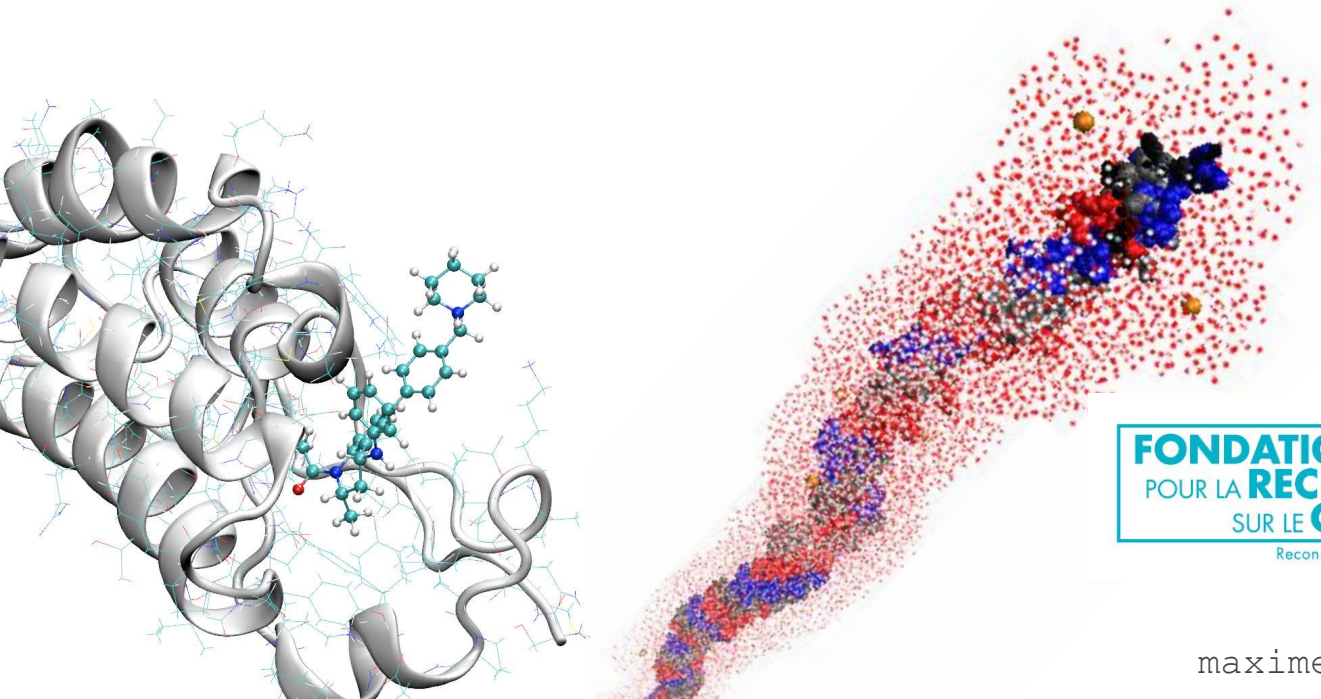


Des simulations moléculaires et multi-échelles fiables pour sonder l'auto-assemblage des fibrilles de collagène



Maxime Vassaux
Institut de Physique de Rennes
Université Rennes 1

FONDATION ARC
POUR LA **RECHERCHE**
SUR LE **CANCER**

Reconnue d'utilité publique

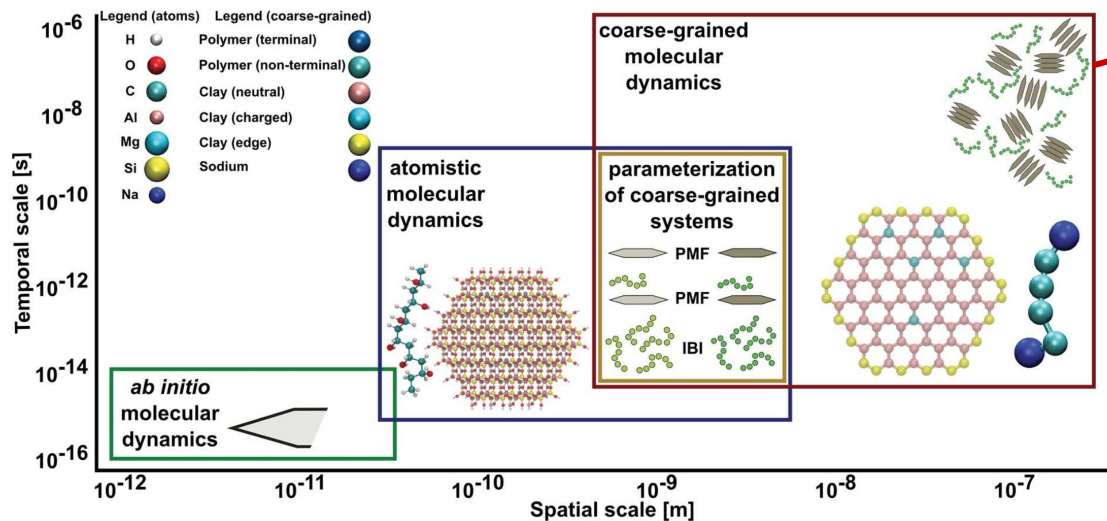
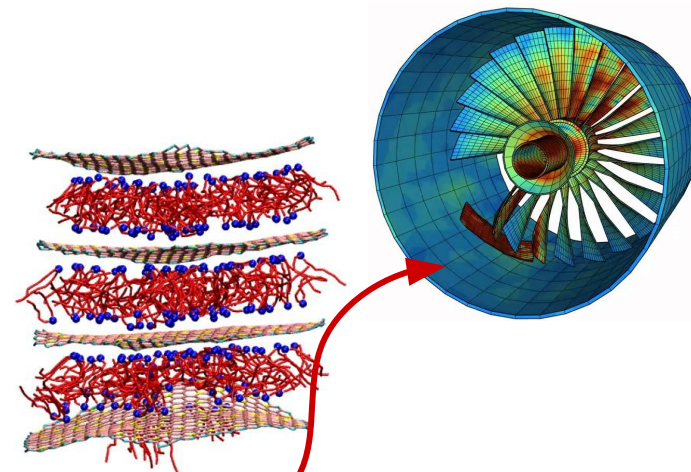


