



UNIVERSITE D'AIX-MARSEILLE

ECOLE DOCTORALE 352

CINaM (Centre Interdisciplinaire de Nanoscience de Marseille)

Département Théorie et Simulation Numérique

Thèse présentée pour obtenir le grade universitaire de docteur

Discipline : Physique et Science de la Matière

Spécialité : Matière condensée et Nanosciences

Léa ATMANI

**Study of Organic Matter Decomposition under Geological
Conditions from Replica Exchange Molecular Dynamics
Simulations**

**Etude de la décomposition de matière organique dans des
conditions géologiques par simulations numériques de
Replica Exchange Molecular Dynamics**

Soutenue le 15/05/2017 devant le jury :

Virginie MARRY	PHENIX	Rapportrice
Marc MONTHIOUX	CEMES	Rapporteur
Edo BOEK	University of Cambridge	Examineur
Jean-Marc LEYSSALE	LCTS	Examineur
Philippe DUMAS	Aix-Marseille Université	Examineur
Roland PELLENQ	UMI CNRS/MIT	Directeur de thèse
Christophe BICHARA	CINaM	Co-Directeur de thèse

Résumé

Pétrole et gaz proviennent de la décomposition de la matière organique dans la croûte terrestre. En s'enfouissant, les résidus organiques se décomposent en un solide poreux et carboné, appelé kérogène et en un fluide composé d'hydrocarbures et de petites molécules telles que de l'eau. Le processus de formation du kérogène n'est pas totalement élucidé et une modélisation aiderait à une meilleure compréhension à la fois de sa structure et de sa composition et serait utile à l'industrie pétrolière.

Dans le présent travail, nous adoptons une approche thermodynamique ayant pour but, à l'aide de simulations numériques, de d'étudier la décomposition de précurseurs de kérogène d'un type donné –ici le type III- dans les conditions d'un réservoir géologique. La méthode dite de *Replica Exchange Molecular Dynamics* (REMD) est appliquée pour étudier la décomposition de cristaux de cellulose et de lignine. Le potentiel d'interaction ReaxFF et le code LAMMPS sont utilisés. La REMD est une façon de surmonter de larges barrières d'énergie libre, en améliorant l'échantillonnage de configurations d'une dynamique moléculaire conventionnelle à température constante, en utilisant des états générés à températures supérieures.

En fin de simulation, les systèmes ont atteint un état d'équilibre entre deux phases : une phase riche en carbone, composée d'amas de macromolécules, que nous appelons « solide » et d'une phase riche en oxygène et en hydrogène, composée de petites molécules, que nous dénommons « fluide ». L'évolution des parties solides de nos systèmes coïncide avec celle d'échantillons naturels de kérogènes de type III.

Mots clés : Kérogène, REMD, Simulations, Gaz, Pétrole, Gaz de schistes, Matière organique, LAMMPS, ReaxFF, Cellulose, Lignine, van Krevelen, Pyrolyse, Hydrocarbures, Simulations Numériques, Carbone, Potentiel réactif, Carbone amorphe, Thermodynamique, Géologie, Roche mère, Maturation, Dynamique Moléculaire

Abstract

In deep underground, organic residues decompose into a carbonaceous porous solid, called kerogen and a fluid usually composed of hydrocarbons and other small molecules such as water, carbon monoxide. The formation process of the kerogen is well understood, but related, accurate modeling is missing. Modeling geological maturation could widen the understanding of both structure and composition of kerogen, and could be useful to oil and gas industry.

In this work we adopt a purely thermodynamic approach in which we aim, through molecular simulations, at determining the thermodynamic equilibrium corresponding to the decomposition of given organic precursors of a specific type of kerogen –namely type III- under reservoir conditions. Starting from cellulose and lignin crystal structures we use replica exchange molecular dynamics (REMD) simulations, using the reactive force field ReaxFF and the open-source code LAMMPS. The REMD method is a way of overcoming large free energy barriers, by enhancing the configurational sampling of a conventional constant temperature MD using states from higher temperatures.

At the end of the simulations, we have reached for both systems, a stage where they can clearly be cast into two phases: a carbon-rich phase made of large molecular clusters that we call here the "solid" phase, and a oxygen and hydrogen rich phase made of small molecules that we call "fluid" phase.

The evolution of solid parts for both systems and the natural evolution of a type III kerogen clearly match. Evolution of our systems follows the one of natural samples, as well as the one of a type III kerogen submitted to an experimental confined pyrolysis.

Key words: Kerogen, REMD, Replica exchange molecular dynamics, Gas, Shale Gas, Oil, Organic matter, LAMMPS, ReaxFF, Cellulose, Lignin, van Krevelen, Pyrolysis, Hydrocarbons, Numerical simulations, Carbon, Reactive potential, Amorphous carbon, Thermodynamics, Petroleum geology, Source rocks, Maturation, Molecular dynamics

Remerciements / Acknowledgements

Le présent travail de thèse a été mené entre le laboratoire CInaM (Marseille, France) et le MIT (Boston, USA). Je tiens à remercier dans leur ensemble ces deux institutions qui m'ont permis, de par les humains qui les composent et leurs installations, de poursuivre ma recherche dans des conditions agréables et adéquates. Je remercie en particulier les directeurs successifs du CInaM, **Claude Henry** et **Frédéric Fages** de m'avoir accueillie dans leur laboratoire, ainsi que le *hub Energy Initiative* du MIT et l'UMI MSe², avec **Franz Ulm** et **Roland Pellenq** en directeurs de l'unité mixte. Merci à mes directeurs de thèse, **Christophe Bichara** et **Roland Pellenq**, de m'avoir offert cette opportunité et suivie durant ces 3 années aussi intenses qu'enrichissantes. Je remercie aussi les membres du jury, qui ont accepté de rapporter, pour V. Marry et M. Monthieux et d'examiner, pour P. Dumas, J-M. Leyssale et E. Boek, le présent travail.

Je remercie l'ensemble des membres de l'UMI pour l'année que j'y ai passée. Merci à **Jean-Marc** pour son aide précieuse en physique et son apport en cannelés, merci aussi à **Henri** pour ses connaissances élargies du domaine des hydrocarbures, un grand merci à **Farhang** et **Mahdi** pour les longues discussions « humaines » et merci à **Amaël** pour son administration du cluster - que j'ai beaucoup monopolisé - et pour tout ce qui se rapporte à un corgi à roulettes ! Merci aussi à **Leïla** qui permettait à toute la troupe de se poser et à **Eloïse**, toujours souriante. Merci à **Jérémie**, **Hadrien**, **Katerina** pour les *Trump's* parties, les survols du Massachussetts, les repas grecs, les matchs de hockey et de basketball. Thank you to **Katharina** for every morning coffee we had. Merci à **Katerina** et **Alice** d'avoir formé un bureau réconfortant en chocolat et autres délicatesses. Thank you to **Mirna**, **Muhannad** (Mumun!), **Jake**, and **Amer** for every peaceful moment I enjoyed, lying on the grass, listening music and eating hummus, thank you for the best cheesecake I have ever eaten!

My stay in Boston wouldn't have been the same without **Monika** (my little bird, thank you sooo much for being you, being such a human person, with such a big heart and consideration for everyone), **Leilani** who took me with her since my very first stay in Boston (merci d'être toujours présente), **Jenny** and **Annalise** (I loved our roommate dinners, brunches, and breakfast), it was very enjoyable to get home and find you after a long day ("Au lit!" "Seriously Mr. Zen"), thank you for answering all my questions about English (I know, Beyoncé has a great influence), and **Marija**, thank you for your always good mood, for wine, tarts, coffee, and everything else. Thank you to **Loni**, **Azadeh**, **MaryAnn**, and **Deirdre** for having me being more confident. Thank you for every Tuesday night I spent with you, dancing and laughing.

Merci à l'équipe TSN du CINaM dans son ensemble pour son support et ses conseils. Merci à **Zapzap** pour son administration de Pancho (dans l'attente des éclairs à la violette), à **Yann** pour son regard posé de la vie, à **Azzam**, à **Céline** pour ses encouragements lors de la rédaction de ce manuscrit et à **Claudio** pour les *eigenvalues*. Toujours au CINaM, merci à **Véro** pour toutes les discussions et encouragements, à **Valérie**, **Dominique** et **Elodie** pour leur aide et à **Nathalie** pour son sourire du matin. Merci à **Fred** pour son aide informatique. Merci à **Ling**, **Philippe**, **Lisa**, **Matthieu**, **Suzanne** et **Vinh** pour l'avant-thèse (mais évidemment, aussi le pendant). Un gros merci à tous les copains-copines, sans qui les 3 ans n'auraient pas été : **Maïssa**, **Maïmouna**, **Charline**, **Yousra**, **Nassima**, **Cheng**, **Astrid**, **Angèle** (marmelaaade), **Raj** (mention spéciale pour ta bonne humeur, tes talents de cuisinier, merci pour les livres, les films, les musiques, merci pour HP), **Benjamin**, **Paul**, **Guillem** le rouge, **Michel**, **Ali**, **Maxime**, **Romain**, **Vasile**, **Vica** (legoushka). Merci à **Lionel** et **Evelyne** aussi. Un énorme merci à **Laurence** pour son soutien et ses paroles, sensées et pourtant nécessaires. Evidemment, un merci spécial pour « les filles », merci pour toutes ces soirées à rigoler, manger et jouer (Pardon Maïmouna :p). Merci pour votre bonne humeur et votre soutien sans faille. Merci pour les photos, les pizzas sodebo (Pardon Youyou). Merci pour les robes, les maquillages (Pardon Maïssa). Merci.

Merci à la **DreamTeam**. Merci pour les licornettes, les retrouvailles ci et là, les poupées bouteilles, les Jean-Georges, les biscuits, les chocolats, les cafés, les baklavas, les crumbles, les cartes postales, les skypes, les fous-rires. Merci pour les dancefloors squattés à 3 jusqu'au lever du Soleil... Merci à **Yahube** pour sa venue surprise à ma soutenance depuis l'autre côté de l'Atlantique, un vrai pirate. Merci pour votre présence, à Marseille ou à Boston, je suis heureuse de clôturer les missions **#arayrospatial**. En attendant le prochain dimanche aux Goudes, merci.

Merci aux **Poulettes**, toujours présentes, depuis si longtemps. Le socle amical. Merci pour votre soutien, les discussions, les coups de fils, les week-ends, les soirées. Merci pour toutes ces années et toutes celles qui arrivent. Merci aux **Riom**, **Bessou** et **Arnaud-Nunes**. Merci aux **Mojito-Girls**, pour votre présence, les restos, les histoires, les rires. Une pensée émue pour la Dédé-mobile qui n'aura pas tenue jusqu'à la fin de la thèse. Merci à ma **Justine** et ma **belette Céline**, toujours régulières dans leurs prises de nouvelles. Merci à **Ambre** d'être là. Merci à mon **club Pantagruel**, parce qu'un verre de vin dans une main et un livre dans l'autre, c'est presque incontournable en fin de thèse. Merci à **Elise** pour tous ses messages lumineux. Merci à toutes mes **copines de Clairval**, toujours présentes et rigolotes. Merci à mon **Burger**, à **Fred**, à **Franck**, à **Latour**, à **Célia**, à **Gégé** et à **Nicolas** pour leur bonne humeur, leur dynamisme et leur joie de vivre contagieuse. Merci d'être présents et de me faire tant rire. Merci à mes **Paupiettes**, sans qui mon parquet ne brillerait pas (c'était facile !) mais surtout sans qui je n'aurais pas tant d'abdos (et de snapchats) à force de me faire rire.

Un grand merci aux **Roule-Gerbail**, de tous bords, pour leur présence, leur support, leur soutien, leurs idées. Merci de si bien m'accueillir. Merci à **Mini-Boule** d'être si bon à PES. Un grand merci à toute ma famille, ci et là, du Nord au Sud, de l'Est à l'Ouest, un arbre sans racine grandit mal. Merci à **Jeannot** pour les dessins, les orties, les balades, merci à **Christiane** pour les fous-rires infinis, les cubis, les histoires si bien racontées. Merci à **Simone** pour sa grande force, une grande pensée pour **Roger** et sa lettre. Merci à tous les **Mazard** élargis. Merci à **Nordine** et **Anissa**, lien indéfectible. Merci à tous les **Atmani-Attemani** élargis. Merci à mon **Papi**, ses beaux yeux, ses sourires, ses bricolages divers, au gré de mes envies, de mes lubies. Merci à ma **Mamie**, qui, avec quelques ingrédients, créait le plus doux des mets, reliait les gens. Les pizzas, les glaces à la fraise, les soupes, les marrons. Merci à mon **Pépé**, qui, la tête haute, a su s'imposer et poser ma base, a su différencier les personnalités. Merci à ma **Mémé**, qui a su trouver un autre langage que le français pour communiquer. Un grand merci à **Léo** et **Jacqueline**, lumières douces pour les cœurs, qui me guident dans leur poésie et leur détermination à aider leur prochain. C'est une grande chance de vous avoir dans ma vie.

Un énorme merci à mon cercle premier. Mon **Papa**, ma **Maman**, qui prouvent que le mélange et la différence sont des richesses et que l'amour est le seul dogme qui mérite d'être suivi. Merci de m'avoir offert une double culture. Merci de faire de nous, non pas ce que vous voudriez que nous soyons, mais ce que nous devons être. Merci d'être mes arcs. Merci d'être mes lumières et mon refuge. Merci pour les rires, les jeux, les débats, les disputes, les livres, les câlins, les repas. Merci pour tout. Merci d'être vous et de continuer à chercher l'amélioration. Merci aussi, de m'avoir appris la confiance, en les autres et en moi-même. Merci de m'avoir inculqué que la condition de naissance, à la fois en termes de condition sociale mais aussi de sexe, ne détermine en aucun cas la valeur de la personne ni l'ampleur de sa réussite. Merci enfin de m'avoir offert une fratrie. Merci à mes frères, **Gabriel** et **Ryane**, d'être mes frères, tout simplement. Merci d'être mes piliers, de me rattraper quand je tombe. Merci pour les innombrables fous-rires, merci de partager ce que personne d'autre ne pourra jamais partager avec moi. Merci pour les « Yololoooo », les « aaah », les XXL. Merci pour tous les jeux, pas merci pour les +4. Je vous aime.

Enfin, quatorze mercis à **Guilhem**. Merci d'être toi et de me laisser être moi. Merci de toujours me tirer vers le haut et de ne jamais me laisser glisser vers le bas. Merci de me rendre tous les jours un peu de plus forte, de me soutenir dans tout ce que j'entreprends. Merci pour tous les voyages, les cartes, les souvenirs, les nems. Merci pour tous les fous-rires, les courses dans l'appart, les bigotelles. Merci pour les débats, les discussions, les grandes questions. Merci pour les négociations, les boules de Noël, le chat, le chien (bientôt parti). Merci d'être un si bon ami, un si bon compagnon de vie. Merci de croire en moi, encore mieux, de me pousser à croire en moi. Je t'aime.

Résumé substantiel en français

Le pétrole, que l'on retrouve, en profondeur, dans des gisements, provient de la décomposition de matière organique. Cette matière est la partie d'organismes morts, qui a échappé à une oxydation en étant enfouie dans le sol et mélangée avec d'autres sédiments pendant des millions d'années. Seulement 1% de ces organismes morts sédimente. Cette matière organique dispersée, qui a sédimenté, est principalement composée de carbone, d'hydrogène et d'oxygène. En quantité moindre, de l'azote et du soufre peuvent être trouvés.

Les résidus organiques se décomposent en un mélange entre un solide carboné et poreux, appelé kérogène et un fluide généralement composé d'hydrocarbures et de petites molécules telles que de l'eau et du monoxyde de carbone. Ce processus de décomposition se déroule en trois principales étapes : la diagenèse, la catagenèse et la métagenèse. Au cours de cette dégradation, la matière organique se déleste en premier lieu principalement de son oxygène, puis perd majoritairement son hydrogène. Durant la deuxième étape, des cycles aromatiques sont formés et la métagenèse correspond principalement à une étape de réarrangement structural¹.

Le kérogène est un état intermédiaire de la matière organique, entre le résidu organique et les hydrocarbures. En 1980, Durand² définissait le kérogène comme « *toute matière organique insoluble* » dans les solvants de l'époque. Pendant son enfouissement, le kérogène est affecté par un processus de craquage. Durant ce phénomène, qui se déroule sur des millions d'années, le kérogène génère du pétrole, du gaz et quelques autres molécules. Les produits formés dépendent de l'état de transformation du kérogène, que l'on appelle son degré de maturité et des précurseurs dont il provient.

Trois principaux types de kérogènes ont été définis³, en fonction de leur source et donc de leur composition initiale.

- Le kérogène de type I provient principalement de dépôts lacustres. Au début du processus de pyrolyse, la matière organique possède un ratio atomique H/C de 1.5 et une faible proportion d'oxygène (ratio O/C inférieur à 0.1). Les échantillons de ce type sont généralement hautement aliphatiques.
- Les kérogènes de type II proviennent généralement d'environnements marins et leur ratios atomiques initiaux sont approximativement de 1.3 pour H/C et de 0.15 pour O/C.
- Le kérogène de type III possède des origines terrestres et provient souvent de plantes supérieures. Une pyrolyse prolongée de ce type mène à une importante formation de charbon. Son ratio H/C, à moins de 0.8, est bien inférieur à celui des deux types précédents tandis que son ratio O/C, d'environ 0.25, est plus important.

Le processus de décomposition de la matière organique lors de son enfouissement dans le sol, est considéré comme une pyrolyse⁴ s'opérant sur des millions d'années. La pyrolyse est une méthode de dégradation thermique non sélective. Des essais pour simuler en laboratoire ce processus naturel ont été menés, utilisant différentes méthodes –pyrolyse ouverte⁵, fermée⁶ et confinée⁷, cette dernière s'avérant être la méthode la plus aboutie. Quelques simulations numériques ont aussi été tentées. L'évolution du kérogène est généralement visualisée à l'aide d'un diagramme de Van Krevelen⁸ qui fournit l'évolution du ratio atomique H/C en fonction du ratio O/C.

La structure du kérogène était et est toujours étudiée en suivant les avancées technologiques en termes d'amélioration de la puissance de calcul, des méthodes, des potentiels et du développement de codes. La formation du kérogène est aussi un sujet intéressant et de nombreuses équipes de recherches ont essayé de la simuler numériquement⁹⁻¹² mais avec une importante question en suspens : comment accélérer cette réaction –qui prend à la nature des millions d'années- sans en changer l'équilibre. La majorité des études menées sur le sujet l'ont été à bien plus hautes températures que celles opérant en sous-sol. Une température plus haute induit un déplacement de l'équilibre et les produits obtenus sont, de fait, différents de ceux présents dans les échantillons naturels. Des méthodes numériques pour accélérer des processus d'événements rares existent et une tentative d'application d'une de ces méthodes à la décomposition de matière organique a été menée¹³. Cette étude a été menée en *ab initio* et fournit des résultats intéressants mais malheureusement limités en taille de système et en temps de calcul.

En recherche, deux façons d'étudier le comportement de systèmes existent : les expériences physiques et les expériences numériques. La première façon utilise des échantillons réels et des outils physiques tandis que la seconde se base sur des modèles numériques et utilise des ordinateurs. Dans ce travail, nous utilisons la seconde façon de faire. Des techniques telles que les éléments finis sont utilisées pour étudier des comportements à l'échelle mésoscopique. Nous sommes intéressés ici à un niveau plus petit, l'échelle atomique. A l'échelle nanoscopique, deux principales méthodes numériques se dessinent : le Monte-Carlo et la dynamique moléculaire.

Les méthodes de Monte-Carlo sont basées sur des essais aléatoires de modification du système considéré. Aujourd'hui, la majorité des méthodes Monte-Carlo utilise l'algorithme de Metropolis¹⁵. Son principe est de tenter un changement aléatoire du système, qui est accepté ou rejeté suivant un critère défini, sur l'énergie par exemple. Ces méthodes peuvent être appliquées à des systèmes de taille importante et le calcul est relativement rapide. Le principe de changement aléatoire implique qu'aucune information sur la dynamique du système n'est accessible.

La Dynamique Moléculaire quant à elle fournit des informations de dynamique et se base sur les équations du mouvement de Newton. Chaque particule du système est considérée comme un point ayant une masse et les équations de Newton sont intégrées¹⁶ afin d'obtenir les trajectoires des particules au cours de la simulation. La Dynamique Moléculaire peut s'appliquer sur des systèmes de taille importante –jusqu'à plusieurs centaines de milliers d'atomes.

Pour reproduire le comportement des systèmes étudiés, les méthodes Monte-Carlo et de Dynamique Moléculaire nécessitent l'utilisation de potentiels d'interaction. Ces potentiels ont pour but de décrire les interactions entre atomes. En dérivant ce potentiel, la force qui leur est appliquée est obtenue. Il est composé de termes répulsif et attractif qui s'additionnent pour former les interactions. L'un des potentiels les plus simples et le plus connu est celui de Lennard-Jones¹⁷, mais de nombreux autres existent. Dans ce travail, nous utilisons celui publié par Van Duin et ses collègues en 2001 et appelé ReaxFF¹⁸. Il s'agit d'un potentiel réactif, qui décrit les ruptures et formations de liaisons, en décrivant les intermédiaires réactionnels. Il est basé sur un formalisme dit d'ordre de liaisons, qui prend en compte non seulement les premiers voisins mais aussi les suivants.

Concernant le code de simulation, nous utilisons le code LAMMPS¹⁹ qui est libre d'accès et qui intègre à la fois le potentiel ReaxFF et la méthode REMD. Cette méthode REMD²⁰⁻²² –pour *Replica Exchange Molecular Dynamics*- est une méthode compilant des critères de la Dynamique Moléculaire et de Monte-Carlo, utilisée pour surmonter de larges barrières d'énergie libre, en améliorant l'échantillonnage de configurations d'une dynamique moléculaire conventionnelle à température constante, en utilisant des états à températures supérieures. Le même système initial est lancé à différentes températures, des échanges entre répliques de ce système sont tentés à intervalles réguliers et sont acceptés ou rejetés en suivant un critère basé sur des différences énergétiques. C'est cette méthode que nous appliquons dans ce travail, afin de simuler numériquement la dégradation de la matière organique s'effectuant sur des millions d'années.

Cette technique nécessite d'importantes ressources de calcul. Plusieurs centaines de milliers d'heures de calculs ne peuvent être lancées sans avoir, au préalable, validé les méthodes utilisées. Une phase de validation, à la fois du potentiel et de la méthode REMD, est essentielle.

Des simulations en dynamique moléculaire et en Monte-Carlo ont été menées sur des systèmes carbonés amorphes. Les systèmes ont été chauffés à haute température - 5000K et 3273K- en utilisant deux potentiels différents. Pour la dynamique moléculaire, le potentiel ReaxFF¹⁸ a été utilisé, tandis que c'est un code de liaisons fortes²³ – disponible au laboratoire- qui a été utilisé pour les calculs Monte-Carlo. Les structures obtenues par ces deux méthodes se sont avérées très différentes, à la fois dans

l'organisation atomique et dans l'hybridation des atomes de carbone. Les configurations obtenues avec le potentiel de liaisons fortes semblaient plus physiques. Nous avons donc décidé de tester plus en détail des implémentations de ReaxFF –en particulier les implémentations dites de Mattsson²⁴ et de Pitman²⁵.

Pour ce test du potentiel ReaxFF, des énergies de dissociation de liaisons à 0K ont été calculées pour plusieurs hybridations d'atomes de carbones et pour des molécules simples telles que H₂O, CO, CO₂ et C₂H₆. Les courbes obtenues présentaient des formes en V –quasi non dérivables en leur minimum, alors qu'un potentiel, dont dérive la forme appliquée aux atomes doit l'être en tout point- et des distances d'équilibres trop importantes. Les implémentations testées n'étaient, de fait, pas satisfaisantes. En décembre 2014, un nouvel ensemble de paramètres de ReaxFF a été publié, offrant une description plus exacte des liaisons carbone-carbone. Après vérification, nous avons décidé d'utiliser cette nouvelle version de ReaxFF, appelée ReaxFF_{C-2013}²⁶ qui s'avérait être la mieux adaptée à notre étude.

Avant de pouvoir être appliquée à un système de plus de 4000 atomes dans notre étude- la méthode REMD a d'abord été testée sur un système de cellulose de petite taille -420 atomes- afin d'en définir les paramètres nécessaires. Tout d'abord, la gamme de températures et leur distribution ont été définies et testées. Pour que la REMD soit efficace, il faut que la valeur haute de la gamme de température corresponde à une température à laquelle les plus importantes barrières énergétiques du système sont passées et avec laquelle aucune réplique du système ne puisse se retrouver coincée dans un minimum énergétique. La température la plus basse de la gamme est celle à laquelle nous nous intéressons dans notre étude.

Pour le petit système de cellulose 33 répliques, allant de 300 à 3500K ont été utilisées. Chacune des distributions énergétiques des différentes températures doivent posséder une zone de recouvrement avec, au moins, leurs premières voisines –distributions énergétiques aux températures juste au-dessus et juste en dessous. 400ps de REMD ont été calculées et, en regardant la configuration finale à la plus basse température -300K- la formation de deux phases distinctes apparaît. La première est un fluide composé de petites molécules telles que H₂O, CO et CO₂ tandis que la seconde est une structure riche en carbone, se rapprochant d'un solide. Ces premiers résultats incitent à prolonger l'investigation sur un système plus gros.

Des simulations courtes de REMD sur le système de cellulose de 4200 atomes ont été menées afin de définir le nombre de répliques nécessaires et leur distribution. Nous avons conclu que l'utilisation de 96 répliques, distribuées en suivant une suite géométrique débutant à 423K et de raison 1.0225 était un choix optimal. Il permet à chacune des distributions énergétiques de recouvrir une partie des distributions voisines –en température- et au système d'évoluer dans un délai de temps raisonnable tout en évitant la surconsommation de ressources de calcul.

La cellulose est le principal constituant des plantes²⁷ et se trouve être la matière organique la plus abondante sur Terre. Il s'agit d'un polymère naturel, composé de chaînes linéaires de molécules de glucoses, liées entre elles par des ponts oxygènes. Sa formule est $(C_6H_{10}O_5)_n$, où n est le degré de polymérisation. Comme mentionné précédemment, le kérogène de type III provient principalement des plantes supérieures. La cellulose étant leur principal composant, il semble donc naturel de la choisir comme modèle de précurseur de kérogène de type III. Nous avons construit un système de cellulose I β de 4200 atomes. Ses ratios atomiques H/C et O/C initiaux étaient respectivement de 1,67 et 0,83.

Dans un premier temps, une première simulation REMD a été lancée sur le système de cellulose de 4200 atomes, en alternant avec des relaxations NPT ayant pour objectif de conserver le système dans une gamme de pression géologique –entre 25 et 125 MPa. Cette première étape de simulation représente 750ps de REMD. L'énergie du système décroît tout au long de la simulation, ce qui indique un processus exothermique vers une configuration du système plus stable à la température d'intérêt -423K. Au cours du processus de décomposition, deux phases sont apparues : un fluide, majoritairement composé d'eau et de monoxyde de carbone, riche en hydrogène et en oxygène et une phase solide, riche en atomes de carbone. Ces deux phases se forment en deux principales étapes : la formation de la phase fluide dans un premier temps, accompagnant la fragmentation du cristal de cellulose, puis la coalescence de la partie solide. La croissance de la phase solide débute réellement une fois que le nombre de petites molécules telles que l'eau commence à se stabiliser. Après 750ps de REMD l'énergie du système converge et l'on en déduit que le système atteint un état d'équilibre. La simulation est alors arrêtée et une deuxième étape débute : une simulation REMD appliquée à la phase solide uniquement.

La partie solide obtenue lors de la première étape de REMD sur le système entier de cellulose a d'abord été relaxée en NPT. Elle a ensuite été soumise à une alternance de simulations REMD et de relaxations NPT, le tout pour une durée de 1800ps de REMD. L'énergie du système décroît durant toute la simulation, jusqu'à converger. Pendant cette simulation REMD, bien que l'énergie décroisse, la composition chimique du système n'est pas drastiquement modifiée puisque, à la fin des 1800ps de REMD, seulement 10 molécules d'eau ont été formées, 6 de monoxyde de carbone et 5 de dihydrogène, alors qu'en comparaison le solide comporte 1033 atomes. Cette décroissance de l'énergie est probablement due à un réarrangement structural, puisque le nombre moyen de cycles aromatiques par amas augmente. Néanmoins, il n'augmente pas suffisamment pour que cette étape de la simulation puisse correspondre à la métagénèse de la matière organique. Cette seconde simulation de REMD sur la cellulose semble plutôt correspondre à la fin de la catagénèse. Pour que la métagénèse soit simulée, un apport énergétique supplémentaire ou un temps de simulation plus long semblent nécessaires.

L'évolution de la phase solide obtenue par décomposition de la cellulose est tracée sur un diagramme de Van Krevelen et comparée avec des résultats expérimentaux. Il apparaît que nos simulations reproduisent précisément l'évolution naturelle de kérogènes de type III. Les points correspondant à notre système se retrouvent compris dans le nuage de points représentant des échantillons naturels de kérogène de type III.

Nous avons alors décidé d'étudier la lignine, comme complément de la cellulose en tant que composant des plantes terrestres. Le terme de lignine réfère à une famille de polymères aromatiques, construits à partir de trois principaux blocs, appelés monolignols. Les structures de lignine possibles sont légion. Banoub et Delmas²⁹ affirment que les structures linéaires sont les plus probables. Nous avons donc cherché une structure linéaire, relativement facile à construire numériquement et avons choisi celle de Crestini *et al.*³⁰ Nous avons construit un cristal de lignine de 3978 atomes, avec pour ratios atomiques H/C et O/C initiaux respectifs de 1.11 et de 0.34. Cette structure a ensuite été relaxée puis soumise à une simulation REMD.

La simulation REMD de la lignine a duré 1300ps, et des relaxations NPT -à 423K et 50MPa- ont été appliquées toutes les 200ps. Ces relaxations sont utilisées pour conserver le système dans une gamme de pression géologique de [25-125MPa]. Les paramètres de la simulation REMD sont les mêmes que ceux utilisés pour la cellulose -96 températures allant de 423 à 3500K, échanges de températures tentés toutes les 10fs. L'énergie du système diminue -ce qui indique une évolution vers une configuration plus stable à 423K- jusqu'à converger. L'évolution de la pression présente une tendance comparable à celle de la cellulose, avec une augmentation importante en début de simulation, correspondant à la formation considérable d'eau. Une décomposition du système en deux phases -fluide et solide- apparaît, comme dans le cas de la cellulose.

La phase fluide est riche en hydrogène et en oxygène et majoritairement composée d'eau, de monoxyde carbone et de dihydrogène, tandis que la phase solide est très riche en carbone et est principalement hybridée sp^2 . La prédominance, dès le début de la simulation, de cette hybridation provient en grande partie des monolignols initiaux, qui sont composés de cycles aromatiques. Contrairement à la cellulose, la lignine n'est pas fragmentée en petites molécules composées de chaînes carbonées mais de cycles aromatiques qui s'assemblent pour former la partie solide. Bien que les aspects qualitatifs correspondent pour les deux systèmes de cellulose et de lignine, certaines différences existent et en particulier pour ce qui est de la composition quantitative du fluide.

Comme attendu³¹, la pyrolyse de la lignine forme plus de méthane que celle de la cellulose. Les structures des parties solides coïncident, avec une hybridation évoluant vers une prédominance de sp^2 . En revanche, les compressibilités isothermes à la fin des premières étapes de REMD - 750ps pour la cellulose et 1300ps pour la lignine- diffèrent

l'une de l'autre. Le solide obtenu à partir de la cellulose, après 750ps de REMD, possède une compressibilité isotherme de $1.64 \cdot 10^{-10} \text{ Pa}^{-1}$, tandis que le solide obtenu depuis la lignine possède une compressibilité isotherme de $5.71 \cdot 10^{-9} \text{ Pa}^{-1}$. Collell *et al.*³² ont trouvé une valeur de $2.7 \cdot 10^{-10} \text{ Pa}^{-1}$ pour un kérogène de type II –défini comme “au milieu de la fenêtre de génération de pétrole”. Les deux phases solides se trouvent, de fait, à des degrés de maturité différents.

Malgré ces différences les chemins de décomposition des solides obtenus à partir de la cellulose et de la lignine sont conformes à l'évolution des échantillons naturels de kérogènes de type III. La méthode REMD reproduit précisément la dégradation thermique de matière organique en conditions géologiques. Nous sommes donc capables de reproduire un phénomène se produisant sur une échelle de temps géologique en utilisant des simulations numériques qui, sans l'apport de la technique d'échange de répliques, sont en général limitées à reproduire des temps physiques de l'ordre de la nanoseconde.

Bien que la méthode reproduise précisément la dégradation thermique de la matière organique en conditions géologiques, elle peut encore être améliorée. Du fait de la petite taille du système, certaines caractéristiques n'ont statistiquement que peu de sens. Certaines informations, telles que celles qui relèvent de l'étude de la porosité du système ne sont pas accessibles. Pour cela, il faudrait avoir accès à de très importantes ressources de calcul. Au-delà de la taille du système, deux grands chemins d'améliorations peuvent être suivis. Le premier porte sur la méthodologie elle-même. Naturellement, le fluide formé lors de la dégradation thermique de la matière organique s'échappe –au moins en partie- laissant au solide la liberté de se réarranger. Dans cette étude, le fluide est conservé en totalité lors de la première REMD, puis éliminé, là aussi en totalité, avant de débiter la seconde REMD. Cela ne correspond pas à ce qui se passe en profondeur. Tout ou partie du fluide pourrait être retiré tout au long de la simulation. Par ailleurs, bien que le potentiel ReaxFF_{C-2013} semble être le meilleur outil pour reproduire cette pyrolyse, certains points peuvent être améliorés, tels que les réactions composant le reformage à la vapeur. Une nouvelle version de ReaxFF est actuellement en développement et les premiers tests que nous avons pu effectuer sont prometteurs.

Le deuxième chemin d'amélioration serait d'élargir le précurseur utilisé, en dégradant des composés. En effet les composants du bois, par exemple, peuvent être combinés afin d'en obtenir des structures plus représentatives et permettre une meilleure simulation de la dégradation thermique de la matière organique. Certains composants peuvent avoir des impacts sur la dégradation des autres. Par exemple, la lignine est connue pour être un antioxydant, du fait de son aromaticité et cette caractéristique pourrait désactiver certains radicaux instables formés lors de la dégradation d'hémicellulose³³. Au-delà des composants des précurseurs, d'autres participants à la dégradation pourraient être pris en compte : les composés inorganiques. Ils pourraient encourager la formation de cycles aromatiques et donc

impliquer une augmentation de la formation de charbon³⁴. Ajouter des composés inorganiques au système pourrait aider à comprendre leur impact et améliorer la simulation.

La méthode REMD appliquée à la simulation de la dégradation thermique de la matière organique peut donc encore être améliorée de différentes façons. Elle permet néanmoins déjà de reproduire convenablement ce processus géologique et d'apporter des informations uniques sur le détail, à l'échelle atomique, des processus mis en jeu lors de la décomposition et des structures associées.

1. Vandenbroucke, M. & Largeau, C. Kerogen origin, evolution and structure. *Org. Geochem.* **38**, 719–833 (2007).
2. Durand, B. *Kerogen: Insoluble organic matter from sedimentary rocks*. (Editions technip, 1980).
3. Durand, B. & Espitalie, J. Evolution de la matière organique au cours de l'enfouissement des sédiments. *Compte rendus l'Académie des Sci.* **276**, 2253–2256 (1973).
4. Trager, Earl, A. Kerogen and its relation to the origin of oil. (1924).
5. Rouxhet, P. G. & Robin, P. L. P. R. Infrared study of the evolution of kerogenes of different origins during catagenesis and pyrolysis. *Fuel* **57**, 533–540 (1978).
6. Lewan, M. D., Winters, J. C. & McDonald, J. H. Generation of oil-like pyrolyzates from organic-rich chales. *Scien* **203**, 1–4 (1978).
7. Monthieux, M., Landais, P. & Monin, J.-C. Comparison between natural and artificial maturation series of humic coals from the Mahakam delta, Indonesia. *Org. Geochem.* **8**, 275–292 (1985).
8. van Krevelen, D. W. *Coal--typology, chemistry, physics, constitution.* **3**, (Elsevier Science & Technology, 1961).
9. Salmon, E., van Duin, A. C. T., Lorant, F., Marquaire, P.-M. & Goddard, W. a. Early maturation processes in coal. Part 2: Reactive dynamics simulations using the ReaxFF reactive force field on Morwell Brown coal structures. *Org. Geochem.* **40**, 1195–1209 (2009).
10. Zou, C., Raman, S. & Van Duin, A. C. T. Large-scale reactive molecular dynamics simulation and kinetic modeling of high-temperature pyrolysis of the gloeocapsomorphaprisca microfossils. *J. Phys. Chem. B* **118**, 6302–6315 (2014).
11. Liu, X. X. *et al.* Initial pyrolysis mechanism of oil shale kerogen with reactive molecular dynamics simulation. *Energy and Fuels* **29**, 2987–2997 (2015).
12. Zhang, Y., Wang, X., Li, Q., Yang, R. & Li, C. A ReaxFF Molecular Dynamics Study of the Pyrolysis Mechanism of Oleic-type Triglycerides. *Energy and Fuels* **29**, 5056–5068 (2015).
13. Agarwal, V., Dauenhauer, P. J., Huber, G. W. & Auerbach, S. M. Ab initio dynamics of cellulose pyrolysis: Nascent decomposition pathways at 327 and 600 °C. *J. Am. Chem. Soc.* **134**, 14958–14972 (2012).
14. Metropolis, N. & Ulam, S. The Monte Carlo Method. *J. Am. Stat. Assoc.* **44**, 335–341 (1949).
15. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1092 (1953).
16. Verlet, L. Computer 'Experiments' on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **159**, 98–103 (1967).
17. Lennard-Jones, J. E. Cohesion. *Phys. Soc.* **43**, 461–482 (1931).
18. Van Duin, A. C. T., Dasgupta, S., Lorant, F. & Goddard, W. A. ReaxFF: A reactive force field for hydrocarbons. *J. Phys. Chem. A* **105**, 9396–9409 (2001).
19. Munday, L. B. & Chung, P. W. Large-Scale Atomic / Molecular Massively Parallel Simulator (LAMMPS) Simulations of the Molecular Crystal α RDX. (2013).
20. Swendsen, R. & Wang, J. Replica Monte Carlo simulation of spin glasses. *Phys. Rev. Lett.* **57**, 2607–2609 (1986).
21. Geyer, C. J. Markov Chain Monte Carlo Maximum Likelihood. *Comput. Sci. Stat. Proc. 23rd Symp. Interface* 156–163 (1991).
22. Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein

- folding. *Chem. Phys. Lett.* **314**, 141–151 (1999).
23. Amara, H., Roussel, J. M., Bichara, C., Gaspard, J. P. & Ducastelle, F. Tight-binding potential for atomistic simulations of carbon interacting with transition metals: Application to the Ni-C system. *Phys. Rev. B - Condens. Matter Mater. Phys.* **79**, 1–17 (2009).
 24. Mattsson, T. R. *et al.* First-principles and classical molecular dynamics simulation of shocked polymers. *Phys. Rev. B - Condens. Matter Mater. Phys.* **81**, 1–9 (2010).
 25. Pitman, M. C. & Duin, A. C. T. Van. Dynamics of Confined Reactive Water in Smectite Clay – Zeolite Composites. *J. Am. Chem. Soc.* **134**, 3042–3053 (2012).
 26. Goverapet Srinivasan, S., Van Duin, A. C. T. & Ganesh, P. Development of a ReaxFF potential for carbon condensed phases and its application to the thermal fragmentation of a large fullerene. *J. Phys. Chem. A* **119**, 571–580 (2015).
 27. Chambon, F. Transformation de la cellulose par catalyse hétérogène. (2013).
 28. Gomes, T. C. F. & Skaf, M. S. Cellulose-builder: A toolkit for building crystalline structures of cellulose. *J. Comput. Chem.* **33**, 1338–1346 (2012).
 29. Banoub, J. H. & Delmas, M. Structural elucidation of the wheat straw lignin polymer by atmospheric pressure chemical ionization tandem mass spectrometry and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J. Mass Spectrom.* **38**, 900–903 (2003).
 30. Crestini, C., Melone, F., Sette, M. & Saladino, R. Milled wood lignin: A linear oligomer. *Biomacromolecules* **12**, 3928–3935 (2011).
 31. Caballero, J. A., Font, R., Marcilla, A. & García, A. N. Flash pyrolysis of Klason lignin in a Pyroprobe 1000. *J. Anal. Appl. Pyrolysis* **27**, 221–244 (1993).
 32. Collell, J. *et al.* Molecular Simulation of Bulk Organic Matter in Type II Shales in the Middle of the Oil Formation Window. *Energy & Fuels* **28**, 7457–7466 (2014).
 33. Rousset, P., Turner, I., Donnot, A. & Perré, P. Choix d ' un modèle de pyrolyse ménagée du bois à l ' échelle de la microparticule en vue de la modélisation macroscopique. *INRA EDP Sci.* **63**, 213–229 (2006).
 34. Collard, F. X. & Blin, J. A review on pyrolysis of biomass constituents: Mechanisms and composition of the products obtained from the conversion of cellulose, hemicelluloses and lignin. *Renew. Sustain. Energy Rev.* **38**, 594–608 (2014).

Table of contents

Résumé	3
Abstract	5
Remerciements / Acknowledgments	7
Résumé substantiel en français	11
Introduction	27
Chapter 1: Context	35
Organic matter decomposition and kerogen	37
Definitions	37
Organic matter	37
Kerogen	39
Organic matter decomposition	42
Stages of the decomposition	42
Conditions of the decomposition	44
Evolution of the structure	46
Hydrocarbons formation	48
Characterization of Kerogen	50
Previous studies	54
Experimental studies	55
Open and closed pyrolysis	55
Confined pyrolysis	56
Numerical studies	58
Structure modelling	58
Modelling the formation of kerogen	65
Summary	69
Chapter 2: Method	75
Interaction potential	78
Principle	79
Reactive potential : ReaxFF	82
Definition	82
Applications	85
Atomistic simulations	86
Thermodynamic Ensembles	87
Monte-Carlo	88
Classical Molecular Dynamics	91
Parallelization	97
Replica exchange molecular dynamics (REMD)	97
Monte-Carlo Parallel Tempering	98
Molecular Dynamics Parallel Tempering: REMD	98
Principle of Parallel Tempering	99
Choice of replicas	104
Applications	105
LAMMPS	106

Definition	106
The use of ReaxFF and the REMD method with LAMMPS	107
Reaxff	107
REMD	108
Summary	108
Chapter 3: Validation	113
Tests on ReaxFF	115
Small cellulose crystal (420 atoms)	125
Parameters	125
Results	127
Parameters for 4200 atoms (cellulose) system	128
Summary	130
Chapter 4: Results	133
Type III Kerogen precursor: Cellulose	135
Choice of cellulose as a type III kerogen precursor	135
Building of cellulose system	136
Confined Pyrolysis	137
Parameters	138
First step: REMD on the whole system	138
Second step: Solid-part relaxation by REMD	155
Comparison of our kerogen evolution on a Van Krevelen diagram	159
Type III Kerogen precursor: Lignin	163
Choice of lignin as a type III kerogen precursor	163
Building a lignin model	165
REMD Confined pyrolysis of lignin	166
Fluid phase formation	168
Solid phase formation	170
Van Krevelen diagram	175
Comparison between cellulose and lignin	177
Summary	183
Conclusion	189

Introduction

Words such as oil, gas, conventional, or unconventional fossil resources, refer to energy resources which all share a common chemical composition. They all are complex mixtures of hydrocarbon rich molecular compounds. Therefore, their main components are carbon and hydrogen. Life on Earth is the only way to transform primary sources of carbon and hydrogen into complex systems made of these two compounds, but what is the link between living organisms and fossil energy resources?

When living organisms die, their rests are consumed by micro-organisms. The major part returns into the Earth surface cycle but a tiny part –less than 1%- is buried and starts to sediment. While this tiny proportion of organic matter is buried, chemical transformations occur, that will transform it into a fossil material called kerogen. This matter is a carbonaceous porous solid, and depending on the primitive organic matter, and on its thermal history and environment, three main types of kerogen can be found. Type I kerogen mainly results from lacustrine deposition, type II from marine environment, and type III from terrestrial origins, such as plants.

Due to tectonic activity, sediments are gathered into sedimentary basins. During their formation, and because of their mass, it is buried deeper and deeper. Those sedimentary intervals when they are rich in kerogen, are called parent rocks. Their burying, up to few kilometers, goes along with an increase of the temperature, due to a geothermal gradient. This temperature increase will lead to organic matter decomposition in three main stages. First, dehydration occurs with the formation of carbon dioxide. This first stage is called diagenesis –at temperature around 300K- and is followed by catagenesis – at temperature above 350K- during which, hydrocarbons are mainly formed. The last step is called metagenesis, and corresponds to hydrocarbons and gas formation –at temperature above 400K. From early diagenesis to late metagenesis, kerogen is submitted to a thermal maturation. Kerogen degree of maturity and its origin –therefore its initial composition- have a significant impact on hydrocarbon formation. This formation is considered as a far-from-equilibrium process powered by thermal energy provided to the matter by its environment.

This hydrocarbon formation, inside the solid called kerogen, will induce a pressure increase which will lead to hydrocarbon release. They will be expelled towards porous rocks. If these products do not meet impermeable formations, they will seep on the surface. On the contrary, if they meet an impermeable formation, hydrocarbons will be stacked and will be hoarded into deposits. Depending on the kerogen type, its degree of maturity and the type of parent rock, this deposit will be mostly made of oil, gas or of a mixture of both. A third possibility for these hydrocarbons is to remain trapped in parent rock porosity. This is what is called unconventional deposit also known as shale-gas or shale-oil.

Despite of all progresses concerning alternative energy resources and biofuels, we still rely on fossil hydrocarbons. Those resources need to be extracted in order to be used in our installations, such as cars. In order to be able to extract them, field study is needed. Indeed, before starting drilling, operators need to have a good idea of where and how deep they have to drill. Geological data, both for physical and chemical aspects, are used to build complex three dimensions models of basins. Until now, basin modelling is based on a set of parallel chemical reactions, which reproduce diverse reactions occurring during kerogen maturation. This set is gradually improved with the help of laboratory-scale experimental simulations. However, natural organic matter thermal degradation takes place over millions of years and is thus highly difficult to reproduce in laboratory. It can be considered as pyrolysis, namely the chemical decomposition of organic matter under the only effect of heat, without any oxidant effect. There are different experimental ways to simulate. Thus, process experiments are limited to conditions which can be applied to samples. Although chemical bonds analysis, such as NMR, can be used, models of kerogen structure are hard to draw from these information only. Kerogen is a mix of diverse macromolecules, and not a single structure. Therefore, a unique configuration cannot be found but an average one could be drawn, assembling together data from numerous analysis.

Hydrocarbons are not only products of the kerogen degradation. They also use kerogen as a reservoir. They are trapped in voids left by released hydrocarbons. All of that plays a role in formation of kerogen structure, and on its porosity development. The way these pores are connected together is not easy to know, despite the obvious role of this porosity on hydrocarbon expulsion and possible recovery.

Despite its importance, formation of kerogen process remains not well understood and a matter of debate. Modeling its geological maturation could widen the understanding of both structure and composition of kerogen, and could be useful to oil and gas industry. Indeed, treating the problem as a set of parallel chemical reactions, although useful, is not perfect and still needs to be improved¹. A global simulation of kerogen maturation, at an atomic scale and under geological temperature and pressure could replace this chemical route and help to have a better understanding of the process and its implications. Some attempts of such a numerical simulation have been performed²⁻⁶ but at high temperatures compared to the one operating in the ground – above 1000K. Therefore, thermodynamic chemical equilibrium was significantly affected, and did not correspond to the one observed in natural samples. An accelerated sampling based approach⁷ has also been applied, but was limited to the early decomposition process.

Besides techniques like finite elements to study mesoscale behaviors, methods such as Monte-Carlo and Molecular dynamics (MD) are used to examine atomic scale processes. The latest rely on the use of interaction potentials. In this work, interatomic potentials are used to simulate the complex evolution of the system at the atomic scale.

Besides this state, information such as dynamics for instance, can be obtained from a numerical simulation. Time order of magnitude of these simulations is usually of few nanoseconds. We feel here the great discrepancy between this order of magnitude and millions of years of the geological process we described above, taking place in millions of years. **We so do wonder how to simulate geological process, taking place on the scale of millions of years, by the use of atomistic numerical simulation?**

In this work, we intend to answer this question by the use of a method, called Replica Exchange Molecular Dynamics (REMD)⁸⁻¹⁰, that is designed to accelerate long times processes by making it possible to overcome large free energy barriers. It works by enhancing configurational sampling of a conventional constant temperature MD, using states of higher temperatures. Temperature swaps between different temperatures replicas are attempted in order to help lower temperatures to overcome energy barriers. We therefore apply this method to type III kerogen precursors –namely cellulose and lignin–, with the goal of simulating their thermal degradation under geologic conditions.

This manuscript is structured in four chapters. The first one develops the context and gives definitions. It supplies key information on kerogen types, composition, and degradation: It highlights known processes of kerogen decomposition and how its structure evolves with maturation. It also gives a review of both experimental and numerical attempts to simulate this geological process of organic matter decomposition. In the second chapter, numerical simulation methods are described. We focus on Monte-Carlo and Molecular Dynamics techniques that are useful to understand how REMD works. We obviously explain in details the REMD method, but also how interaction potential works, with a focus on the one we use in this study: ReaxFF¹¹. We end this chapter by a presentation of the code we use, called LAMMPS. A third chapter is devoted to techniques validations, an important task to do before to run calculation time consuming jobs. We finally present our results in the fourth chapter. We present the degradation of precursors we decided to study –cellulose and lignin. We submitted them to REMD runs, at geological temperature and pressure, and then analyzed products resulting from their decompositions.

1. Van Damme, H. *Source rocks (work in progress)*. (2016).
2. Salmon, E., van Duin, A. C. T., Lorant, F., Marquaire, P.-M. & Goddard, W. a. Early maturation processes in coal. Part 2: Reactive dynamics simulations using the ReaxFF reactive force field on Morwell Brown coal structures. *Org. Geochem.* **40**, 1195–1209 (2009).
3. Zou, C., Raman, S. & Van Duin, A. C. T. Large-scale reactive molecular dynamics simulation and kinetic modeling of high-temperature pyrolysis of the gloeocapsomorphaprisca microfossils. *J. Phys. Chem. B* **118**, 6302–6315 (2014).
4. Liu, X. X. *et al.* Initial pyrolysis mechanism of oil shale kerogen with reactive molecular dynamics simulation. *Energy and Fuels* **29**, 2987–2997 (2015).
5. Zhang, Y., Wang, X., Li, Q., Yang, R. & Li, C. A ReaxFF Molecular Dynamics Study of the Pyrolysis Mechanism of Oleic-type Triglycerides. *Energy and Fuels* **29**, 5056–5068 (2015).
6. Zheng, M. *et al.* Initial reaction mechanisms of cellulose pyrolysis revealed by ReaxFF molecular dynamics. *Fuel* **177**, 130–141 (2016).
7. Agarwal, V., Dauenhauer, P. J., Huber, G. W. & Auerbach, S. M. Ab initio dynamics of cellulose pyrolysis: Nascent decomposition pathways at 327 and 600 °C. *J. Am. Chem. Soc.* **134**, 14958–14972 (2012).
8. Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**, 141–151 (1999).
9. Hansmann, U. H. E. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* **281**, 140–150 (1997).
10. Earl, D. J. & Deem, M. W. Parallel Tempering: Theory, Applications, and New Perspectives. (2005). doi:10.1039/B509983H
11. Van Duin, A. C. T., Dasgupta, S., Lorant, F. & Goddard, W. A. ReaxFF: A reactive force field for hydrocarbons. *J. Phys. Chem. A* **105**, 9396–9409 (2001).

Chapter 1: Context

Organic matter decomposition and kerogen

The main goal of this research work is to understand the mechanisms leading to the formation of kerogen, a carbon-rich porous solid structure that derives from the decomposition of organic matter, that is gradually buried and transformed over geological time scales. We first give some definitions of organic matter and kerogen. We then explain how the first decomposes into the second, what are the main stages of it and the conditions. We explain the link with petroleum formation and we end by giving some tools for kerogen characterization used in industry and laboratory.

Definitions

When discussing about the formation of oil and gas, the simplest way to explain is to say that oil and gas come from the decomposition of organic matter. If we want to understand in details the process of decomposition, we must first define all the components and terms involved such as the organic matter and the kerogen.

Organic matter

When something is said organic, it usually means it has something to do with living matter. It means that it comes from living tissues or from the transformation undergone by products extracted from living bodies. Indeed, **organic matter**, the matter that we study here, comes from living organisms. What is called organic was, at first, only thought of as resulting from the living matter. Some of those components had then been synthetized and now, organic chemistry is part of the chemistry of materials including carbon materials.

Origin

Carbon is an essential constituent of our world. It plays an important role in the life cycle and has its own cycle called the *organic carbon cycle*. This cycle is an overview of how carbon flows from a reservoir to another and is composed by fast and slow exchanges. There are four main reservoirs: atmosphere, biosphere, hydrosphere, and lithosphere.

The short cycle includes all carbon exchanges which take place in less than one hundred years – what we call “fast”. It is mainly represented by photosynthesis and breathing due to living organisms. These exchanges occur between atmosphere and biosphere or hydrosphere.

The long cycle includes processes taking place in few centuries to millions of years – what we call slow exchanges. Flows are going from the three reservoirs mentioned for the short cycle –atmosphere, biosphere, and hydrosphere – to the lithosphere and are of smaller amplitude than for the short cycle but the reservoir here is much larger.

A schematic representation of the organic carbon cycle is given by Figure 1.

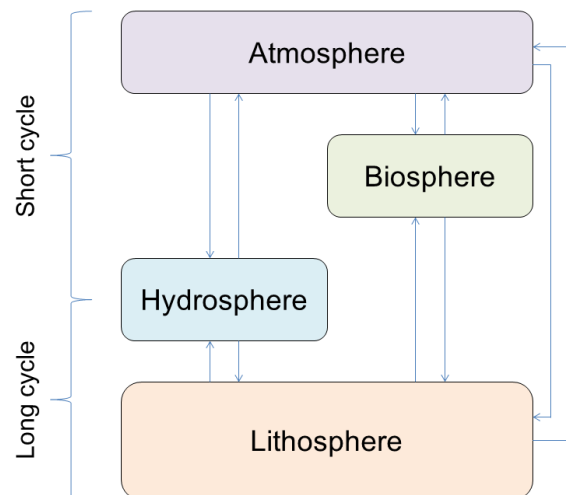


Figure 1. Organic carbon cycle

The main source of organic carbon in the lithosphere is biomass. A biomass is the living matter of a given habitat. For example, on the surface of the earth, higher plants are a dominant biomass. When biomass dies, the majority of the matter is used by small organisms which make energy out of it and the rest -usually less than 1%- is incorporated to sediments and starts to be buried in the ground.

The ground is poor in oxygen and during the first hundreds of meters, remaining biomass is transformed by anaerobic bacteria. They extract a part of the oxygen and nitrogen content. With time, this organic matter goes deeper in the ground. Bacteria stop playing an important role in its transformation and temperature increases as well as the pressure. With this increase, a thermal decomposition starts. This decomposition will transform a percentage of the organic matter into oil and gas. The less reactive compounds can sometimes keep their original structure. They thus have a clear link with the organic matter they came from and therefore, can be used as geochemical markers.

Decomposition of organic matter has a beginning and an end. If the end is not important for us –for the reason that it has no impact on our current time- we do care about the beginning. The decomposition history of the organic matter has a direct impact on what we can extract nowadays.

Composition

Organic matter comes from living organisms, thus it is composed of the same major constituents. Indeed, cellulose, lignin, planktons, and algae, for example, are part of the sedimentary organic matter source.

Organic matter is, as we explained above, mainly composed of carbon. As key elements in the living, oxygen and hydrogen are also present and non-negligible. In less important proportions we can also find nitrogen and sulfur. Cellulose, for instance, is made of 28.6% of carbon, 47.6% of hydrogen and 23.8% of oxygen. All other elements are less than 2%.

Total organic matter can be divided in two parts. The division is made regarding its solubility. If the dispersed matter is soluble in organic solvents, such as carbon bisulfide, it is called bitumen. It is a kind of hydrocarbon, used, for example, for road construction. The second portion, the insoluble one, is called kerogen and is the subject of our research.

Kerogen

In the deep underground, organic residues decompose into a carbonaceous porous solid, called kerogen and a fluid usually composed of hydrocarbons and other small molecules as water, carbon dioxide...

