

Arch. INTERN carton 0710-21

# MRT

MINISTÈRE  
DE LA RECHERCHE  
ET DE LA TECHNOLOGIE

→ JG  
VAL

Département "Biologie, Médecine et Santé"  
Réf. CA.MB/420.89

PARIS, 1e - 5 JUIL 1989

**INSERM**  
DIRECTION GÉNÉRALE  
10 JUIL. 1989  
COURRIER ARRIVÉE

2284.

Monsieur Philippe LAZAR  
Directeur Général  
I.N.S.E.R.M.  
101, rue de Tolbiac  
75654 PARIS CEDEX 13

VALES  
Courrier arrive le  
12 JUIL. 1989

Monsieur le Directeur Général,

Je vous adresse, ci-joint, un document de travail, très informel, établi par M. Daniel COHEN au sujet du C.E.P.H.

La création d'un GIP Génomes y est envisagée, avec entre autres partenaires l'INSERM et le CNRS.

J'aimerais avoir votre opinion sur ce point précis.

Merci d'avance de vos remarques.

Bien à vous.

FB - PHL

- Comment le GIP structure versus laborap (industriel = Amersham)?
- qui finance 50 000 MF/an
- quels industriels? (Biocom ou autres bien sûr)
- Où est le projet scientifique? (à faire évaluer par Ph. Kourbaj)

(le accord 17/1)

Claude Amiel

P.J. 1

Juin 1989

79 06 89  
-----  
Daniel Cohen

## LE PROJET " GENOMES "

Daniel COHEN

C.E.P.H.

### I - GENERALITES

- a) la carte génétique
- b) la carte physique
- c) l'inventaire des gènes

### II - PROPOSITIONS GENERALES

- a) état des idées
- b) état des actions
- c) position française
- d) propositions

### III - LE COUT DU PROJET

## I - GENERALITES

Le génome humain est composé de 24 segments d'ADN constituant chacun un chromosome (visible pendant la mitose) le plus petit chromosome mesure 50 millions de bases, le plus grand 450 millions de bases, le génome totalisant 3.5 milliards de bases .

a) La CARTE GENETIQUE (genetic map = linkage map) d'un génome consiste aujourd'hui à positionner les uns par rapport aux autres des locus polymorphes détectés par des sondes spécifiques . Ces locus sont généralement détectés à l'aide de ces sondes par la technique de Southern Blot . Le polymorphisme peut être dû à une mutation ponctuelle ou l'insertion en nombre variable de séquences répétées de " mini ou microsattellites. La distance entre les locus polymorphes est mesurée par des études familiales multiples en évaluant le nombre de crossing over pouvant intervenir entre ces 2 locus (plus les locus sont proches moins il y a de possibilités de crossing over).

Le CEPH organise l'élaboration de la carte génétique du génome humain en mettant à la disposition de la communauté scientifique l'ADN de familles nombreuses de référence immortalisées sous forme de lignées EBV . Les résultats sont accumulés sur les ordinateurs du CEPH qui organise également l'analyse des données et qui publie régulièrement les cartes mises à jour. L'objectif est d'obtenir dans 5 ans une carte avec des locus polymorphes chaque centimorgan (1 centimorgan = environ 1 million de bases = environ 1% recombinaisons).

En Mai 1989 la carte génétique comporte un marqueur tous les 20 centimorgans en moyenne : grâce au panel de sondes jalonnant ainsi tout le génome, on peut positionner le gène de n'importe quelle maladie génétique . Il suffit de cribler des familles de malades à l'aide de ce panel de sondes et d'analyser le nombre de crossing over (donc la distance génétique ) entre les locus définis par ces sondes et le locus du gène de la maladie .

b) Constituer la CARTE PHYSIQUE de tout ou partie d'un génome consiste à positionner les uns par rapport aux autres des fragments d'ADN chevauchant clonés à partir de ce génome , l'ensemble de ces fragments recouvrant tout ou partie de ce génome .

Les fragments clonés sont généralement insérés dans des cosmides (50 kb); depuis peu on peut cloner 300 kb voire 1 million de bases dans des chromosomes artificiels de levure ce qui pourrait accélérer considérablement le processus . On estime qu'il faudra 5 à 10 ans pour établir la carte physique complète du génome humain .

La cartographie physique est indispensable à l'identification des

gènes des maladies . La cartographie génétique ne permet, en effet que de localiser ces gènes entre deux locus polymorphes . La carte physique permettra, lorsqu'elle sera achevée, de disposer automatiquement de l'ADN cloné de toute la région comprise entre ces deux locus polymorphes . Il ne reste plus qu'à faire l'inventaire des gènes entre ces marqueurs en utilisant ces clones .

c) Faire l'INVENTAIRE DES GENES, tel est le but essentiel du Projet "Genomes". Cet inventaire constituera un outil formidable pour une biologie moderne. Il est évident que la démarche biologique s'effectuera beaucoup plus efficacement dès lors que cet inventaire sera disponible . Faire l'inventaire repose sur un ensemble de méthodes dont le séquençage qui deviendra certainement un outil privilégié . Les informaticiens développent en effet de nouvelles méthodes d'analyse permettant de distinguer dans une séquence inconnue ce qui est codant de ce qui ne l'est pas .