@mxvassaux

maxime.vassaux@univ-rennes1.fr

Construire un microscope virtuel *digne de confiance*

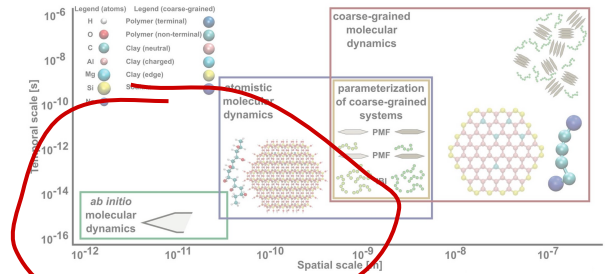
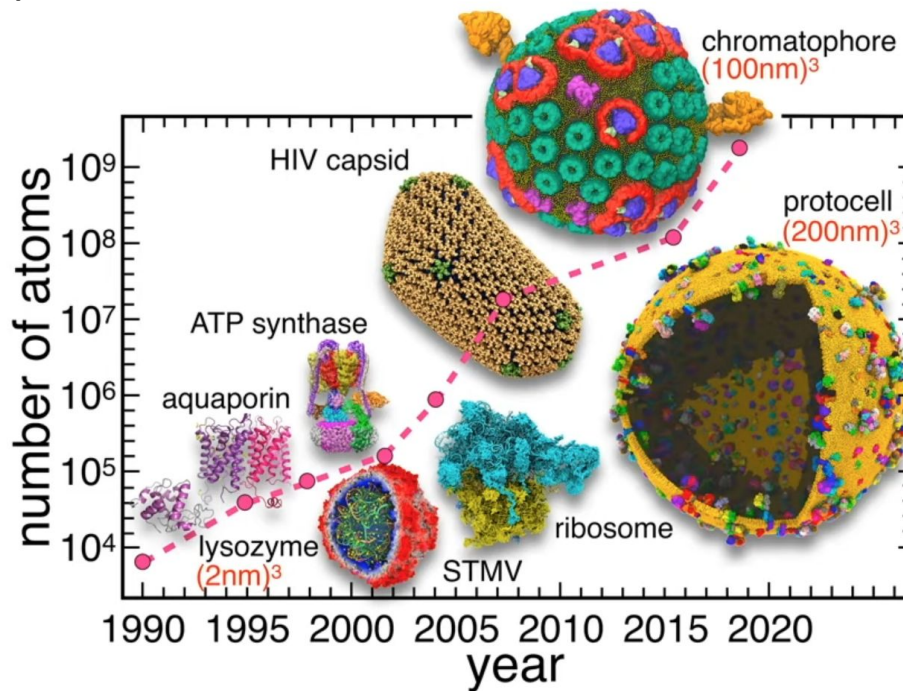
- Complémenter les expériences de laboratoire avec de nouvelles données
- Développer des outils **fiables**, vérifiés et validés
- **Prédire** en tenant compte de la **chimie**
- Intégrer de **multiples échelles**
 - chimie - nanoscopique
 - thermodynamique - microscopique
 - observable - macroscopique



[Suter, et al. Adv. Mat. 27, 6 \(2015\)](#)

La simulation moléculaire au coeur du microscope

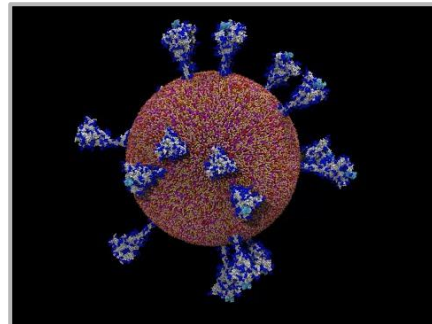
- Simulations biomoléculaires à la limite des capacités de calcul
- La Dynamique Moléculaire (DM) idéale pour les **supercalculateurs** et le **calcul parallèle**



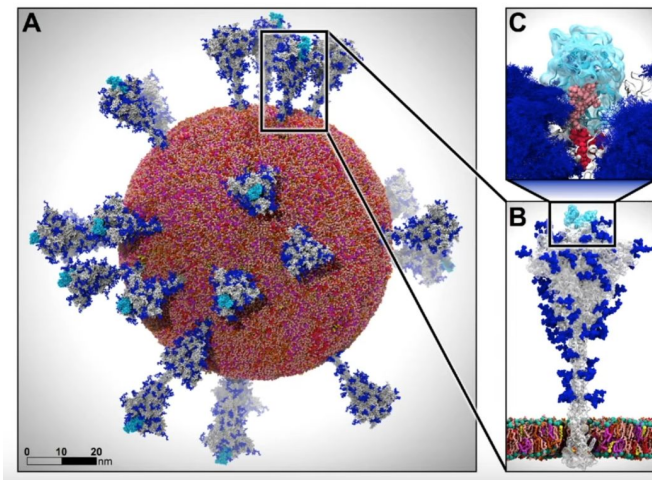
[Philips, et al. J. Chem. Phys. 153, 044130 \(2020\)](#)

