

Queste de savoir

Participer à la recherche contre le covid

5 mars 2021

Table des matières

1.	Une solution médicale	1
1.1.	Qu'est-ce qu'un virus?	2
1.2.	Un médicament contre les virus	2
2.	Simuler une protéine	3
3.	Calcul distribué contre le coronavirus	7
3.1.	Rosetta at Home	8
3.2.	Folding at Home	9
3.3.	World Community Grid	12



Ce texte a été écrit il y a un an, vers le début de l'épidémie. Il n'a pas passé la validation, et depuis j'ai perdu le temps et la motivation de le retravailler. Prenez-le pour ce qu'il est: un document à la fois pas tout à fait mature, et dont les informations d'actualité ne le sont peut-être plus, mais avec pas mal d'explications sympas qui valent le coup d'être lues. C'est pour cette raison que je le publie aujourd'hui en tant que billet.

Vous l'avez sûrement remarqué, il y a une épidémie mondiale qui ravage tout en ce moment. Si non, dites-moi où vous vivez, parce que c'est certainement l'un des derniers endroits épargnés sur Terre. Et peut-être vous êtes vous demandés comment vous pourriez aider, pour atténuer un peu les conséquences de cette crise.

Et bien pour commencer, la manière la plus efficace pour aider, c'est de respecter les consignes de votre gouvernement: qu'elles recommandent de rester chez soit, de porter un masque, ou de ne pas acheter trop de médicaments et de papier toilette en une fois, ces consignes ne seront efficaces que si une grosse majorité les respecte, donc essayez de vous tenir informé.

Mais il y a d'autres manières! L'objet de cet article est de parler un peu de calcul distribué et beaucoup de simulations atomistiques, et de comment ça peut aider contre le coronavirus. Et après l'avoir lu, vous devriez comprendre comment ces calculs sont faits, et les différences entre les programmes qui permettent de participer à la recherche. Allons-y!



Un très grand merci à @pierre_24 pour ses relectures et à @Ekron pour le logo

1. Une solution médicale

Cette section sera courte, parce que ce n'est pas du tout là où j'ai mon expertise. Je souhaite simplement donner un peu de contexte, pour que les explications qui suivent soient plus faciles à comprendre.

1. Une solution médicale

1.1. Qu'est-ce qu'un virus ?

Le coronavirus est donc un virus. Le «but» ultime d'un virus est de se répliquer. C'est pas forcément très recherché, mais on a fait pire, comme vous pouvez le voir sur la vidéo ci-dessous. Le virus est peut-être vivant, sûrement pas, ça dépend un peu des définitions. Et il peut être bénin ou tout dérégler dans un corps humain, ce qui n'est pas idéal, vous en conviendrez.

ÉLÉMENT EXTERNE (VIDEO) —

Consultez cet élément à l'adresse <https://www.youtube.com/embed/wqzLoXjFT34?feature=oembed>.

Un virus, et vous pouvez en lire beaucoup plus sur [wikipedia](#) ou d'autres sites, est composé d'un acide nucléique avec l'information génétique, entouré par une capsid composée de protéines. Cette capsid a plusieurs fonctions, comme protéger l'acide nucléique, favoriser l'attachement à une cellule et permettre de pénétrer à l'intérieur d'une cellule. Et certains virus ont une autre enveloppe tout autour, composée de lipides et de protéines. Dans le cas du coronavirus, il y a une enveloppe, et l'acide nucléique est sous la forme **ARN**.

Pour se répliquer, un virus a besoin d'infecter une cellule, qu'il va corrompre pour qu'elle produise les protéines nécessaires. Il réplique son génome, réassemble tout pour faire un nouveau virus, et sort de la cellule. Il ne peut aucunement se répliquer seul, sans l'aide d'une cellule.

Dans le cas du coronavirus, ou SARS-CoV-2, il y a quatre types de protéines, dont celles appelées protéine S, pour *spike*, protéine de pointe en français. Ce sont celles-ci qui permettent, en résumé, de pénétrer dans la cellule à infecter.

1.2. Un médicament contre les virus

L'idée d'un médicament peut être d'aider à arrêter la multiplication du virus, de l'empêcher de créer des réactions négatives dans le corps humain, ou plus simplement de l'exterminer. Et pour se faire, on va chercher dans un premier temps comment le virus performe ces actions néfastes, et ensuite, comment l'en empêcher.

On vient de le dire, le virus n'est probablement pas vivant, ne cherche pas du tout à vous tuer, mais peut le faire en cherchant à se répliquer. Ses fonctions sont entièrement déterminées par sa composition. Il ne peut s'adapter que grâce à des mutations aléatoires, et qui dans l'extrême majorité des cas, seront néfastes pour le virus. Et donc, des chercheurs très intelligents, certainement plus que moi, ont déterminé que si on connaît la forme et la composition des protéines qui composent le virus et qui participe à son action, on comprendra comment il agit, et donc indirectement, comment le neutraliser.

On a été un peu vite sur la dernière phrase, en particulier sur ce qu'on apprend de la composition et de la forme de la protéine: reprenons. Un virus est composé de nombreuses protéines, comme on l'a vu plus haut. Ces protéines se replient sur elle-mêmes, ce qui va créer des sites topologiques

2. Simuler une protéine

qui vont permettre différentes fonctions. C'est illustré sur l'image juste en dessous. Ces sites vont permettre de bien positionner les réactifs (du virus et de la cellule cible) pour que la réaction correspondant à une fonction (rentre dans la cellule, se répliquer, ...) se produise.

