

> L'infection par des agents pathogènes est à l'origine, chez l'homme, de plusieurs millions de décès chaque année. La connaissance exhaustive des interactions entre les pathogènes et leurs hôtes (l'infectome) est un défi majeur pour le développement de nouvelles stratégies diagnostiques, thérapeutiques et de prévention. L'émergence de méthodes analytiques à haut débit et leur application à l'étude des agents pathogènes révolutionnent notre manière de suivre et de mesurer les modifications moléculaires et physiologiques induites chez l'hôte lors de l'infection. La compréhension globale du fonctionnement du système pathogènes-hôte repose par ailleurs sur le développement de modèles de biologie moléculaire systémique prenant en compte l'ensemble des molécules et leurs interactions. Dans cette optique, nous avons reconstruit un prototype d'une « cellule infectée virtuelle » humaine en passant en revue plus de trente années de recherche en virologie moléculaire. Ce modèle concerne plusieurs centaines de virus (dont les virus des hépatites C et B, de l'herpès et du cancer du col de l'utérus et le VIH) et nous a permis de dégager de nouvelles hypothèses pour le développement de stratégies antivirales innovantes basées sur le contrôle de fonctions cellulaires. <

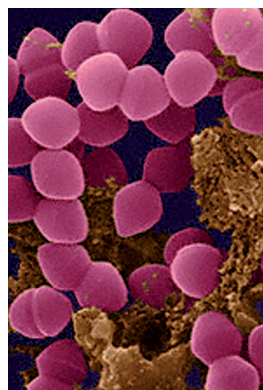
« Infectomique », vers une encyclopédie moléculaire de l'infection

Les développements de la biologie systémique (*Encadré 1*) des infections reposent sur l'essor de l'« infectomique », discipline qui regroupe l'ensemble des technologies « -omiques » à haut débit (*Encadré 2*) appliquées à l'étude systématique du répertoire moléculaire des interactions entre les agents pathogènes et leurs hôtes, l'« infectome » [1]. En ce qui concerne les pathogènes

La cellule infectée virtuelle

Un rationnel de biologie systémique pour la découverte de nouvelles cibles thérapeutiques antivirales

Vincent Navratil, Vincent Lotteau,
 Chantal Rabourdin-Combe



V. Navratil : ENS Lyon,
 5, rue de la Doua,
 69100 Villeurbanne, France ;
 U851 Inserm I-MAP Team,
 21, avenue Tony Garnier,
 F-69007, Lyon, France.
vincent.navratil@ens-lyon.fr
 V. Lotteau, C. Rabourdin-Combe :
 U851 Inserm I-MAP Team,
 21, avenue Tony Garnier,
 F-69007, Lyon, France.

viraux, l'infectome peut être défini comme l'ensemble des interactions moléculaires et fonctionnelles (par exemple les régulations génétiques, les interactions protéine-protéine, les réactions métaboliques) entre les composants des virus (génomme, capsid, enveloppe virale, etc.) et ceux de la cellule hôte (le cytosquelette, le noyau, les membranes, les organites, etc.) (→). **(→) Voir l'Éditorial de Pierre Legrain, p. 559 de ce numéro**

Génomique virale, recherche des composantes du système

Une partie des bases moléculaires des maladies induites par les virus est codée au niveau génomique. Un nombre important de projets de séquençage systématique de génomes viraux a été réalisé. À ce jour, sur les 5 972 génomes d'espèces entièrement séquencés et référencés dans la base de données du NCBI *Entrez genome*, près de 2 292 génomes correspondent à des virus (*Figure 1*)¹. Même si ces projets ne couvrent pas encore l'incroyable diversité des souches et des quasi-espèces pourtant en lien direct avec leur virulence, leur pathogénicité et leur tropisme, ils apportent néanmoins une information unique sur la diversité des

¹ <http://www.ncbi.nlm.nih.gov/genomes/VIRUSES/viruses.html>

1 Introduction aux principaux objectifs de la biologie systémique

La biologie systémique a pour objet de créer des modèles informatiques et mathématiques descriptifs et prédictifs de systèmes biologiques [34, 35]. Ces modèles ont pour but de proposer de nouvelles hypothèses biologiques testables expérimentalement au laboratoire. On peut de manière schématique considérer que l'on doit passer par trois étapes pour atteindre cet objectif :

1. L'énumération exhaustive d'une part des éléments composant le système étudié, et d'autre part des interactions entre ces éléments (Encadré 2). Le traitement bio-informatique des données d'« -omiques » (modélisation, intégration, gestion, accessibilité des données) en est un des aspects majeurs.

