bionumbers, une base de données quantitatives en biologie cellulaire et moléculaire

publié le 25.09.17 |par harold lopparelli

présentation de la base de données bionumbers, qui recense des valeurs utiles en biologie (concentrations de différentes biomolécules dans les cellules, temps de génération, tailles caractéristiques, vitesses de synthèse...)

la base de données *bionumbers* est née en 2007 de l'effort conjoint de plusieurs chercheurs du département de biologie des systèmes de l'université d'harvard ; elle est aujourd'hui développée au sein du milo lab du weizmann institute of science (israël).

elle rassemble des données biologiques quantifiées concernant principalement les échelles moléculaires et cellulaires, ainsi que, dans une moindre mesure, d'autres échelles comme celle de la biosphère. elle a été conçue pour fonctionner de façon collaborative : ses utilisateurs sont encouragés à soumettre des données disponibles dans la littérature, et à commenter les fiches des données déjà enregistrées dans la base.

1. quelques exemples de *bionumbers*

1.1. quantités et concentrations

- **quelle est la concentration en ribosomes d'une cellule d'*escherichia coli* ? de 25 000 à 31 000 ribosomes/μm^3.**
- **quelle est la concentration en atp dans un neurone de rat ? 2,59 mmol/l.**
- **quel est le nombre d'atpases à protons dans une cellule de *saccharomyces cerevisiae* ? $1{,}26.10^6$ molécules/cellule.**
- **quelles sont les concentrations intra et extracellulaires en ions calcium ca^{2+} dans une cellule cardiaque de mammifère ? concentration extracellulaire : 10^{-3} mmol/l, intracellulaire : 10^{-5} mmol/l.**
- **quel est le nombre de molécules de la protéine calmoduline dans une cellule de rat (lignée cellulaire pc12) ? de 5 à 10 millions.**

1.2. durées et vitesses

- **quelle est la demi-vie des arnm dans une cellule de *saccharomyces cerevisiae* ? vingt minutes.**
- **quel est le temps moyen de génération de la chlorelle à la lumière, en conditions aérobies et à 25 °c ? 11 heures.**

- **quelle est la durée minimale des cycles cellulaires lors du développement embryonnaire précoce de *drosophila melanogaster* ? 8 minutes.**
- **quelles sont les vitesses de transcription et de traduction chez d'*escherichia coli* ? environ 45 bases/s et 15 acides aminés/s.**
- **quel est le flux de molécules d'eau traversant une protéine aquaporine de type 1 ? 5.10^{8} molécules/s.**
- **quelle est la durée du renouvellement complet des cellules de la cornée de l'œil humain ? de l'ordre de 7 jours.**
- **quelle est la durée du renouvellement complet des cellules de l'épithélium de l'intestin grêle ? de l'ordre de 3-4 jours.**

1.3. tailles

- **quel est le volume du noyau d'une cellule hela ? 690 μm^3.**
- **quel est le volume occupé par l'amidon dans les chloroplastes d'*arabidopsis thaliana* ? 15 %.**
- **quelle est la taille du génome du bactériophage lambda ? 48 502 bp.**

1.4. valeurs diverses

- **quelle est la consommation moyenne de glucose d'un neurone du cortex cérébral chez la souris ? $1,39.10^{-8}$ µmol/min.**
- **quelle est la consommation de glucose habituelle dans une cellule animale en culture de la lignée cho ? 0,2 pmol/h/cellule.**
- **quelle est la fraction des réactions de la rubisco qui sont des oxygénations du ribulose-1,5-bisphosphate (et non des carboxylations) à 25 °c dans la biosphère ? 35%.**
- **quel est le taux de mutation par substitution dans la lignée germinale humaine ? 10^{-8} par paire de bases par génération.**
- **comment varie la densité de stomates chez d'*arabidopsis thaliana* ? de 45 à 720 stomates/mm^2 selon la concentration de co_2 du milieu.**
- **quel est le pourcentage de la fixation du carbone due à la photosynthèse des diatomées ? de l'ordre de 20 %.**
- **quelle est la biomasse contenue dans la litière végétale à l'échelle de la biosphère ? 122 gt.**