Repousser les limites du calcul haute-performance

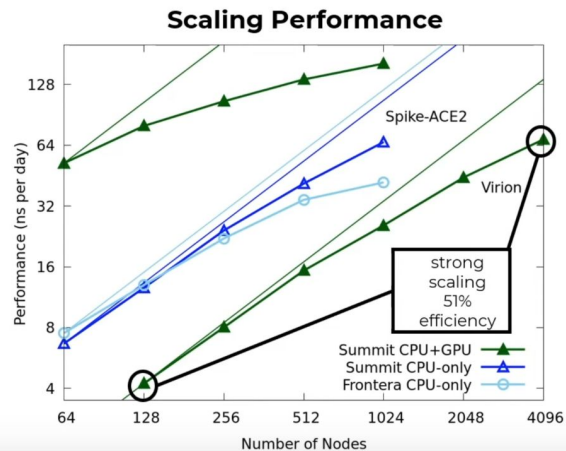
- Simulation lauréate du Gordon Bell award 2020
 - enveloppe virale et protéine de spicule du SARS-CoV-2
 - **305 millions** d'atomes
 - sur le supercalculateur **Summit** (2ème plus large supercalculateur public au monde)
 - 200 petaFLOPS (202,752 CPUs + 27,648 GPUs)
- Nécessité de simulations **multi-échelles!**
 - parallélisme insuffisant



(A) Virion, (B) Spike, (C) Glycan shield conformations

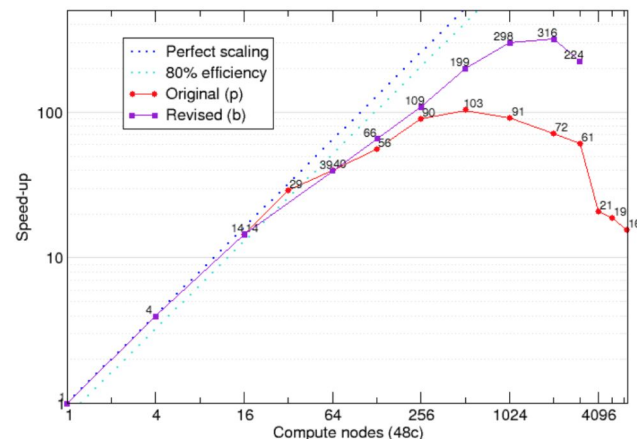
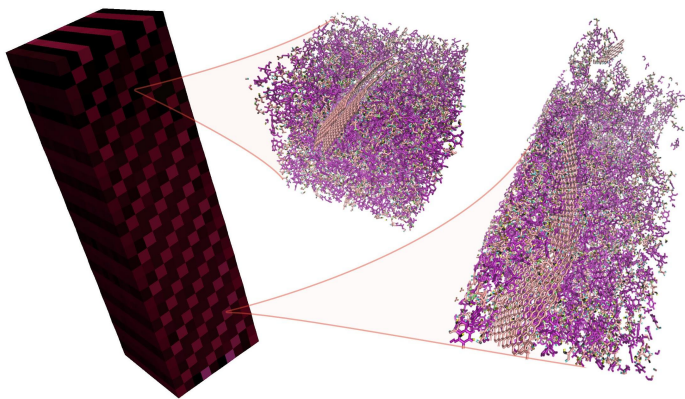


8.5M atom spike: 8k atoms/node @ 1k nodes
 305M atom virion: 12k atoms/GPU @ 4k nodes



Exploiter la séparation des échelles

- **Décomposition** et synchronisation **spatiale**
 - **1 modèle continu** macroscopique (EF) - **N modèles moléculaires** microscopiques (DM)
 - **problèmes** fortement **non-linéaires** (e.g **fissuration** d'un composite sous chargement)
- Potentiel de **parallélisation**
 - 0.1 heure CPU par simulation moléculaire - 10,000 cellules x 8 gps/cellule x 5 répliques/gp
 - **400,000 simulations microscopique indépendantes** par pas de temps macroscopique
- Test de performance (**scalability**)
 - SuperMUC-NG en entier (311,000 CPUs, plus grand ordinateur européen)
 - accélération (vs 1 noeud) maximale **x316** sur 2048 noeuds (98,304 CPUs)



[Vassaux, Richardson and Coveney. *Philos. Trans. R. Soc. London, Ser. A.* \(2019\) 377.2142](#)