2a. La modélisation de l'architecture physique du système d'éléments en interaction. Un des objectifs est de déterminer les propriétés structurales émergentes des réseaux d'interactions en lien avec le fonctionnement du système biologique étudié [33]. L'analyse des graphes est une approche utilisée pour quantifier et explorer la connectivité des éléments au sein du réseau global d'interactions moléculaires (Encadré 3).

2b. La modélisation dynamique spatiale et temporelle des flux et des variations moléculaires (modélisation statistique, équations différentielles, etc.), conduisant à un modèle en principe prédictif du système étudié. Une très bonne revue des méthodes utilisées est disponible dans le numéro spécial de *Médecine/Sciences* sur la biologie systémique [28-30, 34, 35].

3. Le retour à la biologie expérimentale pour la validation de nouvelles hypothèses biologiques obtenues à partir de modèles qualitatifs (2a) et/ou quantitatifs (2b) et l'enrichissement de ces modèles à l'aide de données expérimentales.

Dans le cadre du projet I-MAP nous nous sommes focalisés sur les étapes 1, 2a et 3. Un des objectifs est de préparer des sous-systèmes « simplifiés » pour passer à l'étape 2b.

modes d'organisation génomique du monde viral, comme en témoigne le nombre important de gènes viraux sans homologues connus (gènes orphelins ou *orphan*). Par ailleurs, ces génomes viraux sont extrêmement compacts en comparaison des génomes eucaryotes et codent au total pour plus de 75 000 protéines. Si on considère les génomes viraux mis bout à bout, ces protéines sont réparties au total sur plus de 60 mégabases, soit la taille du chromosome 20 humain. Mais le développement de cibles antivirales repose principalement sur la connaissance fine des structures biochimiques et des fonctions biologiques de ces protéines virales [2] qui, pour la grande majorité d'entre elles, restent encore peu ou pas caractérisées.

Modélisation systémique de la réponse cellulaire aux pathogènes

Au cours de leur coévolution avec les agents pathogènes, les organismes hôtes ont sélectionné des lignes de défense extrêmement sophistiquées. Une partie de ces défenses immunitaires est innée et s'exprime très précocement au sein des cellules de l'organisme en réponse à l'activation de sentinelles moléculaires. C'est le cas des récepteurs membranaires TLR (récepteurs Toll-like) et des senseurs intracellulaires

RLR (récepteurs Rig-like) qui interagissent avec des composants des pathogènes (les lipopolysaccharides, l'ARN simple brin, l'ARN ou l'ADN double brin, etc.) [3]. La réponse cellulaire à ces interactions est alors assurée par un ensemble complexe de voies de signalisation et de voies de régulations génétiques conduisant à la production de molécules à activité antivirale telles que l'interféron. Des travaux ont été entrepris pour reconstruire le réseau moléculaire des voies des récepteurs TLR [4] et de celles qui sont associées à la réponse immunitaire innée [5]. Citons, pour illustrer les progrès réalisés dans ce domaine, les travaux du Dr Ido Amit *et al.* sur le profilage systématique utilisant de manière intégrée des approches de transcriptomique et de criblage d'ARN interférents (pour l'extinction de l'expression de gènes) [6]. Compte tenu de leur fort taux de mutation et de réplication, les virus, et plus particulièrement ceux dont le génome est formé d'ARN, présentent une capacité d'adaptation extrêmement rapide à la réponse immunitaire innée et aux nouvelles molécules antivirales, dont celles qui modulent la réponse à l'interféron [7]. Dans ce contexte, un des défis actuels est le développement de nouvelles stratégies de découverte de molécules antivirales basées sur la connaissance des interactions entre les protéines virales et les protéines cellulaires [8], dans le but d'agir directement au niveau des fonctions cellulaires tout en échappant au processus d'adaptation mutationnelle des virus.

Interactome virus-hôte,

définition de l'interface moléculaire de l'infection

Les virus, parasites intracellulaires obligatoires, utilisent les ressources métaboliques et la machinerie moléculaire de la cellule hôte pour réaliser leur cycle de réplication. Dans le contexte général d'une infection, les composants viraux (les protéines, l'ARN, l'ADN) n'agissent pas de manière isolée. La recherche des partenaires viraux ou cellulaires des protéines virales offre un cadre idéal pour l'étude de leurs fonctions et de leur rôle éventuel dans le développement des pathologies associées à l'infection. Depuis l'avènement de la biologie moléculaire, ce type d'approche a été très utilisé à l'échelle d'une protéine ou même d'une interaction binaire testée *a priori* (par exemple par des techniques de GST pull down, co-immunoprécipitation ou double hybride en levure). Cependant, peu d'efforts ont encore été engagés à ce jour dans la caractérisation systématique à large échelle du vaste espace d'interactions des protéines virales. En effet, plusieurs millions d'interactions binaires devraient encore être testées en théorie si l'on considère les dizaines de milliers de protéines virales et cellulaires. Citons néanmoins les études pionnières de cartographie de

