

PROJET - CONFIDENTIEL :

RAPPORT SUR LE "GENOME HUMAIN"
ELEMENTS D'UN PROJET "GENOMES"

(5 Juillet 1990)

Philippe KOURILSKY

PROJET - CONFIDENTIEL :

RAPPORT SUR LE "GENOME HUMAIN"
ELEMENTS D'UN PROJET "GENOMES"
(5 Juillet 1990)

PREAMBULE.....	3
I.- LES JUSTIFICATIONS DU PROJET GENOME.....	4
A) Les objections.....	4
B) Les retombées.....	5
1. Dans le champ scientifique.....	5
2. Dans le champ technologique.....	6
3. Dans la "sociologie" des milieux de recherche.....	7
4. Dans le domaine économique.....	7
II.- LA SITUATION FRANCAISE.....	9
III.- QUEL PROJET ?.....	11
A) Résumé de l'analyse des "bénéfices".....	11
B) L'étude des autres génomes.....	11
IV.- LES OBJECTIFS.....	12
A) Les objectifs pouvant présenter une importance économique.....	12
1. Le séquençage systématique de cADN.....	12
2. La mise au point d'instruments nouveaux.....	13
3. La commercialisation des banques de données	14
B) Les objectifs scientifiques.....	14
1. La définition des stratégies scientifiques.....	14
2. La recherche de nouveaux gènes humains.....	14
3. L'établissement de la carte génétique de l'homme.....	16
4. L'étude des autres génomes.....	16
5. L'instrumentation.....	16
6. La recherche en informatique.....	17
7. Les banques de données.....	17
a) La constitution des banques de données	17
b) La distribution des données.....	18
c) La situation française.....	18

V.- LE DEFI ORGANISATIONNEL	19
A) L'organisation nationale.....	20
B) La prise en compte des considérations éthiques, légales et sociales.....	21
C) La prise en compte de l'impact éducatif du projet génomique.....	21
D) Les articulations internationales.....	22
1. Les relations européennes.....	22
2. Les autres accords internationaux.....	22
VI.- ACTIONS POSSIBLES ET LEUR CALENDRIER	23
A) Liste des actions	23
B) Formes de financement.....	23
C) Echancier	25
1. Identifier, missionner puis nommer au plus tôt le Directeur du GIP.....	25
2. Entreprendre dès à présent l'analyse de quelques questions urgentes et notamment celle des banques de données.....	25
3. Le reste des actions	25
ANNEXES	26
1.- Liste des personnes consultées.....	
2.- Article du Dr. J.D. Watson sur le programme américain.....	
2'.- Le programme génome humain américain	
3.- Pétition de scientifiques américains hostiles au projet.....	
4.- Réflexions du Dr. B. Jordan sur un programme génome en France et sur le problème des personnels.....	
5.- Rapport du Dr. P. Avner sur le Génome Murin.....	
6.- Table des matières du projet CEE levure (P. Slonimski).....	
7.- Rapport du Dr. J.-M. Claverie sur les aspects mathématiques et informatiques dans le Programme Génome Humain.....	
8.- Rapport du Dr. S. Aymé sur la collecte des échantillons en vue de la cartographie des maladies génétiques humaines.....	
9.- Rapport du Dr. P. Oudet sur la bio-informatique dans les projets génomes.....	
10.- Rapport du Pr. J. Frézal sur Genatlas.....	
11.- Réflexion du Dr. P. Dessen sur les développement en bio- informatique moléculaire.....	

RAPPORT SUR LE "GENOME HUMAIN" ELEMENTS D'UN PROJET "GENOMES"

Philippe KOURILSKY

(Juillet 1990)

PREAMBULE

Le "génomome humain", grand projet biologique de la fin du XXème siècle agite les passions et fait couler beaucoup d'encre. Totem d'une certaine biologie du futur, symbole de la puissance nouvelle de la biologie, instrument de pouvoir pour certains, le "génomome humain" doit être débarrassé de sa gangue mythologico-médiatique pour faire l'objet d'une analyse froide et rigoureuse.

Le projet, en France comme aux Etats-Unis, a ses détracteurs dont on ne peut balayer les arguments en leur faisant le mauvais procès d'être hostiles au progrès ou à la nouveauté. Au contraire, ces arguments doivent être étudiés avec attention. Ils s'avèrent parfois plus constructifs que certains arguments mis en avant par les "défenseurs du génome" lorsque ceux-ci procèdent de la simple autorité (par exemple, "il serait inconcevable que la France ne se lance pas dans le programme génome").

De quoi s'agit-il ? En principe, et à terme, de déterminer l'emplacement respectif de tous les gènes humains sur les chromosomes, ainsi que la séquence totale de l'ADN qui constitue ces derniers. Le projet peut donc être décomposé, pour l'instant du moins, en deux grands axes : la cartographie et le séquençage. Et chacun de ces deux chapitres est décomposé à son tour en sous-problèmes touchant à l'instrumentation, à l'acquisition, au stockage et à la gestion des données. Un programme de séquençage beaucoup moins ambitieux et beaucoup moins difficile, consiste à s'intéresser aux séquences des ARN messagers sous forme de cADN clonés. Ce séquençage des cADN peut être effectué sans cartographie préalable et, à l'inverse, fournir des informations très utiles pour cette dernière. Bien entendu, il paraît peu cohérent de ne pas étendre l'approche à d'autres génomes que le génome humain. Enfin,

comme on le verra, il est impossible de juger sainement d'un tel projet sans examiner ses retombées technologiques et économiques ainsi que son impact "sociologique" dans les milieux de la recherche.

