
EPREUVE ECRITE DE BIOLOGIE**ENS : PARIS - LYON - CACHAN**

Durée : 6 heures - Coefficients : PARIS option Biologie : 7 / option Géologie : 4
LYON option Biologie : 6 / option sciences de la terre : 4
CACHAN : 8

MEMBRES DU JURY : J.-G. BARBARA, N. CAUDRON, A. CORBIN, J.-L. GIRANTON, U. HAZAN, M.-H. KRYSZKE, I. NEGRUTIU, G. PEYROCHE, G. VOGÉ

Comme pour les sessions précédentes, l'épreuve écrite de biologie comportait, pour une durée globale de six heures, un sujet de synthèse et un sujet avec documents, avec pour thème commun, cette année, le génome. Les candidats se devaient de bien gérer leur temps de composition afin de pouvoir aborder correctement les deux sujets. La sanction est toujours sévère pour ceux qui ne traitent qu'un seul des deux sujets, ou qui, faute de temps, ne parviennent à couvrir qu'une petite partie du sujet avec documents. De manière générale, nous encourageons vivement les candidats, dans le sujet avec documents, à améliorer encore leur méthode d'analyse : nous les invitons à procéder à une rapide description des données expérimentales (type de document, paramètres importants, nature des variations observées,...), qui facilite la compréhension des phénomènes étudiés, à approfondir leur réflexion et à essayer d'en tirer un meilleur bénéfice en proposant des conclusions plus achevées.

SUJET DE SYNTHESE

Le sujet de synthèse proposé cette année était centré sur les relations entre le génotype et le phénotype, question tout à fait classique pour un étudiant en biologie. Afin d'en réduire l'étendue, il était demandé aux candidats de se limiter au cas des organismes eucaryotes, de s'intéresser à l'établissement du phénotype, ceci pour ne pas avoir à prendre en compte la grande diversité des variations phénotypiques dépendantes de l'environnement et liées aux conditions physiologiques auxquelles sont soumis les organismes tout au long de leur vie, et de traiter le sujet à l'échelle de la cellule, ce qui permettait d'accorder une place privilégiée à l'étude des interactions moléculaires qui président à la mise en place du phénotype, en évitant les développements anatomiques trop descriptifs. Les indications préliminaires apparaissant dans le libellé du sujet invitaient toutefois les candidats à se référer au besoin au tissu, à l'organe, voire à l'organisme entier, et à replacer leur exposé dans le contexte du développement des êtres vivants.

Après une évocation rapide de la diversité phénotypique des individus, aisément observable (caractères spécifiques et individuels, adaptations fonctionnelles), il était naturel de proposer une définition opérationnelle du phénotype à l'échelle de la cellule : le phénotype est élaboré à partir de l'ensemble des protéines synthétisées par la cellule, des structures et des architectures qu'elles engendrent, des activités qu'elles exercent et des produits résultant de ces activités. Le programme inclut la description de quelques exemples de cellules différenciées, avec même, dans le cas des Végétaux, la possibilité d'une comparaison avec la cellule méristématique. Ce point de départ a été judicieusement utilisé dans les devoirs, par un nombre, hélas trop restreint, de candidats.

A partir de là, pour traiter de l'établissement du phénotype à l'échelle de la cellule, il convenait dans un premier temps de s'interroger sur l'origine des protéines présentes dans une cellule différenciée, c'est-à-dire sur les modalités de biosynthèse des protéines à partir du génome qui les code. La relation gène - caractère est normalement bien connue des candidats, elle a été ici correctement illustrée par la plupart d'entre eux. Au-delà, l'étude des relations entre génotype et phénotype nécessitait d'expliquer les mécanismes permettant de passer des gènes aux protéines fonctionnelles

à travers les étapes de l'expression génique. Les candidats devaient s'appuyer sur leur connaissance de la structure du génome des Eucaryotes pour décrire les processus de transcription de l'ADN et de traduction des ARN messagers en protéines. Les phénomènes de maturation des transcrits primaires ont également été évoqués dans de nombreux devoirs. En revanche, peu de candidats ont pensé que certains polypeptides devaient subir des modifications post-traductionnelles pour devenir actifs.