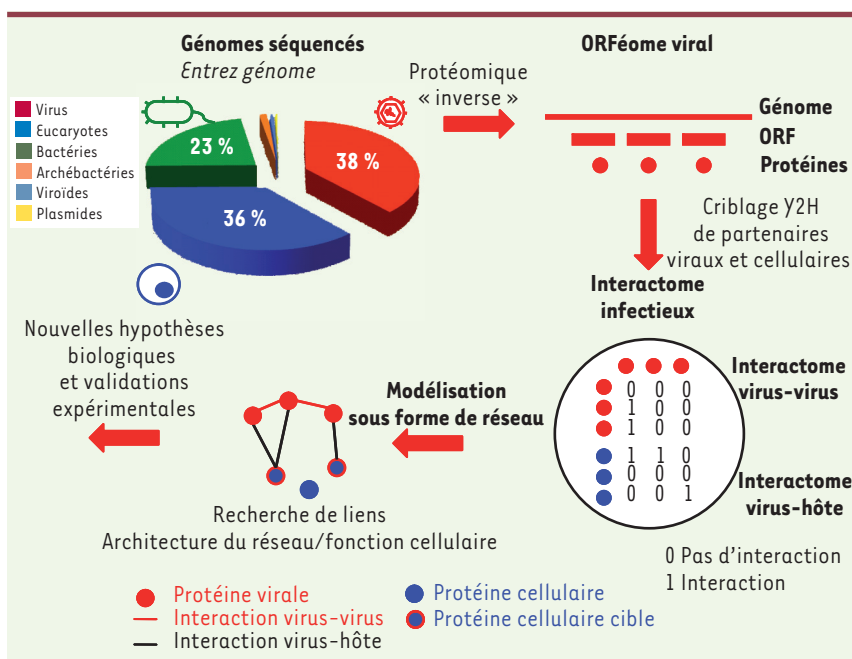


Figure 1. Du génome au réseau d'interactions protéiques du virus, une approche transversale de biologie moléculaire systémique. Une répartition taxonomique des génomes séquencés est tout d'abord donnée. Ces génomes sont utilisés comme un canevas (par une approche de « protéomique inverse ») pour l'annotation des protéines (ou fragments de protéines) encodées et la genèse de l'ORFéome viral, c'est-à-dire l'ensemble des ORF (*open reading frame*). L'ORFéome viral est utilisé pour le criblage systématique de l'interactome virus-virus et virus-hôte. Un exemple théorique d'un espace de recherche de 18 combinaisons concernant 3 protéines virales et 3 protéines cellulaires est présenté (1 si interagit, 0 si n'interagit pas). Les interactions expérimentales (arêtes rouges) sont intégrées au sein d'une modélisation du réseau d'interactions entre les protéines du virus (nœuds

rouges) et celles de l'hôte (nœuds bleus). L'étude systémique des réseaux d'interactions protéine-protéine permet l'identification de liens entre l'architecture du réseau et le fonctionnement cellulaire. Cette approche conduit à la proposition de nouvelles hypothèses biologiques testables expérimentalement.

l'interactome viral par la technologie du double hybride en levure (Y2H : *yeast two-hybrid*) réalisées à l'échelle de plusieurs virus entiers dont celui de l'hépatite C [9], du SARS (*severe acute respiratory syndrome*) [10] et de plusieurs herpèsvirus [11, 12]. Des travaux récents ont également été entrepris pour explorer sans *a priori* l'interactome virus-hôte, dont celui de l'EBV (*Epstein Barr*

virus) [13], du virus de l'hépatite C [14] et du virus de la vaccine [15]. Comme pour les interactomes des organismes modèles, ces avancées ont été rendues possibles par le développement de larges collections d'ORF (*open reading frame*, cadre ouvert de lecture) ou ORFéome [16, 17]. Ainsi, dans le cadre du projet I-MAP (*infectious mapping project*), le premier ORFéome viral est en cours d'assemblage dans le but de faciliter l'étude systématique des fonctions des protéines virales tout en prenant