Word **kerogen** was first used by Crum Brown in a personal communication¹ to describe a Scottish oil shale organic matter producing a waxy oil. Indeed, kerogen comes from **keros** which means wax in Greek. At first, this word was only used in case of rocks containing organic matter and of economic importance. It then has been extended² to all organic matter potentially source of oil, contained in rocks. Nowadays, the widest used definition for kerogen is the extended one of Durand³ who defined it as “***the fraction of sedimentary organic matter which is insoluble in the usual organic solvents***” and this, applied to “*all sedimentary organic matter*” in contrast to previous definitions. This is not really a clear and strict definition, for example, the usual organic solvents are not the same today than the ones used in the 80’.

In Figure 2 we can see the total quantity of kerogen compared to the final resources as oil and gas. We clearly see that all kerogen does not decompose into fossil fuel resources. Indeed, the amount of kerogen, that is the insoluble sedimentary organic matter, is far more important than the fossil fuel resources ones - 10^{16} t versus 1.4×10^{12} t.

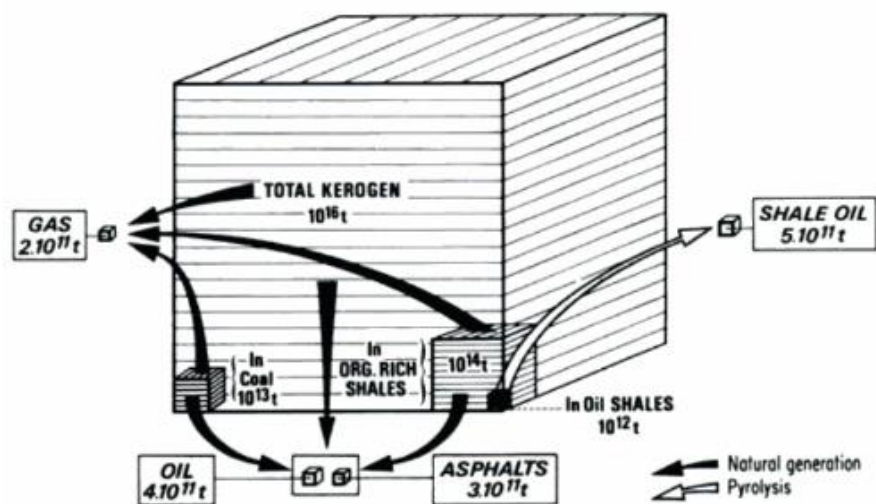


Figure 2. Quantity of kerogen compared to the ultimate resources of fossil fuels (from Durand, 1980)³

There is clearly a continuity between living biomass and the fossil organic matter defined as kerogen since the first decomposes into the second under mild temperatures. During the burying temperature increases and it is known that “*petroleum generation is a consequence of the drive of kerogen to adjust to its [environment]*”⁴. This evolution is called **maturation**. Maturation is defined as the degree of decomposition of the organic matter. A kerogen will be defined as immature at the beginning of decomposition and more and more mature while decomposing. Main steps of its decomposition are explained in the following section.

Even if the main steps of the maturation are the same no matter where the organic matter comes from, there are mainly **three types of kerogen** –defined by Durand and Espitalie⁵ in 1973- depending on the composition and also on the precursor of the kerogen. There exists a fourth type with no potential to produce hydrocarbons.

Type I kerogen principally derives from lacustrine depositions. It comes from lakes and lagoon primitive algae and bacteria. Usually, at the beginning of the degradation process, the matter has a significant H/C atomic ratio, higher than 1.5 and a low oxygen content -O/C atomic ratio lower than 0.1. The samples are generally highly aliphatic¹.

Type II kerogens usually come from the marine environment, mainly from phytoplankton. They are poorer in hydrogen than lipids of type I but richer in oxygen. Indeed, the H/C and O/C atomic ratios are around 1.3 and 0.15 respectively. A sub-type,

¹ Two classes of hydrocarbons are defined in organic chemistry: aliphatic and aromatic. They are both composed of a hydrogen and carbon structure. In the case of aliphatic compounds, structure can be straight-chain, branched chain or cyclic, saturated or unsaturated. At the opposite, aromatic structures are more stable –this system is conjugated and its pi bonds electrons can be delocalized around the carbon ring- and often refer to benzene systems.

called II-S, exists and is a type II kerogen enriched with sulfur, usually because of sulfate-reducing bacteria during its decomposition.

Type III kerogen has terrestrial origins, from higher plantsⁱⁱ, and gives rise to coal chars under further decomposition. The H/C atomic ratio, lower than 0.8, is lower than the ones of the two former types, and the O/C ratio, around 0.25, is higher. Cellulose, the main component of higher plants, with H/C and O/C ratios of 1.67 and 0.83, respectively, is an obvious precursor of type III kerogen.

Kerogen types have differences in their initial composition since they come from different kinds of organic matter sources and they follow different decomposition paths. Due to these differences, they have different characteristics concerning their formation. Indeed, if they have similar qualitative evolution, quantitative part is different. For example, type I produces bigger quantities of effluents than type II which itself produces more than type III with also a more important proportion of hydrocarbons. Type III gas over oil ratio is bigger type II's and type I's⁴.

One of the main tools to visualize the evolution of kerogen is the Van Krevelen diagram –see Figure 3- which is an evolution route map and gives the evolution of H/C and O/C atomic ratios with decomposition.

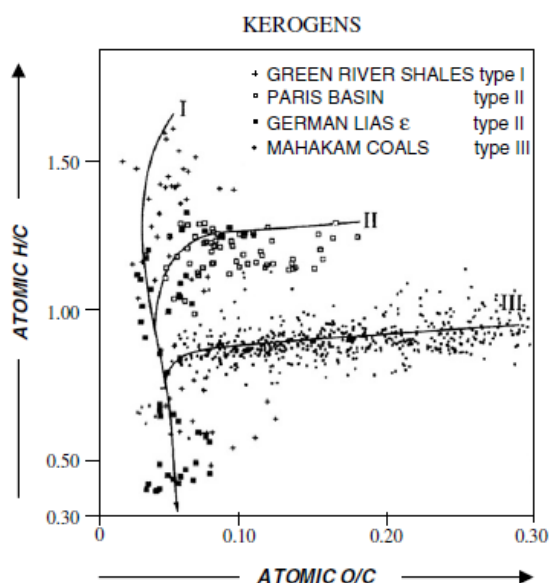


Figure 3. Van Krevelen diagram, taken from Vandenbroucke¹

To follow the decomposition path, we have to go from the upright corner to the bottom left one. We see on this diagram that kerogen first loses mainly oxygen atoms, in

ⁱⁱ Higher plants, also called vascular plants, are plants that can be planted and which have developed vascular tissues to transport water and minerals. They are mainly found on the earth surface but some of them are submerged.

particular for types II and III, which are the types containing the most oxygen. It then loses hydrogen atoms in a more significant proportion.

Organic matter decomposition

A part of biomass decomposes into what we call organic matter. Insoluble part –in usual solvent- of this sedimentary organic matter is called kerogen. A part of this kerogen will then decompose into hydrocarbons.

This decomposition can take place under certain conditions. Indeed, this process is the response of organic matter to a change of its environment. The matter will not change if it is in equilibrium. Some conditions of the environment, temperature, for example, have to change to put the matter in a non-equilibrium state and thus decompose while the organic matter tries to reach a new equilibrium state. This decomposition occurs in three main stages, called diagenesis, catagenesis, and metagenesis.

We can interchangeably use different expressions when we talk about sedimentary organic matter decomposition as transformation, thermal alteration or organic metamorphism.

Stages of the decomposition

There are three main stages of decomposition⁴ explained here: diagenesis, catagenesis, and metagenesis. Every kerogen type follows these three stages.

Diagenesis

Diagenesis is the first stage of decomposition of organic matter. The main process during this step is the loss of oxygen atoms. This loss is illustrated in the Van Krevelen diagram, Figure 3, by the almost horizontal part –mostly for types II and III- on the right part of the diagram. This loss leads to an important decrease of the O/C ratio, a small decrease of the H/C atomic ratio and has been shown to be mainly a result of the elimination of oxygen containing groups^{4,6} such as C-O-, and C=O.

At the end of this stage, a small amount of hydrocarbons has been generated but a larger amount of water and carbon dioxide is found. Indeed, H₂O and CO₂ are the main ways for oxygen atoms to recombine. Kerogen is immature and has usually reached a depth between 800 and 1200 meters. Depending on the geothermal gradient, diagenesis can occur deeper.

Before diagenesis exists another stage called early diagenesis. This stage is driven by biological processes not in the scope of this manuscript.

Catagenesis

Catagenesis is the second main step of sedimentary organic matter's transformation. It is mainly characterized by a significant loss of hydrogen atoms due to hydrocarbon generation and release. Indeed, catagenesis is the main window of oil generation and the beginning of the cracking zone, when a lot of methane is produced. In hydrocarbons there is usually twice as much hydrogen atoms than carbon atoms, consequently, during catagenesis, kerogen loses, in average, two hydrogen atoms for one carbon atom.

Moreover, kerogen keeps losing oxygen atoms -even if it is no longer the most important loss- by still producing usual small molecules as water, carbon dioxide, and carbon monoxide. Catagenesis usually ends at an atomic ratio H/C of 0.5 and from this point on, all kerogen types follow the same decomposition path to a more mature structure.

During this stage, immature kerogen is in a plastic domain and is like a soup of aromatic stacks and aliphatic parts. Aromaticity of the matter increases during catagenesis and aliphatic chains are reduced to reach a brittle matter. Aromatic sheets are randomly distributed. Burial depth is usually between 1000 and 3000 meters and can go up to 4000 meters in the Sahara, for example.

Metagenesis

Metagenesis is the third and last stage of organic matter decomposition. A major part of oxygen and hydrogen atoms has been removed during the first two stages. Consequently, metagenesis is mainly represented by a rearrangement of aromatic layers that were randomly distributed in the last stage. Elimination of impurities and a significant part of hydrogen and oxygen content during the first two stages allows a reorganization of aromatic stacks and a decrease of interlayers distance. Aromatic stacks have a size between 50 and more than 500 Å³ depending on the kerogen type. Indeed, those stacks are larger in a type I than in a type III kerogen. At this stage, matter is less sensitive to degradation and its structure is simpler. Moreover, since the less dense components have been removed, the residue is more and more condensed.

This stage occurs in really deep underground, usually deeper than 4000 meters. It can occur up to 10 000 meters and is totally in the dry gas zone -i.e. no liquid phase coexisting with the gas. It is recognized that kerogen with high hydrogen content will generate mostly oil and decomposition of kerogen with low hydrogen content is dominated by gas generation.

Conditions of the decomposition

Evolution of kerogen during burying is its answer to the change of its environment's conditions. Indeed, if environmental conditions would not have changed, pressure and temperature would have stayed the same than the ones at the surface for millions of years. Sedimentary organic matter would thus have reached its equilibrium state at the given pressure and temperature. Omitting the role of biological organismsⁱⁱⁱ, this equilibrium state would not have been really different than the original one.

In order to change from biomass to hydrocarbons, organic matter needs to be subjected to a certain change of pressure and temperature. In our work we are interested in the decomposition seen as a giant pyrolysis and we do not take into account the role played by biological organisms. The reason for that is practical. It is really complicated to simulate it by numerical simulation. We do not offer an exhaustive view here, but we focus our work on model systems as cellulose and lignin.

Temperature

Natural radioactivity of earth's mantle is the main source of underground heat. Indeed, during their disintegration, radioactive components of the mantle produce energy in the form of heat. After being produced into the mantle, heat is homogeneously flowing to the surface, this heat flux depending on the rocks thermal conductivity.

Due to this, a geothermal gradient exists which is the evolution of the temperature with depth. It is expressed in a temperature unit per distance, like °C/m for example. In Europe, the average geothermal gradient is 3°C/100m it is to say 30°C/km. This gradient depends on the underground composition and is usually comprised between 10 and 100 °C/km. In Figure 4 we can see the average evolution of the temperature with depth, in the case of a geothermal gradient of 35°C/km.

Organic matter burying is really slow since its environmental temperature increases from 0.5 to 20°C per million of years. It means that kerogen can stay at the same temperature for a really long time.

ⁱⁱⁱ For the few first meters below water-sediment interface, living organisms start to degrade organic matter. This process is due to anaerobic bacteria which extract nitrogen and some oxygen of organic matter^{1,64}.

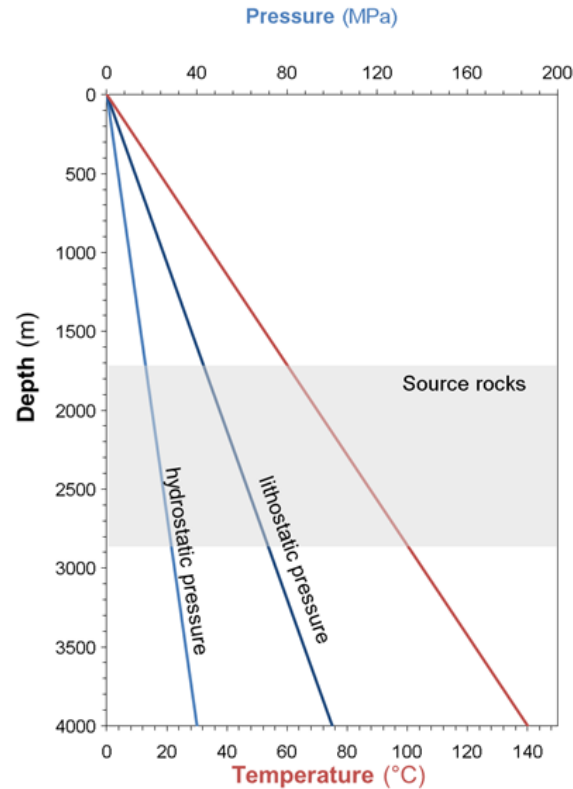


Figure 4. Evolution of lithostatic pressure (dark blue) and temperature (red) with depth (courtesy of J. Berthonneau)^{iv}

Organic metamorphism is a chemical process and due to this fact is a strongly temperature-dependent⁷ effect. At a given temperature, any system has a thermodynamic equilibrium state. If the environment does not change, if we stay at the same temperature and pressure, the system will not change. At equilibrium, every atomic bond of the structure fluctuates around its equilibrium state. Those fluctuations increase with temperature. If the temperature is high enough, fluctuations can be sufficient to overcome activation barriers. In this case, structure will rearrange in order to find a new equilibrium state corresponding to the new temperature. This kind of rearrangements –under increasing temperature– is what happens during the thermal alteration of organic matter.

Pressure

At a certain depth into soil, matter will be subjected to a given pressure. Pressure applied to an element in ground is function of its depth and density of above matter (see

^{iv} Difference between hydrostatic and lithostatic pressures lies on the kerogen position. Hydrostatic pressure will be applied if the kerogen is positioned into an open porosity. In this case, pressure will be equal to $h \cdot \rho_e \cdot g$ where ρ_e is water density, h is the kerogen depth and g is the acceleration due to gravity. In the case of a closed porosity, pressure will be due to the solid as $h \cdot \rho_s \cdot g$. ρ_s is higher than ρ_e . This is the reason of the two different pressure slopes.

footnote iv). Therefore, if the element is deep, pressure applied to it will be important. In the case of kerogen, pressure, resulting from the weight of the rocks above, is called lithostatic pressure. Obviously, it increases with depth. A view of the evolution of this pressure is given in Figure 4.

Pressure role in organic matter decomposition has been less studied than temperature's. High pressure can be obtained in the laboratory but specially on really small samples^v. In addition, pressure can be the lithostatic one –resulting from rocks weight- but can also come from generated products during thermal evolution, as water or gas and it is not simple to reproduce natural conditions for the simple reason that we do not totally know them.

Nevertheless, some studies have been made on the role of pressure during organic matter thermal alteration. Two different points of view emerged from the results of these studies. Some researchers have found no effect of pressure⁸ – in the range 500 to 4000 bars- on the evolution of organic matter under thermal stress, while some others noticed that elevated pressure -1000 bars- had a retardant effect on maturation⁹. A combined vision is to consider pressure to play a minor role in kerogen maturation below a certain value¹⁰ –around 2000 bars.

In our work, the main decomposition factor is the temperature. However, we want to keep our systems pressure in a realistic domain. The lithostatic gradient can vary, as the geothermal ones, depending on the soils compositions, but a commonly accepted value is between 20 and 25 bars/100 meters. On the other hand, many studies have also been made around 250 bars^{11,12}. For those reasons, we decided to keep, in our simulations, our system pressure in the range [250-1250 bars], equivalent of a depth comprised between 1 and 5 km with a gradient of 250 bars/km.

Evolution of the structure

As organic matter decomposes, it loses heteroatoms (non-carbon atoms) as hydrogen and oxygen that form small molecules. This loss induces kerogen to rearrange its structure.

This evolution of sedimentary organic matter structure leads us to distinguish two classes of kerogens. The first class is composed by **immature kerogens**. In this category are the structures with O/C and H/C atomic ratios under 0.1 and above 1 respectively and with a sp^2/sp^3 ratio around $\frac{1}{4}$. They are in the middle of the decomposition in Van Krevelen diagram and they are rich in saturated aliphatic chains. The second class is composed by what we call **mature kerogens**. Their O/C atomic ratio is usually under

^v Multi-anvils press⁶⁵ can reach up to 20GPa and a temperature of 2200°C on few mm³ samples.

0.1 and the H/C one is under 0.5. Their sp^2/sp^3 ratio lies between 5 and 20 which means that the relative number of sp^2 carbons increases with maturation. These kerogens are mainly composed of aromatic structures and some unsaturated chains.

If one looks at the C-C-C bond angle distribution¹³, one finds a peak around 109° for immature sample. This angle value is characteristic of sp^3 carbons. The main peak for mature sample is around 118° when a characteristic angle for sp^2 carbon is 120° . It was also found peaks at 140 and 180° interpreted as unsaturated cumulene chains, so sp carbons. In mature sample, more sp than sp^3 carbons are found and the dominant fraction is composed by sp^2 carbons. Hybridization depends on the hybrid orbital organization. Hybridization can usually be easily deduced from the number of its first neighbors. For example, a carbon atom will usually be hybridized sp if he has two first neighbors –and so two bonds–, sp^2 if it has three and sp^3 if he has four neighbors.

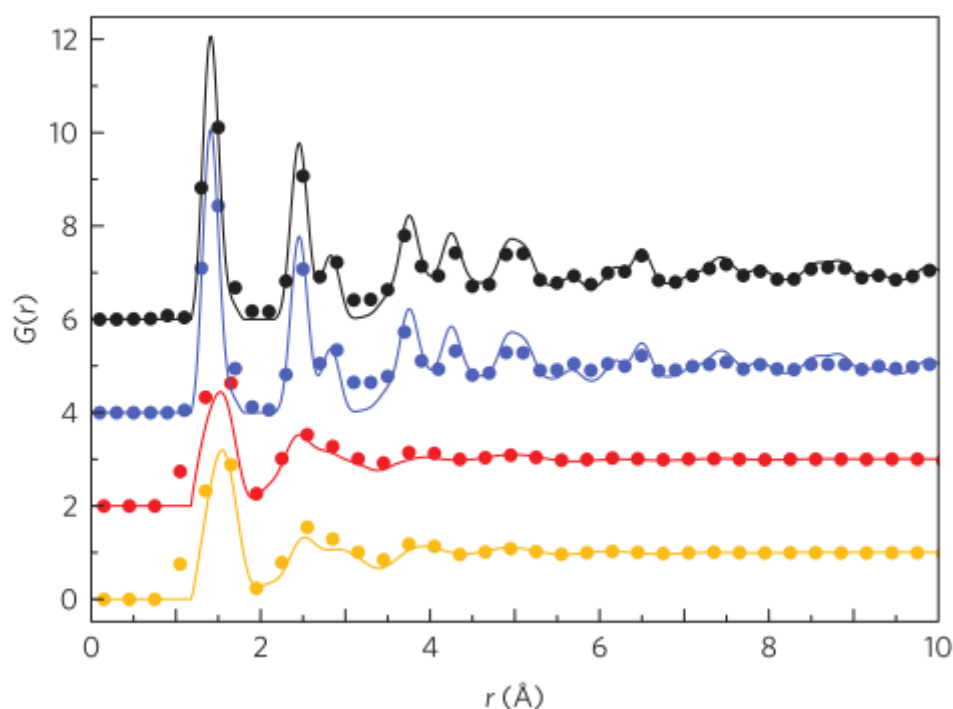


Figure 5. C-C pair distribution functions for different kerogen samples. Lines represent experimental data obtained from neutron diffraction and points represent simulated curves. Black is for *PY02* sample, an over mature kerogen, blue is for *MarK* sample, a mature marine kerogen from Marcellus, red is for *EFK* sample, an immature marine kerogen from Eagle Ford and yellow is for *MEK* sample, an immature Sulphur-rich marine kerogen from Middle East¹³.

A radial distribution function is the representation of the probability distribution of distances between atoms, gives comparable information –see an example in Figure 5. Indeed, for immature sample one find the main distance of $1.54\text{\AA}$, characteristic of sp^3 carbons, and a well-defined peak at $1.41\text{\AA}$, characteristic of sp^2 carbon. This well-

defined peak is an indication of a more ordered structure. Mature kerogens are more ordered than immature ones.

As mentioned in the section on stages of organic matter decomposition, during catagenesis, *i.e.* the middle of the process, aromatic clusters appear but are randomly distributed. They organize during the last stage, metagenesis, during which kerogen gets more ordered. It goes with a condensation of the residue and a lower sensitivity to degradation.

With this arrangement, nanopores size increases and one finds larger pores in mature samples than in immature ones. Larger pores mean larger amount of hydrocarbons and *vice versa*¹³.

Hydrocarbons formation

Petroleum generation is a chemical temperature-controlled process⁷. We have seen that among both parameters –temperature and pressure- temperature is by far the most important. During its thermal decomposition, organic matter releases hydrocarbons. Thereby petroleum formation results from kerogen decomposition. The kind of hydrocarbons and the product proportions mainly depend on kerogen type and thermal history. While decomposing, kerogen is mixed with sediments layers. If those sediments contain more than 1wt% of organic matter, they are called **source rocks**.

Oil and gas formation domains

During Earth's history, six time intervals –called **stratigraphic intervals**- have been identified as petroleum formation periods¹⁴. These six intervals are the main ones since they involve rocks which have provided more than 90% of discovered petroleum reserves.

The first one is the Silurian which has generated 9% of the world's reserves. The second and third ones are Upper Devonian-Tournaisian and Pennsylvanian-Lower Permian which have both generated 8% each of the reserves. However distribution is uneven. For example, the fourth and fifth intervals, called Upper Jurassic and middle Cretaceous, generated 25% and 29% respectively of world's reserves. The last and sixth one is called Oligocene-Miocene and represents 12.5% of global reserves. These stratigraphic intervals differ by the time of generation but also by the geographic position of resources and type of kerogen present in source rocks.

During sedimentation, in all stratigraphic intervals, organic matter decomposed, if conditions were favorable, into hydrocarbons. Hydrocarbons can be oil or gas, depending on their carbon chain length. Short alkanes will be gas and long alkanes will be oil. Depending on hydrocarbon type, the formation domain will be different.

Hydrocarbons formation starts when kerogen becomes mature, and is a result of kerogen thermal degradation and cracking. Two domains of hydrocarbon generation are distinguished.

The first one is called **oil window**, and corresponds –as its name suggests- to the temperature window when oil is generated. It is temperature dependent and occurs on average at a depth comprised between 1000 and 4500 meters when the temperature increases from 60 to 150°C¹⁵. The maximum of oil generation is usually found between 2000 and 3000 meters.

The second domain is called **gas window** and corresponds to the gas generation interval. It occurs between 3000 and 6000 m of depth, when temperature increases above 100 °C¹⁶. Cracking plays also a role in gas generation, since it is the decomposition of heavy hydrocarbons into lighter molecules. Obviously, generation domains are not exactly delimited and can overlap. For example, gas starts to be generated at the end of oil generation, at this moment one may obtain wet gas that is a mixture of gas –mostly methane- and heavier hydrocarbons, sometimes in liquid state –meaning that they come out in a liquid state in the well (at P,T of the well).

Once they are generated, a significant part of these hydrocarbons is expelled from the source rocks and starts its migration to upper levels, through wide pores, fractures, and permeable rocks. This migration is stopped when hydrocarbons meet an impermeable cap-rock. It may form an oil accumulation that is called **reservoir**. This kind of reservoir is what is exploited during conventional extraction. Unconventional one corresponds to the extraction of the hydrocarbon fraction still in source rocks, thus hydrocarbons that have not migrated.

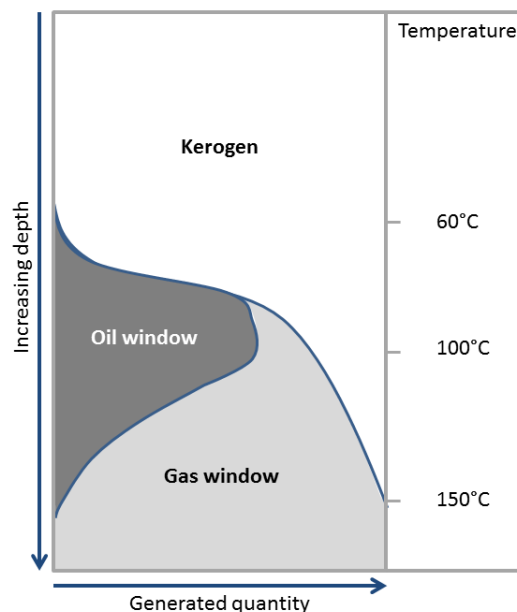


Figure 6. Oil and gas windows

An overview of oil and gas windows is given in Figure 6.

Hypothesis

Hydrocarbon formation is an irreversible process. Indeed, one never saw kerogen transform back into higher plants or phytoplankton. Knowing that temperature is continuously changing while organic matter is buried –on a geological time-, study of oil generation is traditionally based on a kinetic approach. It is considered as a purely kinetic process and industrial basin models are based on a large set of first order reactions. It works pretty well and gives useful information for oil exploration but requires time and resources to adjust. Moreover, this technique does not give accurate information on the kerogen dispersed in the source rock nor on its hydrocarbons deliverability.

A first hypothesis we would like to address with our work is the idea that petroleum generation is a phase separation process. Indeed, it looks as if a soft immature kerogen separates into two phases during its decomposition. From an aliphatic structure, it goes to a mixture of a more mature carbon-rich aromatic kerogen and a hydrogen rich fluid.

Another question that will be interesting to address is whether the mixture ever reaches a thermodynamic equilibrium state or not. In other words, is the fluid in equilibrium with residual kerogen?

This information could allow us to build a phase diagram in which the separation will be made between a miscible phase at a pretty low temperature and a two-phase separation above a critical temperature. This kind of diagram would be a really useful tool by giving, for any kerogen that has reached a given temperature in its basin, information on residual kerogen and hydrocarbon phase and, in particular, ideas on hydrocarbons deliverability of the basin of interest.

Obviously, in real basins, a large variety of elements, from organic matter to rocks contribute to the process, and it is complicated and resources demanding to try to model it. We thus have to limit our work if we want to be sure to get accurate results. We here neither simulate the action of bacteria, nor the possible catalytic role of rocks in the maturation process. We limit our investigations to kerogen decomposition, starting from an organic precursor and we simulate the alteration process as a pyrolysis process.

Characterization of Kerogen

A number of tools is used by petroleum geologists in order to characterize kerogen. Their goal is to understand the history of geological formations and consequently,

knowing the potential of the reservoir. These tools are designed to reduce the risks and cost of exploration and exploitation.

Total organic carbon content

Knowing the quantity of organic matter in the sediment rock is important –but not sufficient- in order to have an idea of commercial quantities of oil and gas. This potential will depend not only on this quantity but also on the kerogen type and the feasibility of the extraction.

We saw that hydrocarbon generation comes from organic matter decomposition. Hence there is a clear link between organic matter content and hydrocarbon extraction potential of the source rock. Total organic carbon, abbreviated **TOC**, is measured as the percentage of organic carbon of the total weight of the sediment. This value is used to characterize the source rock as shown in Table 1. This estimation can be done by neutron-induced gamma ray spectroscopic measurements¹⁷.

Table 1. Denomination of source rocks depending on the total organic carbon value. (based on Cornford, 1998)¹⁸. Denomination used to *rate source rock in terms of amount of sedimentary organic matter*.

TOC (%wt)	Denomination
<0.5	Very poor
0.5-1.0	Poor
1.0-2.0	Fair
2.0-4.0	Good
4.0-12.0	Very good
>12.0	Oil shale or bituminous limestone
>65.0	Coal

The total organic carbon value can be parametrized in order to take into account the effects of hydrocarbon maturation and expulsion¹⁸.

Nuclear Magnetic Resonance

Solid state ¹³C Nuclear Magnetic Resonance (**NMR**) is a non-destructive method widely used to study char structure. It provides structural information such as the fraction of aliphatic and aromatic structures, specific functional groups, and heterogeneities¹⁹. It also gives information on carbon environment.

With Infrared spectroscopy (**IR**), NMR is more appropriate to provide chemical structure details than destructive methods. Moreover, NMR is particularly adapted to handle organic matter and thus is the method of choice²⁰. Infrared spectroscopy is only qualitative. As opposed to infrared, NMR is the only method that can give quantitative estimations on carbon bonds types²¹ under controlled experimental conditions. These estimations are usually given with an uncertainty around 10%.

Nuclear magnetic resonance is a particularity of atom nuclei which have a nuclear spin, like ^{13}C . As a brief explanation, we can say that these nuclei, under an electromagnetic radiation, absorb the energy of the radiation and then release it with a specific frequency. This frequency provides information on the kind of atoms and bonds. Several kinds of NMR exist, both for liquid and solid-state.

In the case of kerogen characterization, since, by definition, kerogen is insoluble, high-resolution liquid NMR methods are not applicable, so, only solid-state NMR can be used. The most used solid-state NMR technique is ^{13}C cross polarization/magic angle spinning (**CP/MAS**)^{22,23}. The combination of these two methods is able to resolve the aromatic and aliphatic spectral zones²⁴.

One concern with the characterization of kerogen by solid-state NMR was the influence of mineral and paramagnetic materials on the signal. Analyses are usually performed on isolated kerogen. Some studies have been performed to address the influence of separation procedure, also called demineralization –commonly made with strong acids such as HF and HCl- and no significant impact has been found. Resolution is even better after isolation but one may keep in mind that acidic procedure might have an impact on kerogen structure²⁵.

^{13}C CP/MAS NMR is also a useful tool due to the fact that it gives information on aliphatic and aromatic distributions. Indeed, an aliphatic-rich kerogen will generally produce more oil than an aromatic-rich one. Moreover, an increase of aromaticity coupled with a decrease of aliphatic fraction is a good indicator of kerogen maturity. The ratio of aromatic to aliphatic carbon and aromatic clusters size are also interesting indicators of maturity²⁶ and increase with it. Other kinds of NMR exist but we do not discuss them here since they are less used.

NMR, and especially ^{13}C CP/MAS NMR, gives useful information on kerogen structure and maturity.

Rock-Eval

The **Rock-Eval** thermal analysis technique is one of the most used in kerogen characterization. It is based on an open pyrolysis and is coupled with infrared detector. It enables a rapid geochemical characterization of kerogen in rocks. It does not require

important resources and provides useful information. It gives indications of kerogen maturity and type, the presence of migrated hydrocarbons and the potential content of hydrocarbon¹⁸. Kinetics of kerogen cracking into petroleum and importance of time-temperature couple led to use experimental pyrolysis to evaluate oil generation potential. This analysis is possible with the Rock-Eval machine.

In order to have a Rock-Eval analysis, a sample of rock powder is pyrolyzed under an inert atmosphere, such as N_2 , following a well-determined sequence. The sample is only submitted to the effect of heat, without any oxidant.

Up to $300^{\circ}C$, the rock releases hydrocarbons and a peak noted S_1 , corresponding to the volatilization of hydrocarbons, appears on the Rock-Eval graph.

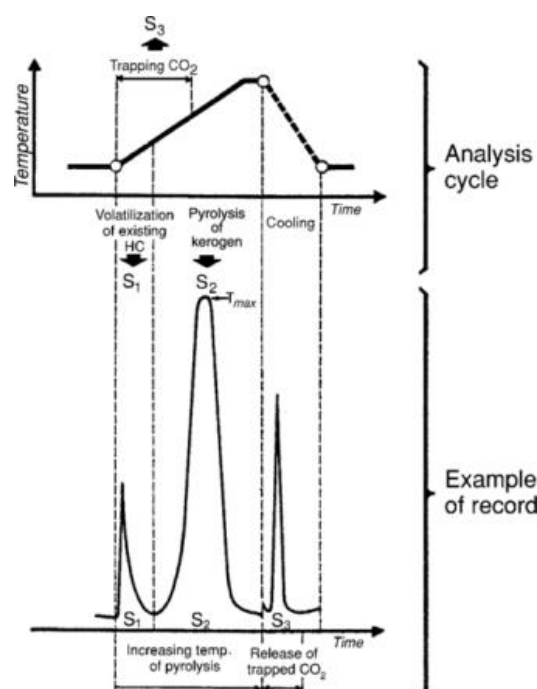


Figure 7. Rock-Eval analysis. From Tissot and Welte⁴

From $300^{\circ}C$ to $550^{\circ}C$, modification of kerogen structure, coarsely mimicking the maturation, is performed and gives the S_2 peak, which is used to evaluate the hydrocarbon generation potential of the rock. The area under S_2 gives an idea of the amount of hydrocarbons possibly extractable. The top of the S_2 peak corresponds to the temperature at which the maximum of hydrocarbons generation occurs –noted T_{max} – and increases with kerogen maturation. Thus, the value of T_{max} provides information on kerogen maturity and depends on its type²⁷. Still between $300^{\circ}C$ and $550^{\circ}C$, trapped CO_2 is released and provide the so-called S_3 peak.

A typical Rock-Eval plot is given in Figure 7.

Vitrinite reflectance

Vitrinite reflectance is used to characterize kerogen thermal maturity. It measures the percentage of light reflected by vitrinite maceral under special conditions. Vitrinite maceral is a component of coal –in the same way than minerals are components of rocks- and derives usually from vascular plants. It has a shiny appearance. Vitrinite mainly comes from lignocellulosic tissues and for this reason, it can be difficult to assess the thermal maturity of some type II and type I kerogens which do not contain enough vitrinite²⁸.

Vitrinite reflectance increases with maturation and thus is used as an indicator of thermal maturity. Measure of vitrinite reflectance is made using an optical microscope. It needs to be equipped with an electron photomultiplier which quantifies the reflected light. This reflectance is given in percentage²⁹.

Other more common characterization methods such as X-ray and electron diffractions for instance are not described in this manuscript.

Previous studies

Thermal decomposition of biomass has been the focus of research for decades, and this for several reasons. Among these, a better understanding of the structure of kerogen would lead to a better extraction, and a better knowledge of degradation process could help processing hydrocarbon.

In 1924, Trager² suggested that kerogen was an intermediate product between organic source and hydrocarbons. He stated that “*Kerogen can be converted into oil, either by the aid of heat or by pressure*” and with this statement formulated the concept of a link between pyrolysis and evolution during natural burial. **Pyrolysis** is the chemical decomposition of an organic source under the only effect of heat, without the action of any oxidant.

A lot of attempts to reproduce natural hydrocarbons formation by heating kerogen of different organic sources was reported in the literature. We give here a non-exhaustive review of studies of this type, first by experiments and then by numerical simulation, in a second section.

Experimental studies

Accepted idea of similarities between pyrolysis and degradation due to burial lead researchers to experimentally simulate in the laboratory organic matter evolution at high temperature over short period of time –few hours to few days. Those experiments took usually place in **open** and rather dry **systems**. Use of high temperature was a way to avoid geological time scale –increase the temperature help to overcome energy barrier involved in chemical reactions- but had an impact on obtained products. Indeed, the solid part was close to its natural counterpart but the fluid part was totally different from natural crude oil.

Lewan^{30,31} built an experiment known as hydrous pyrolysis in which he used a close system maintaining water in contact with organic matter during the whole heating. He actually got better results than in dry systems.

Other researchers³² thought that the decomposition was taking place in a closed and compressed environment, everything staying in contact for a really long time. This kind of technique is called **confined pyrolysis** and allows a change in temperature, pressure and water volume.

Open and closed pyrolysis

The first method that comes to mind when we say “open pyrolysis” is the Rock-Eval one. As we already discussed here above, the Rock-Eval technique is an open pyrolysis used for organic matter characterization, that gives information about its potential of oil generation. **Open pyrolysis** has also been used in attempts to reproduce natural decomposition of organic matter. The principle is to insert the sample into a – usually preheated- oven and heat it under an inert gas –as nitrogen or argon- flow. Once the working temperature is reached, the sample is quickly cooled. The oven is obviously closed and the atmosphere is an inert gas in order to avoid oxygen coming to the experimental reactor in which case it would not be pyrolysis anymore but combustion. With this technique, the obtained solid part is in a quite good agreement with its natural equivalent but the effluents are really different.

The main **closed pyrolysis** method is the **hydrous** pyrolysis, proposed by Lewan³⁰ in 1978. The principle is to fill half a pressure vessel with water and organic matter sample and the other half with helium at a pressure of 1 atm. The whole system is then heated for few days around 300°C. He found a better correlation between his results and natural crude oils than with the anhydrous pyrolysis method.

Both kinds of methods were used and compared. In 1972, Goodarzi and Murchison³³ carbonized coals between 20 and 950°C and studied the evolution of samples vitrinite reflectance. Six years later, Rouxhet and Robin⁶ pyrolyzed natural kerogen samples in open-pyrolysis system and studied their evolution using infrared spectroscopy. In 1987 Bourrat *et al.*³⁴ studied sulfur behavior upon heating heavy petroleum products. In the present work, we are not interested into sulfur but Bourrat's study shows that open pyrolysis was a useful tool. They pyrolysed their "*heavy petroleum fractions*" at a temperature between room and 2900°C and under inert gas flow.

Several studies comparing open, closed and confined pyrolysis have been published: Monthioux *et al.*^{8,35} performed experimental pyrolysis on different coals. They noticed that the open-pyrolysis did not reproduce the natural behavior. The same conclusion was reached by Behar *et al.*³⁶ comparing open and anhydrous closed methods. They noticed a secondary cracking reaction and an underestimation of methane potential with the open pyrolysis. Monthioux *et al.*⁸ clearly stated that "*Open-system pyrolysis [was] not suited for simulation of the natural evolution*" and found "*Best results [...] by confined pyrolysis experiments*".

Confined pyrolysis

We made the choice to oppose **confined-system pyrolysis** to open-system pyrolysis. We consider now as closed pyrolysis, pyrolysis in sealed glass tubes in which the confinement is intermediate -between open and confined systems. Confinement is defined as the inverse of the dead-volume, it is to say, the volume without solid.

The idea was that in the deep underground sedimentary matter was subjected to geological pressure and almost all the chemical reactions were taking place inside the sedimentary rock porosity. Therefore in a closed system, all initial components are staying in contact for a quite long time. Due to this closed environment, secondary reactions were possible. The main difference between closed and confined pyrolysis, beyond the dead-volume, lies on the kind of containers that are used. In the first method, rigid containers are used -that's the reason why a significant dead volume is necessary to avoid explosions- while confined technique requires ductile recipients -usually gold tubes. Samples are placed into these tubes that are then heated into hydraulic pressure bombs. Strong confinement has been proved to be a determining factor in order to have a good simulation of natural organic matter degradation³⁷. Such an experimental simulation has been proven³² to need a certain level of partial pressures due to confinement -pressure of effluents- to be reached in order to be satisfactory. In the same time, the total pressure seems to not have a great impact in the range [500-4000bars].

In 1985, Monthioux *et al.*³² proposed the confined pyrolysis and showed that kerogen behavior during natural evolution could be reproduced in laboratory. They also found

that water had no special effect in their case. Their results are really interesting. They satisfactorily reproduced natural evolution, both following the Van Krevelen diagram and quantities of effluents –see in Figure 8.

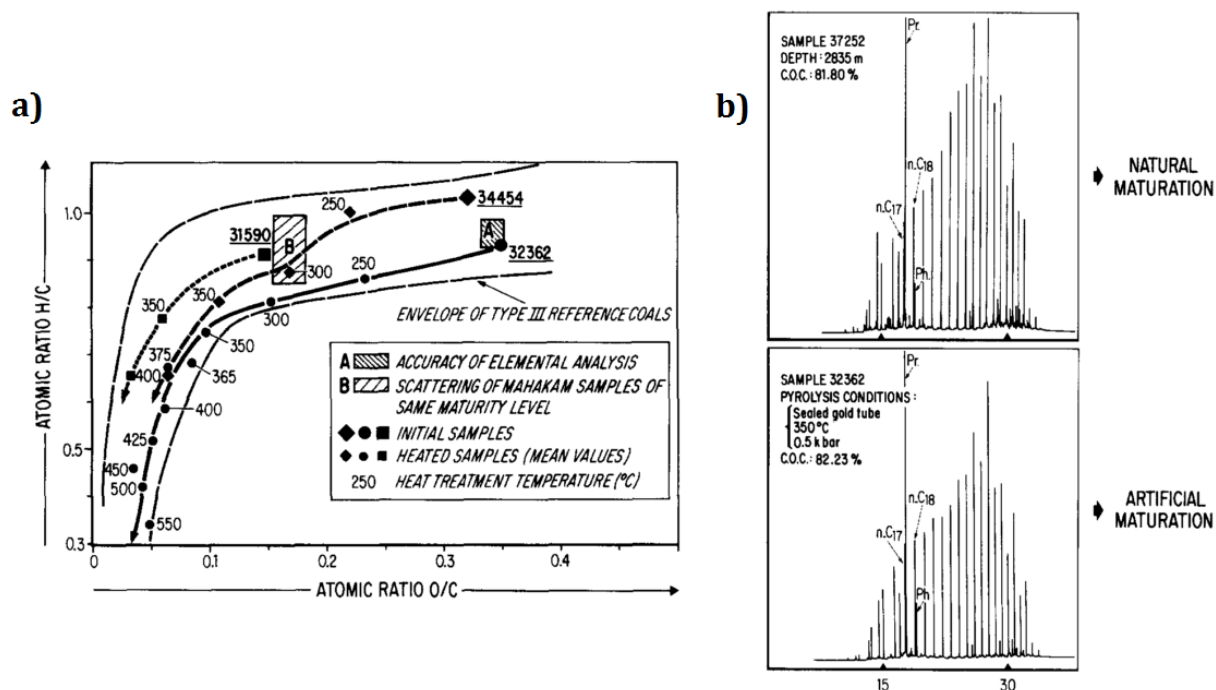


Figure 8. Comparison of natural and artificial evolution obtained by confined pyrolysis, by Monthioux et al.³². a) solid evolution on a van Krevelen diagram and b) comparison of effluents.

Several other studies confirmed that confined pyrolysis gave better results than open and closed –both wet and dry- pyrolysis^{8,35,37-39}.

However, Behar and Hatcher⁴⁰ artificially maturing a fossil wood using the confined pyrolysis method noticed some discrepancies. Confined pyrolysis apparently well-reproduces transformations of aromatic rings but not the reactions involving aliphatic structures. Numerical simulation could be a complement to these experiments in order to get a better understanding.

Numerical studies

As a complement to experimentation, numerical studies give information at the atomic scale. Moreover, it allows comparisons and helps to get a better understanding and to build realistic models.

Two main characteristics of kerogen are explored by numerical simulations. The first one is its structure. Once the macrostructure is known, it is interesting to address the structure at the atomic level. Research teams tried to build molecular models of kerogen using numerical methods deriving from statistical physics. We give a non-exhaustive review of these attempts in a first section. The second interesting aspect of kerogen is its formation. This field is more recent than the study of the structure mainly because of the need of powerful computational tools and resources.

Structure modelling

Kerogen cannot be defined as a simple and single structure. We do not have one defined and exclusive arrangement of atoms. Kerogen structure is indeed a non-crystalline and non-homogeneous system. Thereby, disorder and inhomogeneities have to be taken into account. Kerogen composition and organization depend on a number of parameters such as its initial elemental composition and its degree of maturity. The environment during decomposition, that depends on the location of the deposit, is also important. Kerogen structure is complex and evolves with time and maturation. Immature kerogen matures with the sedimentation of organic matter. Having structure models for several natural samples can help to visualize and understand differences and similarities between kerogen types and also physical characteristics such as mechanical properties for example.

Studies on **coal molecular structure** first appeared in 1942. Fuchs and Sandhoff⁴¹ were interested in coal pyrolysis. Coal comes from organic matter decomposition and, as a matter of fact, type III kerogens give mostly coal as an over mature substance. This model was only 2-dimensional –see in Figure 9.

3-dimensions structures only started to appear in the 1980s⁴². From the sole use of chemistry, researchers⁴³ started to use physical information. Later, in the 1990s, computer simulation democratization helped to construct and visualize coal structure in 3-dimensions⁴⁴ encompassing IR, NMR, and X-ray experimental results.

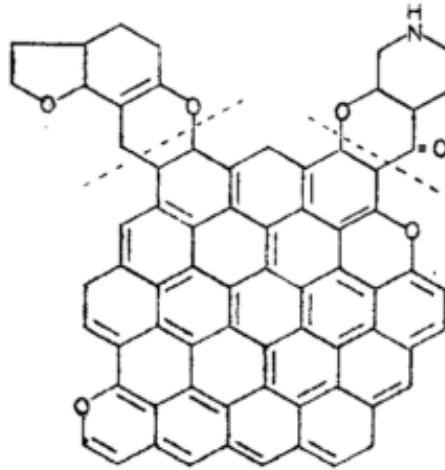


Figure 9. A view of Fuchs and Sandhoff –bituminous- coal structure adapted from ⁴¹

Different kinds of coal exist. Their structures are not always only aromatic as we can notice in Figure 10. Four main categories exist:

- Lignite, composed by 50 to 60% of carbon,
- Subbituminous ones, composed by 60 to 70% of carbon,
- Bituminous ones, composed by 70 to 90% of carbon and
- Anthracite, composed by more than 90% of carbon.

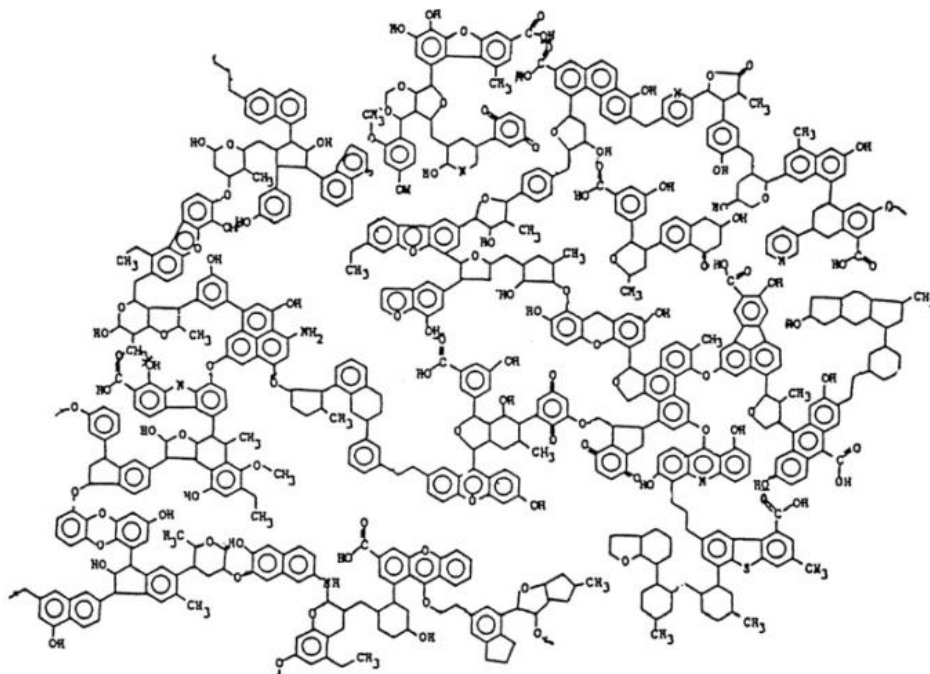


Figure 10. Representation of a subbituminous coal by Shinn ⁴⁵

Kerogen structure elucidation followed the same path in its evolution, improving with techniqueal progress. First studies on kerogen structure have been done on the Green River formation that is in the USA, in northern Colorado. Forsman⁴⁶ first discussed kerogen composition and found two different families of kerogen, that he called coaly and non-coaly depending on the amount of rings and chains in the structure. In 1969, Burlingame *et al.*⁴⁷ studied a carbon-rich sedimentary rock from the Green River formation –around 52 millions of years old- and proposed the first published model –see Figure 11. For its construction, they oxidized samples and used gas chromatography and mass spectrometry to analyze the products. With those results they had an idea of linked, groups and functions they could find in kerogen and they suggested their structure in which they only proposed some building blocks with possible bonds. We can note that they already suggested that compounds could be entrapped into the porosity.

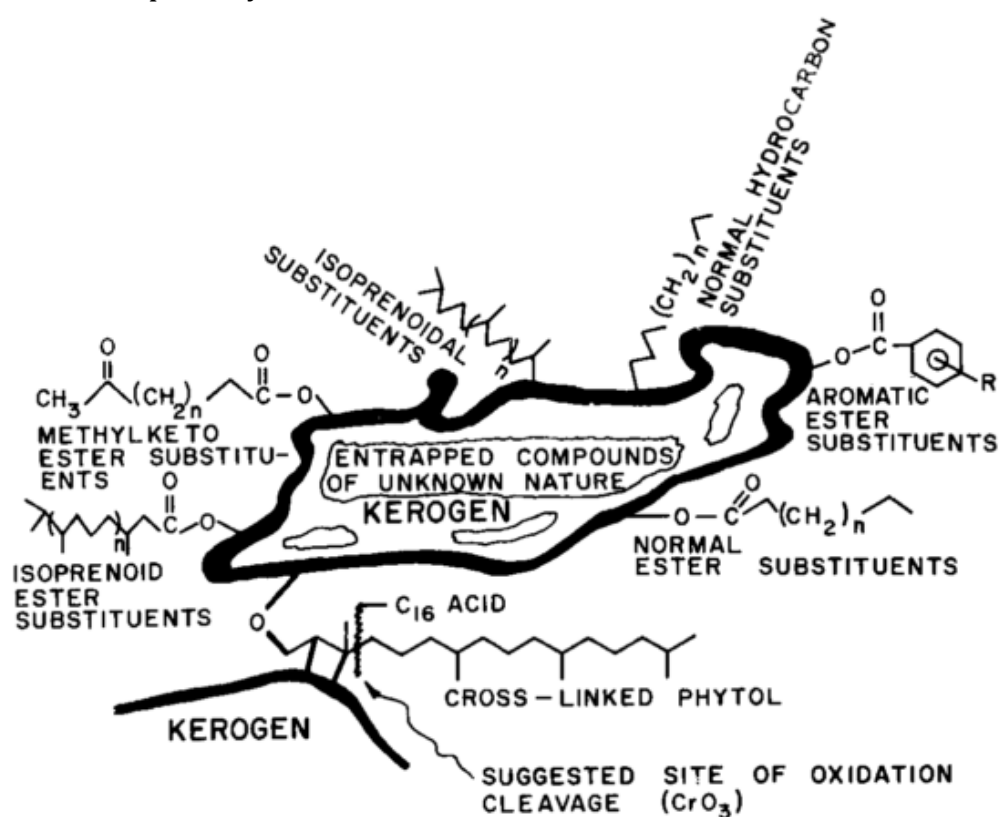


Figure 11. First published kerogen structure, by Burlingmane *et al.*⁴⁷

In 1976 Yen⁴⁸ published the first detailed kerogen structure with a molecular definition. To build it, he used spectroscopic data obtained on a natural sample structure, such as electron spin resonance, IR spectroscopy, and X-ray diffraction. He gave a chemical formula: $C_{235}H_{397}O_{13}N_3S_5$, thus H/C and O/C ratio of 1.69 and 0.06 respectively.

In 1995 Siskin *et al.*⁴⁹ started to study different formations than the Green River one and it was the first use of a ^{13}C and ^{29}Si combined NMR method of characterization. They got more accurate information on constituents and chemical groups presents in their samples.

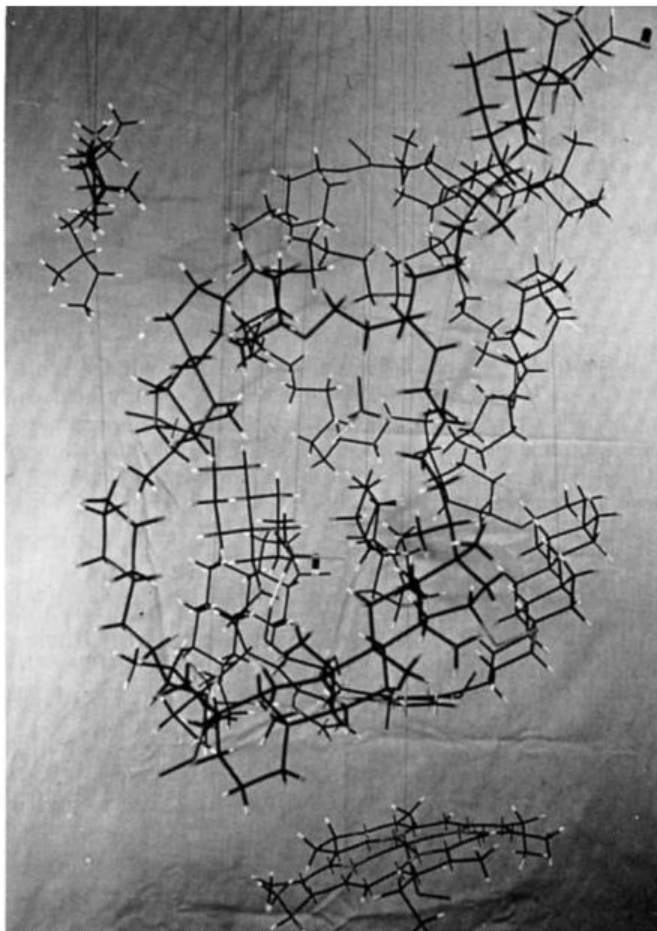


Figure 12. Structure proposed by Yen in 1976⁴⁸

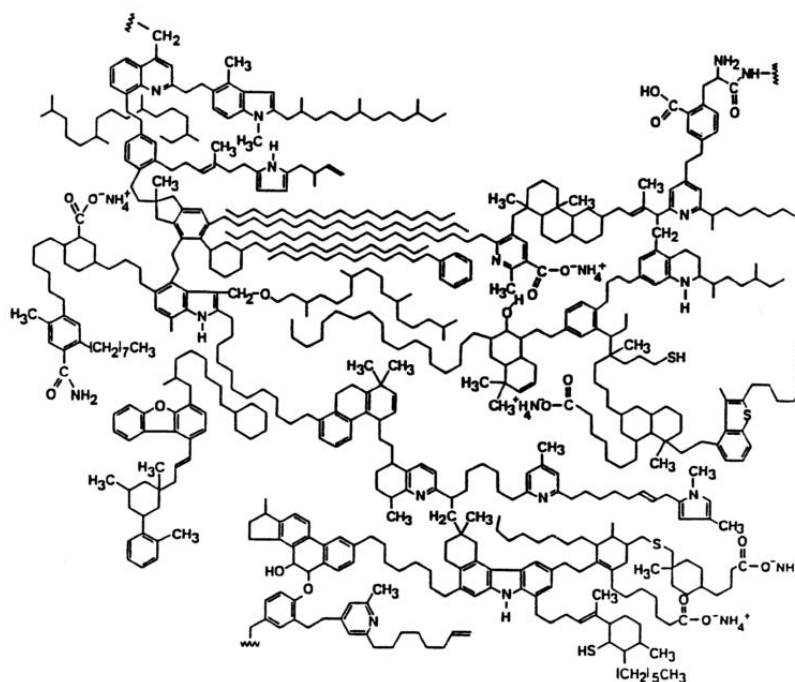


Figure 13. Structure of a Green River shale proposed by Siskin *et al.* in 1995⁴⁹

Researchers then started to suggest molecular models for different kerogen types. All these attempts were not to give a perfect and unique structure since, as stressed before, there is not a one and only configuration for kerogen.

In 1987, Béhar and Vandenbroucke⁵⁰ proposed chemical models of kerogens. They represented the three types of kerogens for three different evolution stages: beginning of diagenesis, beginning of catagenesis and end of catagenesis. They obtained chemical parameters on natural samples using various methods such as elemental analysis, ¹³C NMR, IR spectroscopy or electron microscopy. They could evaluate the H/C and O/C ratios, the proportions of aromatic carbons, the ratios of aliphatic over aromatic carbons and obtain information like the size and orientation of aromatic clusters. Thanks to these data, they were able to build molecular blocks that they connected by covalent bonds, like *Lego* blocks.