Le présent rapport a été établi après consultation des personnes dont la liste est fournie dans l'Annexe 1. Il se veut plus *analytique* que descriptif. Une description éclairante et documentée des divers aspects du projet génome humain figure dans le rapport établi par le Pr. D. Cohen. On pourra lire, par ailleurs, divers documents (dont un article récent du Dr. J. Watson) sur le projet américain (cf. Annexes 2 et 2'), ainsi qu'une pétition des adversaires de ce projet (cf. Annexe 3). Enfin, certaines des personnes consultées ont fourni des documents écrits qui sont, en totalité ou pour leurs parties les plus pertinentes (afin d'éviter les redondances), également annexées (cf. Annexes 4 à 11).

I.- LES JUSTIFICATIONS DU PROJET GENOME

Un projet génome *doit être justifié* et, si possible, faire l'objet d'un certain consensus. C'est pourquoi il est utile d'en analyser les bénéfices putatifs par rapport aux désavantages supposés. Ce rapport se situe dans une perspective de cinq ans. Il conviendra de suivre très attentivement l'évolution du programme et de ses objectifs qui peuvent évoluer de façon considérable et relativement imprévisible au cours des prochaines années.

A) Les objections

(a) Les grands programmes tels que le génome drainent une partie du financement public, au détriment, pensent certains, des autres secteurs. Dans le cas des sciences de la vie en France, il est important de souligner que la croissance des financements publics par les organismes au cours des 7-8 dernières années a été faible (négative même dans certains cas pour ce qui est du financement des laboratoires hors personnels). Les suppléments proviennent de recherches sur objectifs (SIDA) et des associations philanthropiques (ARC, AFM, etc...) qui poursuivent une politique généralement ouverte mais néanmoins ciblée. Il en résulte que

certaines secteurs de la biologie fondamentale vivent dans un état de pénurie semi-chronique. L'opinion dans le monde des chercheurs est largement répandue que hors le SIDA, le cancer, les maladies héréditaires et le génome, les financements sont difficiles à trouver. Même s'il faut nuancer cette appréciation, l'auteur de ce rapport se fait le porte-parole de nombreux collègues pour insister avec force sur le fait qu'un accroissement du budget des organismes (hors personnels) et la défense de certains secteurs de la biologie fondamentale sont importants et parfois urgents. Au-delà des budgets, il faut aussi prendre en compte les compétences des hommes et éviter d'assécher, par un pompage trop vigoureux en direction de certains secteurs, des domaines entiers (microbiologie, physiologie, etc...).

(b) La seconde objection est moins conjoncturelle et plus profonde. Elle peut être résumée en ceci que le *projet génome humain est relativement pauvre en ce qui concerne les objectifs scientifiques explicites*, même si les objectifs technologiques sont importants et évidents. Il est exact que l'accent initialement mis sur le séquençage systématique des génomes a plutôt nui à l'image du projet et que les questions scientifiques qui motivent cet énorme effort de séquençage ont été mal explicitées. A l'inverse, d'innombrables questions scientifiques sont encore sans réponse, même lorsque les séquences sont disponibles. Par exemple, bien que les gènes codant pour des maladies génétiques majeures aient été isolés, un travail considérable reste à effectuer pour comprendre les maladies elles-mêmes. Pour beaucoup l'aventure biologique consiste à exploiter les séquences plus qu'à les obtenir, et le taux d'acquisition actuel des séquences suffit largement à saturer les capacités de recherche.

B) Les retombées

Tout en conservant ces réserves en mémoire, on peut établir une liste des retombées attendues des projets génomes.

1. Dans le champ scientifique

a.- L'identification des gènes responsables de maladies génétiques n'entre pas dans la définition stricte du programme américain. Le

rapporteur propose de l'inclure dans le présent projet. Il s'agit là d'un sujet scientifiquement (et médicalement) très bien fondé.

- b.- La cartographie des chromosomes avec divers types de marqueurs et, plus spécifiquement, la localisation des gènes sur les cartes, constituent des objectifs scientifiquement fondés, surtout si ces opérations sont menées dans plusieurs espèces, ce qui apportera, de façon indubitable, des informations essentielles sur l'évolution. Chez l'homme seul, ces cartes permettront d'analyser les réarrangements chromosomiques qui se produisent dans diverses situations physiologiques et surtout pathologiques.
- c.- Le séquençage à l'aveugle de vastes régions d'ADN présente un coût excessif en temps, en matériel et en produits, et requiert, de l'avis général, que les techniques et les instruments soient améliorés. En outre, l'opération ne présente pas, à ce jour, de justification scientifique évidente. Peut-être serait-il intéressant de séquencer une région importante d'ADN (par exemple, le locus du CMH), même sans avoir de question précise ou de cadre théorique en tête. Ici encore, c'est la comparaison de plusieurs régions homologues dans différentes espèces ou différents individus qui serait susceptible de fournir des données de type évolutif dont l'urgence n'apparaît pas criante.
- d.- Le séquençage systématique (mais mené de façon intelligente) des ARN messagers sous forme de cADN présente un intérêt scientifique certain, ne serait-ce que pour identifier de nouveaux gènes. En outre, bien menées, ces analyses peuvent apporter d'importants renseignements sur des gènes exprimés dans des situations physiologiques ou pathologiques données. Bien que certains doutent que ce soit la meilleure façon de parvenir à des résultats présentant un sens biologique, l'approche doit être retenue, d'autant qu'elle peut, de surcroît, présenter un certain intérêt sur le plan économique (cf. plus bas).