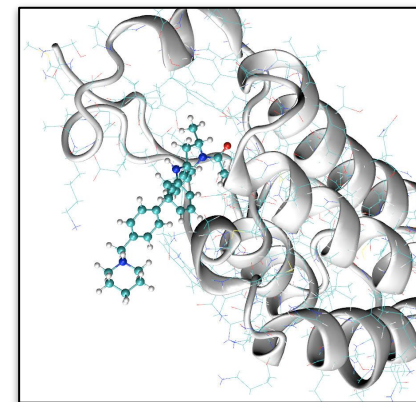
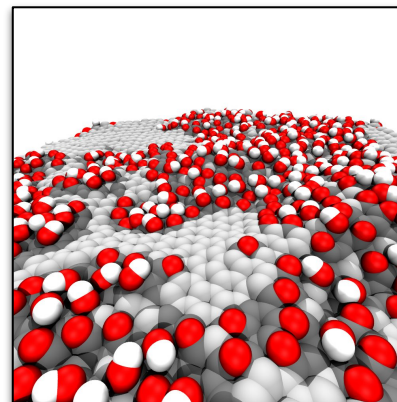
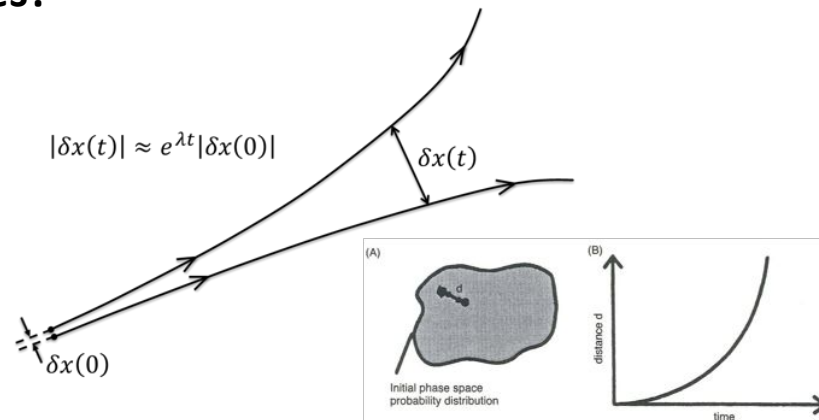
Quelle fiabilité pour les simulations moléculaires?

- Les simulations de DM
 - équations de mouvement de Newton
 - champs de force complexe
 - large nombre de particules
- Un **système chaotique** (instabilité de Lyapunov)
 - petites approximation, grandes déviations
 - sensible aux incertitudes, à quantifier!

$$\mathbf{F} = m \frac{d^2 \mathbf{r}}{dt^2}$$

$$\mathbf{F} = -\nabla V(\{\mathbf{R}\})$$

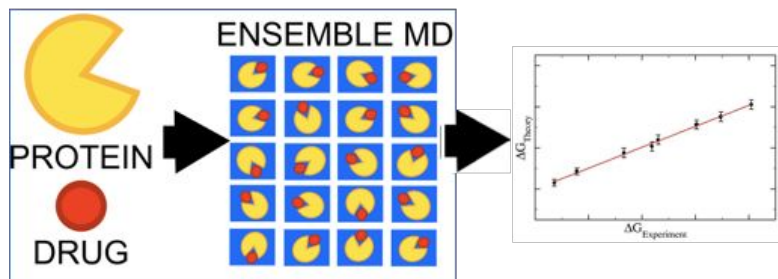
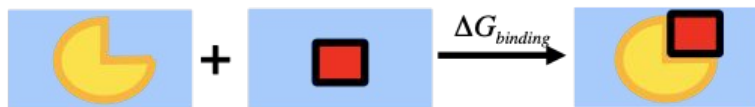
$$\begin{aligned}
 V(\{\mathbf{R}\}) = & \sum_{bonds} K_b (b - b_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 \\
 + & \sum_{dihedrals} K_\chi (1 + \cos(n\chi - \chi_0)) + \sum_{impropers} K_\psi (\psi - \psi_0)^2 \\
 + & \sum_{nonbond} \left[\left(\frac{A_{ij}}{r_{ij}} \right)^{12} - \left(\frac{B_{ij}}{r_{ij}} \right)^6 + \frac{q_i q_j}{\epsilon r_{ij}} \right]
 \end{aligned}$$



[Boghosian, Coveney and Wang. *Adv. Theor. Simul.* \(2019\) 2.12](#)
[Wan, et al. *J. Chem. Theory Comput.* \(2015\) 11, 7, 3346–3356](#)
[Suter et al... *Adv. Mat.* \(2020\) 32.36](#)

Exemple avec la prédiction d'affinité chimiques

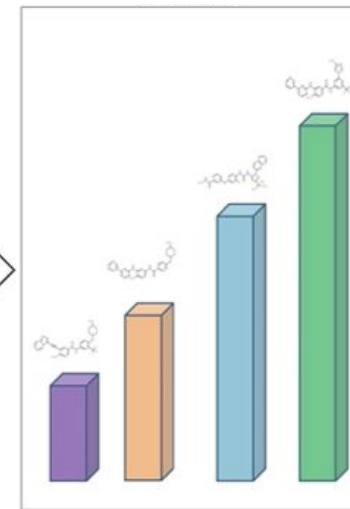
- Prédiction d'**énergie libre** (affinité chimique) de complexe **protéine-ligand**
 - application à la découverte de nouveaux médicaments (drug design)
 - cas d'étude standard en DM
 - comparaison relative
- Protocole **ESMACS** validé (Enhanced Sampling of Molecular Dynamics with Approximation of Continuum Solvent)
 - simulation d'**ensembles de répliques** du complexe



