::Savoirs LE FIL DES IDÉES

Collectionner ou expérimenter ?

Les bases de données bio-informatiques dans les sciences du vivant

Lieux de savoir, 1. Espaces et communautés, Albin Michel, 2007, p. 681-703

Bruno J. Strasser

Ce texte a bénéficié des précieuses remarques de Joe November, Michael Bürgi et Fréderic Joye. Je tiens à remercier les archivistes pour leur collaboration, en particulier à l'American Philosophical Society et au California Institute of Technology Archives, et surtout Robert S. Ledley pour m'avoir accordé l'accès aux archives privées de la National Biomedical Research Foundation, et la famille de Margaret O. Dayhoff pour l'accès à ses archives privées. Je remercie également le Fonds national suisse de la recherche scientifique (subside 105311-109973/1) et la Fondation du 450e anniversaire pour leur soutien.

La révolution moléculaire

La « révolution moléculaire » est sans conteste la transformation la plus profonde dans les sciences de la vie au XXe siècle [1]. Elle marque un déplacement des pratiques et des explications scientifiques, jusqu'alors centrées sur les organismes et les cellules, vers le niveau moléculaire. Cette transformation, bien qu'elle s'inscrive dans un processus plus large, est généralement associée à l'émergence d'une nouvelle discipline : la biologie moléculaire[2]. Cette dernière a été constituée en opposition à la « biologie traditionnelle ». En effet, les partisans de la biologie moléculaire, et parmi eux de nombreux physiciens, ont souligné combien leur nouvelle discipline se distinguait de la biologie telle qu'elle était alors pratiquée[3]. Ils ont brossé le portrait d'une biologie traditionnelle, qui relèverait essentiellement de l'histoire naturelle : le savant se préoccuperait surtout de décrire la diversité de la nature et accumulerait des observations qualitatives ; puis, s'aidant de méthodes comparatives, il conclurait à l'existence de quelques principes d'organisation généraux. Ces résultats seraient publiés dans d'épaisses monographies, atlas ou obscurs essais de *Naturphilosophie*. Les biologistes moléculaires

en revanche auraient recours à l'expérimentation quantitative pour arriver à des explications mécanistes et réductionnistes du vivant, publiées dans des articles scientifiques à l'image de ce qui se faisait en physique expérimentale[4].

Il est évident que l'image présentée par les promoteurs de la biologie moléculaire constitue une caricature grossière des pratiques et des objectifs de la « biologie traditionnelle » du milieu du xxe siècle. Cette caricature servait avant tout la stratégie disciplinaire des biologistes moléculaires soucieux de créer une rupture profonde entre eux, les Modernes, et les autres, les Anciens[5]. Il est, par conséquent, surprenant de remarquer qu'à l'avant-garde de la recherche contemporaine en biologie moléculaire on retrouve une pratique qui s'apparente plus à celle de l'histoire naturelle, décriée par les biologistes moléculaires, qu'aux stratégies expérimentales et épistémiques qui les caractériseraient : l'usage des bases de séquences bio-informatiques.

Les bases de séquences forment aujourd'hui l'un des outils de recherche les plus communs des laboratoires de recherche en biologie, au même titre que le microscope optique il y a un siècle. Elles représentent des collections de séquences de différentes protéines ou d'ADN (acide désoxyribonucléique) provenant d'une variété d'organismes. Les protéines et l'ADN sont des molécules composées d'une chaîne de différents éléments, dont la séquence détermine la signification, tout comme les lettres dans un mot. Les chercheurs accèdent aux bases de séquences grâce à l'Internet afin d'y lire ou d'y inscrire des séquences. Aujourd'hui, plusieurs dizaines de milliers de chercheurs accèdent quotidiennement à la principale base de séquences d'ADN (GenBank)[6]. En utilisant des programmes informatiques permettant de comparer ces séquences entre elles, les chercheurs peuvent inférer la fonction biologique de la séquence étudiée, ainsi que l'histoire évolutive de l'organisme dont elle provient. La recherche biologique *in silicio* vient s'ajouter aux méthodes *in vivo* et *in vitro* traditionnellement utilisées dans les sciences de la vie. Cette manière d'explorer le vivant constitue un mode de production des connaissances qui relève sous certains aspects de l'histoire naturelle et se distingue de l'approche expérimentale, analytique et réductionniste, qui semble dominer les sciences de la vie au xxe siècle [7]. L'usage des bases de données s'inscrit dans un lieu d'un genre particulier, un lieu « délocalisé », un espace fragmenté et distribué sur d'innombrables sites. Cette façon de procéder représente, de par le nombre de chercheurs impliqués, une forme originale de *Big Science* [8] dans les sciences de la vie, où le laboratoire de taille modeste était jusqu'alors la norme. Toutefois, contrairement au laboratoire des physiciens du CERN (Conseil européen pour la

recherche nucléaire) par exemple, les chercheurs contribuant aux bases de données sont dispersés partout dans le monde et reliés grâce aux réseaux informatiques au sein d'un espace virtuel global. Les bases de données constituent donc à la fois le prolongement d'une tradition de l'histoire naturelle et une innovation sociale importante pour les sciences de la vie.

Un musée au laboratoire ?

La littérature secondaire décrivant le rôle des collections, notamment les jardins et les musées, dans le développement de l'histoire naturelle est abondante. En revanche, leur rôle dans l'essor des sciences expérimentales a été largement ignoré au profit d'études sur des objets plus conformes à l'idéal expérimentaliste, tels les grands instruments scientifiques, accélérateurs de particules ou microscopes électroniques[9]. Dans le domaine des sciences de la vie, si les collections botaniques établies à partir du XVIIe siècle ou les collections anatomopathologiques du XIXe continuent à être enrichies au XXe siècle, elles sont souvent considérées comme relevant d'une pratique scientifique archaïque, qui s'incline peu à peu devant l'essor des sciences expérimentales. Pourtant, de nombreuses collections ont également été créées au XXe siècle et elles ont joué un rôle essentiel dans le développement de plusieurs disciplines en biologie expérimentale, comme la génétique ou la microbiologie. L'Eugenics Record Office, par exemple, fondé en 1910 à Cold Spring Harbor, rassemble plus de six millions de descriptions de traits héréditaires[10], et l'American Type Culture Collection, établi à Chicago en 1925, préserve des milliers de souches bactériennes vivantes. Une des premières collections dans le domaine de la recherche sur les protéines, la Cambridge Structural Database, est créée en 1965 par Olga Kennard à l'Université de Cambridge, et comprend la description de structures tridimensionnelles de protéines obtenues par analyse cristallographique. Un an plus tard, le généticien Victor MacKusick fonde Mendelian Inheritance in Man, une base de données des maladies héréditaires, à la Johns Hopkins Medical School de Baltimore. Au même moment, les bases de données bibliographiques se multiplient, en réponse à l'explosion du nombre des publications scientifiques[11]. Les bases de séquences de protéines, qui apparaissent dans la même période, combinent les informations de ces différentes collections : descriptions de structures, de maladies, et références bibliographiques.