A travers cette partie du sujet, le jury souhaitait s'assurer que les candidats possèdent une connaissance solide des mécanismes fondamentaux grâce auxquels une protéine fonctionnelle est synthétisée dans la cellule à partir du gène qui la code. Cette connaissance est essentielle, malheureusement, une fois de plus dans ce domaine, les copies se sont révélées dans l'ensemble très décevantes. Comme il l'était apparu de manière plus ponctuelle, à plusieurs occasions lors de sessions antérieures, les étapes de l'expression génique sont encore mal identifiées et mal définies par beaucoup de candidats, les connaissances restent souvent beaucoup trop approximatives, et des erreurs graves persistent, comme celle qui consiste à confondre les mécanismes de la transcription et de la répllication de l'ADN. Certains exposés révèlent de grandes lacunes, tandis que d'autres, fort détaillés, souffrent d'un manque de pertinence ou de cohérence. Il serait souhaitable, de manière générale mais tout particulièrement dans le domaine de la génétique moléculaire, que des connaissances de base réellement solides soient acquises, et que les principes fondamentaux les plus intéressants de cette discipline soient perçus par les candidats, quitte à abandonner une certaine somme de détails qu'ils rencontreront plus tard dans leur cursus et dont ils profiteront mieux à cette condition. Par exemple, on peut constater que le code génétique est toujours cité dans les devoirs de manière très stéréotypée et rébarbative, sans que l'extraordinaire performance que représentent le codage et le décodage de l'information génétique soit jamais évoquée par les candidats. Ainsi il est peut-être dommage que l'importance de l'appariement des bases soit très vite perdue de vue, que l'intérêt de l'existence d'un intermédiaire d'ARN, quantitativement modulable et transitoire, entre l'ADN génomique (dont la stabilité et les redondances garantissent la pérennité du patrimoine génétique) et les protéines ne soit jamais évoqué, et que le rôle crucial de la spécificité d'action des amino-acyl-ARNt synthétases dans la fiabilité du décodage lors de la traduction des ARN messagers soit totalement occulté.

Un organisme est constitué d'un grand nombre de types cellulaires différents, dont chacun possède le même génome - celui du zygote de départ - mais pourtant synthétise un ensemble distinct de protéines caractéristiques. C'est l'expression sélective d'une partie seulement de ses gènes qui permet à une cellule d'acquérir un phénotype spécifique. Les candidats devaient être capables de conduire cette élémentaire réflexion, puis d'expliquer quels types de mécanismes sont responsables de cette expression sélective. Comme il est décrit dans la plupart des devoirs qui abordent cette partie du sujet, la régulation s'exerce principalement (mais non exclusivement) au niveau la transcription des gènes : au sein du génome, certains gènes sont activement transcrits tandis que d'autres sont silencieux. Les séquences d'ADN impliquées dans ce contrôle de l'expression génique, agissant en *cis*, et les facteurs protéiques, agissant en *trans*, grâce auxquels elles produisent leur effet ont été présentés, avec plus ou moins d'exactitude, par beaucoup de candidats. Dans les devoirs les plus intéressants, les événements intervenant en amont de l'interaction ADN / protéines régulatrices proprement dite ont été, au moins partiellement, évoqués. Ainsi, certains candidats ont bien compris et ont su montrer clairement qu'un contrôle portant sur la synthèse ou sur l'activité des facteurs de transcription, en réponse à divers stimuli, permettait la mise en place d'une régulation intégrée, dans l'espace et au cours du temps, de l'expression des gènes. Les possibilités de régulation en cascade ont été indiquées dans quelques compositions. Les signaux à l'origine de ces régulations ont été décrits (nature, réception, transduction), permettant aux candidats de relier l'expression des gènes à la physiologie de l'organisme étudié. Les notions de contrôle intrinsèque et extrinsèque intervenant au cours du développement, les communications intercellulaires par contact direct ou à distance ont été citées. D'excellents exposés sont pertinemment illustrés par une description précise des mécanismes intervenant dans la détermination des territoires embryonnaires au cours du

développement des animaux, ou dans le contrôle hormonal des métamorphoses des Insectes ou des Amphibiens.

