

[COVID Information Commons \(CIC\) Research Lightning Talk](#)

[Transcript of a Presentation by Marc Riedel \(University of Minnesota\), July 16, 2021](#)



Title: [EAGER: Computationally Predicting and Characterizing the Immune Response to Viral Infections](#)

[Marc Riedel Database Profile](#)

NSF Award #: [2036064](#)

[YouTube Recording with Slides](#)

[July 2021 CIC Webinar Information](#)

Transcript Editor: Julie Meunier

Transcript

Slide 1

Bonjour à tous. Je parlerai aujourd'hui de travaux de calcul sur le problème de la liaison des peptides, en collaboration avec George Vasmatzis de la Mayo Clinic.

Slide 2

En ce qui concerne le COVID-19, nous avons tous entendu dire que la maladie se comporte comme n'importe quelle autre maladie. La gravité des symptômes varie d'une personne à l'autre, mais elle est peut-être plus prononcée avec le COVID-19 qu'avec d'autres maladies. Une bonne partie des personnes qui contractent le virus ne présentent aucun symptôme. La plupart d'entre elles ne présentent aucun symptôme ou présentent des symptômes légers, mais une fraction importante présente des symptômes graves. Et bien sûr, un petit nombre de personnes présentent des symptômes critiques, voire meurent. De nombreux facteurs influencent donc la gravité de la maladie et les experts peuvent en parler mieux que moi. L'âge, le sexe et, en particulier, les comorbidités sont très importants et l'exposition passée à des virus similaires joue un rôle, mais une partie de l'équation en termes de gravité de la maladie est également constituée par les différences innées de nos systèmes immunitaires. En fait, les gènes que nous avons hérités de nos parents jouent un rôle dans la capacité de notre organisme à combattre cette maladie ou la plupart des maladies. Notre sujet d'étude consiste donc à prédire par calcul un aspect de la différence innée dans ce qu'on appelle la réponse cellulaire du système immunitaire d'un individu.

Slide 3

La réponse cellulaire est donc la première ligne de défense de notre organisme face à une infection virale. Il s'agit du mécanisme par lequel les peptides étrangers introduits dans les cellules sont découpés en petits fragments appelés peptides. Ceux-ci peuvent alors être transportés à la surface des cellules où ils peuvent se lier à des récepteurs de la surface cellulaire appelés molécules de classe 1 du CMH

[complexe majeur d'histocompatibilité]. Si elles sont ainsi liées, les cellules infectées deviennent des cibles pour les lymphocytes T tueurs qui peuvent venir tuer les cellules infectées - qu'elles soient tuées par une infection virale ou, comme il s'avère, qu'il s'agit de la défense la plus efficace dont nous disposons contre le cancer. Les cellules cancéreuses sont également éliminées de cette manière. Il s'agit de la première ligne de défense, car si elle se met en place une fois que nous sommes infectés, le système immunitaire - les cellules T tueuses - peut tuer toutes les cellules infectées avant qu'elles n'aient une chance de se développer. Mais si la réponse cellulaire échoue, les cellules infectées deviennent des usines. Elles commencent à produire de nombreuses copies du virus et d'autres aspects du système immunitaire doivent prendre le relais.