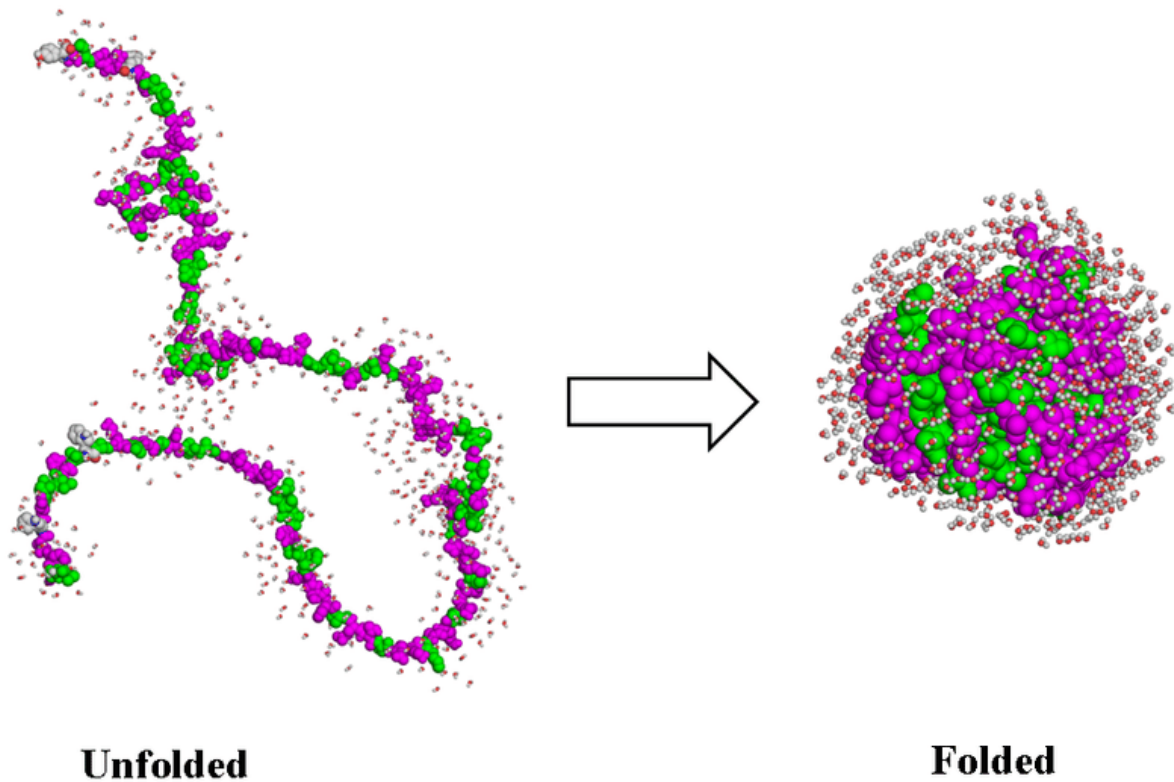


FIGURE 1.1. – Une protéine qui se replie. Source: "ProFAX : A hardware acceleration of a protein folding algorithm", Guidi et al., 2016.

Donc, une fois que ces protéines sont repliées, l'idée est d'observer les atomes qui en forment l'enveloppe, et leur forme, trouver ce qui permet au virus d'être dangereux, et d'ensuite proposer une molécule qui s'y fixera, ou une manière de déplier et replier différemment la protéine, et ce-faisant, rendre le virus inoffensif. Facile, non?

2. Simuler une protéine

Du coup vous l'aurez compris, le nerf de la guerre est de simuler une protéine et peut-être une molécule additionnelle pour voir son effet sur ladite protéine. Et peut-être des millions de fois cette étape, pour trouver une molécule qui pourrait marcher. Et une fois que c'est fait, recommencer sur toutes les protéines utiles du virus. La première question est donc de savoir comment on sait comment une molécule va se replier, et comment on le prédit.

2. Simuler une protéine

Et bien ici, la réponse est à la fois simple et extrêmement complexe: la partie simple, c'est que comme pour toute chose dans l'univers, plus l'énergie est basse, plus c'est stable. Il suffit donc de calculer l'énergie de la protéine dans différentes configurations. La partie compliquée, c'est de calculer l'énergie de la protéine.

On est à l'échelle atomique, là où la physique quantique règne sans partage. On va donc tout simplement partir de l'équation de Schrödinger.

$$H\psi = E\psi$$

Avec H l'Hamiltonien, ψ la fonction d'onde et E l'énergie.

En a rarement vu équation plus simple. Bon, bien sûr, la fonction d'onde ψ est vraiment compliquée à écrire pour une molécule de la taille d'une protéine moyenne. Et l'hamiltonien H , ou l'opérateur de l'énergie (la fonction qui va nous donner la valeur propre correspondant à un état propre) est aussi complexe, avec des termes pas loin d'être impossibles à estimer. Vous n'avez pas besoin de comprendre cette partie, comprenez seulement que H va permettre de trouver l'énergie d'un système) devient aussi extrêmement complexe dans des systèmes larges (plus que quelques particules) à cause des effets multicorps.