L' *Atlas of Protein Sequence and Structure*, publié en 1965, constitue la première base d'envergure de séquences de protéines. Elle résulte du travail mené par Margaret O. Dayhoff (1925-1983) à la National Biomedical Research Foundation (NBRF). Cette institution privée de

recherche sans but lucratif a été fondée en 1960 par Robert S. Ledley (né en 1926) dans les environs de Washington D.C. Dentiste formé en physique et en mathématiques à Columbia University (New York), Ledley avait travaillé dans le domaine de la recherche opérationnelle avant de s'intéresser aux utilisations des ordinateurs dans la recherche biomédicale[12]. En 1956, l'Us Air Force avait demandé à la National Academy of Science de convoquer une conférence sur ce thème. L'U S Air Force déployait des efforts importants pour tenter de justifier « les immenses sommes investies [jusqu'alors] dans les ordinateurs digitaux[13] ». À la suite de la conférence, la National Academy of Science a confié à Ledley une étude extensive sur le sujet. E n 1960, il rendra un manuscrit qui paraîtra sous la forme d'une monographie de plus de neuf cents pages intitulée : *Uses of Computers in Biology and Medicine* [14]. Cet ouvrage constitue une introduction au fonctionnement des ordinateurs et à leur utilisation, destinée aux biologistes et aux médecins. Il se fait écho de la volonté de Ledley de promouvoir les applications des ordinateurs digitaux dans les domaines les plus divers de la recherche biomédicale, allant de la reconnaissance automatique d'images microscopiques au diagnostic médical assisté par ordinateur[15].

En raison de ses compétences dans le domaine des ordinateurs et de la logique formelle, Robert S. Ledley avait été contacté par le physicien George Gamow, en 1954, pour devenir l'un des vingt membres du « RNA Tie Club[16] ». Ce groupe informel de chercheurs avait été constitué après la découverte de la structure en double hélice de l'ADN par James Watson et Francis Crick, un an auparavant. Ses membres cherchaient à déchiffrer, par des méthodes théoriques, le code génétique qui définissait la correspondance entre les séquences d'ADN et celles des protéines. Malgré ses efforts, ce groupe de chercheurs prestigieux allait échouer dans son entreprise. En 1961, en revanche, deux biochimistes parviendront, par des méthodes expérimentales, à déterminer la première règle de correspondance du code génétique. En l'espace de cinq ans, la quasi-totalité du code est déchiffrée de cette manière[17].

L'intérêt pour le code génétique, ainsi que le développement de nouvelles techniques biochimiques, avait stimulé les recherches expérimentales sur les séquences de protéines, après le premier succès obtenu par Frederick Sanger, en 1951, avec l'insuline[18]. La méthode utilisée par les biochimistes consistait à couper la chaîne de protéines en petits fragments et à en analyser les séquences. Il fallait ensuite assembler ces différents éléments dans le bon ordre afin de reconstituer la séquence complète de la protéine d'origine. Ce problème se prêtait particulièrement bien à une résolution par un ordinateur. Margaret O. Dayhoff et Robert S. Ledley collaboreront pour développer des algorithmes qui puissent aider les chercheurs

dans cette tâche[19]. Il s'agira de la première confrontation de Dayhoff avec le domaine de la recherche sur les protéines.

Margaret O. Dayhoff avait obtenu un doctorat de chimie sous la direction de George Kimball, à Columbia University[20]. Dans sa thèse, elle avait développé une méthode permettant d'utiliser des machines à calculer pour déterminer des paramètres chimiques (les énergies de résonance) de petites molécules. Après avoir étudié des problèmes liés à l'évolution de l'atmosphère primitive sur la terre, elle commence, e n 1960, à travailler à la NBRF. Elle deviendra professeur de physiologie et de biophysique à Georgetown University (Washington D.C.) et présidente de la Biophysical Society.

Figure 1. Modèle de protéine avec, en arrière-plan, les résultats des calculs ayant permis d'en déterminer la séquence. Archives NBRF, 1962.

L'intérêt de Margaret O. Dayhoff pour l'évolution chimique, ainsi que pour la détermination des séquences de protéines l'a certainement amenée à prendre connaissance d'un article publié en 1962 par le biologiste Emil Zuckerkandl et le physicien et chimiste Linus Pauling. Les deux auteurs y exposaient comment les séquences des protéines pouvaient être interprétées comme des « documents de l'histoire évolutive » des organismes[21]. En comparant les séquences d'une même protéine chez différentes espèces, il devait être possible d'estimer depuis combien de temps deux espèces s'étaient formées. En effet, les séquences de protéines s'apparentaient à des « horloges moléculaires » : les modifications des séquences sous l'effet des mutations marquaient le temps écoulé. La comparaison des séquences permettait également d'aborder la question des mécanismes de l'évolution, une question très débattue après la présentation de la théorie neutraliste de l'évolution en 1968 [22].

L'approche proposée par Emil Zuckerkandl et Linus Pauling exigeait de comparer deux séquences pour en déterminer les différences. Ce problème se prêtait aussi bien au traitement par un ordinateur à condition de développer un algorithme approprié[23]. Toutefois, afin d'aboutir à des résultats permettant de reconstituer l'histoire évolutive des organismes, il était également nécessaire de disposer de

la séquence d'une même protéine chez de nombreuses espèces différentes. De plus, ces séquences devaient être présentées dans un format identique afin de permettre un traitement uniforme par un ordinateur. Ces exigences ont conduit Margaret O. Dayhoff et ses collaborateurs à rechercher de manière systématique dans les journaux scientifiques toutes les séquences parues jusqu'alors. Le fruit de ce travail ardu est publié en 1965, dans la première édition de l' *Atlas of Protein Sequence and Structure*, qui rassemble soixante-dix séquences de protéines provenant de divers organismes. Cette monographie comprend en outre les résultats de comparaisons de séquences effectuées par Dayhoff à l'aide d'un programme informatique qu'elle avait mis au point, et les conclusions phylogénétiques qu'elle en avait tirées. Charles Darwin avait dû voyager plusieurs années sur le *Beagle* pour collectionner les observations d'organismes qui lui permettraient de développer sa théorie de la sélection naturelle. Un siècle plus tard, il suffisait de passer quelques mois dans les bibliothèques pour collectionner les séquences moléculaires permettant de comprendre l'histoire évolutive du vivant.

Figure 2. Margaret O. Dayhoff, autour de 1952. Archives NBRF.

L' *Atlas* présente également une innovation discrète, mais qui allait jouer un rôle important par la suite. En effet, afin de faciliter le traitement des séquences par un programme informatique, Dayhoff a utilisé une nouvelle nomenclature pour les vingt acides aminés qui composent les protéines. Au lieu des abréviations de trois lettres couramment utilisées, elle a adopté un système désignant les acides aminés par une seule lettre, ce qui permettait de diminuer les ressources informatiques nécessaires au traitement de cette information. L'acide aminé alanine par exemple, abrégé « ala », était simplement remplacé par la lettre « a ». Après quelques résistances initiales, ce système s'est finalement imposé, notamment en raison de son avantage pour un traitement informatisé des séquences, mais sans supplanter le système précédent, plus mnémotechnique[24].