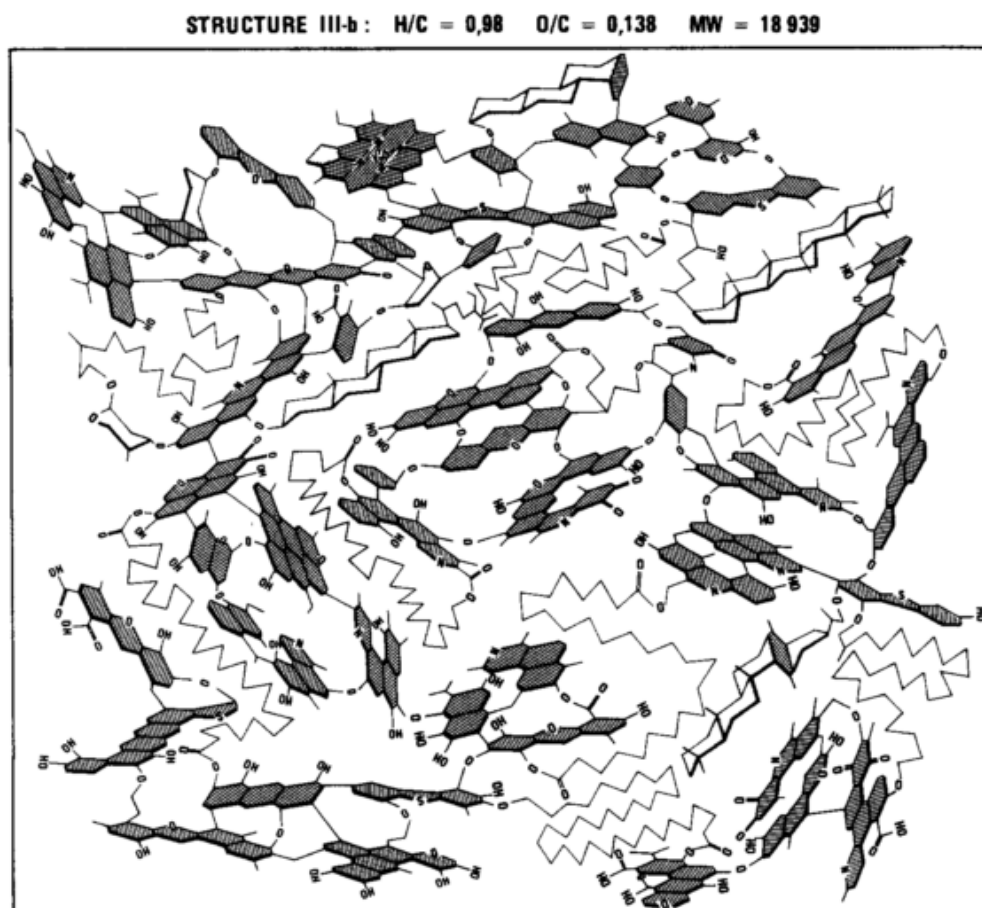


Figure 14. Chemical model of a type III kerogen at the beginning of catagenesis, from Béhar and Vandenbroucke⁵⁰

In Figure 14 and Figure 15 are given chemical models of Béhar and Vandenbroucke for a type III kerogen at the beginning and the end of catagenesis respectively. Between those two models, we note an increase of aromaticity and order and a loss of linear carbon chains.

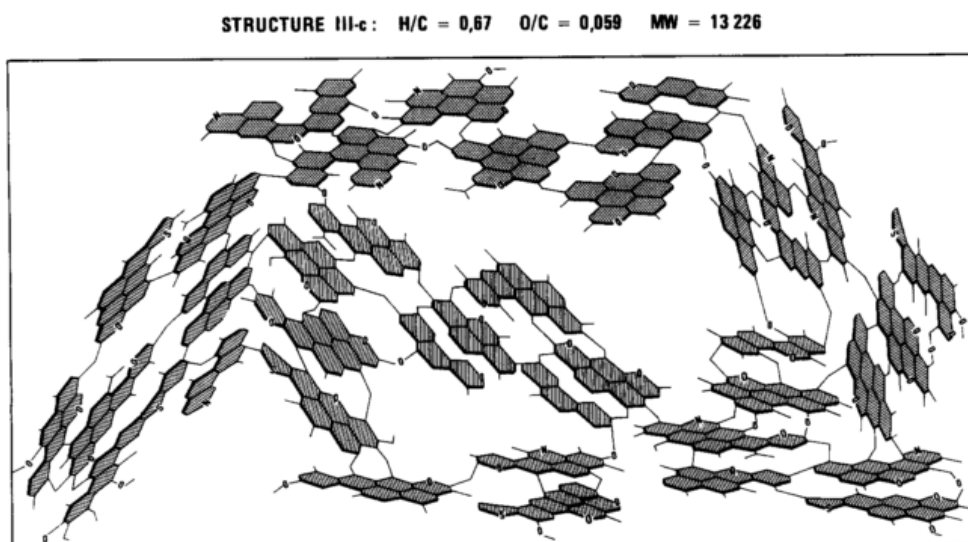


Figure 15. Chemical model of a type III kerogen at the end of catagenesis, from Béhar and Vandenbroucke⁵⁰

Since the 1980s, computational methods have been improved. Attempts to model using computers started with the work of Oka *et al.*⁵¹ in 1977 who published an elucidation of liquefaction products from sedimentary organic matter. They based their work on ¹H NMR and elemental analysis and did not attempt to give a 3D reconstruction.

A more recent study is the one of Ungerer and co-workers⁵², in 2014 is based only on elemental and group analysis. They constructed molecular kerogen models. They used six different samples in order to address the three main types. Their goal was to use these models to predict thermodynamic properties, using molecular dynamics and quantum mechanics. To build their model, they used a molecular builder –in the Medea environment⁵³. A model of a type III kerogen is given in Figure 16.

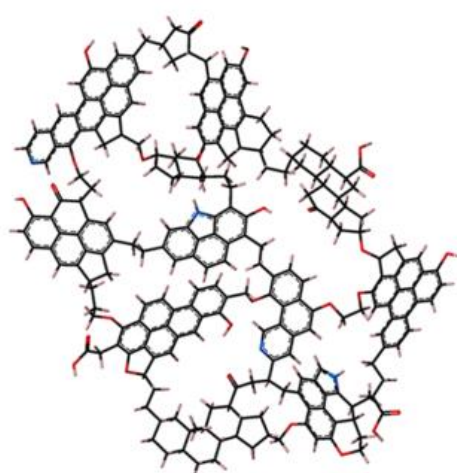


Figure 16. Molecular model of a type III kerogen from Ungerer *et al.*⁵². Its composition is $C_{233}H_{204}O_{27}N_4$.

Another method that can be used for kerogen modeling is the *Molecular Dynamics - Hybrid Reverse Monte Carlo* (MD-HRMC) used by Bousige and co-workers¹³ in 2016. This method relies only on molecular constraints. They use a reactive force field which allows bonds breaking and forming without imposing carbon hybridization or chemical groups distribution. They worked on four different structures and two are given in Figure 17.

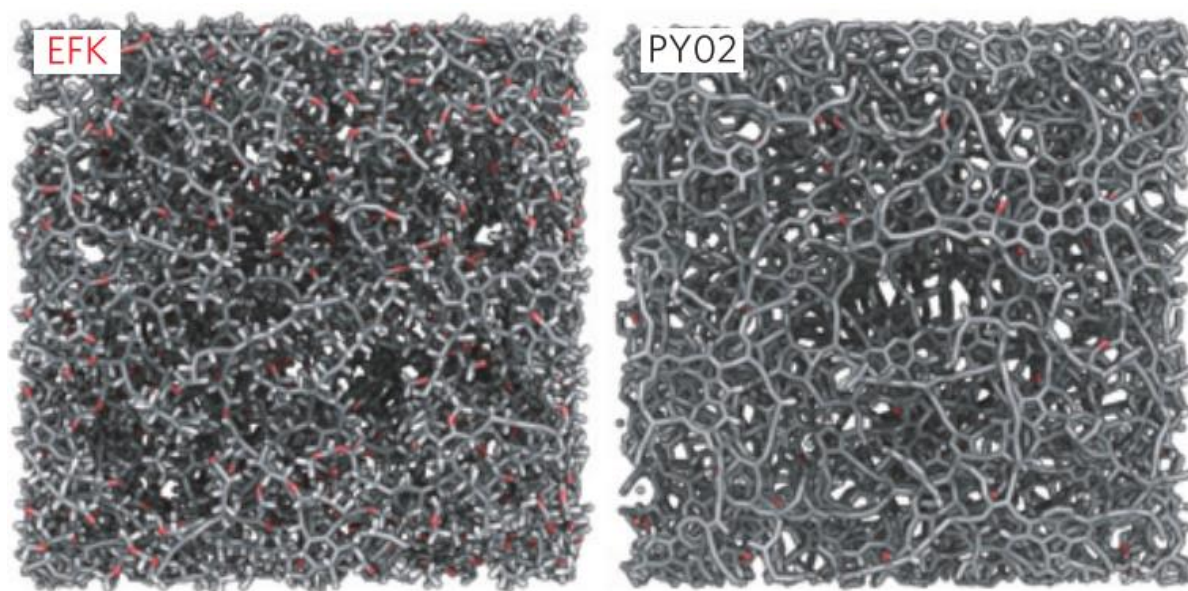


Figure 17. Models for a (Eagle Ford Play Kerogen= EFK) type II and a (mineral-free shungite = PY02) over-mature like kerogens, from Bousige *et al.*¹³

In 50 years models complexity and accuracy have increased a lot. Studies presented so far are based on experimental data fits. In contrast, we base our work on a purely numerical and energy model. In our approach, we propose a way to obtain organic matter decomposition pathway besides the kerogen structure.

Modelling the formation of kerogen

Obtaining atomic scale configurations of kerogen structure is interesting for a number of reasons among which stands the possibility to calculate mechanical properties. It is critical for analyzing hydraulic fracturing, a method used to extract non-conventional hydrocarbons entrapped into the solid. Kerogen formation mechanisms are also crucial since they have a significant impact on its structure and on the products. Moreover, besides this quite industrial point of view, there is also a technic point to address. It would be a great step to be able to reproduce on a computer in an accessible time by human, what nature does in millions of years.

In the last few years, a lot of research teams have tried to study organic matter decomposition by numerical simulations. Among them, Salmon *et al.*⁵⁴ who worked on the decomposition of a lignite geopolymer, the lignite being a sedimentary rock which

contains plant fossils. They submitted their geopolymer model, composed of carbon, hydrogen, and oxygen to pyrolysis at high temperature –between 700 and 3100K. Those temperatures are considered as high. The reason is that into the ground, organic matter faces much lower temperatures, around 400K at 4km depth.

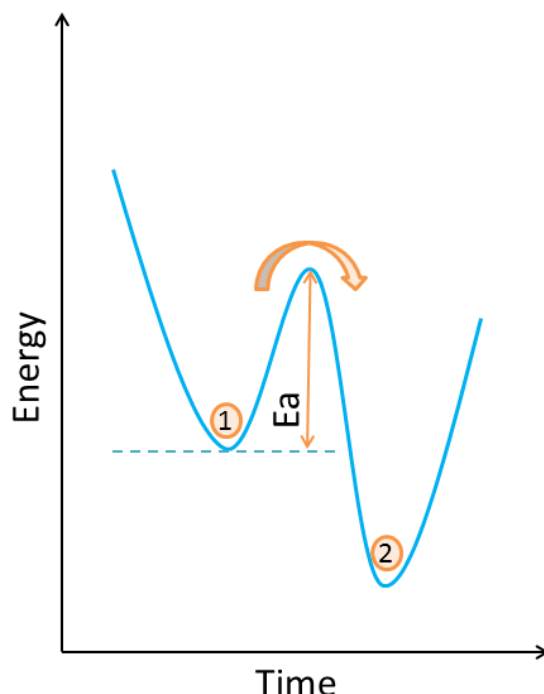


Figure 18. To cross the barrier between energy states 1 and 2, system needs to receive the energy of activation E_a . Speed of the reaction $1 \rightarrow 2$ is function of this energy E_a and of the system temperature T .

Increasing temperature is a way to accelerate a reaction. In fact, at a higher temperature, it becomes easier to overcome any energy barrier, from an initial configuration to another. Indeed, the speed of a reaction is function of temperature T and follows the Arrhenius equation:

$$v = A \cdot e^{-\frac{E_a}{R \cdot T}}$$

Where v is the reaction speed, A is a phenomenological pre-exponential factor, E_a is the activation energy –see Figure 18-, and R is the universal gas constant. E_a corresponds to the energy a system needs to overcome speed. Probability for a system to have energy higher than E_a is $\exp(-\frac{E_a}{R \cdot T})$. Probability for a system to overcome an energy barrier is thus a function of temperature^{vi}.

^{vi} Link between overcoming energy barrier and temperature can also be seen considering the equation $E = a \cdot k_B \cdot T$ where E is the thermic energy, a is a factor which depends on freedom degrees of the system, k_B is the Boltzmann constant and T is the temperature.

This method has a main disadvantage: it shifts equilibrium reactions and leads to higher yields of bio-oils^{55,56}. Indeed, Zou *et al.*⁵⁷ found that ethylene formation was entropically favored at high temperature and Liu *et al.*⁵⁸ noticed that “*the results indicated that the product types and their distribution are strongly influenced by the temperature*”. As shown in Figure 19, Salmon and co-workers indeed found significant differences at different temperatures and almost no decomposition at their lowest temperatures in an accessible simulation time -50ps.

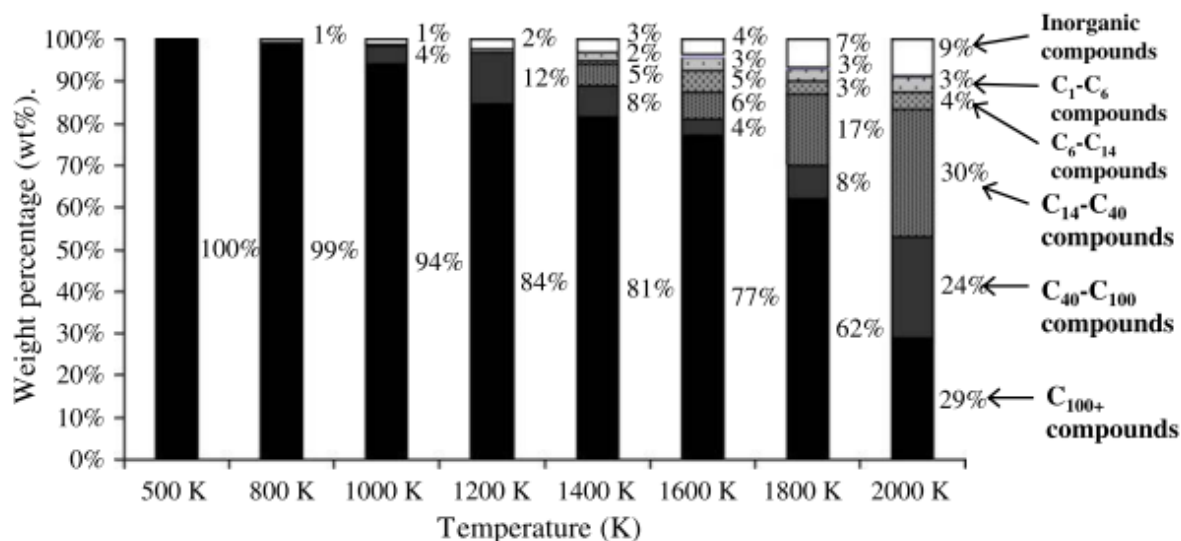


Figure 19. Final composition after 50ps of NVT simulations, at different temperatures, on a lignite geopolymer. From Salmon *et al.*⁵⁴

The vast majority of studies on organic matter decomposition by numerical simulation^{54,57-60} is done at high temperature since it is the easiest –and cheapest– way to accelerate a reaction which takes millions of years in real time. During the last decade, teams mostly used the reactive force field ReaxFF. An overview of this potential is given later in this manuscript –see Chapter2– but the main reason is that it allows reproducing chemical reactions. They all agree that ReaxFF is a “*promising and reliable method*”⁵⁹ for the study of organic matter pyrolysis.

Liu *et al.*⁵⁸ studied the pyrolysis of an oil shale kerogen model. They submitted it to a temperature between 800 and 2800K during molecular dynamics simulations, using the ReaxFF potential. They found a 2-stages decomposition: first a decomposition of kerogen into heavy compounds –with 40 carbon atoms or more– and second the generation of light compounds. They found a strong influence of temperature on the products types and their distributions.

Zhang *et al.*⁵⁹ also found an important dependency on temperature, especially in the formation of carbon monoxide, CO. They reached this conclusion when they studied the pyrolysis process of *oleic type triglyceride* at high temperature –1800 to 2800K– using

the ReaxFF potential. Triglycerides are found into fat –human, animal and vegetal- and are thus a clear component of organic matter.

At the beginning of 2016, Zhen *et al.*⁶⁰ published an article in which they exposed their study of initial reaction mechanisms of cellulose pyrolysis, also with the ReaxFF potential and at rather high temperature -500 to 2800K. To elucidate reactions paths they used a new tool –called VARxMD- which interprets simulation trajectories in order to provide reactional sites and reaction equation. They explained that “*ReaxFF force field [...] is promising for predicting the behavior and chemistry events in biomass pyrolysis*”.

The only study on the decomposition of organic matter by numerical simulation, in which use of high temperature is not the way to speed up the reaction, is the work of Agarwal *et al.*⁶¹. They worked on the initial decomposition of cellulose during pyrolysis and, instead of using high temperature, they accelerated rare events by metadynamics. Metadynamics enhances sampling in molecular dynamics simulation. It uses a bias potential which progressively grows and fills the free energy landscape already visited^{62,63}. This method prevents the structure to go back to visited configurations and, once the energy well is filled, the structure can jump to another well: jumping over the reaction barrier. A schematic view of metadynamics principle is given in Figure 20. The energy landscape is periodically filled with Gaussians.

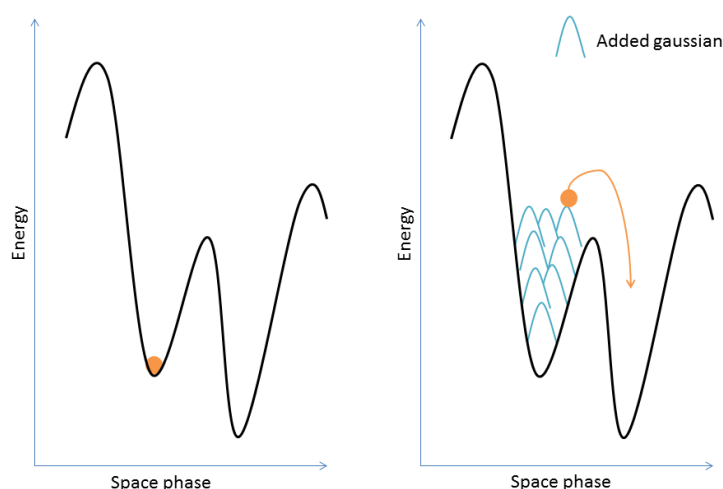


Figure 20. Schematic view of metadynamics principle.

Agarwal and co-workers used *ab-initio* simulation combined with metadynamics. *Ab initio* method uses quantum mechanics principles and is therefore the best suited approach to study reactivity but it has the significant disadvantage to be really expensive in resources and CPU time. For this reason, the computation is limited in size –only small systems of a hundred of atoms can be studied- and in simulation time. Agarwal thus used really small systems of 92 and 184 atoms and limited his work to the initial decomposition. Their work was accurate and they found pathways of the formation of the major pyrolysis products, as levoglucosan and formic acid. They made a

first step in a more realistic description of biomass pyrolysis but they do not seem to have been followed so far.

Challenged by this work and the potential of accelerated events methods, we tried to simulate the geological decomposition of organic matter on larger systems and time scales. That is what we did and what we explain in following sections.

Summary

Hydrocarbons found in subsurface reservoirs results from the decomposition of organic matter. This organic matter is the part of dead organisms that have escaped oxidation by being buried and mixed with sediments for millions of years. This dispersed sedimentary organic matter is mainly composed of carbon, hydrogen, and oxygen, with traces of nitrogen and sulfur.

Organic residues decompose into a carbonaceous porous solid, called kerogen and a fluid usually composed of hydrocarbons and other small molecules as water and carbon monoxide and dioxide. This decomposition process mainly consists of three stages: diagenesis, catagenesis, and metagenesis. During the degradation, organic matter first loses oxygen and then essentially hydrogen. During the second step, aromatic rings are formed and metagenesis is mainly a rearrangement stage.

Kerogen is an intermediate state of the organic matter, between organic residues and hydrocarbons. It has been defined by Durand as “*all insoluble sedimentary organic material*”. Along with burying, the kerogen is affected by a cracking process. During this process, that takes millions of years, the kerogen generates oil, gas and other molecules. What is formed depends on the stage of the kerogen transformation and on the original pristine precursor.

Three significant types of kerogens have been defined, depending on its origin, and thus on its initial composition.

- Type I kerogen derives from lacustrine depositions. Usually, at the beginning of the pyrolysis process, the matter has an H/C atomic ratio higher than 1.5 and a low oxygen content (O/C atomic ratio lower than 0.1). The samples are generally highly aliphatic.
- Type II kerogens usually come from the marine environment. The H/C and O/C atomic ratios are around 1.3 and 0.15 respectively.
- Type III kerogen has terrestrial origins, from higher plants, and gives rise to coal chars under further pyrolysis. The H/C atomic ratio, lower than 0.8, is lower than the ones of the two former types, and the O/C ratio, around 0.25, is higher.

Decomposition process has been seen as a pyrolysis of organic matter, taking place during millions of years. Pyrolysis is a non-selective thermal degradation method.

Attempts to simulate this natural process have been made, with different methods – open-system, closed and confined pyrolysis. Numerical simulations have also been tried. Kerogen structure was and still is, investigated, following computing power improvements, methods, potentials, and codes developments. Kerogen formation is also an interesting subject and a lot of research teams tried to numerically simulate it. A significant methodology question arises: how to accelerate the reaction without changing its equilibrium. The majority of those studies have been performed at much higher temperature than in real systems. Higher temperature displace system equilibrium and products they obtained were different than in natural samples. Numerical methods to speed up rare events exist and an attempt of application to organic matter decomposition is the subject of this work.

Three main documents to read in order to have a large scope on hydrocarbons formation and kerogen:

- B. Tissot and D. Welte, *Petroleum formation and occurrence*, 1978
- B. Durand, *Kerogen*, 1980
- M. Vandenbroucke and C. Largeau, *Kerogen origin, evolution, and structure*, 2007

1. Vandenbroucke, M. & Largeau, C. Kerogen origin, evolution and structure. *Org. Geochem.* **38**, 719–833 (2007).
2. Trager, Earl, A. *Kerogen and its relation to the origin of oil.* (1924).
3. Durand, B. *Kerogen: Insoluble organic matter from sedimentary rocks.* (Editions technip, 1980).
4. Tissot, B. P. & Welte, D. H. *Petroleum formation and occurrence.* (Springer Science & Business Media, 2013).
5. Durand, B. & Espitalie, J. Evolution de la matière organique au cours de l'enfouissement des sédiments. *Compte rendus l'Académie des Sci.* **276**, 2253–2256 (1973).
6. Rouxhet, P. G. & Robin, P. L. P. R. Infrared study of the evolution of kerogenes of different origins during catagenesis and pyrolysis. *Fuel* **57**, 533–540 (1978).
7. Philippi, G. T. On the depth, time and mechanism of petroleum generation. *Geochim. Cosmochim. Acta* **29**, 1021–1049 (1965).
8. Monthieux, M., Landais, P. & Durand, B. Comparison between extracts from natural and artificial maturation series of Mahakam delta coals. *Org. Geochem.* **10**, 299–311 (1985).
9. Sajgo, C., Horvath, Z. A., Szoor, G. & McEvoy, J. Influence of temperature and pressure on maturation processes - II. in *13th International Meeting on Organic Geochemistry* **10**, 163 (1987).
10. Freund, H., Clouse, J. a. & Otten, G. a. Effect of pressure on the kinetics of kerogen pyrolysis. *Energy & Fuels* **7**, 1088–1094 (1993).
11. Naturwissenschaften, D. Der. Experimental Investigation of Gas Storage Properties of Black Shales. (2013).
12. Falk, K., Coasne, B., Pellenq, R., Ulm, F.-J. & Bocquet, L. Subcontinuum mass transport of condensed hydrocarbons in nanoporous media. *Nat. Commun.* **6**, 6949 (2015).
13. Bousige, C. *et al.* Realistic molecular model of kerogen's nanostructure. *Nat. Mater.* **15**, 576–583 (2016).
14. Klemme, H. D. & Ulmishek, G. F. Effective petroleum source rocks of the world: stratigraphic distribution and controlling depositional factors. *Am. Assoc. Pet. Geol. Bull.* **75**, 1809–1851 (1991).
15. Atwater, G. I. petroleum - From kerogen to petroleum | Britannica.com. Available at: <https://www.britannica.com/science/petroleum/From-kerogen-to-petroleum#ref502600>. (Accessed: 18th August 2016)
16. Pedersen, J. H. Oil and Gas Geology. Available at: <http://oilandgasgeology.com/>. (Accessed: 18th August 2016)
17. Gonzalez, J. *et al.* Determination of formation organic carbon content using a new neutron-induced gamma ray spectroscopy service that directly measures carbon. (2013).
18. Cornford, C. *Source rocks and hydrocarbons of the North Sea. Petroleum Geology of the North Sea: Basic Concepts and Recent Advances: Fourth Edition* (1998). doi:10.1002/9781444313413.ch11
19. Mao, J. *et al.* Chemical and nanometer-scale structure of kerogen and its change during thermal maturation investigated by advanced solid-state ¹³C NMR spectroscopy. *Geochim. Cosmochim. Acta* **74**, 2110–2127 (2010).
20. Cao, X., Yang, J. & Mao, J. Characterization of kerogen using solid-state nuclear magnetic resonance spectroscopy: A review. *Int. J. Coal Geol.* **108**, 83–90 (2013).
21. Vandenbroucke, M. Kerogen: from Types to Models of Chemical Structure. *Oil Gas*

- Sci. Technol. – Rev. IFP* **58**, 243–269 (2003).
22. Maciel, G. E., Bartuska, J. & Miknist, F. P. Characterization of organic material in coal by ¹³C nuclear magnetic resonance with magic-angle spinning. **58**, 391–394 (1979).
 23. Hatcher, P. G., VanderHart, D. L. & Earl, W. Use of Solid-State ¹³C-NMR in Structural Studies of Humic Acids and Humin From Holocene Sediments. *Org. Geochem.* **2**, 87–92 (1980).
 24. Resing, H. A., Garroway, A. N. & Hazlett, R. N. Determination of aromatic hydrocarbon fraction in oil shale by ¹³C n.m.r. with magic-angle spinning. *Fuel* **57**, 450–454 (1978).
 25. Maciel, G. E. & Dennis, L. W. Comparison of oil shales and kerogen concentrates by ¹³C nuclear magnetic resonance. *Org. Geochem.* **3**, 105–109 (1981).
 26. Wei, Z., Gao, X., Zhang, D. & Da, J. Assessment of Thermal Evolution of Kerogen Geopolymers with Their Structural Parameters Measured by Solid-State ¹³C NMR Spectroscopy. *Energy & Fuels* **19**, 240–250 (2004).
 27. Tissot, B. P., Pelet, R. & Ungerer, P. Thermal History of Sedimentary Basins, Maturation Indices, and Kinetics of Oil and Gas Generation. *Am. Assoc. Pet. Geol. Bull.* **71**, 1445–1466 (1987).
 28. Requejo, A. G. *et al.* Maturation of petroleum source rocks. 1. Changes in kerogen structure and composition associated with hydrocarbon generation. *Energy & Fuels* **6**, 203–214 (1992).
 29. Prudhon, G., Jacquesson, F., Lete, J. & Paris, S. Combustibles solides . Charbon Origine , constitution , classification. *Tech. l'Ingénieur* **33**, 531 (2005).
 30. Lewan, M. D., Winters, J. C. & McDonald, J. H. Generation of oil-like pyrolyzates from organic-rich chales. *Scien* **203**, 1–4 (1978).
 31. Lewan, M. D. *et al.* Evaluation of Petroleum Generation by Hydrous Pyrolysis Experimentation [and Discussion]. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **315**, 123–134 (1985).
 32. Monthioux, M., Landais, P. & Monin, J.-C. Comparison between natural and artificial maturation series of humic coals from the Mahakam delta, Indonesia. *Org. Geochem.* **8**, 275–292 (1985).
 33. Goodarzi, F. & Murchison, D. G. Optical properties of carbonized vitrinites. **51**, 322–328 (1972).
 34. Bourrat, X. Sulphur behaviour treatment during asphalt. **66**, 542–550 (1987).
 35. Monthioux, M. & Landais, P. Natural and artificial maturation of coal: Non-hopanoid biomarkers. *Chem. Geol.* **77**, 71–85 (1989).
 36. Behar, F., Vandenbroucke, M., Tang, Y., Marquis, F. & Espitalie, J. Thermal cracking of kerogen in open and closed systems: Determination of kinetic parameters and stoichiometric coefficients for oil and gas generation. *Org. Geochem.* **26**, 321–339 (1997).
 37. Landais, P. & Monthioux, M. Closed system pyrolysis: an efficient technique for simulating natural coal maturation. *Fuel Process. Technol.* **20**, 123–132 (1988).
 38. Landais, P., Michels, R. & Elie, M. Are time and temperature the only constraints to the simulation of organic matter maturation? *Org. Geochem* **22**, 146–6380 (1993).
 39. Monthioux, M. Expected mechanisms in nature and in confined-system pyrolysis. *Fuel* **67**, 843–847 (1988).
 40. Behar, F. & Hatcher, P. G. Artificial Coalification of a Fossil Wood from Brown Coal by Confined System Pyrolysis. 984–994 (1996).
 41. Fuchs, W. & Sandhoff, A. Theory of coal pyrolysis. *Ind. Eng. Chem.* 567–571 (1942).

42. Mathews, J. P. & Chaffee, A. L. The molecular representations of coal - A review. *Fuel* **96**, 1–14 (2012).
43. Spiro, C. L. Space-filling models for coal: a molecular description of coal plasticity. *Fuel* **60**, 1121–1126 (1981).
44. Hatcher, P. G., Faulon, J. L., Wenzel, K. a & Cody, G. D. A structural model for lignin-derived vitrinite from high-volatile bituminous coal (coalified wood). *Energy & Fuels* **6**, 813–820 (1992).
45. Shinn, J. H. Visualization of complex hydrocarbon reaction systems. *Prepr. Pap. Am. Chem. Soc. Div. Fuel Chem.* **41**, (1996).
46. Forsman, J. P. in *Organic Geochemistry* 178–182 (1963).
47. Burlingame, A. L., Haug, P. A., Schnoes, H. K. & Simoneit, B. R. Fatty Acids Derived from the Green River Formation Oil Shale by Extractions and Oxidations) — A Review. *Adv. Org. geochemistry* 85–129 (1968). doi:10.1016/B978-0-08-006628-8.50008-3
48. Yen, T. F. in *Oil Shale, Developments in Petroleum Science* **5**, 129–148 (1976).
49. Siskin, M., Scouten, C. G., Rose, K. D., Aczel, T. & Pabst, R. E. Detailed structural characterization of the organic material in rundle Ramsay Crossing and Green River oil shales. *Compos. Geochemistry Convers. oil shales* 143–158 (1995).
50. Behar, F. & Vandenbroucke, M. Chemical modelling of kerogens. *Org. Geochem.* **11**, 15–24 (1987).
51. Oka, M., Chang, H. & Gavalas, G. R. Computer-Assisted Molecular Structure Construction for coal-derived compounds. **56**, 3–8 (1977).
52. Ungerer, P., Collell, J. & Yiannourakou, M. Molecular Modeling of the Volumetric and Thermodynamic Properties of Kerogen: In fl uence of Organic Type and Maturity. *Energy Fuels* **29**, 91–105 (2014).
53. Medea | Materials Design. Available at: <http://www.materialsdesign.com/medea>. (Accessed: 27th September 2016)
54. Salmon, E., van Duin, A. C. T., Lorant, F., Marquaire, P.-M. & Goddard, W. a. Early maturation processes in coal. Part 2: Reactive dynamics simulations using the ReaxFF reactive force field on Morwell Brown coal structures. *Org. Geochem.* **40**, 1195–1209 (2009).
55. Broido, A. & Nelson, M. A. Char yield on pyrolysis of cellulose. *Combust. Flame* **24**, 263–268 (1975).
56. Varhegyi, G., Jakab, E. & Antal, M. J. Is the Broido-Shafizadeh Model for Cellulose Pyrolysis True? *Energy & Fuels* **8**, 1345–1352 (1994).
57. Zou, C., Raman, S. & Van Duin, A. C. T. Large-scale reactive molecular dynamics simulation and kinetic modeling of high-temperature pyrolysis of the gloecapsomorphaprisca microfossils. *J. Phys. Chem. B* **118**, 6302–6315 (2014).
58. Liu, X. X. *et al.* Initial pyrolysis mechanism of oil shale kerogen with reactive molecular dynamics simulation. *Energy and Fuels* **29**, 2987–2997 (2015).
59. Zhang, Y., Wang, X., Li, Q., Yang, R. & Li, C. A ReaxFF Molecular Dynamics Study of the Pyrolysis Mechanism of Oleic-type Triglycerides. *Energy and Fuels* **29**, 5056–5068 (2015).
60. Zheng, M. *et al.* Initial reaction mechanisms of cellulose pyrolysis revealed by ReaxFF molecular dynamics. *Fuel* **177**, 130–141 (2016).
61. Agarwal, V., Dauenhauer, P. J., Huber, G. W. & Auerbach, S. M. Ab initio dynamics of cellulose pyrolysis: Nascent decomposition pathways at 327 and 600 °C. *J. Am. Chem. Soc.* **134**, 14958–14972 (2012).
62. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.*

- 99**, 12562–6 (2002).
63. Michel, C. Etudes théoriques de la réactivité chimique : des méthodes statiques à la métadynamique. (2007).
 64. Formation du pétrole. Available at:
<http://www.connaissancedesenergies.org/fiche-pedagogique/formation-du-petrole>. (Accessed: 1st January 2016)
 65. Presses Hautes Pressions. Available at:
http://www.imPMC.upmc.fr/fr/equipements/plate_forme_hautes_pressions/pres-ses_hautes_pressions.html. (Accessed: 7th November 2016)

Chapter 2: Method

In research, there are two ways to study processes. We can do physical experimentations or numerical ones. The first approach uses real samples with experimental tools while the second one uses numerical models and computers. Often experimentations are defined as practical, giving real information and numerical studies are seen as theoretical. Everything happening in real is a response to a physical rule, even if we do not know it yet. Numerical experimentation can only be based on rules we know, understand and that we are able to formulate in an algorithm. Models are used in science to simplify the reality. Numerical simulations can be used in order to verify these models. Results of such work can be compared with real samples and experimental results and, if needed, help to improve the model.

In the same time, some characteristics or behaviors are unreachable –or it is really complicated- or dangerous to reach by experimentation and numerical simulations can help. Moreover, the majority of laws of nature is not easily exactly solvable in an analytically way. Without numerical simulations, predictions on the behavior of a system are really limited –to well-known and simple models- and are usually inaccurate because based on approximations. Numerical simulations can also be used to try different conceivable mechanisms and help to choose the good one. Different theories can indeed be tested and compared to experiment. If our theoretical mechanism and experiment disagree, it may imply a wrong theory.

We saw in the previous section that organic matter decomposition into kerogen and hydrocarbons takes millions of years. The best experimental method people have found to reproduce this process so far, is confined pyrolysis. Numerical simulations allow testing different configurations of temperatures and pressures, and new methods to span long time scales are developed.

In this section, we present the methods we used. First, we give a description of the interaction potential we use. We then give an overview of different numerical simulation techniques. In a third part, we focus on the *Replica Exchange Molecular Dynamics* technique that is a method allowing to speed-up slow processes. We end by a short presentation of the code that we have used and by a summary of this whole section. Note that from now on, we will use the term simulation only for numerical simulation –not experimental ones.

Interaction potential

A potential of interaction describes the way atoms –or other particles- interact with each other. It provides the forces that are applied to those atoms. Its precision will change the accuracy of the simulation and is the main limitation of the method. Note that both denominations force field and potential of interaction can be used.

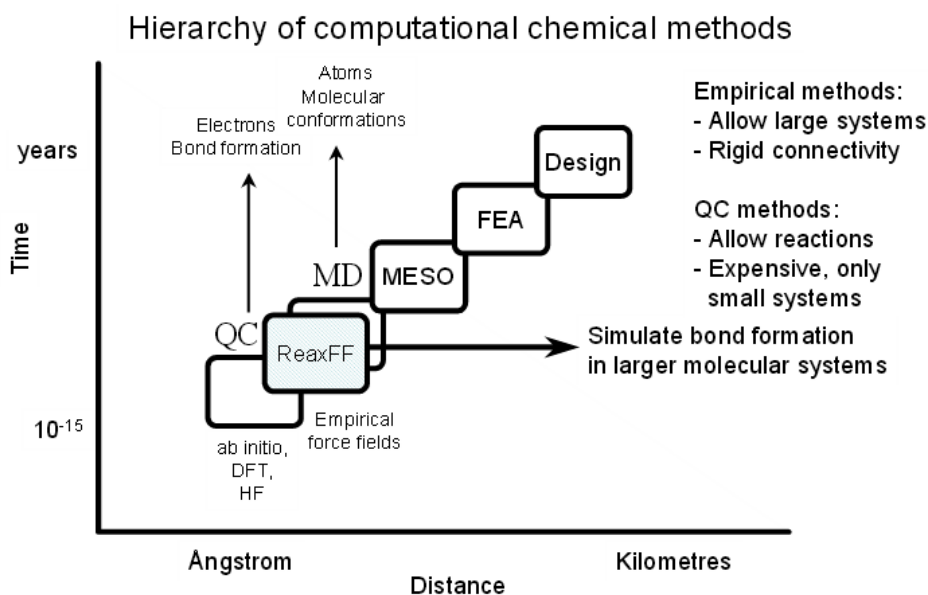


Figure 21. ReaxFF position into computational methods¹

In this work, we use the ReaxFF potential, that is designed to bridge the gap between empirical and quantum methods, as illustrated in Figure 21. As one can see, ReaxFF is positioned between both methods and provides more accurate results than classical molecular dynamics while allowing larger systems than fully quantum mechanically based method. It is able to follow individual chemical reactions, although some reactions are sometimes incorrectly reproduced^{vii}.

We first describe principle and role of potential of interaction before focusing on ReaxFF more in details.

^{vii} The water gas shift reaction, for instance, is incorrectly reproduced. See *Chapter 4: Results* for more information on it.

Principle

The correct way to calculate atomic interactions is to solve the Schrödinger equation of quantum mechanics:

$$H\psi(r) = E\psi(r)$$

where H is the Hamiltonian operator, $\psi(r, t)$ is the wave function, and E is the system energy. This equation defines interactions between charged particles and its solving is highly time consuming and is the base of the *ab initio* method. It can only be applied to few hundreds of atoms systems and on limited simulation time. In order to go further in investigation, classical mechanics is applied, in which atoms are considered as mass points and electrons are not taken into account. In such a way, interactions have to be modeled by the means of force fields.

Force fields contain two components: the intra-molecular potential and the inter-molecular one. The intra-molecular potential describes interactions between atoms belonging to the same molecule while the inter-molecular potential describes interactions between atoms belonging to different molecules. A schematic of intra-molecular interactions is given in Figure 22.

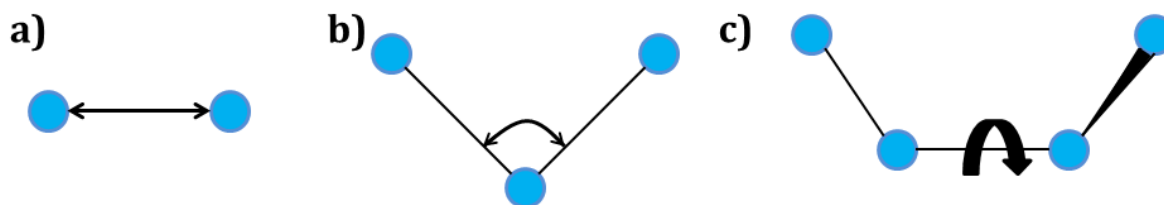


Figure 22. Schematic of intra molecular interactions: **a)** stretching, **b)** bending, and **c)** torsion.

Bond stretching U_s can be related to a spring with a spring constant k and corresponds to a covalent bond. It can be expressed as it follows:

$$U_s = \frac{1}{2} k \cdot r_{12}^2$$

Where r_{12} is the distance between atoms 1 and 2. Bond bending U_B corresponds to angle of valence variation and is expressed as:

$$U_B = \frac{1}{2} k \cdot (\theta_{123} - \theta_{123}^0)^2$$

Where θ_{123} is the angle of valence made by three atoms 1, 2, and 3. Torsion U_T around the covalent bond is expressed as it follows:

$$U_T = \sum_{na} A \cdot \cos(\omega)^{na}$$

Where n_a is the number of dihedral angles, A is the magnitude and ω is the torsion angle.

Concerning inter molecular interactions –or non-bonded interactions- there are two types: electrostatic and Van der Waals. Van der Waals term can be described in a first approximation by a Lennard Jones potential –see below- and the electrostatic U_e one can be expressed between atoms i and j as:

$$U_e = \sum_{i < j} \frac{q_i \cdot q_j}{4 \cdot \pi \cdot \epsilon_0 \cdot r_{ij}}$$

Where ϵ_0 is the dielectric constant and q_i and q_j are partial atomic charges of atoms i and j .

Interaction potentials are usually empirical, it is to say they have a simple and numerically convenient analytical form, sometimes based on an approximate theory, and their parameters are fitted against experimental data. They have the advantage of allowing simulations on large systems –systems composed of thousands of atoms.

The easiest way to understand how a potential works is to take a look at the Lennard-Jones one. In 1931 Lennard-Jones published a paper entitled *Cohesion*², in which he proposed potential energy as a function of interatomic distance for different atoms interactions as He, H or N. A plot of this potential is given in Figure 23.

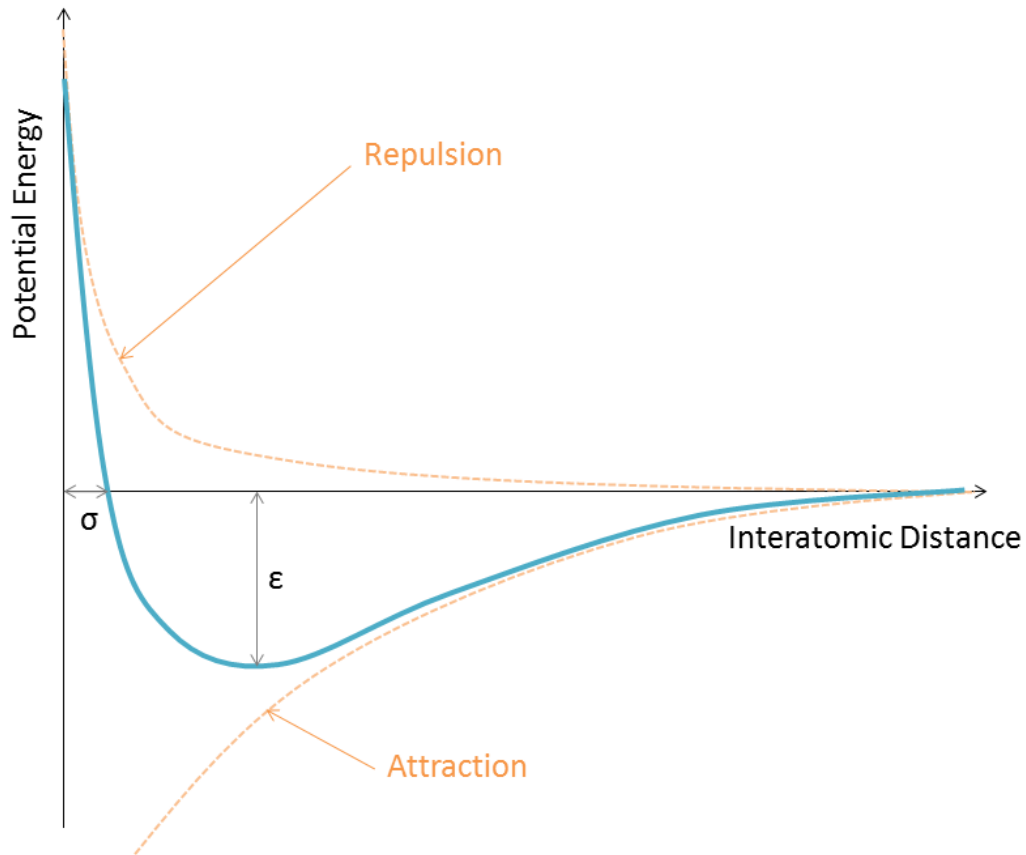


Figure 23. Vision of the Lennard-Jones potential. Sum of a repulsive and an attractive interactions gives the interaction potential.

Van der Waals interactions result of two main interactions: attraction and repulsion. The last one –the repulsive interaction- includes different contributions, among which the repulsion between species with the same electric charge, as well as quantum mechanical contributions (Pauli...) and decreases rapidly as the interatomic distance increases. Attractive interaction results from the dispersion interaction. It is called the London attraction. The Lennard-Jones potential U can be written simply as:

$$U(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right]$$

Where ϵ and σ depend on the system of interest, σ being the distance at which the potential is null ϵ is the value of the potential energy at its minimum, it means at the equilibrium distance. In the equation, the term in r^{-12} represents the repulsive part and the term in r^{-6} represents the attractive one.

Reactive potential : ReaxFF

In 2001, Van Duin *et al.*³ published an article in which they proposed a new force field for reactive systems. The main particularity of this new potential was its ability to properly reproduce bond breaking and forming, in the sense that it describes reaction intermediates.

Besides Brenner's⁴ potential, all force fields were empirically determined and well suited for system simulations with nonreactive interactions. Chemical reactions cannot be described by such formalisms. Van Duin and coworkers then built a potential in the perspective to bridge the gap between empirical interatomic potentials and quantum methods.

Definition

Van Duin's potential is called ReaxFF, for reactive force field and is based on a bond-order formalism. It is a reactive potential since it describes both reactive and non-reactive interactions between atoms. It takes into account several energy terms to compute system energy, as shown by the following equation.

$$U_{system} = U_{bond} + U_{over} + U_{under} + U_{val} + U_{pen} + U_{tors} + U_{conj} + U_{VdW} + U_{Coulomb}$$

Where U_{system} is the system energy, U_{bond} is the energy value of forming atomic bonds and is a function of interatomic distance, U_{over} and U_{under} are corrections to avoid over and under coordination that could occur with the U_{bond} application. U_{val} corresponds to the energy due to the valence angle. This term has to vanish smoothly while bond dissociation. U_{pen} is an energy penalty imposed when the system has two double bonds sharing an atom and is defined to guarantee stability of such systems. U_{tors} corresponds to the energy due to torsion angles. U_{conj} describes the conjugation effect when there are successive bonds. U_{VdW} and $U_{Coulomb}$ complete the U_{bond} by adding attractive interaction for long distances –due to dispersion- and repulsive energy at short distances. Those two contributions are applied to all atom pairs. ReaxFF also includes atomic charges calculations, using the electronegativity equalization method.

ReaxFF is based on a bond-order formalism it means that the 1st, 2nd, 3rd... neighbors are part of the description of bonds. It is obtained directly from interatomic distance and is composed of three terms –for sigma, pi, and double pi bonds^{viii}- as shown by the following equation:

^{viii} There are two kinds of bonds: sigma and pi. Sigma bonds are simple ones and pi bonds are multiple ones which appear as a complement of sigma ones. Pi bonds are weaker than sigma ones but combination

$$BO_{ij} = \exp \left[p_{bo,1} \cdot \left(\frac{r_{ij}^\sigma}{r_o} \right)^{p_{bo,2}} \right] + \exp \left[p_{bo,3} \cdot \left(\frac{r_{ij}^\pi}{r_o} \right)^{p_{bo,4}} \right] + \exp \left[p_{bo,5} \cdot \left(\frac{r_{ij}^{\pi\pi}}{r_o} \right)^{p_{bo,6}} \right]$$

Where BO_{ij} is the bond order between atoms i and j , p_{bo} are calibrated parameters, r_{ij} the distance between atoms i and j and r_o the equilibrium bond length. Bond order is illustrated in Figure 24, for carbon bonds. Contributions of the three terms are visible. The term of the sigma bond is unity for distances under 1.5Å and is null above 2.5Å. The second component –for first pi bond– is equal to one under 1.2Å and to zero above 1.75Å. The last term –for second pi bond– is unity until 1.0Å and null above 1.4Å. Bond orders are corrected to avoid overcoordinations.

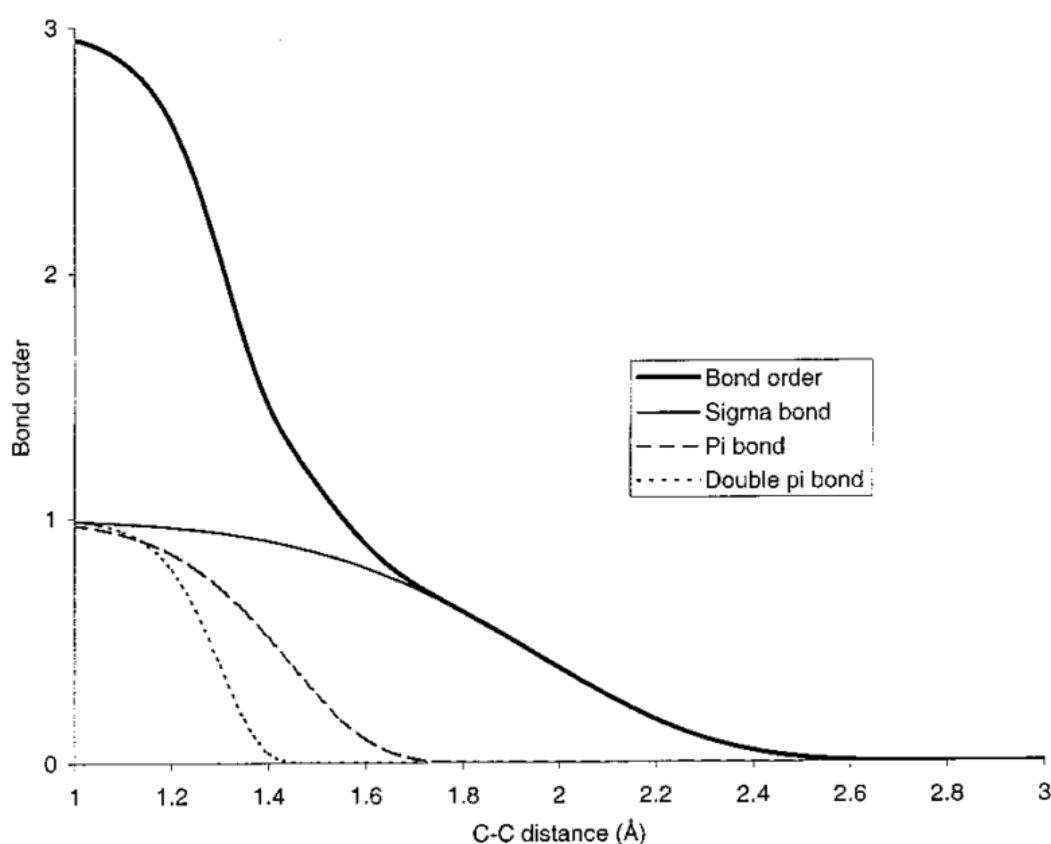


Figure 24. Bond order for a C-C bond, as a function of interatomic distance³

ReaxFF has been parametrized based on both experimental –especially for reaction enthalpies– and quantum calculations. As in quantum mechanics, there is no need to predefine a reaction site, thus simulation is not biased. Moreover, an atom A is treated with the same formalism whether it is in a liquid, solid or gaseous phase. This

of both is stronger than a lonely sigma bond. For example, a triple bond will consist of one sigma and two pi bonds.

transferability allows ReaxFF to model multi-phases systems and complex processes related to it.

Another significant point is that ReaxFF is developed for many different components –almost 50 species, see Figure 25. Obviously, adaptation is needed depending on the studied system and different versions of ReaxFF have been published, following research needs⁵⁻⁷. This reactive potential is *de facto* used in a large variety of research fields.

Energy terms are described in the first ReaxFF paper³ as it follows:

$$U_{bond} = -D_e \cdot BO_{ij} \cdot \exp[p_{be,1} \cdot (1 - BO_{ij})]$$

Where BO_{ij} is the corrected bond order between atoms i and j –see below.

$$U_{over} = p_{over} \cdot \Delta_i \cdot \left(\frac{1}{1 + \exp(\lambda_6 \cdot \Delta_i)} \right)$$

Which imposes energy penalty on the system when the atom is overcoordinated ($\Delta_i > 0$). It also ensures that U_{over} quickly goes to zero for undercoordinated systems ($\Delta_i < 0$).

$$U_{under} = -p_{under} \cdot \frac{1 - \exp(\lambda_7 \cdot \Delta_i)}{1 + \exp(-\lambda_8 \cdot \Delta_i)} \cdot f_6 \cdot (BO_{ij,\pi}, \Delta_j)$$

$$U_{val} = f_7 \cdot (BO_{ij}) \cdot f_7 \cdot (BO_{jk}) \cdot f_8 \cdot (\Delta_j) \cdot \{k_a - k_a \cdot \exp[-k_b (\theta_0 - \theta_{ijk})^2]\}$$

All the functions f_{7-13} are described in the paper. Δ_j is used to take into account over and under coordination of the central atom. θ_{ijk} is the valence angle and θ_0 is its equilibrium value.

$$U_{pen} = \lambda_{19} \cdot f_9(\Delta_j) \cdot \exp[-\lambda_{20} \cdot (BO_{ij} - 2)^2] \cdot \exp[-\lambda_{20} \cdot (BO_{ij} - 2)^2]$$

$$U_{tors} = f_{10} \cdot (BO_{ij}, BO_{jk}, BO_{kl}) \cdot \sin(\theta_{ijk}) \cdot \sin(\theta_{jkl}) \cdot \left[\frac{1}{2} \cdot V_2 \cdot \exp\{p_l (BO_{jk} - 3 + f_{11} \cdot (\Delta_j, \Delta_k))^2\} \cdot (1 - \cos(2 \cdot \omega_{ijkl})) + \frac{1}{2} \cdot V_3 \cdot (1 + \cos(3 \cdot \omega_{ijkl})) \right]$$

Where ω_{ijkl} is the torsion angle.

$$U_{conj} = f_{12} \cdot (BO_{ij}, BO_{jk}, BO_{kl}) \cdot \lambda_{26} \cdot [1 + (\cos^2(\omega_{ijkl}) - 1) \cdot \sin(\theta_{ijk}) \cdot \sin(\theta_{jkl})]$$

$$U_{vdW} = D_{ij} \cdot \left\{ \exp \left[\alpha_{ij} \cdot \left(1 - \frac{f_{13}(r_{ij})}{r_{vdW}} \right) \right] - 2 \cdot \exp \left[\frac{1}{2} \cdot \alpha_{ij} \cdot \left(1 - \frac{f_{13}(r_{ij})}{r_{vdW}} \right) \right] \right\}$$

$$U_{Coulomb} = C \cdot \frac{q_i \cdot q_j}{\left[r_{ij}^3 + (1/\gamma_{ij})^3 \right]^{1/3}}$$

All D_{a-z} , p_x , Δ_x , λ_{6-26} , k_{a-b} , V_{2-3} , Θ_x , ω_x , q_x , C , and γ are calibrated parameters for different couples of atoms and can be found in the paper³. They have been optimized using a successive one-parameter search technique⁸. The goal is to be able to reproduce a set of experimental and/or theoretical data⁵. ReaxFF has been parametrized using a quantum chemistry training set, trying to reproduce data such as heats of formation, and bond angles and lengths³. Parametrization of the potential needs to be reassessed for every new practical application. Therefore, a number of ReaxFF parametrizations exist, depending on the need, such as Chenoweth's⁵ one, designed for simulation of hydrocarbon oxidation, Pitman's⁷ one, designed to study clay-zeolite composite systems, or Mattsson's⁶ one, designed for simulation of shocked polymers.

Applications

The first version of ReaxFF has been developed for hydrocarbons thermal reactivity study³. Since this first attempt, several chemical elements have been added to the available species with ReaxFF, both metal and non-metal.

ReaxFF has been used both in physical processes study and biology. It has been applied to hydrocarbon study^{3,5}, polymer thermal decomposition⁹, nitramines¹⁰ and atomic layer deposition¹¹.

As we highlighted in the first chapter, the ReaxFF potential has also been widely used in organic matter decomposition studies. Indeed, numerous studies of organic matter thermal degradation have been published such as Salmon *et al.*¹², Zou *et al.*¹³ or Liu *et al.*¹⁴. All concluded on a reliability of ReaxFF for studies on organic compounds thermal decomposition.

Thermodynamic Ensembles

Simulation can be performed in different thermodynamic ensembles, imposing external environment constraints to the system. Four main statistical ensembles exist and are described in the following sections: NVE, NVT, NPT, and μ VT.