calculateur d'affinité chimique

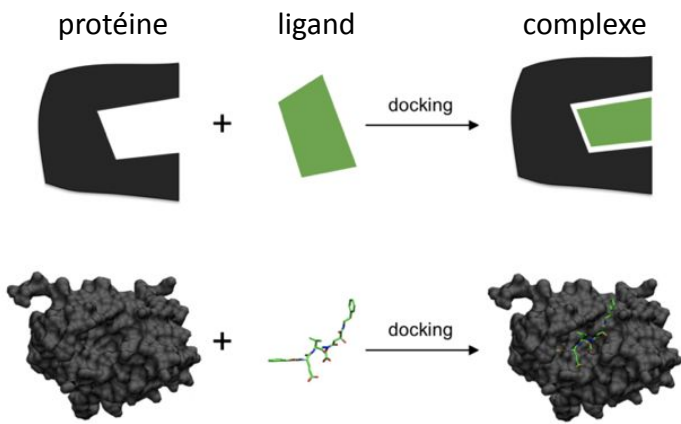


classement des candidats



Une campagne de quantification d'incertitudes à grande échelle

- Campagne de quantification d'incertitude
 - simulations à **175 paramètres**
 - paramètres *physiques* et *méthodologiques*
 - champs de force, solveur, conditions aux limites, etc...
 - restriction à **14 paramètres**
 - **distribution** aléatoire des paramètres
 - uniform, +/- 15% range
- Coût de calcul
 - 63 configurations paramétriques, 25 répliques/conf., simulations de 10ns/réplique
 - **2,000,000 CPUhs** on SuperMUC-NG



Parameter name	unit	default	min	max	group
Set Temperature (setTemperature)	ns	300.00	280.00	320.00	P
Equilibration duration (time_factor_eq)	K	1.86	1.58	2.14	P
Pressure (BerendsenPressureTarget)	bar	1.01325	0.86	1.17	P
Simulation duration (time_sim1)	ns	10.00	8.50	11.50	P
Minimum distance between protein and box edge (box_size)	Å	14.00	11.90	16.10	S
Non-bonded interactions cut-off (cutoff)	Å	12.00	10.20	13.80	S
Timestep duration (timestep)	fs	2.00	1.70	2.30	S
ShakeH bond-length error (rigidtolerance)	Å	0.00001	0.0000085	0.000015	S
Particle Mesh Ewald grid spacing (PMEGridSpacing)	Å	1.00	0.85	1.15	S
Initial temperature (initTemperature_eq1)	K	50.0	42.50	57.50	S
Temperature increment (reassingIncr_eq1)	K	1.00	0.85	1.15	S
Langevin Dynamics damping coefficient (langevinDamping)	ps ⁻¹	5.00	4.25	5.75	S
Barostat compressibility (BerendsenPressureCompressibility)	bar ⁻¹	0.000046	0.000039	0.000053	S
Barostat relaxation time (BerendsenPressureRelaxationTime)	fs	100.00	85.00	115.00	S

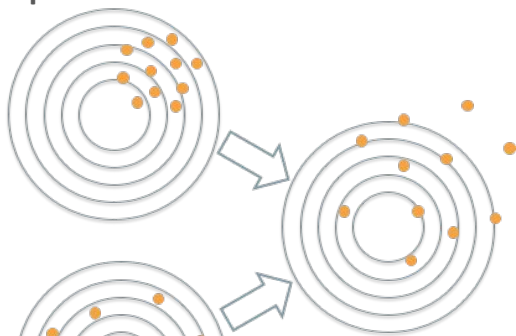
[Vassaux et al., J. Chem. Theory Comp. \(2021\) 17 \(8\), 5187–5197](#)

Le coût de la fiabilité...

- Comparaison des **types** d'incertitudes
- Distribution **non-Gaussienne** des prédictions d'énergie libre (chaoticité)
 - **incertitude aléatoire non-négligeable** vs incertitude paramétrique

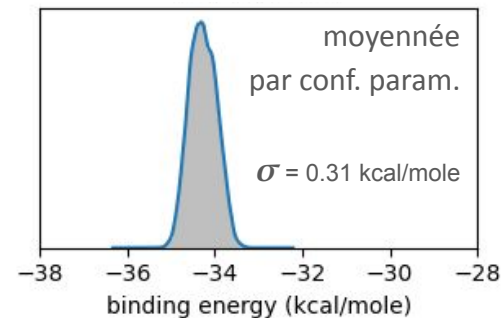
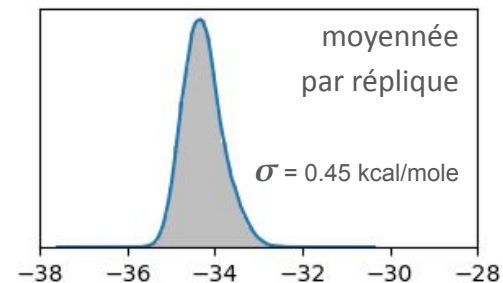
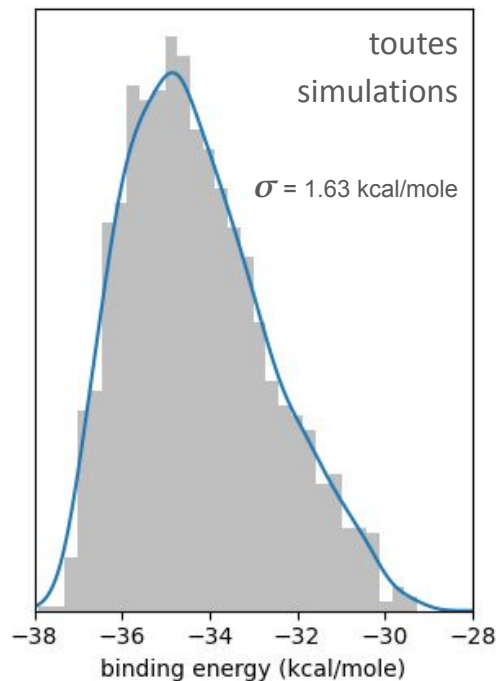
→ Plus de **30 répliques** nécessaires pour $e_{\text{alea}} < e_{\text{param}}$