2. Dans le champ technologique

La nécessité d'acquérir et de gérer des masses importantes de données, tant au niveau des cartes que des séquences, fait des projets génomes des projets de haute technologie. Il en résultera indubitablement un important développement d'instruments nouveaux, qui trouve-

ront d'abord leur place dans les laboratoires de recherche, mais plus tard dans le monde médical, pour l'analyse et le diagnostic notamment. Cet aspect des programmes génomes est essentiel -même aux yeux des plus réservés qui voient là les seules retombées positives assurées de ce type de programme.

3. Dans la "sociologie" des milieux de recherche

Les projets génomes sont aussi l'illustration et le symbole d'une évolution d'une partie de la biologie : concentration, accroissement considérable du degré de technicité, dépendance accrue vis-à-vis des banques de données informatisées, etc... Le fait, pour une communauté scientifique, d'être engagée dans cette opération aura donc une *valeur éducative* certaine. Les évolutions induites par les projets génomes dans les milieux de recherche ne se produiront pas sans soulever un certain nombre de questions, produire des frictions et engendrer des difficultés. Par exemple, il faudra sans doute, dans le cadre de ces projets de haute technologie, réfléchir aux missions et fonctions respectives des ingénieurs et des chercheurs, particulièrement dans les EPST où les fonctions et les salaires des ingénieurs sont souvent *inappropriés* pour ce type de travail, et où la carence en ingénieurs sera peut-être compensée par un afflux de chercheurs (mais lesquels ? et qui les évaluera ?). Peut-être faudra-t-il aussi analyser les répercussions de ces travaux sur le système des publications scientifiques : les données brutes rentreront-elles directement dans les banques de données informatisées, sans support papier ? Quelle sera alors la "valeur" des publications ? Et comment sera évalué, dans un travail très collectif, l'apport de chacun, en vue, notamment, des nominations et des promotions ? Si véritablement l'émergence des projets génome traduit une inflexion dans la recherche biologique, *il faut considérer comme positif le fait de pouvoir réfléchir à l'avance sur des évolutions inéluctables*, et expérimenter, à l'occasion des projets génomes, des voies nouvelles.

4. Dans le domaine économique

L'impact économique des projets génome n'est pas facile à évaluer. A terme, il est bien clair que les connaissances scientifiques et technologiques acquises bénéficieront à la santé de l'homme et fertiliseront le

terrain des biotechnologies nouvelles. Mais qu'en est-il des retombées plus immédiates ?

Le rapporteur estime que, quels que soient les élans généreux visant à mettre en commun les informations génétiques au service de tous, ces informations feront l'objet de commerces, et ce d'au moins trois façons différentes :

- 1) les banques de données génome -comme la plupart des banques de données- seront certainement commercialisées (à la manière de Medline, etc..). On peut aussi imaginer que les accès aux données fasse l'objet de trocs. En particulier, il paraît tout à fait possible que des contributions individuelles (ou nationales) à un effort d'ensemble (ou international) constituent des modes d'accès privilégiés aux données des autres et aux banques de données générales.
- 2) Les informations concernant les nouveaux gènes peuvent-elles être protégées par des *brevets* ? Bien qu'en droit des brevets strict, il semble qu'on ne puisse pas protéger la séquence d'un gène (car il s'agit de la description d'un objet "naturel" sans apport inventif), la protection devient possible dès lors que des hypothèses de fonction et d'utilisation sont formulées. Or, du fait de l'extraordinaire cohérence du monde des molécules biologiques (qui reflète la réutilisation perpétuelle des matériaux génétiques au cours de l'évolution des espèces), il est fréquent que la séquence d'un gène nouveau de fonction inconnue, présente des homologues avec la séquence d'un autre gène de fonction connue, ce qui permet de formuler une hypothèse vraisemblable de fonction. On peut attendre qu'à mesure de la croissance des banques de gènes (de tous organismes et microorganismes), les homologues éclairantes seront de plus en plus nombreuses. Il est donc probable que les séquences de gènes nouveaux seront, pour partie, effectivement brevetées grâce à des hypothèses de fonction qui ouvriront la perspective de cibles plausibles.
- 3) Même si les séquences des nouveaux gènes ne sont pas brevetées, leur exploitation commerciale reste possible. Il est illusoire en effet de croire que les informations de séquence seront immédiatement mises dans le domaine public, et ce pour deux raisons : (a) seules les