For each ensemble, at least 3 constants constraints are defined. One has to be extensive to fix system size. Calculated variables are then denominated as conjugate variables of these constraints. They fluctuate. These fluctuations, for an average value of a property A , are defined equal to $\langle A^2 \rangle - \langle A \rangle^2$.

NVE: micro canonical ensemble

Micro canonical ensemble -abbreviated NVE- is a ensemble in which number N of atoms, volume V , and system energy E are constant. It does not correspond to natural conditions of experiments. Conjugate variables are chemical potential μ , pressure P and temperature T . Newtonian equations maintain total energy and are then considered as the natural method to generate equilibrium states in this ensemble. Therefore, micro canonical ensemble is the natural one of Molecular Dynamics method –see below.

System in micro canonical ensemble is considered as isolated, it means a closed system with no exchange of heat, work, or mater with external environment. Let us consider $\Omega(E)$ as the number of possible states of the system –at the given energy E . In statistical mechanics, for an isolated system at equilibrium, accessible states have the same probability. Therefore, probability W for system to be in a state i of energy E_i –with E_i really close to E - is:

$$W = \frac{1}{\Omega(E)}$$

NVT: canonical ensemble

In canonical ensemble –NVT- constant constraints are the number N of atoms, the volume V and the temperature T and their conjugate variables are respectively the chemical potential μ , the pressure P , and the energy E . A system in the canonical ensemble is considered as closed. It means it does not exchange matter or work but heat with its external environment. Probability W for a system to be in a state i of energy E_i is:

$$W(E_i) = \frac{\exp(-\beta \cdot E_i)}{\sum_i \exp(-\beta \cdot E_i)}$$

Where $\beta=1/(k_B \cdot T)$, with k_B the Boltzmann constant, and the denominator is the partition function.

NPT: isotherm-isobaric ensemble

In isotherm-isobaric ensemble, number N of atoms, pressure P , and temperature T are constant while their conjugate variables are chemical potential μ , volume V , and energy E , respectively. Probability W for a system to be in a state i of energy E_i with a volume V is:

$$W(E_i, V) = \frac{\exp(-\beta(E_i + P.V))}{\sum_i \exp(-\beta(E_i + P.V))}$$

μVT : grand-canonical ensemble

In grand-canonical ensemble, chemical potential μ , volume V and temperature T are constant and have for conjugate variables number N of atoms, pressure P and energy E . Probability W for a system of N atoms in this ensemble to be in a state i of energy E_i is:

$$W(E_i, N) = \frac{\exp(-\beta.E_i). \exp(\mu. \beta. N)}{\sum_N \sum_j \exp(-\beta.E_j). \exp(\mu. \beta. N)}$$

Some molecular dynamics based algorithms have been proposed to sample this ensemble, but the most efficient choice is to use Monte-Carlo simulations, in the so called GCMC²¹ –for Grand Canonical Monte Carlo.

Monte-Carlo

The Monte-Carlo method is used in a wide range of fields as biology, physics, finance, humanities²² science or telecommunications²³. It is based on statistical sampling techniques adapted to every statistical ensemble.

We can trace back the origin of the MC method to the seminal work of Georges Louis Leclerc, Earl of Buffon, in 1733, with his “*Mémoire sur le jeu de Franc-Carreau*”. It was then better formulated in 1777. He presented his work to the Royal Academy of Sciences in Paris in which, he threw a needle of width l on a wooden floor –slats width was d with $l \leq d$. The goal was to find the probability that the needle crossed a floor line and thanks to that, to find an approximation of the value of π . He threw his needle several times and found a probability p as:

$$p=2.l/\pi.d$$

The more throws he did, the more accurate his value of p –and so the value of π – was. Results were based on attempts and repetition. In Figure 26 is given a schematic view of Buffon's experiment.

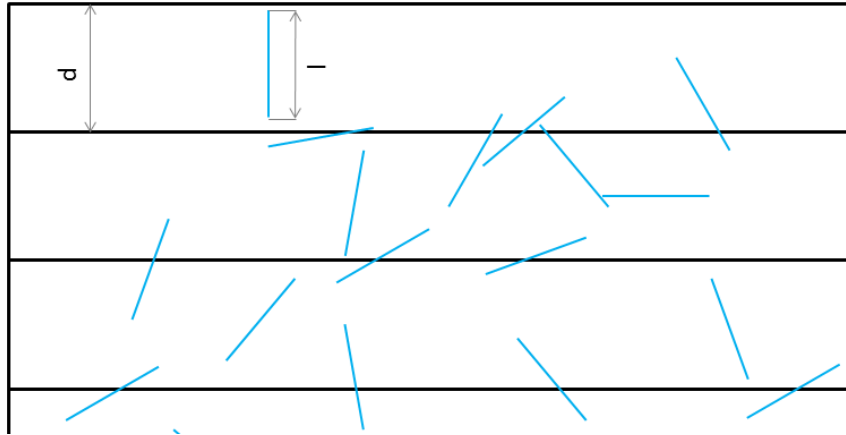


Figure 26. Buffon's needle experiment. In blue the needles and in black the wooden floor.

In 1949 Metropolis and Ulam²⁴ proposed a statistical approach to study “*integrodifferential*” equations. This method is called Monte-Carlo due to the love of Ulam for poker. They took the example of Solitaire game and noted that the best way to calculate the probability of success was to produce a large number of tries and to deduce the “*relative proportion of successes*”. As we noticed in Buffon's needle experiment, the more attempts are done, the more accurate the measured value.

In 1953 Metropolis *et al.*²⁵ proposed the Metropolis algorithm which laid the foundations of contemporary Monte-Carlo method. They applied it to a two-dimensions rigid-sphere system and found it as a reliable method for statistical mechanical problems that were “*not analytically soluble yet*”.

In basic Monte-Carlo method, sampling is homogeneous, hence the whole space of configurations is visited: this is the ergodicity hypothesis. Convergence of a basic Monte-Carlo sampling is rather slow, since the number of considered configurations is large. Average value of a quantity of interest A in the canonical ensemble is given by:

$$\langle A \rangle = \int W(x). A(x). dx$$

Where $A(x)$ is the value of a given property A at the position x in the phase space and $W(x)$ the probability distribution of A –see above.

In order to simplify sampling and to have a better convergence, one can generate configurations according to the probability distribution $W(x)$. For this purpose, a

Markov chain is used to generate probable configurations series. In such a case, average value of A can be computed as it follows:

$$\langle A \rangle = \frac{1}{n} \cdot \sum_{i=1}^n A(x_i)$$

Where x_i are generated configurations. It is the importance sampling method, the principle being to weight more probable configurations over the others.

A new configuration is generated from the previous state to which a random perturbation is applied. Microreversibility –or detailed balance– implies that the transition $i \rightarrow j$ is reversible. It writes:

$$W(i) \cdot pt(i \rightarrow j) = W(j) \cdot pt(j \rightarrow i)$$

Where $pt(j \rightarrow i)$ is the transition probability from i to j .

In Metropolis algorithm, the first step is to generate a trial move from state i to state j . This step is then accepted or rejected with the probability

$$\frac{W(j)}{W(i)} = \exp(-\beta \cdot (U_j - U_i))$$

Where U_i is the potential energy of state i and $\beta = 1/(k_B \cdot T)$, k_B being the Boltzmann constant and T the temperature.

Metropolis criterion is then the following:

$$pt(i \rightarrow j) = \begin{cases} \frac{W(j)}{W(i)} & \text{if } W(j) < W(i) \\ 1 & \text{if } W(j) \geq W(i) \end{cases}$$

Metropolis algorithm can be synthetized in 4 main steps:

- i) Computation of potential energy U_i of state i ,
- ii) Generation of a configuration $j=i+1$. The easiest way is to proceed to the displacement of a random atom,
- iii) Computation of potential energy U_j of new state j ,
- iv) Acceptance of the displacement with a probability of $pt(i \rightarrow j) = \min[1, \exp\{-\beta(U_j - U_i)\}]$.

Note that the way one obtains trajectories with Monte-Carlo method is not physical, MC steps do not follow the time line. Therefore Monte-Carlo method is not used to study kinetics behaviors. Other of Monte-Carlo methods exist –in all statistical ensembles such as Grand Canonical (μ, V, T), and isothermal and isobaric (N, P, T)- but are not developed here.

Classical Molecular Dynamics

Since the 1950s, Molecular Dynamics (**MD**) method is used to simulate the time evolution of systems under certain conditions of interest. It is a determinist method, based on classical mechanics and Newton equations of motion. Computer power has increased so much during the last 50 years that we are today able to compute evolution of systems at the atomic level and improvement is still going on.

In this work, we used classical MD –combined into a method called *Replica Exchange* explained in a following section. We first explain its principle and then give an overview of input information required. Note that Molecular Dynamics can be applied in *ab initio*, using quantum mechanics to compute forces but we have chosen here only the classical ones based on empirical potentials.

Principle

Principle of MD simulations is to use equations of motion of Newtonian mechanics in order to deduce trajectories of system of interest components –usually atoms. Indeed, as its name suggests, Molecular Dynamics provides time evolution and, on the contrary of Monte Carlo methods, for example, gives information on the system dynamics. Information such as diffusion coefficients is then accessible.

Numerical simulations enable to calculate average values representative of the sampling. Ergodicity principle states that an ensemble average is equal to the temporal one. During an MD run, any average value of a quantity A is expressed as:

$$\langle A \rangle_{time} = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \cdot \int_0^{\tau} A(\Gamma(t)) \cdot dt$$

Where $A(\Gamma(t))$ is the value of A at the point Γ of the phase space, explored at the time t , and τ is the time length of the trajectory. Equations of motion are stepwise integrated, the above equation is usually replaced by a discrete sum.

In classical MD, atoms are considered as whole objects, without dissociating electrons from nuclei –as in *ab initio*. These objects can be in movement or not and they follow Newtonian classical mechanics laws. By applying equations of motion, the configuration at time $t+\Delta t$ can be deduced from the one at time t , the aim being to minimize the system thermodynamic potential in order to reach the most favorable state in the given conditions.

Interaction Potential

As explained above, potential of interaction describes the way atoms –or other particles- interact with each other. It provides the energy and forces (as $F=-\nabla U$) which are applied to those atoms. It is a key element of a numerical simulation.

Integration

Integration in time is done using classical Newtonian mechanic. Molecular dynamics simulations consist in resolving Newton equations of motion, that state that the sum of forces applied to a system is equal to the system mass times its acceleration :

$$\sum \vec{F}_i = m_i \cdot \vec{a}_i = m_i \cdot \frac{d^2 \vec{r}_i}{dt_i^2}$$

Where F_i is the force applied to each atom i , m_i is its mass, a_i is its acceleration, r_i is its position, at time t . This resolution has to be done for every atom i in the system.

By default, the mass m is known since the system and its composition are known as an initial input. Every force derives from an interaction potential $U(r)$ as:

$$\vec{F}(r) = -\nabla U = -\frac{dU}{dr}$$

By integration of the acceleration, velocities and positions of particles are deduced.

Equations of motion are stepwise integrated. Integration time step dt has to be small enough in order to resolve the equation with the wanted accuracy. A schematic view of the basic way of integration is given in Figure 27.

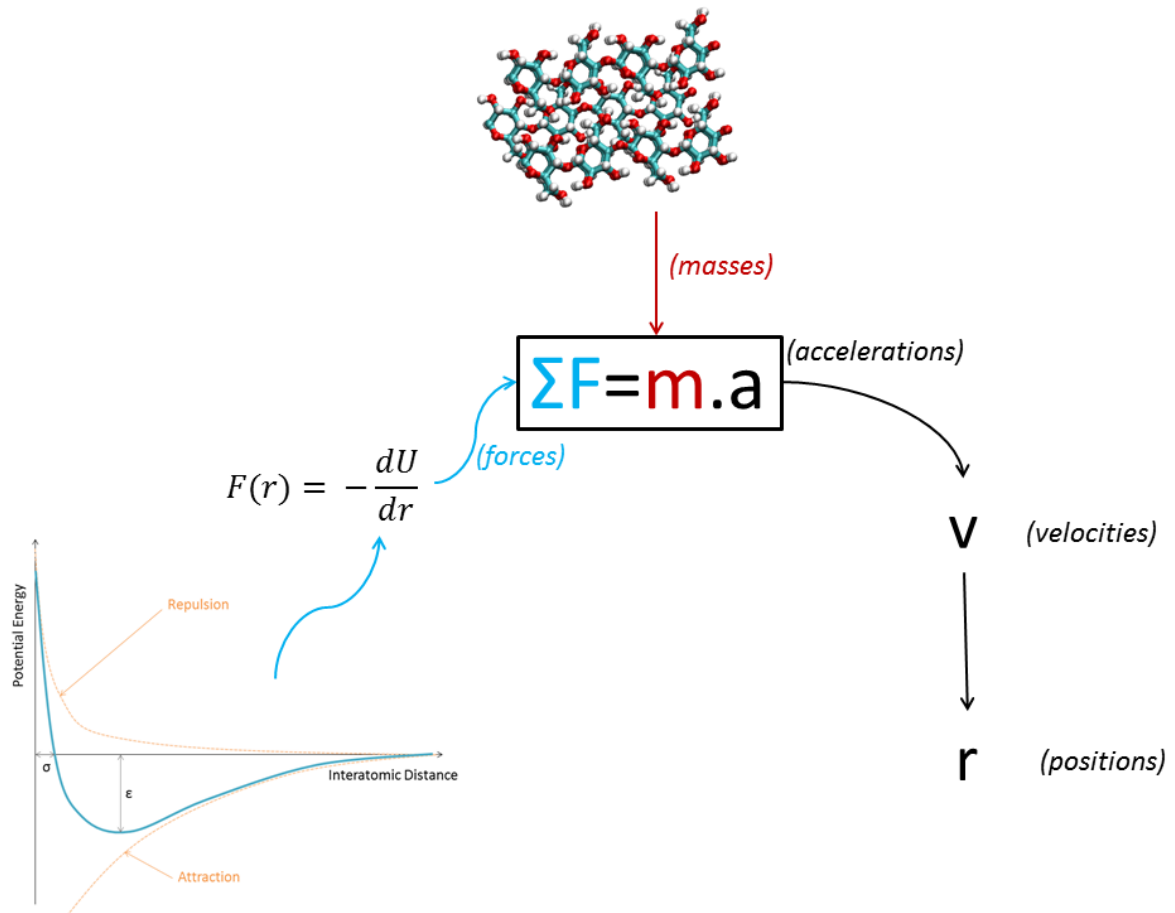


Figure 27. The classical MD scheme.

An integration algorithm has to be stable on long simulation time, and accurate enough –for example, δt^4 on positions is considered as enough-, to see that the total energy of the system is conserved.

Integration algorithms in MD are usually based on finite differences. Position $\mathbf{r}$ at time $t+dt$, is deduced from a Taylor expansion:

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \delta t \cdot \dot{\mathbf{r}}(t) + \frac{\delta t^2}{2!} \cdot \ddot{\mathbf{r}}(t) + \frac{\delta t^3}{3!} \cdot \dddot{\mathbf{r}}(t) + O(\delta t^4)$$

We know that

$$\mathbf{v}(t) = \dot{\mathbf{r}}(t)$$

And

$$\mathbf{a}(t) = \ddot{\mathbf{r}}(t)$$

Which implies that

$$r(t + \delta t) = r(t) + \delta t \cdot v(t) + \frac{\delta t^2}{2!} \cdot a(t) + \frac{\delta t^3}{3!} \cdot \ddot{r}(t) + O(\delta t^4)$$

In the same way $r(t - \delta t)$ can be written as it follows:

$$r(t - \delta t) = r(t) - \delta t \cdot v(t) + \frac{\delta t^2}{2!} \cdot a(t) - \frac{\delta t^3}{3!} \cdot \ddot{r}(t) + O(\delta t^4)$$

Summing both last equations we get:

$$r(t + \delta t) + r(t - \delta t) = 2 \cdot r(t) + \delta t^2 \cdot a(t) + O(\delta t^4)$$

This is known as the position Verlet algorithm, published 1967²⁶. It is an algorithm based on a step to step limited development of Taylor, as demonstrated above. The velocities can be then obtained:

$$r_i(t + \delta t) = 2r_i(t) - r_i(t - \delta t) + a_i(t) \cdot \delta t^2 + O(\delta t^4)$$

$$v_i(t) = \frac{r_i(t + \delta t) - r_i(t - \delta t)}{2\delta t} + O(\delta t^2)$$

Where r_i is the position of the particle i , v_i is the velocity of the particle i and δt is the time increment. Concerning the positions, the first negligible term is on δt^4 which is rather accurate. Velocities are less accurate (δt^2) but are mainly used for temperature T computation:

$$T = \frac{1}{3 \cdot N \cdot k_B} \cdot \sum_{i=1}^N \frac{p_i^2}{m_i}$$

Where N is the number of particles, k_B is the Boltzmann constant, p_i is the particle i momentum, and m_i the particle i mass, with $v=p/m$.

Nosé-Hoover thermostat

As mentioned before, the natural ensemble to run a molecular dynamics simulation is the micro canonical one –NVE. If we wish to simulate a system evolution under constant temperature, a thermostat is needed. Different kinds of thermostats exist, such as the Andersen one that randomly modifies velocities of particles in the system every Δt time, according to the Maxwell distribution of velocities –or the Nosé-Hoover one –first published by Nosé^{17,18} then modified by Hoover²⁰– that modifies equation of motion by addition of a friction term evolving as a function of the difference between measured

kinetic energy and the targeted one. In this work we have chosen the Nosé-Hoover thermostat.

With a Nosé-Hoover thermostat, system temperature fluctuates around the wanted one. Temperature exchanges with the thermostat are performed following these relations:

$$\begin{aligned}\frac{dr_i}{dt} &= \frac{p_i}{m_i} \\ \frac{dp_i}{dt} &= -\frac{dU(r^N)}{dr_i} - \xi \cdot p_i \\ \frac{d\xi}{dt} &= \frac{\sum_i \frac{p_i^2}{m_i} - \frac{L}{\beta}}{Q} \\ \frac{ds}{s \cdot dt} &= \frac{d \ln s}{dt} = \xi\end{aligned}$$

Where r is particle position, p particle momentum, m particle masses, U potential energy, ξ a friction term -which is linked via a feedback equation to the kinetic energy²⁷- L is the number of freedom degrees of the system, Q is an effective mass -linked to the heat bath inertia-, and s is an additional coordinate and results of the Sundman transformation:

$$s = \frac{d\tau}{dt}$$

Where τ is a virtual time, and t the real one. Note that with the use of Nosé-Hoover thermostat, the whole system -system + thermostat- samples the micro canonical ensemble.

Input information

Interaction potential and integration algorithm are used to reproduce a system evolution. Obviously one needs to input the initial configuration and choose a statistical ensemble. The initial configuration is a file in which initial atomic coordinates are given. System size is limited by computation capacity. In MD, even if the number of atoms able to be simulated is significantly larger than the one with quantum methods -up to millions of atoms against few hundred- this amount remains small compared to atoms numbers in real systems. In order to limit effects due to finite size of the system, configurational box is usually periodically repeated in the three dimensions. In this way, any atom leaving the simulation box on one side is reintroduced by the opposite side, with the same velocity. It is called periodic boundary conditions -see in Figure 28. The minimum image convention is also used to calculate interactions between atoms.

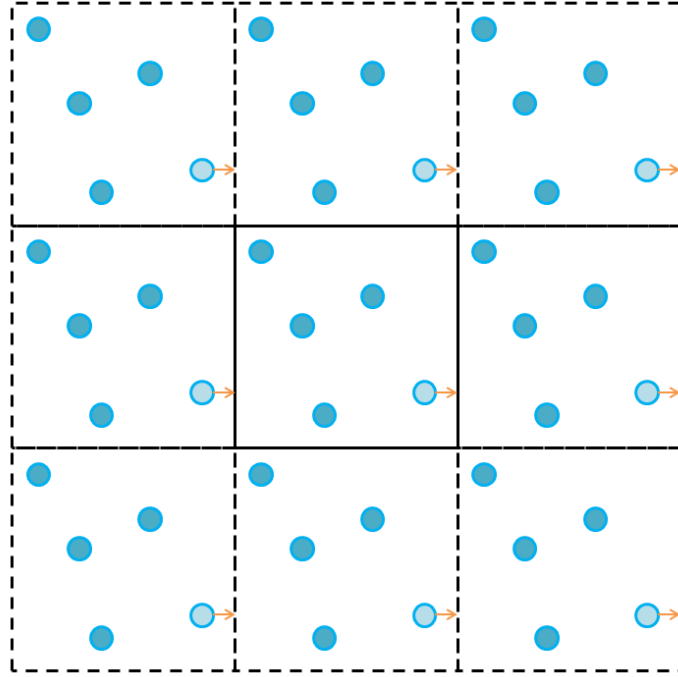


Figure 28. Schematic of periodic boundaries conditions, or minimum image convention.

In periodic boundary conditions, an atom must not interact with itself or several images of another atom, but only with the one at the closest distance. This is introduced by using a cutoff distance that has to be equal or less than half of the simulation box size.

As mentioned above, integration time step is also an important parameter. It has to be not too small, in which case the solution would be very accurate but simulation will take too much time. It has to be neither too large, in which case integration algorithm would become unstable²⁸. It is usually taken around the femtosecond, depending on the system and the potential. Time step has indeed to be chosen smaller than the period of the highest vibrational frequency in the system, in order to have a stable dynamics. A sp^2 C-C bond has a vibrational frequency around $4.9 \cdot 10^{13}$ Hz which corresponds to a period of $2.0 \cdot 10^{-14}$ seconds, thus 20 femtoseconds. Initial conditions –as temperature and pressure- are introduced as inputs of the simulation, depending on the chosen statistical ensemble.

Once integration algorithm, interaction potential, ensemble, and initial configuration are chosen, depending on the simulation goal, simulation can be run. If system is big, run can be long. In order to decrease the simulation time, parallelization can be used.

Parallelization

Processors power increases but the need for large calculations increases even faster. Processors power has increased for a long time following the Moore's law but the increase in computing power nowadays results from the use of parallelization. The most common way to keep a reasonable simulation time is to associate several processors on a same task –they can be assigned to execute a certain part of the work load. The most efficient way is called parallelization.

Its principle is to explicitly distribute the work load on several processors. Every processor does its own part of the job. The main question to address is communication between them. Synchronization of data transfer between processors has to be efficient since some processors may need results from others. They need to communicate, to exchange data and sometimes wait for each other.

Goal of parallelism is to run large jobs in a reasonable time. Software libraries are available such as MPI –Message Passing Interface. Parallelization can be applied to the job by itself –distributing tasks- or spatially –dividing simulation box. Large jobs are usually launched on clusters, to a master that will spread tasks or bow parts to other processors and manage communications between them. Note that it is not possible to have an efficiency of M times the one of a single processor with M processors because of the communication time between them. Moreover, time gain by adding more processors to a parallelization stops increasing after a critical number of processors, when the communication time starts to be more significant than the time dedicated to calculation.

Replica exchange molecular dynamics (REMD)

Concerning study of organic matter decomposition on geological times, the main problem is the difficulty to overcome really long physical times. It is indeed highly challenging to reproduce a geological phenomenon in a human time base. In order to numerically simulate this process, one needs to find a method capable of accelerating time. We here decided to adopt the *Replica Exchange Molecular Dynamics (REMD)* method. It is also called *Parallel Tempering*. We first explain its principle before to emphasize the significance of replicas choice and give a view of application domains. We develop the principle of REMD –or Parallel Tempering Molecular Dynamics (**PTMD**)- and its history.

Monte-Carlo Parallel Tempering

The first publication containing an idea of Parallel Tempering method is the one of Swendsen & Wang, in 1986²⁹. They presented a new Monte-Carlo method with which they reduced computational times for random interactions problems and allowed faster study of low-temperature slow processes. The originality in their Replica Monte Carlo is the use of independent replicas of a system at different temperatures and the swap of these replicas when configurations are energetically favorable to be transferred to lower temperatures. One key condition in their technique was to have a non-null probability of connection between two configurations in order to have a possible transition. The swap is applied to every replica with its neighboring temperature replicas. In this initial version of Replica Exchange MC, they did not yet exchange the whole configuration –by the way of temperature exchange- but only partial information such as configurational degrees of freedom.

In 1991, Geyer³⁰ was the first to attempt whole configuration information swapping. He called his method *Metropolis-coupled Markov chain Monte Carlo* in which he run Markov chains in parallel –with different but related equilibrium distributions- and he attempted swaps between two chains with an acceptance ratio. In 1996, Hukushima and Nemoto³¹ proposed a whole algorithm to simulate what they qualified as “hardly-relaxing” systems. They simultaneously simulate replicas at different temperatures and introduced a virtual exchanging configurations process. Their tests showed that their algorithm allowed systems to relax really fast, even at low temperatures.

At first, only simple model used in statistical physics problems were solved using the Parallel Tempering method but application fields range rapidly increased. For instance, in 1997, Hansmann published a paper³² in which he demonstrated the effectiveness of the method on a peptide structure. He used seven temperatures –from 50 to 1000K- and showed the technique could be efficiently applied to bio-molecules avoiding configurations to get trapped in local minima as it can often be the case with classical Monte-Carlo and Molecular Dynamics simulations of complex systems.

Molecular Dynamics Parallel Tempering: REMD

In 1999, Sugita and Okamoto³³ proposed an algorithm of Parallel Tempering applied to Molecular Dynamics. They tested this new algorithm on a peptide system, using the force-field AMBER³⁴ and eight replicas. They found that their method could overcome multiple minima problem and that from a single run, they could obtain various thermodynamic quantities for a wide temperature range. Note that in REMD, each

replica is run at constant temperature. As mentioned before, NVT ensemble is not the natural one for MD. Therefore a thermostat needs to be used to keep a constant temperature. In our simulations –with the LAMMPS code- we use a Nosé-Hoover thermostat.

One may keep in mind that every single exchange during Replica Exchange Molecular Dynamics is a non-physical move. Pathway is accelerated –and that is the reason of using such a method- and not reversible. The main goal of this technique is to quickly evolve towards the thermodynamic equilibrium corresponding to the temperature of interest.

Principle of Parallel Tempering

REMD method is a way of overcoming large free energy barriers, or equivalently slow kinetics, by enhancing configurational sampling of a conventional constant temperature MD, using states from higher temperatures. It uses high temperatures replicas to help lower ones to overcome energy barriers. Replicas are copies of the initial system in the canonical ensemble. Each one of those is run at a different temperature in a way that there is exactly one replica at each temperature –of the chosen temperature set.

To help lower temperatures replicas to overcome energy barriers, temperature exchanges between neighbor -in temperature- replicas are allowed. Indeed, in a state i at a given temperature T , a molecular configuration will exist with a certain distribution $\exp(-E_i/k_B T)^{ix}$, as shown in Figure 29. At a little bit higher temperature than the previous one, another molecular configuration will exist and its energy distribution will be slightly shifted compared to the first one. If these two temperatures are close enough, their energy distributions will overlap. It means that the probability that a same configuration exists for both temperatures is nonzero.

^{ix} Boltzmann distribution for a system at temperature T , in canonical ensemble –number N of particles, volume V and temperature T constant.

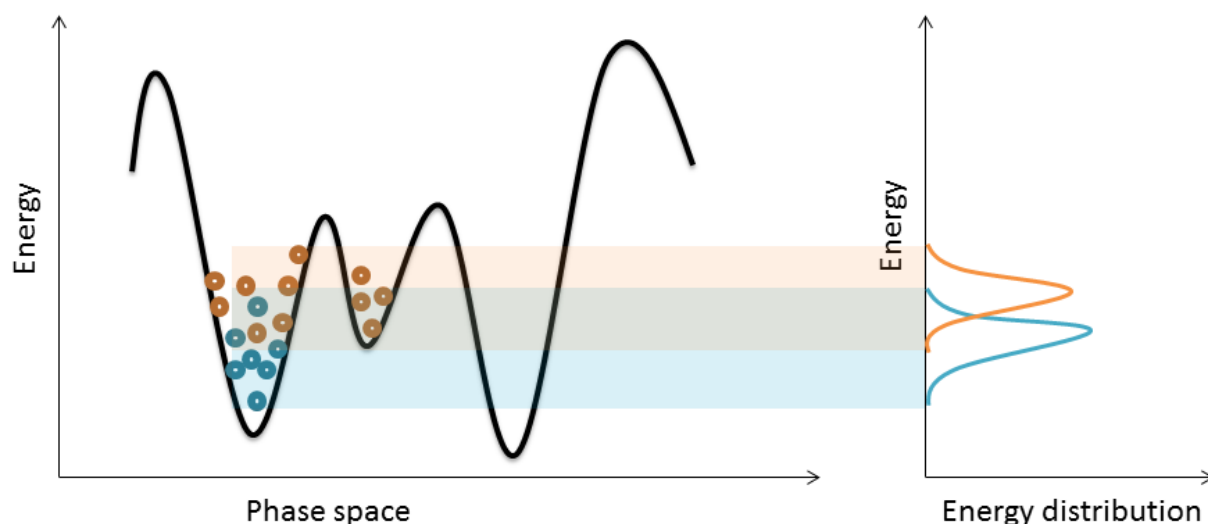


Figure 29. Relation between configuration energy and its distribution. Blue circles represent a molecular configuration of lower temperature than orange ones. Blue configuration can exist in a certain energy domain and its energy distribution is shown on the right side. It is the same for the orange one. As these two temperatures are close enough, both distributions overlap.

In REMD method, replicas of a same initial system are launched in the NVT ensemble at different temperatures, relatively close from each other. Replicas of higher temperatures are used to help lower temperatures replicas to evolve faster, by allowing temperature swaps between neighbor replicas. The temperature of interest is the lowest one and will change several times of replica during the simulation. Through the replica exchanges machinery, system configuration will evolve, only through possible structures towards its thermodynamic equilibrium.

In Figure 30 is given a schematic representation of temperature swaps between neighbor replicas. Every replica can only swap with a neighbor one. If one follows the initial lowest temperature, represented in blue, one may see its evolution through different replicas which have been visited by highest temperatures and which thus has more evolved configurations. All molecular configurations on its pathway are possible at the given temperature.

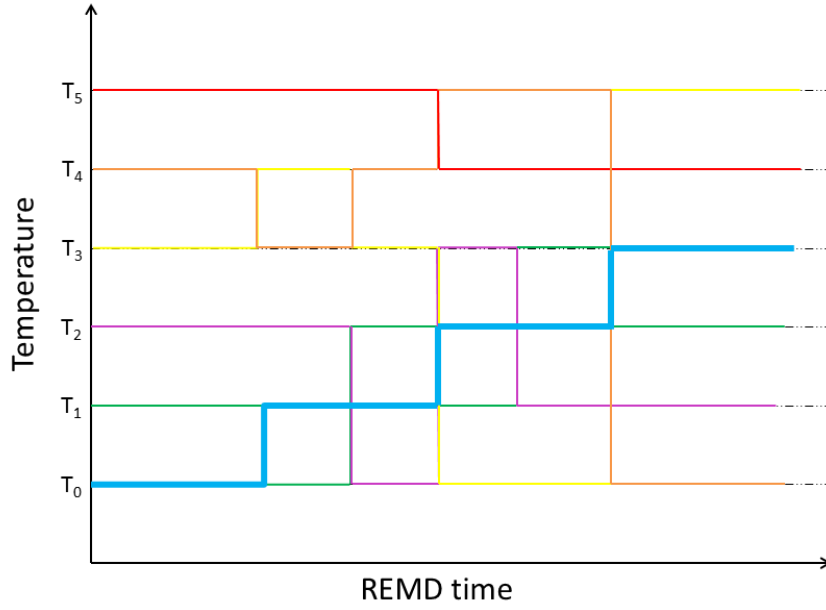


Figure 30. Schematic representation of swaps between neighbor replicas during a REMD run. Blue line represents the initial lowest temperature system evolution. One can note that configurations which have been at higher temperatures are found at T_0 .

Hamiltonian of a system composed of N atoms of mass m_k ($k=1\dots N$), of coordinate vectors $q=\{q_1\dots q_N\}$ and of momentum vectors $p=\{p_1\dots p_N\}$ is described as follows:

$$H(q, p) = K(p) + U(q)$$

Where $K(p)$ is kinetic energy and $U(q)$ is potential energy. Kinetic energy is defined as:

$$K(p) = \sum_{k=1}^N \frac{p_k^2}{2 \cdot m_k} = \frac{3}{2} \cdot N \cdot k_B \cdot T$$

In the canonical ensemble, probability for system to be in state x at the temperature T is given by:

$$W(x, T) = \exp(-\beta \cdot H(q, p))$$

Where $\beta = 1/k_B \cdot T$, where k_B is the Boltzmann constant.

Since replicas do not interact, probability of the system to be in state x can be generalized as:

$$W(x) = \exp\left\{-\sum_{m=1}^M \beta_m \cdot H(q^{[i(m)]}, p^{[i(m)]})\right\}$$

Where m is the label of a given replica.

We now consider to exchange two replicas i and j at temperatures T_m and T_n respectively. Note that exchanging configurations or temperatures between two replicas leads to the same result³³. Therefore, the exchange can be done as follows:

$$\begin{cases} x_m^{[i]} \equiv (q^{[i]}, p^{[i]})_m \rightarrow x_n^{[i]'} \equiv (q^{[i]}, p^{[i]'})_n \\ x_n^{[j]} \equiv (q^{[j]}, p^{[j]})_n \rightarrow x_m^{[j]'} \equiv (q^{[j]}, p^{[j]'})_m \end{cases}$$

Nonetheless temperature swap requires lower computational effort. Assignment proposed by Sugita and Okamoto³³ concerning momentum evolution during a swap is the following:

$$\begin{cases} p^{[i]'} = \sqrt{\frac{T_n}{T_m}} \cdot p^{[i]} \\ p^{[j]'} = \sqrt{\frac{T_m}{T_n}} \cdot p^{[j]} \end{cases}$$

Detailed balance on transition probability $pt(x \rightarrow x')$ gives:

$$W(x) \cdot pt(x \rightarrow x') = W(x') \cdot pt(x' \rightarrow x)$$

We thus have:

$$\begin{aligned}
\frac{pt(x \rightarrow x')}{pt(x' \rightarrow x)} &= \frac{W(x')}{W(x)} = \frac{\exp\{-\beta_m \cdot (K(p[j']) + U(q[j])) - \beta_n \cdot (K(p[i']) + U(q[i]))\}}{\exp\{-\beta_m \cdot (K(p[i]) + U(q[i])) - \beta_n \cdot (K(p[j]) + U(q[j]))\}} \\
&= \exp\{-\beta_m \cdot (K(p[j']) - K(p[i]) + U(q[j]) - U(q[i])) \\
&\quad - \beta_n \cdot (K(p[i']) - K(p[j]) + U(q[i]) - U(q[j]))\} \\
&= \exp\left\{-\beta_m \cdot \left(\frac{T_m}{T_n} \cdot K(p[j]) - K(p[i])\right) - \beta_n \cdot \left(\frac{T_n}{T_m} \cdot K(p[i]) - K(p[j])\right) \right. \\
&\quad \left. - \beta_m \cdot (U(q[j]) - U(q[i])) - \beta_n \cdot (U(q[i]) - U(q[j]))\right\} \\
&= \exp\left\{-\frac{1}{k_B \cdot T_n} \cdot (K(p[j]) - K(p[i])) + \frac{1}{k_B \cdot T_m} \cdot (K(p[i]) - K(p[j])) \right. \\
&\quad \left. - \beta_m \cdot (U(q[j]) - U(q[i])) - \beta_n \cdot (U(q[i]) - U(q[j]))\right\} \\
&= \exp\{(\beta_n - \beta_m) \cdot (U(q[j]) - U(q[i]))\} = \exp\{-\Delta\}
\end{aligned}$$

If we reformulate it in the Metropolis criterion –cf. Monte-Carlo section– the transition probability writes:

$$pt(x \rightarrow x') = \begin{cases} \exp\{-\Delta\} & \text{for } \Delta > 0 \\ 1 & \text{for } \Delta \leq 0 \end{cases}$$

We therefore find that temperature swaps are accepted or rejected with the following acceptance probability criterion:

$$p_{ij}^{acc} = \min\left[1, \exp\left(\left[\frac{1}{k_B T_i} - \frac{1}{k_B T_j}\right] \cdot [U_i - U_j]\right)\right]$$

Where P_{ij}^{acc} is the acceptance criterion, k_B is the Boltzmann constant, T_i and T_j are the imposed temperatures to replicas i and j respectively and U_i and U_j are their potential energies. In this equation, if $T_i < T_j$ it implies that, in average, $U_i < U_j$. The criterion will accept a temperature swap only if replicas are close enough in energies –or temperatures– thanks to instantaneous fluctuations. In other words, REMD will only provide enhanced sampling if the distributions of instantaneous temperature/potential

energy of the different replicas present some overlap. Choice of replicas temperatures is then critical for the simulation's efficiency.

Choice of replicas

The choice of both number of replicas and their associated temperatures aims to achieve the best possible sampling by using the minimum computational effort. Considering the temperature range, there are two conditions:

- the lowest value is the temperature of interest –and is a temperature at which regular simulation might get trapped into a local minimum–
- the highest must be chosen high enough in order to have no replica trapped into a local energy minimum and like every energy barrier will be crossed at this temperature. It has to not be too high in order to not lose computation time in non-physical trajectories.

This condition on the highest temperature is the first requirement in order to run a successful REMD. The second requirement is to use enough replicas to be sure that each of them overlaps with its neighboring ones.

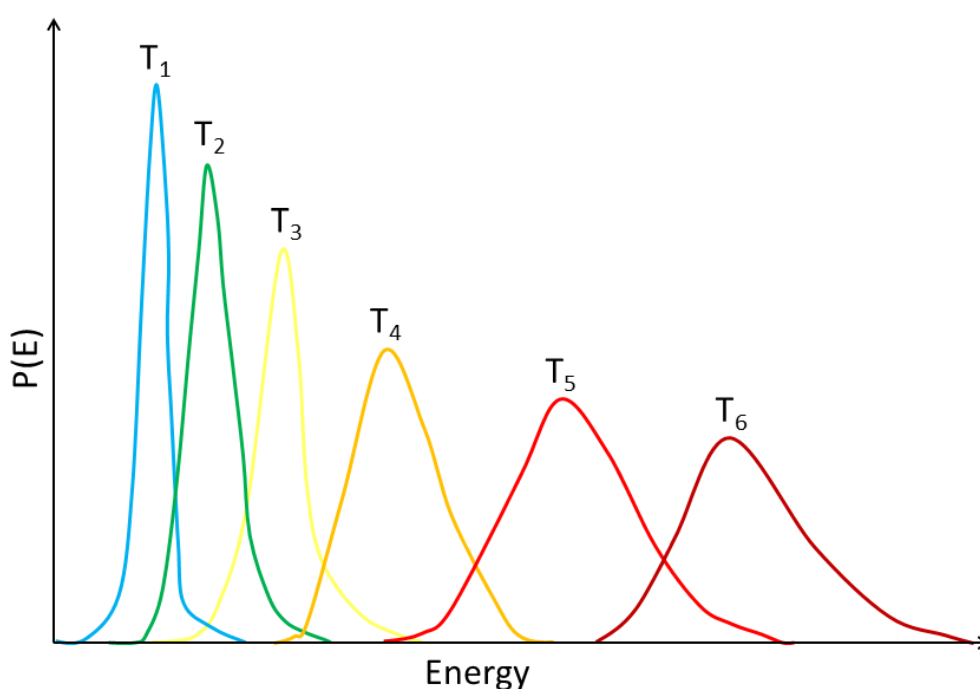


Figure 31. Histogram of energy distribution in the case of a six-replicas REMD run. One may note that distributions overlap with their neighbors in order to allow exchanges.

In Figure 31 it is shown how the energy distributions should overlap with their neighbors in order to allow temperature –or configuration– swaps during the run. This

overlap means that the acceptance criterion can be validated with the energy U_j of a replica j sometimes lower than the potential energy U_i of a replica i , where $T_i < T_j$.

Upon choosing the number of replicas, one has to take into account the fact that the width of the system energy distribution, for a given system, increases with the temperature. It means that, for a given system size, at low temperature, energy distribution will be thin and of high amplitude, while at higher temperature, it will be of smaller amplitude but with larger spread. Spacing between two neighboring replicas can therefore be more important for high temperatures than for low ones.

Sugita and Okamoto³³ and, later, Kofke^{35,36} proposed that each replica stays the same amount of time at the same temperature in order to get a uniform acceptance probability. Kofke^{35,36} also showed that, for systems in which the heat capacity is constant, a progression between two following temperature i and j as T_i/T_j is constant will lead to a uniform acceptance ratio.

Applications

Systems in biology usually have complex energy landscapes. Deep energy wells with high energy barriers make difficult the sampling. This is particularly complicated to simulate their evolution with classical Molecular Dynamics or Monte-Carlo simulations. Indeed, residence time of a system in energy well can be really long, too long in fact for an evolution to be accessible in the time scale of classical simulation. Parallel Tempering has been showed to be more efficient and to overcome possible slowdowns during simulations. Hansmann³² applied a Monte-Carlo Parallel Tempering to a peptide system –he was the first to do so- and found a much better effectiveness than traditional methods. He indeed showed that when 1 000 000 steps are not enough to thermalize to his peptide at 50K in regular canonical Monte-Carlo, less than 10 000 are needed to Parallel Tempering. Sugita and Okamoto³³ applied a REMD simulation to study a protein folding. Their study showed that the technique allowed the system to escape from local energy minima and to avoid to get trapped into it, by contrast to a classical molecular dynamics simulation into the NVT ensemble.

Other research teams also worked on protein folding, such as Garcia and Onuchic³⁷ who worked on a protein folding mechanism and computed free energy, entropy, and surface enthalpy of their system. Zhou³⁸ also studied protein folding with Replica Exchange Molecular Dynamics and revealed details about intermediate states structures and thermodynamic properties. Crystal structure determination is also a field in which Parallel Tempering is widely used. In 1999, Falcioni and Deem³⁹ determined zeolites structures from powder diffraction data. They tested the method on all known zeolites structures and succeed every time. REMD has also been successfully applied to crystal

structure determination by Dova *et al.*⁴⁰ and Garcia-Cuesta *et al.*⁴¹. In 2001, Auer and Frenkel⁴² applied Parallel Tempering method to study crystal nucleation rate and were the first to prove the possibility of calculating such a value from simulation. Rate of crystal nucleation was then difficult to predict, although it was a widely studied phenomenon.

Parallel Tempering is also useful in the study of polymers. In 1999, Yan and de Pablo⁴³ applied the method to a polymeric system and demonstrated it was more efficient than previously available methods. They highlighted that the method was extremely simple to be incorporated into existing simulation codes. Other teams^{44,45} worked on polymer melts and found a significant improvement with the Parallel Tempering method, compared to regular simulations.

Parallel Tempering –or REMD- method has been widely applied to different fields as biology, crystals structures determination, polymers, spin glasses, but to my knowledge, no simulation study has been published so far focusing on organic matter degradation with REMD techniques.

LAMMPS

The Large-scale Atomic/Molecular Massively Parallel Simulator (**LAMMPS**)⁴⁶ is an open-source code.

All the simulations of present work are performed using the LAMMPS code. LAMMPS is certainly the most widely used molecular dynamics package in the context of Materials Science. It is a highly parallel code using domain decomposition algorithm and both supports the ReaxFF force field and REMD algorithms.

Definition

Plimpton and co-workers started to develop LAMMPS during the 1990s and it was publicly released for the first time in 2004. It is massively parallelized but can also be used sequentially. LAMMPS is coded in C++. Since the first release, LAMMPS has been significantly improved and new features have been added. Since 2004 more than thirty versions have been released⁴⁷. Numerous interaction potential –as ReaxFF- and method –as REMD- are implemented.

The use of ReaxFF and the REMD method with LAMMPS

LAMMPS integrates several different interaction potentials and methods. We briefly give the way of using the ReaxFF potential and the REMD method within LAMMPS. Further information can be found on the LAMMPS website⁴⁶.

Reaxff

In order to run a simulation with the reactive force field ReaxFF, user has to specify some entries to the LAMMPS code. Important instructions in the input script are the following.

```
pair_style reax/c cfile keyword value
pair_coeff * * ffield.reax C H O
fix ID group-ID qeq/reax Nevery cutlo cuthi tolerance params
```

Where `pair_style`, `pair_coeff`, and `fix` are names of LAMMPS commands. `pair_style` indicates to LAMMPS which kind of potential will be used and the style `reax/c` computes the ReaxFF potential. Note that the corresponding package must be compiled during LAMMPS installation in order to be able to use it. `cfile` can define a file in which some parameters will be changed compared to the default ones. Keyword and its value can be used for instance to increase size for array allocation but is optional.

`pair_coeff` informs on where the potential parameters file is, on its name `ffield.reax`, and on the order of elements. In our example atom type 1 will be C, type 2 will be H and type 3 will be O.

The `fix qeq/reax` is used for charge equilibration. It is based on neighboring interactions and partial charge is adjusted depending on it. `ID` corresponds to the fix name –which can be used later in the script to recall the command– and `group-ID` informs to which particles group the command is applied. If one wants to apply it to the whole system one will use `all` as `group-ID`. `Nevery` correspond to the frequency at which equilibration is performed, `cutlo` and `cuthi` are cutoff limits of the adjustment, `tolerance` corresponds to the precision of the charge equilibration and `params` specifies the parameters file. More information can be found on the corresponding web pages^{48,49} for these three commands.

REMD

First of all, if one wants to use the Replica Exchange Molecular Dynamics method, one has to compile LAMMPS with the **REPLICA** package. The main line to add into the input script is:

```
temper N M temp fix-ID seed1 seed2 (index)
```

Where `temper` is the name of the LAMMPS command, `N` is the number of time steps to run for the simulation, `M` is the frequency to attempt a tempering swap, `temp` is the initial temperature for the ensemble and `fix-ID` is the name given to this command –we can recall the command by using this name. `Seed1` and `seed2` are seeds used for randomization. `(index)` is used when restarting a parallel tempering run. It gives information on which replica was at which temperature. `temp` is usually a variable giving the temperatures at which replicas are run. For that, a line like the following needs to be added.

```
variable temp world T1 T2 T3 T4 T5 T6
```

Where `variable` is the name of the LAMMPS command, `temp` is the name of the variable defined by the user, `world` is the style of the command and means that one string is defined for each partition of processors. `Ti` are the defined temperatures.

In the submission bash script, a complement has to be added in order to inform the clusters of the number of processors requested and their allocations.

```
[...] -partition  $\alpha \times \beta$  -in file.in [...]
```

Where `-partition` indicates that the run will be running with multiple partitions and assigns one string of the previous variable to every processor partition. α is the number of replicas and β is the number of processors used per replica. `file.in` corresponds to the input file to read.

Summary

In research, two ways of studying processes exist: physical experimentations or numerical ones. The first way uses real samples with physical tools while the second one uses numerical models and computers. In this work, we use the latest. We are here interested into the atomic level. In this section, we presented the methods we used. First,

we gave a description of the interaction potential we use, then an overview of different numerical simulation techniques –Monte-Carlo and Molecular Dynamics. In a third part, we focused on the *Replica Exchange Molecular Dynamics* technique.

Simulations are performed in different thermodynamic ensembles, imposing external environment constraints to the system. Four main statistical ensembles exist: NVE, NVT, NPT, and μ VT.

Monte-Carlo methods are based on statistical sampling techniques, adapted to every statistical ensemble. The Metropolis algorithm is an importance sampling method, the principle being to weight more probable configurations over the others. Changes on the system of interest are randomly attempted, and accepted or rejected following an acceptance criterion. It can be applied on significant size systems and calculus is relatively quick.

Molecular Dynamics simulations use equations of motion of Newtonian mechanics in order to deduce trajectories of system of interest components –usually atoms. MD provides time evolution. Every particle of the system is treated as a point mass and Newton's equation is integrated in order to get their trajectories. To resolve equations, interactions between particles must be known and are given by an interaction potential.

Interaction potential describes interactions between atoms. It provides forces applied to them and is a composition of both attractive and repulsive terms. The basic potential is the Lennard-Jones one but a number of different interaction potentials exist. In our work, we use ReaxFF which has been published by Van Duin in 2001 and is reactive. It means it is able to properly reproduce chemical bond breakings and formations. It is based on a bond-order formalism which means that not only the first neighbor is part of the bond description but several.

We use the LAMMPS code which is open-source and integrate both the ReaxFF potential and the REMD technique. The REMD method is a way of overcoming large free energy barriers, by enhancing the configurational sampling of a conventional constant temperature MD using states from higher temperatures. Same initial system configuration is launched at different temperatures and swaps between replicas are attempted with an acceptance criterion based on energy difference.

Two main documents to read in order to have a large vision on ReaxFF potential and Replica Exchange Molecular Dynamics:

- Senftle *et al.*, *The ReaxFF reactive force-field: development, applications and future directions* (2016)
- Earl and Deem, *Parallel tempering: theory, applications, and new perspectives* (2005)

1. Michael, R. & van, D. A. ReaxFF. (2011). doi:doi:10.4231/D3WH2DF8S
2. Lennard-Jones, J. E. Cohesion. *Phys. Soc.* **43**, 461–482 (1931).
3. Van Duin, A. C. T., Dasgupta, S., Lorant, F. & Goddard, W. A. ReaxFF: A reactive force field for hydrocarbons. *J. Phys. Chem. A* **105**, 9396–9409 (2001).
4. Brenner, D. W. Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamonds films. *Phys. Rev. B* **42**, 9458–9471 (1990).
5. Chenoweth, K., Van Duin, A. C. T. & Goddard, W. A. ReaxFF Reactive Force Field for Molecular Dynamics Simulations of Hydrocarbon Oxidation. *J. Phys. Chem.* **112**, 1040–1053 (2008).
6. Mattsson, T. R. *et al.* First-principles and classical molecular dynamics simulation of shocked polymers. *Phys. Rev. B - Condens. Matter Mater. Phys.* **81**, 1–9 (2010).
7. Pitman, M. C. & Duin, A. C. T. Van. Dynamics of Confined Reactive Water in Smectite Clay – Zeolite Composites. *J. Am. Chem. Soc.* **134**, 3042–3053 (2012).
8. Van Duin, A. C. T., Baas, J. M. A. & van de Graaf, B. Delft Molecular Mechanics : A New Approach to Hydrocarbon Force Fields. *J. Chem. soc. Faraday Trans.* **90**, 2881–2895 (1994).
9. Chenoweth, K. *et al.* Simulations on the Thermal Decomposition of a Poly(dimethylsiloxane) Polymer Using the ReaxFF Reactive Force Field. *J. Am. Chem. Soc.* 7192–7202 (2005). doi:10.1021/ja050980t
10. Strachan, A. *et al.* Shock Waves in High-Energy Materials: The Initial Chemical Events in Nitramine RDX. *Phys. Rev. Lett.* **91**, 98301 (2003).
11. Lee, J. S. *et al.* Atomic imaging of nucleation of trimethylaluminum on clean and H 2O functionalized Ge(100) surfaces. *J. Chem. Phys.* **135**, 0–7 (2011).
12. Salmon, E., van Duin, A. C. T., Lorant, F., Marquaire, P.-M. & Goddard, W. a. Early maturation processes in coal. Part 2: Reactive dynamics simulations using the ReaxFF reactive force field on Morwell Brown coal structures. *Org. Geochem.* **40**, 1195–1209 (2009).
13. Zou, C., Raman, S. & Van Duin, A. C. T. Large-scale reactive molecular dynamics simulation and kinetic modeling of high-temperature pyrolysis of the gloeocapsomorphaprisca microfossils. *J. Phys. Chem. B* **118**, 6302–6315 (2014).
14. Liu, X. X. *et al.* Initial pyrolysis mechanism of oil shale kerogen with reactive molecular dynamics simulation. *Energy and Fuels* **29**, 2987–2997 (2015).
15. T. P. Senftle *et al.* The ReaxFF Reactive Force-field: Development, Applications, and Future Directions. *npj Comput. Mater.* (2016). doi:10.1038/npjcompumats.2015.11
16. Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* **72**, 2384 (1980).
17. Nose, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **52**, 255–268 (1984).
18. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **81**, 511–519 (1984).
19. Nosé, S. An extension of the canonical ensemble molecular dynamics method. *Mol. Phys.* **57**, 187–191 (1986).
20. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695–1697 (1985).
21. Frenkel, D. & Smit, B. *Understanding Molecular Simulation From Algorithms to Applications.* (1996).
22. Nylund, K. L., Asparouhov, T. & Muthén, B. O. Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation

- Study. *Struct. Equ. Model. A Multidiscip. J.* **14**, 535–569 (2007).
23. Frost, V. S. & Melamed, B. Traffic modeling for telecommunications networks. *IEEE Commun. Mag.* **32**, 70–81 (1994).
 24. Metropolis, N. & Ulam, S. The Monte Carlo Method. *J. Am. Stat. Assoc.* **44**, 335–341 (1949).
 25. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1092 (1953).
 26. Verlet, L. Computer ‘Experiments’ on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **159**, 98–103 (1967).
 27. Hoover, W. G. Nose-Hoover Nonequilibrium Dynamics and Statistical Mechanics. *Mol. Simul.* **33**, 13–19 (2006).
 28. Kim, S. Issues on the choice of a proper time step in molecular dynamics. *Phys. Procedia* **53**, 60–62 (2014).
 29. Swendsen, R. & Wang, J. Replica Monte Carlo simulation of spin glasses. *Phys. Rev. Lett.* **57**, 2607–2609 (1986).
 30. Geyer, C. J. Markov Chain Monte Carlo Maximum Likelihood. *Comput. Sci. Stat. Proc. 23rd Symp. Interface* 156–163 (1991).
 31. Hukushima, K. & Nemoto, K. Exchange Monte Carlo method and application to spin glass simulations. *Journal of the Physical Society of Japan* **65**, 1604–1608 (1996).
 32. Hansmann, U. H. E. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* **281**, 140–150 (1997).
 33. Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**, 141–151 (1999).
 34. Weiner, P. K. & Kollman, P. A. AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions. *J. Comput Chem* **2**, 287–303 (1981).
 35. Kofke, D. A. On the acceptance probability of replica-exchange Monte Carlo trials. *J. Chem. Phys.* **117**, 6911–6914 (2002).
 36. Kofke, D. A. Erratum: On the acceptance probability of replica-exchange Monte Carlo trials. *J. Chem. Phys.* **120**, 10852 (2004).
 37. García, A. E. & Onuchic, J. N. Folding a protein in a computer: an atomic description of the folding/unfolding of protein A. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 13898–13903 (2003).
 38. Zhou, R. Exploring the protein folding free energy landscape: Coupling replica exchange method with P3ME/RESPA algorithm. *J. Mol. Graph. Model.* **22**, 451–463 (2004).
 39. Falcioni, M. & Deem, M. W. A Biased Monte Carlo Scheme for Zeolite Structure Solution. *J. Chem. Phys.* **110**, 28 (1999).
 40. Dova, E. *et al.* Structures of FeII spin-crossover complexes from synchrotron powder-diffraction data. *Acta Crystallogr. Sect. B Struct. Sci.* **60**, 528–538 (2004).
 41. García-Cuesta, M. C. *et al.* Structure determination of nitrato-κO-bis[2-(2-pyridyl-κN)amino-5,6-dihydro-4H-1,3-thiazine-κN]copper(II) nitrate via molecular modelling coupled with X-ray powder diffractometry. *J. Appl. Crystallogr.* **37**, 993–999 (2004).
 42. Auer, S. & Frenkel, D. Prediction of absolute crystal-nucleation rate in hard-sphere colloids. *Nature* **409**, 1020–1023 (2001).
 43. Yan, Q. & De Pablo, J. J. Hyperparallel tempering Monte Carlo simulation of

- polymeric systems. *J. Chem. Phys.* **113**, 1276–1282 (2000).
44. Bedrov, D. & Smith, G. D. Exploration of conformational phase space in polymer melts: A comparison of parallel tempering and conventional molecular dynamics simulations. *J. Chem. Phys.* **115**, 1121–1124 (2001).
 45. Doxastakis, M., Mavrantzas, V. G. & Theodorou, D. N. Atomistic Monte Carlo simulation of cis-1,4 polyisoprene melts. II. Parallel tempering end-bridging Monte Carlo simulations. *J. Chem. Phys.* **115**, 11352–11361 (2001).
 46. LAMMPS Website. Available at: <http://lammps.sandia.gov/index.html>. (Accessed: 14th November 2013)
 47. LAMMPS history. Available at: <http://lammps.sandia.gov/history.html>. (Accessed: 14th November 2013)
 48. LAMMPS pair_style. Available at: http://lammps.sandia.gov/doc/pair_reax_c.html. (Accessed: 14th November 2016)
 49. LAMMPS fix_qeq. Available at: http://lammps.sandia.gov/doc/fix_qeq_reax.html. (Accessed: 14th November 2016)

Chapter 3: Validation

REMD technique requires important calculation resources. One cannot run hundreds of thousands hours calculation without a thorough validation of the method. Methods used in this work were presented in the previous section. We first tested the reactive potential ReaxFF, for which we found some issues and had to select the best set of parameters. We then applied the Replica Exchange Molecular Dynamics to a small cellulose system.