paramétrique



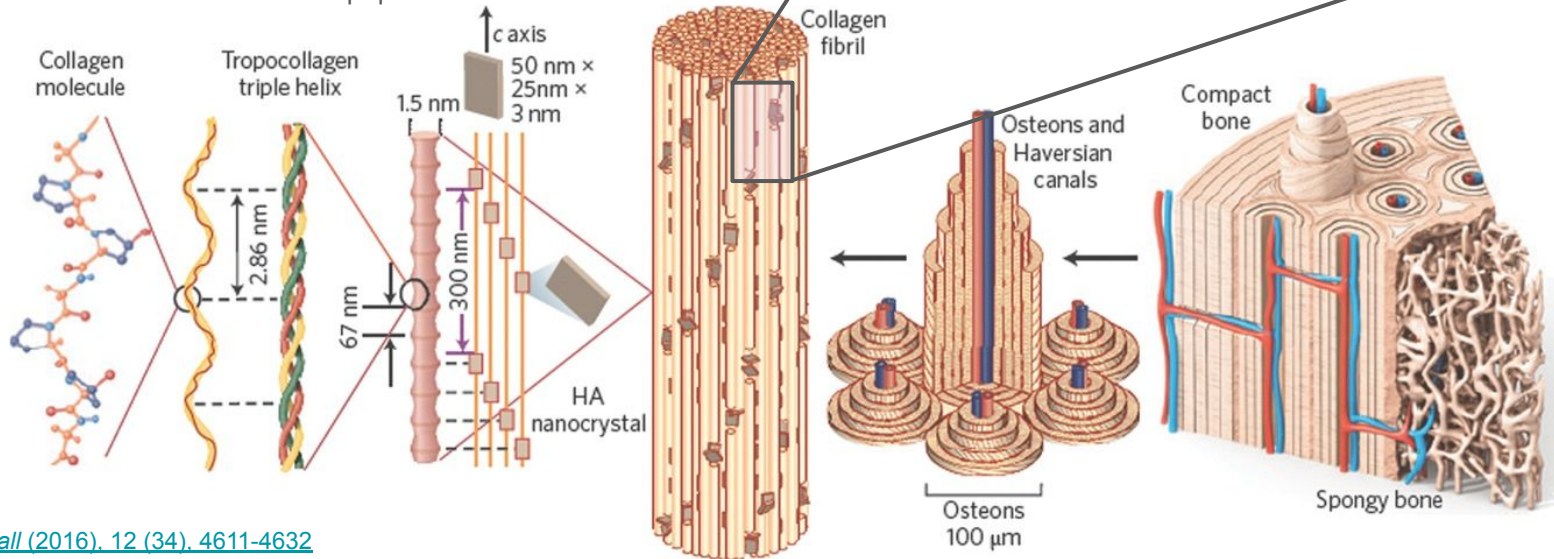
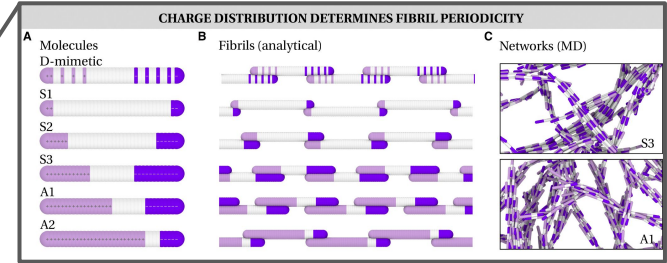
aléatoire

totale



Un microscope virtuel du tissu osseux

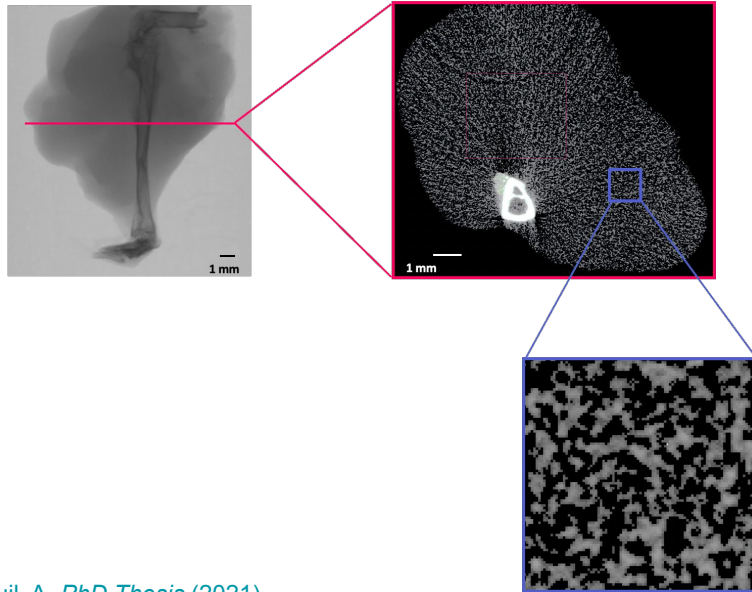
- Complémenter les expériences de laboratoire avec de nouvelles données
- Développer des outils **fiables**, vérifiés et validés
- **Prédire** en tenant compte de la **chimie**
- Intégrer de **multiples échelles**
 - chimie - nanoscopique
 - thermodynamique - microscopique
 - observable - macroscopique