séquences vérifiées seront diffusées et ceci est conforme aux règles déontologiques habituelles ; (b) le délai de diffusion d'une séquence vérifiée est d'au moins six mois dans le système de publication actuel. Un accord au sein de la communauté internationale pourrait fournir une recommandation forte concernant le délai de diffusion. On voit mal comment cette recommandation serait véritablement contraignante, et l'on comprend bien, à l'inverse, que des chercheurs ou ingénieurs qui auront accompli un travail long et ingrat, voudront recueillir quelques fruits au niveau de l'interprétation des données, avant de laisser à d'autres la partie la plus intéressante du travail. En résumé, on peut estimer qu'il s'écoulera une période de 12 à 18 mois entre le moment où une séquence brute, non vérifiée, est obtenue et le moment où elle est mise en libre circulation. Pendant cette période, l'*information restera captive*, tout en étant utilisable : si une homologie de séquence fait apparaître une hypothèse de fonction exploitable (e.g. la possibilité d'un nouveau canal ionique, d'un nouveau récepteur hormonal, etc...), l'information pourra être livrée à un industriel qui pourra de ce fait bénéficier d'une priorité utile.

II.- LA SITUATION FRANCAISE

La France bénéficie de plusieurs atouts importants.

- 1) Le CEPH constitue un pôle actif et reconnu dans plusieurs des secteurs importants concernant le génome humain. Des projets associant Bertin (dont certains dans le cadre d'un contrat Euréka impliquant l'ICRF, en Grande-Bretagne) touchent aux questions d'automatisation.
- 2) La France possède une recherche de très bon niveau dans les domaines de la génétique moléculaire.
- 3) La France dispose, grâce aux formations multidisciplinaires, des Universités et des Grandes Ecoles, de ressources humaines appropriées.

- 4) Le public (mobilisé par des opérations comme le Téléthon) porte un intérêt certain aux maladies génétiques et une certaine contribution de fonds privés est envisageable.
- 5) Certains industriels français semblent manifester de l'intérêt pour le projet.
- 6) La France pourrait jouer un rôle moteur en Europe, où seule la Grande-Bretagne est vraiment active. La RFA est hors du courant pour plusieurs raisons (attitude négative des courants écologistes, problèmes liés à la réunification des deux Allemagnes, etc...). Le projet CEE est géré de façon un peu lourde et n'a pas toujours pour objet d'être "leader" dans le domaine (au moins pour le génome humain), mais plutôt de se préparer à utiliser l'information lorsqu'elle sera disponible. L'EMBL est, en revanche, un pôle actif (dont les axes sont distincts de ceux de la CEE).

La conclusion de cette brève analyse est que la France devrait être en mesure de mettre sur pied un programme intelligent à coût modéré visant non seulement à acquérir des connaissances nouvelles mais aussi à faire évoluer le mode d'acquisition des connaissances. La France pourrait être membre du "Club" (comme dans le nucléaire). Elle conserverait une place et une capacité de dialogue avec d'autres grands partenaires (particulièrement les Etats-Unis qui, en gros, se déclarent ouverts à des collaborations et à un dialogue pour autant que cela ne ralentisse pas leur effort). Enfin, si le coût reste modéré, l'impact négatif sur d'autres types de recherche ne saurait être significatif. (Ceci est important pour éviter que la communauté des biologistes ne développe une forte hostilité contre ce projet). Enfin, la France pourrait jouer un rôle d'entraînement en Europe (peut-être en nouant des relations bi-, puis pluri-latérales, comme cela a été le cas pour Airbus ou l'espace) (cf. Annexe 4).

III.- QUEL PROJET ?

A) Résumé de l'analyse des "bénéfices" (section I). Les bénéfices à attendre proviennent :

- de développements *scientifiques* liés à l'identification de nouveaux gènes (notamment des gènes associés à des maladies héréditaires) et à la cartographie. Le séquençage à grande échelle n'est pas, à ce stade, une priorité, alors que le séquençage de cADN est scientifiquement fondé ;
- de développements *technologiques* liés à la mise au point et à l'utilisation d'instruments nouveaux ;
- du développement de *nouveaux comportements* dans le monde des chercheurs, confrontés à un phénomène de concentration dans certaines de leurs activités ;
- de retombées *économiques* au niveau (1) des banques de données et (2) de la protection et/ou de l'accès privilégié à des informations présentant un intérêt industriel potentiel (à plus long terme, d'une fertilisation des sphères de la biotechnologie et de la santé).