Il est important de préciser le rôle joué par l'ordinateur dans le développement de l' *Atlas*. Sa fonction première était de préserver les séquences sur un support permettant de les imprimer sans devoir les saisir dans une machine à photocomposition. En effet, chaque étape manuelle était susceptible d'introduire des erreurs dans les séquences, ce qui pouvait rendre leur comparaison insignifiante. Aussi, avec les

cartes perforées, on pouvait à la fois conserver les données, les reproduire à l'identique lors des mises à jour de l' *Atlas* et les introduire dans un ordinateur puissant qui effectuerait les calculs nécessaires aux analyses phylogénétiques.

Dans les années 1960, l'ordinateur n'était pas utilisé pour distribuer les données récoltées. En effet, la plupart des biologistes et des chimistes qui s'intéressaient aux séquences de protéines ne possédaient pas d'ordinateur[25]. Dans la seconde moitié des années 1960, les biologistes avaient parfois accès à de gros ordinateurs, le plus souvent logés dans des départements de physique. Toutefois, travailler avec ces machines demandait encore un investissement considérable et ne faisait pas partie de la culture de la recherche en biologie, ce qui explique également que les rédacteurs de l' *Atlas* l'aient distribué sur un support imprimé[26]. La révolution micro-informatique, qui introduisit des ordinateurs « personnels » dans les laboratoires au début des années 1980, permit d'envisager de nouveaux modes de distribution[27].

Figure 3. Carte perforée, moyen de stockage des programmes et des données utilisé pour les ordinateurs jusqu'au début des années 1970.

L' *Atlas* a été mis à jour moins d'une année après sa parution initiale (1965), puis à plusieurs reprises depuis. Le principal obstacle rencontré par Margaret O. Dayhoff et ses collaborateurs dans cette tâche était l'augmentation vertigineuse des séquences disponibles, et par conséquent la somme de travail nécessaire pour les rassembler. Alors que la première édition de l' *Atlas* comportait moins de cent références bibliographiques à des séquences publiées, sept ans plus tard, elle en réunissait plus de mille[28]. Ce problème pratique allait prendre une ampleur telle que le succès de l'entreprise en était menacé. Trouver une solution dépassait les capacités du seul laboratoire de Dayhoff et nécessitait la participation active de tous les chercheurs impliqués dans le séquençage de protéines. Aussi, dès la première édition de l' *Atlas*, les rédacteurs ont instamment prié les chercheurs de leur soumettre toute nouvelle séquence, même incomplète et non publiée, qu'ils auraient déterminée. Ce projet impliquait de mobiliser un grand nombre de chercheurs dans une entreprise scientifique collective. En effet, de 160 auteurs dans l'édition initiale de l' *Atlas*, on passe à près de 2 500 auteurs en 1972 [29]. Le projet managérial consistant à discipliner ce collectif d'auteurs allait se heurter à des difficultés considérables, qui révèlent combien l'entreprise de Dayhoff remettait en question plusieurs idées

fondamentales et tacitement acceptées sur la nature de l'activité scientifique expérimentale, le statut de l'auteur et la propriété du savoir. Ces notions, élaborées dans le contexte d'une science pratiquée individuellement au laboratoire, s'accommodaient mal de leur transposition dans un espace plus vaste réunissant un collectif de chercheurs.

La production du savoir dans un espace collectif

Pour récolter les données nécessaires à la publication de l' *Atlas*, Margaret O. Dayhoff et ses collaborateurs se sont appuyés sur une économie du don et de la réciprocité[30]. En effet, le premier volume de l' *Atlas* a été envoyé gracieusement à tous les auteurs des séquences publiées. Dans une lettre d'introduction, Dayhoff insistait pour qu'à l'avenir ces chercheurs lui envoient directement les séquences, avant même leur publication dans un journal. Dayhoff a également offert l' *Atlas* à de nombreuses personnalités, comme Melvin Calvin, John Kendrew, Max Perutz, Joshua Lederberg ou Richard Synge, tous récipiendaires du prix Nobel[31]. S'ils n'étaient pas directement impliqués dans des projets de séquençage de protéines, ils pouvaient néanmoins apporter un soutien précieux pour légitimer le projet dans son ensemble, notamment envers les agences de financement de la recherche. Plus de cinq cents copies de la première édition de l' *Atlas* ont ainsi été distribuées gratuitement. L'édition suivante, imprimée à trois mille exemplaires, a été proposée à la vente et immédiatement épuisée[32].

Les réactions des chercheurs à la publication de cet ouvrage ont été le plus souvent enthousiastes. Ils ont souligné l'immense valeur du travail entrepris et son utilité pour la communauté scientifique. En effet, l' *Atlas* constituait non seulement un ouvrage de référence pratique, mais également un outil de travail pour les chercheurs impliqués dans le séquençage de protéines, dans l'étude de leur fonction ou dans la reconstruction de l'histoire évolutive du vivant. Plusieurs chercheurs ont immédiatement décidé d'envoyer à la rédaction de l' *Atlas* des informations encore non publiées sur des séquences qu'ils étaient en train de déterminer. Toutefois, l'appel de Margaret O. Dayhoff ne rencontra pas entièrement le succès escompté. Certains chercheurs se sont vraisemblablement interrogés sur l'opportunité de rendre publiques des informations qui n'étaient pas encore publiées dans un journal scientifique. En effet, les rédacteurs de l' *Atlas* avaient prévenu leurs lecteurs et leurs contributeurs futurs qu'ils ne voulaient « pas être impliqués dans des questions de priorité [33] ». Or, la raison première qui amenait les chercheurs à publier était précisément l'établissement de la priorité donnant droit au statut d'auteur et au crédit scientifique qui lui est attaché. En refusant

d'établir la priorité pour la détermination d'une séquence, les rédacteurs de l' *Atlas* se coupaient du principal mécanisme qui incitait

les chercheurs à rendre publics les résultats de leur travail[34]. En effet, le dévoilement d'une séquence incomplète dans l' *Atlas*, avant sa publication dans un journal scientifique, non seulement ne donnait pas droit au crédit scientifique, mais pouvait conférer un avantage compétitif à d'autres chercheurs qui tentaient de séquencer la même protéine. Le système mis en place par Dayhoff allait donc à l'encontre de pratiques et de normes communément admises au sein des communautés scientifiques auxquelles elle s'adressait.

Dans les sciences biologiques expérimentales de la seconde moitié du xxe siècle, la connaissance scientifique, tant qu'elle n'était pas publiée, appartenait à celui qui l'avait produite. Les réactions indignées de nombreux chercheurs à la publication par James Watson du livre *The Double Helix*, un récit dans lequel il révèle de manière candide comment il a utilisé des données non publiées d'une collègue, la cristallographe Rosalind Franklin, pour le travail qui l'a rendu célèbre, témoignent de la force de cette norme morale. Grete Kellenberger-Gujer, une généticienne des phages qui connaissait la plupart des protagonistes mentionnés par James Watson, décrit de manière lapidaire ce qu'elle a ressenti lors de sa lecture. Elle écrit : « Ce livre m'a rendue triste, c'est comme de la pornographie : tout est vrai, mais quel manque de goût, quelle laideur[35]. » Le biologiste Richard C. Lewontin, dans son compte rendu du livre, reconnaissait que ces transgressions étaient devenues communes, même si elles demeuraient condamnables :

> Ce que chaque scientifique sait, mais que peu sont prêts à admettre, c'est que la condition d'un grand succès est une grande ambition. En outre, l'ambition est celle d'un triomphe personnel sur d'autres hommes, pas seulement sur la nature. La science est un genre d'activité compétitive et agressive, une compétition des hommes contre d'autres hommes qui produit du savoir comme effet secondaire[36].