2 Biologie systémique et « -omiques », vers un avatar moléculaire de la cellule

L'avènement des techniques de séquençage, de criblage de génomique fonctionnelle (ARN interférents, criblage de petites molécules, etc.) et de profilage dit à « haut débit » (transcriptomique, protéomique, métabolomique, etc.) regroupées sous le terme de technologies « -omiques » a bouleversé en quelques années l'horizon de la recherche fondamentale, biomédicale et pharmaceutique. Il est désormais envisageable de caractériser le répertoire moléculaire d'une cellule eucaryote « saine » (génome, protéome, transcriptome, protéome, métabolome, etc.) ainsi que le mode d'interaction de ces molécules (régulation génétique, interaction protéine-protéine ou interactome, réaction enzymatique). Le suffixe « -ome » employé vient du mot sanscrit « OM » signifiant complétude. Il est utilisé en biologie pour les mots désignant un ensemble ou un système de manière exhaustive (par exemple le génome : ensemble des gènes d'un organisme ; le transcriptome : ensemble des transcrits ; le métabolome : ensemble des métabolites). En utilisant le suffixe « -omique » ou « -omics » en anglais, le mot peut désigner soit les lois issues de la connaissance des ensembles (« -omes ») analysés, soit les

technologies à haut débit utilisées pour obtenir ces ensembles. Ainsi, le séquençage du génome humain a permis l'identification de plus de 25 000 gènes [31]. Ces gènes codent potentiellement pour près de 50 000 à 100 000 protéines matures distinctes dont les fonctions restent très mal connues. Ces protéines s'expriment au sein d'un réseau cellulaire extrêmement complexe comprenant plusieurs centaines de milliers d'interactions moléculaires directes comme le suggèrent les premières estimations réalisées à partir des grands projets sur l'interactome cellulaire humain [32]. Ainsi, à partir des données publiées et de celles qui sont disponibles dans les bases de données publiques (dont IntAct, MINT, HPRD), nous avons reconstruit un réseau d'interactions protéine-protéine cellulaire, véritable modèle virtuel du fonctionnement d'une cellule. L'expression dynamique des réseaux moléculaires façonne la spécificité de fonctionnement des types cellulaires mais également celle des tissus et des organes. Cette dynamique prédétermine également le niveau de plasticité de la cellule en réponse à des modifications endogènes (mutations héréditaires ou somatiques) ou exogènes (médicaments, stress chimique, UV, infections, etc.).

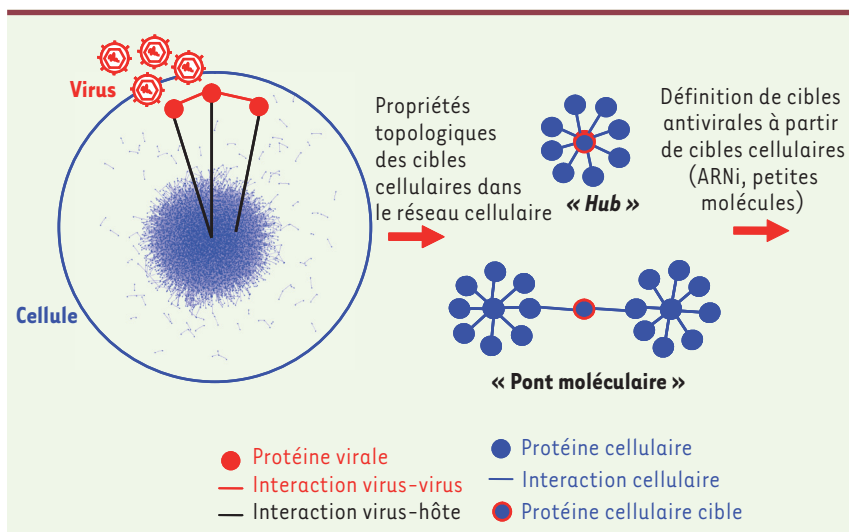


Figure 2. La cellule infectée virtuelle, vers un modèle systémique de l'infection. Projection des cibles cellulaires du virus au sein du réseau d'interactions protéine-protéine de la cellule. Les interactions entre les protéines cellulaires sont représentées en bleu et viennent compléter le formalisme présenté dans la Figure 1. L'étude du graphe (Encadré 3) représentant le réseau cellulaire infecté par le virus de l'hépatite C a permis d'étudier la proximité des protéines ciblées dans le réseau (plus court chemin entre deux protéines, traduisant une voie fonctionnelle potentielle). Les protéines du VHC interagissent de manière préférentielle avec les *hubs* et les « ponts moléculaires ».

en compte la diversité génétique liée aux différentes souches existantes et à leur virulence. Cet ORFéome viral est accessible au sein de la base de données publique ViralORFéome² [18]. En complément de l'approche I-MAP, de nombreuses initiatives internationales de criblage d'interactomes virus-hôte sont en train de bouleverser la discipline. Citons les travaux en cours du *Center for cancer system biology* sur le virus VP5³, et ceux de l'*Institute of toxicology and genetics* sur les herpèsvirus⁴.

