index](http://gopher://biox.embnet.unibas.ch:13021/77/.index/xembl/index)

National Institute of Genetics (Japan): <http://www.nig.ac.jp/>

International Nucleotide Sequence Database Collaboration: <http://www.ncbi.nlm.nih.gov/collab/>

## 2.4. Journaux

La recherche de journaux en science passe avant tout par l'étude des catalogues d'éditeurs disponibles sur Internet tels que:

- Elsevier: <http://elsevier.nl/>
- Springer-Verlag: <http://www.springer.de/server/listjournals.html>

Les moteurs de recherche ont aussi permis de découvrir, par les mots-clés "electronic journals", la liste Medweb: <http://www.cc.emory.edu/WHSCCL/medweb.ejs.html>.

De la même manière que pour les bases de données, l'aide d'une professionnelle a été utile. Anne Dujol, conservateur de l'UPM (Unité Pédagogique Médicale), m'a conseillé le serveur du CHU de Rouen. J'y ai découvert une liste, spécialisée en médecine, recensant des journaux de faible diffusion: <http://www.chu-rouen.fr/docmed/textel.html>.

Ces différentes listes ont été plus ou moins transformées (pour celle de Springer-Verlag, par exemple, les titres sont suivis des numéros effectivement accessibles, seule la liste des titres a été conservée), puis présentées aux différents chercheurs qui ont souligné ou coché les titres les intéressant.

Une fois les différentes listes soulignées, cochées, il a fallu:

- faire attention aux redondances entre les différentes listes
- récupérer les sommaires des journaux. Les listes sont constituées avec des liens hypertexte, certes, mais lorsque le lien ne répond pas, il faut essayer de trouver quand même le journal. Pour cela, deux méthodes, remonter la hiérarchie de l'URL, ce qui est souvent efficace, ou même utiliser les moteurs de recherche pour retrouver le journal. Malgré tout, certains journaux n'ont pu être retrouvés (Applied Genetic News et Protein Data Bank Newsletter notamment) !

- conserver uniquement les journaux offrant leur sommaire, et parfois des résumés ou des articles en texte intégral. Ceux nécessitant un abonnement ont été immédiatement évincés, sauf si l'abonnement était gratuit et nécessitait seulement de remplir un formulaire de renseignements.

La liste complète obtenue comporte 131 journaux, mais les listes sont personnalisées pour chaque chercheur, chacun ayant des intérêts plus ciblés (ex: la microbiologie ou la pédiatrie). Chaque chercheur doit toutefois avoir une liste, d'une soixantaine de journaux, qu'il a fallu constituer séparément ! Je ne sais si tous ces sommaires seront suivis, mais je ne pouvais, par moi-même, faire le tri !!! Comme il est très simple de supprimer un lien, j'ai considéré que ce n'était qu'à l'usage que cette liste pourrait être réduite.

## 2.5. Divers

Cette partie "Divers" a été constituée par des réponses à des questions précises et spontanées, exprimées lorsque quelqu'un passait avec une préoccupation à l'esprit et demandait « Tu crois que tu trouverais quelque chose sur..... » ou « J'ai trouvé cette adresse, est ce qu'on peut aller voir ? ».

## 2.6. Listes de discussion et groupes de news

Ces deux types de méthodes de communication diffèrent essentiellement par le fait qu'un abonné à une liste de discussion reçoit les messages directement dans sa boîte aux lettres alors que les news sont rassemblés quelque part sur le réseau, administrés et consultables à l'envie.

### 1. Listes de discussion

Il existe des serveurs de listes de discussion, j'ai trouvé quatre adresses dans L'Internet professionnel. Bien que les adresses aient un peu changé depuis l'année dernière, j'ai retrouvé ces listes en remontant la hiérarchie de l'URL:

- <http://www.tile.net:2001/tile/listserv/>: ce serveur recense les listes par description, par nom, par sujet ou les groupe par pays et organisme sponsor. J'y ai trouvé une liste sur la biologie moléculaire Molbio-l (Molecular Biology Discussion) située à l'Université du Missouri. Mais comme je l'ai découverte en juillet, elle ne fonctionnait plus et je n'ai pu l'évaluer !
- <http://www.neosoft.com/internet/paml/>: ce serveur regroupe les listes par nom ou par sujet.
- <http://www.univ-rennes1.fr/LISTES/>: les listes sont des listes francophones classées par sujets ou alphabétiquement.
- <http://www.ai.sri.com/mailling-lists/>: ce serveur regroupe les listes par sujet mais aucune ne m'a semblé intéressante et la recherche est plutôt pénible.

### 2. Groupes de news

J'avais trouvé la liste des groupes "bionet" \* mais comme le laboratoire ne possède pas de logiciel capable de lire ceux-ci.....!

D'autre part, il existe une très longue liste de newsgroups dans "The Internet complete reference"

\*Bionet newsgroups: <http://www.w3.org/pub/DataSources/News/groups/bionet.html>

## 2.7. Conclusion

La constitution de cette liste a été très longue et relativement fastidieuse, plusieurs raisons peuvent être invoquées:

1. les chercheurs n'avaient pas toujours le temps de parcourir les listes et le retour de la liste traînait.

2. la réalisation de listes personnifiées a duré fort longtemps. Pour les bases de données, une seule liste a été constituée puis copiée dans les autres répertoires, mais pour les journaux, c'était moins facile. En effet, si certains journaux sont en commun, il faut savoir lesquels (et parfois, un journal sélectionné sur une liste ne l'aura pas été dans une autre !!) et de toute façon, comme les listes avaient déjà été en partie personnifiées, pour les dernières listes parcourues la sélection a été faite pour chaque liste séparément.

3. la constitution d'une liste sur cette version de Mosaic se fait par "Add Current" lorsque l'on est sur le site voulu. Il faut donc aller sur le site et ensuite l'enregistrer, ce qui est, selon l'heure plutôt long !

S'il fallait estimer la durée de réalisation avec un matériel moins vieux, il faudrait une à deux semaines, avec Mosaic, comptez une semaine de plus !!!!

Avec une telle liste de sites Web, il est possible d'avoir accès à une très grande quantité d'information. Mais encore faut-il savoir y accéder !

En effet, il faut savoir avant tout se servir du logiciel d'accès à Internet, en l'occurrence Mosaic, et parvenir à la liste. Une fois cette liste atteinte, si arriver aux sommaires des journaux ne demande rien d'autre que de cliquer sur leur titre, il n'en est pas de même avec les moteurs de recherche ou les bases de données. Effectivement, face à un formulaire d'interrogation, comment faire lorsque l'on est novice ? Chercher dans l'aide, si toutefois elle existe et est facilement compréhensible. Or cela demande beaucoup de temps, ce dont ne disposent pas les chercheurs.

**Il m'a donc paru nécessaire de réaliser un manuel à leur intention, leur décrivant les procédures pour interroger les différents sites.**

### **3 .REALISATION D'UN MANUEL D'UTILISATEURS**

Dans ce chapitre, certaines pages du manuel seront présentées telles quelles. Afin de les reconnaître, la police des caractères sera différente de celle utilisée jusqu'à présent. Ce sera celle ci: police atlanta.

#### **3.1. Que décrire ?**

Avant toute chose, il faut expliquer

1. ce qu'est Internet et évoquer rapidement son vocabulaire spécifique
2. comment utiliser Mosaic et parvenir à la liste
3. comment procéder pour interroger les différents sites recensés

Comme les liens rattachés aux journaux ont été constitués afin de parvenir directement aux listes des sommaires disponibles, leur utilisation ne nécessite aucune explication. En revanche, l'interrogation des moteurs de recherche et des bases de données ou outils associés nécessite une aide générale sur l'utilisation des opérateurs booléens et une aide spécifique à chaque site. Les autres sites étant à vocation plus informative, ils ne seront pas décrits.

#### **3.2 Comment présenter le manuel ?**

Comme ce manuel doit être un support pour une interrogation à partir d'un ordinateur, il m'a semblé plus pratique d'utiliser un classeur. En effet, un classeur a l'avantage d'être facilement manipulable même s'il comporte 120 pages de format A4, ce qui est loin d'être le cas d'un fascicule !!! De plus, l'utilisation de pochettes plastiques permet non seulement de maintenir les pages du manuel propres, même si celui-ci est très utilisé mais aussi de les remplacer lorsqu'une modification est intervenue sur la page Web.

### 3.3. Sommaire du manuel obtenu

Ce sommaire permet de visualiser la progression de l'aide apportée.

---

<b>AVANT-PROPOS : Internet et son vocabulaire</b>	2
<b>PRECAUTIONS</b>	3
<b>ACCES A INTERNET</b>	3
Procédure pour accéder à Mosaic	3
Utilisation de Mosaic	3
<b>ACCES AUX SITES SELECTIONNES : Utilisation de «Hotlist»</b>	4
<b>COMMENT OPTIMISER UNE INTERROGATION</b>	9
<b>LES MOTEURS DE RECHERCHE</b>	10
<b>ACCES AUX DONNEES</b>	14
Données de séquences	15
Données protéiques	42
Autres	79
<b>BIBLIOGRAPHIE</b>	110
<b>ANNEXES</b>	113

---

### 3.4. Internet et son vocabulaire

Comme les chercheurs ne pratiquaient pas Internet, il m'a semblé important de faire un petit historique de son émergence, mais surtout indispensable de définir les termes les plus employés. Pour cela, j'ai utilisé le Dictionnaire du Multimédia<sup>3</sup> et l'Internet professionnel<sup>4</sup>.

-----

Internet a été créée en 1969 par la DARPA (Defense Advanced Research Projects Agency) comme un réseau expérimental de partage de données situées sur des ordinateurs distants. Ce réseau a ensuite été scindé en Milnet, réseau militaire et NSFnet (National Science Foundation network), réseau de la recherche. Depuis des centaines d'ordinateurs de par le monde se sont reliés pour donner une collection de réseaux.

Internet, réseau mondial qui s'agrandit tous les jours, possède une organisation originale, car non centralisée. L'Internet Society, créée en 1992, discute de l'avenir du réseau; l'IAB (Internet Architecture Board) est responsable des standards, et des groupes de travail spécialisés sont organisés par l'IETF (Internet Engineering Task Force) et émettent des recommandations. L'InterNIC (Internet Network Information Center) attribue les numéros IP (Internet Protocol): chaque ordinateur est reconnu grâce à un numéro et un nom propre (Spectrum a 194.167.35.11 comme numéro).

Le vocabulaire d'Internet est tout à fait spécifique :

- *Serveur*: machine ou logiciel qui offre des services à un client consommateur (utilisateur, machine ou logiciel) généralement serveur d'information. Les utilisateurs peuvent visualiser, récupérer des documents par des moyens variés.
  - *Lien*: relation établie entre différentes informations qui permet lorsque l'on accède à l'une d'accéder à l'autre rapidement.
  - *Gopher*: protocole et application qui permettent d'accéder à des informations sur Internet de manière conviviale.
  - *WWW ou Web*: le plus récent des services d'information, sur ensemble de Gopher avec HTML, l'hypertext, URL.
  - *HTML*: HyperText Markup Language, langage de description des pages Web.
  - *Hypertext*: langage qui permet la création de liens.
  - *URL*: Uniform Resource Location qui contient la méthode d'accès, le nom du serveur et le chemin d'accès au document.
  - *Adresse électronique*: tout utilisateur de messagerie doit avoir une adresse qui l'identifie personnellement. Sur Internet, elle est de la forme: *nom@nom\_organisation.pays*.
- 

<sup>3</sup> NOTAISE J., BARDA J., DUSANTER O. Dictionnaire du multimédia. Audiovisuel - Informatique - Télécommunication. AFNOR, 1995,- 886 p.

<sup>4</sup> L'Internet Professionnel. Témoignages, expériences, conseils pratiques de la communauté enseignement et recherche (choix, installation et utilisation de logiciels sur Macintosh, PC et stations Unix). CNRS & Universités, CNRS Editions, Paris, 1995,-448p.



### 3.5. Accès à Internet

Cette partie a été réalisée dans l'intention de permettre à l'utilisateur de parvenir à Internet par la liste de liens sans aucune connaissance préalable.

#### PRECAUTIONS

L'activation des fenêtres apparaissant en contour rouge se fait en cliquant sur le bouton gauche de la souris.

Attention pour écrire dans une fenêtre, il faut que la flèche de la souris se trouve dans la-dite fenêtre.