2. les fiches *bionumbers*

chaque nombre de la base de données est accompagné de son unité ainsi que d'une attribution biologique précise (le plus souvent un organisme modèle, un type ou une lignée cellulaire). la référence correspondante dans la littérature primaire est indiquée et fréquemment accompagnée d'extraits du texte de l'article (qui le plus souvent n'est pas lui-même en accès libre) relatifs à la caractérisation précise de la grandeur en question, à la méthodologie suivie pour la calculer ou la mesurer, ainsi que des commentaires permettant de cerner son domaine de validité. ces informations sont complétées par une liste des *related bionumbers* disponibles.

exemple d'une fiche de la base de données *bionumbers* relative à l'arn polymérase iii, enzyme qui intervient dans la synthèse de l'arnr 5s, des arnt et d'autres petits arn dans

les cellules eucaryotes.

id	103657
grandeur	vitesse d'élongation de l'arn polymérase iii
organisme	levure bourgeonnante *saccharomyces cerevisiae*
intervalle	21-22
unité	nucléotides/seconde
référence	matsuzaki h, kassavetis ga, geiduschek ep. analysis of rna chain elong termination by *saccharomyces cerevisiae* rna polymerase iii. *j mol bio* jan 28 235(4): 1173-92.
référence pubmed id	8308883
méthode	transcription *in vitro* et libération du transcrit. p. 1175, colonne de ga paragraphe : « des autoradiographies ont été scannées avec un densit laser et analysées avec un logiciel développé par ambis (san diego). »
commentaire	dans le résumé de l'article : « […] à 20 °c avec 1 mmol/l de ntp [nucléo phosphates], les chaînes dont l'élongation est la plus rapide atteigne vitesse moyenne de 29 nucléotides/seconde, la valeur médiane est de nucléotides/seconde » […]
saisie	uri m.
date d'ajout	15 mars 2009, 10h10
date de dernière modification	7 mai 2015, 7h00
version	11
lien permanent	http://bionumbers.hms.harvard.edu/bionumber.aspx?s=n&id=103657&

différentes catégories (*popular bionumbers, recent bionumbers, amazing bionumbers*, etc.) permettent d'avoir un aperçu des ressources disponibles ; les fonctions « rechercher » (*search*) et « parcourir » (*browse*) permettent d'utiliser efficacement la base de données de façon plus systématique.

plusieurs collections réalisées à partir de la base bionumbers permettent d'accéder facilement à des données-clé correspondant aux ordres de grandeurs importants en

biologie moléculaire et cellulaire, dont une synthèse est proposée dans le tableau suivant.

	propriétés	*e. coli*
tailles	longueur	2–4 µm
	volume cellulaire	0,3–5 μm^3
	épaisseur des membranes	4–10 nm
	diamètre moyen d’une protéine	3–6 nm
	paire de bases	2 nm (diamètre)
	molécule d’eau	~ 0,3 nm
génomes	taille du génome	4,6 mb
	nombre de gènes codant une protéine	4 300
	longueur des sites régulateurs	10–20 bp
	longueur des promoteurs	~ 100 bp
	longueur des gènes	~ 1 000 bp
concentrations	protéines par μm^3 de cellule	2–4 × 10^6
	concentration d’une protéine donnée par cellule	~ 1 nm
	protéines par cellule	~ 10^6
	ribosomes/cellule	~ 10^4
vitesses et durées	temps de fixation d’un facteur de transcription à l’adn	~ 1 s
	vitesse de réplication par l’adn polymérase	200–1 000 nt/s
	vitesse de transcription par l’arn polymérase	10–100 nt/s
	temps de transcription d’un gène	< 1 min
	durée de vie d’un arnm	3 min
	vitesse de traduction d’un ribosome	10–20 aa/s
	temps de traduction d’une protéine (300 aa)	~ 1 min
	temps de repliement d’une protéine	1 ms – 1 min
	durée de vie d’une protéine	1 h