Les écritures de H et ψ sont très en dehors du champ de cet article, mais modéliser quelques centaines d'atomes en utilisant cette méthode peut prendre plusieurs heures sur quelques centaines de processeurs. Il existe des algorithmes qui permettent d'atteindre l'énergie la plus basse de manière systématique, mais ils ne fonctionnent que si on a une bonne idée de ce qu'on veut trouver (ou si on est d'accord pour faire des milliers de simulations).

Résoudre tout ça, vous l'aurez compris, se fait sur des superordinateurs, et une fois de plus, vous ne pouvez pas aider. Mais qu'en serait-il si on simplifiait tout? Si au lieu de la physique quantique, on utilisait de la physique classique? Qu'on se contentait d'utiliser un potentiel qui donne une idée correcte sans être exceptionnelle de ce qui se passe? Et bien dans ce cas-ci, oui, une configuration pourrait être calculée en une fraction de seconde même sur un ordinateur portable bas de gamme. Et ça tombe bien, puisqu'il faudrait calculer des centaines de milliers de cas, au bas mot. Différentes méthodes sont disponibles et seront détaillées plus tard, mais on va déjà voir les principes généraux.

Et pour ces principes généraux, on va se concentrer sur un atome (contre un électron en physique quantique). Si vous avez besoin d'un rappel sur le sujet, je vous invite à aller lire [ce court tutoriel](#) [↗](#). On a donc un atome, entouré d'autres atomes, et ce qu'on aimerait bien savoir, c'est son énergie. Et on va l'obtenir grâce à un potentiel.

Un potentiel que vous connaissez instinctivement, c'est celui d'un ressort. Si vous tirez dessus, ça vous demande de l'énergie, et pareil si vous voulez le compresser. Il y a une position d'équilibre où l'énergie est minimale, et où le ressort est stable.

2. Simuler une protéine

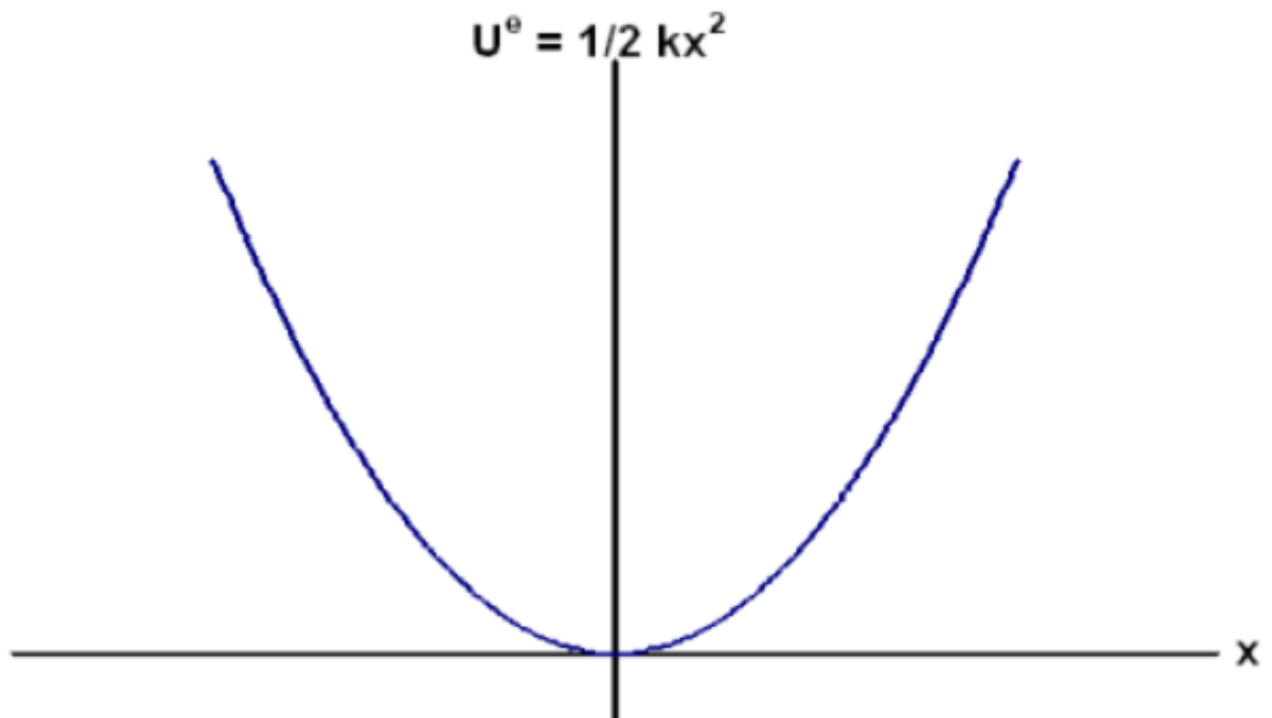


FIGURE 2.2. – Énergie potentielle d'un ressort.

Maintenant, retournons à l'échelle atomique. Un potentiel tout simple est de mettre des ressorts entre tous les atomes.

Pour un cristal, ça donnerait quelque chose comme ça.