La cellule infectée virtuelle, une simulation *in silico* de l'infection

L'accumulation de perturbations moléculaires et cellulaires induites au cours d'une infection aiguë (grippe, rougeole, Chickungunya, etc.) ou chronique (VHC, VHB, VIH, EBV, etc.) joue, en combinaison avec d'autres facteurs de risque dont les prédispositions génétiques, un rôle prépondérant dans le développement de pathologies. Nous avons dans ce contexte introduit un modèle formel d'infection basé sur des données d'interactions moléculaires impliquées au cours d'une infection et une représentation sous forme de graphe (Encadré 3), véritable ossature physique de la cellule infectée virtuelle (Figure 2).

Dans le cadre d'une étude pilote menée au sein du projet I-MAP, le modèle de cellule infectée virtuelle a été appliqué à la simulation de l'infection virale par le VHC (génotype 1b). Une approche de criblage double hybride systématique nous a tout d'abord permis d'établir une première carte dite de « haute confiance » de l'infection

par le génotype 1b du VHC. Cette carte comprend 481 interactions directes entre les 11 protéines du VHC et 421 protéines cellulaires humaines. L'analyse des fonctions cellulaires des cibles cellulaires du VHC a montré une attaque préférentielle de plusieurs voies cellulaires dont la voie d'adhésion focale des cellules à leur substrat ; la perturbation de cette voie par les protéines NS3 actives et NS5A du VHC a été validée *in vitro*. Ces résultats suggèrent que cette voie intervient dans la mise en place du phénotype tumoral lié à la perte d'adhésion des cellules qui est associée à l'infection chronique par le VHC (par exemple dans les hépatocarcinomes).

Une étude sans *a priori* des cibles protéiques du VHC dans le contexte du réseau cellulaire humain (Encadré 3) a révélé une interaction préférentielle des protéines virales avec des protéines fortement connectées à d'autres protéines cellulaires (ce sont les *hubs* moléculaires qui agissent comme des nœuds centraux) et des protéines centrales dans le réseau (Figure 2) [19]. Nous avons par ailleurs montré une interaction préférentielle et massive des protéines virales du VHC avec des protéines de type « goulot d'étranglement » (ce sont des *bottleneck* moléculaires) dans le réseau (Figure 2). Ces protéines présentent des caractéristiques topologiques particulières : elles ont un nombre restreint de partenaires protéiques et une centralité importante dans le réseau cellulaire. Elles joueraient un rôle crucial dans la coordination des régulations entre différentes voies fonctionnelles.

Dans une première tentative exploratoire, un interactome panviral a été reconstruit à partir des résultats de plus de trente années de recherche sur les interactions des protéines virales incluant les premiers résultats de cribles systématiques. Plus de 4 000 interactions protéine-protéine ont été compilées (2 700 interactions virus-virus ; 1 600 interactions virus-hôte) [20] (Tableau 1). Un des résultats marquants est le nombre de fonctions cellulaires ciblées par les virus, avec plus de 5 % (1 600 cibles protéiques) du protéome potentiellement impliqué au cours de l'épisode infectieux. L'ensemble des données est accessible dans la base de connaissances VirHostNet⁵. Même si cette

² <http://pbildb1.univ-lyon1.fr/viralorfeome>

³ <http://ccsb.dfci.harvard.edu/web/www/ccsb/Resources/Data.html>

⁴ <http://itgmvl.fzk.de/itg/uetz/>

⁵ <http://pbildb.univ-lyon1.fr/virhostnet>



3 **Biologie systémique et étude des graphes, numériser l'architecture des réseaux moléculaires**

Le réseau d'interactions moléculaires peut être formalisé mathématiquement sous la forme d'un graphe $G = (N, A)$ formé (1) d'un ensemble de nœuds (N), représentant les molécules (dans notre cas les protéines virales et cellulaires) et (2) d'un ensemble d'arêtes (A) reliant ces nœuds et représentant les interactions moléculaires (dans notre cas les interactions protéine-protéine). Le cadre analytique de la théorie des graphes peut être appliqué à l'étude des caractéristiques topologiques de connectivité et de centralité des protéines dans le réseau cellulaire [33]. Le degré d'un nœud correspond au nombre de partenaires d'une protéine. La distribution des valeurs de connectivité de l'ensemble des protéines humaines du réseau suit approximativement une loi de puissance. Sur la base des données expérimentales d'interactome, l'étude de cette distribution montre que peu de protéines interagissent avec de nombreux partenaires, ce sont des *hubs* dans le réseau. La *betweenness* d'une protéine correspond à une mesure globale de centralité. Elle donne une estimation du flux d'informations passant par une protéine dans le réseau cellulaire. Le degré et la *betweenness* d'une protéine sont deux mesures étroitement corrélées dans l'interactome [14], à ceci près qu'une proportion significative des protéines faiblement connectées présente une position centrale dans le réseau ; on parle de protéines de type « goulot d'étranglement » (*bottleneck*), telles que les ponts moléculaires. Ces ponts moléculaires articulent donc différentes régions du réseau et présentent une position-clé dans la régulation croisée de processus biologiques de la cellule.