Les rédacteurs de l' *Atlas* ont été confrontés à d'autres défis pour mener à bien leur projet. La première partie du travail lié à l' *Atlas* était soutenue par le National Institute of Health (NIH) dans le cadre d'un subside destiné à développer l'usage de l'ordinateur pour la détermination et la comparaison de séquences. Toutefois, si cette activité relevait clairement de la recherche scientifique, la mise à jour de l' *Atlas* avait un statut plus ambigu. En effet, pour certains, la compilation bibliographique, son organisation et sa publication s'apparentaient plus à une entreprise éditoriale qu'à une véritable recherche scientifique. Aussi, le financement de l' *Atlas* a toujours été

inclus dans divers autres projets, où cet ouvrage figurait comme une ressource indispensable à la recherche sur la structure et la fonction des protéines ou sur l'histoire évolutive. Toutefois, au fur et à mesure que la part du budget dévolu à la mise à jour de l' *Atlas* augmentait, le NIH devenait de plus en plus hésitant à soutenir le projet. Plusieurs réductions de budget, des transferts dans d'autres programmes, notamment la National Library of Medicine, et plusieurs refus de poursuivre le financement du projet témoignent des difficultés rencontrées[37]. En prévision de ces dernières, Robert S. Ledley et Margaret O. Dayhoff ont diversifié leurs sources de soutien, faisant appel non seulement au NIH, mais également à la National Science Foundation (NSF) et à la National Aeronautics and Space Administration (NASA), qui s'intéressait à la possibilité d'une vie extraterrestre, et donc à l'étude des formes les plus simples de l'évolution moléculaire[38].

Le soutien chancelant dont Margaret O. Dayhoff bénéficiait pour son *Atlas* a eu des conséquences significatives sur son développement à long terme. En effet, pour couvrir les frais de publication, les rédacteurs de l' *Atlas* ont protégé leur ouvrage par un *copyright* et l'ont proposé à la vente. Le prix modeste, de 3,50 dollars pour l'édition de 1966, puis de 6 dollars pour celle de 1967-1968, avait toutefois une portée symbolique qu'il ne faut pas sous-estimer. En effet, l' *Atlas* représentait une compilation de données déjà publiées dans des journaux scientifiques. Le fait de vendre ces données, y compris à leurs auteurs, pouvait apparaître comme une appropriation illégitime d'un savoir relevant du domaine public. Un chercheur faisait remarquer à Margaret O. Dayhoff : « Vous êtes en quelque sorte dans la position d'un collectionneur de chansons populaires qui dépose un *copyright* sur le matériel qu'il publie ; dois-je le payer si je chante *John Henry* [39] ? » D'autres signes montrent que la vente de l' *Atlas* par une fondation privée provoquait un malaise chez certains scientifiques. Quinze ans plus tard par exemple, lorsqu'un physicien tenta de constituer une base de données d'ADN concurrente de celle de Margaret Dayhoff, il fit remarquer :

> Il est important que la communauté des spécialistes de biologie moléculaire [...] soit convaincue que nous fournirons un accès libre et gratuit aux informations et aux programmes que nous collecterons [...] puisque notre principal concurrent [Margaret O. Dayhoff] est de plus en plus empêtré dans des arrangements propriétaires[40].

Un autre signe témoigne du fait que les données contenues dans l' *Atlas* était considérées comme relevant du domaine public. Lorsque le contenu de l' *Atlas* est proposé sur des bandes magnétiques, plusieurs chercheurs le reprennent intégralement pour constituer leur propre base de données, sans que l'équipe de Dayhoff objecte à cette pratique

[41].

Le travail fourni par Margaret O. Dayhoff pour organiser, développer et enrichir l' *Atlas* correspondait mal à l'idée communément admise d'une contribution scientifique individuelle. Lorsque Dayhoff a posé sa candidature pour devenir membre de l'American Society of Biological Chemists, le biochimiste John T. Edsall lui a répondu de manière un peu embarrassée :

> Personnellement, je crois que vous êtes le genre de personne qui devrait devenir membre de l'American Society of Biological Chemists [...] mais connaissant la politique générale du Comité de nomination, je dois ajouter que j'ai quelques doutes sur les chances de votre élection. L'élection est presque immanquablement basée sur [...] des travaux récents publiés par le candidat qui démontrent qu'il ou qu'elle a effectué des recherches qui sont clairement les siennes. La compilation de l' *Atlas of Protein Sequence and Structure* ne rentre guère dans cette catégorie [42].

Les remarques de John T. Edsall, tout comme les hésitations des agences de financement, témoignent du statut incertain de l' *Atlas* et de l'entreprise scientifique qui avait conduit à sa création.

La difficulté d'obtenir une reconnaissance proprement scientifique pour la création de l' *Atlas* permet également de comprendre son mode de distribution. Alors qu'une version imprimée de l' *Atlas* suffisait pour la plupart des utilisations, concernant des recherches sur l'histoire évolutive, où un grand nombre de séquences devaient être comparées, il était indispensable de disposer des séquences sur un support permettant un traitement informatisé (des cartes perforées, puis des bandes magnétiques). Aussi, on conçoit que les rédacteurs de l' *Atlas* n'aient pas été particulièrement désireux de diffuser les données sur ces supports. En effet, les rédacteurs de l' *Atlas* auraient perdu le seul moyen dont ils disposaient pour obtenir un crédit scientifique récompensant leur travail de compilation des séquences disponibles. Grâce à leur collection de séquences, Dayhoff et ses collaborateurs ont produit des connaissances sur l'histoire évolutive des organismes, qui ont été publiées soit dans l' *Atlas*, soit dans des revues scientifiques. Toutefois, cette autre pratique d'appropriation du savoir s'opposait à nouveau à la norme implicite qui faisait des connaissances publiées un bien public.