Slide 4

La réponse cellulaire est donc - comme tous les aspects de l'immunologie - complexe. Il est très important de la comprendre. Elle est essentielle pour comprendre et prédire la gravité des nouveaux virus tels que le SRAS-CoV-2, pour le développement de vaccins ciblant les virus, ainsi que pour comprendre l'impact des mutations virales. Je reviens sans cesse sur le sujet de l'immunothérapie du cancer, qui est le domaine d'expertise de mon collaborateur. Comme beaucoup de gens, nous nous sommes intéressés à COVID-19 et nous avons réorienté les compétences. Ainsi, dans ce cas, les prédictions informatiques ciblent à l'origine l'immunothérapie du cancer. Bien entendu, la même théorie s'applique aux maladies auto-immunes. La réponse immunitaire cellulaire est donc, à la base, un problème de calcul. Si nous avons les peptides bleus dans la figure de droite, il s'agit du fragment de protéine associé au virus. La question est la suivante : ce peptide bleu va-t-il se lier à l'intérieur d'un sillon ou d'une fente à l'intérieur de la molécule de surface de la cellule jaune, la molécule CMH 1 ? La réponse immunitaire dépendra alors de la fixation de ce fragment de protéine et du fait qu'il se fixe suffisamment bien pour que les cellules T tueuses le reconnaissent. Les fragments bleus provenant du virus sont nouveaux. Ainsi, avec un nouveau virus, nous aurons un ensemble de peptides complètement nouveau. Les molécules jaunes de la surface cellulaire sont déterminées par nos gènes et chaque individu possède jusqu'à six molécules différentes du CMH 1 déterminées par notre héritage - trois de chaque parent. Les molécules de surface cellulaire, les molécules du CMH, sont parmi les plus diverses de notre génome ; il existe environ 21 000 variantes dans la population humaine. Ce n'est pas un hasard. L'évolution a fait en sorte que cet aspect de notre système immunitaire soit très diversifié afin que nous puissions survivre aux infections virales passées.

Slide 5

Mais cela pose également un défi informatique très important, comme je vais le décrire. Le problème central auquel nous nous attaquons est donc de savoir comment appliquer l'informatique pour prédire si le peptide bleu se liera à la molécule jaune de la surface cellulaire, et ce pour une très large gamme de peptides, tous ceux qui sont associés au virus, et pour une très large gamme de récepteurs de la surface cellulaire, les molécules du CMH 1. Il y a eu beaucoup de travaux antérieurs - des travaux computationnels antérieurs. Les travaux expérimentaux ont bien sûr fourni des informations moléculaires sur la structure de ces molécules, et les travaux informatiques qui ont été très fructueux ont consisté à appliquer un apprentissage automatique - ce qui n'est pas une surprise pour ceux qui ont une formation en informatique et en réseaux neuronaux. Il existe donc un logiciel appelé NetMHC qui a été formé de manière intensive sur les données expérimentales avec la force de liaison pour les paires connues de molécules bleu-jaune - les paires de peptides du CMH 1. Sur la base d'une structure de

réseau neuronal, ce programme peut prédire la force de liaison d'un nouveau peptide. Cette inférence est basée sur une simple séquence peptidique, de sorte que la séquence de lettres, c'est-à-dire les acides aminés du peptide, est associée à une étiquette correspondant à la molécule de surface cellulaire, la molécule du CMH 1. Il y a une force. Ce sont toutes les données expérimentales. Ainsi, avec plusieurs milliers de paires de ce type, le réseau neuronal peut être entraîné. Une fois qu'il est entraîné, il peut prédire, à partir d'une nouvelle séquence peptidique, d'une nouvelle séquence d'acides aminés, dans quelle mesure ce peptide se liera à une molécule donnée du CMH 1. Ici, la séquence peptidique serait la molécule peptidique bleue dans le dessin, et l'étiquette correspondrait au récepteur jaune de la surface cellulaire, la molécule du CMH 1. Les réseaux neuronaux sont puissants dans la mesure où ils peuvent être facilement entraînés et où ils peuvent très efficacement faire des prédictions sur la base des données qui leur sont fournies. Mais la limite ici est qu'il n'y a absolument aucune donnée moléculaire, nous formons simplement des étiquettes et des lettres, et les données de formation proviennent d'un ensemble très diversifié de données expérimentales. Une grande partie de ces données provient par exemple du VIH, un virus différent, et des peptides associés au VIH. Le problème est qu'avec un tout nouveau virus comme le SARS-CoV-2, la plupart des peptides n'ont jamais été vus auparavant, et un réseau neuronal fera des prédictions erronées parce qu'il fait des déductions à partir d'une région très différente de l'espace peptidique.