Certains candidats, et c'est ce qui était attendu dans cette épreuve, ont donc su mobiliser des connaissances issues de diverses parties de leur programme et les intégrer avec pertinence en une présentation personnelle et bien construite. L'étroitesse du temps imparti au sujet de synthèse faisait que tous ces aspects ne pouvaient être développés de manière exhaustive. Les candidats étaient donc amenés à faire des choix dans la conception de leurs devoirs, ce qui laissait la place à la manifestation d'une relative originalité de leur part, qualité toujours appréciée par le jury, à condition que de tels choix découlent d'une réflexion intéressante et fassent l'objet d'une bonne justification. Malheureusement beaucoup de devoirs ne traitent que trop partiellement le sujet, ce qui est développé n'étant pas pour autant forcément de bonne qualité, et donnent au jury l'impression qu'ils résultent d'un manque réel de réflexion plus que d'un choix justifiable.

SUJET AVEC DOCUMENTS

Il était proposé aux candidats cette année d'explorer le génome des Eucaryotes, mettant à profit le fait qu'un coin du voile ait été levé à l'occasion du séquençage presque complet du génome humain, tout en sachant que celui-ci est encore loin de nous avoir livré par là tous ses secrets. La première partie du sujet, basée sur les observations préliminaires concernant le génome humain, tentait de faire ressortir, au fil d'une démarche encore un peu abstraite, quelques caractéristiques générales de ce génome susceptibles de jeter une lumière utile sur son fonctionnement. Les deux parties suivantes illustraient la mise en application des données de la génomique dans l'étude du contrôle de l'expression des gènes, d'une part chez un Végétal modèle, *Arabidopsis thaliana*, d'autre part, de manière plus succincte, au cours du développement chez le Nématode *Caenorhabditis elegans*. De façon globale, les candidats semblent avoir bien appréhendé ce sujet, dont ils ont la plupart du temps abordé les trois parties, même si rares demeurent ceux qui ont essayé de traiter toutes les questions. Le grappillage est resté une exception dans les copies, et c'est davantage la progression dans la difficulté des questions qui a conduit à l'abandon de chacune des parties en cours de route, plus ou moins tôt selon les cas. Le jury a apprécié l'effort fait par beaucoup de candidats pour mener le plus loin possible une réflexion cohérente sur l'ensemble des documents soumis à leur analyse. Dans les détails, certaines questions nous semblent mériter quelques commentaires particuliers.

Cartographie et séquençage du génome humain

Question 1. La courbe de la figure 1 pouvait être analysée de manière globale, et éventuellement plus finement après quelques calculs, mais c'est avant tout la logique et la cohérence de la démarche adoptée et la justification des conclusions tirées qui ont retenu l'attention des correcteurs. Certains candidats ont brillamment traité la question. La représentation de la distance génétique en fonction de la distance physique par rapport au centromère n'est pas une droite mais une courbe dont la pente de la tangente augmente, sur chaque bras du chromosome, du centromère vers le télomère. La valeur moyenne de cette pente est plus élevée sur le bras court que sur le bras long, ce qui permet d'augmenter sur le bras court la probabilité de survenue d'un enjambement au cours de la méiose, en compensant une longueur d'ADN plus faible par une fréquence de recombinaison plus importante.

Question 2. Diverses conclusions pouvaient être tirées à partir de l'allure du graphe de la figure 2. Le pourcentage de paires de bases G:C du génome humain n'est pas constant et sa distribution n'est pas homogène sur l'ensemble du génome. La plupart des régions analysées possède cependant une teneur en G:C proche de la moyenne. Le pourcentage en G:C de l'ADN n'est jamais inférieur à 31% ni supérieur à 68%, ce qui suggère que des contraintes physiologiques doivent exister pour définir

ces limites. La limite supérieure s'éloigne beaucoup de la moyenne, mais on constate que les régions d'ADN correspondantes restent relativement rares. Le raisonnement tenu par les candidats sur la relation qui existe entre la composition en bases de l'ADN bicaténaire et la stabilité de la double hélice devait les amener à conclure qu'une teneur moyenne en G:C de 41% signifie en fait une relative richesse en paires de bases A:T du génome, en accord avec la nécessité d'ouverture de la double hélice lors de la réplication, de la recombinaison et de la transcription de l'ADN. Les candidats ont en général du mal à établir des liens entre structure et fonctions des molécules et leur raisonnement reste souvent incomplet. De la même façon, l'observation des figures 3 et 4 est restée trop souvent partielle et les réponses aux questions 3 et 4 trop incomplètes.