première image de l'infection est loin d'être complète, elle propose un point de référence inédit pour des études ultérieures de prédiction *in silico* d'interaction et de criblage systématique des interactions pathogènes-hôte. D'autres projets internationaux de haute confiance viennent épauler VirHostNet dans l'annotation des interactions virus-hôte. Citons la base de données spécialisée VirusMint⁶ [21] ou IntAct⁷, plus généraliste [22]. Alors que la couverture offerte par VirHostNet est actuellement nettement plus large⁸, VirusMint et IntAct s'efforcent d'apporter une documentation plus précise des méthodes expérimentales à l'origine de l'identification des interactions.

Conclusion et perspective

L'intégration des données issues des technologies haut débit appliquées aux relations pathogènes-hôte permettra de modéliser une cellule infectée virtuelle. L'architecture actuelle de cette cellule infectée virtuelle est encore à l'état embryonnaire mais elle constitue une base inestimable pour une compréhension synthétique des bases moléculaires des infections et des maladies infectieuses associées. Par ailleurs, elle offre un canevas pour une interprétation et une intégration rationnelles du flot exponentiel

de données issues des projets d'infectiologie en cours et à venir.

En complément des technologies de cartographie des interactomes (par exemple par Y2H), les technologies de criblage pangénomique par ARN interférents (banques de siARN ou de shARN) se révèlent d'une valeur inestimable dans l'étude mécanistique des relations virus-hôte. Elles permettent d'identifier, par une approche systématique, les cofacteurs cellulaires indispensables à la réplication du virus mais également ceux potentiellement impliqués dans les processus de division et de prolifération anormales des cellules infectées [23-26]. À ce jour, plus de 1 500 facteurs cellulaires essentiels pour la réplication virale ont été isolés. La mise en connexion de ces facteurs-clés avec le réseau d'interactions protéine-protéine virus-hôte permettra de proposer une convergence de rationnels systémiques pour l'identification de nouvelles cibles cellulaires antivirales.

Bien qu'indispensable dans la phase de reconstruction « ascendante » (*bottom-up*) de modèles mécanistiques qualitatifs (reconstruction de voies de signalisation moléculaire et de complexes protéiques), l'aspect quantitatif et prédictif des modèles s'appuyant sur ces données reste encore limité [27]. La caractérisation d'empreintes (signature moléculaire globale) et de profils (catalogue de molécules) transcriptomique et métabolomique par une approche de biologie systémique dite « descendante » (*top-down*) permet néanmoins de pallier ce problème. En effet, l'étude des variations des concentrations d'ARN messagers ou de métabolites mesurées dans des échantillons cellulaires ou tissulaires permet d'élaborer des modèles beaucoup plus prédictifs, par exemple pour la classification des tumeurs, des maladies ou des molécules thérapeutiques.

Ainsi, dans un futur proche, l'intégration multi-échelles des données d'« - omiques » - fusionnant des approches de biologie moléculaire systémique « ascendante » et « descendante » - devrait permettre une modélisation fine des différents modes d'organisation et de réponse fonctionnelle de la cellule au cours de l'infection. Dans ce contexte, le modèle de cellule infectée virtuelle pourra alors être décliné selon différentes pathologies, différents virus, différentes familles de protéines, etc. Finalement, l'étude systématique panpathogène et comparative inter-souches (souches virulentes/non virulentes, résistantes/non résistantes aux traitements, etc.) et des empreintes et profils moléculaires devrait conduire au développement d'outils diagnostiques, notamment pour les maladies dont l'étiologie virale est incertaine (cancers, maladies auto-immunes, etc.) et

⁶ <http://mint.bio.uniroma2.it/virusmint/Welcomedo>

⁷ <http://www.ebi.ac.uk/intact/main.xhtml>

⁸ <http://pbildb1.univ-lyon1.fr/virhostnet/statistics.php>

	Interactome virus-virus	Interactome virus-hôte	Protéines cellulaires ciblées
VHC	28	444	392
VHB	2	91	88
VIH 1	13	336	284
VHH-1	164	78	69
VHH-4	261	403	304
VAA-2	6	220	202
VV	56	216	198
VPH-16	5	166	142
Autres	1 070	843	462
TOTAL	1 662	2 797	1 605