Tests on ReaxFF

In the first chapter, we highlighted that the main component of organic matter is carbon. We run a first test in order to check the accuracy of the interaction potential. Indeed, empirical potentials are developed for a given purpose, under specific conditions in order to reproduce systems behaviors in particular cases. For instance, ReaxFF has been parametrized on chemical reactions data bases, and not on condensed carbonaceous phases. Specific parametrization implies specific application.

A number of interaction potentials reproducing carbon interactions exist, among which we can name Stillinger and Weber's¹, Tersoff's², Brenner's REBO³, Stuart's AIREBO⁴, Van Duin's ReaxFF⁵, or tight binding from Amara and co-workers⁶, and Shan's COMB⁷. Although they all reproduce carbon interactions, they are based on different principles. ReaxFF reproduces energy barriers while COMB focuses more on elastic properties. Stillinger and Weber's, and Tersoff's potentials use a pairwise additive description of the total energy and, REBO, and AIREBO, like ReaxFF, are based on a bond order concept. As mentioned in the previous chapter, we use the ReaxFF potential in this work. The tight binding method is a quantum semi-empirical method that provides a good descriptions of electronic structures. Amara and co-workers's tight binding potential has been developed to describe bonds and environment of Ni-Ni, Ni-C, and C-C bonds. Carbon-Carbon interactions have been fitted to reproduce the competition between sp, sp², and sp³ bonds, using an *ab initio* set of structures, among which an infinite linear chain, a graphene sheet, diamond, and different crystallographic configurations. The total-energy curves⁶ match very well the *ab initio* data, and that for the three carbon bonding states –sp, sp², and sp³. For this reason, we decided to use it to validate the ReaxFF potential. Note that this tight binding potential does not include hydrogen and oxygen.

We started our validation from an amorphous carbon structure that we submitted to different simulation pathways. The first step was to heat the structure at 5000K for 100ps. We then cooled it down to 1000K in 310ps -300ps from 5000K to 2000K and 10ps from 2000K to 1000K. Structure was maintained at 1000K for 100ps. The whole

simulation was done with LAMMPS and the Mattsson's⁸ implementation of ReaxFF. A configuration noted A resulted from this run.

Configuration B was produced by first putting the amorphous carbon structure at 3273K and then cooling it down to 273K. It was then maintained at 2200K for 3 million of Monte-Carlo steps per atom and afterwards cooled to 300K in 2 million of Monte-Carlo steps per atom. This first part was performed with a Monte-Carlo code and the tight binding potential⁶ of Amara *et al.* developed at CINaM. The resulting configuration was then maintained at 1000K for 100ps with LAMMPS and the Mattsson's implementation of ReaxFF.

Configuration C was obtained following the same first part of simulation than configuration B but the structure was then maintained at 1000K during 1,5 million of Monte-Carlo steps per atom, using the Amara tight binding potential and a in-house Monte-Carlo code –instead of ReaxFF and LAMMPS. A schematic representation is given in Figure 32.

Different densities have been studied and similar results have been found. Studied densities are 1.00, 1.50, 2.00, 2.50 and 3.00g.cm⁻³.

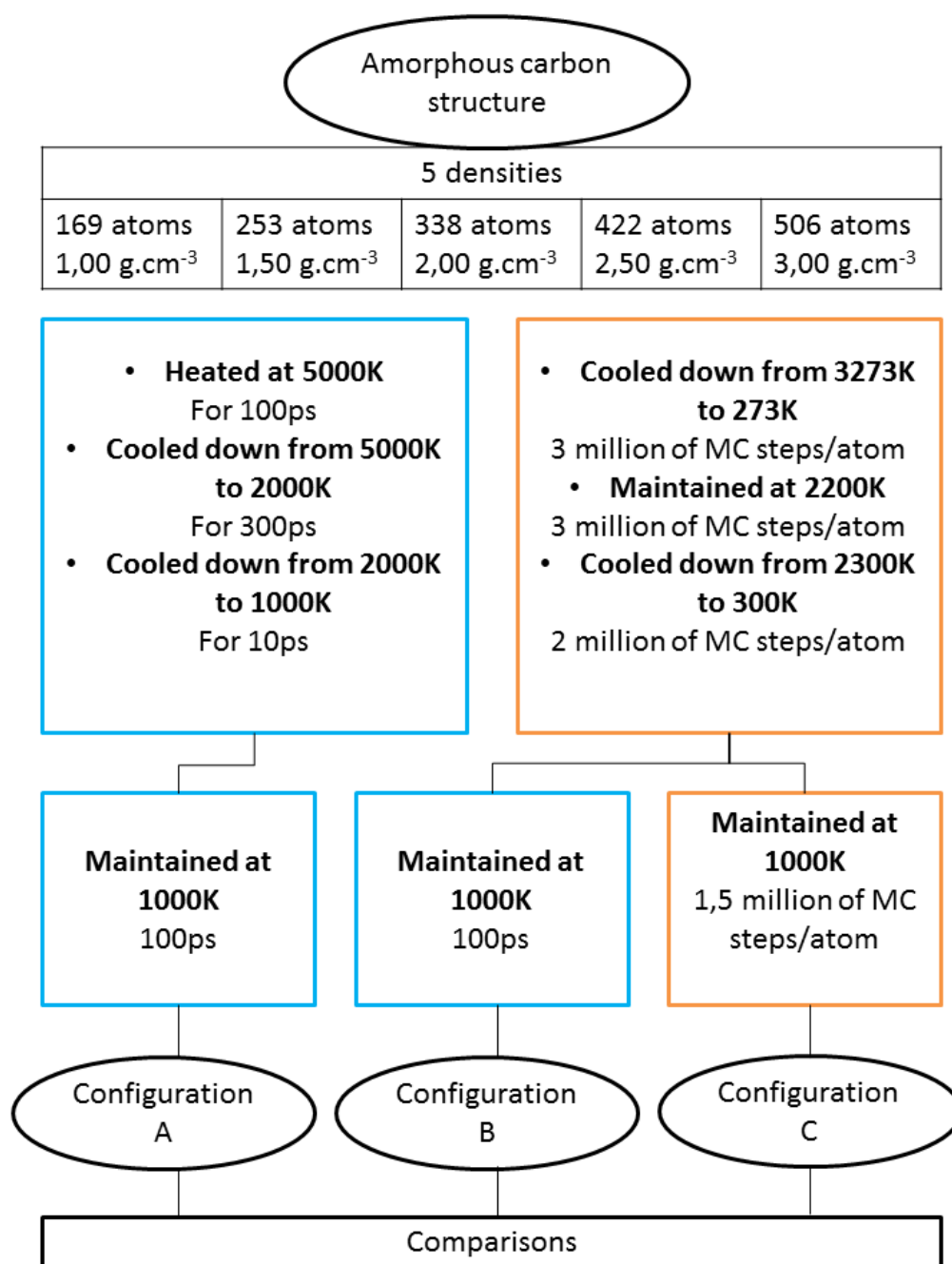


Figure 32. Schematic view of test simulation pathways. In blue simulations run with LAMMPS and the Mattsson's implementation of ReaxFF. In orange simulations run in Monte-Carlo with a tight binding potential⁶.

If one looks at structures snapshots for 169 atoms –density of 1.00 g.cm⁻³- in Figure 33, one can notice significant differences. It is extremely marked between configurations A and C. Configuration C offers an aromatic layer curved on two sides of the simulation box. The structure seems to be physical. Atoms are mainly hybridized sp². On the contrary, configuration A is not well ordered and presents triangles of carbon atoms, which is not a very stable carbonaceous formation. We also noticed that the first trial –

with only ReaxFF and LAMMPS- systematically provided more sp^3 carbons and less sp^2 than the third one –with only the tight binding potential.

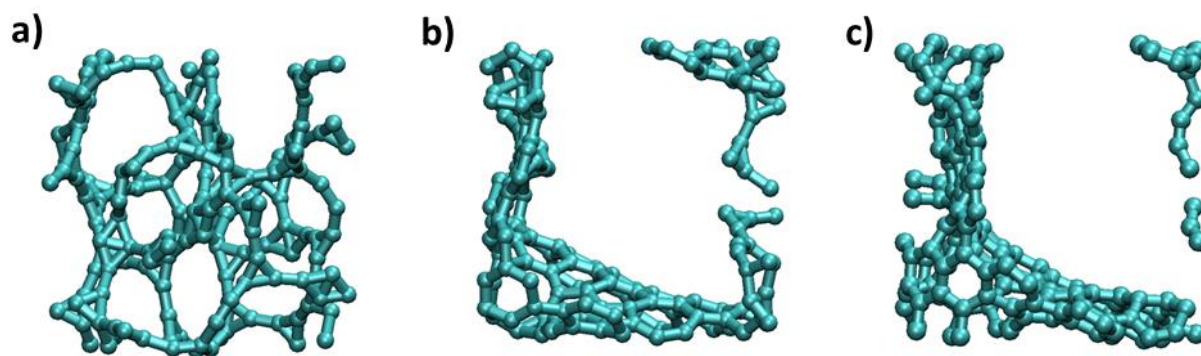


Figure 33. Snapshots of configurations a) A, b) B, and c) C for 169 carbon atoms.

At higher densities, we noticed that some carbon atoms of configurations A were overcoordinated –more than 4 first neighbors- which is not physical at all. Indeed, carbon can be hybridized in order to have 2, 3 or 4 neighbors but not more. In Figure 34 is presented the radial distribution function for C-C bonds in the case of configurations of 169 carbon atoms -density of 1.00 g.cm^{-3} . First peaks for configurations A and B are slightly shifted and are smaller than configuration C first peak. Second and third peaks are more shifted towards higher distances for configuration A, compared to B and C. Configuration B is also shifted compared to the C one.

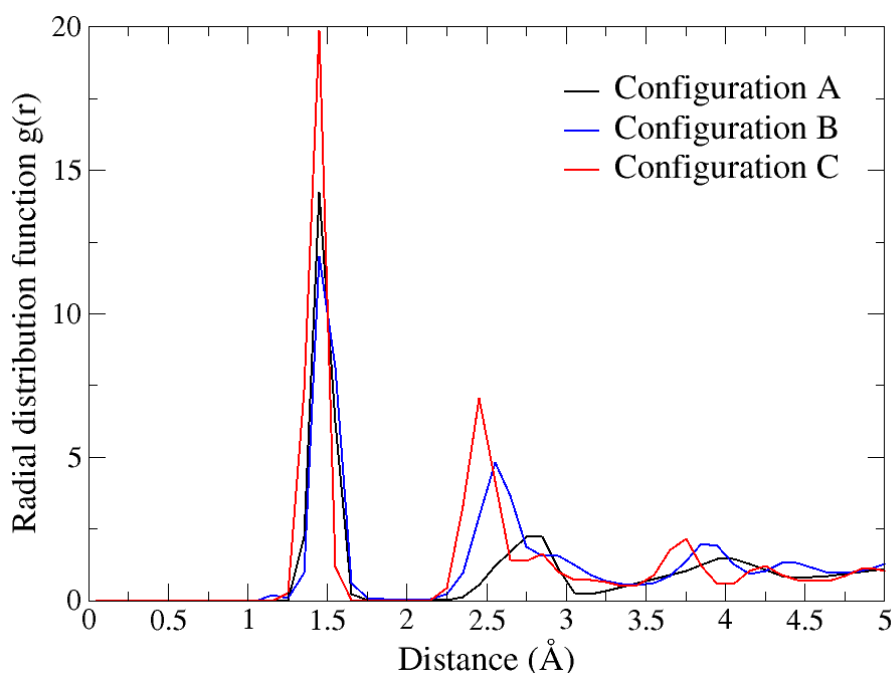


Figure 34. Radial distribution function $g(r)$ for carbon-carbon atom bonds for configurations A (black), B (blue) and C (red) for 169 atoms.

These results led us to have a closer look at the ReaxFF implementation. As recalled in the method section, a number of different sets of parameters have been released for ReaxFF. Those sets have been parametrized for specific studies. For this first test we used the Mattsson's⁸ implementation, that has been developed for shocked polymers.

We thus started a second test, comparing ReaxFF implementations with different sets of parameters. We generated configurations of graphene and diamond, with a variable lattice parameter, close to the equilibrium bond distance. In the case of graphene, carbon atoms are hybridized sp^2 and their equilibrium bond distance should be around $1.42\text{\AA}$ while in diamond, carbons are sp^3 and should be at a distance around $1.53\text{\AA}$ from each other. Energy per atom at 0K is then computed as a function of bond distance.

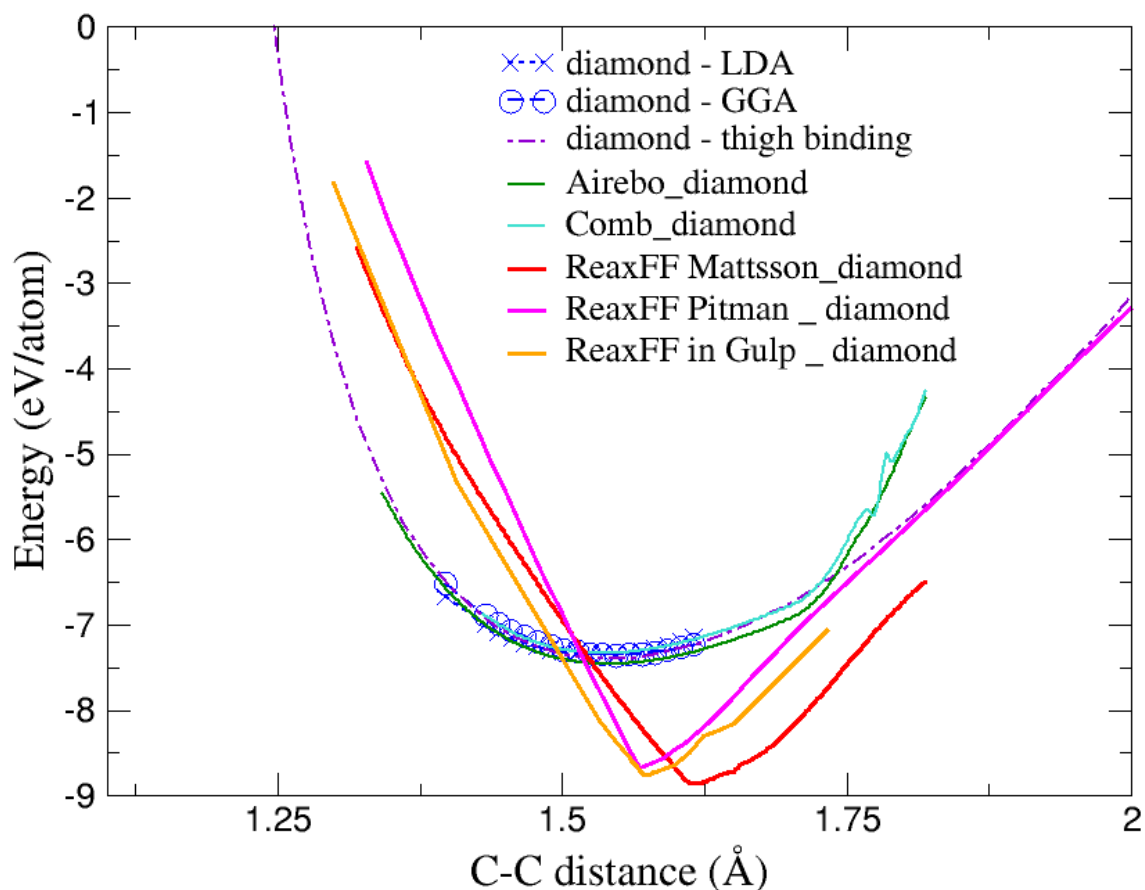


Figure 35. Energy per atom as a function of C-C bond distance in a diamond structure. Implementations of Mattsson (red solid line) and Pitman (pink solid line) ReaxFF in LAMMPS, ReaxFF in GULP⁹(orange solid line), Airebo potential¹⁰ (green solid line), COMB potential¹¹ (turquoise solid line), Tight Binding (violet dashed line), and *ab initio* DFT^x calculations LDA and GGA (respectively crosses and circles, blue dashed lines)⁶.

This energy calculation has been performed for several different potentials –ReaxFF, Airebo, COMB, Tight Binding- and using different simulation methods –Molecular Dynamics, Monte-Carlo and *ab initio*. All results for a diamond structure are given in Figure 35. We clearly note a difference between ReaxFF curves and the other ones, both in shape and values.

^x DFT –for Density Functional Theory- is a quantum mechanical based method used to study electronic structures of a system –and its evolution. To apply it, approximations need to be used. Two kinds of approximations exist: the LDA –for Local Density Approximation- and the GGA –for Generalized Gradient Approximation.

ReaxFF curves have a V-shape. As explained in the “Methods” chapter, forces used to integrate movement equations derive from the potential as:

$$F(r) = -\frac{dU}{dr}$$

Where F is the force, U the potential and r the inter particles distance. It implies the potential to be derivable everywhere, in the domain of considered distances –few Angstroms for atoms. With their V-shape, ReaxFF curves do not seem derivable everywhere. Besides, their minima are respectively at 1.616Å for Mattsson’s implementation and 1.568Å for Pitman’s. Those values are too high compared to 1.53Å for classical equilibrium distance between carbon atoms in diamond structure – hybridized sp^3 . Concerning energy value at equilibrium distance for C-C bonds in diamond, ReaxFF gives a value around -8.86eV/atom while tight binding and DFT give value of -7.34eV/atom⁶. Note that we tested two different versions of the Mattsson’s implementation -2012 and 2014- and no noticeable difference has been found. We also performed some tests using the Chenoweth¹² version of ReaxFF in LAMMPS, called ReaxFF_{CHO}, and for which we also found a V-shape.

We also performed the test on graphene for which, the potential minimum was found at a distance of 1.484Å in the Mattsson’s implementation, at variance from the true equilibrium distance of 1.42Å.

In the first paper of ReaxFF⁵, in 2001, and in different presentations of this reactive potential¹³, C-C bond stretching in ethane is given as an example. We performed the test on a C-C bond in ethane with both, Mattsson’s and Pitman’s, implementations, see in Figure 36.

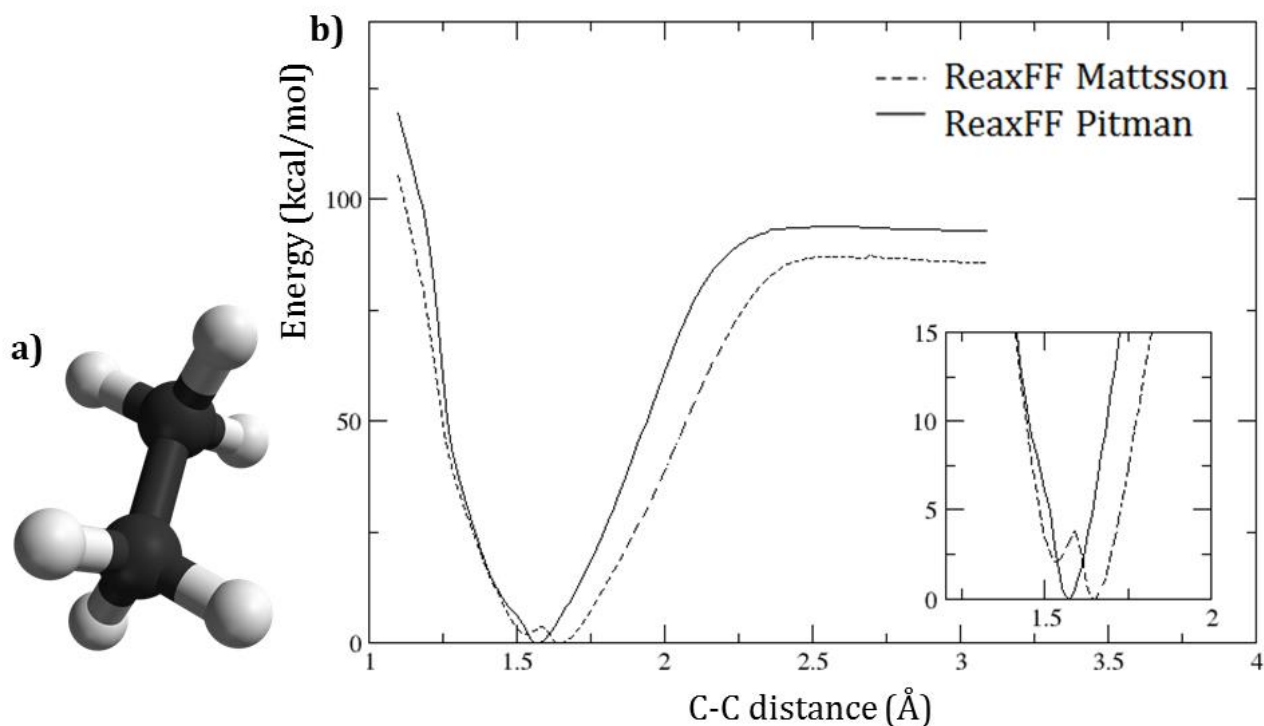


Figure 36. Test on an ethane molecule. a) representation of an ethane molecule. Carbon atoms in black and hydrogen atoms in white. b) C-C bond dissociation energy at 0K in an ethane molecule (scaled at zero). Mattsson's implementation of ReaxFF represented by dashed line. Pitman's implementation of ReaxFF represented by solid line.

Pitman's implementation well is narrow at its minimum. Mattsson's implementation has another particularity: it has two minima. This is clearly wrong. In our test, only distance between the two carbon atoms varies. A single minimum should appear in dissociation energy. Minima are at $1.52\text{\AA}$ and $1.65\text{\AA}$ and their corresponding energies are -4.14 and -4.15 eV/atom respectively –instead of 3.78^{xi} eV/atom¹⁴.

ReaxFF has a high and really interesting ambition of accurately describe a large number of species, in various situations. This goal is not easy to reach and obviously needs time and model improvement to achieve. For this purpose, new sets of parameters are regularly released.

In December 2014, Srinivasan *et al.*¹⁵ developed a version of ReaxFF for carbon condensed phases. It is called ReaxFF_{C-2013} and is the one we decided to use in our work on the degradation of cellulose and lignin (see Chapter 4). They reparametrized the ReaxFF_{CHO} using DFT data and based their work on equations of state of graphite and diamond, formations energies of defects in different carbonaceous structures, as

^{xi} At 298K, dissociation energy for C-C bond in ethane is 368kJ/mol. Value at 0K is obtained by subtracting $3/2RT$, with $R=8.32 \cdot 10^{-3}\text{kJ/mol}$. At 0K dissociation energy is then $E_{0K}=364.28\text{kJ/mol}=3.78\text{eV/atom}$.

graphene and small fullerene^{xii} and on amorphous carbon phases. They also applied their potential to fragmentation kinetics of large fullerene and concluded their paper by highlighting the importance of their potential for decomposition of carbon rich structures modelling. Comparison of their new parameters with the ones of Chenoweth¹² version is given see Figure 37.

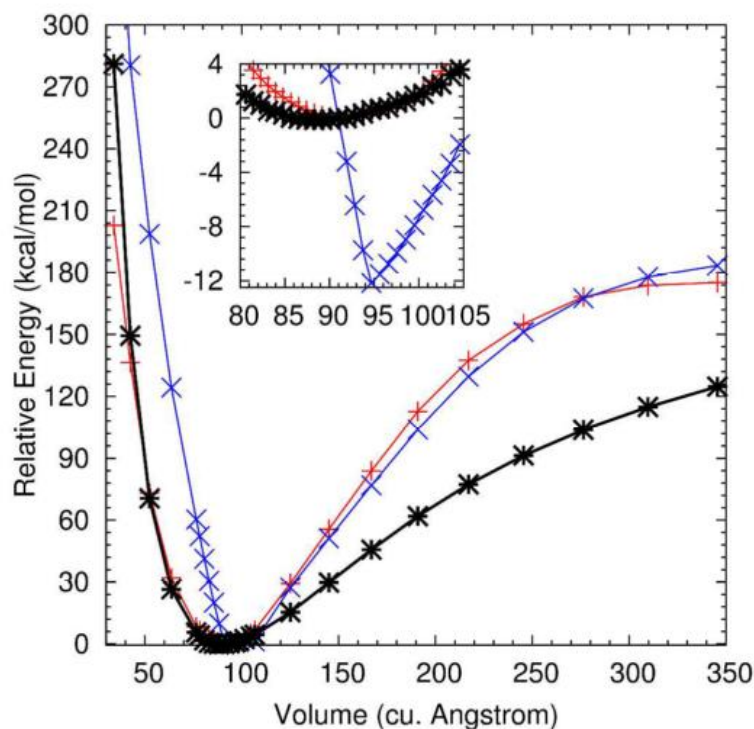


Figure 37. Comparison of ReaxFF_{C-2013}¹⁵ (red line) with Chenoweth¹² version ReaxFF_{CHO} (in blue) and DFT data (in black), for a cubic expansion of a diamond cell.

This new set of parameters looked really interesting and accurate. We therefore computed dissociation energy in graphene and diamond with this new set of parameters and compared them with those obtained with Pitman's implementation –see Figure 38.

^{xii} Fullerenes are carbon structure with a tube, sphere, or ellipsoid shape. Tube fullerenes are commonly called nanotubes of carbon. Carbon atoms in fullerenes are hybridized sp^2 .

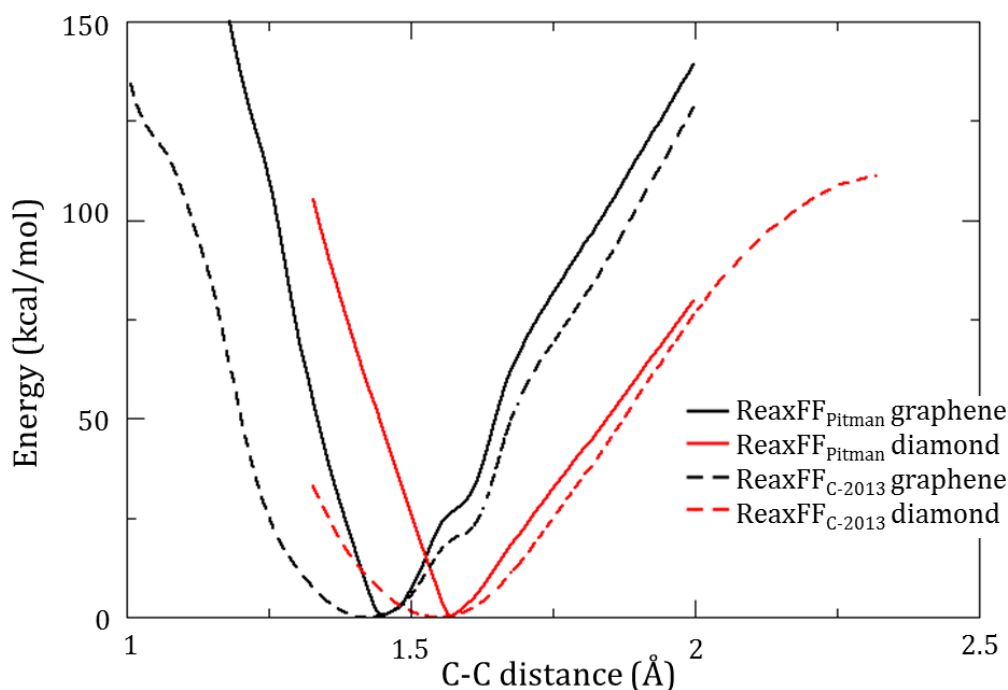


Figure 38. Dissociation energy at 0K in graphene (black lines) and diamond (red lines) with ReaxFF_{Pitman} (solid lines) and ReaxFF_{C-2013} (dashed lines). Curves are scaled at zero.

We indeed found continuous potential lines –and not the V-one obtained with previous versions. Equilibrium distances are also in better agreement. For graphene, ReaxFF_{C-2013} gives an equilibrium distance of 1.418Å and of 1.548Å for diamond -1.448 and 1.568 respectively with Pitman's. The first version of ReaxFF_{C-2013} was only taking carbon in account. A. Van Duin, from Penn State University, kindly provided us the version incorporating hydrogen and oxygen.

We calculated bond dissociation energy on different and well-known molecules, with ReaxFF_{C-2013}, such as ethane, ethylene, dihydrogen, propane, carbon monoxide and dioxide. Results are relatively close to the experimental values, for CO and H₂ for instance –see in Table 2. However, some bond dissociations are less well reproduced, as the C-O bond in carbon dioxide.

Table 2. Comparison of bond dissociation energies at 0K for CO, H₂, and CO₂ molecules, given by ReaxFF_{C-2013} and literature¹⁶.

Molecule	ReaxFF _{C-2013}	Literature ¹⁶
CO	11.35 eV	11.11 eV
H ₂	4.71 eV	4.48 eV
CO ₂	8.11 eV	5.45 eV

During our REMD runs, we realized the water gas shift reaction –which turns water and carbon monoxide into dihydrogen and carbon dioxide- was not well reproduced – see next chapter. Since our simulations were already well under way, we kept this set of parameters for the rest of our study.

Nevertheless, ReaxFF_{C-2013} is pretty well parametrized and appears as the best option we have. It offers a very satisfactory approach to study organic matter –mainly composed of carbon- decomposition. Besides, although some inaccuracies exist, the ReaxFF potential is the only one to be able to properly reproduce bond breaking and formation, in the sense that it described reaction intermediates, and to take into account carbon, hydrogen, and oxygen bonds. We thus chose to use this set of parameters for the present work.

Small cellulose crystal (420 atoms)

Once the potential has been successfully tested on comparable species –carbonaceous structures in particular- simulation method has to be tested as well. Indeed, REMD method requires important computational resources. It needs, at least, one processor per replica. The reason for that is that every replica has to run a NVT simulation and, moreover, communication between replicas is involved for temperature swaps. Besides, computation time increases with system size. As we are limited in computational resources, we have to test our method on a small system first.

Parameters

In the first chapter, we explained there are three main types of kerogens. Those types vary in their origins and compositions. As a first work, we chose to study cellulose as a model of organic matter. Indeed, cellulose is the main component of higher plants. Type III kerogen derives from those plants. Cellulose was thus an obvious precursor to test.

Cellulose is an organic polymer. Its formula is $(C_6H_{10}O_5)_n$. To build a cellulose system, we used a cellulose builder code published by Gomes and Skaf¹⁷ in 2012. Thanks to that code, we constructed a system of 420 atoms composed of:

- 120 carbon atoms,
- 200 hydrogen atoms and,
- 100 oxygen atoms.

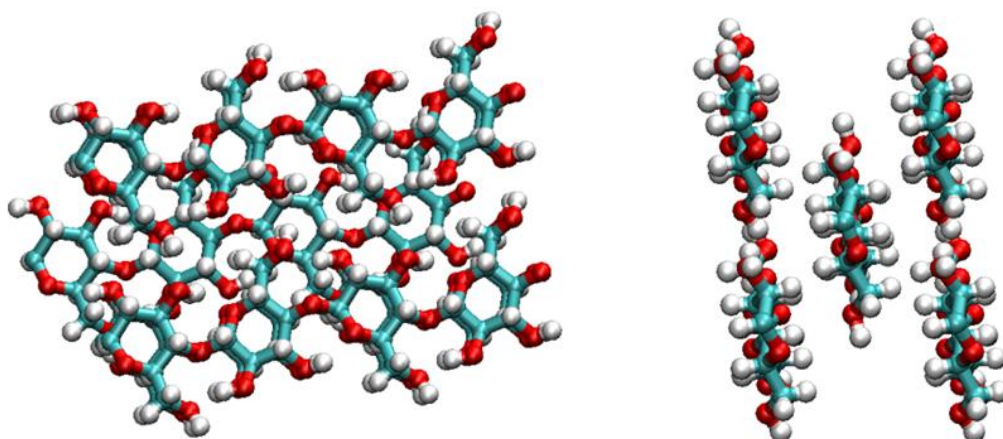


Figure 39. Snapshot of the cellulose system of 420 atoms.

We run some simple NVT simulations, at different temperatures, on this structure. The goal was to find the appropriate upper temperature. This temperature must be chosen high enough in order to have no replica trapped into a local energy minimum. At this temperature, every energy barrier has to be overcome. In the same time, it has to be not too high in order to not waste computation time in non-physical trajectories. We found that, to be sure to have a fast enough evolution of the system, the upper temperature had to be close to 3500K. This upper side of our temperature range was found suitable to quickly initiate the decomposition of the crystal. Our lowest temperature was set at 300K.

We tried different numbers of replicas -20, 33, 40 and 48- and found the best trade off was to use 33 replicas, from 300 to 3500K, with 100K between neighbor replicas.

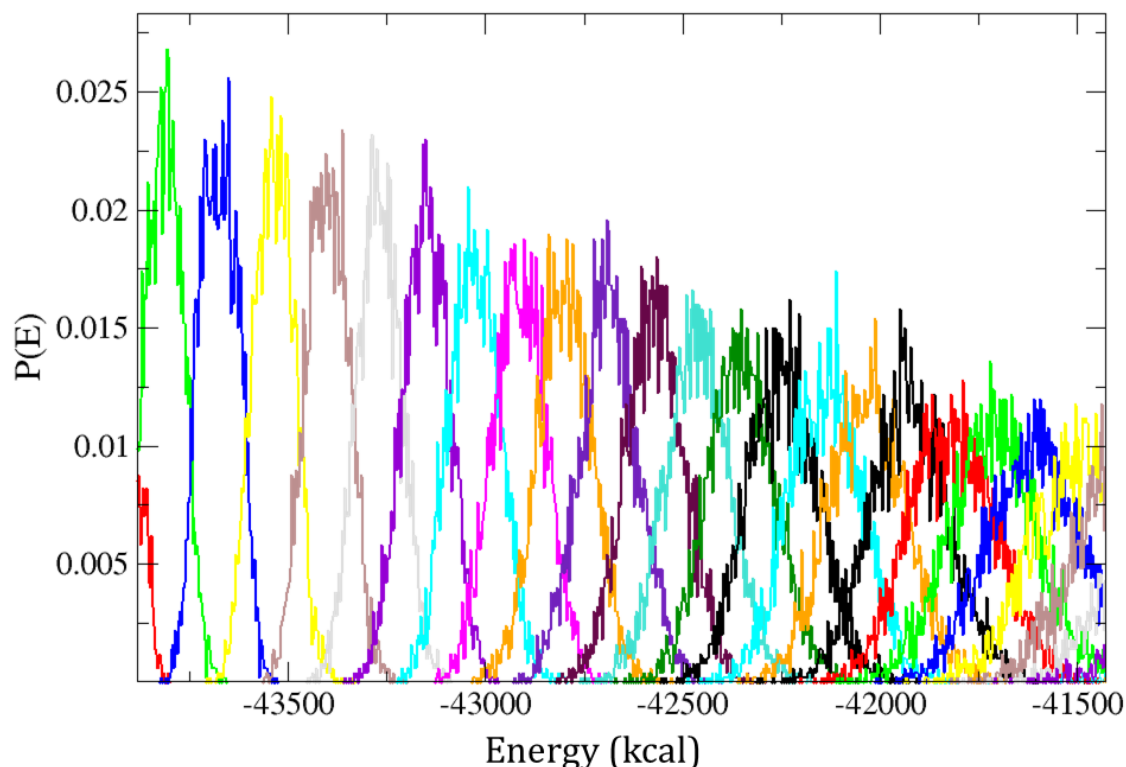


Figure 40. Energy distributions during REMD of 420 atoms cellulose systems. During this run, 33 replicas are used. Each distribution corresponds to a temperature.

Energy distributions during a REMD run on the cellulose system of 420 atoms, using 33 replicas, are given in Figure 40. Each distribution corresponds to a given temperature. One clearly notices overlaps between neighbor replicas. These overlaps allow temperature swaps between replicas during REMD and thus evolution of the lowest temperature system.

Results

We run 372 ps of REMD for this test. In only 372ps of REMD time, we noticed a significant evolution. Two main steps have been found. The first one is the fragmentation of initial cellulose structure. This first step goes along with production of small molecules, such as water, carbon monoxide and carbon dioxide. The second stage is the development of a carbon rich phase. This carbonaceous solid-like phase is poor in hydrogen and oxygen and keeps growing with time. It tends to organize in an aromatic structure.

A snapshot of this structure, at 300K, after 372ps of REMD is given in Figure 41. Even at the lowest temperature considered in the simulation, the initial cellulose crystal has

been decomposed. One can notice two phases: the carbon rich one, with ring structures, and the hydrogen and oxygen rich one, in which one can clearly see CO and H₂O molecules.

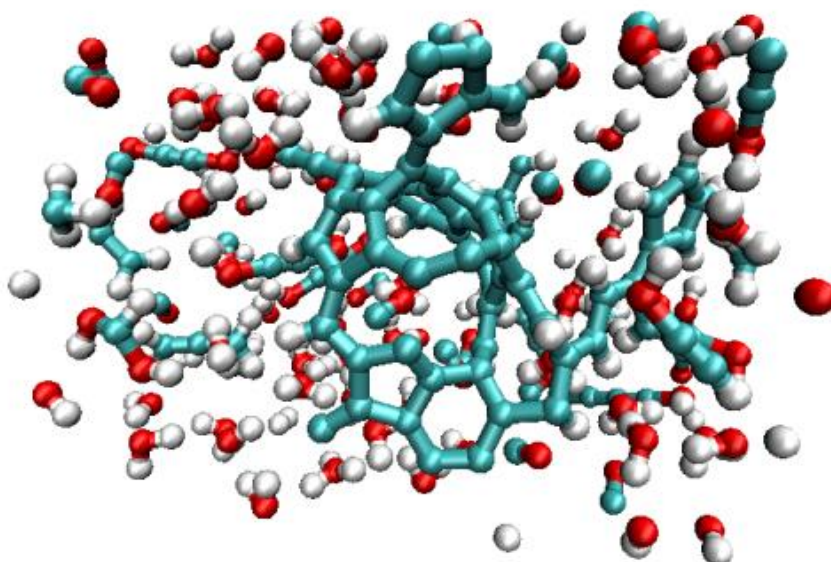


Figure 41. Snapshot of final structure of the cellulose system after 400ps of REMD. Structure at 300K. Carbon atoms are in blue, hydrogen atoms in white and oxygen atoms in red.

A short *ab initio* relaxation of the system after 372ps of REMD time was performed. The goal was simply to check its stability. During this relaxation, one only notes some bond relaxations and fluctuations but nothing breaks or forms. It supports the fact that the obtained structure is physically stable and possible.

Parameters for 4200 atoms (cellulose) system

A REMD test on a cellulose structure composed of 420 atoms has been performed. This simulation showed applicability of the method on such an organic system. We then wanted to extend our study to a larger cellulose system composed of 4200 atoms. This size increase has a direct impact on temperature distributions on replicas.

As discussed above, the acceptance probability depends on the overlap between the energy distributions of adjacent -in terms of temperature- replicas. This overlap depends on the spacing of temperatures and on the magnitude of the energy fluctuations $\Delta U = \sqrt{\langle U^2 \rangle - \langle U \rangle^2}$ where $\langle U \rangle$ is an ensemble average -i.e the internal energy of

the system. Indeed, energy distribution is narrower when system size is larger and at low temperatures. Statistical mechanics tells us that (i) ΔU scales as the inverse of the square root of the number of atoms ($\Delta U \propto \frac{1}{\sqrt{N}}$) and (ii) increases with temperature ($\Delta U \propto T$). In other words, (i) a much larger number of replicas will be needed for a 4200 atoms system than was necessary for a 420 atoms system and (ii) the spacing of temperatures has to increase with T.

We are also limited in computation capacity and need to limit the number of replicas. We thus need to find the appropriate distribution of temperatures, which addresses following criterions:

- Limits the number of replicas,
- Covers the range of temperature from 423 to 3500K,
- Replicas closer at low temperatures than at high ones, in order to obtain an overlapping distribution at all temperatures,.

Spacing between two replicas at high temperatures can be larger since energy distributions are wider. After few attempts, we found that proper overlapping can be achieved with 96 replicas and a temperature distribution following a geometric series formula:

$$T_{n+1} = T_n * q$$

With $T_0 = 423\text{K}$ and $q = 1.0225$ (q is the common ratio of the geometric series).

Energy distributions during a REMD run of the 4200 atoms cellulose systems are given in Figure 42. One clearly sees distributions overlaps. This indicates that temperature swaps are possible between neighbor replicas.

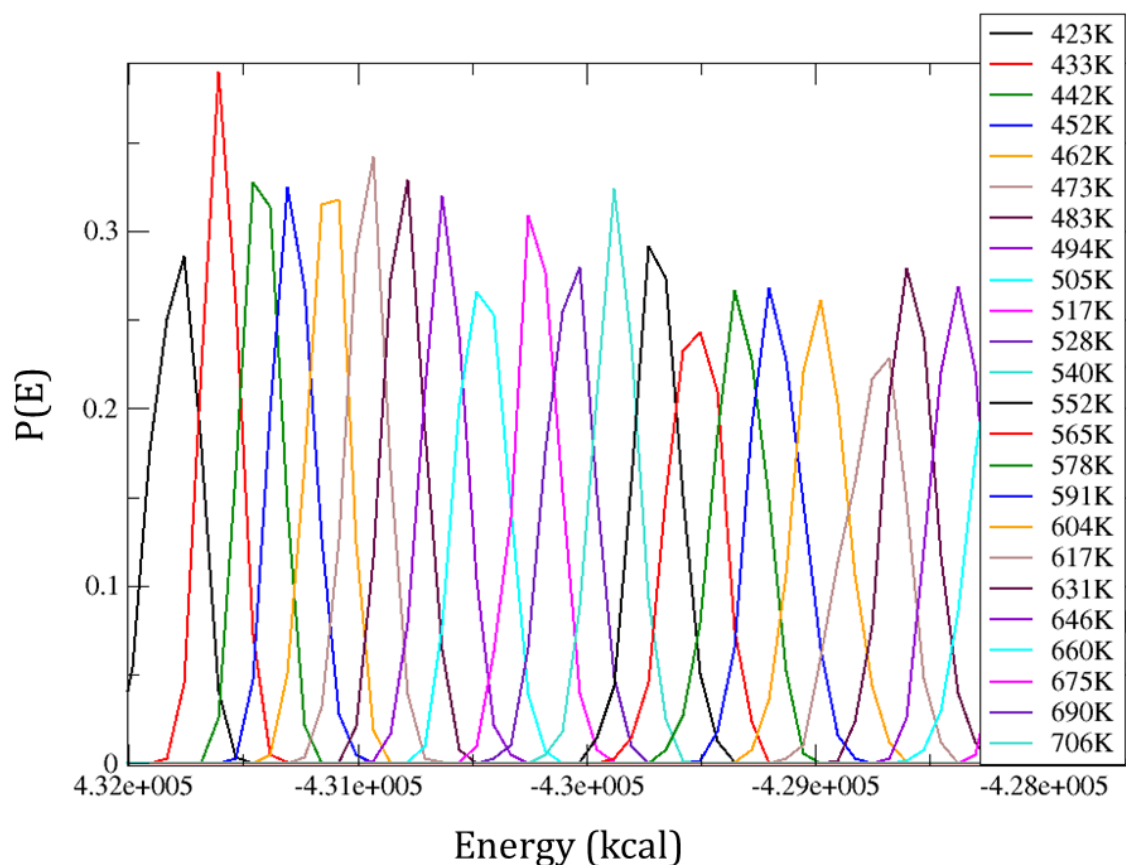


Figure 42. Energy distribution during REMD of 4200 atoms cellulose system, using 96 replicas. Each distribution corresponds to a temperature.

Summary

REMD technique requires important calculation resources. One cannot launch hundreds of thousands hours calculation without a first validation of methods. We briefly present in this section some tests we run in order to validate the interaction potential and the REMD method we wanted to use.

We first run simple NVT molecular dynamics and Monte-Carlo run on amorphous carbonaceous systems. We heated it at high temperatures -5000K and 3273K- using two different potentials: the tight binding potential of Amara developed at CINaM and the ReaxFF one. Resulting structures were really different, both on atomic organization and on carbon atoms hybridizations. Configurations obtained with the tight binding potential looked more physically probable. Therefore we tested implementations of ReaxFF in LAMMPS –in particular Pitman’s¹⁸ and Mattsson’s⁸ implementations.

We studied bond dissociation energies at 0K with ReaxFF in the case of different hybridizations C-C bonds and on simple molecules as H₂O, ethane, CO and CO₂. Resulting curves presented a V-shapes and too large equilibrium distances. These implementations of ReaxFF were not satisfying. In December 2014, a new set of ReaxFF parameters¹⁵ was published, offering a more accurate description of C-C bonds. After some verification we decided to use this new version of ReaxFF, called ReaxFF_{C-2013}.

Replica exchange molecular dynamics method was tested on a small cellulose system, composed of 420 atoms. First, temperature range and distribution were tried out and defined. REMD simulation was then run, using 33 replicas, going from 300K to 3500K. In this configuration, energy distributions well overlap with their neighbor replicas. We run 400ps of REMD and, looking at the final configuration at the lowest temperature - 300K- we noticed formation of two different phases. The first one was composed of small molecules as H₂O, CO and CO₂ while the second one was a carbon rich structure. Those first results incited us to continue on a larger system of 4200 atoms.

We run some REMD short simulations in order to find the appropriate number and distributions of replicas. We found that 96 replicas distributed following a geometric series, starting at 423K and with a common ratio of 1.0225 was an optimal choice. It allows every replica to overlap with its neighbors, for a reasonable CPU time.

1. Stillinger, F. H. & Weber, T. A. Computer simulation of local order in condensed phases of silicon. *Phys. Rev. B* **31**, 5262–5271 (1985).
2. Tersoff, J. Empirical Interatomic Potential for Carbon, with Applications to Amorphous Carbon. *Phys. Rev. Lett.* **61**, 2879–2882 (1988).
3. Brenner, D. W. Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamonds films. *Phys. Rev. B* **42**, 9458–9471 (1990).
4. Stuart, S. J., Tutein, A. B. & Harrison, J. A. A reactive potential for hydrocarbons with intermolecular interactions. *J. Chem. Phys.* **112**, 6472–6486 (2000).
5. Van Duin, A. C. T., Dasgupta, S., Lorant, F. & Goddard, W. A. ReaxFF: A reactive force field for hydrocarbons. *J. Phys. Chem. A* **105**, 9396–9409 (2001).
6. Amara, H., Roussel, J. M., Bichara, C., Gaspard, J. P. & Ducastelle, F. Tight-binding potential for atomistic simulations of carbon interacting with transition metals: Application to the Ni-C system. *Phys. Rev. B - Condens. Matter Mater. Phys.* **79**, 1–17 (2009).
7. Shan, T.-R., Devine, B. D., Kemper, T. W., Sinnott, S. B. & Phillpot, S. R. Charge-optimized many-body potential for the hafnium/hafnium oxide system. *Phys. Rev. B* **81**, 125328 (2010).
8. Mattsson, T. R. *et al.* First-principles and classical molecular dynamics simulation of shocked polymers. *Phys. Rev. B - Condens. Matter Mater. Phys.* **81**, 1–9 (2010).
9. Gale, J. D. GULP: A computer program for the symmetry-adapted simulation of solids. *J. Chem. Soc. {,} Faraday Trans.* **93**, 629–637 (1997).
10. Brenner, D. W. *et al.* A second-generation reactive empirical bond order (REBO) potential energy expression for. *J. Phys. Condens. Matter* **14**, 783–802 (2002).
11. Liang, T. *et al.* Classical atomistic simulations of surfaces and heterogeneous interfaces with the charge-optimized many body (COMB) potentials. *Mater. Sci. Eng. R* **74**, 255–279 (2013).
12. Chenoweth, K., Van Duin, A. C. T. & Goddard, W. A. ReaxFF Reactive Force Field for Molecular Dynamics Simulations of Hydrocarbon Oxidation. *J. Phys. Chem.* **112**, 1040–1053 (2008).
13. van Duin, A. C. T. *ReaxFF training program at NASA AMES.* (2006).
14. Hooper, A. B. & Lipscomb, J. D. Bond Dissociation Energies. Available at: <https://labs.chem.ucsb.edu/zakarian/armen/11---bonddissociationenergy.pdf>.
15. Goverapet Srinivasan, S., Van Duin, A. C. T. & Ganesh, P. Development of a ReaxFF potential for carbon condensed phases and its application to the thermal fragmentation of a large fullerene. *J. Phys. Chem. A* **119**, 571–580 (2015).
16. Darwent, B. *Bond dissociation energies in simple molecules.* (1970).
17. Gomes, T. C. F. & Skaf, M. S. Cellulose-builder: A toolkit for building crystalline structures of cellulose. *J. Comput. Chem.* **33**, 1338–1346 (2012).
18. Pitman, M. C. & Duin, A. C. T. Van. Dynamics of Confined Reactive Water in Smectite Clay – Zeolite Composites. *J. Am. Chem. Soc.* **134**, 3042–3053 (2012).

Chapter 4: Results

This chapter is dedicated to the results obtained using *Replica Exchange Molecular Dynamics* to study organic matter decomposition under geological conditions. We studied two type III kerogen precursors, cellulose and lignin, that are both main components of terrestrial plants.

Type III Kerogen precursor: Cellulose

The first part of our work was the study of a cellulose system decomposition under geological conditions. In this section, we first give an overview of cellulose, its composition, and structure. We then explain how we constructed a numerical structure in our simulation. Step one of our work is the decomposition of the whole organic system at relevant (T,P) conditions, along with the formation of two different phases: a solid carbon rich phase and a hydrogen rich liquid phase. Simulation of the second stage is the REMD of the solid structure, formed during the first step. We end this section with a comparison between our results and experimental data.

Choice of cellulose as a type III kerogen precursor

Cellulose is the main constituent of plants and is the most abundant organic matter on Earth^{xiii}. Plants produce between 50 and 200 Gt per year. It is used in a wide variety of fields, such as papers, tissues, and energy with fire wood or biofuel.

Cellulose is a natural polymer, composed of linear chains of glucose molecules rings. Cellulose linear chains are composed of repeated monomers –see in Figure 43. Those units are cellobiose, it means two anhydroglucose elements linked through an oxygen bridge. This bond is called β 1-4. Cellulose is thus composed of carbon, hydrogen and oxygen atoms. Its formula is $(C_6H_{10}O_5)_n$, where n corresponds to the degree of polymerization and depends on the cellulose source. In our work we start from a system with a degree of polymerization of 4, to which we apply periodic conditions. Different polymorphs of cellulose exist –I, II, III and IV- but only the first one –I- interests us since it is the natural one –produced by natural organisms. Cellulose I density is about $1.5\text{g}\cdot\text{cm}^{-3}$. Its H/C and O/C ratios are 1.66 and 0.83 respectively.

^{xiii} Cellulose represents up to 50% of the biomass⁵⁷.

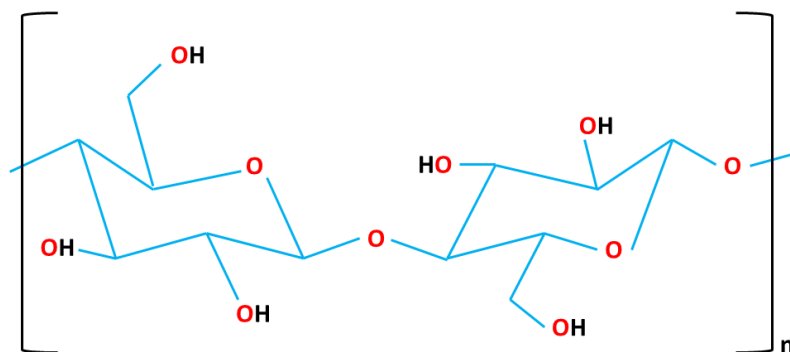


Figure 43. Monomer of cellulose. Blue lines correspond to carbon bonds, oxygen atoms are in red and hydrogen atoms are in black.

Cellulose I has two polymorphs. Its structure can be triclinic -cellulose called $I\alpha$ - or monoclinic - $I\beta$ ¹. $I\alpha$ cellulose is more common in algae and bacteria while $I\beta$ is the main component in higher plants².

As we mentioned in the chapter 1 -context- type III kerogen mainly results from higher plants. Since cellulose $I\beta$ is the major constituent of higher plants cells, it appears as an obvious representative of type III kerogen precursor.

Building of cellulose system

Cellulose structure is well known. In order to run simulations one needs to build an initial numerical configuration.

To build this initial cellulose configuration, we used the cellulose-builder code³. It is a program that generates Cartesian coordinates for all atoms of a specified cellulose structure. It offers the possibility to build cellulose polymorphs $I\alpha$, $I\beta$, II and III. Data used in cellulose-builder are taken from studies of Nishiyama and co-workers who studied structures of the fourth polymorphs⁴⁻⁷. It is a free downloadable code^{xiv}.

^{xiv} To download the code: <https://code.google.com/archive/p/cellulose-builder/>

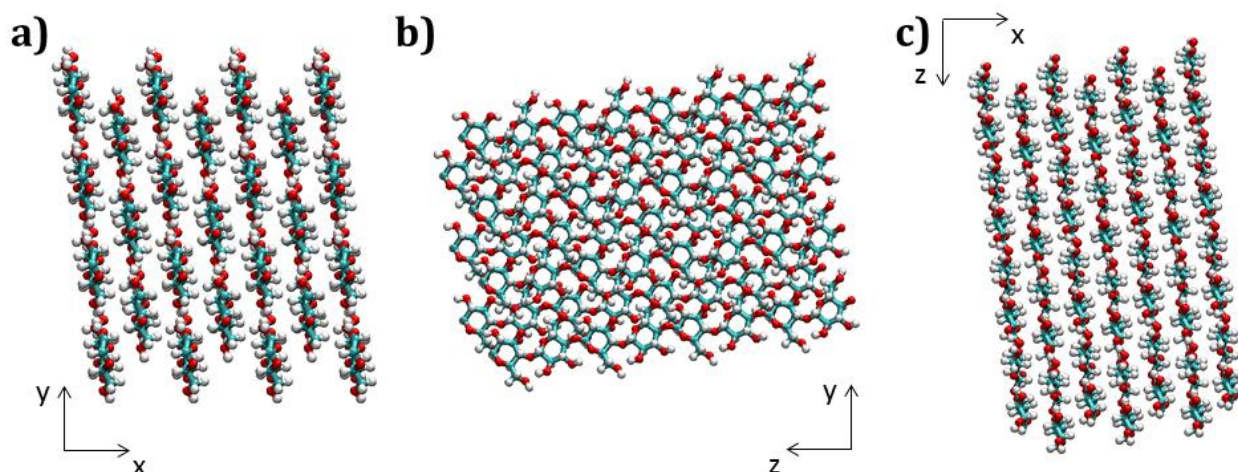


Figure 44. Initial I_β cellulose structure. View on planes (xy) (a), (yz) (b), and (zx) (c).

To obtain a structure, one only has to specify the phase - I_α , I_β , II or III- and three integers. Those three numbers correspond to the number of unit cells to replicate in each crystallographic orientation.

With the cellulose-builder code, we constructed a I_β cellulose structure of 4200 atoms -1200 carbon atoms, 2000 hydrogen atoms and 1000 oxygen atoms- and with a density of 1.29g.cm^{-3} . Upon thermal degradation, we will see that crystalline order quickly disappears with cellulose fragmentation and system relaxations.

Confined Pyrolysis

As we explained in Chapter 3 –Validation- we first performed an REMD simulation on a smaller cellulose system –composed of 420 atoms. This first test demonstrated us that the REMD method coupled with the ReaxFF potential was an appropriate method to study cellulose decomposition. After only 400ps of REMD and even at the lowest temperature considered in the simulation -300K-, the initial cellulose crystal was decomposed to form a carbon-rich solid / O/H-rich fluid equilibrium. The solid phase was made of sp^2 carbon, both under the form of aromatic rings and aliphatic chains, typical features of chars. On the other side, the fluid essentially consists of small oxygen containing molecules like water, CO_2 , and CO. We describe here a larger simulation, of cellulose with 4200 atoms. The same initial cellulose crystal is stable when submitted to a classical NVT run at a constant temperature of 300K.

Simulation on 4200 atoms cellulose system was performed in two stages. In a first step, the whole crystal was submitted to a REMD simulation during which the system followed the same way of decomposition than the small system. In a second step, REMD

was only applied to the “solid” part in order to obtain a kerogen-like structure. We explain these different steps before comparing our results with experimental ones.

Parameters

As we explained in chapter 3 –Validation- we found a proper overlapping between neighbor replicas of REMD for 4200 atoms cellulose system, using 96 temperatures distributed following a geometric series. This series starts at 423K and has a common ratio of 1.0225. Temperatures are comprised in the range [423K:3500K] –see chapter 3.

In classical statistical mechanics, the equipartition theorem states that, at thermal equilibrium, the energy of a system is equally shared by its components. Kinetic energy E_k can therefore be defined by the following equation:

$$\frac{E_k}{N} = \frac{3}{2} k \cdot T = \frac{1}{2} m \cdot v^2$$

Where E_k is the kinetic energy of a given system, N the number of particles, k is Boltzmann’s constant, equal to $1.38 \cdot 10^{-23} \text{ J.K}^{-1}$, T is the temperature of the system, m is the system mass and v the average particles velocity.