Question 5. L'analyse de la figure 5 permettait de relever un ensemble de données qui apparaissent contradictoires avec l'hypothèse selon laquelle la complexité des organismes serait reliée à la taille de leur génome (comparaison de plusieurs groupes ou variations au sein d'un même groupe). L'exploitation des données numériques disponibles conduisait à proposer que la complexité des organismes est davantage corrélée au nombre de gènes que contient leur génome. Cette conclusion est cohérente avec le fait qu'un organisme qui possède davantage de gènes a la capacité de synthétiser un plus grand nombre de protéines différentes. On pouvait en déduire que les grands génomes contiennent beaucoup de séquences extragéniques (dont la fonction est encore en grande partie inconnue). Sur cette question, les candidats ont surtout été jugés sur la clarté de l'analyse et la pertinence des conclusions proposées.

Question 6. La séquence codante d'un gène procaryote est facilement repérable, constituée d'une succession continue de codons traduisibles en acides aminés qui forment un cadre de lecture ouvert du codon initiateur aux codons "stop". L'outil informatique identifie très bien ces cadres de lecture au sein des séquences génomiques qu'il permet d'analyser. Dans le cas des Eucaryotes, les gènes sont morcelés et les séquences codantes apparaissent fragmentées en exons relativement courts, séparés par des segments d'ADN qui peuvent être très longs (introns), ce qui rend beaucoup plus difficile leur identification dans une séquence génomique. De façon surprenante, pratiquement personne n'a su répondre à cette question, ce qui montre que, bien que connaissant l'existence des introns dans les gènes eucaryotes, l'immense majorité des candidats ne semble pas avoir réfléchi à ses conséquences.

Question 7. L'observation des trois graphes permettait assez aisément de conclure que lorsqu'on passe de *Caenorhabditis* à la *Drosophile* et à l'Homme, la taille des exons est relativement constante, tandis que celle des introns (globalement, et aussi pour les plus courts d'entre eux) augmente sensiblement, contribuant à l'accroissement du rapport entre la taille du génome et le nombre de gènes au cours de l'évolution (cf. question 5).

Question 8. La structure secondaire caractéristique en feuille de trèfle de tous les ARN de transfert découle de la présence au sein de ces ARN de segments de séquences complémentaires deux à deux, complémentarités que l'on retrouve dans l'ADN qui les code. Des programmes informatiques capables de prédire et de modéliser les structures spatiales les plus probables à partir de la seule structure primaire des acides nucléiques peuvent repérer de telles séquences dans l'ADN génomique. Il s'agit là encore d'un raisonnement simple que les candidats auraient dû savoir conduire puisque la structure des ARN de transfert et la complémentarité des bases leur sont familières. Pourtant, l'immense majorité des réponses proposées était totalement incohérente.

Etude du profil d'expression des gènes d'*Arabidopsis thaliana*

Questions 13 et 14 : Si la description des résultats présentés dans les figures 10 à 12 ne semble avoir posé aucune difficulté, leur interprétation s'est révélée plus problématique, donnant lieu dans la

plupart des copies à des développements confus, contradictoires ou inachevés. Les données fournies dans l'énoncé suggèrent que la protéine CAB2 est localisée dans le mésophylle des feuilles du végétal, associée aux thylakoïdes au sein des chloroplastes, et qu'elle est surtout utile à la plante en présence de lumière, sa synthèse étant diminuée à l'obscurité. La répression de la synthèse de CAB2 à l'obscurité est perdue chez le mutant *doc*, restaurée chez le double mutant *yucca doc*, indiquant que la mutation *yucca* compense l'effet de la mutation *doc*. Si on tente d'établir un lien entre la diminution d'expression du gène *CAB2* à l'obscurité et le transport de l'auxine, le raisonnement le plus simple, très rarement rencontré dans les copies, consiste à dire qu'une quantité suffisante d'auxine est nécessaire dans les feuilles du végétal pour que la synthèse de CAB2 soit correctement réprimée en l'absence de lumière. Cette quantité est réduite chez le mutant *doc* par suite du défaut de transport de l'auxine (l'auxine est produite à l'apex de la plante, transportée vers la base du végétal puis distribuée dans les feuilles), d'où une surexpression de *CAB2* à l'obscurité. Elle augmente à nouveau chez le double mutant *yucca doc* du fait de la surproduction d'auxine liée à la mutation *yucca* qui compense au moins partiellement le défaut de transport de l'hormone.