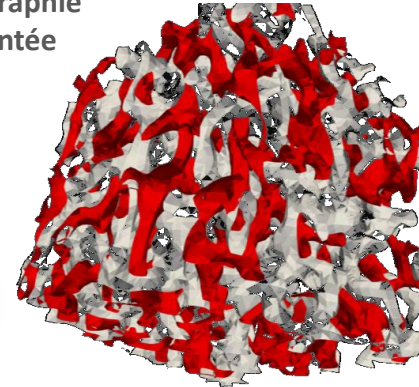
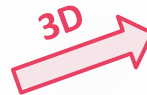
[Liu, W. et al. *Small* \(2016\), 12 \(34\), 4611-4632](#)

Comprendre la croissance et la morphologie de l'ostéosarcome

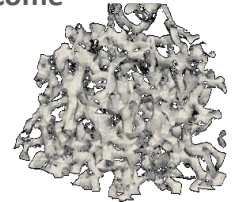
- Cancer du tissu osseux en croissance
- Une morphologie *spongieuse* bi-continue
 - tissu osseux remodelé
 - tumeur (sarcome)
- **Reconstruction** tomographique et **segmentation** (par A. Bardouil)



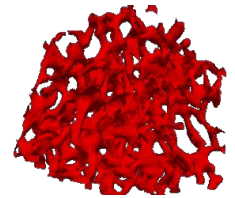
tomographie
segmentée



sarcome

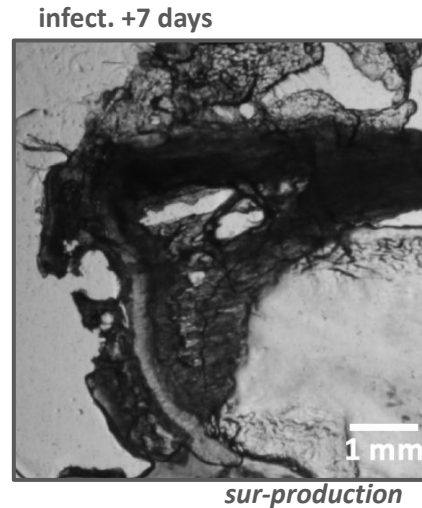
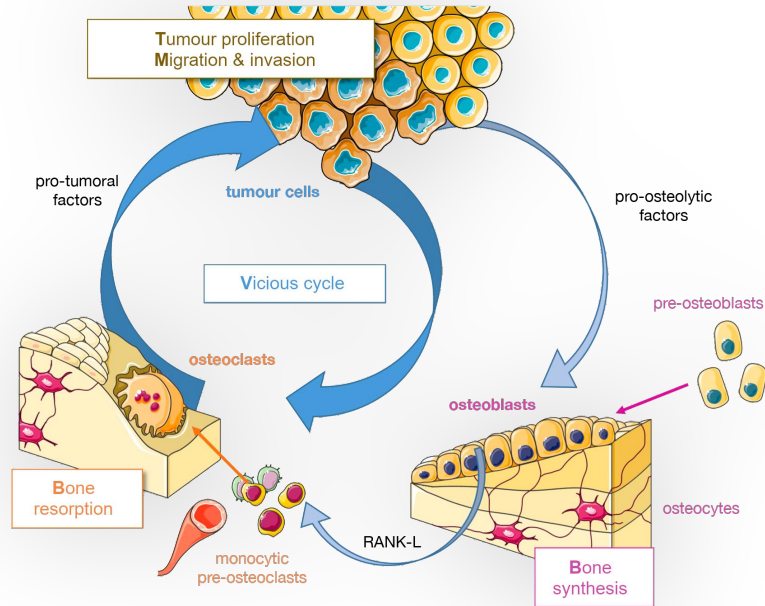


tissu osseux



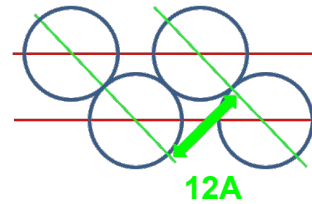
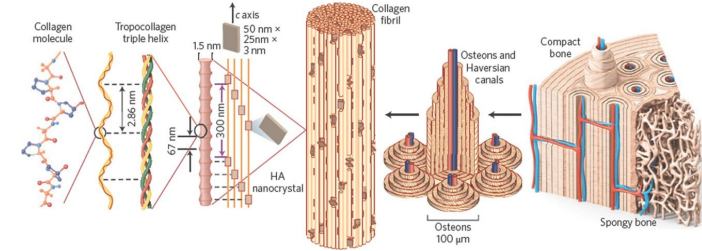
Comprendre la croissance et la morphologie de l'ostéosarcome

- Un cercle vicieux (double catalyse)
 - cellules osseuses catalysent activité (prolifération) cellules tumorales
 - cellules tumorales catalysent activité (remodelage) des cellules osseuses (ostéoblastes & ostéoclastes)

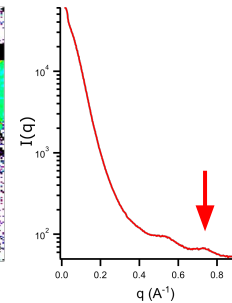
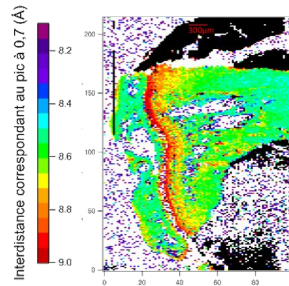
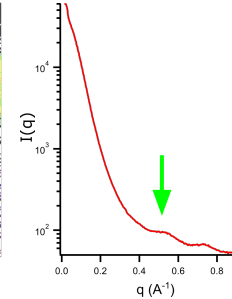
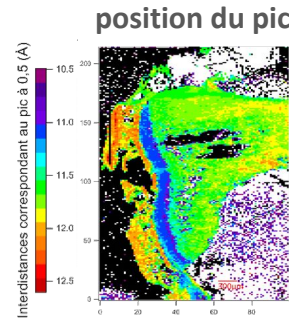
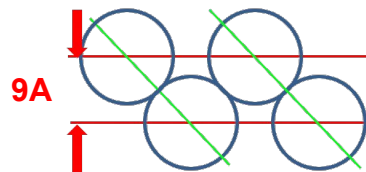