Let’s consider a water molecule at $T=3500\text{K}$ –upper temperature in our simulation. Its mass is $m = 8.97 \cdot 10^{-26} \text{ kg}$. With these data, we can compute a molecule average velocity of $1.27 \cdot 10^{-2} \text{ Å.fs}^{-1}$. It means that, between two temperature swap attempts separated by 10fs, the molecule will have moved by 1.27 Å in average. Since swaps are not accepted at each attempt, particles have time to move and structure can evolve between two swaps. Temperature swaps between replicas are thus attempted every 10fs of REMD. Every time, swaps for every replica are attempted. This frequency seems therefore to be adequate. In our cellulose system, acceptance rate is of 1/23 for the 423K replica, of 1/12 for the 3500K replica, and in average between 1/7 and 1/4 for other replicas. The lowest and the highest replicas have only one neighbor replica with which they can swap. Therefore, their acceptance ratio is lower than for other replicas.

First step: REMD on the whole system

Once initial configuration and parameters are set, simulations of cellulose decomposition under geological conditions can be launched. REMD runs alternate with NPT MD relaxation. These relaxations help to keep the system in a geological pressure range [25:125MPa]. Pressure and energy evolution during simulation are represented in Figure 46. These values are the ones of the lowest temperature -423K- which is the one we are interested in because it corresponds to the geological temperature in the ground,

around 4 and 5km depth. A diagram of the numerical simulations scheme applied to our cellulose model during this first stage of REMD is given in Figure 45.

We run REMD simulation on the whole cellulosic system for 750ps. Note that this time lag does not correspond to classical time in molecular dynamics since swaps between adjacent temperature replicas are used to speed up evolution. Therefore, although decomposition pathway and structures correspond to the natural ones, there is no equivalence rule linking simulation time to real time. We try here to simulate millions of years processes taking place in nature.

Final equilibrium state should correspond to the natural one since the goal of REMD simulations is to lead the system towards its thermodynamic equilibrium, but the followed kinetic path, does not necessarily correspond to the real one. Nevertheless, the route followed by the system, connecting initial configuration to final equilibrium state is supposedly thermodynamically accurate⁸.

From Figure 46, one can notice that energy is decreasing with time. This decrease means that the system is evolving towards a more stable configuration at the given temperature -423K. It also indicates an exothermic decomposition process. After 750ps, the system seems to have reached equilibrium. We noticed a two-phase separation, with a fluid and a solid-like phases. The second phase structure is close to a kerogen configuration –see below. Thus, we decided to stop this first stage of our simulation after 750ps of REMD and to focus on the solid phase only. At this point, the energy curve reaches a plateau, that can be considered as an indication of convergence to an equilibrium state.

Note that REMD method was not easily applicable in isotherm-isobaric –NPT-ensemble with LAMMPS. Exchange probabilities are calculated from ratios of Boltzmann probabilities between states in adjacent replicas, that are only valid for NVT. For NPT, new terms of PV need to be added. Therefore, we conducted REMD run with replicas in the canonical ensemble –NVT only. We thus tried to maintain our system pressure in a geological range of [50-125MPa] by the way of NPT relaxations, during which no swaps were attempted. Between two relaxations, pressure evolves thanks to the system decomposition.

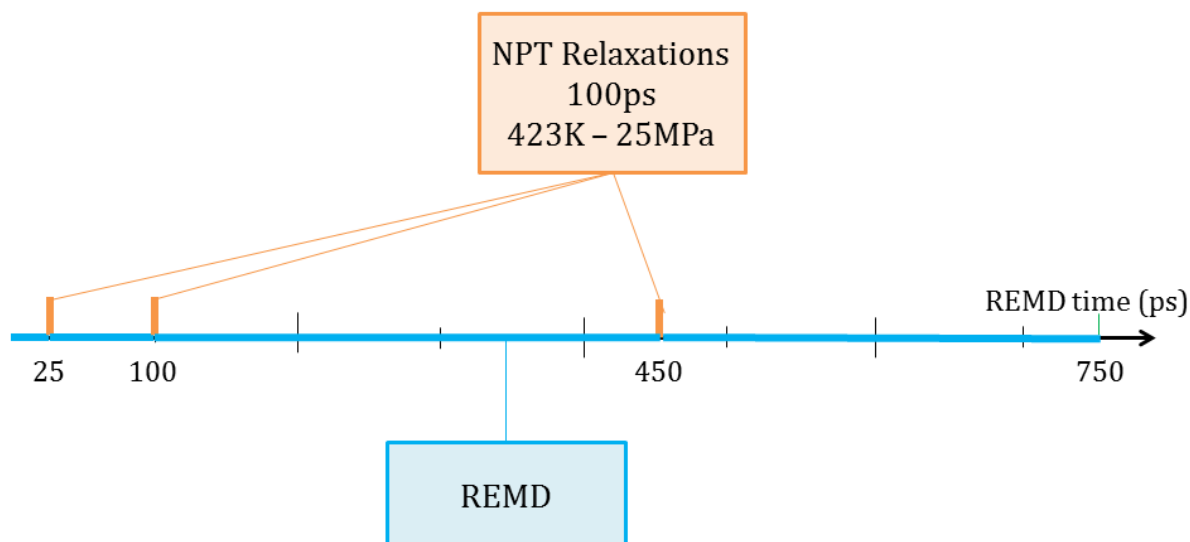


Figure 45. Diagram of the numerical simulations applied to the crystal of cellulose during the first stage of REMD. This stage of REMD is 750ps long and NPT relaxations are performed after 25, 100, and 450ps of REMD.

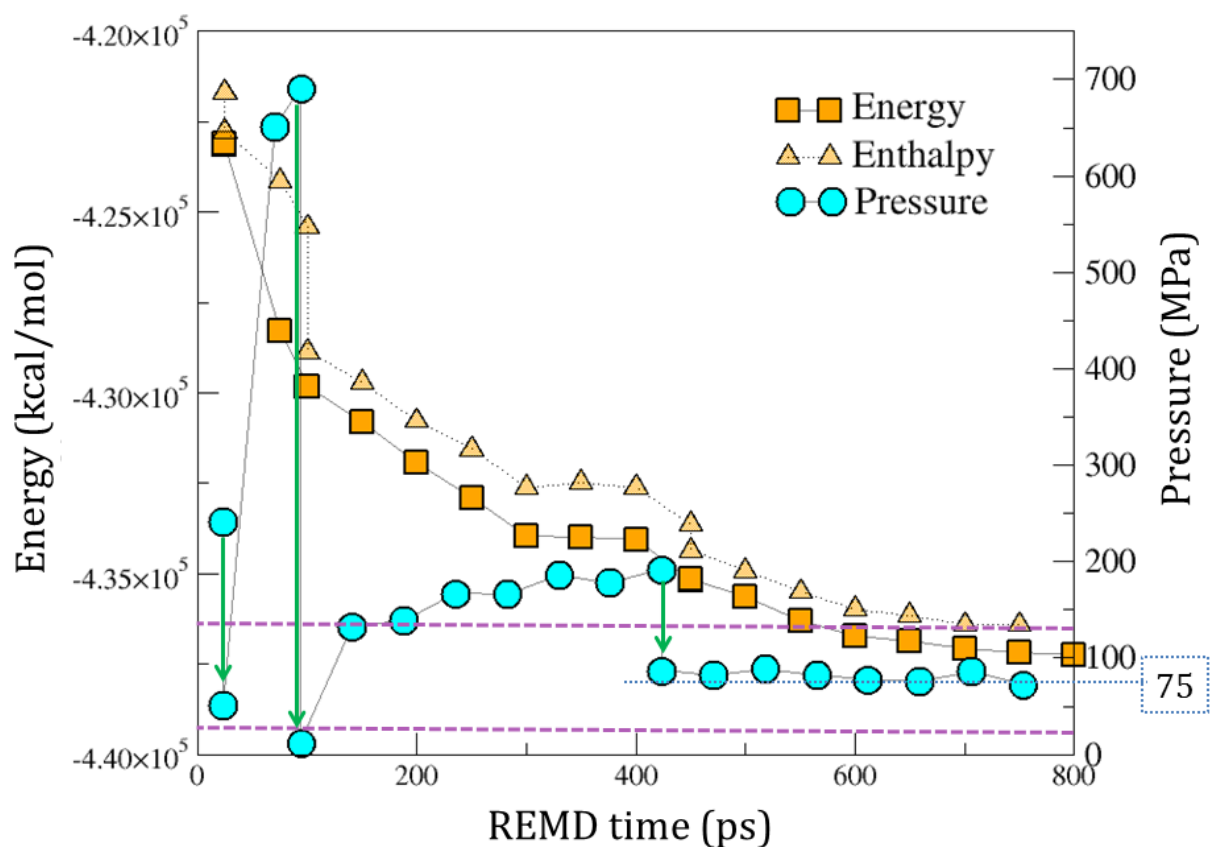


Figure 46. Whole cellulosic system energy (orange squares), enthalpy (orange triangles), and pressure (blue circles) evolution during REMD simulation, for the replica at 423K. Violet dashed lines represent limits of geological pressure between which we try to keep our system. Green arrows represent NPT relaxations.

System significantly evolves during this REMD simulation, as we can see in Figure 47. Indeed, one can notice a great difference between the initial system -Figure 47 a)- and the system after 750ps of REMD -Figure 47 b). At the later stage, system is more disordered –compared to the crystallographic arrangement at the beginning of the simulation.

During this first stage, a two-phase separation appeared. Indeed, two well defined systems have been formed from the initial cellulose crystal. This formation followed a two-stage pathway. We now develop further these two aspects of decomposition.

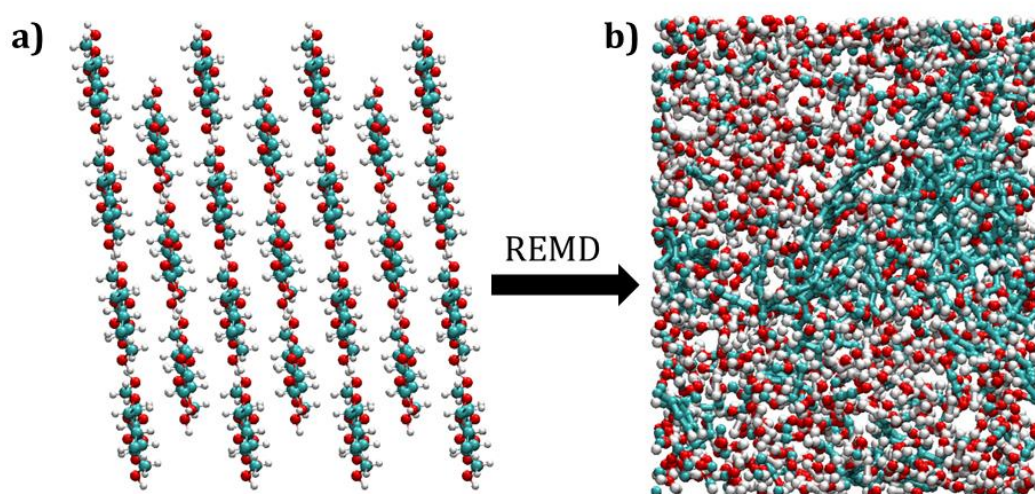


Figure 47. Cellulosic system at the initial step (a) and after 750ps of REMD time (b).

Two-phase separation

After 750ps –and since around 500ps of REMD- we have reached a stage where we clearly have a two-phase system -see in Figure 48. The first phase is a solid one. It is mainly composed by carbonaceous molecules. Their H/C and O/C atomic ratios are low, 0.55 and 0.04 respectively -i.e. typical of mature kerogens⁹⁻¹¹. The second phase is a fluid one and is hydrogen and oxygen rich, with H/C and O/C ratios of 3.14 and 1.89 respectively.

As shown in Figure 48, molecules with less than 12 carbon atoms are considered as part of the fluid phase while molecules with 12 or more carbon atoms are part of the solid phase. This choice results from the fact that a separation occurs in molecule sizes around 12 carbon atoms. Indeed, all along the simulation, a number of C₀, C₁, C₂, C₃, C₄, C₅, C₆, and C₇ molecules exist, with a large proportion of hydrogen and oxygen, while, for instance, after 500ps of REMD, there is no molecule with a number of carbon atoms

between 12 and 30, after 650ps, no molecules between C_{11} and C_{48} , and after 750ps, no molecules with a number of carbon atoms between 13 and 235. A separation occurs around a number of carbon atoms between 11 and 13. Above, the number of carbon atoms increases and the content of hydrogen and oxygen decreases. The differentiation is thus not done on a basis of crystalline order but on the number of carbon atoms, since the kerogen is known to be a carbon-rich phase.

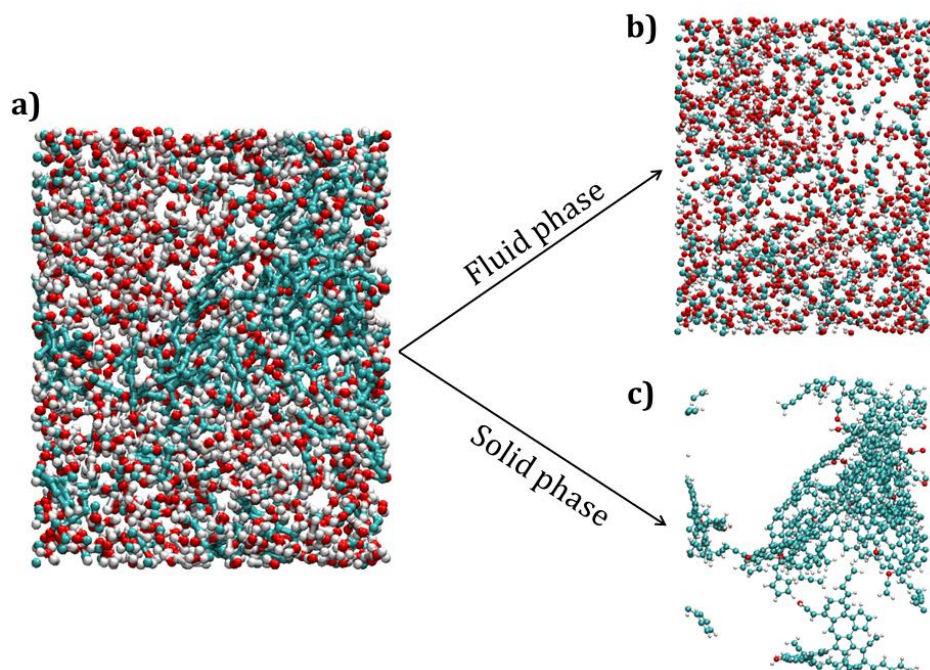


Figure 48. Two-phase system after 750ps of REMD. **a)** snapshot of the whole system, **b)** snapshot of the fluid phase, and **c)** snapshot of the solid phase. Carbon atoms are blue, hydrogen atoms white and oxygen atoms red.

In Figure 48, snapshots of the two phases are shown. One clearly sees a large concentration of carbon atoms in the solid phase and of hydrogen and oxygen in the fluid one. The fluid phase is composed of small molecules –such as H_2O , H_2 , CO and CO_2 – while the solid phase is more compact and regrouped in large clusters. This carbonaceous system can be related to a kerogen structure.

In order to identify the different molecules, we developed a deep first search –or depth first search– algorithm¹²-based code. This algorithm starts from a first atom – considered as the root of the search– and explores neighbor atoms as far as possible. Once all the neighbors –and neighbors of neighbors– have been explored –the whole being considered as a molecule– the search restarts from another atom that has never been explored yet, and a new molecule is hence defined. The idea of neighbor is based on the interatomic distance. This criterion is set up using a radial distribution function. For instance, the cutoff distance below which two carbon atoms are considered as neighbors is of $1.9\text{\AA}$.

Two-stage decomposition

1.1.1.1.1 Cellulose fragmentation: small molecules formation

Fragmentation of the cellulose crystal is the first step of decomposition and starts at the very beginning. After only 25ps of REMD no cellulose unit remains. We remind that no equivalence rule exists between simulation and real times. Although 25ps can seem really short, decomposition of cellulose is already well underway. At this point, 38% of the fragments do not contain any carbon atoms and 9% contain only one carbon atom. The first proportion is composed by water and H_2 molecules. The loss of all this water by cellulose is called cellulose dehydration. Half of the fluid phase has almost no carbon atoms. Carbon is clearly not the major component of this phase.

Concerning the fluid composition, in Figure 49, evolution of the amount of the main molecules constituting the fluid phase: H_2O , CO , CO_2 and H_2 is shown. During initial steps of the simulation, we observe a strong increase of the number of the small molecules considered here as the fluid phase -essentially water and carbon monoxide. Cellulose dehydration is obvious from this plot.

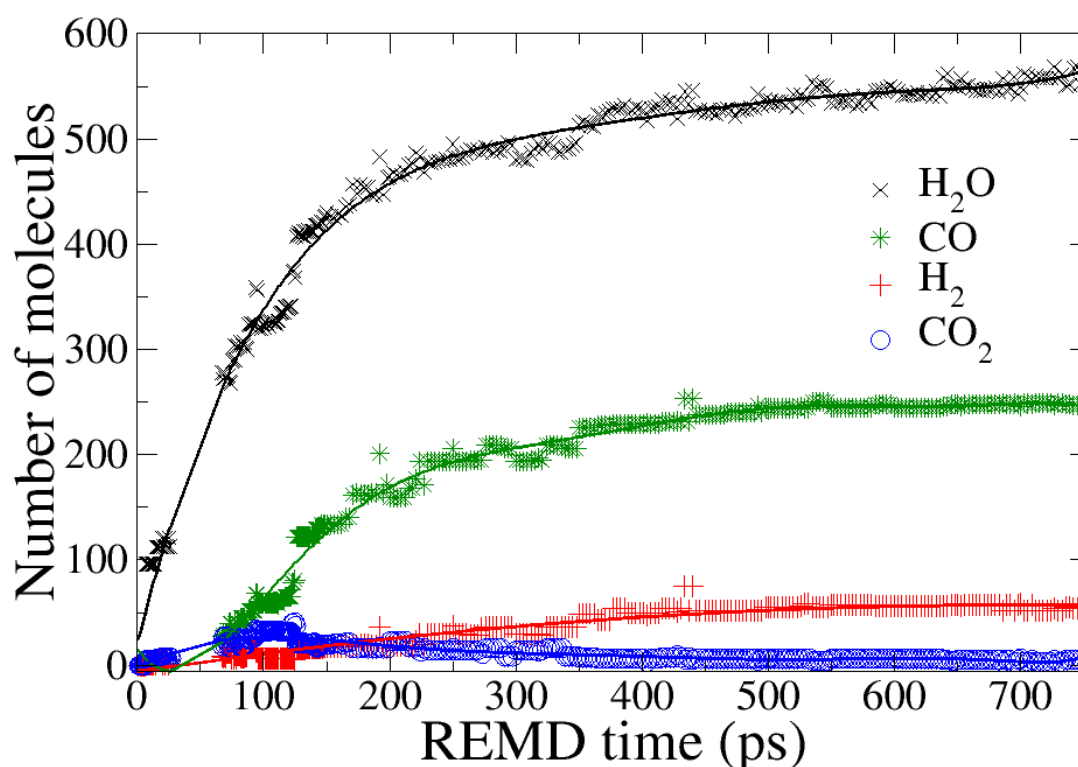


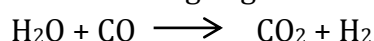
Figure 49. Evolution of small molecules during cellulose decomposition REMD simulation. H_2 in red, H_2O in black, CO in green, and CO_2 in blue.

From 400 and 500 REMD ps, evolutions of small molecules reaches a plateau. At the end of this first simulation step -750ps- 88% of the fluid phase is composed by H₂, H₂O, CO and CO₂. Other molecules compose this phase –see in Table 3. A number of acetylene-based alcohols as acetylenediol -30 molecules- and ethynol -9 molecules- are present. There are also 14 hydronium ions at 423K. In Table 3, numbers of molecules are given at different temperatures. Between 423K and 1407K, differences exist but no clear trend can be seen. Dihydrogen proportion clearly increases with temperature while acetylenediol decreases. For instance, we notice an effect of temperature on effluents. We notice that at 1407K, CO₂ does not exist.

Table 3. Fluid phase composition at different temperatures, after 750ps of REMD.

Molecule	Usual name	423K	604K	901K	1407K
H ₂ O	Water	554	563	552	577
CO	Carbon monoxide	248	245	245	302
H ₂	Dihydrogen	55	57	59	155
C ₂ H ₂ O ₂	Acetylenediol	30	28	29	20
H ₃ O ⁺	Hydronium	14	11	13	17
C ₂ H ₃ O	Ketenium	12	10	12	3
C ₂ H ₂ O	Ethynol	9	12	10	15
CH ₄ O ₂	Methanediol	8	6	7	2
CH ₄ O	Methanol	6	9	9	3
CO ₂	Carbon dioxide	5	7	7	0
CH ₄	methane	5	5	5	6

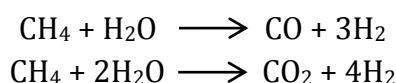
As it can be seen in Figure 49, H₂O and CO are far more represented in the fluid phase than CO₂ and H₂. Fluid equilibrium seems to go against the water gas shift reaction:



Experimentally¹³⁻¹⁵, its reaction enthalpy is comprised between -9.8 and -8.3 kcal.mol⁻¹. Following this reaction, water and carbon monoxide should react in order to form dihydrogen and carbon dioxide. In our system, it does not happen since we have a large overrepresentation of H₂O and CO -802 molecules- compared to CO₂ and H₂ -60 molecules. ReaxFF implementation that we use in this work –ReaxFF_{C-2013}- does not reproduce well this reaction. It indeed presents a reaction enthalpy^{xv} of +29.23kcal.mol⁻¹. This implies the opposite reaction than the one expected.

^{xv} To compute reaction enthalpy, we minimized systems of CO+H₂O and CO₂+H₂ using the *fire* algorithm in LAMMPS, with tolerances on energy and force of 1.0e-8.

Besides the water gas shift reaction, we checked how the ReaxFF_{C-2013} potential reproduced other essential reactions of the steam reforming. Steam reforming¹⁶ consists of the water gas shift reaction combined with the following two:



Their experimental reaction enthalpies¹⁶ are respectively of +49.24 kcal.mol⁻¹ and +39.44 kcal.mol⁻¹. Reaction enthalpies of these reactions computed using ReaxFF_{C-2103} are different but of the same order of magnitude than experimental ones: +61.06 kcal.mol⁻¹ and +80.05 kcal.mol⁻¹ respectively. Values for the three reactions of steam reforming are given in Table 4. More important, both reactions are reproduced in the appropriate way.

Table 4. Comparison of experimental¹³⁻¹⁶ of reaction enthalpies and reaction enthalpies computed with ReaxFFC-2013 for the three reactions of steam reforming.

Reaction	Experimental (kcal.mol ⁻¹)	ReaxFF _{C-2013} (kcal.mol ⁻¹)
$\text{H}_2\text{O} + \text{CO} \longrightarrow \text{CO}_2 + \text{H}_2$	-9.8 / -8.3	+29.23
$\text{CH}_4 + \text{H}_2\text{O} \longrightarrow \text{CO} + 3\text{H}_2$	+49.24	+61.06
$\text{CH}_4 + 2\text{H}_2\text{O} \longrightarrow \text{CO}_2 + 4\text{H}_2$	+39.44	+80.05

Although water gas shift reaction is not well reproduced by the interaction potential, fluid composition may correspond, at least in part, to what is experimentally observed. In Figure 50 is given the evolution of product yield –CO, H₂O and CO₂– with increasing temperature during a cellulose pyrolysis¹⁷ taken from a paper of Hajaligol *et al.*¹⁷ who studied the primary decomposition of cellulose by pyrolysis. Between 573 and 873K, light volatiles are dominated by H₂O and some CO, CO₂, and light oxygenated liquids, but above 1023K CO is by far the most abundant gaseous product. H₂O was the first product to form, around 603K followed by CO₂ around 673K and CO around 813K. Around 423K, water is the most represented product. In our study –see in Figure 49– during the first picoseconds –until around 75 REMD ps– there are more CO₂ molecules than CO ones. Afterwards, this tendency reverses and the proportion of CO largely increases while that of CO₂ one smoothly decreases. After 750ps of REMD, there are fifty times as many CO as CO₂. This trend also corresponds to the one found by Hajaligol *et al.* Although the evolution in this experiment is driven by an increasing temperature –while in our study it goes with time– product proportions indicate that the water gas shift reaction might not always be predominant. This correspondence supports our results.

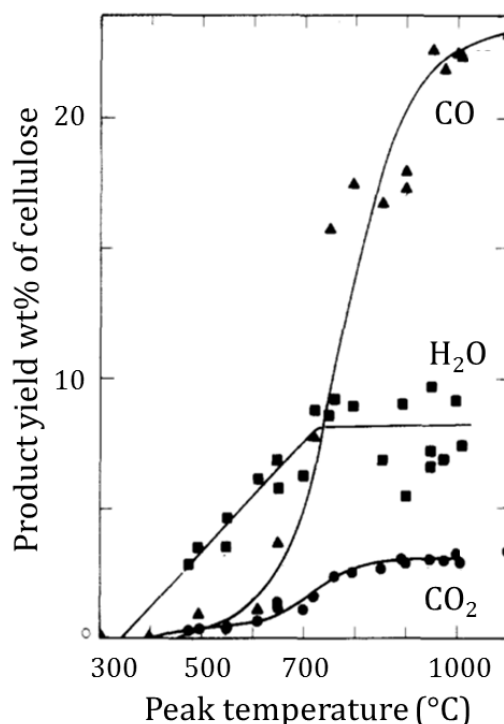


Figure 50. Yields evolution of CO, CO₂, and H₂O from cellulose pyrolysis with temperature, from Hajaligol *et al.* (1982)¹⁷.

In our simulation, the fact that the system is confined might lead to reverse the water gas shift reaction. This idea has to be addressed with a potential that better reproduces the water gas shift reaction^{xvi}. A better version of ReaxFF has been recently provided by A. van Duin –ReaxFF₂₀₁₆.

Formation of the fluid phase induces a significant increase of the system pressure. We relaxed the system by performing NPT simulations, after 25, 100 and 450 REMD ps –as represented by green arrows in Figure 46– with the goal to keep the pressure at a geological value –between 25 and 125 MPa –violet dashed lines in Figure 46.

After approximately 100 REMD ps, a carbon-rich macromolecular phase^{xvii} starts to be observed. It starts by fairly small fragments, with 20 carbon atoms and grows with time.

^{xvi} A new version of ReaxFF –ReaxFF₂₀₁₆– which seems to well reproduce the water gas shift reaction, with a reaction enthalpy of -17.92 kcal.mol⁻¹, is under development.

^{xvii} We consider as carbon-rich phase, or solid-like phase, molecules containing 12 or more carbon atoms.

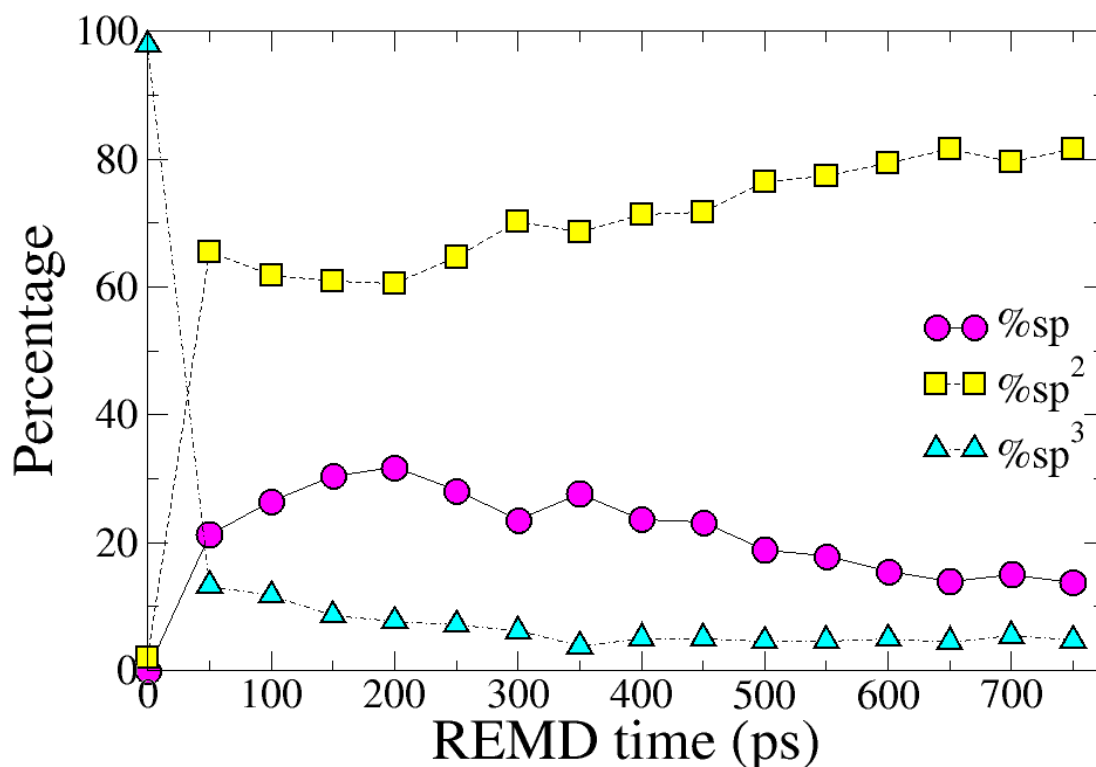


Figure 51. Evolution of carbon atoms hybridization in the solid part during REMD simulation. Percentage of sp hybridized carbon atoms (magenta circles), of sp² hybridized carbon atoms (yellow squares), and of sp³ hybridized carbon atoms (cyan triangles).

In Figure 51, we can see evolution of the hybridization of carbon atoms in the solid-like phase -note that we define hybridization on the sole basis of its number of neighbors. sp, sp² and sp³ carbon atoms are considered to have two, three and four neighbors respectively. In cellulose, all carbon atoms are sp³. In the first stages of the decomposition, bonded carbon atoms lose some neighbors and change their hybridization state. We can think that this is mainly due to the formation of water -see in Figure 49. After that, the solid part of the system seems to slowly become more and more composed of sp² aromatic carbon atoms. Cellulose decomposes directly to a sp² structure, with rings, without passing by an immature step composed by sp³ chains.

1.1.1.1.2 Coalescence of the solid-like part

The fluid phase is mostly formed during the first 200ps of the REMD, and thanks to the cellulose crystal fragmentation. The number of small molecules such as H₂O and CO quickly increases at the beginning before reaching a plateau. While the fluid phase is formed at the beginning, the solid one starts to grow right after the fluid phase starts to reach its equilibrium structure and composition.

In Figure 52, percentage of carbon atoms taking part of the solid-like phase is given as a function of simulation time. At the initial stage, all carbon atoms are in the solid phase since the system is the cellulose crystal. Once the simulation is launched, this initial system is fragmented and the number of carbon atoms into the solid part -defined as all the molecules composed by at least 12 carbon atoms- drops to 5.7%, at 100ps. Percentage of carbon in the solid-like phase then starts to progressively increase until it reaches a plateau between 55 and 60% of the total number of carbon atoms present in the system.

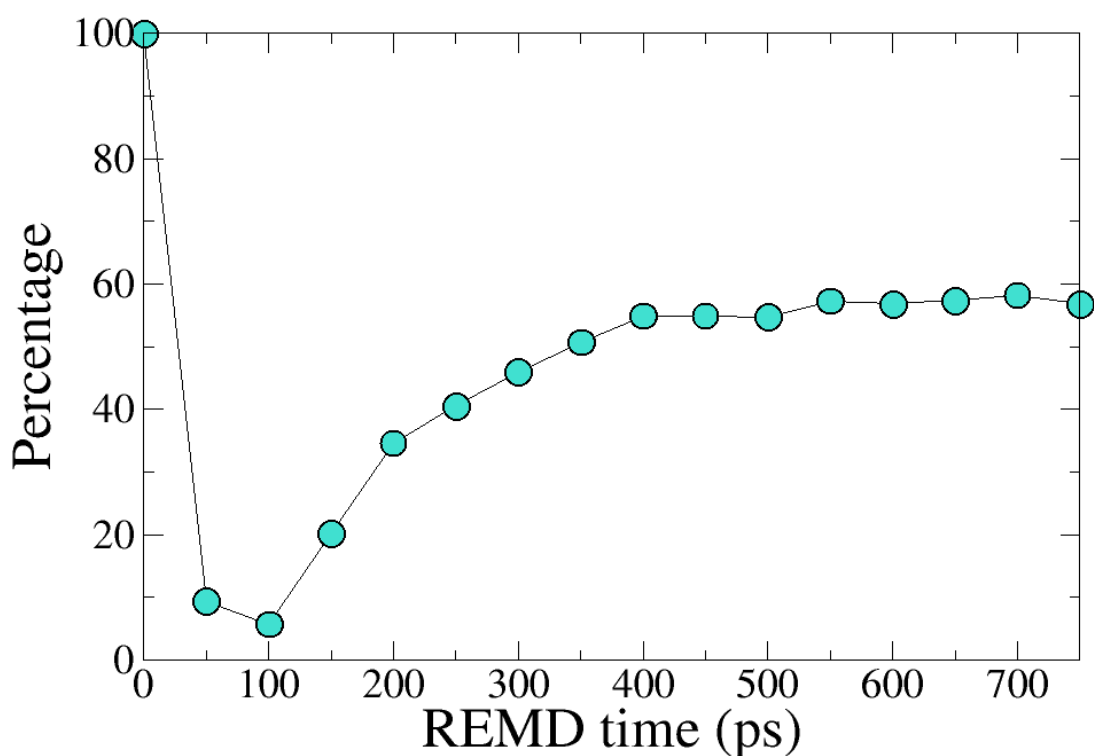


Figure 52. Evolution of the percentage of carbon atoms in the “solid” phase during REMD run of the whole cellulosic system.

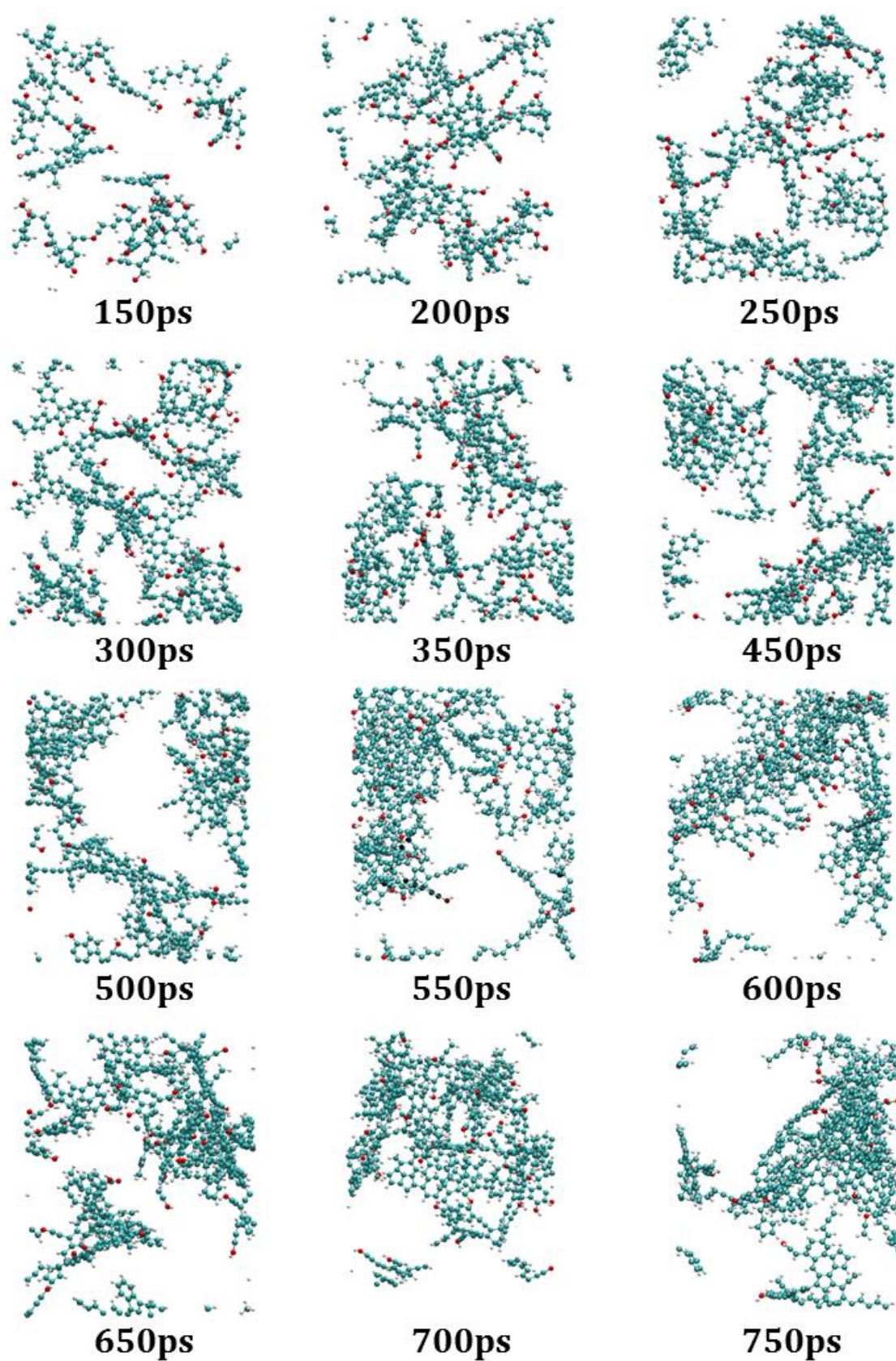


Figure 53. Evolution of the solid-like phase during the REMD run. Carbon atoms in blue, hydrogen atoms in white and oxygen atoms in red.

Snapshots of the solid-like part of the system are given in Figure 53. One clearly sees the phase growing with REMD time. The macromolecule structure also evolves, from predominant chains to ring clusters. This observation correlates well with the data presented in Figure 51. The percentage of sp hybridized carbon decreases with time while proportion of sp^2 hybridized carbon atoms increases up to 80%. The number of carbon atoms in rings and chains is given in Figure 56 a). We clearly see a predominance of chains over rings until 450ps of REMD and an inversion of this tendency beyond this time.

In Figure 54, H/C and O/C ratios are given for both fluid and solid-like phases. Separation between ratios for both phases is clear. H/C and O/C ratios decrease for the solid-like part, reaching plateaus around 0.55 and 0.045 respectively. Such a decay is observed in the formation of natural kerogens –see comparison with experiments in Figure 63. On the contrary, H/C and O/C ratios of the fluid phase increase until they reach plateau values at 3.15 and 1.90 respectively. These ratios evolutions support the composition of both phases. Indeed, high H/C and O/C atomic ratios correspond to a H and O rich fluid phase while low atomic ratios correspond to a C-rich solid phase.

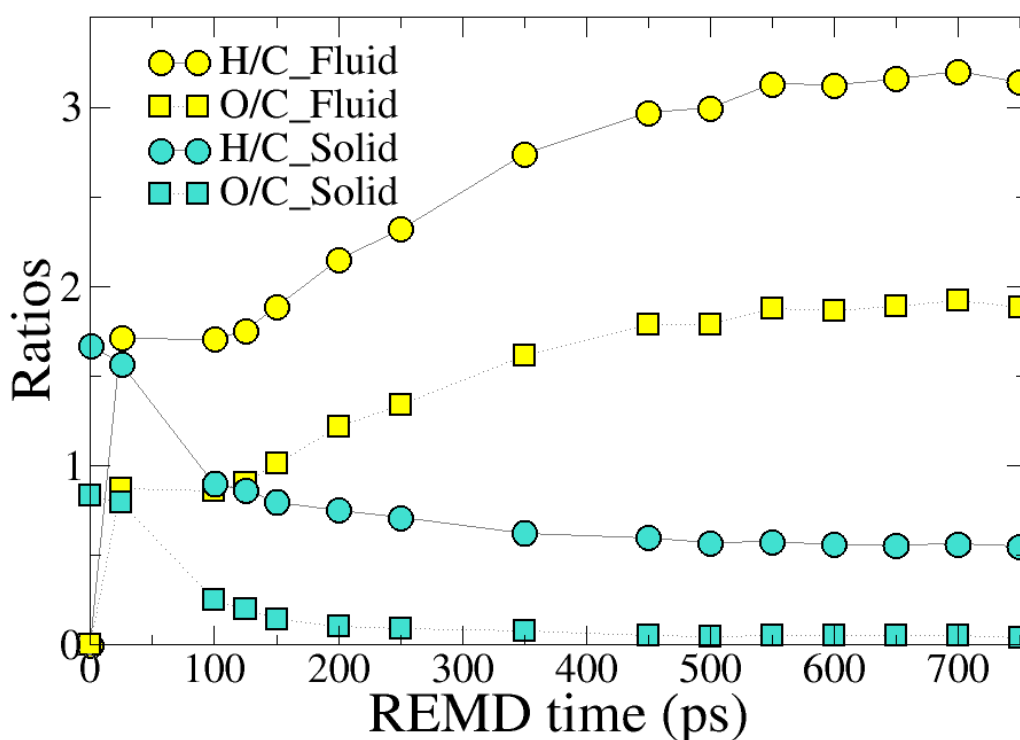


Figure 54. Evolution of H/C (circles) and O/C ratios (squares) for fluid (yellow) and solid-like (cyan) phases, during the REMD simulation.

H/C and O/C ratios for the fluid phase quickly increase at the beginning of the simulation. It corresponds to the significant formation of this phase at early stage. These H/C and O/C ratios of the solid phase show a sharp drop at the beginning, due to the

cellulose fragmentation, and then smoothly decrease. This smooth evolution comes from the fact that a number of molecules with 12 or more carbon atoms are formed from smaller molecules formed during initial fragmentation and they slowly lose hydrogen and oxygen atoms while they coalesce to form a solid macromolecule.

In Figure 55, a close-up of the last solid phase configuration at 750ps is given,. Ring clusters are colored in order to distinguish one from another. Clusters do not seem to be organized relative to each other. It seems to correspond to a mid-diagenesis stage in which aromatic structures have been formed but are not ordered yet.

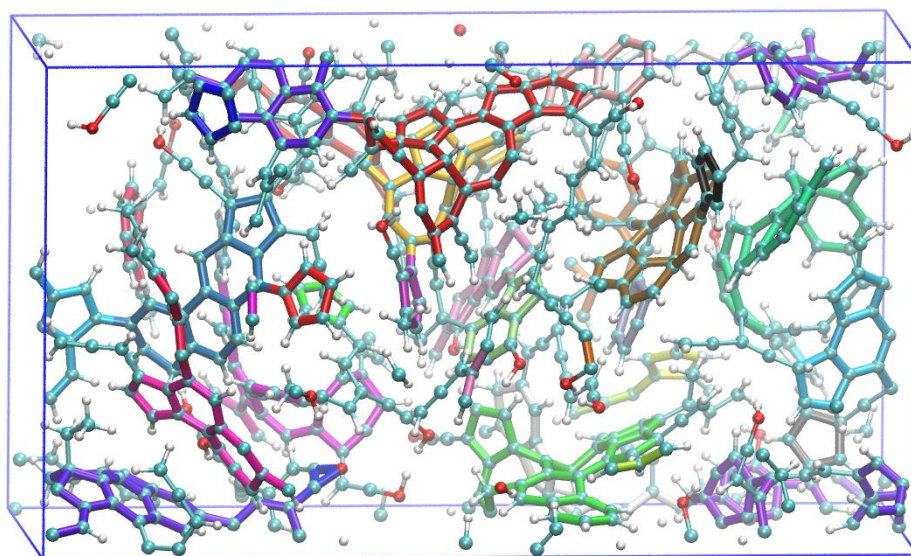


Figure 55. Snapshot of the solid-like phase after 750ps of REMD. Each ring cluster is represented with a different color.

In Figure 56, detailed topology of the solid-like phase is given. Kerogen topology is interesting since it evolves during its decomposition. It is known that a mature kerogen will be more aromatic than an immature one⁹. Thus, we can check our system evolution towards a more mature configuration. In Figure 56 a) the number of carbon atoms in rings and in chains, and the ring/chain ratio are plotted. The number of carbon atoms in chains increases for REMD time up to 450ps and then decreases. At the same time, number of carbon atoms in rings increases throughout the simulation, until it reaches a plateau around 450ps. This evolution corresponds to the previous observation of the presence of chains and sp hybridized carbons in the first half of the run, and the presence of rings and sp^2 hybridized carbons in the second half. Total number of rings and number of clusters composed of rings are represented in Figure 56 b). They both increase during almost the first half of the simulation run, before oscillating around 125 and 30 respectively.

In Figure 56 c), the numbers of rings and atoms per cluster are given. They both increase all along the simulation, as the solid-like phase grows along the REMD run. Kerogen aromaticity is known to increase once hydrocarbons start to form^{9,18}. Finally, in Figure 56 d) ring topology is given. Evolution of percentage of pentagons, hexagons, heptagons and octagons is plotted. The first two correspond to the largest amount. At the end, hexagons represent 45% of the total rings, pentagons around 40%, heptagons around 10% and octagons around 5%. All these elements support the idea of a structural organization of the solid phase all along its formation from chains containing fragments.

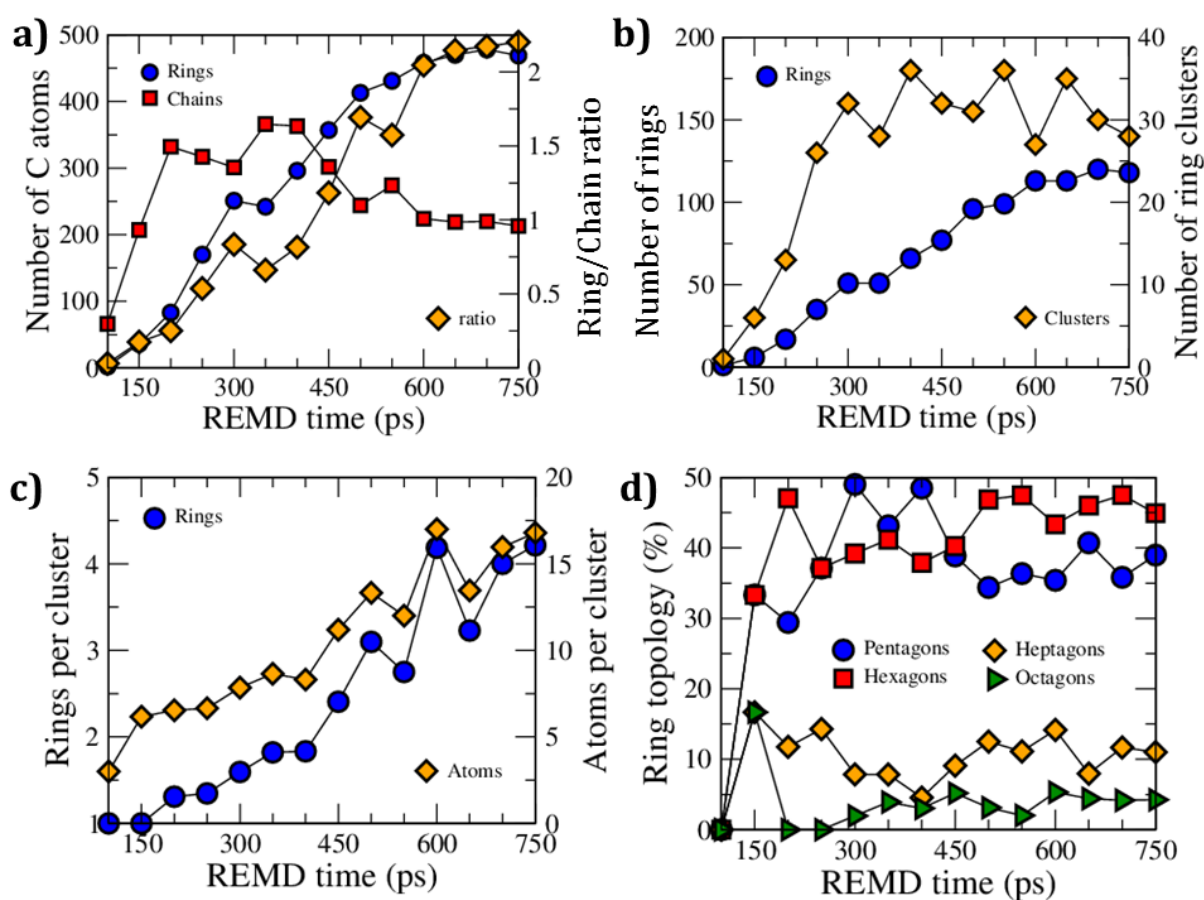


Figure 56. Solid-like phase topology. **a)** Number of carbon atoms in rings (blue circles) and in chains (red squares) and ring/chain ratio (orange diamonds). **b)** number of rings (blue circles) and number of ring clusters (orange diamonds). **c)** number of rings (blue circles) and atoms (orange diamonds) per cluster. **d)** Percentage of pentagons (blue circles), hexagons (red squares), heptagons (orange diamonds) and octagons (green triangles) in rings.

Beside topology, another interesting characteristic would be the solid-like phase density. We here define the density as the mass per unit volume, using the following equation:

$$d = \frac{m}{V}$$

Where d is the density in g.cm^{-3} , m is the mass in g, V is the volume in cm^3 . To calculate the whole system density, we divide the whole system mass per the whole box volume, while for the solid part density, we divide the solid-phase mass per the whole box volume. A specific volume of the solid part cannot be easily defined.

Kerogen density range is quite wide, depending on its type and maturation stage, and their relation remains ambiguous¹⁹. Nevertheless, some studies^{18,20} intended to measure kerogen density, in the range $[1.12\text{-}1.32 \text{ g.cm}^{-3}]$. In this first REMD run, density value for both phases is not meaningful. Although system evolves, both phases are constrained by the other. Solid-like phase density does not correspond to the one it would have if a fluid phase portion could escape, as in real conditions. Nevertheless, we plotted density evolution for whole system and solid-like part, in Figure 57, in order to follow its evolution trend. Even lower than the literature range, the whole system density is not too different, and the difference might be due to the great content of fluid that would, in part, escape during the pyrolysis process. Solid part density decreases before it increases while whole system density decreases. They both reach a plateau after 500ps of REMD, when the system converges. It corresponds to the point when the number of carbon atoms in rings clusters reaches a plateau, and rings quicker coalesce into clusters.

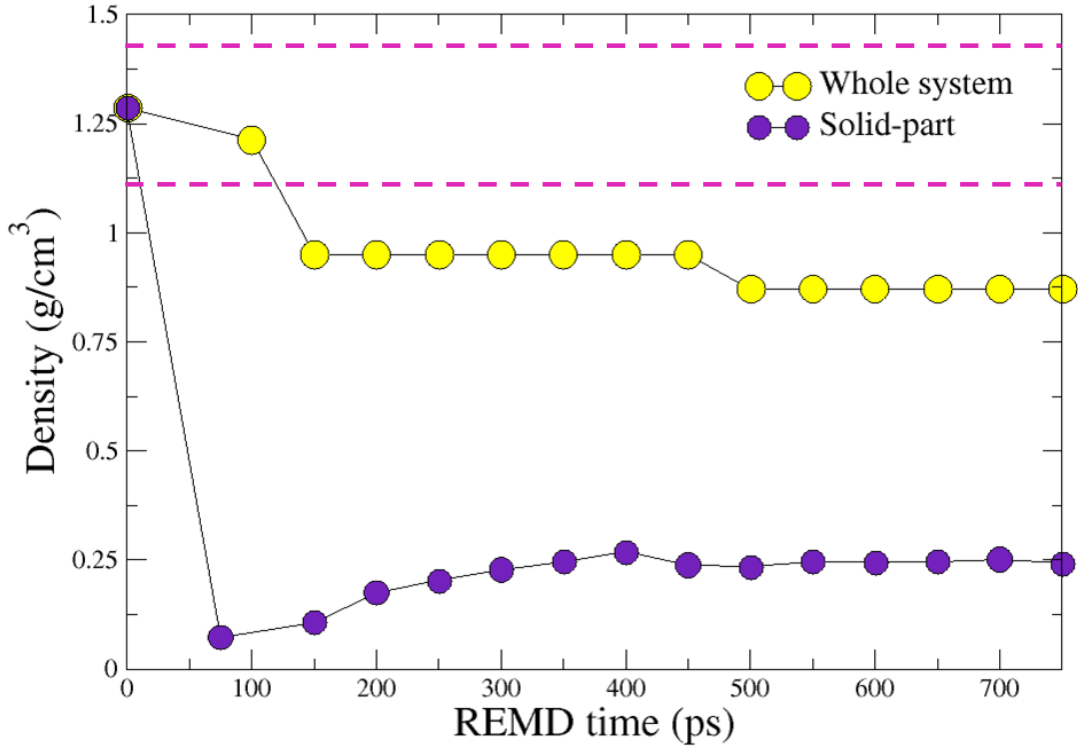


Figure 57. Density evolution of the whole system (yellow circles) and of the solid-like phase (purple circles). Violet dashed lines represent literature range^{18,20}. The essential data here is the whole system density and the solid part density is given for information purposes only.

Density evolution only occurs during (NPT) pressure relaxations. Even if the solid part density value does not have real meaning, we can note that it evolves with the carbon atoms percentage in the solid phase. The whole system density decreases with the pressure relaxation.

After 750ps of REMD, there is a clear distinction between solid and fluid phases. We computed isothermal compressibility on the solid-like phase. Isothermal compressibility measures the relative volume change of a matter under a given pressure, and at a constant temperature. It is defined as following:

$$\chi_T = -\frac{1}{V} \cdot \left(\frac{\partial V}{\partial P} \right)_T$$

Isothermal compressibility χ_T can be computed, using volume fluctuations in the NPT ensemble, as following^{21,22}:

$$\chi_T = \frac{\langle V^2 \rangle - \langle V \rangle^2}{k_B T \langle V \rangle}$$

Where $\langle V \rangle$ is the average system volume, k_B the Boltzmann constant and T the temperature.

We found an isothermal compressibility of $1.64 \cdot 10^{-10} \text{ Pa}^{-1}$. Collell *et al*²². found values of $2.7 \cdot 10^{-10} \text{ Pa}^{-1}$ for a type II kerogen –defined as “*in the middle of oil generation window*”- and $2.9 \cdot 10^{-8} \text{ Pa}^{-1}$ for a bulk fluid. A very mature kerogen has a Young modulus around 14GPa¹⁹, thus^{xviii} a bulk modulus around 9GPa. An isothermal compressibility of $1.64 \cdot 10^{-10} \text{ Pa}^{-1}$ implies a bulk modulus^{xix} of 6.1GPa. Our solid phase is thus really close in its compressibility to the one of Collell *et al*. We decided to start from this solid-like system to run the second step of our simulation: an REMD run on the solid-like part only.

Second step: Solid-part relaxation by REMD

During the first step of 750ps of REMD on the whole cellulosic system, two distinct phases appeared: a fluid one, composed by small molecules such as water and carbon monoxide, rich in hydrogen and oxygen, and a solid-like one, rich in carbon atoms. We have seen in previous section that the latter can be considered as a kerogen in formation.

NPT relaxation

We first relaxed the final configuration of the solid-like phase –after 750ps of REMD on the whole cellulosic system. For that, we run a constant pressure and temperature simulation during 1050ps –at 423K and 50MPa. In Figure 58 a) and b) are given snapshots of the structure before and after relaxation. As all the fluid phase has been removed, the solid part is not constrained anymore by it, and therefore, condenses under the pressure applied by the NPT simulation. Figure 58 a) corresponds to the solid part as it was at the end of the first REMD stage, fluid removed, and Figure 58 b) corresponds to the solid part which has condensed under 50MPa.

Density evolution of the system is given in Figure 58 c). It increases, while the system condenses, until it reaches the density of 1.31 g.cm^{-3} , which is representative of that of a kerogen phase.

^{xviii} Young modulus E and bulk modulus K are linked by the following equation : $K = \frac{E}{3(1-2\nu)}$ where ν is the Poisson ratio, with $\nu=0.25$.

^{xix} Bulk modulus is the inverse of isothermal compressibility. It measures how compressible the system is, how a material pressure increases as a result of an isostatic decrease of volume. For example, bulk modulus of glass, steel, and diamond are around 45, 160 and 440 GPa respectively.