Il est clair par ailleurs que plus l'on connaîtra la fonction de gènes nouveaux, plus, par comparaison, on sera capable de déceler des motifs "géniques" d'où l'absolue nécessité de ne pas freiner la recherche classique au profit du Projet "Genomes" . A l'inverse ne pas participer au projet mettrait notre recherche biologique en danger car même si les résultats produits ailleurs seront communiqués, ils le seront d'autant plus tardivement que nous n'aurons rien à donner en échange .

Il faut donc impérativement être prêt à contribuer activement à l'effort international , finalement à la fois collaboratif et compétitif .

## II - PROPOSITION GENERALE POUR LA MISE EN PLACE D'UN PROJET "GENOMES" EN FRANCE

Le projet "Génomes" fait l'objet de discussions scientifiques, éthiques et politiques depuis près de 3 années. Cette initiative n'est pas uniquement considérée comme une démarche considérable pour la connaissance mais comme une source de bienfaits pour l'homme, notamment dans le domaine de la santé, de l'agro-alimentaire et de l'environnement.

### a) état des idées :

Le projet initial était d'établir la séquence complète du génome humain. Les idées ont depuis évolué et le projet s'est progressivement rationalisé.

- 1) Un séquençage systématique du génome humain n'est pas réalisable avant au moins cinq ans. Il semble qu'en tout état de cause :

. qu'il faille d'abord établir les cartes dites génétiques et physiques du génome sans lesquelles son séquençage est



- . Marché de la santé
- . Marché agro-alimentaire
- . Problèmes liés à l'environnement

Aussi, il est temps de définir la position française sur ce sujet.

c) position française :

1) Les atouts français :

1° Un projet EUREKA de développement technologique pour des automates spécialisés, débuté en 1987 ; il s'agit d'une collaboration entre les Sociétés BERTIN (France) et AMERSHAM (Grande Bretagne) et les deux institutions de recherches CEPH et ICRF (Grande Bretagne).

Ce projet de 400 MF sur 4 ans est le seul de cette ampleur en Europe.

Les compétences de l'équipe constituée sont équivalentes à celles que l'on peut trouver aux USA .

2° Une bio-informatique naissante non encore reconnue nationalement mais plus développée que dans les autres pays d'Europe ou aux USA .

3° Une position du meilleur niveau pour l'analyse du Génome de souris .

4° L'aide logistique considérable apportée par l'AFM pour l'organisation du prélèvement des familles de malades indispensable à la recherche des gènes des maladies génétiques.

5° Une expérience quasi unique de l'activité de services dans ce domaine acquise depuis 5 ans au CEPH .

2) Les points faibles français : (peuvent être exploités ou supprimés)

1° Proportionnellement moins d'équipes sérieusement impliquées que USA ou Grande Bretagne . IL serait peut-être erroné d'essayer d'en recruter davantage dans un premier temps. Il vaudrait mieux financer correctement un petit nombre d'équipes plutôt que de saupoudrer un grand nombre (c'est ce qui va se passer en Angleterre). Par ailleurs un petit nombre d'équipes travaillant sur des régions génomiques différentes collaborerait très facilement dans le domaine technologique.

2° Pas de reconnaissance du projet par les Commissions des Organismes : considéré comme trop technologique .

d) Propositions :

Nous proposons de mettre en place une structure forte et souple qui permettra d'atteindre un certain nombre d'objectifs bien définis. Cette proposition, qui est le fruit de réflexions et de discussions variées au cours des 12 derniers mois, a plusieurs caractères :

- . Elle tient compte au maximum des structures et compétences existantes ;
- . Elle tient compte des intérêts manifestés par les scientifiques français ;
- . Elle tient compte des capacités qu'elle accentue progressivement ;
- . Elle permet à la France d'avoir une position originale et compétitive ;
- . Elle est compatible avec les projets internationaux.

Il s'agit d'un groupement d'intérêt public (GIP) associant :

du côté Secteur Public : CEA, CNRS, INRA, INRIA, INSERM,

du côté Secteur Privé :

pour la Recherche : AFM, CEPH, Institut PASTEUR,

pour l'Industrie : BERTIN, BIOCOM, autres industriels ?