Les arbres phylogénétiques publiés par Margaret O. Dayhoff et ses collaborateurs se distinguaient des travaux empiriques sur lesquels ils étaient établis par le fait qu'ils résultaient d'une activité essentiellement théorique pratiquée à l'aide d'un ordinateur. Or, si l'activité théorique bénéficiait d'un statut respectable en physique et

dans certains domaines des sciences de la vie, comme la biologie évolutionniste, il n'en allait pas de même en biologie moléculaire, où l'expérimentation était particulièrement valorisée, car elle distinguait précisément la « biologie moderne » des pratiques de l'histoire naturelle et de la spéculation philosophique. La plupart des chercheurs impliqués dans la détermination de séquences s'inscrivaient dans la tradition biochimique qui valorisait tout particulièrement l'expérimentalisme. Le biochimiste Frederick Sanger, par exemple, qui avait déterminé la première séquence d'une protéine, résumait bien plus tard : « "faire" pour un scientifique, implique faire des expériences[43] ». On comprend mieux les tentatives répétées des rédacteurs de l' *Atlas* pour inscrire leur travail dans cette tradition. La préface de chacune des éditions de cet ouvrage débute immanquablement par : « Cet *Atlas* illustre de manière imposante le triomphe des techniques expérimentales sur les secrets de la nature [44]. »

Les résultats obtenus à partir de l' *Atlas* se distinguaient de ceux obtenus dans les laboratoires, non seulement par le fait qu'ils étaient essentiellement de nature théorique, mais aussi par la figure de l'auteur scientifique qu'ils invoquaient. En effet, comme nous l'avons vu, le système de distribution du crédit scientifique reposait sur la notion de l'auteur individuel[45]. Toutefois, les résultats scientifiques obtenus par Margaret O. Dayhoff et ses collaborateurs représentaient de manière tellement évidente le produit d'une activité collective, reposant sur des données établies par les centaines de chercheurs qui avaient déterminé les séquences, qu'il était difficile d'entretenir l'illusion d'un auteur scientifique individuel. Le statut des recherches scientifiques conduites dans les laboratoires de la *Big Science*, comme ceux de la physique des hautes énergies, avait posé exactement le même problème. En effet, les résultats étaient obtenus par des équipes comprenant des centaines de chercheurs, et la distribution du crédit scientifique entre eux posait des problèmes inextricables. Aussi, dans cette communauté scientifique particulière, de nouvelles normes d'attribution du crédit scientifique ont été élaborées, qui s'appuyaient sur une notion originale d'auteur collectif[46]. L'absence de cette notion dans les sciences de la vie mettait les travaux tels ceux de Dayhoff dans une position incommode. Comme s'ils cherchaient à prévenir cette objection, les rédacteurs de l' *Atlas* rappelaient d'ailleurs dans leurs préfaces que « certaines découvertes ne peuvent être attribuées à aucun chercheur ou école en particulier[47] ».

Ces conflits étaient étroitement liés aux lieux où était produite la connaissance scientifique[48]. La notion d'auteur scientifique individuel, qui s'incarne dans la figure de Newton en génie solitaire, est associée au travail du philosophe naturel reclus dans son cabinet ou dans son laboratoire[49]. Le savoir produit dans ces lieux particuliers circulait vers d'autres lieux, où des chercheurs pouvaient tenter de le

reproduire[50]. La circulation du savoir se caractérise alors par un mouvement du centre vers la périphérie[51]. À l'inverse, les entreprises de production du savoir liées aux grandes collections comme l' *Atlas* reposaient sur la coordination de nombreux lieux périphériques, dont les résultats convergeaient vers un espace central. Toutefois, alors que la constitution des grandes collections du xixe siècle, comme celles du British Museum, pouvait s'appuyer sur le pouvoir et l'étendue d'un empire, la création des bases de séquences devait reposer sur la participation volontaire des différents laboratoires où étaient séquencées les protéines. Et cette participation ne pouvait se passer de discipline, comme le montre, par exemple, la nécessité de standardiser les notations des séquences. Mais surtout, si les musées et les jardins constituaient depuis le xviie siècle des lieux légitimes de production du savoir dans la tradition de l'histoire naturelle, les bases de séquences développées au milieu du xxe siècle se trouvaient en porte à faux avec les valeurs morales, les pratiques sociales et les normes épistémiques qui prédominaient alors dans la tradition de la recherche expérimentale.

L'alignement sur l'économie morale des sciences expérimentales

La notion d'économie morale a été développée par l'historien E. P. Thompson soucieux de comprendre les révoltes frumentaires en Angleterre au xviiie siècle. Au lieu d'analyser ces dernières en termes de psychologie des foules ou en termes strictement économiques, sa lecture repose sur le sentiment d'injustice provoqué par la hausse injustifiée des prix du blé dans le contexte des valeurs morales qui sous-tendaient jusqu'alors les échanges marchands. L'ensemble de ces valeurs et leurs relations constituent une économie morale, où le mot *économie* prend son sens ancien de « structure ». La notion d'économie morale a été utilisée dans diverses études sur les sciences avec des sens assez différents[52]. Il est essentiel de relever que les économies morales ne sont pas nécessairement identiques pour toutes les sciences, mais peuvent largement varier en fonction des communautés scientifiques ou des lieux de science considérés. Robert Kohler, dans son étude sur la communauté de chercheurs travaillant sur la génétique de la drosophile au début du xxe siècle, a souligné l'importance de trois éléments pour l'économie morale de la science : l'accès aux ressources matérielles, l'équité dans la distribution du crédit et l'autorité pour décider de la valeur des contributions

intellectuelles.

Si les bases de séquences sont finalement parvenues à s'établir dans les sciences de la vie, c'est grâce à plusieurs transformations qui ont permis leur intégration à l'économie morale des sciences expérimentales. L'examen de la constitution de nouvelles bases de séquences permet de mettre en évidence trois transformations cruciales. Premièrement, au lieu de reposer sur une économie du don désintéressé, comme l'avait envisagé Margaret O. Dayhoff, la participation des différents laboratoires a été assurée grâce au crédit scientifique octroyé par les journaux scientifiques. En 1980, le Nih a mis au concours un subside pour la création et le maintien d'une base de séquences nationale d'ADN [53]. Deux groupes ont déposé une candidature : Margaret O. Dayhoff et le physicien Walter Goad de Los Alamos. Un des avantages décisifs du projet de Walter Goad, qui a finalement remporté le concours, a été le système qu'il a mis en place pour s'assurer la collaboration des laboratoires qui séquençaient de l'Adn. Il a négocié avec les éditeurs des vingt journaux les plus établis dans le domaine pour qu'ils exigent des auteurs qu'ils soumettent leurs séquences dans la base de données de Los Alamos avant de pouvoir publier un article[54]. Les éditeurs évitaient ainsi de devoir imprimer de longues séquences dans leurs pages, puisqu'elles étaient disponibles dans une base de données publique, et les auteurs, en accédant à la publication dans un journal, étaient récompensés pour avoir mis immédiatement à disposition les résultats de leur travail de séquençage. Un éditorial de la revue *Proceedings of the National Academy of Science* annonçant l'obligation pour les auteurs de soumettre leurs séquences dans une base de données publique reconnaissait que cette mesure était dans l'intérêt des chercheurs, tout en masquant l'enjeu réel pour les chercheurs :

> les scientifiques qui génèrent des séquences [...] utilisent aussi les informations disponibles sur les séquences. Ce serait donc servir leur propre intérêt que de respecter les règles[55].

Une deuxième transformation a permis d'assurer la coopération des chercheurs impliqués dans des activités de séquençage. La soumission de séquences dans une base de données avant la publication d'un article pouvait conférer, comme nous l'avons vu, un avantage à un groupe concurrent. Aussi, plusieurs bases de données ont prévu une clause de confidentialité permettant de maintenir les séquences déposées hors d'accès du public, jusqu'à ce que l'article qui s'y référait soit publié. Un système similaire a été adopté pour d'autres bases de données, comme la Protein Data Bank comprenant les structures tridimensionnelles de protéines, qui prévoit un maintien des informations hors du domaine public jusqu'à une année après leur dépôt[56].