Modification du tissu osseux à l'échelle moléculaire par les cellules tumorales

- **Auto-assemblage** protéines de tropocollagène en Fibrilles de Collagène (FC)
 - FC servent de **supports mécaniques** pour le **tissu osseux** à l'échelle **moléculaire**
- **Croissance du sarcome** corrélée avec maturation structurale des FC
 - évolution de **distances inter-moléculaires**



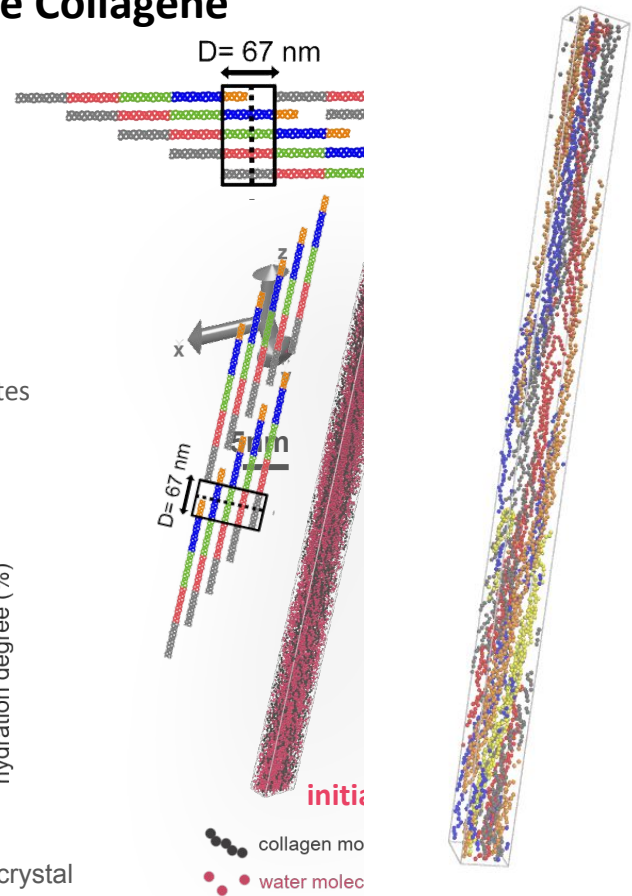
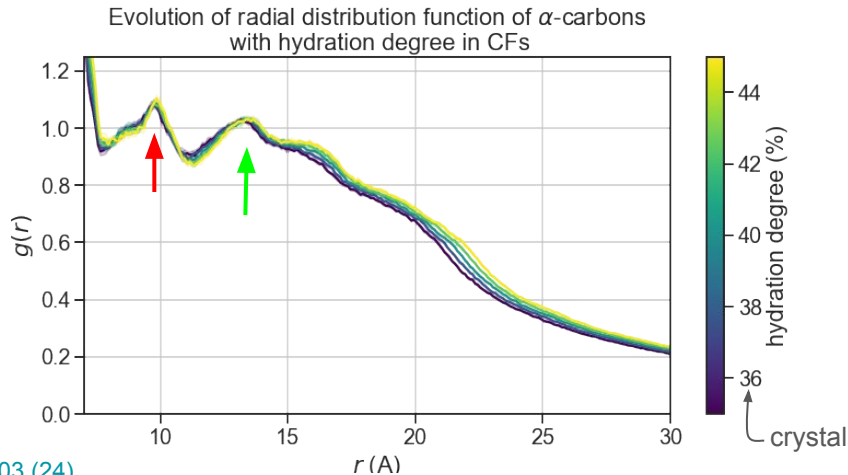
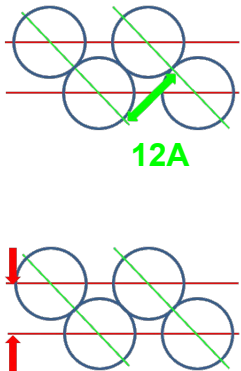
décalage distance
inter-moléculaire



Quels facteurs modifient
l'auto-assemblage de
l'empilement du collagène?

Ingénierie inverse de l'auto-assemblage des Fibrilles de Collagène

- Reconstitution virtuelle cristal de microfibrille de collagène hydraté
 - à partir de données de cristallographie aux rayons X
- Expérimentation *in silico* des effets:
 - locaux, chimie (e.g methylation)
 - globaux, thermodynamiques (e.g pH, hydratation)
- **D'abord (!)** déterminer les règles d'auto-assemblage des FCs
 - **désassemblage** du cristal microfibrille
 - hydratation/dilatation du cristal → interactions électrostatiques dominantes



Conclusions

- **Simulations (bio)moléculaires fiables** coûteuses
- **Supercalculateurs** capables simuler dynamique **virus** pendant quelques **microsecondes**
- Ingénierie inverse auto-assemblage FCs pour comprendre interactions entre tissu osseu et tumeur
- **Vers des simulations multi-échelles** (gros-grains et machine learning) pour explorer auto-assemblage des FCs à l'échelle **microseconde/micromètre**