#### ACCES A INTERNET

Celui-ci est possible grâce à un logiciel. Sur Spectrum, il s'agit de Mosaic.

#### Procédure pour accéder à Mosaic

- Se connecter sur la machine<sup>(1)</sup>, c'est-à-dire donner son nom et son mot de passe.
- Aller dans le menu en haut à gauche: Tools puis Shell (dorénavant noté Tools/Shell).
- Lorsque la fenêtre "Winterm" apparaît ainsi que "spectrum 1%", taper "**mosaic**"

#### Utilisation de Mosaic

Seules les commandes vraiment utiles seront examinées, les autres pouvant être explorées si cela vous tente.

- Activer un lien : les mots ou expressions en bleu dans les pages Web représentent des liens. Lorsque la petite main blanche apparaît à la place de la flèche de la souris, cliquer du bouton gauche.

- Arrêter une recherche : cliquer sur le bouton en haut à droite (la petite terre qui tourne).

---

(1) Pour les personnes extérieures, il faut d'abord parvenir à la machine par une session «Telnet». La machine s'appelle « spectrum.iurc.inserm.fr » et son IP adresse est « 194.167.35.11 »

- Le menu File : Les commandes utiles sont :  
*Open URL* : Permet d'accéder à un serveur par son adresse.

*Save as*  
*Print*  
*Mail to*

Pour ces dernières commandes, le texte peut présenter plusieurs formes :  
Plain text : texte simple ASCII  
Formatted text : texte avec les formats (gras, italique...)  
Postscript : programme conservant la structure de la page Web  
HTML : programme sous-jacent à la page Web. (avec les liens....)

**! Pour imprimer**, changer la commande lpr en lp, sinon vous allez attendre longtemps l'impression, car la machine ne reconnaîtra pas la commande.

- Le menu Navigate : Les commandes utiles sont :

*Back* : Retour au site visité précédemment  
*Forward* : Avance au site suivant.  
*Home Document* : Renvoie à la page Mosaic NCSA.  
*Window History* : Historique des sites visités (mais sans lien pour y retourner).  
*Hotlist* : Liste des sites sélectionnés et mis en mémoire pour y accéder directement.  
*Add current to hotlist* : Ajoute le site courant sur la liste.

## ACCES AUX SITES SELECTIONNES

Utilisation de "Hotlist"

La liste des sites a été constituée au fur et à mesure des navigations sur Internet puis, une fois jugée convenable, réaménagée par thèmes pour faciliter les recherches.

Pour accéder à un site à partir de la liste, il faut le sélectionner et cliquer sur *Go To*.

Pour ajouter un site, il faut cliquer sur *Add Current* dans la fenêtre *Hotlist* ou dans le menu *Navigate*.

Les listes diffèrent un peu selon les utilisateurs, chacune devant spécifiquement répondre aux besoins et désirs exprimés par chacun. La liste ci-dessous est donc la plus complète possible.

-----

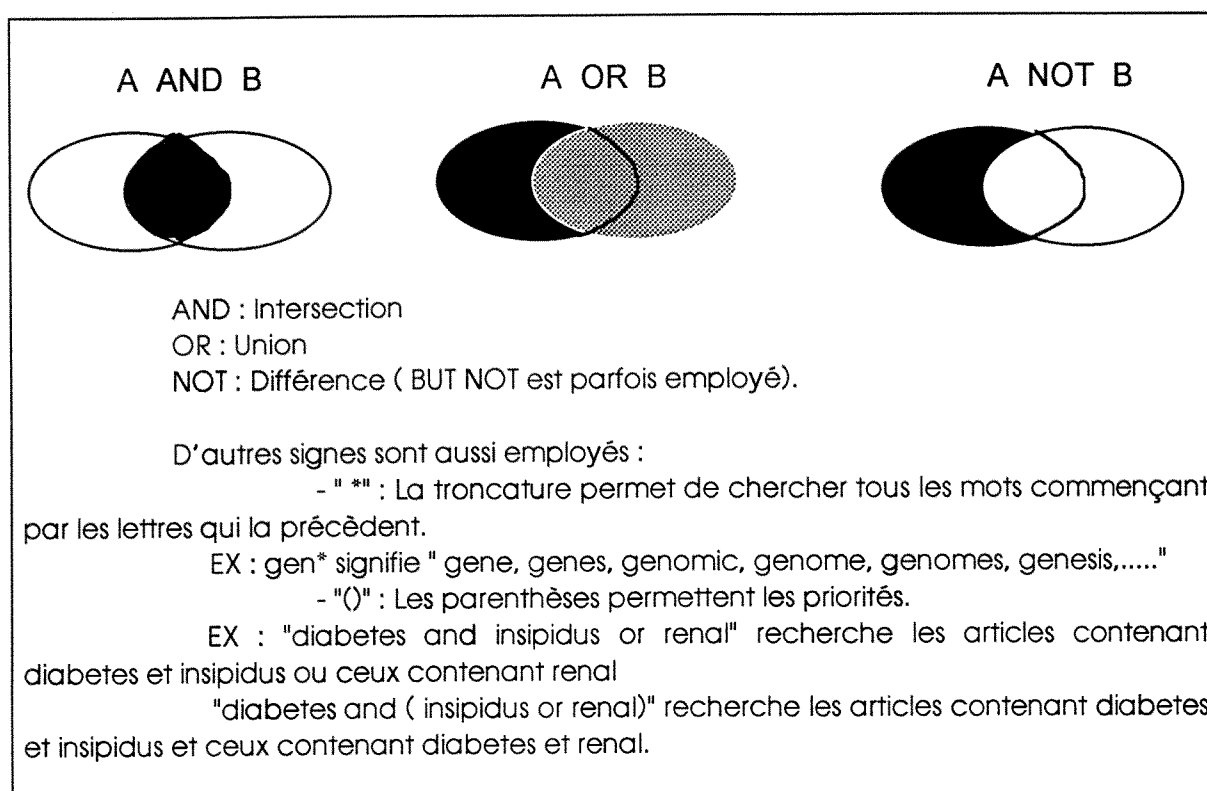
La liste complète avec toutes les adresses URL se trouve en annexe.

### 3.6. Optimisation d'une interrogation grâce aux opérateurs booléens

---

Plutôt qu'une simple interrogation avec un seul mot, il est possible d'associer plusieurs mots grâce à des "opérateurs". Comme la majorité des sites que vous aurez à interroger seront en anglais, nous allons raisonner dans cette langue.

Le schéma ci-dessous illustre le fonctionnement des différents opérateurs



Selon les bases de données ou les index, les mots opérateurs sont remplacés par des signes. Chez OMIM par exemple, and = &, or = |, not = -.

---

### 3.7. Description des moteurs de recherche

Les pages suivantes décrivent les moteurs de recherche en général puis, plus précisément, ceux sélectionnés. Pour chaque logiciel, après une petite phrase sur les types de sites recensés, les consignes d'interrogation sont exposées.

Les sources utilisées pour la réalisation de ces pages sont le rapport de stage de DESS de Laurence Rodrigues<sup>5</sup> et l'article de Jian Liu<sup>6</sup> sur les outils de recherche.

---

## LES MOTEURS DE RECHERCHE

Ces moteurs de recherche sont des logiciels qui permettent de retrouver ou de trouver un site par l'utilisation de mots-clés.

Ils fonctionnent tous globalement de la même manière: inscription des mots-clés et clic sur le bouton de recherche, et, à part Altavista, tous permettent de restreindre le champ des recherches par la sélection de catégories.

Yahoo, qui est en réalité un catalogue, est très perfectionné dans cette démarche puisqu'il est possible d'atteindre les sites en sélectionnant des sous catégories de plus en plus précises

Il faut savoir que ces logiciels ne recensent pas toujours les mêmes sites et donc, lorsque l'on veut retrouver tous les sites portant sur un sujet, il vaut mieux interroger tous les moteurs de recherche.

Ces différents outils vont être présentés ci-après:

**Altavista:** <http://www.altavista.digital.com/>

**Infoseek:** <http://www2.infoseek.com/>

**Lycos:** <http://lycos.cs.cmu.edu/>

**Webcrawler:** <http://www.webcrawler.com/>

**Yahoo:** <http://www.yahoo.com/>

---

<sup>5</sup> RODRIGUES L. Intérêt d'utiliser Internet pour les besoins documentaires et la veille technologique. Michelin. Rapport de stage DESSID 95.

<sup>6</sup> LIU J. Understanding WWW Search tools: <http://www.indiana.edu/~librcsd/search/>

## **ALTAVISTA**

Ce moteur recense uniquement les serveurs WWW et indexe les documents complets.

### ***Consignes d'interrogation:***

- le logiciel est sensible à la casse, c'est-à-dire que: "paris" permettra de retrouver des documents contenant paris, Paris, PARIS ou même pAris, mais "Paris" ne retrouvera que Paris.
- entre les mots l'intersection (et) est implicite
- pour chercher une expression, mettre celle-ci entre guillemets (ex: "petite galerie")
- pour rendre un mot obligatoire, le faire précéder, sans espace, du signe + (ex: +paris)
- pour interdire un mot, le faire précéder, sans espace, du signe -
- la troncature correspond au signe \*

## **INFOSEEK**

Cet outil recense uniquement les serveurs WWW et permet une recherche sur tout le document.

### ***Consignes d'interrogation:***

- le logiciel est sensible à la casse, c'est-à-dire que: "paris" permettra de retrouver des documents contenant paris, Paris, PARIS ou même pAris, mais "Paris" ne retrouvera que Paris.
- pour rechercher plusieurs mots (opérateur OU), les séparer d'une virgule (ex: holmes, sherlock, baskerville)
- pour chercher une expression, mettre celle-ci entre guillemets (ex: "petite galerie") ou bien, séparer les mots d'un tiret (ex: galerie-roi), cette dernière technique permettra aussi de retrouver galerie du roi.
- [sécurité ascenseur] recherchera ces deux termes dans un ensemble de 100 mots.

## LYCOS

Une petite anecdote pour commencer: le terme Lycos vient du mot Lycosidae qui désigne une famille d'araignées connues pour leur rapidité à la course.

Ce moteur recense tous les types de documents disponibles sur Internet et recherche sur l'ensemble du document

Lycos présente trois types de possibilités d'interrogation: Lycos, le catalogue A2Z et Point review qui recense les journaux et articles de journaux.

### **Consignes d'interrogation:**

- la troncature est implicite. Pour supprimer celle-ci, mettre un point à la fin des mots. Sinon, le signe \$ sert à exprimer une troncature.

- pour interdire un mot, le faire précéder, sans espace, du signe -

- Les deux catalogues sont interrogeables par catégories, qui sont ensuite sous-divisées pour A2Z (Arts & Humanities, Business & Investing, computers, education, Entertainment, Leisure, Government, Health & Medecine, Internet, Just for kids, News & Information, Science & Technology, Shopping the Net, Social & Community Affairs, Sports, The Road Less Traveled, The World)

## WEBCRAWLER

Webcrawler ne recense que les serveurs WWW et permet une recherche sur tout le document

### **Consignes d'interrogation:**

- il est possible d'utiliser les opérateurs booléens directement dans la requête. Ceux-ci sont AND, OR, NOT.

- pour chercher une expression, mettre celle-ci entre guillemets (ex: "petite galerie") ou bien, séparer les mots de "ADJ" (ex: petite ADJ galerie)

- pour rechercher deux termes séparés de n mots, utiliser l'expression suivante terme1 NEAR/n terme2 (ex: assistance NEAR/5 environnement trouvera "l'assistance à la maîtrise de l'environnement").

- l'utilisation des parenthèses est possible dans n'importe quelle requête.

- Webcrawler est aussi interrogeable par catégories, 16 existent, elles-mêmes ensuite sous-divisées (Arts & Entertainment, Business, Computers, Daily News, Education, Government & Politics, Health & Medecine, Humanities, Internet, Life & Culture, Personal Finance, Recreation & Hobbies, Science & Technology, Sports).

## YAHOO

Yahoo est un catalogue et, de ce fait, divisé en catégories (Arts, Business & Economy, Computers & Internet, Education, Entertainment, Government, Health, News, Recreation & Sports, Reference, regional, Science, Social Science, Society & Culture), mais, contrairement aux moteurs de recherche, les premières sous divisions sont accessibles dès la première page.