B) L'étude des autres génomes

On notera ici que l'étude d'autres génomes que le génome humain présente des bénéfices potentiels très significatifs (et parfois du même ordre). C'est ainsi que l'étude du génome de certaines plantes (maïs par exemple) ou de certains animaux domestiques (porc) pourrait avoir des retombées économiques importantes. Il est essentiel de faire une nouvelle fois mention de la profonde cohérence du monde vivant qui a pour conséquence exemplaire que l'un des meilleurs moyens d'aborder l'étude du développement de l'être humain est, à l'heure actuelle, de travailler sur la drosophile, etc... En particulier, les limites évidentes de l'expérimentation humaine imposent le recours à des modèles animaux, généralement de mammifères, le plus souvent la souris. Un effort en direction du "génome murin" paraît donc indispensable (cf. Annexe 5) et l'analyse de génomes plus petits (drosophile, nématode, levure, *B. subtilis*, *E. coli*) est tout aussi justifiée (cf. Annexe 6). Considérant l'homogénéité des techniques de saisie et de gestion des données (en gros, et bien qu'il puisse exister quelques différences, les mêmes ins-

truments nouveaux sont requis, les mêmes structures de banques de données sont nécessaires, etc...), il est logique de *prévoir des efforts coordonnés impliquant d'autres génomes.*

IV.- LES OBJECTIFS

A) Les objectifs pouvant présenter une importance économique

Le rapporteur propose, à ce stade, d'analyser les quelques objectifs qui peuvent présenter une importance économique et d'examiner comment les traiter. De deux choses l'une, en effet : ou bien ces objectifs sont jugés stratégiques, et *il faut s'engager dans une compétition à visée économique avec les moyens appropriés*, ou ils ne sont pas jugés tels et il n'est pas nécessaire de mettre en place des dispositifs exceptionnels.

1. Le séquençage systématique de cADN est le plus évident de ces objectifs stratégiques éventuels. La mise en place d'un ou plusieurs petits centres se livrant à cette activité aurait l'avantage de couvrir un créneau ayant une importance économique probable, tout en livrant des informations scientifiques conséquentes à la communauté. Ces centres assureraient aussi la veille scientifique en ce qui concerne le séquençage et serviraient de terrains d'expérimentation pour d'éventuels nouveaux instruments. Cette activité de séquençage peut être entreprise dès à présent. Elle justifie de faire l'impasse totale ou quasi-totale sur le séquençage chromosomique à grande échelle dans les années qui viennent. Elle peut enfin, tout en étant systématique, rester "intelligente" et être orientée vers l'analyse de problèmes de santé importants. Un effort parallèle, mais moins intensif, pourrait être entrepris chez la souris. On pourrait imaginer, par exemple, que l'homologue murin de tout cADN important chez l'homme soit immédiatement recherché.

Si cet objectif est jugé stratégique, il doit être poursuivi avec les moyens *ad hoc*. Dans l'esprit du rapporteur, il devrait être traité *dans un état d'esprit industriel* (où les notions de rendement et de rentabilité ont

une valeur princeps) et, peut-être, dans un contexte industriel. L'accès préférentiel aux données brutes serait possible, avec priorité aux industriels français et européens. La publication des données vérifiées aurait lieu dans un délai (12 à 18 mois) strictement défini. Il serait souhaitable de pratiquer l'exercice dans au moins deux centres dont les performances pourraient être comparées et il n'est pas à exclure que l'un de ces centres émerge au secteur public et l'autre au secteur privé.

Il n'existe, aux yeux du rapporteur, aucun inconvénient à traiter ce domaine comme stratégique pendant plusieurs années, même s'il apparaît à terme que son importance économique possible a été surévaluée. Les données scientifiques acquises ne seront pas perdues et la veille technologiques concernant le séquençage organisé à moyenne échelle aura été assurée. On ne devra pas omettre d'articuler les banques de cADN avec les banques des protéines pour lesquelles les cADN codent (par exemple après synthèse *in vitro*).

2. La mise au point d'instruments nouveaux constitue un deuxième objectif stratégique possible. La situation est ici plus complexe : le marché des instruments de laboratoire n'est pas considérable et ne bénéficie pas, en général, des mêmes investissements que l'industrie pharmaceutique est parfois prête à effectuer très en amont. Enfin, et surtout, un instrument non ou peu vendu constitue, pour l'essentiel, une perte sèche. Par conséquent, il convient ici d'être circonspect. Autant il est justifié de développer des instruments fondés sur des idées originales ou bénéficiant d'une avance certaine, autant il serait illusoire de vouloir attaquer des créneaux déjà fort encombrés sans autre atout particulier que de vouloir y monter hardiment, mais avec retard. Les compétences en instrumentation se trouvent principalement dans le monde des physiciens et des ingénieurs. On ne devra pas oublier que la chimie peut jouer un rôle important, par exemple dans la diminution des coûts de séquençage. On n'oubliera pas non plus que des "instruments" informatiques nouveaux (machines et logiciels) seront probablement développés.

- de pour les années - J
- avec protéines

3. La commercialisation des banques de données est une affaire des plus complexes. On distinguera une hypothétique commercialisation de l'accès aux banques de données qui, quoi qu'il arrive, devra être organisé. Le problème majeur en matière de commercialisation est de disposer d'un produit de bonne qualité, fondé sur une base quasi-exhaustive (même si le produit n'est qu'un sous-ensemble extrait de la base) et très rapidement mis à jour. Ces problèmes ne sont pas minces et les bases de données nationales (Pascal, etc...) ont, par le passé, révélé quelques défaillances. La montée en puissance de l'INIST permettrait d'aborder la question sous un angle nouveau.