La troisième transformation qui a permis la généralisation des bases de séquences a reposé sur une modification de l'espace dans lequel était produit le savoir. Les premières bases de séquences comme l' *Atlas* constituaient toujours des lieux centralisés de production du savoir, puisque la distribution des séquences sur un support imprimé ne permettait pas encore d'en exploiter toutes les potentialités. Dès 1973, la diffusion des informations contenues dans l' *Atlas* sur des bandes magnétiques allait contribuer à équilibrer quelque peu la situation, même si la vente de ces dernières est toutefois restée limitée. Le cinquième volume de l' *Atlas*, par exemple, a été vendu à 1 765 exemplaires alors que seulement 7 exemplaires des bandes magnétiques ont été distribués[57]. Ce n'est qu'à partir de 1982, lorsque toutes les séquences d'ADN déposées dans la base de données de Los Alamos deviennent accessibles à une grande partie de la communauté scientifique, par l'intermédiaire du réseau Arpanet (un des précurseurs de l'Internet), que la production du savoir s'inscrit dans un espace virtuel global. Dans ce nouvel espace, les notions de centre et de périphérie perdent de leur pertinence, puisque chacun peut accéder à l'intégralité de la base de séquences depuis le lieu où il travaille physiquement, généralement un laboratoire.

L'histoire des bases de séquences témoigne de l'avènement d'un mode de production du savoir distinct de la tradition expérimentaliste, au moment même où cette dernière connaît un essor sans précédent dans les sciences de la vie. Les difficultés auxquelles ont été confrontés les promoteurs des bases de données provenaient en partie du lieu dans lequel s'inscrivait leur activité. Alors que la recherche expérimentale avait fait du laboratoire et du chercheur individuel les éléments constitutifs de toute recherche scientifique, les auteurs des bases de séquences proposaient une organisation collective et délocalisée du travail scientifique remettant en question la notion même d'auteur individuel. On peut comprendre cette tension comme résultant de différences entre des économies morales, autrement dit, comme des conflits au sujet de valeurs essentielles qui fondent l'échange des biens, l'attribution du crédit, et la définition des buts méritant d'être poursuivis. La résolution de cette tension est passée par l'abandon de plusieurs valeurs sur lesquelles reposaient les premiers projets de bases de séquences et l'adoption de celles en vigueur dans les sciences expérimentales.

Notes

1. Sur l'histoire de la biologie moléculaire, voir Morange, 2003.

2. L'émergence de la biologie moléculaire participe d'un processus plus large, Chadarevian et Kamminga, 1998 ; Gaudillière, 2002 ; Strasser et Chadarevian, 2002.

3. Sur ce point, voir Wilson, 2000, chap. 14.

4. On se souviendra de la remarque de Lord Thompson, que nombre de physiciens convertis à la biologie moléculaire aiment rappeler : « En matière de science, soit on fait de la physique, soit on est collectionneur de timbres » (Birks, 1962, p. 108).

5 . Sur la formation disciplinaire de la biologie moléculaire, voir Strasser, 2006.

6. Benson *et al.*, 1999.

7. Sur les modes de production du savoir, voir Pickstone, 2000, et sur les régimes de production des savoirs, Pestre, 2003.

8. Voir Galison et Hevly, 1992.

9 . À propos d'une autre tradition d'histoire naturelle en biologie moléculaire, voir Chadarevian, 1998, et sur l'histoire naturelle et la biologie expérimentale autour de 1900, Nyhart, 1996.

10. Allen, 1986.

11. Cf. le constat de Derek de Solla Price, en 1963.

12. Sur la recherche opérationnelle, voir Fortun et Schweber, 1993.

13. Weinrauch et Pattishall, 1958. L'US Air Force avait développé une expertise considérable dans le domaine des ordinateurs et des réseaux informatiques, notamment pour le programme de surveillance aérienne Sage. Voir Campbell-Kelly et Aspray, 2004.

14. Ledley, 1965.

15. Joe A. November, *Rise of the Digital Organism : How Computers Changed Biology*, thèse de doctorat soutenue à Princeton University en 2006.

16. Georges Gamow à James Watson, 6 décembre 1954, reproduit in Watson, 2001, annexe 12.

17. Kay, 2000.

18. Sanger, 1988.

19. Dayhoff, 1964.

20. Hunt, 1984.

21. Zuckerkandl et Pauling, 1962 ; Zuckerkandl et Pauling, 1965.

22. Dietrich, 1994 et 1998.

23. Sur la phylogénie moléculaire, voir Hagen, 1999.

24. IUPAC-Iub, Commission On Biochemical Nomenclature, 1968.

25. Les ordinateurs étaient plus communs chez les physiciens ou les chimistes impliqués dans la recherche cristallographique, voir Chadarevian, 2002, chap. 4.

26. Sur l'utilisation précoce des ordinateurs en systématique, voir Hagen, 2001.

27. Sur les premiers ordinateurs personnels et leur utilisation en biologie, voir November, 2004.

28. « Comprehensive Progress Report », 23 août 1973 (Archives NBRF, LM 01206).

29. « Comprehensive Progress Report », 23 août 1973 (Archives NBRF, LM 01206).

30. Sur l'économie du don, voir Mauss, 1923-1924, et Bourdieu, 1972, et sur l'utilisation de cette notion dans l'analyse des communautés scientifiques, voir Hagstrom, 1982 ; Biagioli, 1993 ; Findlen, 1994.

31. Correspondance (Archives NBRF).
32. Margareth O. Dayhoff à Gordon B. Ward, 30 avril 1978 (Archives NBRF).
33. Dayhoff *et al.*, 1965, p. xiv.
34. Sur ce point, Hagstrom, 1982.
35. Grete Kellenberger-Gujer à Max Delbrück, 17 mars 1969 (Archives du California Institute of Technology).
36. R. C. Lewontin, *Chicago Sunday*, 25 février 1968, p. 1-2, reproduit in Watson, 1968.
37. Archives NBRF, LM 01206-04 B.
38. Wolfe, 2002 ; Strick, 2004.
39. Burton S. Guttman à Margaret O. Dayhoff, 10 juin 1968 (Archives NBRF). *John Henry* est le titre d'une chanson populaire célèbre aux États-Unis qui illustre le combat d'un jeune homme contre une machine lors du forage d'un tunnel ferroviaire.
40. Walter Goad à Peter Carruthers et Milton Slaughter, 3 novembre 1981 (Archives American Philosophical Society).
41. Doolittle, 1980.
42. John T. Edsall à Margaret O. Dayhoff, 4 novembre 1969 (Archives NBRF).
43. Sanger, 1988, p. 1.
44. Dayhoff *et al.*, 1965, p. 1.
45. La notion d'auteur scientifique est fondamentalement paradoxale, puisqu'un fait scientifique n'a, par définition, pas d'auteur, sinon il s'agit d'un artefact et non d'un fait de nature. Pour une étude sociologique de la signature scientifique, voir Pontille, 2004.
46. Biagioli, 2003 ; Galison, 2003.
47. Dayhoff *et al.*, 1965, p. 1.
48. Voir par exemple Livingstone, 2003 ; Smith et Agar, 1998 ; Ophir et Shapin, 1991. Dans le domaine de l'histoire de la médecine, selon le travail de Michel Foucault, la problématique des lieux occupe une place essentielle.
49. Fara, 2002, chap. 6. Sur la solitude comme condition du savoir, voir Shapin, 1990.
50. Sur les problèmes de la réplication expérimentale, voir Collins, 1985.
51. Sur la problématique de la circulation du savoir, voir Latour, 1987.
52. Voir en particulier Kohler, 1994 ; Daston, 1995, et Porter, 1995.
53. Lewin, 1982.
54. Walter Goad à Richard Roberts, 17 mars 1983 (Archives de l'American Philosophical Society).
55. Dawid, 1989.
56. Dans le contexte du Projet génome humain, voir Hilgartner, 1998.
57. « Comprehensive Progress Report », 23 août 1973 (Archives NBRF, LM 01206).