Tableau 1. Synthèse de l'interactome virus-virus et virus-hôte. Les huit virus majoritaires dans la base de connaissances sont référencés : le VHC (virus de l'hépatite C), VHB (virus de l'hépatite B), VIH-1 (virus de l'immunodéficience humaine de type 1), VHH-4 (virus de l'herpès humain de type 4), VHH-1 (virus de l'herpès humain de type 1), VAA-2 (adénovirus de type 2), VV (virus de la vaccine), VPH-16 (virus du papillome humain de type 16). Un bilan du nombre d'interactions protéine-protéine, virus-virus et virus-hôte ainsi que le nombre de cibles cellulaires est donné pour chacun de ces huit virus. D'après les données issues de la base de connaissances VirHostNet (<http://pbildb.univ-lyon1.fr/virhostnet>).

celles qui sont associées à la chronicité d'une infection (par exemple le syndrome métabolique).

Comme tout modèle, cette description holistique du système infectieux n'est qu'un reflet très partiel de la réalité et de la complexité biologiques. Elle ne pourra en aucun cas se substituer aux modèles réduits classiquement développés en virologie. Ces deux approches, l'une réductionniste (l'étude des propriétés des composants), l'autre holistique (l'étude des propriétés émergentes du système d'interaction des composants) devront être utilisées de manière conjointe pour sans cesse s'enrichir mutuellement. Le développement synergique d'approches interdisciplinaires et transversales est l'une des clés de la réussite, comme l'atteste le nombre croissant d'instituts de recherche couplant les aspects de biologie systémique et de biologie analytique. ♦

SUMMARY

The virtual infected cell: a systems biology rational for antiviral drug discovery

Infection caused by pathogens kills millions of people every year. Comprehensive understanding of molecular pathogen-host interactions, i.e. the infectome, is

one of the key steps towards the development of novel diagnostic, therapeutic and preventive strategies. In this quest, progress in high-throughput « omics » technologies applied to pathogens, i.e. infectomics, opens new perspectives toward systemic understanding of perturbations induced during infection. Deciphering the pathogen-host system also relies on the analytical and predictive power of molecular systems biology and by developing *in silico* models taking into account the whole picture of the molecules and their interactions. In this context, we have reconstructed a prototype of the human virtual infected cell based on 30 years of intensive research in the field of molecular virology. This model contains more than one hundred viral infectomes, including major human pathogens (HCV, HBV, HIV, HHV, HPV) and has led to the generation of novel systems-level hypotheses that could be suitable for the development of innovative antiviral strategies based on the control of cellular functions. ♦

REMERCIEMENTS

Ce travail a été financé par l'Anrs, l'Inserm et le ministère français chargé de l'Industrie. Vincent Navratil a été financé par l'Inra. Tous nos remerciements au Dr Abdel Aouacheria, à Sandy Navratil et aux experts de Médecine/Sciences pour leurs critiques et leurs commentaires sur le manuscrit.

CONFLIT D'INTÉRÊTS

Les auteurs déclarent n'avoir aucun conflit d'intérêts concernant les données publiées dans cet article.

RÉFÉRENCES

- Huang SH, Triche T, Jong AY. Infectomics: genomics and proteomics of microbial infections. *Funct Integr Genomics* 2002 ; 1 : 331-44.
- Viswanathan K, Fruh K. Viral proteomics: global evaluation of viruses and their interaction with the host. *Expert Rev Proteomics* 2007 ; 4 : 815-29.
- Medzhitov R. Toll-like receptors and innate immunity. *Nat Rev Immunol* 2001 ; 1 : 135-45.
- Oda K, Kitano H. A comprehensive map of the toll-like receptor signaling network. *Mol Syst Biol* 2006 ; 2 : 2006 0015.
- Lynn DJ, Winsor GL, Chan C, et al. InnateDB: facilitating systems-level analyses of the mammalian innate immune response. *Mol Syst Biol* 2008 ; 4 : 218.
- Amit I, Garber M, Chevrier N, et al. Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science* 2009 ; 326 : 257-63.
- Gago S, Elena SF, Flores R, Sanjuan R. Extremely high mutation rate of a hammerhead viroid. *Science* 2009 ; 323 : 1308.
- Navratil V, de Chasse B, Meyniel L, et al. Systems-level comparison of protein-protein interactions between viruses and the human type 1 interferon system network. *J Proteome Res* 2010, May 11 (*Epub ahead of print*) PMID : 20459142.
- Flajole M, Rotondo G, Daviet L, et al. A genomic approach of the hepatitis C virus generates a protein interaction map. *Gene* 2000 ; 242 : 369-79.
- Von Brunn A, Teepe C, Simpson JC, et al. Analysis of intraviral protein-protein interactions of the SARS coronavirus ORF6. *PLoS One* 2007 ; 2 : e459.
- Uetz P, Dong YA, Zeretzke C, et al. Herpesviral protein networks and their interaction with the human proteome. *Science* 2006 ; 311 : 239-42.
- Fossum E, Friedel CC, Rajagopala SV, et al. Evolutionarily conserved herpesviral protein interaction networks. *PLoS Pathog* 2009 ; 5 : e1000570.
- Calderwood MA, Venkatesan K, Xing L, et al. Epstein-Barr virus and virus human protein interaction maps. *Proc Natl Acad Sci USA* 2007 ; 104 : 7606-11.
- De Chasse B, Navratil V, Tafforeau L, et al. Hepatitis C virus infection protein network. *Mol Syst Biol* 2008 ; 4 : 230.