2. Simuler une protéine

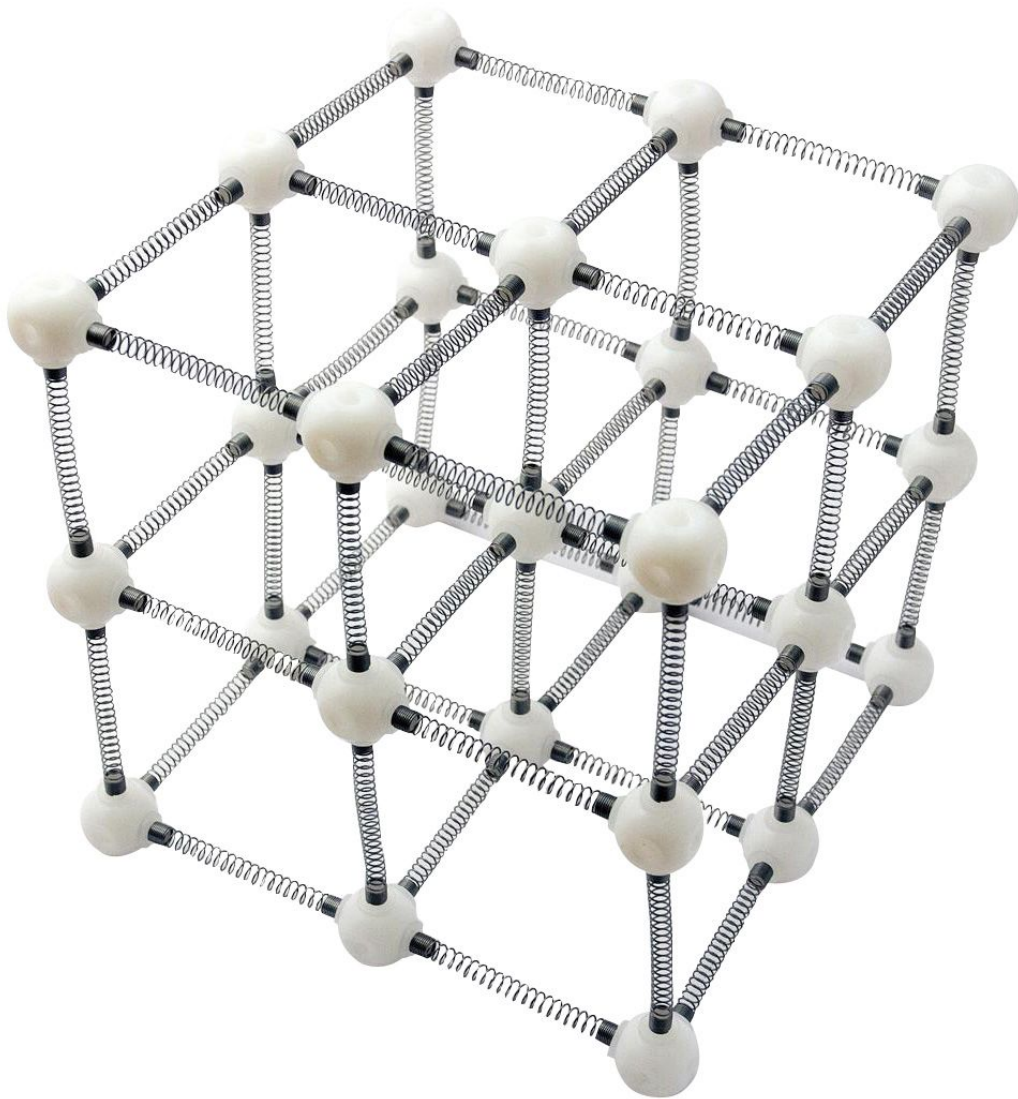


FIGURE 2.3. – Un cristal avec des ressorts entre les atomes. Achetable [ici](#) .

Ce serait donc un potentiel tout simple, que l'on appelle "binaire", qui va uniquement associer une énergie à deux éléments chimiques (qui peuvent être les mêmes) et une distance. Par exemple, si deux atomes de fer sont à une distance de 5 Ångströms, il contribueront une énergie X au total du système, et $X+dX$ si il sont à une distance plus petite ou plus grande que 5Å. D'autres sont plus compliqués, et prendront en compte des triplets d'atomes, les angles entre les atomes, différents types d'interaction, différentes distances maximales d'interaction, etc. Au final, on cherche à avoir une représentation de l'énergie en tout point de l'espace, comme sur la figure ci-dessous (sauf que c'est pour un plan, parce que j'ai du mal à dessiner en 4D). Les atomes chercheraient à être dans les parties où le potentiel est le plus bas.