Slide 6

Notre approche consisterait donc à effectuer des simulations au niveau moléculaire et, bien entendu, de nombreux travaux ont déjà été réalisés sur ce sujet. Des techniques de simulation moléculaire très sophistiquées sont connues et largement utilisées. L'une d'entre elles s'appelle la dynamique moléculaire. Il existe également des simulations basées sur la méthode de Monte Carlo. Elles utilisent une technique appelée recuit simulé, et le docking moléculaire est une autre approche. Les logiciels disponibles pour ces simulations au niveau moléculaire sont largement utilisés, mais ils sont spéciaux. Ils sont d'usage général. Ils ont été développés pour de larges classes de molécules à lier et sont extrêmement gourmands en ressources informatiques. Si l'on prend un peptide, une paire du CMH 1, et que l'on utilise les logiciels existants pour le simuler, il faut des jours, voire des semaines, pour simuler un seul événement de liaison. Il faut donc des semaines de supercalculateur pour faire une seule prédiction. L'ampleur du problème auquel nous sommes confrontés est la suivante : nous avons 21 000 variantes des molécules de surface des cellules jaunes, les molécules CMH1, et pour le SARS-CoV-2, si nous nous concentrons uniquement sur la protéine de pointe et que nous la découpons en petits morceaux pour les peptides, nous en avons environ 38 000. Nous parlons donc d'un milliard de combinaisons que nous voulons simuler en termes de souche, et s'il faut une semaine de temps de supercalculateur pour chacune, nous ne disposons évidemment pas d'un milliard de semaines pour étudier cela.

Slide 7

Notre approche est donc double. D'une part, nous créons un logiciel hautement personnalisé pour la simulation moléculaire et nous utilisons les détails du domaine dans lequel nous travaillons. Nous commençons donc par le peptide. Les peptides ne varient pas tellement en termes de longueur ou de forme. Nous commençons avec les peptides correctement alignés à l'intérieur de la fente de la molécule de surface cellulaire, la molécule du CMH 1, de sorte que nous ne passons pas beaucoup de temps à faire tourner l'ensemble du peptide dans l'espace. Nous le plaçons exactement là où il doit être et nous

effectuons toute la recherche dans l'espace de torsion. Ainsi, au lieu de déplacer la molécule dans l'espace tridimensionnel, nous nous contentons de tordre et de tourner ses liaisons pour essayer de trouver la configuration optimale. L'autre contribution est que nous déployons ce système à grande échelle. Nous utilisons donc des GPU et, à terme, une infrastructure informatique en nuage, pour mettre toute la puissance de calcul au service du problème. Et notre objectif, comme je l'ai indiqué dans le titre, est de transformer un milliard de jours en un million de minutes ou peut-être un mois de temps de calcul en nuage.

Slide 8

Le défi est donc là et je vais en parler brièvement - je n'ai presque plus de temps. Une chose : les données expérimentales ne sont pas complètes. Nous ne disposons pas de modèles moléculaires complets, non seulement pour les 21 000 variantes, mais nous ne disposons même pas d'une bonne répartition géographique. Cela rejoint certains des exposés précédents. La plupart des données expérimentales concernent la variance des molécules du CMH 1 issues de la démographie caucasienne occidentale. L'autre approche consiste donc à déduire la structure des molécules, puis à les simuler.

Slide 9

Dernière diapositive : l'impact de ces travaux. Il est relativement simple de déterminer les gènes d'un individu qui codent pour les molécules concernées, les molécules du CMH 1. Cela peut se faire par le biais du typage HLA, qui est utilisé pour les tests de paternité. Grâce à ces informations et à notre infrastructure informatique, nous serons en mesure de prédire la réaction d'un individu à un nouveau pathogène, à un virus, à des variantes du virus pour différents individus, ainsi que l'effet de différents vaccins sur différents virus pour différents individus. Et bien sûr, comme je l'ai dit au début, ce travail ne s'appliquera pas seulement aux virus, mais aussi potentiellement à l'immunothérapie du cancer et aux maladies auto-immunes. Je m'arrêterai donc là. Je vous remercie de votre attention.