Question 15. L'abondance des informations fournies dans le tableau 1 devait amener les candidats, guidés par l'énoncé, à identifier quelques exemples significatifs et à repérer des ensembles de gènes dont l'expression montre des caractéristiques communes remarquables permettant de répondre de manière synthétique à la question posée. Il convenait de distinguer les gènes dont l'expression à l'obscurité augmente ou diminue en présence de la mutation *doc*, et d'observer l'effet de la mutation *yucca* chez le double mutant : compensation totale, partielle ou inexistante. Comme *CAB2*, les gènes dont on peut supposer qu'ils sont activement transcrits en présence de lumière (codant par exemple des sous-unités des photosystèmes I et II, une protéine précoce inductible par la lumière ou une protéine liant la chlorophylle) voient leur expression à l'obscurité augmenter chez le mutant *doc*, et l'effet est supprimé lorsque la mutation *yucca* est également présente. Le niveau relatif d'expression du gène codant une protéine de réponse à l'auxine apparaît bien corrélé à la quantité d'auxine disponible (diminution chez le mutant *doc*, augmentation chez le mutant *yucca*). Très peu de candidats ont traité ces deux derniers aspects de la question.

Questions 16 à 20. L'étude des effets du NPA et de la BFA sur la localisation du transporteur PIN1 a souvent été menée de façon partielle et peu rigoureuse. Peu de candidats ont vu que le NPA ne provoque un changement de localisation de PIN1 que chez le mutant *doc*, ou que l'effet de la BFA est réversible. Si le blocage de l'exocytose par la BFA suffit à expliquer que PIN1 nouvellement synthétisé n'atteigne jamais la membrane plasmique, il n'explique pas que le transporteur PIN1 déjà associé à la membrane plasmique avant le traitement disparaisse en présence de BFA, sauf si sa demi-vie est très courte (hypothèse exclue par l'énoncé), ou s'il est continuellement soumis à un processus d'endocytose / exocytose (recyclage), sa localisation dans la membrane plasmique résultant d'un équilibre dynamique. Il nous semble que cette analyse était tout à fait accessible aux candidats qui auraient dû obtenir davantage de points sur cette partie du sujet. La fin de l'exercice était de difficulté supérieure et l'établissement d'un modèle complet du système étudié n'était pas chose aisée. Toutefois, un certain nombre de candidats a fait l'effort d'en proposer une ébauche, plus ou moins poussée, que les correcteurs ont récompensée à la condition qu'elle soit en accord avec les réponses apportées précédemment, et correctement justifiée.

Génomique fonctionnelle du développement chez *Caenorhabditis elegans*

Question 21. La figure 14 a rarement été bien analysée par les candidats, dont beaucoup n'ont pas compris que les deux régions du chromosome I comportant des redondances correspondaient aux portions 2 d'une part, 8 + 9 d'autre part, et que si deux copies légèrement divergentes d'un même gène sont présentes sur le chromosome, l'une peut être inactivée par interférence par ARN, tandis

que l'autre reste active, la compensation de fonction qui en découle faisant qu'aucun phénotype mutant n'apparaît.

Question 23. L'analyse de la figure 15 pouvait s'effectuer à partir de l'impression d'ensemble donnée par son observation. On constate que dans les détails les sept profils d'expression représentés sont tous distincts, ce qui traduit une relative spécialisation des protéines étudiées. Cependant, la majorité des gènes de cyclines montre une expression minimale aux stades L1, L2, L3, et une expression forte aux stades E d'une part, (L4 et) A d'autre part. On pouvait proposer que les cyclines sont synthétisées en abondance aux stades où les divisions cellulaires sont les plus fréquentes (développement embryonnaire, développement de la lignée germinale).

Question 24. D'excellents schémas de synthèse des données numériques fournies ont été présentés dans certaines copies. Peu de candidats ont conclu à une corrélation entre l'observation d'un phénotype mutant lors de l'inactivation d'un gène et la conservation de ce gène au cours de l'évolution, donc l'importance physiologique de sa fonction.