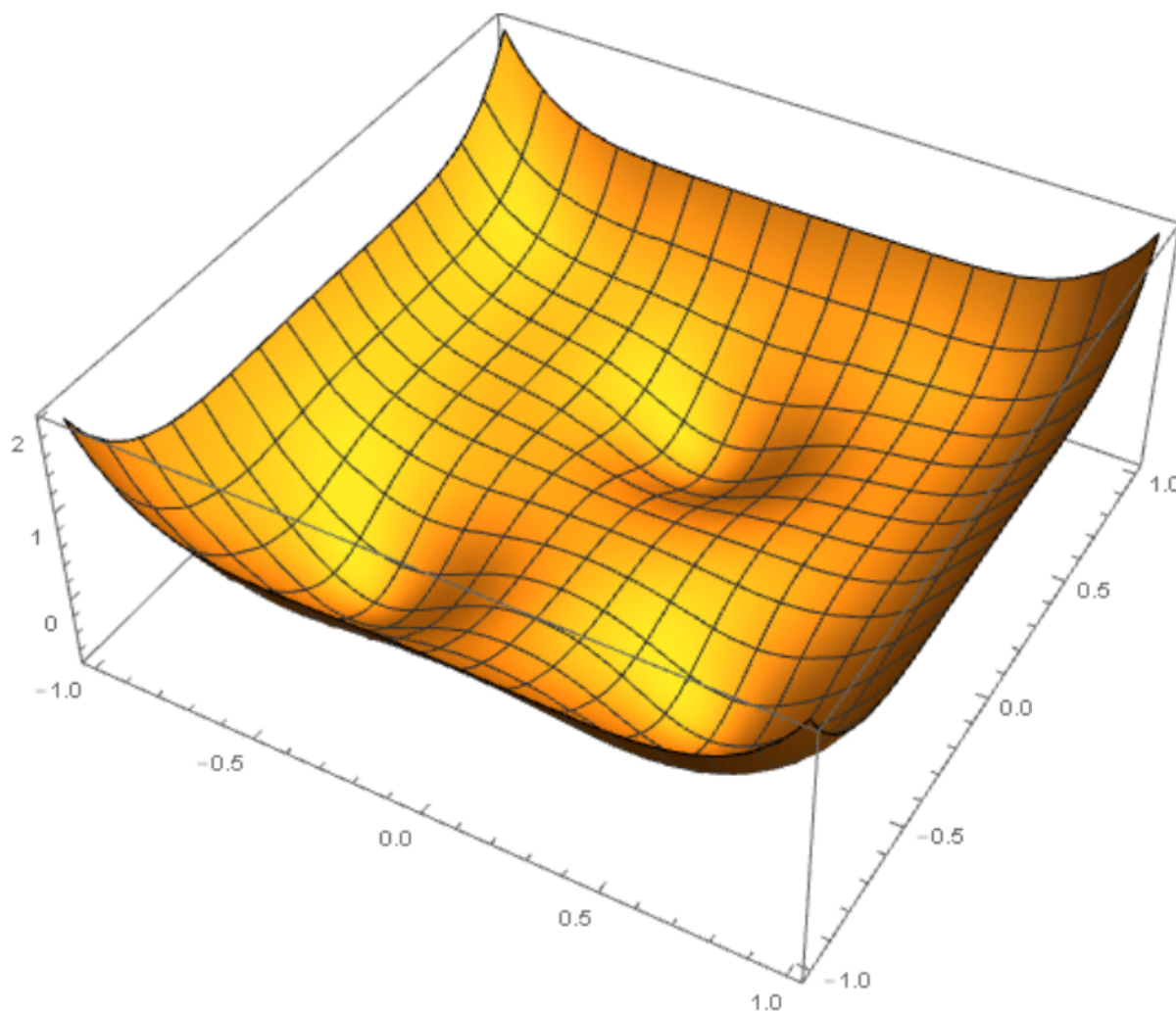


FIGURE 2.4. – Représentation de l'énergie potentielle

Un potentiel de bonne qualité peut prendre des années ou des décennies à être créé, et bien souvent, ce qu'on recherche, c'est quelque chose de "suffisamment bon" pour donner une idée. Ensuite, les meilleures idées seront simulées avec une méthode beaucoup plus lente, mais beaucoup plus précise.

3. Calcul distribué contre le coronavirus

Le calcul distribué consiste à prendre une tâche gargantuesque et à la séparer en de nombreuses petites tâches. Ensuite, ces tâches sont envoyées vers les ordinateurs de volontaires, qui vont travailler dessus dès qu'ils seront oisifs. Vous pourrez ainsi utiliser votre ordinateur de manière transparente, au maximum de ses possibilités, mais quand vous irez manger ou aux toilettes, lui ne se reposera pas et travaillera pour faire avancer la science.

Il existe à ce jour plusieurs initiatives de calcul distribué, pour trouver un médicament contre le coronavirus.

i

Ces initiatives continueront bien après, sur d'autres maladies.

Je vais ici en citer trois, parce qu'elles sont les plus connues et les plus à même d'obtenir des résultats exploitables.

3.1. Rosetta at Home

Rosetta at Home (RaH) va essayer de déterminer l'état fondamental (c'est à dire l'état avec l'énergie la plus basse, l'état dans lequel le système se met spontanément) des différentes protéines qui constituent le coronavirus. Pour ce faire, il va employer la méthode **Metropolis Monte Carlo**. Cette méthode part d'une configuration initiale, et essaie de changer la position de certains atomes. À chaque fois, l'énergie totale du système est recalculée, et de là, deux possibilités:

- L'énergie a baissé, et la nouvelle configuration est retenue.
- L'énergie n'a pas baissé, et soit on rejette la nouvelle configuration, soit on l'accepte avec une probabilité basse, qui dépend de l'augmentation de l'énergie.

$$a_{AB} = \begin{cases} 1, & \text{if } E_B \leq E_A \\ \exp(-\beta(E_B - E_A)), & \text{if } E_B > E_A \end{cases}$$

FIGURE 3.5. – Probabilité d'acceptation d'une nouvelle configuration

Il est utile de ne pas avoir une probabilité nulle quand l'énergie augmente pour éviter de se trouver dans un minimum énergétique local, qui passerait toutes les vérifications, mais donnerait des résultats non-exploitable (voir la figure plus bas). Cette méthode est "rapide", mais ne donne aucune information sur la cinétique de la protéine, seulement sur l'état fondamental. La parallélisation extrême est faite d'une part avec des configurations initiales différentes, et d'autre part avec des algorithmes différents pour favoriser la baisse de l'énergie.