RÉFÉRENCES

15. Zhang L, Villa NY, Rahman MM, et al. Analysis of vaccinia virus-host protein-protein interactions: validations of yeast two-hybrid screenings. *J Proteome Res* 2009 ; 8 : 4311-8.
16. Lamesch P, Li N, Milstein S, et al. hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes. *Genomics* 2007 ; 89 : 307-15.
17. Uetz P, Rajagopala SV, Dong YA, Haas J. From ORFeomes to protein interaction maps in viruses. *Genome Res* 2004 ; 14 : 2029-33.
18. Pellet J, Tafforeau L, Lucas-Hourani M, et al. ViralORFeome: an integrated database to generate a versatile collection of viral ORFs. *Nucleic Acids Res* 2010 ; 38 : D371-8.
19. Yamasaki C, Murakami K, Fujii Y, et al. The H-Invitational Database (H-InvDB), a comprehensive annotation resource for human genes and transcripts. *Nucleic Acids Res* 2008 ; 36 : D793-9.
20. Navratil V, de Chasseay B, Meyniel L, et al. VirHostNet: a knowledge base for the management and the analysis of proteome-wide virus-host interaction networks. *Nucleic Acids Res* 2009 ; 37 : D661-8.
21. Chatr-aryamontri A, Ceol A, Peluso D, et al. VirusMINT: a viral protein interaction database. *Nucleic Acids Res* 2009 ; 37 : D669-73.
22. Aranda B, Achuthan P, Alam-Faruque Y, et al. The IntAct molecular interaction database in 2010. *Nucleic Acids Res* 2010 ; 38 : D525-31.
23. Tai AW, Benita Y, Peng LF, et al. A functional genomic screen identifies cellular cofactors of hepatitis C virus replication. *Cell Host Microbe* 2009 ; 5 : 298-307.
24. Krishnan MN, Ng A, Sukumaran B, et al. RNA interference screen for human genes associated with West Nile virus infection. *Nature* 2008 ; 455 : 242-5.
25. Hao L, Sakurai A, Watanabe T, et al. Drosophila RNAi screen identifies host genes important for influenza virus replication. *Nature* 2008 ; 454 : 890-3.
26. Brass AL, Dykxhoorn DM, Benita Y, et al. Identification of host proteins required for HIV infection through a functional genomic screen. *Science* 2008 ; 319 : 921-6.
27. Forst CV. Host-pathogen systems biology. *Drug Discov Today* 2006 ; 11 : 220-7.
28. Carvunis AR, Gomez E, Thierry-Mieg N, et al. Biologie systématique : des concepts d'hier aux découvertes de demain. *Med Sci (Paris)* 2009 ; 25 : 578-84.
29. Barillot E, Calzone L, Zinovyev A. Biologie des systèmes appliqués aux cancers. *Med Sci (Paris)* 2009 ; 25 : 601-7.
30. Thieffry D. Biologie des systèmes : des réseaux d'interactions moléculaires à la médecine personnalisée. *Med Sci (Paris)* 2009 ; 25 : 547-8.
31. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature* 2001 ; 409 : 860-921.
32. Stumpf MP, Thorne T, de Silva E, et al. Estimating the size of the human interactome. *Proc Natl Acad Sci USA* 2008 ; 105 : 6959-64.
33. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet* 2004 ; 5 : 101-13.
34. Lesne A. Biologie des systèmes : l'organisation multiéchelle des systèmes vivants. *Med Sci (Paris)* 2009 ; 25 : 585-7.
35. Haiech J, Galli T. Biologie systématique, biologie synthétique et nanobiotechnologies : le tryptique du XXI^e siècle. *Med Sci (Paris)* 2009 ; 25 : 576-7.

TIRÉS À PART

V. Navratil