### **Consignes d'interrogation:**

- le logiciel ne tient pas compte de la casse, il retrouvera tous les documents sur Paris quelle que soit la manière de mélanger minuscules et majuscules.
  - pour pouvoir interroger sur plusieurs mots, il faut choisir "Options"
    - \* choisir le lien existant entre les mots, AND= All keys, OR= At least one.
    - \* la troncature peut être activée en choisissant: Substrings ou désactivée par: Complete words.
- 

## **3.8. Description des bases de données et des outils associés**

Cette partie, bien évidemment la plus conséquente du manuel, reprend les pages visualisées à l'écran et explicite leur utilisation.

### **1. Sites répertoriés**

En tout, 24 sites ont été conservés. Ils ont été classés en fonction des données concernées:

- séquences d'ADN,
- séquences protéiques,
- autres: c'est-à-dire, bibliographie, données annexes (Vertebrate Comparative Database , par exemple), sites permettant l'interrogation de plusieurs bases d'ADN et/ou de protéines.

La liste ci dessous est la page de sommaire des bases de données. Il n'est pas nécessaire de décrire les bases de données autrement que par un titre, suffisamment explicite pour les chercheurs.

---

### **DONNEES DE SEQUENCES**

GenBank Database	16
GDB - Genome Database	18
GSDB - Genome Sequence Database	20
HGMD - Human Genome Mutation Database	27

GENOTK - Human cDNA Database	37
DDBJ - DNA Data Bank of Japan	23
TBase - Transgenic/Targeted Mutation Database	24
BLAST - Basic Local Alignment Search Tool	28
NDB - Nucleic Acid Database	34

## DONNEES PROTEIQUES

Swiss Prot	44
PIR - Protein Identification Ressource	50
OWL - Non Redundant Composite Database (PIR + SWISS-PROT)	43
PDB - Protein Database	42
Protein Kinase Database Project	75
MOOSE - Macromolecular Object Oriented Search Engine	53
Dali - 3-D Protein Structure Database	47

## AUTRES DONNEES

Medline	80
OMIM - Online Mendelian Inheritance in Man	82
GENDIAG - Syndrome Database	86
Vertebrate Comparative Database	84
Entrez - Retrieval System	89
SRS - Sequence Retrieval System	101

## 2. Présentation des bases de données

Toutes les pages de présentation et d'explication sont constituées de la manière suivante.

1. le titre, bien sûr
2. quelques lignes sur la provenance de la base de données: l'organisme producteur et l'origine des données
3. la liste des informations obtenues dans les réponses
4. la procédure à suivre pour l'interrogation
5. une référence bibliographique, décrivant la base de données, et

l'URL

Il n'a pas toujours été évident de retrouver ces différentes données. Pour en rendre compte, les différents points évoqués ci-dessus vont être repris et les difficultés rencontrées décrites.

2. La provenance: Tout d'abord, le serveur permettant l'accès à la base de données n'est pas toujours situé dans le lieu où celle est constituée. Et lorsqu'elle est le résultat d'une collaboration, c'est encore moins évident! Quant à l'origine des données, cette information est souvent disponible dans les pages d'aide sur la base de données.



Pour répondre à ces questions, plusieurs sources possibles:

- les sites des bases de données elles-mêmes
- des sites associés présentant des manuels d'utilisation
- les articles sur les bases de données
- les ouvrages suivants: World Databases in Medecine<sup>7</sup> et

CD-Rom Directory<sup>8</sup>

3. Les réponses: Pour obtenir ces informations, un seul moyen: interroger la base de données et observer les réponses, ce qui implique aussi parfois, une phase de décryptage des abréviations employées pour définir les différents champs. En effet, autant AU ou SQ sont faciles à décrypter comme correspondant à "auteur" et "séquence" autant FT (EMBL) ou DR (Prosite) ont été difficiles à traduire, même en ayant les informations correspondantes. Heureusement, il existe parfois des manuels d'utilisateurs expliquant la signification des intitulés de champs. Pour EMBL, voir Nucleotide Sequence Database User Manual.<sup>9</sup> Grâce à ce manuel, voici la traduction de FT: Feature Table, qui inclue des informations sur les gènes et les produits génomiques et sur les régions importantes du gène. Quant à DR, je pense qu'il s'agit de "Domain region".

4. La procédure: celle-ci est importante, car, lorsqu'une étape est oubliée, les réponses peuvent être erronées voire nulles. Cette procédure a été encadrée afin de la "faire sauter aux yeux" des utilisateurs.

5. La référence: Un certain nombre de références sont fournies avec les explications sur les bases de données, d'autres ont été trouvées grâce à Medline, en interrogeant par le nom de la base, d'autres encore ont été obtenues à la suite d'un email à l'adresse indiquée pour les éventuels commentaires ou problèmes.

Une autre difficulté a été l'édition de la page visualisée à l'écran. En effet, il est impossible d'imprimer les boîtes de dialogue ou les listes de choix ou même les boutons à cliquer. Or toutes les pages possèdent au moins les boutons équivalents à "rechercher" ou "effacer".

Afin de rendre le plus fidèlement possible les pages d'Internet, les pages qui ne présentait pas trop de listes de choix ont été éditées en postscript, parfois directement, mais généralement après sauvegarde et remaniement par ajout d'espaces pour rendre la page du manuel la plus proche possible de celle sur l'écran. Les pages présentant de nombreuses listes de choix ont été sauvegardées en HTML, seul moyen d'avoir accès à toutes les informations présentes sur la page. Pour le remaniement de ces pages, tous les "tags" HTML ont été effacés (un par un !) et la présentation modifiée.

Comme il est impossible de présenter tout le manuel, seuls deux exemples seront décrits, l'un dont la page a été obtenue avec la première méthode et l'autre avec la deuxième méthode. Il s'agit de GENBANK et de la PDB (Protein Data Bank).

<sup>7</sup> World Databases in Medecine. Armstrong C.J. ed. Bowker-Saur, London, 1993, -875p..

<sup>8</sup> CD-Rom Directory 7<sup>th</sup> edition. Finlay M., Mitchell J. eds. TFPL, London, 1992.,-868

<sup>9</sup> Nucleotide Sequence Database User Manual: [http://www.ebi.ca.uk/ebi\\_docs/embl\\_db/usrman/usrman.html](http://www.ebi.ca.uk/ebi_docs/embl_db/usrman/usrman.html).

#### *A. GENBANK*

Genbank est une des trois bases de données de séquences d'ADN les plus utilisées par les chercheurs en génétique; avec EMBL (European Molecular Biology Laboratory Database) et DDBJ (DNA Data Bank of Japan). C'est aussi une des seules bases qui, grâce à J. Derancourt, ont été sélectionnées après une recherche sur leur nom.

Les pages 32 et 33 sont les pages du manuel consacrées à Genbank, la page explicative et la page visualisée à l'écran, éditée en postscript et sur laquelle ont été rajoutées les boîtes de dialogue ou de sélection des paramètres.

#### *B. PDB*

La PDB (Protein Data Bank) de Brookhaven, est la seule banque qui ait été utilisée régulièrement au laboratoire, car elle était sur place (téléchargement en 1993).

Les pages 34 et 35 correspondent aux pages du manuel relatives à cette banque; la page provenant d'Internet a été éditée en texte simple après sauvegarde en HTML.

# GENBANK

Genbank est produite par le NIH (National Institutes of Health ) basé à Bethesda (Maryland). Les informations proviennent de journaux mais aussi directement des soumissions d'auteurs .

Chaque entrée présente les informations suivantes :

- Description concise de la séquence
- Nom scientifique
- Taxonomie de la source
- Caractéristiques de la séquence
- Références bibliographiques (liens avec Medline).

Les bases dbEST et dbSTS sont des sous-produits de GENBANK (voir SRS p103)

- Entrer les mots clés
- Appliquer les opérateurs si nécessaire
- Limiter la recherche à certains champs (cliquer "fr" et sélectionner les champs)
- Sélectionner la base
- Décider du nombre maximum de réponses (1)
- Cliquer " Run Query"

NB: Voir "The GenBank DNA Sequence Database" en annexe

Source: BENSON D., BOGUSKI M., LIPMAN D.J., OSTELL J. (1994) GenBank. Nucleic Acids Research, 22 : 3441-3444.  
[http://www2.ncbi.nlm.nih.gov/genbank/query\\_form.html](http://www2.ncbi.nlm.nih.gov/genbank/query_form.html)

- (1) Choix entre
- 1st 100
  - 2nd 100
  - 3rd 100
  - 4th 100
  - 5th 100
  - 6th 100
  - 7th 100
  - 8th 100
  - 9th 100
  - All



### Query GenBank Database

Enter your query in the boxes below. A document matches a box if all words within it appear in one field of the document. The documents matching a box are combined with boxes following it using the selected operator to its right. The AND and BUT NOT operators limit retrieval and the OR operator expands it.

Checking the "fr" (Field Restriction) box limits the match to the fields checked in the table below.

Additional help is [here](#).

fr	<input type="text"/>	<input type="radio"/> and	<input type="radio"/> or	<input type="radio"/> but
not				
fr	<input type="text"/>	<input type="radio"/> and	<input type="radio"/> or	<input type="radio"/> but
not				
fr	<input type="text"/>	<input type="radio"/> and	<input type="radio"/> or	<input type="radio"/> but
not				
fr	<input type="text"/>			

Return up to  documents.\*\*

\*\* Additional documents can be retrieved using the (More) button from the document list. \*\*

Select Database:  GenBank  GenBank Updates  Both

Field Restriction: (Applied only on query lines where "(fr)" is checked)

- Locus
- Definition
- Accession No.
- NID
- Keywords
- Source
- Reference
- Comment
- Features

Rev. 06/27/96



# PDB

## The Protein Data Bank

PDB est qualifiée d'archive de structures biologiques macromoléculaires 3D par le laboratoire producteur: Brookhaven National Laboratory (Upton NY).

Chaque entrée présente les informations suivantes :

- Titre
- Composant
- Source
- Auteur
- Date
- Journal
- Remarques (référence, résolution, raffinage, description...)
- Séquence (formule, hélices, angles, liens cystéine, origine, échelle)

- Sélectionner les options d'affichage (Display options)
- Remplir les champs à explorer
- Cliquer sur "Send Request"
  
- Plus bas sur la feuille, dans la section "List of n Matching ID codes", sélectionner une réponse à la question précédemment posée
- Cliquer "Fetch the Entry with ID Code"

Source: ABOLA E.E., BERNSTEIN F.C., KOETZLE T.F. (1988) The Protein Data Bank in *Computational molecular biology. Source and methods for sequence analysis*, (LESK A.M., ed), Oxford University Press, Oxford.  
<http://pdb.pdb.bnl.gov/cgi-bin/browse>

PDB

Protein Data Bank

This is the PDB Browser, providing fast access to the entire structure data base. Help is available at the click of a mouse.

(Please note - some Macintosh browsers have trouble with this page. You may have to try Mosaic, MacWeb, or Netscape.) If you are having problems with a search involving the characters \*, ?, or +, please see this description of regular expressions.

Display Options

List ID	} a selectionner	Sort by ID	}	Search Full PDB	}
List Title		Sort by Text		Search List Only	
List Compound		Do Not Sort			
List Author					
List Experiment					
List Class					
List Source					
List Journal		And Constraints	}	Apply Constraints	}
List Het		Or Constraints		Invert Constraints	
List Formula					

---

Search Criteria

(See next section for results.)

ID:

Class:

Title:

Compound:

Author:

Source:

Experiment:

Journal:

Cryst1:

Het:

Formula:

Resolution Worse Than  
Resolution Better Than

Accession on or After (dd-mmm-yy)  
Accession on or Before (dd-mmm-yy)

---

Please see the display of matching ID-Codes below

### 3. Présentation des outils de recherche

Une des difficultés pour présenter les systèmes de recherche comme SRS ou Entrez, est que ceux-ci permettent l'accès à un grand nombre de bases de données. Toutefois, la procédure pour les interroger est identique, ce qui facilite les choses, mais avec le seul nom des bases de données, surtout lorsque c'est un sigle, il n'est pas évident de savoir ce qu'elles contiennent.

L'exemple explicité sera SRS, car c'est le plus complexe des deux. En effet, ce système est utilisé à l'Université de Heidelberg (Allemagne) mais aussi au centre Sanger (Grande-Bretagne). Les deux centres ne sont pas miroirs l'un de l'autre même si certaines bases de données sont, heureusement, communes.