B) Les objectifs scientifiques

1. La définition des stratégies scientifiques

Comme il a été noté dans le préambule, l'un des traits les plus surprenants (et les plus critiqués) des projets génomes est la relative faiblesse des justifications scientifiques : pour une part, ces projets reposent sur l'*a priori* selon lequel la connaissance des cartes, des gènes et des séquences jouera un rôle capital dans le développement de la biologie et de la médecine. Ce présupposé n'est pas accepté de tous, la question étant moins celle de l'objectif final que celle de la priorité à accorder à ce type d'approche. Ce sur quoi il faut ici insister est que, *même si l'on accepte les prémisses, il reste essentiel de bien analyser et de bien critiquer les stratégies scientifiques* et ce d'autant plus que l'approche "génomes" des problèmes biologiques est particulière onéreuse. Peut-être apercevra-t-on alors que certaines approches ont des bornes théoriques inhérentes, auxquelles on n'avait pas songé, qui les rendent impropres à la résolution du problème posé. Il y a place ici pour certain type de recherches qui associeraient des biologistes à des mathématiciens et des logiciens (cf. Annexe 7).

2. La recherche de nouveaux gènes humains

On rappellera d'abord, en exergue, qu'une part significative de la recherche biologique et médicale est, en dehors de tout projet génome, consacrée à la recherche de nouveaux gènes, dans une approche phénoménologique et fonctionnelle plutôt que systématique et aveugle. Cette approche de fourmi a conduit à identifier plus de mille gènes humains

en une dizaine d'années, tout en révélant leur fonction. Elle a donc fait ses preuves et l'on ne saurait purement et simplement lâcher la proie pour l'ombre, au profit d'une approche différente dont l'efficacité reste à prouver.

a) *L'analyse systématique de cADN* constitue une voie privilégiée de ce projet (cf. plus haut).

b) La deuxième voie privilégiée est la *recherche de gènes impliqués dans des maladies héréditaires*. Cette voie est suivie dans plusieurs laboratoires de recherche très actifs, soutenus par les organismes de recherche et par l'AFM qui juge cette démarche prioritaire et y investit des fonds importants. Cette part du programme recouvre en fait trois types d'activité.

- La recherche et l'analyse clinique des malades porteurs de maladies héréditaires (ou de leurs familles) et la collecte des échantillons (généralement une prise de sang).
- La conservation des échantillons (avec immortalisation éventuelle des cellules) pose le délicat problème des banques (centralisées ou décentralisées, destinées à la seule recherche ou à la clinique, etc...).
- L'analyse, par les voies moléculaires, des échantillons requiert un important travail de biologie moléculaire que l'AFM, en conjonction avec le CEPH, se propose de soutenir vigoureusement.

Les gènes responsables des principales maladies héréditaires provoquées par la déficience d'un seul gène sont maintenant identifiés. Il reste à découvrir les gènes responsables de maladies héréditaires monogéniques rares ou hétérogènes, ou encore ceux responsables des maladies à hérédité complexe. Dans tous les cas, la qualité de l'expertise génétique en milieu clinique joue et jouera un rôle essentiel. Cette simple considération milite contre une concentration excessive au niveau de la collecte et du stockage des échantillons. La gestion de ces deux activités pourrait être organisée de façon conjointe par l'INSERM et l'AFM, selon les principes énoncés dans l'Annexe 8 (rapport de Madame S. Aymé). Ces principes font l'objet d'un large consensus et le rapporteur recommande de s'y conformer.

3. L'établissement de la carte génétique de l'homme recouvre une série d'activités bien détaillées dans le rapport du Pr. D. Cohen. Aux thématiques précédentes et à l'étude des polymorphismes génétiques, vient se greffer la construction et l'exploitation des banques de YAC, l'hybridation *in situ* sur chromosome, etc...

4. L'étude des autres génomes se heurte à des problèmes méthodologiques et techniques très différents selon qu'il s'agit de génomes complexes ou peu complexes. La cartographie et le séquençage du génome de la souris sont analogues en difficulté aux mêmes opérations menées sur l'homme. Il en va de même pour les génomes végétaux de haute complexité. Pour ceux-là, les priorités seront donc, comme pour le génome humain, axées sur la cartographie et le séquençage des cADN. Pour les plus petits génomes (*e.g.* levure, *B. subtilis* et *E. coli*), la complexité est bien moindre et la densité d'information génomique codante très élevée (ceci est illustré pour la levure par les travaux récents de l'équipe du Pr. P. Slonimski). Dans ce cas, bien évidemment, la cartographie est beaucoup plus facile et la priorité accordée au séquençage de cADN doit être remplacée par le séquençage génomique. Ceci rejoint les préoccupations de la CEE qui soutient le séquençage des petits génomes (cf. Annexe 6).

5. L'instrumentation est, de l'avis général, un élément tout à fait critique des projets génomes, les progrès de l'instrumentation constitueront certainement une de leurs retombées importantes. Les instruments dont il est question sont, soit des automates, reproduisant de façon automatique des opérations répétitives faites à la main, soit des instruments tout à fait nouveaux. Les extracteurs d'ADN, machines à faire des "Southern blots", chaînes automatiques de préparation d'échantillons en vue de leur distribution, ainsi que certains séquenceurs d'ADN, sont des automates. En revanche, et surtout dans le domaine du séquençage, des instruments fondés sur des principes nouveaux (*e.g.* spectrométrie de masse au CEA) pourraient émerger à l'état de prototypes. Ce domaine de l'instrumentation doit être suivi avec une vigilance particulière, en raison de son importance propre, mais

aussi en raison des coûts élevés de recherche et développement, et enfin en raison de la relative difficulté d'évaluation de ces activités pluridisciplinaires.