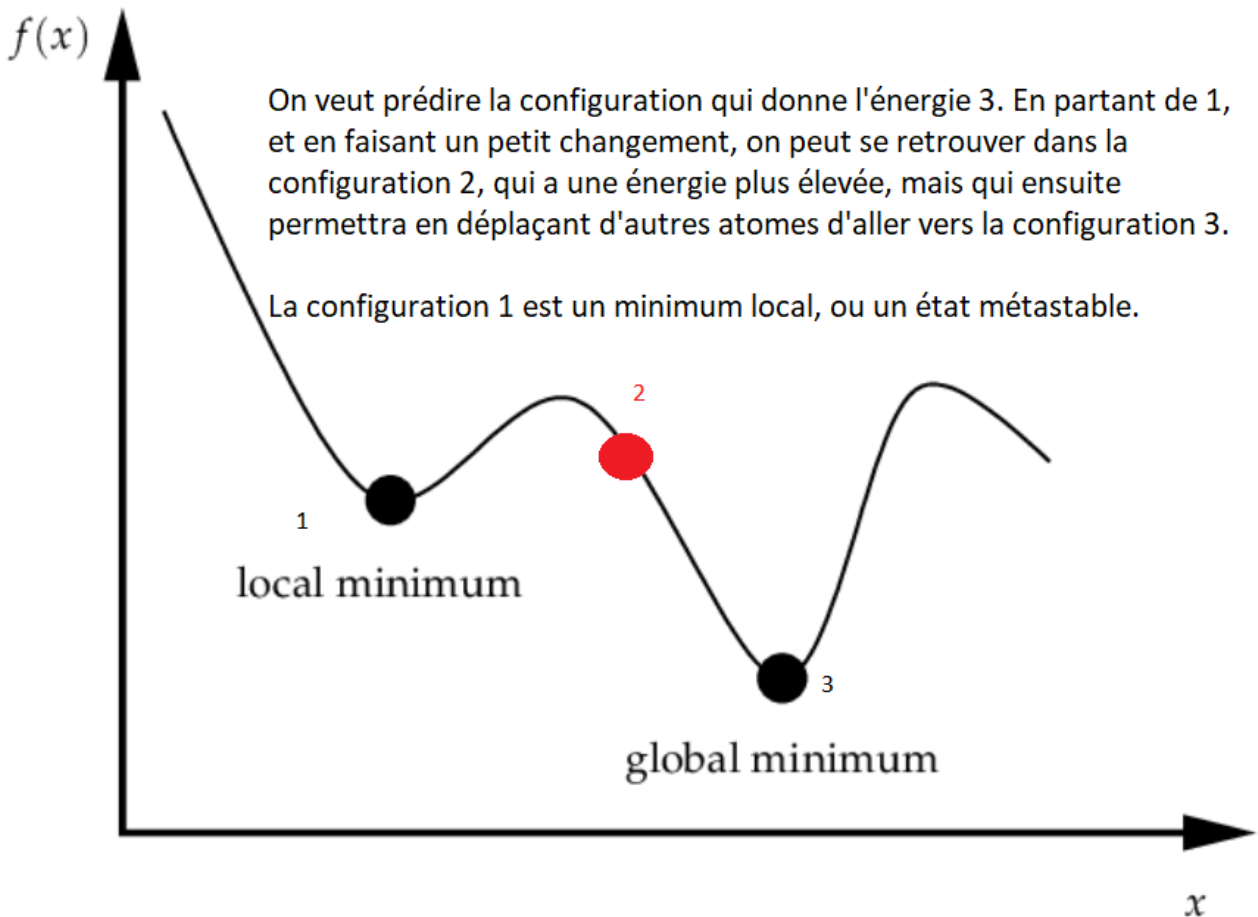


FIGURE 3.6. – illustration d'un minimum local

Grâce à RaH, l'état fondamental des protéines (la manière de se replier) peut être déterminé. Vous pouvez y contribuer à travers BOINC, disponible par [ici](#) .

3.2. Folding at Home

Folding at Home (FaH) va pour sa part essayer de simuler l'évolution d'une protéine avec le temps et la température. De cette manière, l'évolution des différents sites peut être connue, et une molécule qui ne serait peut-être pas efficace sur l'état fondamental pourrait l'être sur un état presque fondamental, que l'on rencontrera parfois.

Ces simulations ont pour petit nom "dynamique moléculaire", et prennent une éternité pour atteindre ce qui n'est qu'une micro-seconde. Même un millième de ça. Du coup, demander de l'aide était tout naturel. Pour aller plus loin, vous pouvez consulter les liens suivants sur les différents champs de force (plus ou moins les potentiels définis plus haut) utilisés par FaH: [AMBER](#) and [CHARMM](#) .