En outre, les deux sites ne permettent pas l'accès au même nombre de bases ni même aux mêmes catégories. Le site du Centre Sanger offre 65 bases de données différentes alors que celui de l'Université de Heidelberg n'en présente que 31. Voici les catégories de recherche communes aux deux sites:

- Search sequence libraries
- Search libraries with protein structure information
- Search a library linked to sequence libraries
- Search bibliographic libraries
- Search other libraries
- Search in one of the Transfac Files

Et celles spécifiques au Centre Sanger:

- Search a genome library
- Search a metabolism library
- Search mapping libraries

Il a donc fallu trouver comment parvenir à regrouper les deux sites. Ceci a été réalisé grâce à un tableau présenté ci après.

Il n'a pas été possible de retrouver la signification de tous les sigles, car le dernier recours: un message envoyé à l'administrateur du système, n'a toujours pas obtenu de réponse. La plupart des titres complets ou une explication du contenu des bases existent dans la page liée au différents titres. En outre, les références associées y ont été découvertes.

# SRS

## Sequence Retrieval System

SRS n'est pas une base de données mais un système qui permet d'en interroger un grand nombre. Deux serveurs s'appellent SRS, l'un est basé à l'Université de Heidelberg et l'autre à l'Institut Sanger.

Les pages de présentation sont à peu près les mêmes dans leur contenu et leur fonctionnement est identique.

Le tableau ci dessous recense tous les produits (bases ou programmes spécifiques d'exploration de celles ci). La colonne SRS indique la localisation du produit Heidelberg (H) ou Sanger (S). Le nombre entre parenthèses correspond au numéro de la référence bibliographique.

	Nom complet ou description	SRS
<b>SEQUENCE</b>		
EMBL	European Molecular Biology Laboratory Database	S,H
EMNEW	Index of new EMBL Sequences	S,H
GBRNA	Great Britain RNA Database	S
GENBANK (11)	Nucleotide Sequence Database	S
NRL3D (39)	Protein Sequence in Brookhaven PDB	S,H
PIR (24,25,52)	Protein Information Resources	S,H
SBASE (43)	Annotated Protein Domain Sequence Segment	S
SPTREMBL (10)	Swiss-Prot's Translated Peptide Sequences from EMBL	S
SWISS-PROT (8,10)	Protein Sequence Database	S,H
SWISS-NEW	New Swiss-Prot Sequences	S,H
TREMBL	Translated Peptide Sequences From EMBL	S
TREMBLNEW	New Translated Peptide Sequences From EMBL	S

Source: ETZGOLD T., ARGOS P. (1993) SRS, an indexing and retrieval tool for flat file data libraries. *Computer Applications in the Biosciences*, 9 : 49-57.

<http://www.sanger.ac.uk/srs/srsc>

<http://www.embl-heidelberg.de/srs/srsc>



	Nom complet ou description	SRS
<b>STRUCTURES</b>		
ALI (29)	Database Merging protein structures and sequences	H
FSSP (30)	Families of Structural Similar Protein	H
HSSP (46,47)	Homology-derived Secondary Structure of Protein	H
PDB SHORT		S
PDB (1)	Protein Database of Brookhaven	H
<b>LINKED SEQ</b>		
BLOCKS (27,42)	Protein Nlocks Database	S,H
CPGISLE (36)	Human CpG Islands Database	S
DBEST (15)	Database for Expressed Sequence Tags	S
DBSTS (14)	Database of Sequence Tagged	S
ECDC (35)	E. coli K12 Chromosome	S,H
ENZYME (6)	Database of EC Nomenclature	S,H
EPD	Eucaryotic promoter Database	S,H
FLYGENE (23)	Partie de FLYBASE(Droso Genetic Map Database)	H
MIMMAP (37)	Maps from OMIM	H
MOLPROBE (2)	Molecular Probe Database	H
OMIM (37)	Online Mendelian Inheritance in Man	H
PFAM	Protein Family Alignment	S
PIRALN	Database of Protein Sequence Alignment	S,H
PMD	Protein Mutant Database	H
PRINTS (5,4)	Protein Motif Fingerprints Database	S
PRODOM	Families of Homologous Protein Domains	S,H
PROSITE (7,9)	Protein Pattern Database	S,H
PROSITEDOC	Documents from Prosite	S,H
REBASE (44)	Restriction Enzyme Database	S,H
SWISSDOM	Swiss-Prot entries in a Prodom map	S,H
SWISSPFAM	SWISS-PROT Protein Family Alignment	S
YPD	Yeast Protein Database	H
<b>BIBLIO</b>		
FLYREF (23)	Flybase Bibliography	S
MEDLINE	Genetics Bibliography	H
SEQANALREF	Sequence Analysis References	S
SEQANALABS	Sequence Analysis Abstracts	S
YPDREF	YPD Bibliography	S,H

	Nom complet ou description	SRS
<b>TRANSFAC</b>	Database on eucaryotic cis-acting regulatory DNA elements and transacting factors	
TFSITE	Index of the site file of Transfac	S,H
TFFACTOR	Index of the transcription factor file	S,H
TFMATRIX	Index of the matrix file	S
<b>GENOME</b>		
ECDCONTIG(55)	Complete, non redundant sequence of E.coli K12	S
EC2D	E.coli proteins obtained through 2D gels	S
FLYCLONES(23)		S
LISTA (18)	Nucleotide Sequence from Saccharomyces	S
LISTAHON (18)	Nucleotide Homology section of LISTA	S
LISTAHOP (18)	Protein Homology section of LISTA	S
NRSUB (41)	Non Redundant Database of Bacillus Subtilis	S
SACLOC	Saccharomyces Genome Database Locus Information	S
SACMAP	Sacch. Gen. DB Genetic Map Data	S
SACORF	Sacch. Gen. DB Open-reading Frames Data	S
SACPMAP	Sacch. Gen. DB Physical Map Data	S
SACSEQ	Sacch. Gen. DB Sequence region Data	S
UNIEST	All human EST Sequences	S
UNIGENE	Unique Human Gene Collection (From GenBank)	S
VIRGIL	Library of links between human GDB & Genbank	S
WORMPEP	Predicted protein sequences from C. elegans	S
<b>MAPPING</b>		
RHDB	Radiation Hybrid Database: Result file	S
RHEXP	Radiation Hybrid Database: Experimental conditions	S
RHPANEL	Radiation Hybrid Database: Panel data	S
RHMAP	Radiation Hybrid Database: Map	S
<b>METABOLIS(33)</b>		
ECOCYCCOLI	E. coli reactions components	S
ECOCYCCOMP	Compounds	S
ECOCYCENZ	Enzymatic reactions	S
ECOCYCGENE	Genes	S
ECOCYCPROT	Enzyme (protein) components	S
ECOCYCPATH	Metabolic pathway	S

	Nom complet ou description	SRS
<b>OTHER</b>		
PDBFINDER	Protein Data Bank Finder=PDB+DSSP+HSSP	S,H
LIMB	Listing of Molecular Biology Databases	S,H
NAKAI	Amino Acid Index Database	S
NAKAI MATRIX		S

Il existe deux méthodes d'utilisation de SRS, selon si les bases sont sélectionnées (1) ou non (2) sur la page d'accueil.

(1)

- Sélectionner la bibliothèque
- Cliquer sur "Search ....."

(2)

- Cliquer directement sur "Search ....."
- Sélectionner la ou les bases sur la page suivante (exemple avec les bases de séquences)

- Sélectionner le(s) champ(s) à explorer
- Rentrer le texte
- Choisir l'opérateur pour combiner les champs
- Sélectionner les champs qui doivent apparaître dans la liste des résultats
- Sélectionner le format des séquences (spécifique aux séquences)
- Sélectionner le type de résultat (spécifique aux séquences)
- Choisir le nombre maximum de réponses à récupérer
- Choisir le nombre de réponses à apparaître sur une même page
- Cliquer "DO QUERY"

-----

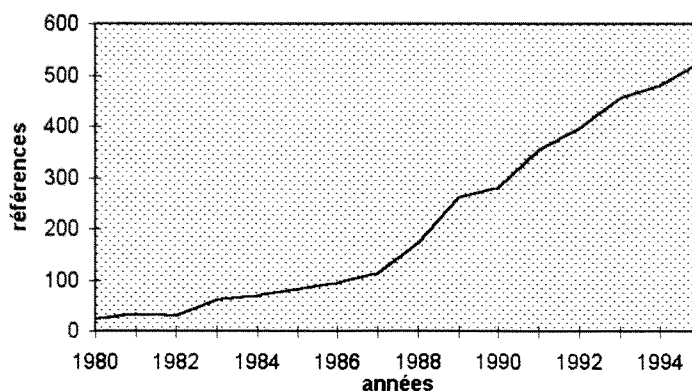
La bibliographie spécifique au manuel, où seront retrouvées les références relatives aux bases ci dessus, est une sous partie de la bibliographie générale, p 53.

## 4. RESULTATS

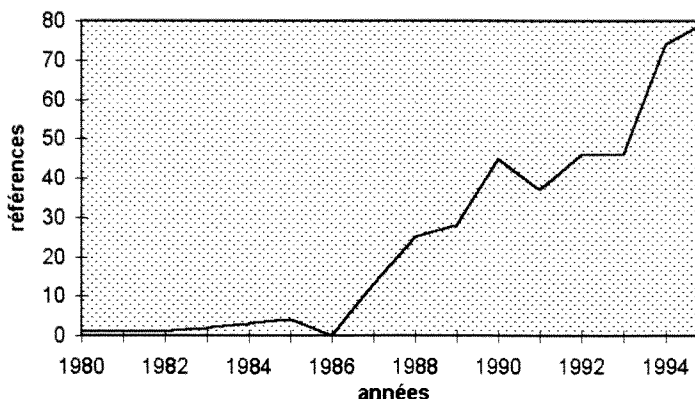
Le manuel a, pour l'instant, surtout été utilisé pour la mise à jour de la bibliographie, longtemps négligée. Ainsi, Medline, OMIM ont été interrogés à propos du récepteur de l'insuline, de ses mutations et Gendiag sur le syndrome d'Alström; ainsi pour cette dernière interrogation, de nombreuses références ont été découvertes. En effet, Medline génétique, interrogé par Alström renvoie 4 références et Gendiag en offre 12, dont la moitié était inconnues.

D'autre part, j'ai interrogé Medline génétique, année par année, afin d'observer l'évolution du nombre de références sur les sujets clés du laboratoire. Les deux interrogations ont été effectuées en moins d'une heure. Les deux graphes ci dessous en sont le résultat.

**Tableau 2: Evolution du nombre de références concernant le récepteur de l'insuline sur une période de 15 ans (1980-1995)**



**Tableau 3: Evolution du nombre de références concernant les mutations du récepteur de l'insuline sur une période de 15 ans (1980-1995)**



Les résultats obtenus serviront pour démontrer l'importance de la recherche sur les mutations dans la recherche sur le récepteur de l'insuline, lors d'un congrès prochain.

Un autre type de recherche, toujours sur Medline génétique, a été effectuée sur les différents sujets de recherche du laboratoire. Ils concernent le récepteur de l'insuline et les protéines substrat, les syndromes cliniques relatifs à l'insulino-résistance, une drogue anti-diabétique et le modèle théorique d'étude: un batracien. Ces résultats sont présentés dans le tableau ci après.