Bibliographie

- Allen, 1986 : Garland E. Allen, « The Eugenics Record Office at Cold Spring Harbor, 1910-1940 : an Essay on Institutional History », *Osiris*, 2, p. 225-264.
- Benson *et al.*, 1999 : Dennis A. Benson *et al.*, « Genbank », *Nucleaic Acids Research*, 27, p. 12-17.
- Biagioli, 1993 : Mario Biagioli, *Galileo, Courtier. The Practice of Science in the Culture of Absolutism*, Chicago-Londres.
- Biagioli, 2003 : M. Biagioli, « Rights or Rewards ? Changing Frameworks of Scientific Authorship », in M. Biagioli et P. Galison (éd.), *Scientific Authorship. Credit and Intellectual Property in Science* , New York, p. 253-279.
- Birks, 1962 : John Betteley Birks (éd.), *Rutherford at Manchester*, Londres.
- Bourdieu, 1972 : Pierre Bourdieu, *Esquisse d'une théorie de la pratique* , Paris ; rééd. 2000.
- Campbell-Kelly et Aspray, 1996 : Martin Campbell-Kelly et William Aspray, *Computer. A History of the Information Machine* , Boulder ; rééd. 2004.
- Chadarevian, 1998 : Soraya De Chadarevian, « Following Molecules : Haemoglobin between the Clinic and the Laboratory », in Chadarevian et Kamminga, 1998, p. 171-201.
- Chadarevian, 2002 : S. De Chadarevian, *Designs for Life. Molecular Biology after World War Ii*, Cambridge-New York.
- Chadarevian et Kamminga, 1998 : Soraya De Chadarevian et Harmke Kamminga (éd.), *Molecularizing Biology and Medicine. New Practices and Alliances, 1910s-1970s*, Amsterdam.
- Collins, 1985 : Harry M. Collins, *Changing Order. Replication and Induction in Scientific Practice*, Chicago ; rééd. 1992.
- Daston, 1995 : Lorraine Daston, « The Moral Economy of Science », *Osiris*, 10, p. 2-24.
- Dawid, 1989 : Igor B. Dawid, « Editorial Submission of Sequences », *PNAS*, 86, p. 407.
- Dayhoff, 1964 : Margaret O. Dayhoff, « Computer Aids to Protein Sequence Determination », *Journal of Theoretical Biology* , 8, p. 97-112.
- Dayhoff *et al.*, 1965 : M. O. Dayhoff *et al.*, *Atlas of Protein Sequence and Structure*, Silver Spring.
- Dietrich, 1994 : Michael R. Dietrich, « The Origins of the Neutral Theory of Molecular Evolution », *Journal of the History of Biology* , 27, p. 21-59.
- Dietrich, 1998 : M. R. Dietrich, « Paradox and Persuasion : Negotiating the Place of Molecular Evolution within Evolutionary Biology », *Journal of the History of Biology* , 31, p. 85-111.
- Doolittle, 1980 : Russell F. Doolittle, « Similar Amino Acid Sequences : Chance or Common Ancestry ? », *Science*, 214, p. 149-159.
- Fara, 2002 : Patricia Fara, *Newton : The Making of a Genius* , New York.
- Findlen, 1994 : Paula Findlen, *Possessing Nature : Museums, Collecting and Scientific Culture in Early Modern Italy*, Berkeley-Los Angeles-

Londres.
- Fortun et Schweber, 1993 : Michael Fortun et Silvan S. Schweber, « Scientists and the Legacy of World War Ii : The Case of Operations Research », *Social Studies of Science*, 23, p. 595-642.
- Galison, 2003 : Peter Louis Galison, « The Collective Author », in M. Biagioli et P. Galison (éd.), *Scientific Authorship. Credit and Intellectual Property in Science*, New York, p. 325-355.
- Galison et Hevly, 1992 : Peter Louis Galison et Bruce William Hevly (éd.), *Big Science. The Growth of Large-Scale Research*, Stanford.
- Gaudillière, 2002 : Jean-Paul Gaudillière, *Inventer la biomédecine. La France, l'Amérique et la production des savoirs du vivant (1945-1965)*, Paris.
- Ghiselin et Leviton, 2000 : Michael T. Ghiselin et Alan E. Leviton, *Cultures and Institutions of Natural History. Essays in the History and Philosophy of Science*, San Francisco.
- Grmek, 1990 : Mirko D. Grmek, *La Première Révolution biologique. Réflexions sur la physiologie et la médecine du xvii e siècle*, Paris.
- Hagen, 1999 : Joel B. Hagen, « Naturalist, Molecular Biology, and the Challenge of Molecular Evolution », *Journal of the History of Biology*, 32, p. 321-341.
- Hagen, 2001 : J. B. Hagen, « The Introduction of Computers into Systematic Research in the United States During the 1960s », *Studies in the History and Philosophy of Biological and Biomedical Sciences*, 32, p. 291-314.
- Hagstrom, 1982 : Warren Hagstrom, « Gift Giving as an Organizing Principle in Science », in B. Barnes et D. Edge (éd.), *Science in Context : Readings in the Sociology of Science,* Cambridge, p. 21-34.
- Hilgartner, 1998 : Stephen Hilgartner, « Data Access Policy in Genome Research », in A. Thackray (éd.), *Private Science. Biotechnology and the Rise of the Molecular Sciences*, Philadelphie, p. 202-217.
- Hunt, 1984 : Lois Hunt, « Margaret Oakley Dayhoff 1925-1983 », *Bulletin of Mathematical Biology*, 46, p. 467-472.
- Iupac-Iub, Commission on Biochemical Nomenclature, 1968 : International Union Of Pure And Applied Chemistry et International Union Of Biochemistry, Commission on Biochemical Nomenclature, « A One-Letter Notation for Amino Acid Secruences », *The Journal of Biological Chemistry*, 234, p. 3557-3559.
- Jardine, 1996 : Nicholas Jardine, James A. Secord et Emma C. Spary, *Cultures of Natural History*, Cambridge.
- Kay, 2000 : Lily E. Kay, *Who Wrote the Book of Life ? A History of the Genetic Code*, Stanford.
- Kohler, 1994 : Robert E. Kohler, *Lords of the Fly : Drosophila Genetics and the Experimental Life*, Chicago.
- Kohler, 2002 : R. E. Kohler, *Landscapes and Labscapes : Exploring the Lab-Field Border in Biology*, Chicago.
- Latour, 1987 : Bruno Latour, *La Science en action*, trad. de l'anglais par M. Biezunski, Paris, 1989 ; rééd. 1995.