3. Calcul distribué contre le coronavirus

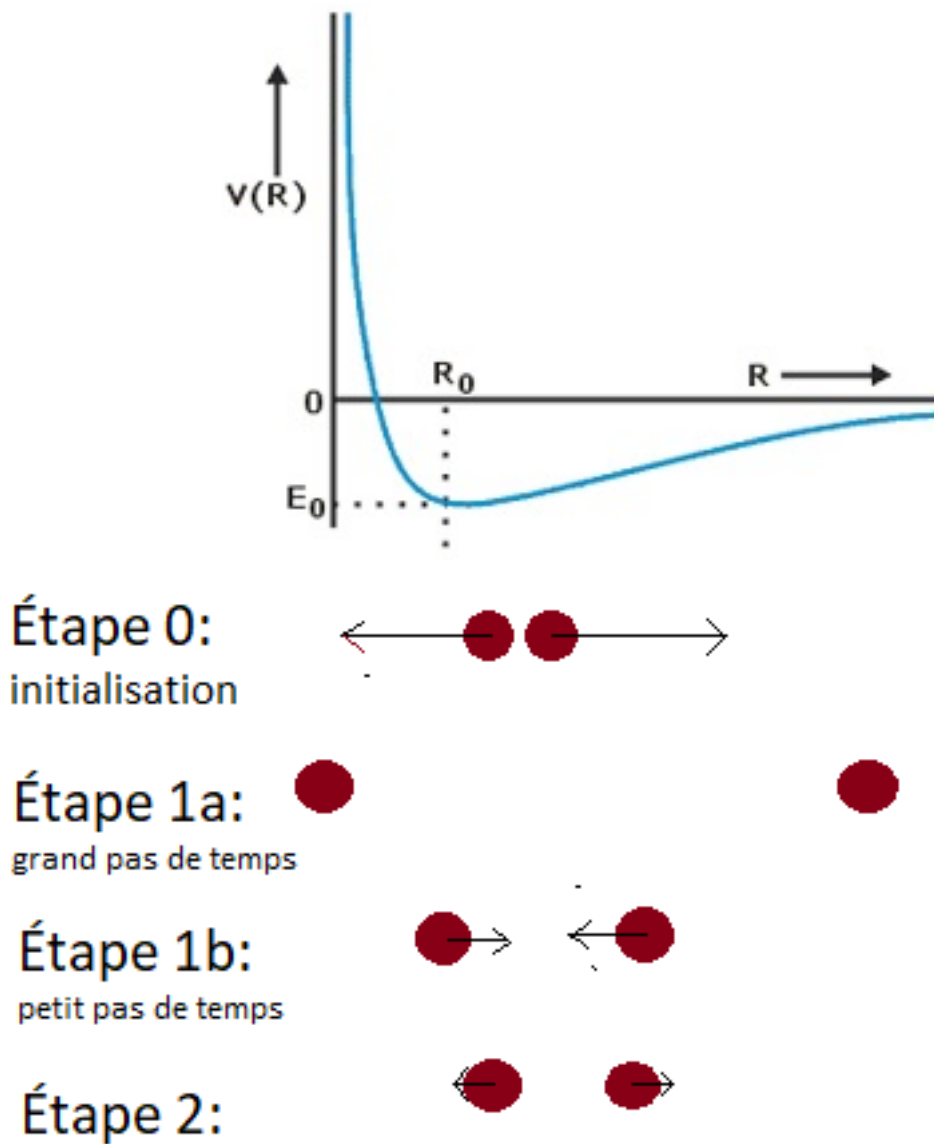


FIGURE 3.7. – illustration d'un procédé de dynamique moléculaire pour deux atomes

Sur l'image au dessus, vous pouvez voir comment la dynamique moléculaire fonctionne. Dans un premier temps, les atomes sont placés là où vous le souhaitez. En utilisant le potentiel, des forces sont calculées, et grâce à celles-ci, on va pouvoir déplacer les atomes. Comme vous pouvez le voir sur l'étape 1a, si votre pas de temps est trop grand, les atomes vont être déplacés sur de trop grandes distances, et arriver à une situation peu réaliste: dans ce cas-ci, le lien entre les atomes est cassé, et ils sont indépendants. Par contre, plus votre pas de temps est petit, plus les positions de l'étape suivante seront réalistes. Mais ça prendra bien plus d'étapes, et donc de puissance et de temps de calcul, pour arriver à des durées qui nous intéressent. Il faut trouver le juste milieu, et en général on le fait en essayant des pas de temps différents, et en regardant quand une grandeur physique converge, c'est à dire à partir de quel pas de temps elle ne change plus assez pour justifier le coût de calcul additionnel.

3. Calcul distribué contre le coronavirus



Notez que j'ai passé sous silence la vitesse initiale des atomes, qui est aussi très importante pour calculer les futures positions. On n'a pas besoin de tout mentionner ici, mais gardez à l'esprit que c'est un tout petit peu plus compliqué.

Pour la parallélisation, c'est plus subtil. En effet, en dynamique moléculaire, pour calculer l'étape suivante, il faut les positions et forces de l'étape présente. Le fait qu'il faille les position empêche une parallélisation temporelle simple, tandis que les forces dépendent de l'environnement et empêchent là aussi de simplement découper le système et d'en envoyer une partie à chacun, simplement.