**Tableau 4: Réponses de Medline génétique à diverses interrogations sur deux durées différentes**

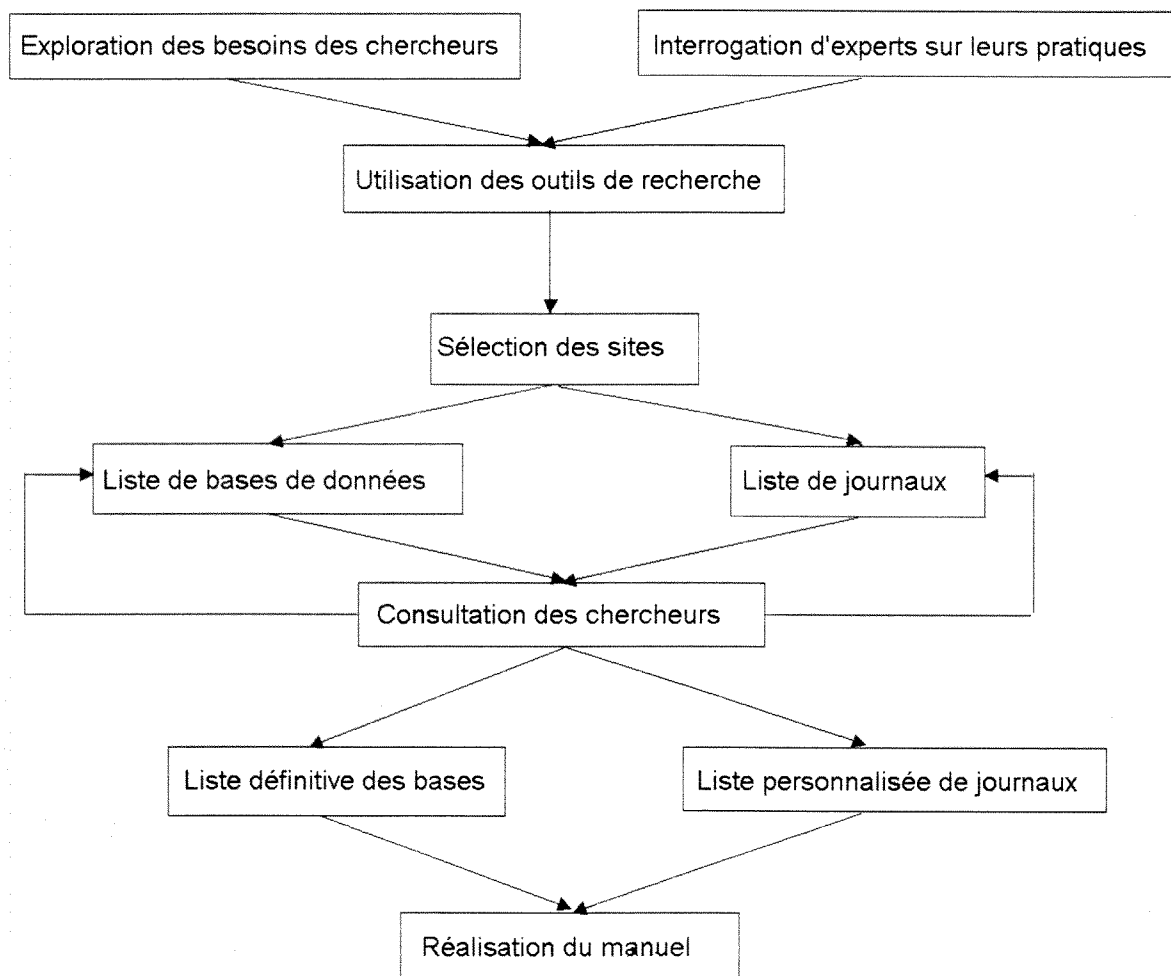
	1996	1987-96
	août	
"insulin receptor"	47	2944
"insulin receptor mutation"	7	199
"insulin resistance"	10	465
"IRS-1"	11	243
"IRS-2"	2	137
"polycystic ovary"		29
"alstöm"		4
"leprechaunism"		49
"NIDDM"	11	375
"metformin"	1	17
"xenopus oocytes"	4	893

Sur une période de 10 ans, la quantité de publications à lire est impressionnante, mais si l'on considère seulement la mise à jour du mois de septembre, ça l'est encore plus. En effet, il aurait fallu lire 93 articles dans le mois, soit au moins quatre articles par jour ouvrable !

Pour faire face à une telle situation et ne pas se laisser déborder par la masse d'information, il faut être très rigoureux dans ses interrogations. Le mieux serait d'en pratiquer une tous les mois, afin de bénéficier de la mise à jour mensuelle de Medline et de la PDB notamment.

## 5.CONCLUSION

En guise de conclusion, le schéma suivant qui résume la stratégie suivie:



## **TROISIEME PARTIE :**

# **CREATION D'UN SITE WEB SUR LE LABORATOIRE D'ENDOCRINOLOGIE MOLECULAIRE**

# INTRODUCTION

La réalisation d'un serveur Web est prévue pour tout l'Institut, mais elle dépend de la venue d'un objecteur de conscience spécialisé en graphisme !! Chaque laboratoire devait donc, en prévision de sa venue, réfléchir aux informations à mettre à disposition de la communauté sur Internet.

Florin Grigorescu, voyant ma connaissance du réseau et de ses usages m'a demandé de travailler à la réalisation du site consacré au Laboratoire d'Endocrinologie Moléculaire. Il voulait, en effet, pouvoir présenter le site au congrès de l'IFLA qui aura lieu à partir du 26 septembre 96 en Sicile. Il avait prévu de réaliser une page sur la Metformine, drogue anti-diabétique, produite par ce laboratoire pharmaceutique.

Nous nous sommes donc penchés sur ce projet et sur ce qui était nécessaire avant de réaliser un site Web.

1. Connaître HTML, HyperTexte Markup Language, pour programmer les pages.
2. Sur quelle machine et sous quel logiciel travailler ?
3. Quelles informations ?

Les pages qui suivent vont tenter de répondre à ces divers points.

## 1. INTRODUCTION A HTML

HTML ou HyperText Markup Language a été créé pour réaliser l'organisation logique d'un document comportant de nombreuses extensions en hypertexte. Une des qualités de ce langage est qu'il peut être lu par n'importe quel logiciel approprié (en anglais "browser"): il n'y a pas de problème de compatibilité.

Les "étiquettes" HTML, entre < et >, divisent le texte d'un document en blocs appelés éléments qui peuvent eux même être divisés.

La structure primaire d'un document HTML est la suivante:

```
<HTML>
  <HEAD>
    <TITLE> Titre du document </TITLE>
  </HEAD>
  <BODY>
    Texte du document, qui peut avoir la même apparence qu'un texte
    produit par un éditeur de texte.
  </BODY>
</HTML>
```

Chaque étiquette de début d'élément < > possède une étiquette miroir de fin d'élément </ >.



Le corps du document peut contenir les étiquettes suivantes:

**<BODY>**

**<H1>** Titre du premier chapitre **</H1>**  
    **<H2>** Titre du premier sous chapitre **</H2>**  
        Texte du premier paragraphe  
    **<p>** Texte du second paragraphe qui en plus possède un  
    **<A HREF="http://.....">**lien **</A>** hypertexte.  
    **<H2>** Titre du deuxième sous chapitre **</H2>**  
        Texte  
    **<IMG SRC=/ligne/horizontale.gif>**  
**<H1>** Titre du second chapitre **</H1>**

⋮  
⋮

**</BODY>**

Le texte entre les deux étiquettes BODY va être détaillé:

- Les étiquettes de titre ( H comme heading) se déclinent sur 6 niveaux, qui ont des tailles de caractères différentes.
- Le saut de ligne qui débute un paragraphe est commandé par **<p>** uniquement.
- Un lien hypertexte est indiqué par **<A REF=" >** et **</A>**
- Une image peut être insérée dans le texte par **<IMG SRC=....>**, il est même possible de définir son alignement par rapport à la page.

Le formatage du texte est possible grâce à **<b>..</b>** pour gras ou **<i>...</i>** pour italique.

Des commandes pour réaliser des listes existent. Pour des listes:

- non numérotées, il faut utiliser **<UL>**  
    **<LI>**  
    **<LI>**  
    **<LI>**  
    **</UL>**
- numérotées, il faut utiliser **<OL>** et **</OL>**.

Avec ces quelques notions de HTML, il est possible d'entreprendre la création de pages HTML simples. Toutefois, il existe de très nombreux guides sur Internet. Deux d'entre eux m'ont permis de me familiariser avec les différentes étiquettes HTML:

- A Beginner's Guide to HTML:  
    <http://www.ncsa.uiuc.edu/General/Internet/WWW/HTMLPrimer.html>
- Introduction to HTML:  
    <http://www.utoronto.ca/webdocs/HTMLdocs/NewHTML/htmlindex.html>

Une autre source m'a été utile: le rapport de stage d'Annie Feyfant<sup>10</sup>.

---

<sup>10</sup> FEYFANT A. 20 ans après...de l'informatique documentaire à l'ingénierie documentaire. DESSID 95.

## **2. QUEL MATERIEL ?**

### **2.1. Quelle machine et quel logiciel utiliser ?**

Le serveur qui devait accueillir les pages du Laboratoire est celui du CRBM (Centre de Recherche en Biochimie Macromoléculaire) et non celui de l'IURC, non encore accessible.

Pour réaliser des pages Web, il faut non seulement pouvoir programmer en HTML mais aussi posséder un logiciel, dit éditeur, capable de traduire ce langage. Or, aucun éditeur ne pouvait fonctionner sur la machine Unix du laboratoire.

La solution envisagée était: de réaliser les pages sur un éditeur de texte puis de les amener au CRBM et les tester là bas. Mais comme cela s'est déroulé début août, et, que pour la plupart des personnes, c'est la période des vacances, j'ai pu avoir accès à un PC et même à Netscape Gold 2.0 sous Windows 95.

### **2.2. Netscape Gold 2.0.**

Ce logiciel, Netscape Gold 2.0, possède un éditeur-assistant HTML et est très facile à prendre en main. En outre, pour l'utilisateur, son fonctionnement ressemble à celui d'un simple éditeur de texte: le choix de police et de format des caractères, de la hiérarchie des titres....se fait comme dans Word, en sélectionnant le texte et en choisissant une commande dans le menu ou la barre d'outil. Le logiciel transforme toutes ces instructions en HTML, sans oublier les déclarations préliminaires à l'écriture du texte: head, body...

Dès les premiers instants, il est possible de visualiser sa page et de vérifier le fonctionnement des liens.

## **3. QUEL BUT ?**

Ce site Web devait servir aux tâches suivantes:

1. Offrir une liste de peptides dans le cadre d'un projet de collaboration européenne (action COST B5)
2. Donner les instructions pour les envois de prélèvements sanguins. Ces échantillons doivent être utilisés pour la réalisation de la banque d'ADN.
3. Présenter l'activité du laboratoire
4. Préparer la réalisation de la base IRMDB: Insulin Receptor Mutation Database

## **4. QUELLES INFORMATIONS ?**

### **4.1. Le laboratoire d'Endocrinologie Moléculaire**

Il a été décidé de présenter

- la localisation
- les missions de recherche
- les différents membres du laboratoire, actuels et anciens,
- les centres d'intérêts en recherche

Pour réaliser cette partie, nous nous sommes appuyés sur des documents administratifs de présentation du laboratoire déjà existants (réalisés à l'occasion de la formation de l'IURC ou pour des candidatures diverses).

### **4.2. Les sites sélectionnés**

Ces sites sont parmi ceux sélectionnés dans le manuel. Le critère de choix a été leur aptitude à intéresser la plupart des chercheurs en génétique ainsi que leur renommée.

### **4.3. Les données liées aux activités de recherche**

Ces données sont en fait celles qui ont motivé la création de ce site Web. En effet, il s'agit d'un reflet direct de l'activité du laboratoire et une manière d'obtenir des contacts.

En effet, la liste des peptides disponibles dans le laboratoire devrait amener des chercheurs, à la recherche de ces peptides, à contacter le laboratoire pour commander.

Les autres listes

- liste des mutations du récepteur
- liste des noms mondains des malades (en quelque sorte leur nom de code,

mon3 par exemple, signifie le troisième malade identifié à Montpellier)

mont3 sont d'un grand intérêt pour les chercheurs travaillant sur le récepteur de l'insuline, car, il est fort probable que tous les papiers qu'ils ont collectés sur les mutations soient dans un dossier et qu'aucune liste claire ne les recensent.

Une phrase sur la future base de données sur le récepteur de l'insuline permettra d'attirer l'attention sur celle-ci.

### **4.4. Les activités de services du laboratoire**

Comme le laboratoire réalise de nombreux séquençages d'ADN, il a été décidé de proposer ce service à l'extérieur.

#### **4.5. Les positions par rapport aux étudiants**

Cette partie explicite les démarches à effectuer

- pour réaliser un stage de DEA Biologie-Santé (Montpellier I) dans le laboratoire
- pour poser sa candidature pour une thèse ou un post-doctorat

#### **4.6. La liste des publications**

### **5. QUEL RESULTAT ?**

Il a été décidé de réaliser ce site en anglais. Mais sur les conseils des informaticiens du CRBM, lors de la présentation de la première ébauche, la traduction en français a été programmée et amorcée.

Après avoir visité les sites de l'INTD<sup>11</sup> et du CNAM<sup>12</sup> qui m'ont séduite par leur clarté, je me suis attachée à réaliser des pages courtes, tenant sur un écran, afin que le visiteur n'ait pas à toucher à la barre de défilement du texte.


La page qui suit est la première page du site anglais. Malheureusement, elle ne tient pas sur un écran, mais sur une page A4 !