On sait que le séquençage est un goulot d'étranglement des projets génomes. Dans le projet proposé ici, l'accent sera mis sur les cADN et les petits génomes qui sont les seuls immédiatement accessibles. On n'oubliera pas que les facteurs limitants sont à la fois le *temps de travail* et le *coût*. Dans l'optique d'une diminution des coûts, il importe de ne pas négliger la *chimie* qui pourrait, à partir de méthodes relativement conventionnelles, fournir des procédés moins chers.

6. La recherche en informatique doit être développée (cf. Annexe 9) -en dehors même des problèmes de banque de données, cf. ci-dessous- et ceci pour plusieurs raisons. D'une part, le problème de l'interprétation automatique des séquences brutes (recherche des gènes, des promoteurs, des introns, exons et éléments signifiants divers) est loin d'être résolu. Les comparaisons de séquences à grande échelle posent des problèmes non négligeables. D'autre part, beaucoup de recherches concernant la cartographie (chevauchement des clones de YAC) ou les séquences (chevauchement, détection des erreurs) ou encore la saisie et l'interprétation automatique des données (*e.g.* Southern blots) requièrent une informatique intelligente. D'autres recherches devraient viser à améliorer l'exploration des banques de données. Dans tous les cas, on veillera à ce que les logiciels nouveaux soient vus ou revus par des professionnels de l'informatique, afin que les utilisateurs disposent d'outils fiables et conviviaux.

7. Les banques de données constituent un problème important et difficile (cf. Annexes 2', 9, 10, 11).

a) *La constitution des banques de données*

Des dispositifs d'une ampleur considérable sont mis en place, surtout aux Etats-Unis (cf. Annexe 2'). Le rapporteur pense qu'il est totalement irréaliste pour la France de viser à développer seule une ou plusieurs bases de données exhaustives. L'analyse faite par plusieurs experts consultés est que même si la saisie des données peut être (avec le bon vou-

loir des uns et des autres, et notamment celui des éditeurs) automatisée, la *qualification de l'information saisie* présente des difficultés considérables et requiert l'intervention directe de personnels qualifiés. En conséquence, si véritablement les banques de données croissent de façon considérable, ce problème deviendra rapidement impossible à maîtriser (on sait avec quelles difficultés les banques de séquences actuelles sont mises à jour et qualifiées). *La solution préconisée consiste à trouver des aires de spécialisation*, au niveau de la qualification de l'information, pour constituer quelques banques de données spécialisées de haute valeur ajoutée. Ces banques pourront ensuite servir de modules dans des banques internationales plus complètes. De ce point de vue, il convient de maintenir en France une activité raisonnable mais non nulle dans ce secteur, afin de demeurer en position de dialogue et d'échange.

b) *La distribution des données*

La mise à disposition des données pose des problèmes d'une autre nature. La décentralisation liée à l'emploi de la microinformatique performante permet à chacun, à des coûts accessibles, de recevoir et de traiter l'information. Par exemple, à l'heure actuelle, toutes les analyses de séquences de gènes peuvent être faites sur un micro-ordinateur, à l'exception notable d'une opération importante, la comparaison d'une séquence à la totalité de la banque. Il ne faut pas, néanmoins, céder trop vite aux sirènes d'une micro-informatique exclusive. Une certaine centralisation reste, pour l'instant, indispensable (pour la mise à jour des banques, le maintien des programmes d'analyse, etc...) (cf. Annexes 9 et 11).

c) *La situation française*

Le terrain semble bon pour dégager, en France, des solutions efficaces et harmonieuses. D'une part, la France dispose de plusieurs groupes de bio-informatique performants (au CEPH, à l'Institut Pasteur, à Gif, à Strasbourg, à Montpellier, etc...). D'autre part, diverses initiatives ont engendré une activité non négligeable en matière de banques de données. La banque Genatlas (J. Frézal) (cf. Annexe 10) mérite, selon l'argument développé plus haut, d'être soutenue, mais à terme spécialisée. Le réseau Bisance, créé par le MRT en 1981 (cf. Annexe 11), implan-

té au Citi2 et, depuis, soutenu par le MRT et le MEN, a rendu de nombreux services, et doit être intelligemment développé dans le cadre d'une structure en réseau sur le territoire national. Un groupe d'experts devrait être rapidement réuni afin d'élaborer des propositions constructives et consensuelles qui permettraient d'utiliser ces bonnes bases de départ. Le rapporteur ne saurait trop insister sur le fait que l'accès facile à de bonnes bases de données conditionnera, en partie, la qualité de la recherche et de toutes les retombées positives des opérations génomes. Les groupes travaillant de façon compétitive sur les projets génomes ont tous besoin, d'un petit noyau local d'informaticiens et d'un accès facile aux banques. On attachera donc une attention particulière :

- au développement d'une ou quelques bases de données spécialisées à haute valeur ajoutée ;
- au développement d'un serveur ou d'un réseau de serveurs distribuant les banques de données exhaustives et actualisées ;
- à l'implantation ou au renforcement de petits groupes de bio-informaticiens dans tous les centres où se développent des projets "génomes".