Ce GIP dont le titre serait "GENOMES" aurait comme caractéristique de faire progresser de façon synergique deux pôles d'activités :

... le développement technologique pour la cartographie et la séquence des génomes d'une part,

. l'inventaire génique de tout ou partie de certains génomes d'autre part .

Premier pôle d'activité : DEVELOPPEMENT TECHNOLOGIQUE:

Du développement technologique dépendra irréversiblement la position de la France dans ce projet et ses enjeux économiques . En effet la plupart des opérations de cartographie et de séquençage s'appuient sur des tâches répétitives faisant manipuler des milliers, voire des millions, d'échantillons ce qui ralentit considérablement le projet . La réalisation d'automates spécialisés est donc indispensable . Il s'avère que la Société BERTIN et le CEPH collaborent à cet effet dans le cadre d'un projet EUREKA franco-britannique nommé LABIMAP 2001.

La constitution de ce GIP renforce le projet EUREKA en :

- . associant les compétences de l'INRIA, le CEA et la Société BIOCOM
- . intégrant l'INRA qui pourra définir les besoins spécifiques au domaine végétal et bénéficier rapidement de possibilités technologiques de pointe.

De cette collaboration naîtront des prototypes spéciaux adaptés aux besoins spécifiques du projet GENOMES et qui conféreront certainement aux autres membres du GIP un élément sûr de compétitivité surtout si

ces instruments sont accessibles avant la mise sur le marché.

Deuxième pôle d'activité : vers un INVENTAIRE GENIQUE de TOUT ou PARTIE de CERTAINS GENOMES :

Cette activité se déroulera autour de plusieurs thèmes :

- 1) la découverte des gènes de maladies génétiques humaines
- 2) l'analyse du génome de souris
- 3) la cartographie du génome de certaines plantes
- 4) le séquençage de Chlamydia trachomatis\*
- 5) la participation à des projets internationaux
  - . cartographie systématique du génome humain
  - . séquence de subtilis de la levure
  - . banque normalisée de cDNA humain ou murin et leurs séquences, leur positionnement génomique

Ces thèmes ont un dénominateur commun - ils nécessitent tous les mêmes technologies : cartographie et séquence. Les laboratoires impliqués devront bénéficier des instruments les plus performants, idéalement ceux développés par le groupe de développement technologique (ou d'autres instruments du marché si nécessaire). Aussi les équipes impliquées (environ une vingtaine) pourront ainsi au cours des cinq prochaines années acquérir la compétence utile à ce type de méga-projet; certains d'entre eux pourraient alors faire face dans quelques années, si cela s'avère nécessaire, à un projet de séquençage de grande ampleur.

Parallèlement à ces activités de recherches, des activités de services devront être développées notamment pour la distribution nationale ou internationale des produits de ces recherches (clones ou données informatiques) .

Enfin il apparait essentiel d'envisager dès à présent d'accentuer les efforts entrepris sur les recherches physiopathologiques et thérapeutiques sur les gènes de maladies génétiques . Un GIP "Thérapeutique" devrait être créé parallèlement au GIP "Genomes" afin de souligner d'emblée une des finalités essentielles du Projet GENOMES: le traitement ou la prévention des maladies génétiques mono ou poly-factorielles .

\* C'est l'un des plus petit génome : environ 800.000 bases contre 5.000.000 pour Escherishia Coli sur laquelle Japonais et Américains sont en compétition . Deux à trois années seront nécessaires à une seule équipe correctement dotée pour séquencer le génome du Chlamydia .



## CARACTERISTIQUES ET AVANTAGES DE LA PROPOSITION

- . ELLE TIENT COMPTE AU MAXIMUM DES STRUCTURES EXISTANTES
- . ELLE TIENT COMPTE DES INTERETS MANIFESTES PAR LES SCIENTIFIQUES FRANÇAIS
- . ELLE TIENT COMPTE DES CAPACITES QU'ELLE ACCENTUE PROGRESSIVEMENT
- . ELLE PERMET A LA FRANCE D'AVOIR UNE POSITION ORIGINALE ET COMPETITIVE
- . ELLE EST COMPATIBLE AVEC LES PROJETS INTERNATIONAUX, NOTAMMENT EUROPEENS, SUR LAQUELLE ELLE S'APPUIE (EUREKA)
- . ELLE FAIT UNE GRANDE PLACE A L'AUTOMATISATION
- . ELLE S'APPUIE SUR LE GIP QUI :
  - . OBLIGE A DEFINIR CLAIREMENT PROJETS ET MOYENS ENTRE PARTENAIRES DIVERS
  - . PERMET UNE INTERACTION DES FINANCEMENTS PUBLIQUES ET PRIVES
  - . PERMET SUIVI DES RESULTATS ET ECHANGES D'IDEES
  - . PEUT SERVIR COMME INTERLOCUTEUR EUROPEEN