---

<sup>11</sup> INTD: <http://www.cnam.fr/INTD/index.html>

<sup>12</sup> CNAM: <http://www.cnam.fr/>

# Molecular Endocrinology Laboratory

Vous préférez le français, cliquez sur le drapeau 

The Molecular Endocrinology Laboratory (MEL) of the University Institute of Clinical Research (IURC), University of Montpellier I, was created under the responsibility of the Dean of the Faculty of Medicine and the Chief of the Departement of Endocrinology. It promotes research programs in clinical diabetes based on recent progress in molecular biology and genetics of insulin action and insulin resistance.

---

## ● Who we are

- Location and commitments
- Members

## ● Our scientific interest

---

## ● Web selected sites

## ● **Specific activities of our research on diabetes**

### Specific information

- 3D molecular structures
- List of peptides available here
- List of known insulin receptor mutations
- List of patients by their mondain name
- Antidiabetic drug - metformin

### Data analysis

**IRDB insulin receptor database in acedb language** 

---

## ● **How to send biological samples**

---

## ● Positions

## ● Publications

---

Page created by Agnès Bladier  
webmaster@spectrum.iurc.montp.inserm.fr

Les autres pages ne sont toujours pas terminées car il faut faire des modifications afin de pouvoir satisfaire tous les types de visiteurs !

## 5. CONCLUSION

Lors de cette réalisation de pages Web avec Netscape, je me suis, osons le dire, vraiment amusée. J'ai eu l'impression de pouvoir jouer avec les couleurs, de créer quelque chose tout de suite. Ça change de la réflexion qui commande toute réalisation de longue haleine.

De plus, seuls quelques problèmes sont survenus lors de la création des pages:

- il a été impossible de définir le titre de la page (<title>) directement dans l'éditeur, j'ai du passer par WordPad pour y parvenir !!!

- HTML 2 n'accepte pas les commandes <tab> pour former les tableaux. Or, la majorité des informations dites "spécifiques" sur la page d'accueil sont sous forme de tableau ! Il n'a pas non plus accepté la commande <pre> qui permet d'insérer du texte formaté sous Word par exemple. En parcourant les sites sur Internet, je suis arrivée sur le site HTML Widget<sup>13</sup>, qui disposait ses informations sous forme de tableau réalisé avec des tirets et des barres verticales (pipe | ), et je me suis inspirée de leur méthode. Dans ce cas, la commande <pre> fonctionnait !

Ces quelques problèmes m'ont agacés, certes, mais lorsque je suis parvenue à les contourner, je me suis à nouveau amusée !!!

---

<sup>13</sup> <http://www.zdv.de/www/zdv-uir/www/software/libraries/htmlw/htmlwidget.html>

# CONCLUSION

Tout d'abord, le but du stage a, il me semble, été atteint. En effet, grâce à l'utilisation des bases de données de type bibliographiques, les chercheurs ont actualisé leurs dossiers sur les différents sujets d'étude. Il aurait, toutefois été fort intéressant qu'ils évaluent la pertinence des bases de données génétiques et des journaux sélectionnés ainsi que la facilité d'utilisation du manuel. Malheureusement, faute de temps, cela n'a pas été réalisé comme prévu.

Pour ce qui est du site Web, l'installation sur le serveur du CRBM, n'a pas été poursuivie car, à ce jour, le texte en français n'est pas finalisé et comme celui-ci sera fort différent du texte existant en anglais, il faudra aussi modifier ce dernier. Comme je ne serai certainement plus présente pour la mise à jour, il est prévu que je forme Florin Grigorescu sur l'utilisation de Netscape et son éditeur.

Ensuite, pour que l'amélioration de l'utilisation d'Internet et de la collecte d'information soit complète, je vais exposer ci-dessous ce que je pense nécessaire.

Tout d'abord du point de vue matériel, il faudrait:

- "rajeunir" le système d'exploitation de la station Silicon Graphics, afin de rendre possible l'utilisation de Netscape Gold 2.0
- mettre en réseau les trois autres ordinateurs du laboratoire

Toutes ces modifications doivent être réalisées avec l'accord et l'aide de l'administrateur réseau du bâtiment.

Du point de vue recherche et de l'utilisation d'information maintenant, il faudrait:

- évaluer les listes de sites Web
- interroger Medline et OMIM vers le 15 de chaque mois, une fois la mise à jour effectuée, afin d'être toujours au point pour la bibliographie.
- acquérir un logiciel rendant les Macintosh capables d'accepter les disquettes PC (AccessPC, par exemple).
- acquérir un logiciel bibliographique pour réaliser une base de données de références (BiblioMacPcC, par exemple)

Enfin, quelques mots sur ce que m'a apporté ce stage. Jusqu'à présent, tous les stages effectués dans le domaine de l'écologie, avaient demandé un travail relativement individuel, pour celui-ci, en revanche; il m'a fallu travailler avec et pour d'autres personnes sur un sujet qui ne m'était, de plus, pas familier. Malgré les difficultés que le travail de service impose, j'ai grandement apprécié ce stage qui m'a permis de réaliser quelque chose en tenant compte des avis et exigences de chacun et en argumentant afin de parvenir à un compromis. J'espère que ce travail aura été apprécié.

# BIBLIOGRAPHIE

## RAPPORT DE STAGE

FEYFANT A. 20 ans après...de l'informatique documentaire à l'ingénierie documentaire. DESSID 95.

NOTAISE J., BARDA J., DUSANTER O. Dictionnaire du multimédia. Audiovisuel - Informatique - Télécommunication. AFNOR, 1995,- 886 p.

RODRIGUES L. Intérêt d'utiliser Internet pour les besoins documentaires et la veille technologique. Michelin. Rapport de stage DESSID 95.

CD-Rom Directory 7<sup>th</sup> edition. Finlay M., Mitchell J. eds. TFPL, London, 1992,-868p.

L'Internet Professionnel. Témoignages, expériences, conseils pratiques de la communauté enseignement et recherche (choix, installation et utilisation de logiciels sur Macintosh, PC et stations Unix). CNRS & Universités, CNRS Editions, Paris, 1995,-448p.

World Databases in Medecine. Armstrong C.J. ed. Bowker-Saur, London, 1993, -875p.

## MANUEL

1. ABOLA E.E., BERNSTEIN F.C., KOETZLE T.F. The Protein Data Bank. *in* Computational molecular biology. Source and methods for sequence analysis, (LESK A.M. ed.), 1988, Oxford University Press, Oxford.
2. ARESU O., CAMPI M.G., ROMANO P., PARODI B., MANNIELO A., THUROFF E., MOLINA F., SAGUATO F., IANNOTTA B., RONDANINA G. Molecular Probe Data Base (MPDB). *Nucleic Acids Research*, 1994, 22: 3474-3480.
3. ATSCHUL S.F., GISH W., MILLER W., MYERS E.W., LIPMAN D.J. Basic local alignment search tool. *Journal of Molecular Biology*, 1990, 215: 403-410.
4. ATTWOOD T.K., BECK M.E., BLEASY A.J., PARRY-SMITH D.J. PRINTS - A database of protein motif fingerprints. *Nucleic Acids Research*, 1994, 22: 3590-3596.
5. ATWOOD T.K., BECK M.E., BLEASY A.J., DEGTYARENKO K., PARRY SMITH D.J. Progress with the PRINTS protein fingerprints database. *Nucleic Acids Research*, 1996, 24: 182-188.
6. BAIROCH A, APWEILER R. The SWISS-PROT protein sequence data bank and it's new supplement TREMBL. *Nucleic Acids Research*, 1994, 24: 21-25.
7. BAIROCH A. The ENZYME data bank. *Nucleic Acids Research*, 1993, 21: 3155-3156.
8. BAIROCH A., BOECKMAN B. The SWISS-PROT protein sequence data bank, current status. *Nucleic Acids Research*, 1994, 22: 3578-3580.



9. **BAIROCH A., BUCHER P.** PROSITE: recent developments. *Nucleic Acids Research*, 1994, 22: 3583-3589.
10. **BAIROCH A., BUCHER P., HOFMANN K.** The PROSITE database, its status in 1995. *Nucleic Acids Research*, 1996, 24: 189-196.
11. **BENSON D., BOGUSKI M., LIPMAN D.J., OSTELL J.** GenBank. *Nucleic Acids Research*, 1994, 22: 3441-3444.
12. **BERMAN H.M., OLSON W.K., BEVERIDGE D.L., WESTBROOK J., GELBIN A., DEMENY T., HSIEH S.H., SRINIVASAN A.R., SCHNEIDER. M.** The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophysical Journal*, 1992, 63: 751-759.
13. **BLEASBY A.J., AKRIGG D., ATTWOOD T.K.** OWL--a non redundant composite protein sequence database. *Nucleic Acids Research*, 1994, 22: 3574-3577.
14. **BOGUSKI M., SCHULER G.** ESTablishing a Human Transcript Map. *Nature Genetics*, v, 10: 369-371.
15. **BOGUSKI M.S.** Bioinformatics. *Current Opinion in Genetics and Developments*, 1994, 4: 383-388.
16. **BOGUSKI M.S., LOWE T.M.J., TOLSTOSHEV C.M.** dbEST - a database for "expressed sequence tags". *Nature Genetics*, 1993, 4: 332-333.
17. **COLLIER B., DANIELSEN M.** GRBase, a database linking information on proteins involved in gene regulation. *Nucleic Acids Research*, 1996, 24: 219-220.
18. **DOLZ R., MOSSE M.O., SLONIMSKI P.P., BAIROCH A., LINDER P.** LISTA, LISTA-HOP and LISTA-HON: a comprehensive compilation of protein encoding sequences and it's associated homology databases from the yeast *Saccharomyces*. *Nucleic Acids Research*, 1996, 24: 50-52.
19. **EMMERT B., STOEHR P.J., STOESSERT G., CAMERON G.N.** The European Bioinformatics Institute (EBI). *Nucleic Acids Research*, 1994, 22: 3445-3449.
20. **ETZGOLD T., ARGOS P.** SRS, an indexing and retrieval tool for flat file data libraries. *Computer Applications in the Biosciences*, 1993, 9: 49-57.
21. **FACCHIANO A., FACCHIANO A., FACCHIANO F., RAGONE R., COLONNA G.** FAST (Flexible Analysis by Software Tool) and CHAMP (CHemico-physical AMinoacidic Parameter data bank): a new tool to investigate protein structure. *Computer Applications in the Biosciences*, 1989, 5: 299-303.
22. **FASMAN K.H., LETOVSKY S.I., COTTINGHAM R.W., KINGSBURY D.T.** Improvements to the GDB Human Genome Data Base. *Nucleic Acids Research*, 1996 ,24: 57-63.
23. **FLYBASE CONSORTIUM** FlyBase: the *Drosophila* database.The Flybase consortium. *Nucleic Acids Research*, 1996, 24: 53-56.
24. **GEORGE D.G., BARKER W.C., HUNT L.T.** The protein identification resource (PIR). *Nucleic Acids Research*, 1986, 14: 11-15.
25. **GEORGE D.G., BARKER W.C., MEWES H.W., PFEIFFER F., TSUGITA A.** The PIR-International Protein Sequence Database. *Nucleic Acids Research*, 24: 17-20.
26. **GOVERNET J., CARABOEUF M., AYME S.** (1985) GENDIAG: a computer assisted facility in medical genetics based on belief functions. *Methods in informatics and medecine* , 1996 ,24: 177-180.
27. **HENIKOFF S., HENIKOFF J.G.** Automated assembly of protein blocks for database searching. *Nucleic Acids Research*, 1991, 19: 6565-6572.
28. **HOLM L., SANDER C.** Protein structure comparison by alignment of distance matrices. *Journal of Molecular Biology*, 1993 ,233: 123-138.