- Ledley, 1965 : Robert S. Ledley, *Use of Computers in Biology and Medicine*, New York.
- Lewin, 1982 : Roger Lewin, « Long-Awaited Decision on DNA Database », *Science*, 217, p. 817-818.
- Livingstone, 2003 : David N. Livingstone, *Putting Science in its Place*, Chicago.
- Mauss, 1923-1924 : Marcel Mauss, « Essai sur le don : forme et raison de l'échange dans les sociétés archaïques », *Année sociologique*, nouvelle série, 1, p. 30-186.
- Morange, 2003 : Michel Morange, *Histoire de la biologie moléculaire*, Paris.
- November, 2004 : Joe November, « Linc : Biology's Revolutionary Little Computer », *Endeavour*, 28, p. 125-131.
- Nyhart, 1996 : Lynn K. Nyhart, « Natural History and the "New" Biology », in N. Jardine, J. A. Secord et E. C. Spary (éd.), *Cultures of Natural History*, Cambridge, p. 426-443.
- Ophir et Shapin, 1991 : Adi Ophir et Steven Shapin, « The Place of Knowledge. A Methodological Survey », *Science in Context*, 4, p. 3-21.
- Pestre, 1995 : Dominique Pestre, « Pour une histoire sociale et culturelle des sciences », *Annales HSS*, 50, p. 487-522.
- Pestre, 2003 : D. Pestre, *Science, argent et politique. Un essai d'interprétation*, Paris.
- Pickstone, 2000 : John V. Pickstone, *Ways of Knowing : A New History of Science, Technology and Medicine*, Manchester.
- Pontille, 2004 : D. Pontille, *La Signature scientifique. Une sociologie pragmatique de l'attribution*, Paris.
- Porter, 1995 : Theodore M. Porter, *Trust in Numbers. The Pursuit of Objectivity in Science and Public Life*, Princeton.
- Price, 1963 : Derek John de Solla Price, *Little Science, Big Science... and Beyond*, New York-Londres, rééd. 1986 ; *Science et suprascience*, traduction de l'américain par G. Lévy, Paris, 1972.
- Sanger, 1988 : Frederick Sanger, « Sequences, Sequences, and Sequences », *Annual Review of Biochemistry*, 57, p. 1-28.
- Shapin, 1990 : Steven Shapin, « "The Mind in Its Own Place" : Science and Solitude in Seventeenth-Century England », *Science in Context*, 4, p. 191-218.
- Shapin et Schaffer, 1993 : Steven Shapin et Simon Schaffer, *Léviathan et la pompe à air : Hobbes et Boyle entre science et politique* , trad. de l'anglais par Th. Piélat et S. Barjansky, Paris.
- Smith et Agar, 1998 : Crosbie Smith et Jon Agar, *Making Space for Science*, Basingstoke.
- Spary, 2000 : Emma C. Spary, *Utopia's Garden : French Natural History from Old Regime to Revolution*, Chicago.
- Strasser et Chadarevian, 2002 : Bruno J. Strasser et Soraya De Chadarevian (éd.), *Molecular Biology in Post-War Europe*, volume de *Studies in the History and Philosophy of the Biological and Biomedical Sciences*, 33C.
- Strasser, 2006 : Bruno J. Strasser, *La Fabrique d'une nouvelle science. La*

biologie moléculaire à l'âge atomique, 1945-1964, Florence.

- Strick, 2004 : James E. Strick, « Creating a Cosmic Discipline : the Crystallization and Consolidation of Exobiology, 1957–1973 », *Journal of the History of Biology* , 37, p. 131-180.
- Thompson, 1971 : Edward P. T hompson, « The Moral Economy of the English Crowd in the Eighteenth-Century », *Past and Present*, 50, p. 76-136.
- Watson, 1968 : James D. Watson, *The Double Helix. A Personal Account of the Discovery of the Structure of DNA. Text, Commentary, Reviews, Original Papers*, New York, rééd. 1980 ; *La Double hélice ou Comment fut découverte la structure de l'ADN*, trad. de l'américain par H. Joël ; rééd. 2003.
- Watson, 2001 : J. D. Watson, *Genes, Girls and Gamow*, Oxford ; *Gènes, génomes et société*, trad. de l'américain par J. Mouchard, 2003.
- Weinrauch et Pattishall, 1958 : Harry Weinrauch et Evan Pattishall, « The Air Force and the Application of Computers to Medicine and Biology », *Military Medicine*, 122, p. 178-180.
- Wilson, 2000 : Edward O. Wilson, *Naturaliste*, trad. de l'américain par C. Chichereau, Paris.
- Wolfe, 2002 : Audra J. Wolfe, « Germs in Space. Joshua Lederberg, Exobiology and the Public Imagination, 1958-1964 », *Isis*, 93, p. 183-205.
- Zuckerkandl et Pauling, 1962 : Emil Zuckerkandl et Linus Pauling, « Molecular Disease, Evolution and Genic Heterogeneity », in M. Kasha et B. Pullman (éd.), *Horizons in Biochemistry*, New York, p. 189-224.
- Zuckerkandl et Pauling, 1965 : E. Zuckerkandl et L. Pauling, « Molecules as Documents of Evolutionary History », *Journal of Theoretical Biology*, 8, p. 357-366.

Nos partenaires

Le projet *Savoirs* est soutenu par plusieurs institutions qui lui apportent des financements, des expertises techniques et des compétences professionnelles dans les domaines de l'édition, du développement informatique, de la bibliothéconomie et des sciences de la documentation. Ces partenaires contribuent à la réflexion stratégique sur l'évolution du projet et à sa construction. Merci à eux !

L'ECOLE DES HAUTES ETUDES EN SCIENCES SOCIALES

haStec Laboratoire d'Excellence Histoire et anthropologie des savoirs, des techniques et des croyances

fondation maison des sciences de l'homme

- CONCEPTION : ÉQUIPE *SAVOIRS*, PÔLE NUMÉRIQUE RECHERCHE ET PLATEFORME GÉOMATIQUE (EHESS).
- DÉVELOPPEMENT : DAMIEN RISTERUCCI, IMAGILE, MY SCIENCE WORK DESIGN : WAHID MENDIL.

PDN
Pôle Document Numérique
Maison de la Recherche en Sciences Humaines
CNRS - UNIVERSITÉ DE CAEN

métopes
méthodes et outils
pour l'édition structurée

EPFL

bnu
strasbourg

enssib
école nationale supérieure
des sciences de l'information
et des bibliothèques

CAK
Centre Alexandre-Koyré
Histoire des sciences et des techniques
UMR 8560 EHESS-CNRS-MNHN

ANHIMA