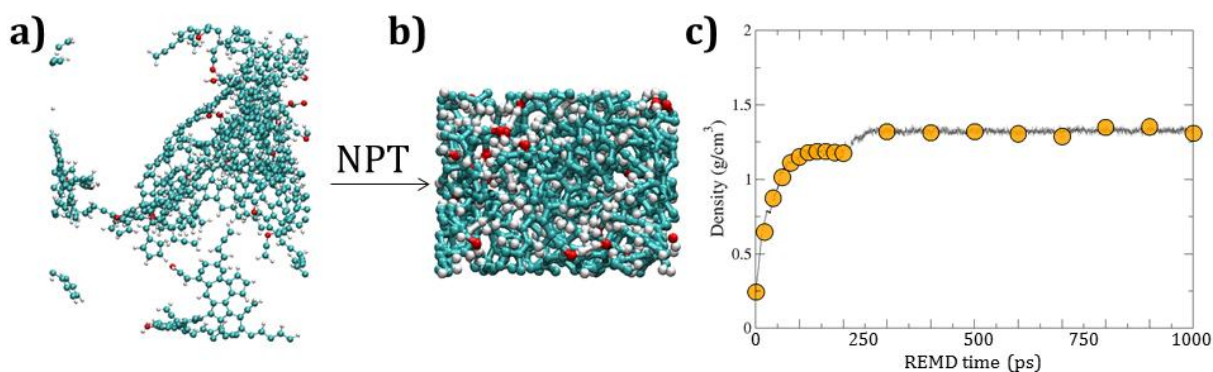


Figure 58. Solid-like phase relaxation, after 750ps of REMD. Relaxation at 423K and 50MPa. Snapshots of the solid phase before **(a)** and after **(b)** relaxation. **(c)** evolution of density.

REMD

Then, we run 1800ps of REMD, alternating 200ps of REMD and 100ps of NPT –at 423K and 50 MPa. Therefore, we made 9 NPT relaxations, one every 200ps of REMD. In Figure 59 a diagram depicting numerical simulations applied to the solid part obtained from the first stage of cellulose pyrolysis is given.

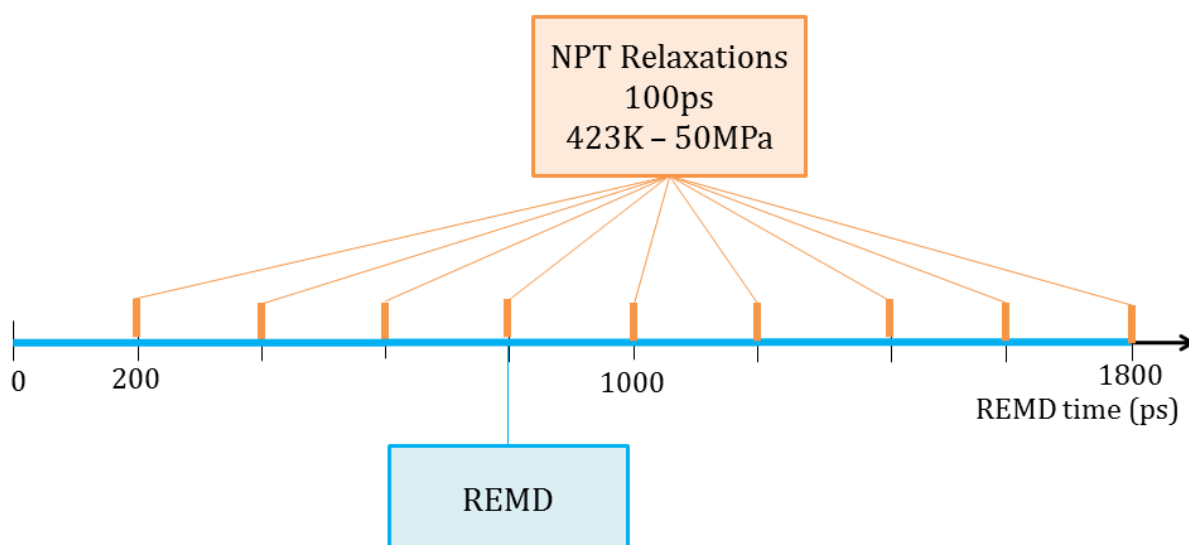


Figure 59. Diagram of the numerical simulations applied to the solid part obtained from the first stage of REMD applied to a crystal of cellulose. This second stage of REMD is 1800ps long and NPT relaxations are performed every 200ps of REMD.

In Figure 60 is given the system energy evolution. Energy decreases all along the simulation, up to converge. This convergence indicates an equilibrium state. After 1500ps, system has reached its equilibrium and does not evolve anymore, even during NPT relaxation –at 1600ps.

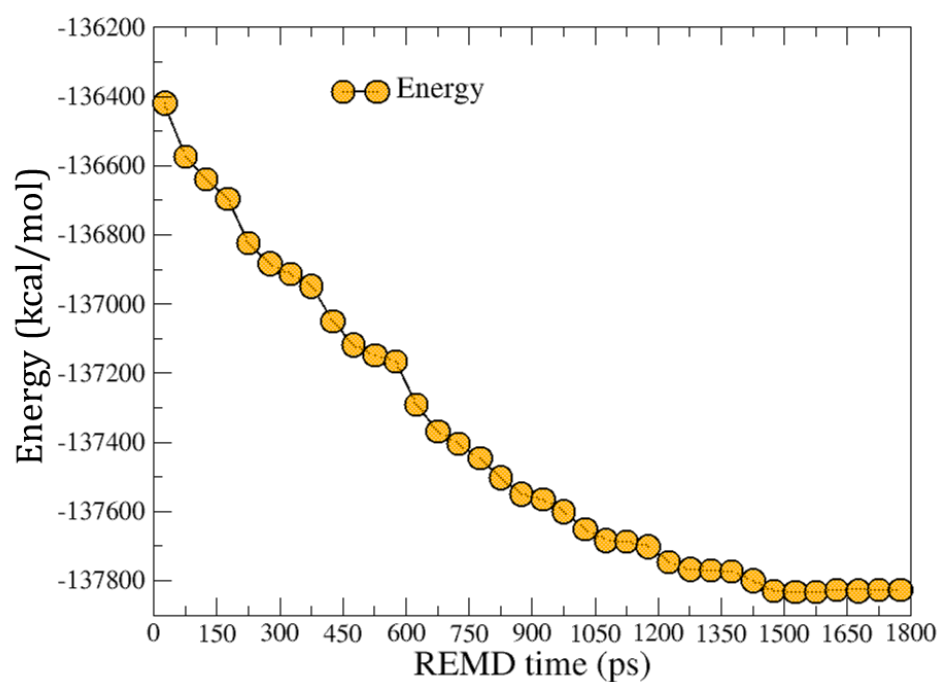


Figure 60. System energy evolution at 423K during REMD simulation.

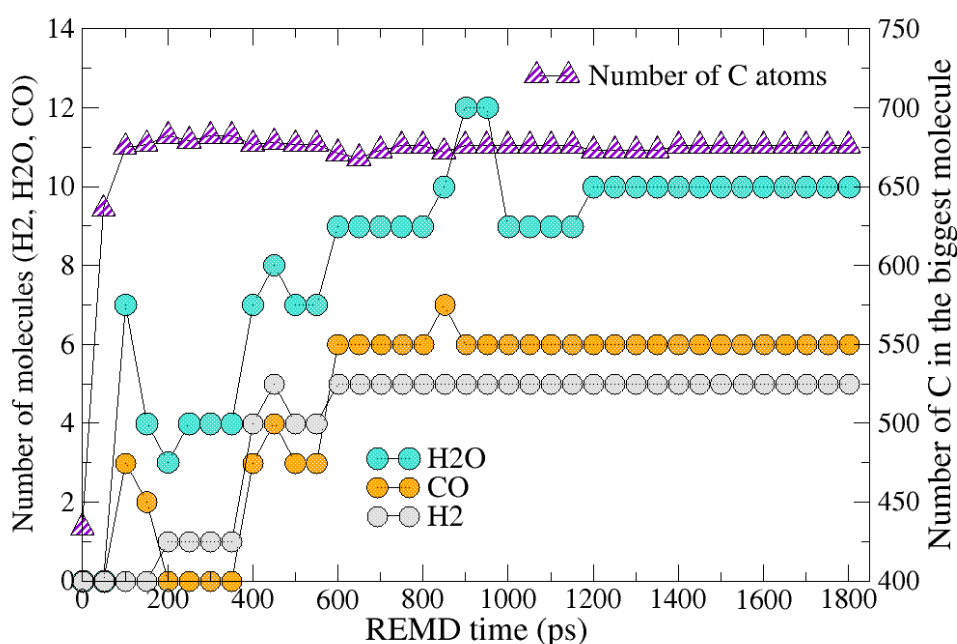


Figure 61. Evolution of the number of H₂ (grey circles), H₂O (cyan circles) and CO (orange circles) molecules, and of the number of carbon atoms in the biggest molecule of the system (violet triangles), during the REMD run.

Some small molecules –as H₂, H₂O and CO– are still formed during the first 600ps of REMD on the solid-like phase. This formation is however limited since, as one can see in

Figure 61, the maximum amount of water is 12 molecules, 7 for CO and 5 for H₂. The most formed molecule is still H₂O, followed by CO and H₂. No CO₂ is formed during this run. After few picoseconds, the only components of the system are these molecules and the solid-like part. Number of carbon atoms in the solid –there is a single large carbon cluster composing the solid phase after 50ps of REMD- does not evolve once the two initial larger clusters have coalesced.

After 1200ps of REMD, numbers of H₂, H₂O and CO do not evolve anymore. Neither the number of carbon atoms in the solid part does. The solid part is then composed of 676 carbon atoms, 345 hydrogens, and 12 oxygens. H/C and O/C atomic ratios are then of 0.51 and 0.018 respectively.

Kerogen density which can be found from different studies¹⁸⁻²⁰ is in the range [1.12-1.32 g.cm⁻³]. At the end of the first REMD stage, we let the solid-like phase relax until it reaches its equilibrium density, which was 1.31g.cm⁻³. In Figure 62, we plotted the evolution of density during the second REMD run.

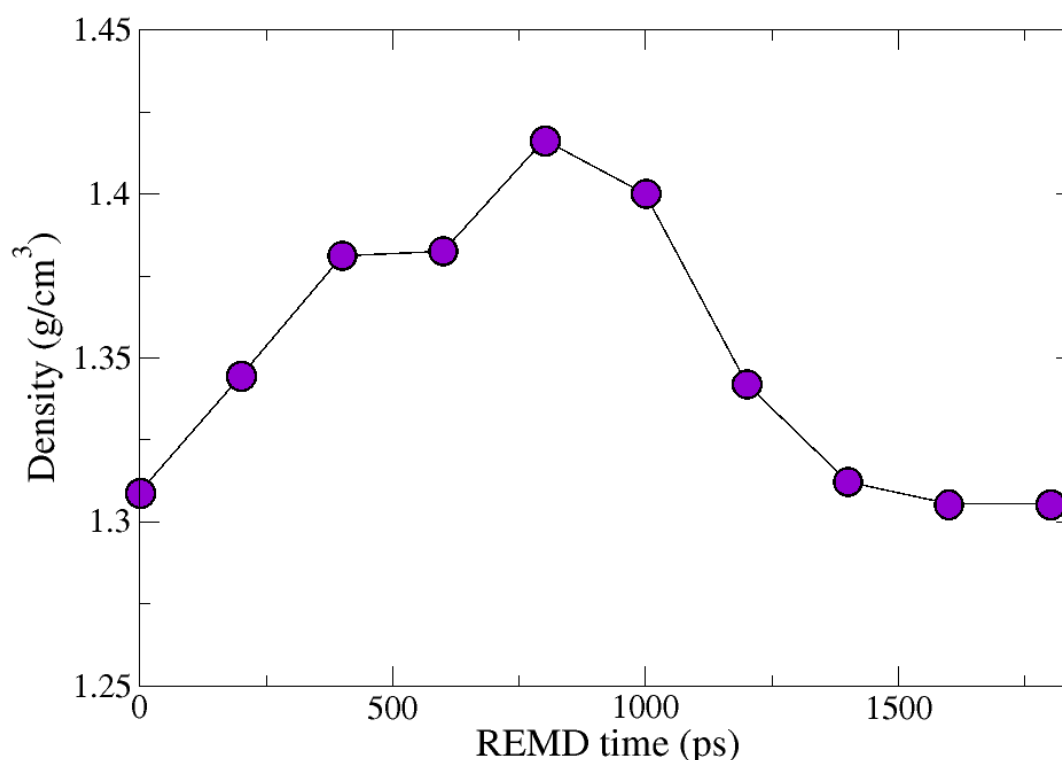


Figure 62. System -solid part- density during the -second- REMD simulation.

Density increases to 1.41 g.cm⁻³ and then decreases until to stabilize around 1.31 g.cm⁻³. In the case of type II kerogen, density is known¹⁸ to be almost constant for immature samples –mostly comprised between 1.20 and 1.25 g.cm⁻³- and then to increase up to 1.35 g.cm⁻³ during petroleum generation and can reach 1.40 g.cm⁻³ at late

oil generation. Density of our solid part is comprised in the experimental range of type II kerogen. Evolution of a type III kerogen density is unknown, and it might be not comparable to that of a type II. Evolution of our solid part density during this second REMD stage does not seem to evolve following the formation of gaseous molecules –see in Figure 61- but in both cases there is a peak around 900ps of REMD. The increase might be due to the increase of the structure aromaticity, and we assume that the decrease could be due to a creation of porosity. At the end of this second simulation, isothermal compressibility of the solid part is $1.40 \cdot 10^{-10} \text{Pa}^{-1}$, which corresponds to a bulk modulus of 7.14GPa. It did not change much during this simulation.

This second REMD run, applied to the solid-like phase, seems to correspond to the end of the second stage of kerogen evolution: **end of catagenesis**. Indeed, as one can notice in Figure 60, system energy decreases until to reach a plateau, around 1500ps, while chemical composition of the structure does not significantly change. As one can see in Figure 61, at the end, when system has reached its thermodynamic equilibrium, only 10 water molecules, 6 carbon monoxide molecules and 5 dihydrogen molecules have been formed. It only fluctuates around its equilibrium value of 676 atoms. Nonetheless, energy decreases which means that the system is evolving towards a more stable configuration: the structure rearranges with an increase of the cluster size. This increase of the cluster size is not significant enough to correspond to a metagenesis stage. Indeed, number of rings per cluster does not exceed 6 and it is known that during catagenesis, aromatic domains do not exceed 10 fused rings⁹. We therefore think that this evolution corresponds to the end of kerogen catagenesis only.

Comparison of our kerogen evolution on a Van Krevelen diagram

Experiments on organic matter pyrolysis, and in particular on cellulose, have been conducted for a long time. In 1918, Pictet et al.²³ reported the transformation of cellulose during pyrolysis into a molecule called levoglucosan. Depolymerization of cellulose to form this component has been found in many other studies, both experimental^{24–26} and numerical^{27,28} and cellulose pyrolysis is still an intense field of research. Applications are numerous, ranging from flame resistant fabric²⁹ to energy³⁰ production.

All previous results presented in this chapter, help to build a view of kerogen formation and agree with experimentations. Cellulose pyrolysis at low temperature indeed leads to formation of a two-phase system. A fluid one mainly composed of water and carbon monoxide, and a solid-like one -carbon rich- which, once relaxed and submitted to a new pyrolysis runs, corresponds to a kerogen phase.

We explained earlier in this manuscript that the main tool to visualize and follow kerogen evolution is the Van Krevelen diagram. In Figure 63 is given a Van Krevelen

diagram on which we plotted our simulation data along with experimental values for a type III kerogen confined pyrolysis³¹ and for a cellulose open pyrolysis³².

Our simulated decomposition pathway is included in the type III kerogen cloud of experimental points although a bit lower in term of H/C. Until mid-catagenesis –around O/C equal to 0.08- our system follows exactly the same path as the type III kerogen submitted to a confined pyrolysis of Monthioux *et al.*³¹. Cellulose open pyrolysis of Tang and Bacon³² follows a same trend but at a lower H/C atomic ratio –see also in Figure 64.

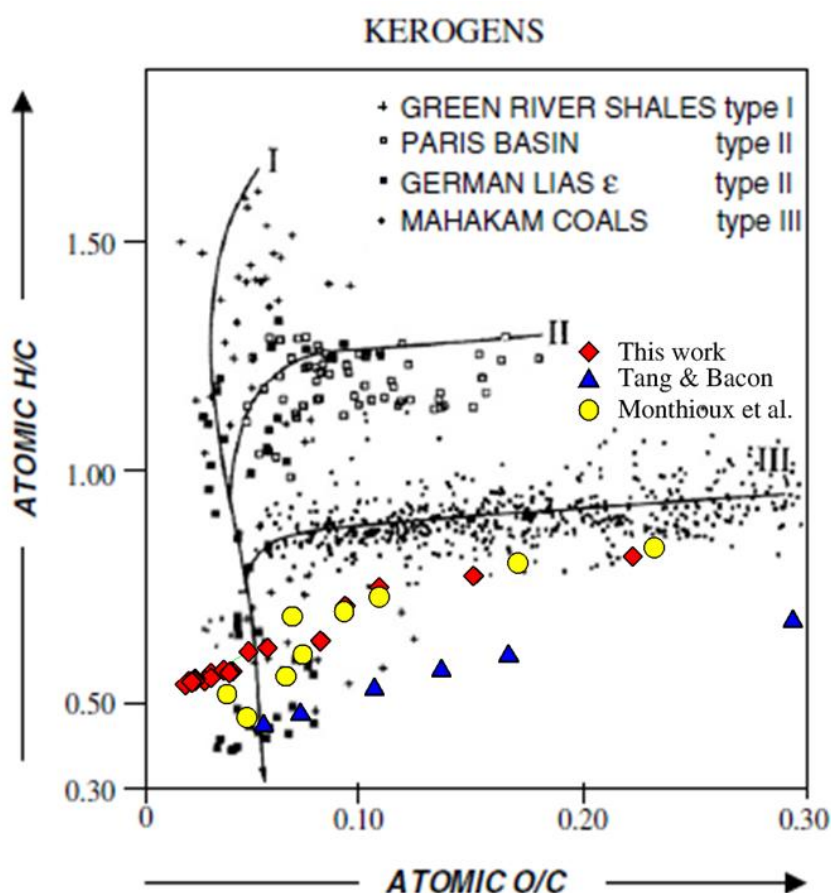


Figure 63. Van Krevelen diagram from Vandenbroucke and Largeau⁹ with values for our work (red diamonds), values of cellulose open pyrolysis from Tang and Bacon³² (blue triangles), and values from type III kerogen confined pyrolysis from Monthioux *et al.*³¹ (yellow circles).

In 1985 Monthioux *et al.*³¹ compared artificial and natural type III organic matter evolution. For this purpose, they pyrolysed concentrated organic matter in sealed gold tube –confined pyrolysis- for 24h at temperatures between 250 and 550°C -523 and 823K. They performed those experiments at different pressures –between 0.5 and 4kB but did not find any change in this pressure range. They compared their results with natural type III kerogen and with other experiments performed in open and closed

environments and found their method to be more accurate. In Figure 63, we see that Monthioux *et al.* results have a better correspondence with type III kerogens envelope than Tang and Bacon's. Tang and Bacon pyrolysis is indeed performed in open medium. The fact that our results are close to Monthioux *et al.*'s supports our numerical approach.

One can note that both results, ours and Monthioux *et al.*'s, are closer to natural evolution than the ones of Tang and Bacon. Fluid formed during kerogen thermal maturation is supposed⁹ to migrate and this is the case during the open pyrolysis of Tang and Bacon. A question thus arises: why Monthioux *et al.*'s experiment and our numerical simulation do give closer results to natural decomposition of organic matter, than the ones of Tang and Bacon? This observation supports that reaching a certain level of partial pressures by confinement –pressure of effluents- leads to a satisfactory simulation of the natural evolution, as experimentally proven by Monthioux *et al.* in 1985³¹. It also raises the idea that the fluid confinement, present in both works, might help to cross kinetic barriers which are time-overcome in Nature by a poromechanics effect.

In 1964 Tang and Bacon³² studied the pyrolysis of cellulose up to 500°C -773K- and found four stages in the process of decomposition:

- desorption of physically adsorbed water (25-150°C);
- dehydration from the cellulose (150-240°C);
- chain scissions, or depolymerization, accompanied by evolution of more water, CO, and CO₂, (240-400°C), and
- aromatization, or formation of graphite-like layers.

Evolution of their cellulosic system is given in Figure 64. We also plotted evolution of our simulated system. Decomposition pathways of both systems display some similarities. The first stage described by Tang and Bacon –desorption of physically adsorbed water- does not appear in our system since we did not have any adsorbed molecules in our initial molecular model. However, cellulose dehydration is highly visible, even if less marked due to the great formation of CO in particular –at 280°C Tang and Bacon report a value of 1.8% of CO for pyrolyzed cotton cellulose when in our system after 750ps of REMD there are more than 25% of CO. Depolymerization and chain scission stage appears in our system with cellulose fragmentation, during its dehydration. We also observe the formation of aromatic clusters as in Tang and Bacon last stage.

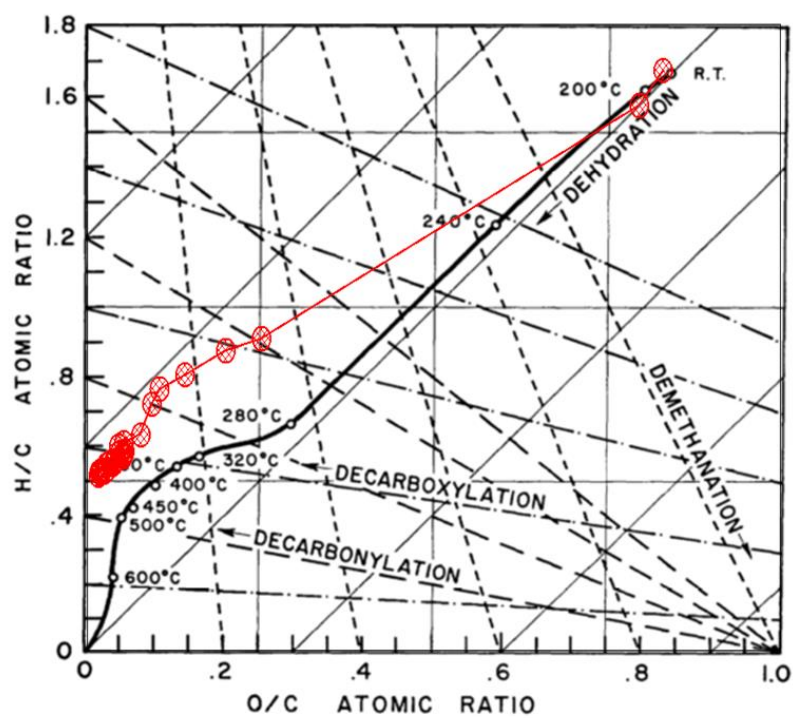


Figure 64. Van Krevelen diagram from Tang and Bacon³² (cellulose open pyrolysis). Our system evolution is in red.

Type III Kerogen precursor: Lignin

As a second precursor of type III kerogen, we now focus on lignin decomposition under geological conditions, using the REMD simulation technique.

Choice of lignin as a type III kerogen precursor

Lignin, as cellulose, is a type III kerogen precursor since it is a complement of cellulose in higher plants structure. Its role is mainly to ensure the rigidity of the biological tissues in wood. The interlinking of cellulose, lignin, and hemicellulose^{xx} is shown in Figure 65.

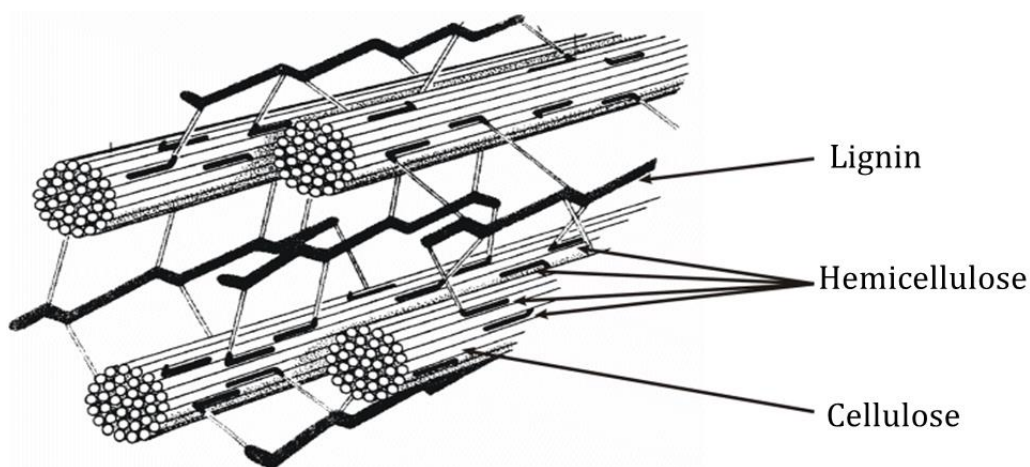


Figure 65. Lignocellulosic biomass, from Almardottir and Gunnarsson (2009)³³

Term lignin is used for a large family of aromatic polymers³⁴. It is found into a number of plant cells and ensure both rigidity and impermeability –because of its hydrophobicity. By contrast to cellulose, there is no unique well defined structure for lignin^{35,36}. A large variety of lignin structures exist but they are all based on three main blocks –called monolignols. These blocks are called coumaryl alcohol, coniferyl alcohol and sinapyl alcohol, and are respectively designated as H, G and S. They are represented in Figure 66. As one can see, these blocks are based on an aromatic ring, feature that will induce a high char formation. Proportion of these components differs depending on the bio-source. For instance, conferyl alcohol will be present with a great fraction in evergreens –up to 80%- while sinapyl will represent half of the lignin composition in flowering plants³⁷.

^{xx} Hemicellulose is one of the three major wood components with cellulose and lignin. It acts as bridges between cellulose fibers. Hemicellulose is a polymer composed of different sugars such as glucose and xylose –not only anhydrous glucose as it is the case for cellulose. For this reason we decided to study lignin instead of hemicellulose.

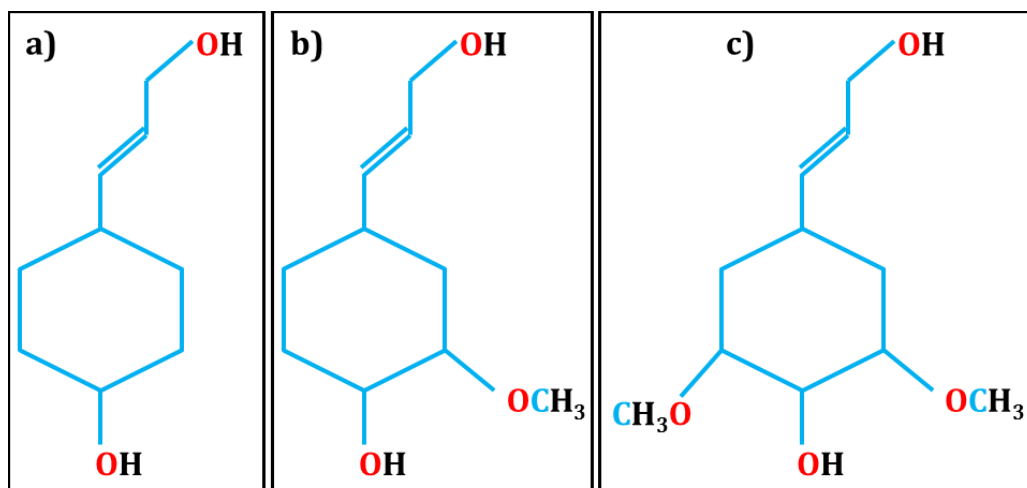


Figure 66. Lignin blocks –monolignols. a) coumaryl alcohol, b) coniferyl alcohol, and c) sinapyl alcohol.

Over the last 50 years, several lignin structures have been proposed³⁸, such as the widely known ones of Freudenberg³⁹ in 1968 and of Adler in 1977⁴⁰. They are shown in Figure 67.

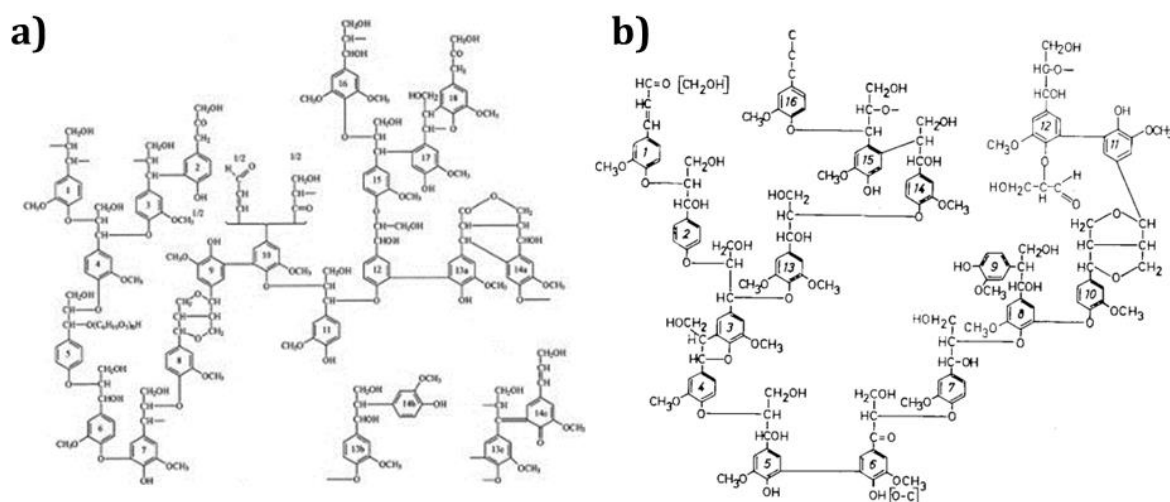


Figure 67. Lignin structures proposed by Freudenberg³⁹ (a) and Adler⁴⁰ (b).

In 2003, Banoub and Delmas⁴¹ proposed a more linear structure. They explained that, lignin structures used until then were incorrect for the reason that they result from wrong analysis and/or from structures which have been modified during extraction. We adopted their view and looked for a more linear, and easier to build, structure of lignin.

Building a lignin model

Among all lignin structures published, we decided to use the one of Crestini⁴² and co-workers, showed in Figure 68, which is both linear and relatively easy to construct.

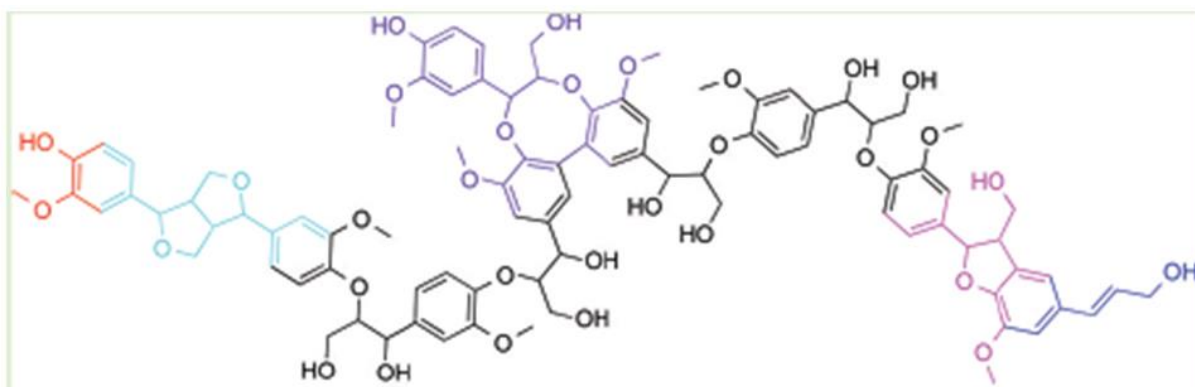


Figure 68. Lignin structure model from Crestini *et al.*⁴²

We first reproduced the structure given in Figure 68 in two dimensions, that we relaxed to let it arrange and fold as it needed to. We replicated it in order to get a box simulation of 3978 atoms, since we wanted a lignin structure with a number of atoms comparable with the one of our cellulose system. This lignin formula is $(C_{90}H_{100}O_{31})_{18}$, and thus, is composed of 1620 carbon atoms, 1800 hydrogen atoms, and 558 oxygen atoms. Initial H/C and O/C atomic ratios are 1.11 and 0.34 respectively.

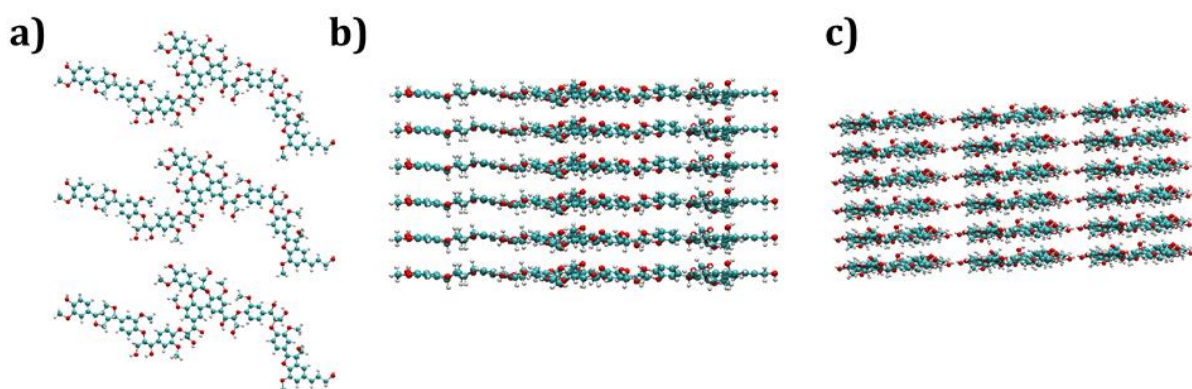


Figure 69. Initial lignin structure, built from Crestini *et al.*⁴² (a), (b), and (c) represent the structure from different views.

We relaxed the whole structure at 423K and 50MPa, using the ReaxFF_{C-2013}. We then used the obtained structure to run our REMD simulation, and simulate lignin pyrolysis under geological conditions.

REMD Confined pyrolysis of lignin

Like for the first stage of cellulose pyrolysis study, lignin system was submitted to a REMD run. We run a total of 1300ps of REMD, with pressure relaxations every 200ps. These NPT relaxations –at 423K and 50MPa- are used to keep the system in a geological pressure range of [25-125MPa].

We performed REMD simulations using the same parameters as the ones used for cellulose system: 96 temperatures distributed following a geometric series, starting at 423K and with a common ratio of 1.0225. Temperatures span the range [423K:3500K]. Temperature swaps between neighbor replicas are attempted every 10fs of REMD, a frequency we showed to be reasonable, in the cellulose section.

In Figure 70 whole system energy and pressure evolutions are plotted. Energy decreases all along the simulation meaning that the system is evolving towards a more stable configuration via an exothermic process. It converges around 1100ps of REMD, that is a little bit longer than for cellulose. This difference in time convergence might be due to the high stability of aromatic structures. Lignin carbon has been showed to be resistant to thermal degradation^{43–45}, because of its high aromaticity. At this point, system presents a disordered structure. Its structure has significantly evolved compared to the initial one, as shown in Figure 71.

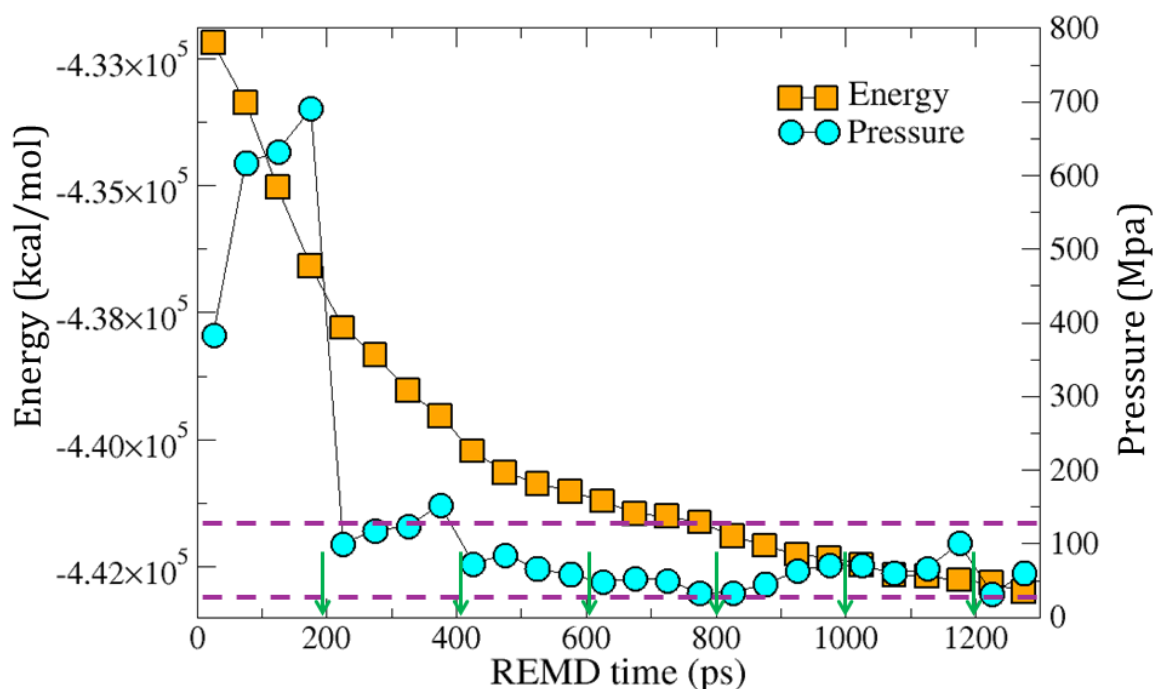


Figure 70. Energy (orange squares) and pressure (blue circles) evolution during REMD simulation of lignin. Violet dashed lines represent limits of geological pressure between which we try to keep our system. Green arrows represent the times at which NPT relaxations are performed.

Concerning the pressure, we can note a trend similar to that of whole cellulose system one during the first REMD run. Indeed, we see a significant increase of the pressure at the beginning –until the NPT relaxation after 200ps of REMD- corresponding to the large formation yield of water. Relaxation at 200ps of REMD allows the system to release pressure due to all the accumulated water, and pressure is brought into the geological range.

During this REMD simulation, as for cellulose, initial system is decomposed into two phases: a fluid one, mainly composed by water and carbon monoxide, and a solid-like one, a carbon-rich macromolecule. Such a separation can be seen in Figure 71.

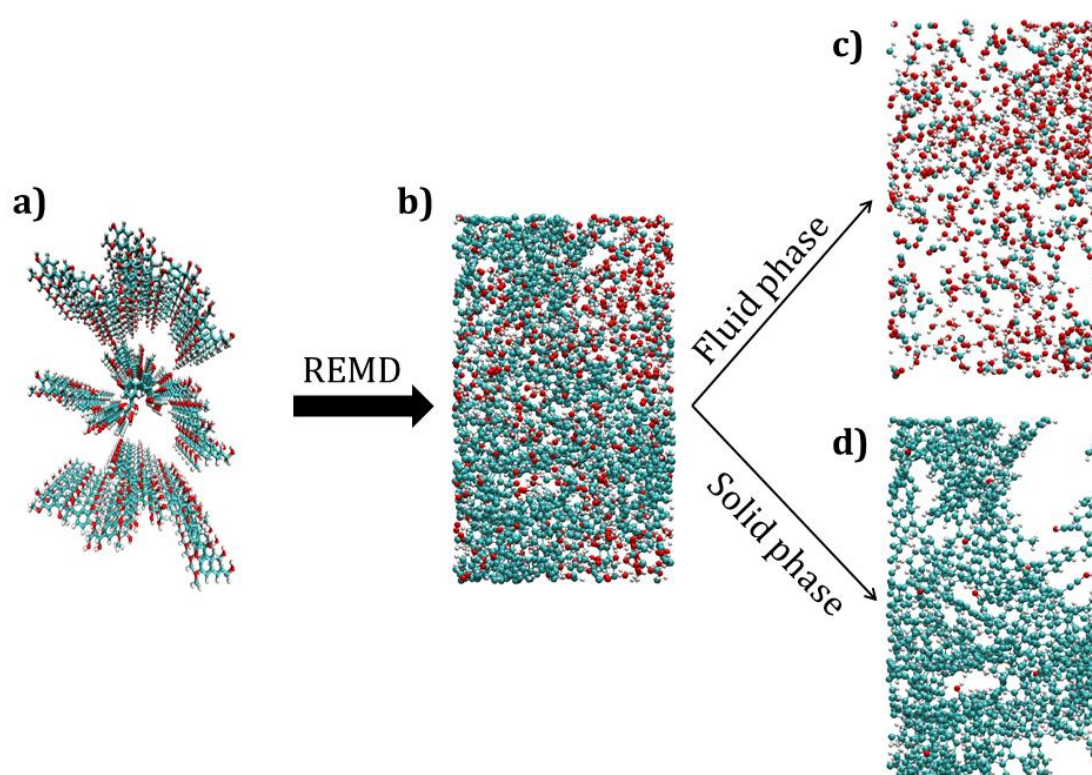


Figure 71. Lignin system at the initial step **(a)** and after 1300ps of REMD time **(b)**. Two-phase system after 1300ps of REMD: **(c)** snapshot of the fluid phase, and **(d)** snapshot of the solid phase. Carbon atoms are blue, hydrogen atoms white and oxygen atoms red.

Lignin follows a similar decomposition pathway under pyrolysis to the one observed for the cellulosic system. It indeed leads to an equilibrium between a hydrogen and oxygen rich fluid phase, and a carbon rich solid-like phase. Looking at the evolution of H/C and O/C ratios for both phases –in Figure 72- gives an interesting overview of this equilibrium formation. Both H/C and O/C atomic ratios for the fluid increase all along the simulation, while those of the solid part decrease. They all saturate once the system

converges, after 1100 ps of REMD. Fluid gets richer and richer in hydrogen and oxygen, at the same time solid gets poorer in those components. This evolution is well known, with the ultimate solid stage being char, composed only by carbon. It validates our results for both cellulose and lignin.

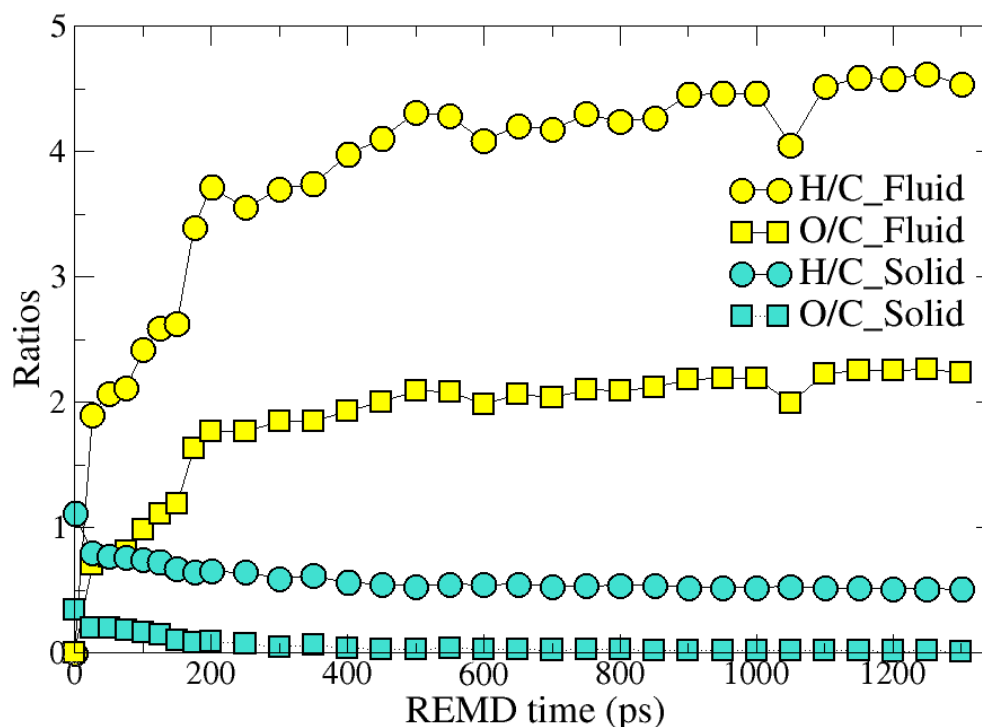


Figure 72. Evolution of H/C (circles patterns) and O/C ratios (squares patterns) for fluid (in yellow) and solid-like (in cyan) phases, during the REMD simulation of lignin decomposition.

Although decomposition pathways of lignin and cellulose are similar, fluid composition and solid-like part formation are quite different. We first describe the fluid phase formation before focusing on the solid one.

Fluid phase formation

The evolution of common molecules during REMD simulation is given in Figure 73. As for cellulose, the main component of the fluid phase is water. Dehydration of lignin is the first process to take place. The number of water molecules quickly jumps from 0 to almost 200 in 200 ps of REMD. At this point, water molecules represent 46% of the total number of molecules in fluid, and carbon monoxide 13%. After 1200 ps of REMD, water molecules represent 54%, and carbon monoxide 20%, thus two thirds of the fluid is composed by these two components.

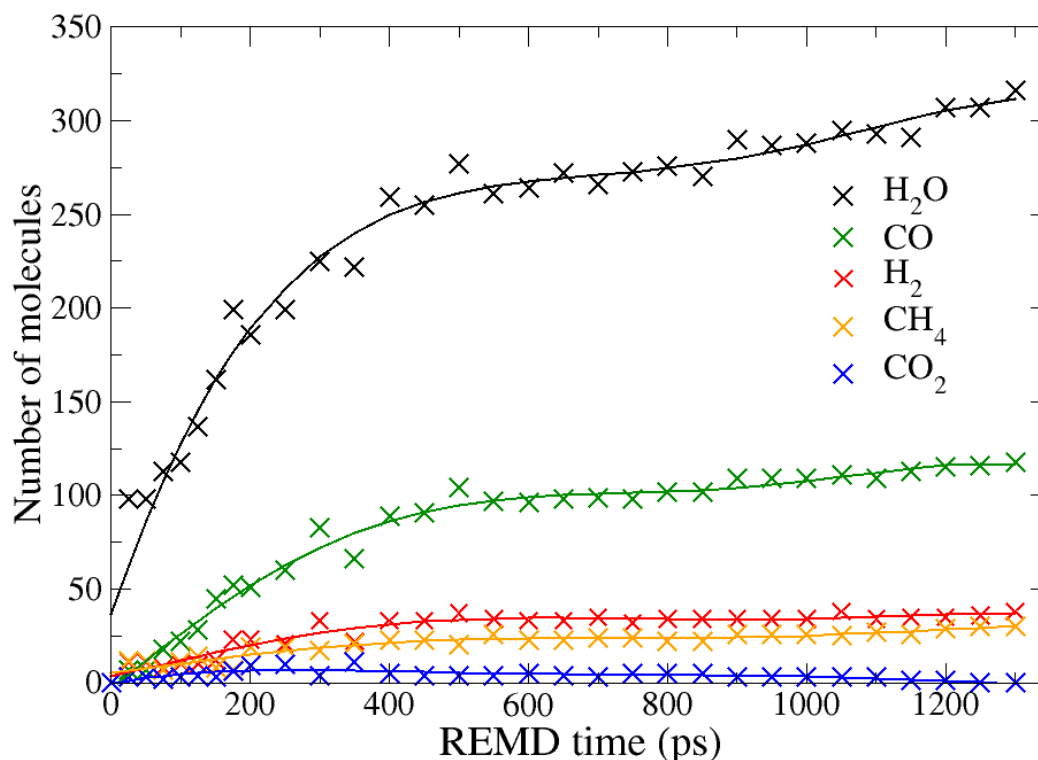


Figure 73. Evolution of small common molecules during cellulose decomposition REMD simulation. H₂O in black, CO in green, H₂ in red, CH₄ in orange, and CO₂ in blue.

Besides common molecules, the evolution of which is given in Figure 73, other stable elements are present in fluid phase. We identify them in

Table 5. We only specify those that present five molecules or more at 423K. CO₂ does not appear in this table since, after 1300ps of REMD, no CO₂ molecule remains in the system. There is only one molecule at 604 and 1407K. CO₂ production is assumed^{46–48} to increase with temperature but this trend has not been found in our simulations. Even if we look at the highest temperature -3502K- we do not find any CO₂ molecule. CO₂ is known⁴⁹ to be part of volatile elements formed during lignin pyrolysis, and this disappearance can be due to the potential used in this study –see discussion in the cellulose section- or to the lignin model, we run our simulation from. Indeed, as we mentioned before, lignin represents an ensemble of different polymers. Studies^{50,51} showed that gas components and their proportions obtained from lignin pyrolysis are strongly dependent on temperature, heating rates, and on the type of lignin. Ferdous *et al.*⁵⁰ also suggested that kinetic parameters, and therefore formed products, were dependent on the type of equipment used for pyrolysis reactions. For those reasons, we limit our comparisons with experimental literature to trends.

Table 5. Fluid phase composition at different temperatures, after 1300ps of REMD

Formula	Usual name	423K	604K	901K	1407K
H ₂ O	Water	316	301	313	302
CO	Carbon monoxide	118	115	117	115
H ₂	Dihydrogen	38	35	35	35
CH ₄	Methane	30	29	31	29
CH ₄ O	Methanol	21	21	27	28
CH ₂ O ₂	Formic acid	8	5	6	6
H ₃ O	Hydronium	7	5	9	9
CH ₅ O	Methyloxonium	6	12	6	4
CH ₄ O ₂	Methanediol	5	17	6	12

Although proportion of CO₂ can be questioned, the presence of other molecules in the fluid phase validates our pyrolysis simulation. The six major components of our fluid phase at 423K –water, carbon monoxide, dihydrogen, methane, methanol, and formic acid- are known^{49,51} to be formed during lignin pyrolysis. No special trend on fluid composition evolution with temperature can be found, except maybe an increase of methanol formation. At the end, only 15% of carbon atoms compose the fluid phase, which is a rather low value. The rest composes the solid-like phase.

Solid phase formation

At the same time as the fluid is formed, a solid-like part –defined here as molecules with twelve carbon atoms or more- develops. Percentage of carbon atoms in the solid phase is given in Figure 74. The drop at the beginning of simulation corresponds to lignin fragmentation along with the start of water formation. After this drop to 72%, percentage of carbon atoms in the solid phase increases until to stabilize around 85%.

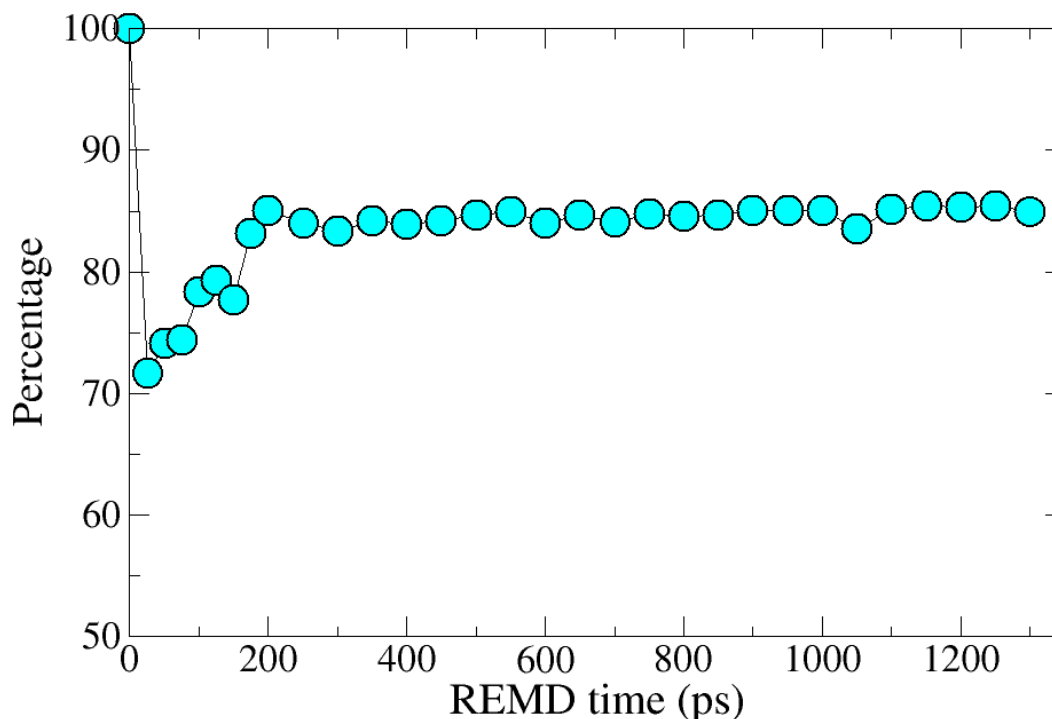


Figure 74. Evolution of the percentage of carbon atoms in the “solid” phase during REMD run of the lignin system.

We can see in Figure 75 how the percentage of sp , sp^2 , and sp^3 hybridized carbon in solid phase evolves. Proportions seem to converge after only 100 ps of REMD. A slight evolution seems to occur concerning sp and sp^2 percentage. The first slowly decreases, after a peak around 200ps of REMD, while the second smoothly increases. Both curves look symmetric and it seems that sp carbon atoms are evolving towards a sp^2 hybridization. It corresponds to the known⁹ trend of type III kerogen to evolve towards an aromatic structure during its maturation. The fact that sp^2 proportion is already high in lignin initial sample and stays high suggests that initial aromatic rings from monolignols remain essentially intact during the run, and new ones are formed right from the beginning of the decomposition process. Percentage of sp^3 carbons stays almost constant all along the simulation.

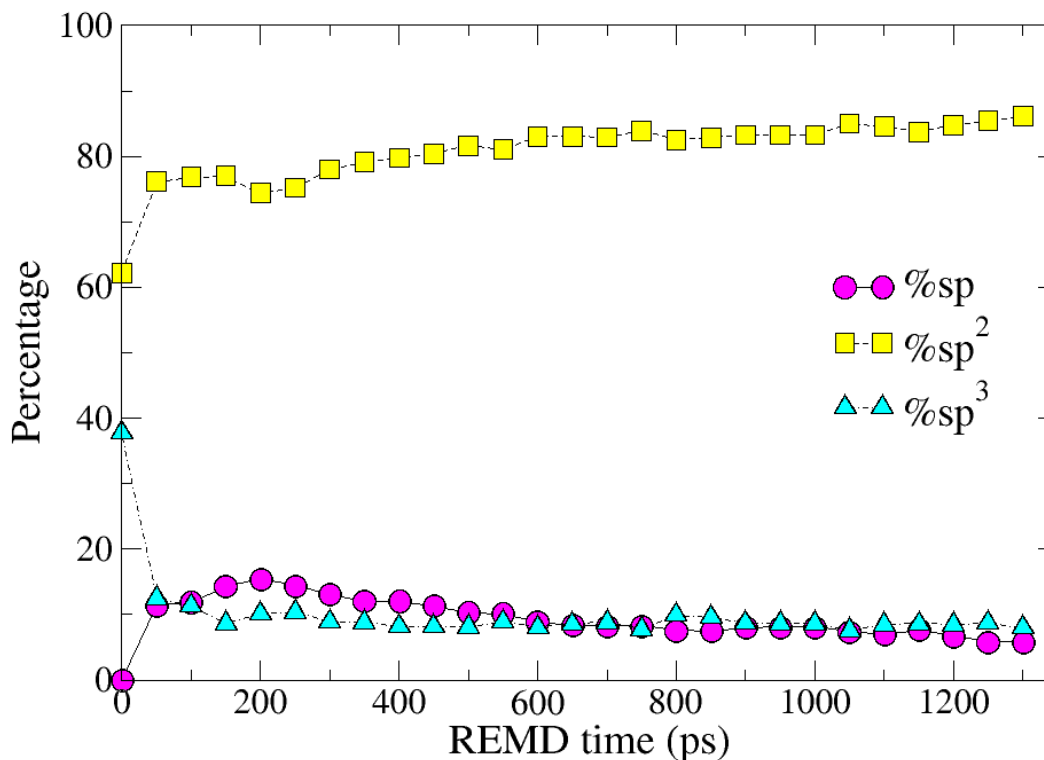


Figure 75. Evolution of carbon atoms hybridization in the solid part during REMD simulation. Percentage of sp hybridized carbon atoms (magenta circles), of sp² hybridized carbon atoms (yellow squares), and of sp³ hybridized carbon atoms (cyan triangles).

Snapshots of the solid-like part all along the simulation are shown in Figure 76. Size evolution is not significant and according to Figure 74, after 200 ps of REMD, number of carbon atoms in the solid phase does not significantly evolve. It goes from 1201 carbon atoms at 50ps to 1364 at 350ps. It then saturates around 1375 carbon atoms until the end of simulation. We can note the important decrease of oxygen atoms in the lignin polymeric chain –in red in Figure 76– during the run. There are indeed 237 oxygen atoms at 50ps of REMD, and only 15 remain after 1300ps. In the same time, hydrogen atom number decreases from 932 at 50ps to 697 at 1300ps of REMD. These evolutions can be mainly explained by water formation. Half of oxygen atoms leaving solid, will combined all the hydrogen. The other half will find hydrogen –to form mainly water– and carbon –to form mainly carbon monoxide. This trend correlates with the one observed in Figure 72: the increase of H/C and O/C atomic ratios for the fluid phase and the decrease of them for the solid.