29. HOLM L., SANDER C. Dali: a network tool for protein structure comparison. *Trends in Biochemical Sciences*, 1995, 20: 478-490.
30. HOLM L., SANDER C. The FSSP database: fold classification based on structure-structure alignment of proteins. *Nucleic Acids Research*, 1996, 24: 206-209.
31. KARP P., PALEY S. Representation of metabolic knowledge: Pathways. in Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, (ALTAM R, BRUTLAG D, KARP P., LATHROP R., SEARLS D eds.), 1994, AAAI Press, Menlo Park, CA
32. KARP P., RILEY M. Representation of metabolic knowledge. in Proceedings of the First International Conference on Intelligent Systems for Molecular Biology, (HUNTER L., SEARLS D., SHAVLIK J. eds.), 1993, pp 207-215, AAAI Press, Menlo Park, CA
33. KARP P.D., RILEY M., PALEY S.M., PELLIGRINI-TOOLE A. EcoCyc: an encyclopedia of Escherichia coli genes and metabolism. *Nucleic Acids Research*, 1996, 24: 32-39.
34. KEEN G., BURTON J., CROWLEY D., DICKINSON E., ESPINOSA-LUJAN A., FRANKS E., HARGER C., MANNING M., MARCH S., MCLEOD M., *et al.* The Genome Sequence Database (GSDB): meeting the challenge of genomic sequencing. *Nucleic Acids Research*, 1996, 24: 13-16.
35. KROGER M., WAHL R. Compilation of DNA sequences of Escherichia coli K12 (ECD and ECDC; update 1995). *Nucleic Acids Research*, 1996, 24: 29-31.
36. LARSEN F., GUNDERSEN G., LOPEZ L., PRYDZ H. CpG island as gene markers. *Human Genome Genomics*, 1992, 13: 1095-1107.
37. MCKUSICH V.A. Catalogs of autosomal dominant, autosomal recessive, and X-linked phenotypes, Tenth Edition. 1991, Baltimore.
38. OLSON M., HOOD L., CANTOR C., BOTSTEIN D. A common language for physical mapping of the human genome. *Science*, 1989, 254: 1434-1435.
39. PATTABIRAMAN N., NAMBOODIRI K., LOWREY A., GABER B.P.
40. NRL-3D: a sequence-structure database derived from the protein data bank (PDB) and searchable within the PIR environment. *Protein Sequence Data Analysis*, 1990, 3: 387-405.
41. PEARSON P.L. The genome data base (GDB)--a human gene mapping repository. *Nucleic Acids Research*, 1991, 19: 2237-2239.
42. PERRIER G., MOSZER I., GOJOBORI T. NRSub: a non-redundant data base for Bacillus subtilis. *Nucleic Acids Research*, 1996, 24: 41-45.
43. PIETROKOVSKI S., HENIKOFF J.G., HENIKOFF S. The BLOCKS database--a system for protein classification. *Nucleic Acids Research*, 1996, 24: 197-200.
44. PONGOR S., SKERL V., CSERZO M., HATSAGI Z., SIMON G., BEVILACQUA V. The SBASE protein domain library, release 2.0: a collection of annotated protein sequence segments. *Nucleic Acids Research*, 1993, 21: 3111-3115.
45. ROBERTS R.J., MACELLIS D. REBASE--restriction enzymes and methylases. *Nucleic Acids Research*, 1996, 24: 223-235.
46. RODRIGUEZ-TOME P., STOEHR P.J., CAMERON G.N., FLORES T.P. The European Bioinformatics Institute (EBI) database. *Nucleic Acids Research*, 1996, 24: 189-196.
47. SATOU K., FURUICHI E., HASHIMOTO S., TSUKAMOTO Y., KUHARA S., TAYAGI T., USHJIMA K. Development of a deductive database system for computing closures of similarity relationships among protein structures. *Journal of Japanese Society of Artificial Intelligence*, 1996, 11: 440-450.

48. SCHNEIDER R., SANDER C. The HSSP database of protein structure-sequence alignments. *Nucleic Acids Research*, 1994, 22: 3597-3599.
49. SCHNEIDER R., SANDER C. The HSSP database of protein structure sequence alignment. *Nucleic Acids Research*, 1996, 24: 201-205.
50. SCHULER G.D., EPSTEIN J.A., OHKAWA H., KANS J.A. Entrez: molecular biology database and retrieval system. *Methods in Enzymology*, 1996, 266: 141-162.
51. SHIN'I T., GOJOBORI T. DDBJ database and genetic information analysis software. *Tanpakushitsu Kakusan Koso*, 1994, 39: 1927-1943.
52. SIDMAN K.E., GEORGE D.G., BARKER WC, HUNT L.T. The protein identification resource (PIR). *Nucleic Acids Research*, 1988, 16: 1869-1871.
53. SONNHAMMER E.L.L., KAHN D. Modular arrangement for proteins as inferred from analysis of homology. *Protein Science*, 1994, 3: 482-492.
54. SUZUKI T., TAGAKI T., KUHARA S., KANEHISA M. Development of an integrated database for genome mapping and nucleotide sequence. in *Proceedings of the 27th Annual Hawaii International Conference on System Sciences*, 1994, pp 68-76,
55. WAHL R., RICE P., RICE C.M., KROEGER M. ECD - a totally integrated database of Echerichia coli K12. *Nucleic Acids Research*, 1994, 22: 3450-3455.
56. WINSHIP I.R. World Wide Web searching tools - an evaluation. *VINE*, 1995, 99: 49-54.
57. WOODSMALL R.M., BENSON Information resources at the National Center for Biotechnology Information. *Bulletin of the Medical Library Association*, 1993, 81: 282-284.
58. WOYCHIK R.P., WASSON J.S., KINGSBURY D. TBASE: a computerized database for transgenic animals and targeted mutations. *Nature*, 1993, 363: 375-376.

# **ANNEXE**

Liste des sites sélectionnés

p 58

## MOTEURS ET CATALOGUE

**AltaVista:** <http://www.altavista.digital.com/>

**Infoseek:** <http://www2.infoseek.com/>

**Lycos:** <http://lycos.cs.cmu.edu/>

**WebCrawler:** <http://www.webcrawler.com/>

**Yahoo!:** <http://www.yahoo.com/>

## SEQUENCES

**GENBANK:** [http://www2.ncbi.nlm.nih.gov/genbank/query\\_form.html](http://www2.ncbi.nlm.nih.gov/genbank/query_form.html)

**GSDB - Genome Sequence Database:** <http://www.ncgr.org/cgi-bin/cansql/keyword>

**GDB - Genome Database:** <http://gdbwww.gdb.org/wais/gdb-wais.html>

**HGMD - Human Genome Mutation Database:** <http://www.cf.ac.uk/search/mg/allgenes?>

**DDBJ - DNA Data Bank of Japan:**

<gopher://ftp2.ddbj.nig.ac.jp/7waisrc%3a/search/ddbj>

**GENOTK - Human cDNA Database:** <http://genotk.genome.ad.jp:8010/>

**NDB - Nucleic Acid Database:** <http://ndbserver.ebi.ac.uk:5700/NDB/interface/index.html>

**TBASE - Transgenic/Targeted Mutation Database:** <http://www.gdb.org/Dan/tbase/tb.html>

**BLAST :** <http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nph-blast?Jform=0>

## PROTEINES

**DALI - 3-D protein structure database:** <http://www.embl-heidelberg.de/dali/dali.html>

**MOOSE - Macromolecular Structure Query:** <http://db2.sdsc.edu/moose/>

**OWLREF:** <http://www.gdb.org/bio/search/FILT/owlref.html>

**PDB - Protein Data Bank:** <http://pdb.pdb.bnl.gov/cgi-bin/browse>

**PIR - The Protein Information Resource:** <http://www.gdb.org/Dan/fields/pir.form.html>

**SWISS-PROT:** <http://expasy.hcuge.ch/sprot/sprot-top.html>

**Protein Kinase Database:** [http://www.sdsc.edu/CompSci/Biomed/Kinases/kinase\\_intro.html](http://www.sdsc.edu/CompSci/Biomed/Kinases/kinase_intro.html)

## AUTRES

**Medline:** [http://www2.ncbi.nlm.nih.gov/medline/query\\_form.html](http://www2.ncbi.nlm.nih.gov/medline/query_form.html)

**OMIM - Online Mendelian Inheritance in Man:**

<http://www3.ncbi.nlm.nih.gov/Omim/searchomim.html>

**GENDIAG:** <http://www.infobiogen.fr/gendiag/>

**Vertebrate Comparative Database:** <http://www.hgmp.mrc.ac.uk/Comparative/home.html>

**SRS (Sanger):** <http://www.sanger.ac.uk/srs/srsc>

**SRS (Heidelberg):** <http://www.embl-heidelberg.de/srs/srsc>

**Entrez:** <http://www3.ncbi.nlm.nih.gov/Entrez/>

## CENTRES

**INFOBIOGEN:** <http://www.infobiogen.fr/>

**CRBM - Centre de Recherche en Biochimie Macromoléculaire:**

<http://www.crbm.cnrs-mop.fr/home.html>

**EMBL - European Molecular Biology Laboratory:** <http://www.embl-heidelberg.de/>

**NCBI - National Center for Biotechnology Information:** <http://www.ncbi.nlm.nih.gov/>

**NIH - National Institutes of Health:** <http://www.nih.gov/>

**National Institute of Genetics (Japan):** <http://www.nig.ac.jp/>

**International Nucleotide Sequence Database Collaboration:**

<http://www.ncbi.nlm.nih.gov/collab/>

## JOURNAUX

**Springer-Verlag Journals:** <http://www.springer.de/server/listjournals.html>

**Elsevier Science:** <http://www.elsevier.nl/>

**UnCover:** <http://www.carl.org/uncover/>

**MedWeb:** <http://www.cc.emory.edu/WHSC/medweb.ejs.html>

**Acta Diabetologia:** <gopher://trick.ntp.springer.de/11/TOC/592/>

**AJP : Endocrinology and Metabolism:** <gopher://oac.hsc.uth.tmc.edu:3300/11/publications/em>

**AJP: Cell physiology:** <gopher://oac.hsc.uth.tmc.edu:3300/11/publications/cp>

**American Scientist:** <http://www.amsci.org/amsci/amsci.html>

**Annual Review of Cell and Developmental Biology:**

<http://www.annurev.org/series/cell/cell.htm>

**Annual Review of Biochemistry:** <http://www.annurev.org/series/biochem/biochem.htm>

**Annual Review of Genetics:** <http://www.annurev.org/series/genetics/genetics.htm>

**Applied Microbiology and Biotechnology:** <gopher://trick.ntp.springer.de/11/TOC/253/>

**BBA-Biochimica et Biophysica Acta/Reviews on Cancer:**

<http://www.elsevier.nl/cas/estoc/contents/SAB/0304419X.html>

**Bio/Technology:** [http://www.eneews.com/magazines/bio\\_tech/](http://www.eneews.com/magazines/bio_tech/)

**Biotechnology and Development Monitor:** <http://www.bdt.org.br/bioline/mo>

**Biotechnology Bulletin:**

[http://www.sebiotech.org.uk/cgi-bin/cdbi/biotech.html?page?wq=&h\\_page=biotech/cover](http://www.sebiotech.org.uk/cgi-bin/cdbi/biotech.html?page?wq=&h_page=biotech/cover)

**Biotechnology Letters:** <http://www.bdt.org.br/bioline/bl>

**Biotechnology Techniques:** <http://www.bdt.org.br/bioline/bt>

**British Medical Journal (BMJ):** <http://www.tecc.co.uk/bmj/archive/curr.htm>

**Cancer Genetics and Cytogenetics:**

<http://www.elsevier.nl/cas/estoc/contents/SAB/01654608.html>

**Cell:** <http://www.cell.com/cell/>

**Center for Genome Research newsletter:**

[http://www-genome.wi.mit.edu/genome\\_data/newsletters/newsletter.0894.html](http://www-genome.wi.mit.edu/genome_data/newsletters/newsletter.0894.html)

**Cliniweb:** <http://www.ohsu.edu/clinweb/>

**Computer Methods and Programs in Biomedecine:**

<http://www.elsevier.nl/cas/estoc/contents/SAB/01692607.html>

**Computer Methods and Programs in Biomedicine:**

<http://www.elsevier.nl/cas/estoc/contents/SAB/01692607.html>

**Computerized Medical Imaging and Graphics:**

<http://www.elsevier.nl/cas/estoc/contents/SAB/08956111.html>

**Computers in Biology and Medicine:**

<http://www.elsevier.nl/cas/estoc/contents/SAB/00104825.html>

**Current Genetics:** [gopher://trick.ntp.springer.de/11/TOC/294/](http://trick.ntp.springer.de/11/TOC/294/)

**Current Microbiology:** [gopher://trick.ntp.springer.de/11/TOC/284/](http://trick.ntp.springer.de/11/TOC/284/)

**Current Opinion in Biotechnology:** <http://biomednet.com/cbiology/bio.htm>

**Diabetes Research and Clinical Practice**

<http://www.elsevier.nl/catalogue/SAH/190/05300/04102/505949/505949.html>

**Diabetes Research and Clinical Practice (Special Issues)**

<http://www.elsevier.nl/catalogue/SAH/190/05300/04102/522280/522280.html>

**Diabetes Self-Management:** <http://www.eneews.com/magazines/diabetes/>

**Diabetologia:** [gopher://trick.ntp.springer.de/11/TOC/125/](http://trick.ntp.springer.de/11/TOC/125/)

**Differentiation:** [gopher://trick.ntp.springer.de/11/TOC/258/](http://trick.ntp.springer.de/11/TOC/258/)

**Embnet.news - The European Molecular Biology Network Newsletter:**

<http://www.ch.embnet.org/embnet.news/info.html>

**EMBO Journal:** <http://www.oup.co.uk/jnls/list/embojo/>

**E-Med News:** <http://www.pjbpubs.co.uk/a/emedhome.html>

**European Journal of Biochemistry:** [gopher://trick.ntp.springer.de/11/TOC/225/](http://trick.ntp.springer.de/11/TOC/225/)