temps de diffusion d'une protéine à travers une cellule (d ~ 10 µm²/s)	**~ 0,01 s**	**~ 0,2 s**
temps de diffusion d'une petite molécule à travers une cellule (d ~ 100 µm²/s)	**~ 0,001 s**	**~ 0,03 s**
transitions entre états protéiques (actif/inactif)	**1–100 µs**	
temps pour atteindre l'équilibre de liaison petites molécules/proté	**1–1000 ms (affinité : 1 µm – 1 nm)**	
temps de génération	**20–40 min (en phase de croissance exponentielle)**	**1–2 h**
erreurs	**taux de mutation**	**~ 10^{-8}–10^{-10}/bp/**
	erreur d'incorporation lors de la transcription	**10^{-4}–10^{-5} par nu**
	erreur d'incorporation lors de la traduction	**10^{-3}–10^{-4} par ac**

taille des génomes

- ***e. coli*** **~ 5 mb**
- ***s. cerevisiae*** **(levure de boulanger) ~ 12 mb**
- ***c. elegans*** **(nématode) ~ 100 mb**
- ***d. melanogaster*** **(drosophile) ~ 120 mb**
- ***a. thaliana*** **(arabette des dames) ~ 120 mb**
- ***m. musculus*** **(souris) ~ 2,5 gb**
- ***h. sapiens*** **(être humain) ~ 2,9 gb**
- ***t. aestivum*** **(blé) ~ 16 gb**

3. documents à télécharger

valeurs utiles en biologie
document au format pdf

valeurs utiles en biologie
document au format odt, pour édition

sources (contenant également des infographies téléchargeables gratuitement) :

- **typical cellular parameter values for a bacterial e. coli cell, the single-celled eukaryote s. cerevisiae (budding yeast), and a mammalian hela cell line**
- **key numbers for cell biologists**
- **snapshot: key numbers in biology, uri moran, rob phillips, and ron milo, cell volume 141, issue 7, p1262–1262.e1, june 25, 2010 https://doi.org/10.1016/j.cell.2010.06.019**
- **snapshot: timescales in cell biology, maya shamir, yinon bar-on, rob phillips, and ron milo, cell volume 164, issue 6, p1301–1302.e1, march 10, 2016, http://dx.doi.org/10.1016/j.cell.2016.02.058**

4. les enjeux des données quantifiées en biologie

pour ses concepteurs, une telle base de données répond à un besoin pour les chercheurs en biologie moléculaire. en effet, alors que les nouvelles techniques de mesure et l'accroissement de leur productivité permettent la mise à disposition de grandes quantités de données biologiques quantifiées de toute nature – disponibles pour partie dans la littérature primaire –, il n'est pas évident d'accéder directement à ces valeurs chiffrées qui sont pourtant souvent nécessaires pour conduire des expériences et en interpréter les résultats, et indispensables pour élaborer de nouveaux modèles quantitatifs.

les informations précises et référencées accompagnant chaque *bionumber* permettent de prendre en compte les enjeux spécifiques de la quantification en biologie que sont la variabilité du vivant (un intervalle ou une distribution sont souvent plus informatifs qu'une moyenne ou une médiane), la diversité du matériel biologique utilisé au laboratoire, celle des conditions expérimentales adoptées, ou encore des méthodes et procédures de mesure suivies. l'usage qui en peut être fait est ainsi variable : on peut chercher à en tirer des ordres de grandeur, ou extraire des valeurs plus spécifiques dont la précision et la validité dépendent de conditions données – on trouvera ainsi dans certains cas des fiches correspondant à une « même » grandeur déterminée dans des conditions expérimentales ou avec des méthodes différentes.