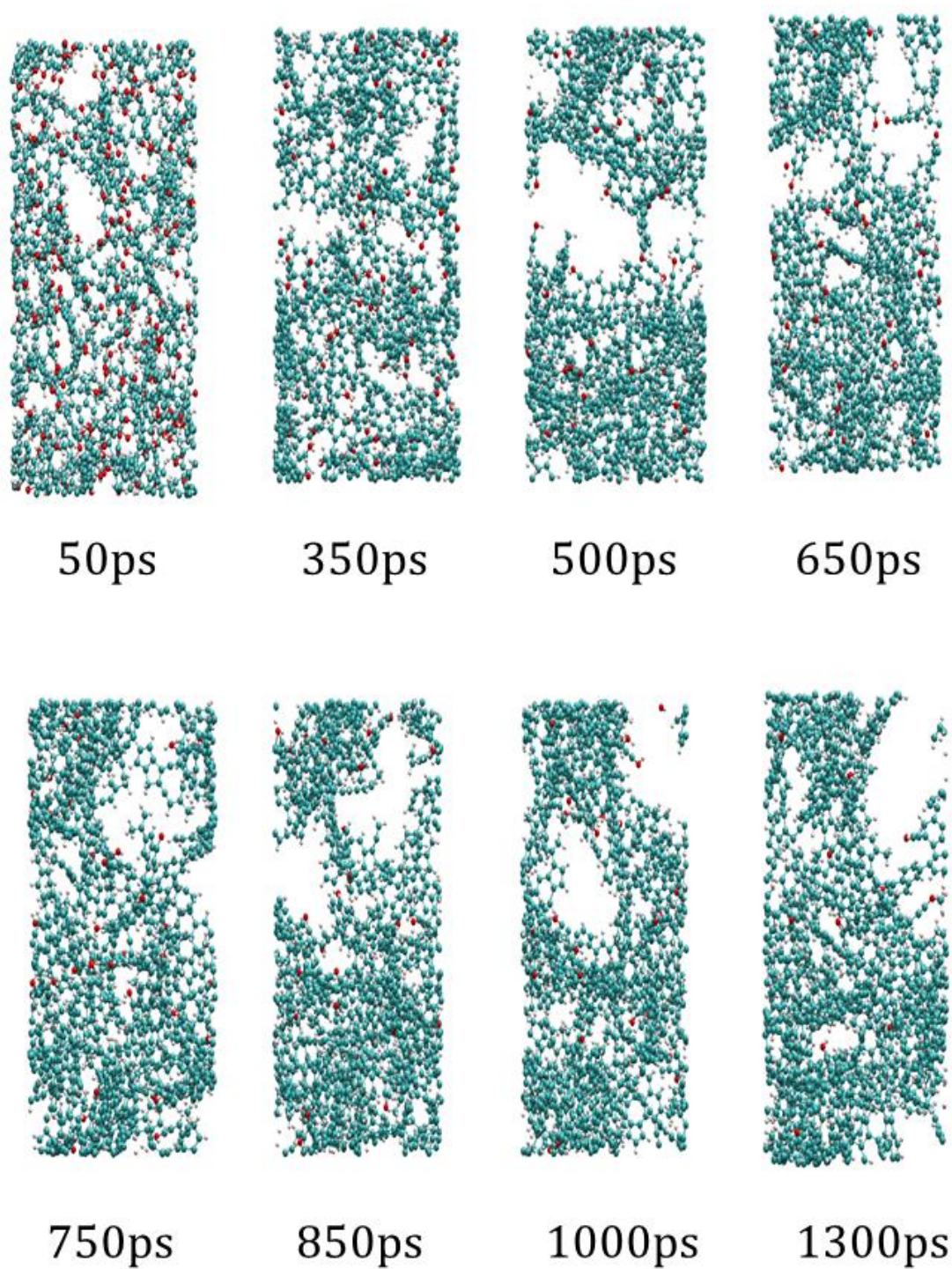


Figure 76. Evolution of the solid-like phase during REMD run on the cellulosic system. Carbon atoms in blue, hydrogen atoms in white, and oxygen atoms in red.

Density is also an interesting characteristic of the solid phase. We already explained – see cellulose section- that kerogen density range is wide, and generally comprised in a range of [1.12-1.32 g.cm⁻³]. In our lignin REMD run, solid part density value is not meaningful since the phase is constrained by the fluid presence. Obviously, the density we compute does not correspond to the one if the fluid phase found a way to escape, like in reality. Despite these facts, we plotted density evolution for both the whole system and the solid-like part, in order to compare their trends –see in Figure 77. Both densities, for solid part and the whole system, decrease during the first 400ps of REMD, before reaching a plateau. The whole system density then fluctuates around 1.12g.cm⁻³, and the solid phase density around 0.65g.cm⁻³. The whole system density corresponds to experimental data range, although it converges towards a lower value than experiment. The whole lignin system density is larger than the whole cellulose one, and it might be explained by the formation of more solid in the case of lignin than in the case of cellulose, and therefore the lesser formation of fluid from lignin.

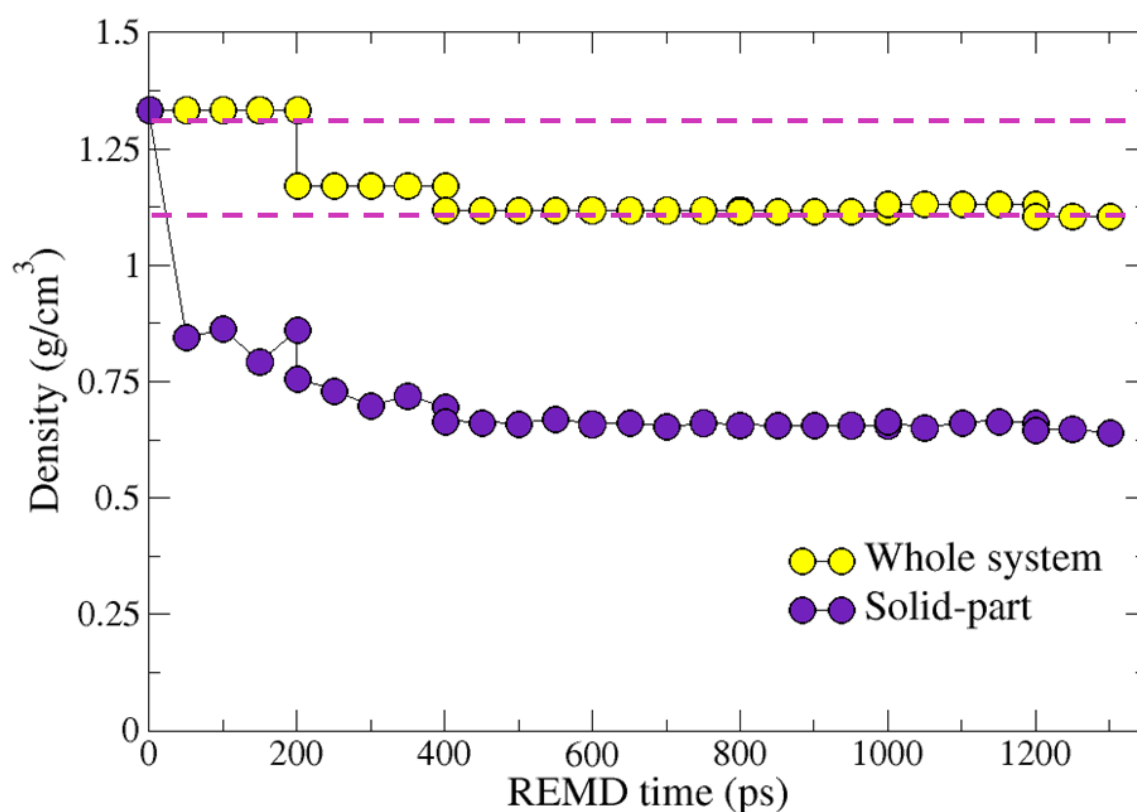


Figure 77. Density evolution of the whole system (yellow circles) and of the solid-like phase (purple circles). Violet dashed lines represent literature range^{18,20}.

At the end of this REMD simulation, isothermal compressibility of the solid-like part is of $5.71 \cdot 10^{-9} \text{Pa}^{-1}$, which corresponds to a bulk modulus of 0.18GPa. This value is close to that found by Collell *et al*²². $-2.7 \cdot 10^{-10} \text{Pa}^{-1}$ for a type II kerogen –defined as “*in the middle*

of oil generation window"- and $2.9 \times 10^{-8} \text{ Pa}^{-1}$ for a bulk fluid. Our solid is thus to be considered as an immature kerogen. A second REMD simulation, only on the solid-part would be interesting for a further work, in order to simulate its maturation.

Van Krevelen diagram

Although lignin is a significant component of higher plants, its pyrolysis has been less studied than cellulose and hemicellulose ones. One of the reasons is that lignin is less utilized than these two other major component of biomass. Anyway, studies and review exist^{45,52-54}.

In Figure 78 are compared H/C versus O/C evolutions of (i) our lignin system during the REMD pyrolysis, (ii) of Sharma *et al.*'s⁴³ -open pyrolysis-, and (iii) of Monthioux *et al.*'s³¹ type III kerogen -confined pyrolysis. We also plotted H/C versus O/C evolution from another numerical simulation of lignin pyrolysis⁵⁵.

Sharma *et al.*⁴³ pyrolyzed lignin samples in order to study resulting chars. Their experiments were made under atmospheric pressure, and at temperatures in the range [423K-1023K]. They found, as expected, a decrease of hydrogen and oxygen content in char, along with an increase of the carbon one. Evolution of H/C and O/C atomic ratios are plotted on a Van Krevelen diagram –in Figure 78. Going from top right of the graph to bottom left, experiment evolves with temperature. We note that, even if evolution trend of ratios –both decreasing- is qualitatively consistent with natural evolution, values are lower than expected for a type III kerogen.

Monthioux *et al.*³¹, got results in better agreement with the evolution of natural samples. Our simulated lignin decomposition follows a similar pathway as that found by Monthioux *et al.*

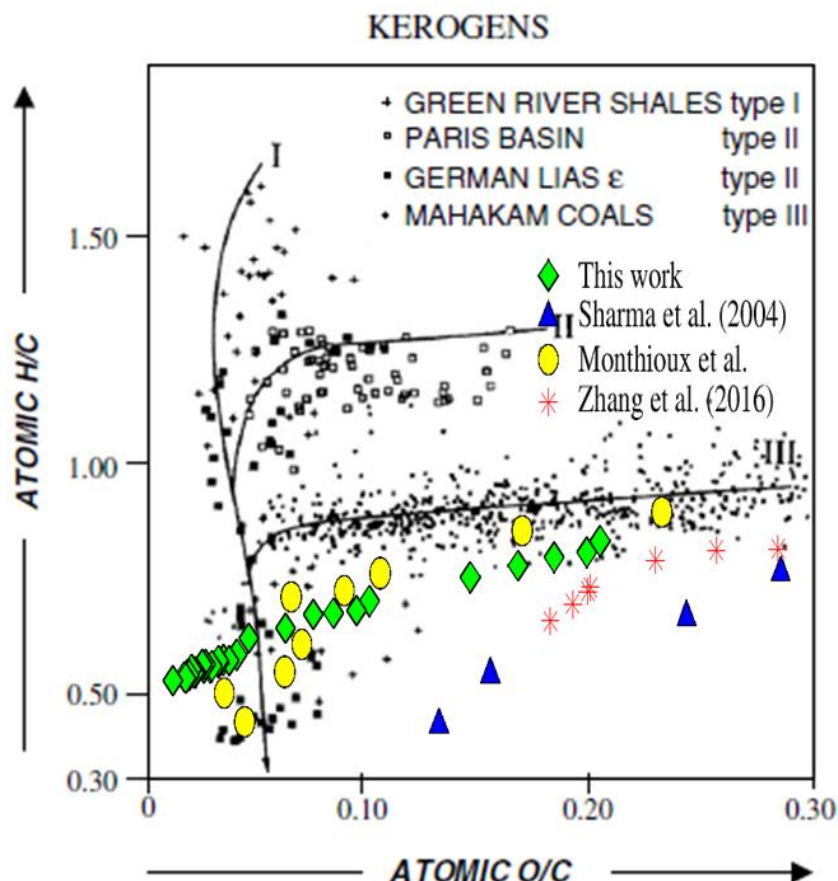


Figure 78. Van Krevelen diagram from Vandenbroucke and Largeau⁹ with values for our work on lignin (green diamonds), values of lignin open pyrolysis from Sharma *et al.*⁴³ (blue triangles), values from type III kerogen confined pyrolysis from Monthioux *et al.*³¹ (yellow circles), and from numerical simulation of lignin pyrolysis by Zhang *et al.*⁵⁵ (red stars).

Our simulated solid, obtained from lignin REMD pyrolysis, also follows a decomposition pathway close to natural samples ones. It is remarkable. Note that the simulations of lignin decomposition by Zhang *et al.*⁵⁵ using conventional high temperature MD with ReaxFF –Mattsson’s implementation- did not come as close to the experimental data as ours. This underlines the creed of using advanced statistical physics method. Zhang and co-worker studied initial lignin pyrolysis. This is the reason why their O/C atomic ratio does not decrease below 0.15. They performed both heat-up simulations –from 300 to 2100K- and NVT runs –at 500 to 2100K- and found a three stages decomposition. First stage, between 700K and 1100K, is lignin decomposition with formation of water and three carbon atoms molecules. Stage two, between 1100K and 1500K, is the period where methane and methanol, in particular, are formed. Stage three starts around 1500K and corresponds to formation of dihydrogen, carbon dioxide, and carbon monoxide. We do not see the same sequence in our system decomposition. Except for the first stage –lignin decomposition with formation of water- major steps in molecules formation as inferred by Zhang *et al.*⁵⁵ do not clearly appear. Methane, carbon monoxide, and dihydrogen all form around the same time –see Figure 73. The reason

why our simulations –and more generally all experiments here- have lower values of H/C atomic ratio than natural samples is unclear. It seems to be due to a higher loss of hydrogen –and not a higher composition in carbon- since the O/C atomic ratio seems to follow the natural decomposition pathway.

Comparison between cellulose and lignin

In this section, we give a comparison of solid and fluid parts obtained from whole cellulose and lignin systems pyrolysis.

In our simulation the solid-like part considered as kerogen in maturation, evolves during pyrolysis, gaining carbon atoms; These atoms hybridization evolving towards a majority of sp^2 aromatic bonds. The evolution of carbon atoms percentage in the solid phase and percentage of carbon atoms hybridized sp^2 are given in Figure 79. If we look at the proportions of carbon in the solid phase, we note a clear difference between lignin and cellulose. They both start at 100% and drop at the very beginning of simulation but this drop is larger for cellulose. It indeed drops to almost 5% while lignin drops only to 75%. Fragmentation of cellulose is therefore much more important than lignin one. It can be explain by the larger stability of lignin, due to its higher aromaticity. Both carbon atoms percentage curves then converge around values of 85% and 60% for lignin and cellulose respectively.

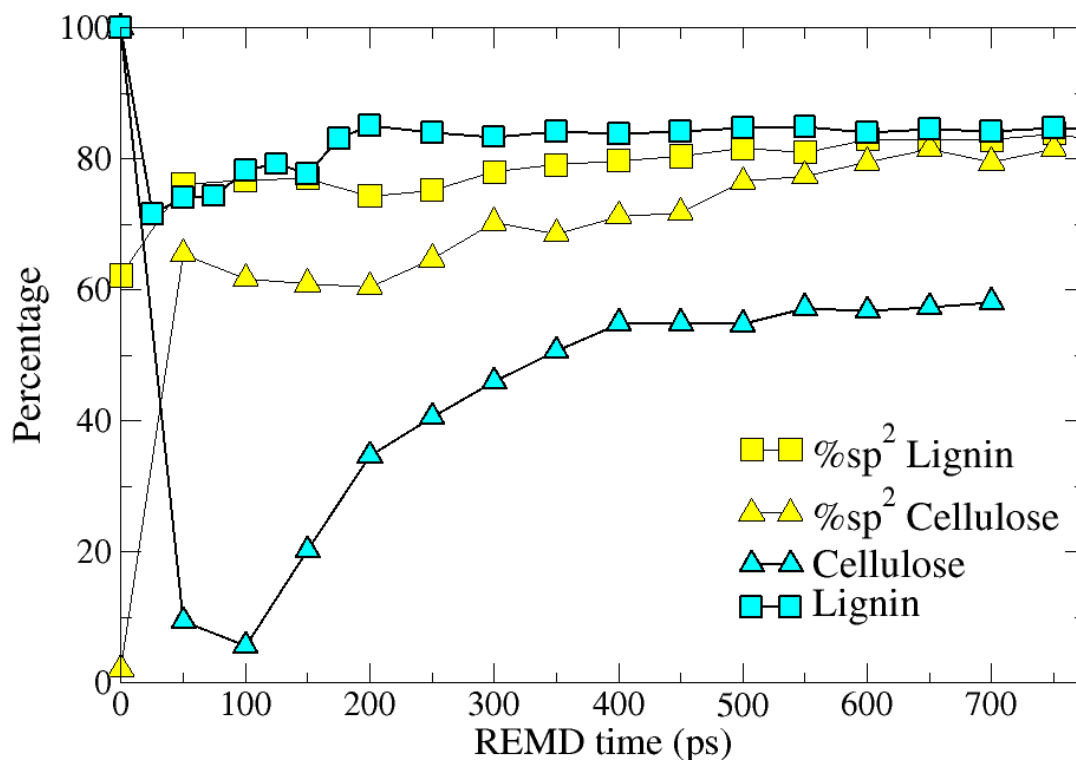


Figure 79. Evolution of the percentage of carbon atoms in the solid like phase (cyan) and of sp^2 hybridized carbon atoms (yellow) during REMD runs (at 423K) on the whole systems of lignin (squares) and cellulose (triangles).

If we look at radial distribution functions of solid phases for both systems after 750ps of REMD –in Figure 80- we can see a comparable qualitative trend but with different quantitative aspects. We note a greater second peak, around $1.47\text{\AA}$, equilibrium distance for sp^2 carbon-carbon bond, for the lignin solid. Cellulose second peak has a hump around $1.33\text{\AA}$ that might correspond to sp carbon-carbon bonds. Cellulosic solid shows off a more intense first peak -around $1.1\text{\AA}$. No carbon-carbon bond is known at this distance. It might be an issue from the interaction potential.

We also note there is not a great difference between lignin system radial distribution functions after 750 and 1300ps of REMD. It is mainly a smooth increase of sp^2 proportion –increase at $1.47\text{\AA}$ - with a slight decrease of the sp one –decrease at $1.35\text{\AA}$. This trend confirms evolution visible in Figure 75. Proportion of sp^3 –around $1.54\text{\AA}$ - does not change with time.

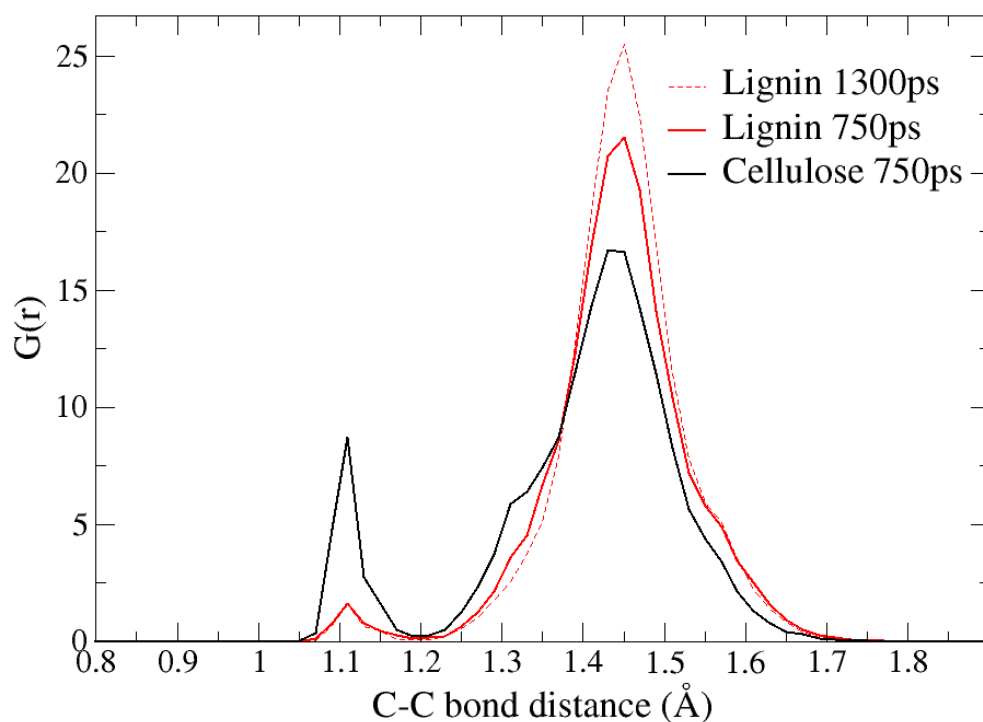


Figure 80. Radial distribution functions of carbon-carbon bonds in cellulose after 750ps of REMD (plain black line), lignin after 750ps (plain red line), and lignin after 1300ps of REMD (dashed red line).

Aromaticity of lignin is visible thanks to the high proportion of sp^2 carbons, and is due to initial structures of monolignols, constituents of lignin. Lignin aromatic carbon structure has been shown to be more stable than the carbon cellulose structure^{43–45}. We note that percentage of carbon atoms hybridized sp^2 converges towards the same value –around 80%– for both lignin and cellulose systems. It supports the idea of a common kerogen for every type III kerogen precursors. Organic matter with high level of cellulose and low level of lignin would eventually decompose into the same carbonaceous structure than organic matter with low level of cellulose and high level of lignin.

This common structure does not depend on the char yield of the system during its pyrolysis. Indeed, char yield of cellulose pyrolysis is lower than lignin one but they both lead to an equivalent structure. Snapshots of solid part for cellulose and lignin, at different REMD times are given in Figure 81.

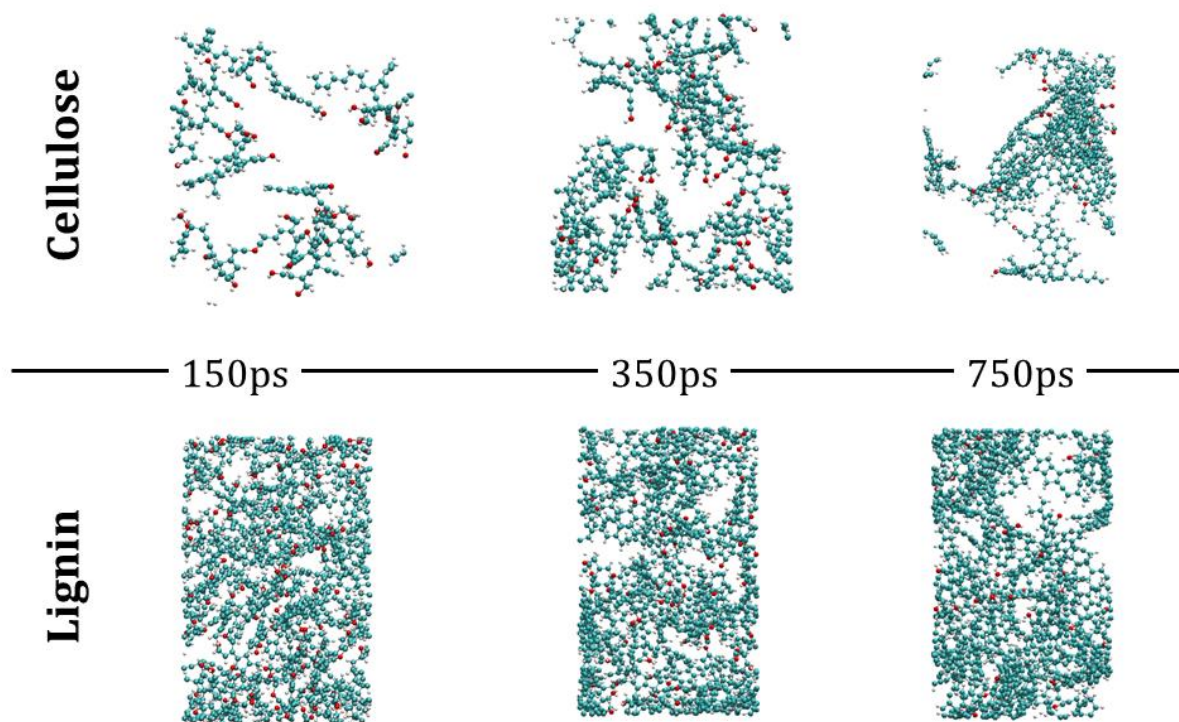


Figure 81. Snapshots of the solid phase at different times (150, 350, and 750ps) of REMD on the whole systems of cellulose and lignin.

We clearly note that the lignin solid part is much substantial than the cellulose one. It supports the difference between both systems percentages of carbon atoms in the solid phase noticeable in Figure 79. Evolution of carbons hybridization is also obvious in snapshots of cellulosic system, with clear carbon chains at 150ps of REMD, and aromatic structures at 750ps.

In Table 6, compositions of solid phases for both cellulose and lignin systems are given. We clearly note the difference in molecules sizes. The largest molecule for lignin is composed by 1014 carbon atoms while the one for cellulose has only 42 of them, after 150ps of REMD.

This difference in carbon atoms number consolidates information given in Figure 79 and in Figure 81, and describes a more dispersed solid for the cellulose system than for the lignin one.

Table 6. Composition of solid-like phases for cellulose and lignin systems at different times of pyrolysis. Only the largest molecule is reported. The number of carbon atoms is given for the others.

Solid part system	150ps of REMD	350ps of REMD	750ps of REMD
Cellulose	C₄₂H₃₄O₆ C ₃₄ , C ₃₀ , C ₂₄ , C ₂₃ , C ₁₆ , C ₁₅ , C ₁₃	C₉₀H₅₄O₁₃ C ₅₈ , C ₄₀ , C ₃₉ , C ₃₁ , C ₂₈ , C ₂₆ , C ₂₃ , C ₂₂ , C ₂₀ , C ₁₈ , C ₁₅ , C ₁₃ , C ₁₂	C₄₃₄H₂₂₉O₁₆ C ₂₃₅ , C ₁₃
Lignin	C₁₀₁₄H₆₇₇O₉₆ C ₆₄ , C ₄₀ , C ₃₀ , C ₂₈ , C ₂₅ , C ₁₇ , C ₁₅ , C ₁₃ , C ₁₂	C₁₂₇₆H₇₆₃O₇₁ C ₂₉ , C ₁₉ , C ₁₆ , C ₁₂	C₁₃₆₁H₇₃₀O₃₈ C ₁₂

Although differences exist on solid-part formation during lignin and cellulose pyrolysis, its evolution, plotted in a Van Krevelen diagram –see in Figure 82- is comparable for both systems.

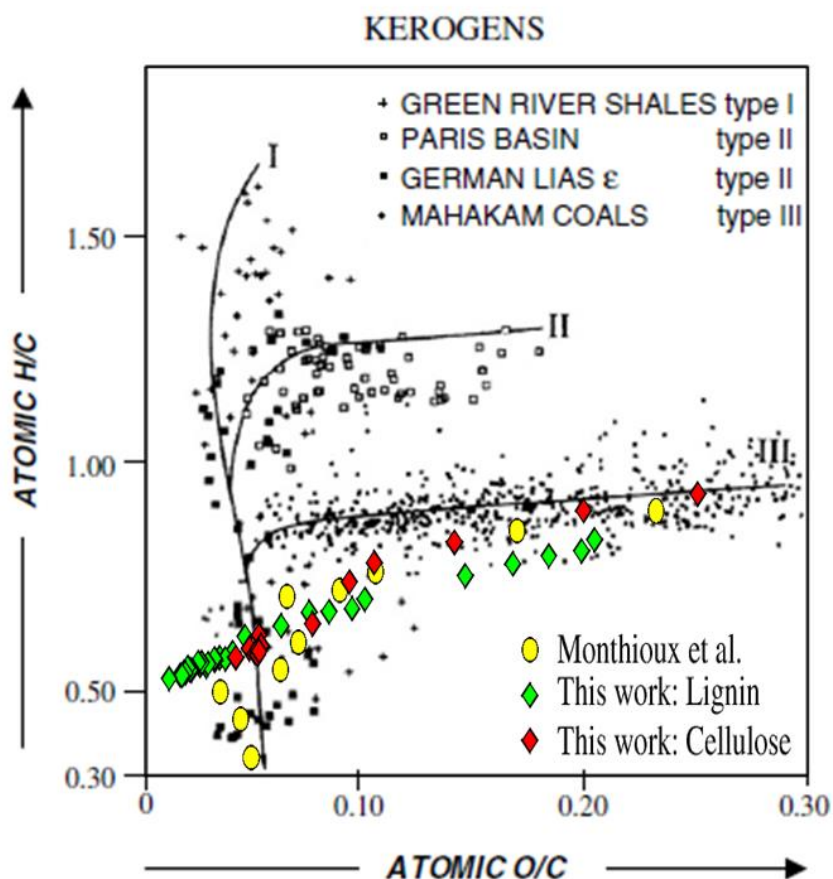


Figure 82. Van Krevelen diagram from Vandenbroucke and Largeau⁹ with values from Monthieux *et al.*³¹ (yellow circles), our work on lignin (green diamonds), and our work on cellulose (red diamonds), both at 423K.

Data for both cellulose and lignin systems are found within the cloud of points of natural samples for type III kerogen. They are in good agreement with experimental

confined pyrolysis data of Monthioux *et al.*³¹ for diagenesis and catagenesis stages. In our systems metagenesis is not reproduced. We think that, since metagenesis mostly consists in a structural rearrangement, the systems need more energy, or a longer REMD simulation to be able to reproduce this latest stage of decomposition. Cellulose evolves at a slightly higher H/C atomic ratio but it is consistent with the different initial H/C and O/C atomic ratios for both systems. Initial cellulose has initial H/C and O/C ratios of 1.67 and 0.83 respectively, while lignin starts with H/C and O/C ratios of 1.11 and 0.34 respectively. The lower H/C ratio for lignin explains why this system evolves, in a first instance, at a lower H/C ratio. Cellulose and lignin curves then cross and join a same evolution pathway, around an O/C atomic ratio of 0.09. O/C atomic ratio of 0.09 corresponds to a REMD simulation time of 200ps for lignin and of 250ps for cellulose. Those times do not correspond to any special occurrence in both systems.

Besides differences between solid phases compositions of the fluid ones are also different. In Figure 83 are compared numbers of common molecules –H₂, H₂O, CO, CO₂, and CH₄– for both lignin and cellulose, and at different simulation times. These values have to be analyzed keeping in mind the difference in initial numbers of carbon, hydrogen, and oxygen atoms in both systems. Cellulose is initially composed of 1200 C, 2000 H, and 1000 O. Lignin is initially composed of 1620 C, 1800 H, and 559 O.

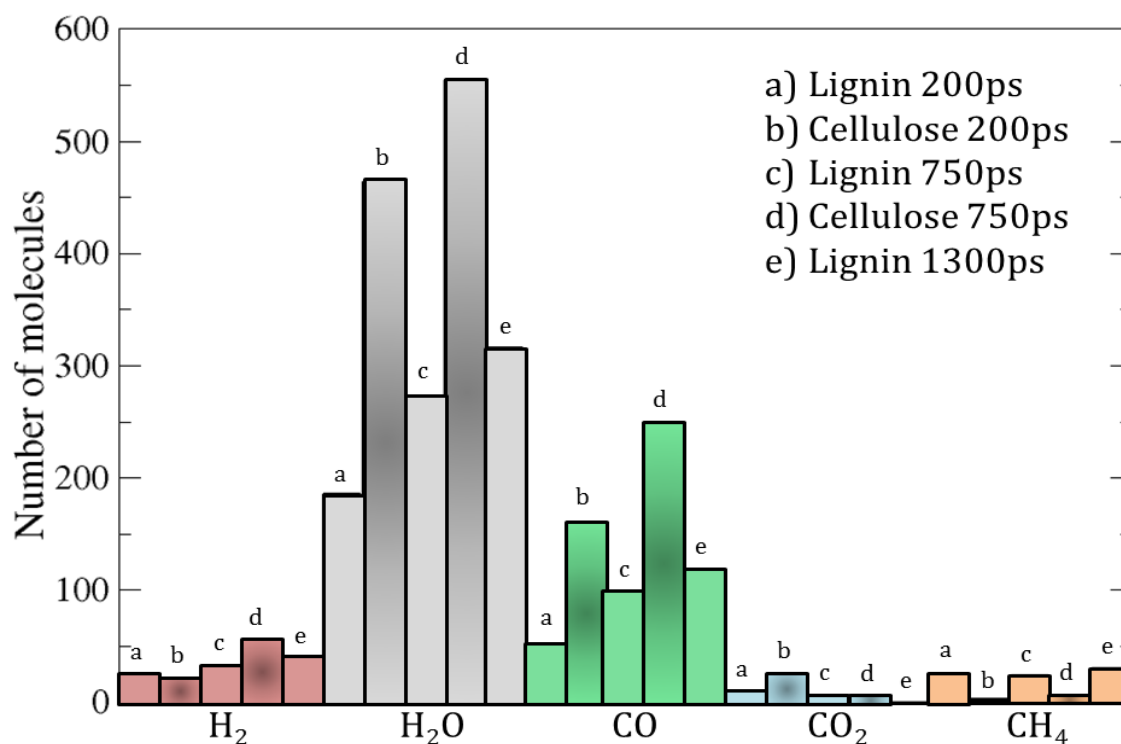


Figure 83. Number of common molecules present in the the fluid phase, at different times of REMD on the whole systems of lignin and cellulose. **(a)** Lignin after 200ps, **(b)** cellulose after 200ps, **(c)** lignin after 750ps, **(d)** cellulose after 750ps, and **(e)** lignin after 1300ps of REMD.

For both systems proportions of H_2 , H_2O , CO , and CH_4 increase during the REMD simulation, and CO_2 one decreases. Cellulose pyrolysis forms far more H_2O , CO_2 , and CO than lignin while lignin pyrolysis goes along a higher formation of CH_4 . This difference in methane formation validates our method since it is known⁴⁷ that lignin pyrolysis forms more methane than the cellulose one.

These comparisons between cellulose and lignin systems pyrolysis, show a qualitatively similar decomposition pathway. They both decompose into a two-phase system. Fluid phase is mainly composed of dihydrogen, water, and carbon monoxide. In both cases few molecules of carbon dioxide appear, but not as much as expected. It may be due to the reactive potential used for this study. On the other hand, as expected, lignin degradation leads to a higher formation of methane than cellulose. Solid phase is a disordered, mainly sp^2 , carbonaceous structure. Its formation is different for both systems. Early cellulose fragmentation forms small chains of carbons –with heterogeneous atoms- that coalesce once fluid formation starts to plateau. Initial lignin aromatic rings are not fragmented but assemble to form the solid part. These differences are due to initial structure –mainly sp^2 carbon for lignin and sp^3 for cellulose- and composition –initial H/C and O/C atomic ratios of 1.67 and 0.83 respectively for cellulose and of 1.11 and 0.34 for lignin respectively. Despite those initial differences, we noted both solid phases converge toward a similar carbon phase after a given REMD time, and they both follow natural type III kerogen evolution. Therefore, our numerical simulations well reproduce the natural evolution.

Summary

Cellulose is the main constituent of plants and is the most abundant organic matter on Earth. It is a natural polymer, composed of linear chains of glucose molecules rings linked through an oxygen bridge and of formula $(C_6H_{10}O_5)_n$. As we mentioned before, type III kerogen mainly results from higher plants. Since cellulose is their main component, it appears as a natural type III kerogen precursor model.

To build our numerical model of cellulose –which will be initial configuration of our simulation- we used the cellulose-builder code. It is a freely available code that allows easily constructing different polymorphs of cellulose. We built a 4200 atoms system of cellulose I β .

We run a first REMD simulation on the whole cellulosic system of 4200 atoms, alternating with NPT relaxation in order to keep system pressure in a geological range –

between 25 and 125 MPa. This first simulation step was 750ps of REMD long. System energy was decreasing, indicating an exothermic process toward a more stable configuration at the given temperature -423K. During the decomposition process, two phases appeared: a fluid one, mainly composed by water and carbon monoxide, hydrogen and oxygen rich, and a solid-like phase, rich in carbon atoms. These two phases are formed in two main steps: first, the formation of the fluid phase, with the fragmentation of cellulose crystal, and then the coalescence of the solid-like part. Growth of the solid phase really starts once the number of small molecules as water starts to stabilize. After 750ps energy converges to an equilibrium state. We stopped the simulation and started the second step: an REMD simulation on the solid-like phase.

Solid-like phase of the first REMD stage was relaxed. It was then submitted to an alternation of REMD and NPT relaxations, and that, for 1800ps. System energy decreases during the whole simulation, until convergence. During this REMD run, although energy decreases, chemical composition of the system does not drastically change since, at the end, only 10 molecules of water have been formed, 7 of carbon monoxide and 5 of dihydrogen. Energy decrease is related to structural rearrangement, since average number of rings per cluster increases. It does not increase enough to describe metagenesis. This second cellulose REMD step seems to correspond more to late catagenesis.

Evolution of the solid-like phase of decomposed cellulose is plotted on a Van Krevelen diagram and compared with experimental results. It appears to well reproduce natural evolution of a type III kerogen. Points corresponding to our system are comprised within the type III kerogens cloud of experimental data.

We then took the step to study lignin, as a complement of cellulose in terrestrial plant constitution. Term lignin refers to a family of aromatic polymers, composed of three main blocks called monolignols. Lignin possible structures are abundant. Banoub and Delmas^{38,41} affirm that linear structures are the most probable. We thus looked for a linear, quite easy to build, structure and chose the one from Crestini *et al.*⁴². We built a 3978 atoms structure, with initial H/C and O/C atomic ratios of 1.11 and 0.34 respectively. This structure was relaxed and then submitted to a REMD simulation.

REMD simulation on the whole system of lignin was 1300ps long, and NPT relaxations –at 423K and 50MPa- were performed every 200ps. Those relaxations are used to keep the system in a geological pressure range of [25-125MPa]. REMD parameters were the same than for cellulose -96 temperatures going from 423 to 3500K, swaps attempted every 10fs. System energy decreased –indicating an evolution towards a more stable configuration at 423K- at convergence. Pressure evolution shows a similar trend than for cellulose, with the formation of water. A decomposition of the system into two phases –fluid and solid- appears, like for cellulose.

Fluid phase is hydrogen and oxygen rich, and mainly composed by water, carbon monoxide, and dihydrogen, while the solid-like part is carbon rich, made of sp^2 bonds. This predominant hybridization mainly comes from initial monolignols units that are composed by aromatic rings. Unlike cellulose, lignin is not fragmented into small carbon chain molecules, but aromatic rings assemble to form the solid part. Although qualitative aspects are the same for both lignin and cellulose systems, differences remain regarding the quantitative composition of fluid. As expected, lignin pyrolysis forms more methane than cellulose. Solid structural configurations are similar, with hybridization heading towards sp^2 domination, but isothermal compressibilities at the end of both first REMD simulations -750ps for cellulose and 1300ps for lignin- differ from each other. Cellulose solid part, after 750ps of REMD, has an isothermal compressibility of $1.64 \cdot 10^{-10} \text{ Pa}^{-1}$, while lignin solid phase isothermal compressibility is of $5.71 \cdot 10^{-9} \text{ Pa}^{-1}$. Collell *et al*²². found a value of $2.7 \cdot 10^{-10} \text{ Pa}^{-1}$ for a type II kerogen –defined as “*in the middle of oil generation window*”. Both solid phases are thus at different maturity levels.

Despite those differences, both cellulose and lignin solid systems decomposition pathways are consistent with type III kerogen natural samples evolution. Replica exchange molecular dynamics well reproduces thermal degradation of organic matter, under geological conditions. Therefore, we are able to reproduce a geological time scale process, using numerical simulations, with characteristic time-scale of few nanoseconds.

1. Yamamoto, H. & Horii, F. Analysis of the Crystal Transformation Induced for Valonia cellulose by Annealing at High Temperatures. *Macromolecules* **26**, 1313–1317 (1993).
2. Belton, P., Tanner, S. & Cartier, N. High-resolution solid-state carbon-13 nuclear magnetic resonance spectroscopy of tunicin, an animal cellulose. *Macromolecules* **22**, 1615–1617 (1989).
3. Gomes, T. C. F. & Skaf, M. S. Cellulose-builder: A toolkit for building crystalline structures of cellulose. *J. Comput. Chem.* **33**, 1338–1346 (2012).
4. Nishiyama, Y., Sugiyama, J., Chanzy, H. & Langan, P. Crystal Structure and Hydrogen Bonding System in Cellulose I α from Synchrotron X-ray and Neutron Fiber Diffraction. *J Am Chem Soc* **125**, 14300–14306 (2003).
5. Nishiyama, Y., Sugiyama, J., Chanzy, H. & Langan, P. Crystal Structure and Hydrogen Bonding System in Cellulose I β from Synchrotron X-ray and Neutron Fiber Diffraction. *J Am Chem Soc* **124**, 9074 – 9082 (2002).
6. Langan, P., Nishiyama, Y. & Chanzy, H. X-ray structure of mercerized cellulose II at 1 Å resolution. *Biomacromolecules* **2**, 410–416 (2001).
7. Wada, M., Chanzy, H., Nishiyama, Y. & Langan, P. Cellulose III I crystal structure and hydrogen bonding by synchrotron X-ray and neutron fiber diffraction. *Macromolecules* **37**, 8548–8555 (2004).
8. Zhang, W., Wu, C. & Duan, Y. Convergence of replica exchange molecular dynamics. *J. Chem. Phys.* **123**, 154105 (2005).
9. Vandenbroucke, M. & Largeau, C. Kerogen origin, evolution and structure. *Org. Geochem.* **38**, 719–833 (2007).
10. Durand, B. *Kerogen: Insoluble organic matter from sedimentary rocks*. (Editions technip, 1980).
11. Tissot, B. P. & Welte, D. H. *Petroleum formation and occurrence*. (Springer Science & Business Media, 2013).
12. Tarjan, R. Depth-first search and linear graph algorithms. *12th Annu. Symp. Switch. Autom. Theory (swat 1971)* **1**, 146–160 (1971).
13. Satterfield, Jr., H. & Longwell. Product Distribution From Iron Catalysts in Fischer-Tropsch Slurry Reactors. *Ind. Eng. Chem. Process Des. Dev.* **21**, 465–470 (1982).
14. Graboski, M. S. a Mathematical Model for Stratified Downdraft Gasifiers. 410–420
15. Demirel, E. & Azcan, N. Thermodynamic Modeling of Water-Gas Shift Reaction in Supercritical Water. *Proc. World Congr. Eng. Comput. Sci.* **II**, 24–27 (2012).
16. Liu, J. a. *Kinetics, catalysis and mechanism of methane steam reforming*. (2006).
17. Hajaligol, M. R., Howard, J. B., Longwell, J. P. & Peters, W. a. Product compositions and kinetics for rapid pyrolysis of cellulose. *Ind. Eng. Chem. Process Des. Dev.* **21**, 457–465 (1982).
18. Okiongbo, K. S., Aplin, A. C. & Larter, S. R. Changes in type II Kerogen density as a function of maturity: Evidence from the Kimmeridge clay formation. *Energy and Fuels* **19**, 2495–2499 (2005).
19. Bousige, C. *et al.* Realistic molecular model of kerogen's nanostructure. *Nat. Mater.* **15**, 576–583 (2016).
20. Stankiewicz, B. A., Kruger, M. a., Crelling, J. C. & Salmon, G. L. Density gradient centrifugation: application to the separation of macerals of Type I, II, and III sedimentary organic matter. *Energy & Fuels* **8**, 1513–1521 (1994).
21. Lagache, M., Ungerer, P., Boutin, a. & Fuchs, a. H. Prediction of thermodynamic derivative properties of fluids by Monte Carlo simulation. *Phys. Chem. Chem. Phys.* **3**, 4333–4339 (2001).

22. Collell, J. *et al.* Molecular Simulation of Bulk Organic Matter in Type II Shales in the Middle of the Oil Formation Window. *Energy & Fuels* **28**, 7457–7466 (2014).
23. Pictet, A. & Sarasin, J. Sur la distillation de la cellulose et de l'amidon sous pression réduite. *Helv. Chim. Acta* **1**, 87 (1918).
24. Lin, Y.-C., Cho, J., Tompsett, G. A., Westmoreland, P. R. & Huber, G. W. Kinetics and Mechanism of Cellulose Pyrolysis. *J. Phys. Chem.* **113**, 20097–20107 (2009).
25. Luo, Z., Wang, S., Liao, Y. & Cen, K. Mechanism Study of Cellulose Rapid Pyrolysis. *Ind. Eng. Chem. Res.* **43**, 5605–5610 (2004).
26. Banyasz, J. L., Li, S., Lyons-Hart, J. & Shafer, K. H. Gas evolution and the mechanism of cellulose pyrolysis. *Fuel* **80**, 1757–1763 (2001).
27. Agarwal, V., Dauenhauer, P. J., Huber, G. W. & Auerbach, S. M. Ab initio dynamics of cellulose pyrolysis: Nascent decomposition pathways at 327 and 600 °C. *J. Am. Chem. Soc.* **134**, 14958–14972 (2012).
28. Zheng, M. *et al.* Initial reaction mechanisms of cellulose pyrolysis revealed by ReaxFF molecular dynamics. *Fuel* **177**, 130–141 (2016).
29. Schwenker, R. F. & Pacsu, E. Chemically modifying cellulose for flame resistance. *Ind. Eng. Chem.* **50**, 91–96 (1958).
30. Demirbas, A. Gaseous products from biomass by pyrolysis and gasification: Effects of catalyst on hydrogen yield. *Energy Convers. Manag.* **43**, 897–909 (2002).
31. Monthieux, M., Landais, P. & Monin, J.-C. Comparison between natural and artificial maturation series of humic coals from the Mahakam delta, Indonesia. *Org. Geochem.* **8**, 275–292 (1985).
32. Tang, M. & Bacon, R. Carbonization of cellulose fibers—I. Low temperature pyrolysis. *Carbon N. Y.* **2**, 211–220 (1964).
33. Almardottir, A. R. & Gunnarsson, I. B. *Framleiðsla etanóls úr flóknum lífmassa með hitakærum bakteríum.* (2009).
34. Vanholme, R., Demedts, B., Morreel, K., Ralph, J. & Boerjan, W. Lignin biosynthesis and structure. *Plant Physiol.* **153**, 895–905 (2010).
35. How about lignin. Available at: <http://howaboutlignin.blogspot.fr/>. (Accessed: 14th August 2016)
36. Icfar Lignin. Available at: <http://www.icfar.ca/lignoworks/content/what-lignin.html>. (Accessed: 14th August 2016)
37. Lignine. Available at: <http://www.societechimiquedefrance.fr/lignine.html>. (Accessed: 1st November 2016)
38. Banoub, J. *et al.* A critique on the structural analysis of lignins and application of novel tandem mass spectrometric strategies to determine lignin sequencing. *J. Mass Spectrom.* **50**, 5–48 (2015).
39. Freudenberg, K. & Neish, A. C. *Constitution and biosynthesis of lignin.* (Springer-Verlag, 1968).
40. Adler, E. Lignin chemistry-past, present and future. *Wood Sci. Technol.* **11**, 169–218 (1977).
41. Banoub, J. H. & Delmas, M. Structural elucidation of the wheat straw lignin polymer by atmospheric pressure chemical ionization tandem mass spectrometry and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J. Mass Spectrom.* **38**, 900–903 (2003).
42. Crestini, C., Melone, F., Sette, M. & Saladino, R. Milled wood lignin: A linear oligomer. *Biomacromolecules* **12**, 3928–3935 (2011).
43. Sharma, R. K. *et al.* Characterization of chars from pyrolysis of lignin. *Fuel* **83**, 1469–1482 (2004).

44. Xie, X. *et al.* Characterization of carbons derived from cellulose and lignin and their oxidative behavior. *Bioresour. Technol.* **100**, 1797–1802 (2009).
45. Brebu, M. & Vasile, C. Thermal degradation of lignin- a review. *Cellul. Chem. Technol.* **44**, 353–363 (2010).
46. Caballero, J. A., Font, R., Marcilla, A. & García, A. N. Flash pyrolysis of Klason lignin in a Pyroprobe 1000. *J. Anal. Appl. Pyrolysis* **27**, 221–244 (1993).
47. Caballero, J. A., Font, R. & Marcilla, A. Pyrolysis of Kraft lignin: yields and correlations. *J. Anal. Appl. Pyrolysis* **39**, 161–183 (1997).
48. Iatridis, B. & Gavalas, G. R. Pyrolysis of a Precipitated Kraft Lignin. *Ind. Eng. Chem. Prod. Res. Dev.* **18**, 127–130 (1979).
49. Liu, Q., Wang, S., Zheng, Y., Luo, Z. & Cen, K. Mechanism study of wood lignin pyrolysis by using TG-FTIR analysis. *J. Anal. Appl. Pyrolysis* **82**, 170–177 (2008).
50. Ferdous, D., Dalai, A. K., Bej, S. K. & Thring, R. W. Pyrolysis of lignins: Experimental and kinetics studies. *Energy and Fuels* **16**, 1405–1412 (2002).
51. Wang, S. *et al.* Comparison of the pyrolysis behavior of lignins from different tree species. **27**, 562–567 (2009).
52. Pandey, M. P. & Kim, C. S. Lignin Depolymerization and Conversion: A Review of Thermochemical Methods. *Chem. Eng. Technol.* **34**, 29–41 (2011).
53. Mu, W., Ben, H., Ragauskas, A. & Deng, Y. Lignin Pyrolysis Components and Upgrading-Technology Review. *Bioenergy Res.* **6**, 1183–1204 (2013).
54. Collard, F. X. & Blin, J. A review on pyrolysis of biomass constituents: Mechanisms and composition of the products obtained from the conversion of cellulose, hemicelluloses and lignin. *Renew. Sustain. Energy Rev.* **38**, 594–608 (2014).
55. Zhang, T. *et al.* Initial Mechanisms for an Overall Behavior of Lignin Pyrolysis through Large-Scale ReaxFF Molecular Dynamics Simulations. *Energy & Fuels* **30**, 3140–3150 (2016).
56. Rousset, P., Turner, I., Donnot, A. & Perré, P. Choix d'un modèle de pyrolyse ménagée du bois à l'échelle de la microparticule en vue de la modélisation macroscopique. *INRA EDP Sci.* **63**, 213–229 (2006).
57. Chambon, F. Transformation de la cellulose par catalyse hétérogène. (2013).

Conclusion

In this study, we are interested in the thermal decomposition of organic matter leading to the formation of hydrocarbons. Oil and gas indeed result from sedimentation and thermal degradation of organic matters of various origins. This degradation leads to the formation of a two phase system: a fluidic one, in which hydrocarbons are found, and a carbonaceous solid one, called kerogen. This kerogen can be of three main types, depending on its organic origin. Type I kerogen mainly results from lacustrine deposition, type II from marine environment, and type III from terrestrial origins, such as terrestrial plants.

Three of the most important components of organic matter are carbon, hydrogen, and oxygen, and proportions of these elements in a kerogen are used to assess its evolution on a chart called the van Krevelen diagram. In this diagram, displays the H/C atomic ratio as a function of the O/C atomic ratio. During its degradation, kerogen loses hydrogen and oxygen. Therefore, both H/C and O/C ratio decrease. This evolution takes place over millions of years and is thus highly difficult to reproduce in laboratory.

Geological thermal degradation can be considered as a pyrolysis process, namely the chemical decomposition of organic matter under the only effect of heat, without any oxidant effect. Attempts to experimentally simulate this degradation have been conducted, and three main routes have been explored: open, closed, and confined pyrolysis. The last one is the most efficient, and the one that better reproduces the natural kerogen evolution.

Numerical simulation can be used to provide access to fundamental mechanisms. It can be used to try different theories and to compare them to experiment, or as a complement to simulate processes that are difficult to achieve in the laboratory. In the present study, difficulty focuses on the time scale. The challenge relies in the simulation of a slow process: thermal decomposition of organic matter that takes geological times.

We have worked at the atomic level, using a method called replica exchange molecular dynamics (REMD). At this scale, interactions between atoms of the system need to be taken into account, and for that, we used the well-known ReaxFF. ReaxFF is a reactive potential, in the sense it reproduces bond formations and breakings -by describing reaction intermediates-, based on a bond-order formalism. REMD method is used to accelerate time. It overcomes large free energy barriers, by enhancing configurational sampling of a conventional constant temperature molecular dynamics (MD), using states of higher temperatures. A set of classic MD replicas is run at different temperatures, starting from the same initial configurations. The lowest temperature is the one of interest -423K in our case, corresponding to the temperature in the ground between 4 and 5km depth- and the highest is chosen such that no replica is trapped in a

local energy minimum and every energy barrier can be crossed at this temperature. Temperature swaps between neighbor replicas are then attempted in order to help lower temperatures to overcome energy barriers. We therefore applied this method to kerogen precursors, in order to simulate the geological thermal degradation of organic matter.

As the major component of terrestrial plants, cellulose has been chosen to be our first type III kerogen precursor. We built an initial configuration of cellulose molecular crystal and run REMD simulations. At a first step, the whole cellulose system has been submitted to this thermal degradation, with some NPT relaxations, in order to keep the system in a geological pressure range. It converged towards a two phase system. The first phase was fluid and mainly composed by water and carbon monoxide, hydrogen and oxygen rich molecules. The second one was solid and carbon rich phase. The fluid part is the first to form, along with the cellulose fragmentation. Once the number of small molecules, such as CO and H₂O, starts to stabilize, the solid part starts growing, with the coalescence of small carbonaceous molecules. Once the whole system has converged, the solid part only, was relaxed and then again submitted to an alternation of REMD runs and NPT relaxations. A structural rearrangement of the solid phase took place to a state in which it can be considered as a mature kerogen, at the end of catagenesis stage, early metagenesis.

To complete this study, lignin, was chosen as another type III kerogen precursor. Lignin is indeed one of the three major components of higher plants, with cellulose and hemicellulose. Therefore, a numerical model of lignin was built then submitted to an alternation of REMD runs and NPT relaxations, in order to keep the system in a geological pressure range. Thermal degradation of the lignin system lead to the formation of two phases, as for cellulose. Its degradation is however different than that for cellulose. Indeed, initial aromaticity of the lignin makes it more resistant to thermal decomposition. Nevertheless, same elements are found, sometimes in different proportions. Fluid consists mainly in water and carbon monoxide, whereas the solid part grows rich in carbon. By comparing the first REMD stages for cellulose and lignin, we notice that lignin degradation forms more methane whereas the solid obtained from cellulose is more mature than the one from lignin, that is in an immature state. Differences found between cellulose and lignin pyrolysis, such as a greater formation of methane from lignin, quantitatively agree with experimental data.

The evolution of solid parts for both systems and the natural evolution of a type III kerogen clearly match. Evolution of our systems follows the geological path of natural samples, as well as the one of a type III kerogen submitted to a confined pyrolysis. Our systems decompositions therefore follow a realistic evolution as shown on the van Krevelen diagram. The method applied in this study –Replica exchange molecular dynamics- is therefore an original and validated numerical method, able to reproduce geological thermal degradation, occurring in Nature on a millions of years time-scale.

This technological breakthrough will help to understand fossil resources formation, to model oil and gas basins, and is an important step in the study of kinetically slow processes.

Although this method well reproduces the thermal degradation of organic matter under geological conditions, it still can be improved. A fully representative structure of kerogen is not accessible because of the small size of our system. Consequently, features such as the porosity cannot be statistically described. In order to increase the system size, huge numerical resources are needed. Nevertheless, two other improvement paths can be envisaged. The first one is related to the methodology itself. In Nature, a part of the fluid is released from the kerogen during the whole process and therefore its impact on kerogen formation and maturation might be lower than in our simulations. All or part of the fluid in the simulation could then be released during the REMD run. Moreover, although the ReaxFF potential does well in reproducing organic matter chemistry, we highlighted some possible improvements, among which, a better account reproduction of the water gas shift reaction. A new version of ReaxFF is currently parametrized in order to answer this issue, and the first tests are really promising.

A second interesting improvement path would be to widen the precursors data base, in the sense that cellulose and lignin are not found alone. They are part of a larger system. Combining those elements together, with other components such as hemicellulose, in order to have a representative description of higher plants system, might even better reproduce the natural thermal degradation. For instance, lignin is known to be antioxidant, because of its aromaticity, and this characteristic could deactivate some unstable radicals formed during hemicellulose degradation. Besides, interactions between wood components are poorly understood, in particular between lignin and other elements, and could be elucidated with this simulation technique. Other aspects can impact the degradation: inorganics. They could promote aromatic rings rearrangement and therefore induce an increase in char and dihydrogen production yields. Adding inorganics to organic matter could help to understand their impact and could improve the simulation.

Therefore, the replica exchange molecular dynamics method can be improved by many ways, but is already an accurate technique to numerically simulate geological processes, naturally evolving on millions of years time-scale.