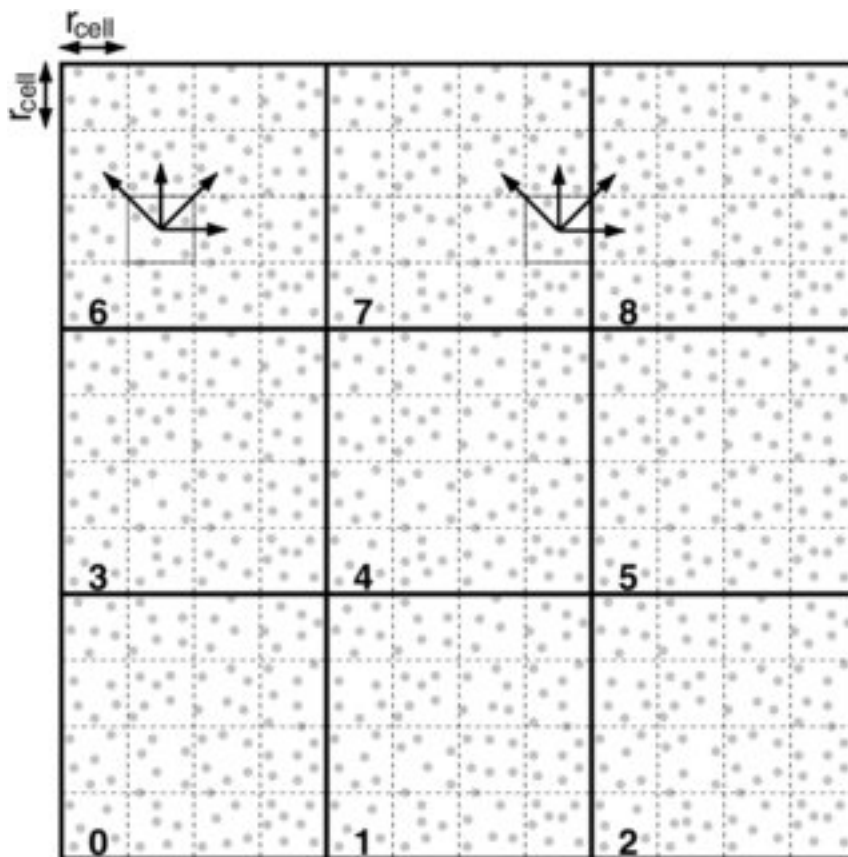


FIGURE 3.8. – illustration d'une découpe spatiale pour paralléliser la dynamique moléculaire. À chaque frontière, il faut des traitements sophistiqués qui limitent le gain et sont presque impossibles en calcul décentralisé, puisqu'il faut que les simulations communiquent. Extrait de Kadau, Germann et Lomdahl, (2004), "Large-Scale Molecular-Dynamics Simulation of 19 Billion Particles", International Journal of Modern Physics C. 27. 193–201.

Ici, il y a une technique qui consiste à simuler beaucoup de systèmes, mais avec des vitesses initiales sur chaque atome différentes. Et ensuite, selon la fréquence de transitions (aller d'une configuration à une autre) observée, il est possible de retrouver l'évolution la plus probable de la molécule pour une durée bien plus longue que chaque volontaire ne l'aura calculée.

Vous pouvez commencer à contribuer en vous rendant [ici](#) ↗.

3.3. World Community Grid

Le projet [OpenPandemics](#) vient apparemment de commencer (NDLR : au moment où l'auteur écrit cet article, vers la mi-avril), et se base sur AutoDock. Cette méthode prédit comment une molécule va se fixer sur une structure 3D connue. Il va utiliser la méthode Force Field, mentionnée au dessus, pour cette fois essayer des millions de molécules différentes sur des structures prédéfinies. La méthode étant la même, nous ne nous attarderons pas dessus.

Tout au plus, nous mentionnerons que la structure est rigide, c'est à dire qu'elle ne se déforme pas, pour que les simulations aillent plus vite ([référence](#)). Ensuite, l'idée est de définir un score reflétant la compatibilité à partir de la simulation pour chaque molécule, et d'en essayer d'autres. Ici, des tas d'optimisations sont possibles, pour ne pas essayer les molécules de manière complètement aveugle. Par exemple, des algorithmes génétiques sont fréquemment utilisés, pour proposer des molécules proches de celles qui ont eu un bon score.

Pour participer, c'est par [ici](#) !

i

Je veux juste attirer votre attention quelques secondes sur le projet [Foldit](#). Ce projet pense que les humains sont meilleurs que les ordinateurs pour trouver les meilleures configurations pour une protéine, et en a fait un jeu. Essayez si le cœur vous en dit!

Pour conclure, vous l'aurez peut-être remarqué, toutes ces approches sont complémentaires. Il serait possible en théorie de procéder comme suit:

- Utiliser RaH pour trouver les états fondamentaux les plus probables.
- utiliser une méthode plus précise pour les affiner.
- Les envoyer à FaH qui va calculer différentes configurations qui dérivent de l'état fondamental.
- Une fois de plus, affiner tout ça.
- Envoyer les résultat à WCG, pour trouver les molécules qui seraient le plus efficace pour contrer les virus.

En résumé, VOUS pouvez être un héros! Et continuer à utiliser ces logiciels une fois que l'urgence CoVid est passée.

Liste des abréviations

ARN acide ribonucléique. 2