Il faudra, bien entendu, analyser très attentivement la situation européenne et internationale, et mandater, à cette fin, des experts ayant une bonne connaissance du milieu (par exemple, P. Oudet pour l'Europe et J.-M. Claverie -en stage sabbatique aux Etats-Unis- pour les Etats-Unis).

V.- LE DEFI ORGANISATIONNEL

Au-delà des défis d'ordre scientifique, technologique ou économique, le projet génome soulève un défi d'ordre organisationnel qui doit être analysé avec attention et lucidité. Le projet est en effet, plus que multi-disciplinaire, trans-disciplinaire, en ce qu'il fait intervenir des aspects très divers de la science et de la technologie et de nombreux partenaires institutionnels. On ajoutera à ceci l'absolue nécessité d'entretenir un important réseau de collaboration ou de relations internationales, et avant tout européennes. Et l'on imagine bien que si une

organisation appropriée n'est pas mise sur pied, le concert espéré pourrait tourner à une cacophonie qui ne manquerait pas d'être largement perçue du fait de la médiatisation de ce champ d'activité. On tentera ci-dessous de tracer les lignes d'une organisation nationale, avant de donner les axes et les principes de relations et de coopérations internationales.

A) L'organisation nationale

Le projet génome représente avant tout un défi pour les EPST. Dans le très vaste domaine des Sciences de la Vie et de la Santé, la recherche est justement (et nécessairement) répartie entre plusieurs EPST qui doivent ici trouver une forme de coopération forte. Et comme celle-ci devra probablement opérer sous le feu des projecteurs médiatiques, il conviendra qu'elle soit particulièrement transparente. Le rapporteur tient à souligner ici qu'au cours de ses différentes consultations, il a rencontré le meilleur accueil et recueilli de tous un désir affirmé de collaboration. Le climat est donc particulièrement propice.

Un Groupement d'Intérêt Public (GIP) semble être la structure la plus appropriée de gestion d'un tel programme. Un GIP présenterait, entre autres, l'avantage de pouvoir associer de façon transparente les organismes de recherche, ainsi que d'autres partenaires éventuels, comme des associations philanthropiques (AFM) et des industriels. Même si une forme juridique différente était adoptée, les principes de gestion suivants, énoncés dans l'hypothèse d'un GIP, devraient être respectés.

- 1) Le Directeur du GIP doit occuper cette fonction à *temps plein*. Il serait désastreux de confier le pilotage de l'ensemble à une personnalité ne disposant pas du temps nécessaire.
- 2) *Le Directeur du GIP doit avoir de solides qualités de "manager" et une certaine connaissance des problèmes industriels et commerciaux.* Ces qualités sont beaucoup plus importantes que l'expertise scientifique dont il pourra s'entourer par ailleurs.
- 3) Il découle de ces prémisses que *le projet doit se donner les moyens (si nécessaire sur son propre budget) de recruter à temps plein, un homme de grande valeur, a priori* choisi en fonction de son expé-

rience de "manager" qu'il aura si possible acquise dans une sphère proche des milieux de recherche, publique ou industrielle.

- 4) Le Directeur (qui, dans l'optique précédente, est peu qualifié sur le plan strictement scientifique) sera entouré d'un *Conseil Scientifique*.
- 5) Le programme est divisé en un certain nombre de projets dont la liste est fournie dans le Tableau I, avec une indication des efforts à fournir dans les divers secteurs.
- 6) Certains de ces projets seront placés sous la responsabilité de personnalités scientifiques mais certains projets stratégiques, ainsi qu'il est indiqué dans le Tableau I, devront être, au moins dans la phase initiale, de la responsabilité immédiate du Directeur.
- 7) Un soin particulier sera apporté dans la constitution du Conseil Scientifique et dans la nomination des responsables de projet, à *distinguer les opérateurs des acteurs, et à éviter que nul, dans le processus d'évaluation des travaux, ne soit juge et partie.*

B) La prise en compte des considérations éthiques, légales et sociales

Il sera important de prendre très attentivement en compte les problèmes d'ordre éthique, légal et social, que pourront engendrer les recherches sur les génomes et particulièrement sur le génome humain. Il conviendra donc de consulter chaque fois qu'il sera utile le Comité National d'Éthique (et, éventuellement, la C.N.I.L.). Il conviendra aussi d'informer le public sur le déroulement et les progrès des recherches. Enfin, il serait probablement très utile de confier à des *économistes* l'étude des retombées possibles, dans le domaine social notamment, afin de s'y préparer au mieux.

C) La prise en compte de l'impact éducatif du projet génome

L'importance de cet impact "éducatif" sur les milieux de recherche a été soulignée plus haut. Il incombe donc au MRT et aux EPST (et aux institutions de recherche para-publiques) d'observer attentivement l'évolution de la situation. La création d'un "*observatoire*" chargé d'analyser l'évolution des techniques et leur impact sur la recherche (coût, comportement des chercheurs, missions des chercheurs *versus*