il est intéressant de remarquer que, de façon générale, les données disponibles sont de deux types : certaines résultent directement de mesures expérimentales (et d'une analyse statistique), tandis que d'autres sont le produit d'une modélisation plus ou moins complexe. toutes peuvent être affinées voire remises en question, au gré de l'amélioration des appareils et des techniques de mesure, ou de la prise en compte de nouvelles données et de nouvelles hypothèses dans les modèles existants, voire de l'élaboration de nouveaux modèles. on peut donner comme exemple d'une telle démarche les travaux concernant le nombre de cellules du corps humain, et celui des bactéries du microbiote qui le colonisent (sender et al., 2016).

les données de la base *bionumbers* peuvent enfin être utilisées de façon heuristique pour des estimations (*bioestimates*) permettant de mettre en évidence des insuffisances ou des points aveugles de la compréhension plutôt qualitative des phénomènes qui a cours habituellement en biologie moléculaire. la mise à disposition de données quantifiées variées permet d'essayer de transposer en biologie moléculaire ce que

physiciens et ingénieurs appellent classiquement les « problèmes de fermi » : en raisonnant à partir d'ordres de grandeur et d'approximations, on peut développer un point de vue renouvelé sur de nombreux problèmes biologiques, et y apporter des solutions valables.

comment la bactérie *escherichia coli* a-t-elle le temps de répliquer son génome ?

son génome comporte environ 5 millions de paires de bases azotées (bp), et sa vitesse de réplication varie entre 200 et 1000 bp/seconde. par conséquent, les deux réplisomes (associés aux deux fourches de réplications) auraient en théorie besoin d'au moins 2500 secondes (soit environ 42 minutes) pour répliquer le génome entier, durée bien supérieure au temps de génération minimum de cette bactérie, qui est d'environ 20 minutes. cela semble impossible… il s'avère en fait que, dans des conditions de culture idéales, il existe chez *e. coli* des fourches de réplications imbriquées au niveau desquelles commence à être répliqué l'adn qu'on retrouvera dans les cellules qui se formeront lors des divisions n+1 et n+2, alors même que la réplication du génome pour les cellules filles de la génération immédiatement suivante (n) n'est pas achevée.

adapté de http://dx.doi.org/10.1016/j.cell.2010.06.019

les concepteurs de la base *bionumbers* espèrent qu'elle facilitera le développement d'une « numératie biologique » (*biological numeracy*) dans le domaine de l'enseignement des sciences du vivant. la publication par ron milo et rob phillips en 2016 de l'ouvrage *cell biology by the numbers* (disponible gratuitement en ligne en version auteur) y contribue également.

5. références

5.1. sur bionumbers

1. **bionumbers, the database of useful biological numbers**
2. **ron milo, paul jorgensen, uri moran, griffin weber, michael springer, « bionumbers – the database of key numbers in molecular and cell biology », *nucleic acids research* (2010) 38 (suppl_1): d750-d753. https://doi.org/10.1093/nar/gkp889**
3. **rob phillips and ron milo, « a feeling for the numbers in biology » , *pnas*, vol. 106 no. 51, 21465-21471, december 22, 2009 https://doi.org/10.1073/pnas.0907732106**
4. **ron milo, rob phillips, *cell biology by the numbers*, new york, ny : garland science, taylor & francis group, 2016, 356 pages**

5.2. d'autres bases de données plus spécialisées

1. **données sur les tailles des génomes animaux, le nombre de chromosomes**
2. **données sur les enzymes**
3. **données concernant exclusivement *escherichia coli***
4. **données concernant la biologie moléculaire et la génétique de *saccharomyces cerevisiae***
5. **plant dna c-values database**

5.3. pour aller plus loin

1. captations vidéo de cours de ron milo « cell biology by the numbers »
2. quelques documents associés à ces cours

CRÉDITS

AUTEUR(S)/AUTRICE(S)

harold lopparelli

professeur agrégé de svt en classe préparatoire bcpst

MISE EN LIGNE

pascal combemorel

agrégé de svt, il est le responsable éditorial du site planet-vie depuis septembre 2016.

LICENCE DU TEXTE DE L'ARTICLE

creative commons - attribution - pas d'utilisation commerciale