**European Journal of Clinical Chemistry and Clinical Biochemistry:**

<http://www.degruyter.de/journals/ejcccb.html>

**European Journal of Clinical Pharmacology:** [gopher://trick.ntp.springer.de/11/TOC/228/](http://trick.ntp.springer.de/11/TOC/228/)

**European Journal of Pediatrics:** [gopher://trick.ntp.springer.de/11/TOC/431/](http://trick.ntp.springer.de/11/TOC/431/)

**Experimental and Clinical Psychopharmacology:** <http://www.apa.org/journals/pha.html>

**FASEB Journal:** <http://www.faseb.org/fj/index.html>

**FASEB Newsletter:** [gopher://gopher.faseb.org/11/Journal](http://gopher.faseb.org/11/Journal)

**FOCUS:** <http://www.med.harvard.edu/publications/Focus/.index.html>

**Gene Therapy Weekly:** <http://www.newsfile.com/1g.htm>

**GENE:** <http://www.elsevier.nl/cas/estoc/contents/SAH/03781119.html>

**GENE-COMBIS:** <http://www.elsevier.nl:80/journals/genecombis/Menu.html>

**Genes to Cells:** <http://www.blacksci.co.uk/products/journals/gtc.htm>

**Genetic Engineer and Biotechnologist:** <http://www.bdt.org.br/bioline/ge>

**Genetic Engineering News:** [http://www.eneews.com/magazines/geng\\_news/](http://www.eneews.com/magazines/geng_news/)

**Genetics:** <http://www.faseb.org/genetics/gsa/jou-menu.htm>

**Gynecological Endocrinology:** <http://www.parthpub.com/gynend/home.html>

**Human Genetics:** [gopher://trick.ntp.springer.de/11/TOC/439/](http://trick.ntp.springer.de/11/TOC/439/)

**Human Genome News:**

[http://www.ornl.gov/TechResources/Human\\_Genome/publicat/publications.html](http://www.ornl.gov/TechResources/Human_Genome/publicat/publications.html)

**Human Genome Newsletter:** [gopher://merlot.gdb.org/11/Genome/hgnews](http://merlot.gdb.org/11/Genome/hgnews)

**Human Molecular Genetics:**

<http://www.informatik.uni-rostock.de/HUM-MOLGEN/journals/HMG/>

**Human Reproduction:** <http://www.oup.co.uk/jnls/list/humrep/hdb/>

**Immunogenetics:** [gopher://trick.ntp.springer.de/11/TOC/251/](http://trick.ntp.springer.de/11/TOC/251/)

**International Journal of Biomedical Computing:**

<http://www.elsevier.nl/cas/estoc/contents/SAB/00207101.html>

**International Journal of Experimental Pathology :**

<http://www.ccc.nottingham.ac.uk/~mpzjlowe/tocs/ijepath.html>

**International Journal of Gynecology and Obstetrics:**

<http://www.elsevier.nl/cas/estoc/contents/SAB/00207292.html>

**Journal of the American Medical Association (JAMA):**

<http://www.ama-assn.org/sci-pubs/journals/standing/jama/jamahome.htm>

**Journal of the American Medical Informatics Association (JAMIA):**

<http://amia2.amia.org/default.html>

**Journal of Applied Physiology:** [gopher://oac.hsc.uth.tmc.edu:3300/11/publications/jap](http://gopher://oac.hsc.uth.tmc.edu:3300/11/publications/jap)

**Journal of Biblical Ethics in Medicine:** <http://www.usit.net/public/CAPO/jbem.html>

**Journal of Biochemistry:** <http://www.bcasj.or.jp/jbindex.html>

**Journal of Biological Chemistry:** <http://www-jbc.stanford.edu/jbc/>

**Journal of Clinical Investigation:** <http://www.rockefeller.edu/RUPress/JCI/JCI.html>

**Journal of Clinical Pathology:** <http://www.bmj.com/bmj/bmjpubs/jcp.htm>

**Journal of Computational Biology:** <http://brut.gdb.org/compbio/jcb/>

**Journal of Computer-Aided Molecular Design:** <http://wucmd.wustl.edu/jcamd/jcamd.html>

**Journal of Diabetes and its Complications**

<http://www.elsevier.nl/catalogue/SAH/190/05300/04102/505770/505770.html>

**Journal of Ethics:** <http://condor.depaul.edu/ethics/ethg16.html>

**Journal of Genetic Counseling:**

<http://www.informatik.uni-rostock.de/HUM-MOLGEN/journals/JGC/>

**Journal of Heredity:** <http://www.oup.co.uk/jnls/list/jhered/>

**Journal of Molecular Biology (JMB):** <http://www.hbuk.co.uk/jmb/>

**Journal of Molecular Evolution:** [gopher://trick.ntp.springer.de/11/TOC/239/](http://gopher://trick.ntp.springer.de/11/TOC/239/)

**Journal of Molecular Medicine:** [gopher://trick.ntp.springer.de/11/TOC/109/](http://gopher://trick.ntp.springer.de/11/TOC/109/)

**Journal of Molecular Modeling:** [gopher://trick.ntp.springer.de/11/TOC/894/](http://gopher://trick.ntp.springer.de/11/TOC/894/)

**Journal of NIH Research:** <http://www.eneews.com/magazines/nih/>

**Journal of Nutrition:** <http://www.edoc.com/nutrition/journal/journal.html>

**Journal of Physiology:** <http://physiology.cup.cam.ac.uk/Publications/jphymenu.html>

**Lyon Pharmaceutique:** <http://www.elsevier.nl/cas/estoc/contents/SAB/00247804.html>

**Mammalian Genome:** [gopher://trick.ntp.springer.de/11/TOC/335/](http://gopher://trick.ntp.springer.de/11/TOC/335/)

**Mathematical and Computer Modelling:**

<http://www.elsevier.nl/cas/estoc/contents/SAB/08957177.html>

**Mayo Clinic Proceedings:** <http://www.mayo.edu/publication/proceedings/prologue.html>

**Médecine Clinique et Expérimentale:**

[http://www.hwc.ca:8080/cma/journals/cim/index\\_f.html](http://www.hwc.ca:8080/cma/journals/cim/index_f.html)

**Médecine et Droit:** <http://www.elsevier.nl/cas/estoc/contents/SAB/12467391.html>

**Medical Microbiology and Immunology:** [gopher://trick.ntp.springer.de/11/TOC/430/](http://gopher://trick.ntp.springer.de/11/TOC/430/)

**Molecular Aspects of Medicine:**

<http://www.elsevier.nl/cas/estoc/contents/SAB/00982997.html>

**Molecular and Biochemical Parasitology**

<http://www.elsevier.nl/cas/estoc/contents/SAH/01666851.html>

**Molecular and Cellular Biology:** <http://www.asmtusa.org/jnls/mcb1.htm>

**Molecular and Cellular Endocrinology**

<http://www.elsevier.nl/catalogue/SAH/190/05300/04100/506028/506028.html>

**Molecular and General Genetics (MGG):** [gopher://trick.ntp.springer.de/11/TOC/438/](http://gopher://trick.ntp.springer.de/11/TOC/438/)

**Molecular Biology of the Cell:** [gopher://gopher.faseb.org/11/Societies/ASCB/Journal](http://gopher://gopher.faseb.org/11/Societies/ASCB/Journal)

**Molecular Pharmacology:** [http://www.wwilkins.com/molec\\_pharm/](http://www.wwilkins.com/molec_pharm/)

**Mouse Genome:** <http://www.oup.co.uk/jnls/list/mousej/>

**Mutagenesis:** <http://www.oup.co.uk/jnls/list/mutage/>

**Nature:** <http://www.nature.com/>

**Nature Genetics:** <http://genetics.nature.com/content>

**Nature Medicine:** <http://medicine.nature.com/>



**Network Science:** <http://edisto.awod.com/netsci/archives.html>  
**New England Journal of Medicine:** <http://www.nejm.org/JHome.htm>  
**Nucleic Acids Research:** <http://www.oup.co.uk/oup/smj/journals/ed/titles/nar/>  
**Obstetrics and Gynecology:** <http://www.elsevier.nl/cas/estoc/contents/SAB/00297844.html>  
**OMNI:** <http://omni.ac.uk/>  
**PCR Methods & Applications:** <http://www.sheridan.com:443/cshl/journals/pcr/>  
**Pediatric Cardiology:** [gopher://trick.ntp.springer.de/11/TOC/246/](http://trick.ntp.springer.de/11/TOC/246/)  
**Pediatric Nephrology:** [gopher://trick.ntp.springer.de/11/TOC/467/](http://trick.ntp.springer.de/11/TOC/467/)  
**Pharmacological Reviews:** <http://www.wwilkins.com/PTS/>  
**Pharmacovigilance et Interactions Médicamenteuses:** <http://www.upml.fr/pharmaco.htm>  
**Physiological Reviews:** [gopher://oac.hsc.uh.tmc.edu:3300/11/publications/pr](http://oac.hsc.uh.tmc.edu:3300/11/publications/pr)  
**Proceedings of the National Academy of Sciences:** <http://journals.at-home.com/PNAS/>  
**Protein Engineering:** <http://www.oup.co.uk/jnls/list/proeng/hdb/>  
**Protein Science:** <http://www.prosci.uci.edu/homepages/>  
**Revue de Médecine Interne:** <http://www.elsevier.nl/cas/estoc/contents/SAB/02488663.html>  
**Science:** <http://science-mag.aaas.org/science/>  
**Scientific American:** <http://www.sciam.com/WEB/index.html>  
**Scientific Computing & Automation:** <http://www.scam.ag.com/>  
  
**Scientific Computing World:** <http://www.ioppublishing.com/Mags/SCW/index.html>  
**Seminars in Cell Biology:** [gopher://ukoln.bath.ac.uk:7070/11/Link/Tree/Publishing/AcademicPress/APJournals/APSemCel](http://ukoln.bath.ac.uk:7070/11/Link/Tree/Publishing/AcademicPress/APJournals/APSemCel)  
**Transgenica:** [http://pharminfo.com/pubs/transgen/tg\\_hp.html](http://pharminfo.com/pubs/transgen/tg_hp.html)  
**Trends in Biochemical Sciences:** <http://www.elsevier.nl/cas/estoc/contents/SAH/09680004.html>  
**Trends in Biotechnology:** <http://www.elsevier.nl/cas/estoc/contents/SAH/01677799.html>  
**Trends in Cell Biology** <http://www.elsevier.nl:80/catalogue/SAH/185/05050/05050/422552/422552.html>  
**Trends in Endocrinology and Metabolism** <http://www.elsevier.nl/catalogue/SAH/190/05300/04100/505783/505783.html>  
**Trends in Genetics** <http://www.elsevier.nl:80/catalogue/SAH/185/05450/05450/405918/405918.html>  
**Trends in Microbiology** <http://www.elsevier.nl:80/catalogue/SAH/195/05100/05100/424100/424100.html>  
**Trends in Pharmacological Sciences:** <http://www.elsevier.nl:80/catalogue/SAH/150/05500/05500/405920/405920.html>  
**Vidal du particulier:** <http://www.vidal.fr/>  
**WWW Journal of Biology:** <http://epress.com/w3jbio/>



## DIVERS

**Alström Syndrome:** [http://www.familyvillage.wisc.edu/lib\\_alst.htm](http://www.familyvillage.wisc.edu/lib_alst.htm)  
**Comité Ethique Français:** <http://www.infobiogen.fr/SDV/cope.html>

BIBLIOTHEQUE DE L'ENSSIB



810792B