Journal of Molecular Graphics and Modelling 53 (2014) 13–22

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Reaction analysis and visualization of ReaxFF molecular dynamics simulations

Jian Liu[a,b], Xiaoxia Li[a,*], Li Guo[a,**], Mo Zheng[a,b], Junyi Han[a,b], Xiaolong Yuan[a], Fengguang Nie[a], Xiaolong Liu[a,b]

[a] State Key Laboratory of Multiphase Complex Systems, Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100190, PR China
[b] University of Chinese Academy of Sciences, Beijing 100049, PR China

ARTICLE INFO

Article history:
Accepted 3 July 2014
Available online 11 July 2014

Keywords:
ReaxFF molecular dynamics
Large scale simulation
VARxMD
Chemical reaction analysis and visualization
Chemical reaction mechanisms
Cheminformatics
Chemical reaction parsing
Reaction pathway analysis
Chemical structure processing

ABSTRACT

ReaxFF MD (Reactive Force Field Molecular Dynamics) is a promising method for investigating complex chemical reactions in relatively larger scale molecular systems. The existing analysis tools for ReaxFF MD lack the capability of capturing chemical reactions directly by analyzing the simulation trajectory, which is critical in exploring reaction mechanisms. This paper presents the algorithms, implementation strategies, features, and applications of VARxMD, a tool for Visualization and Analysis of Reactive Molecular Dynamics. VARxMD is dedicated to detailed chemical reaction analysis and visualization from the trajectories obtained in ReaxFF MD simulations. The interrelationships among the atoms, bonds, fragments, species and reactions are analyzed directly from the three-dimensional (3D) coordinates and bond orders of the atoms in a trajectory, which are accomplished by determination of atomic connectivity for recognizing connected molecular fragments, perception of bond types in the connected fragments for molecules or radicals, indexing of all these molecules or radicals (chemical species) based on their 3D coordinates and recognition of bond breaking or forming in the chemical species for reactions. Consequently, detailed chemical reactions taking place between two sampled frames can be generated automatically. VARxMD is the first tool specialized for reaction analysis and visualization in ReaxFF MD simulations. Applications of VARxMD in ReaxFF MD simulations of coal and HDPE (high-density polyethylene) pyrolysis show that VARxMD provides the capabilities in exploring the reaction mechanism in large systems with complex chemical reactions involved that are difficult to access manually.

1. Introduction

The combination of classical molecular dynamics (MD) and the Reactive Force Field (ReaxFF) proposed by van Duin et al. [1], known as reactive molecular dynamics simulation, is a useful approach for molecular simulation of a complex system with chemical reactions [2–5]. ReaxFF MD uses energy terms based on a general bond-order potential that fully addresses the chemistry of dynamic bonds. In addition to non-bonded interactions, ReaxFF MD employs the dynamic charge equilibration using Mortier's electronegativity equalization method (EEM) [6] at each time-step to account for polarization effects. Thus, ReaxFF MD allows the description of the formation, transition, and dissociation of chemical bonds in a molecular system with accuracy close to that of DFT [7] and with reduced computational costs [8]. Particularly, the potential functions in ReaxFF can automatically handle coordination changes associated with reactions; thus, pre-definitions of reactive sites or reaction pathways are unnecessary. Moreover, the recent progress of high performance computing programs for ReaxFF MD has enabled simulations of large molecular systems on supercomputers [9,10], clusters [11] and desktop workstations [12]. The ReaxFF parallelization scheme in the more recent ADF platform has been rewritten to remove the global limit on the number of atoms in simulation systems [13]. The first Graphic Processing Unit (GPU) enabled ReaxFF molecular dynamics program is available and offers significantly increased computational capability on a single PC with a C2050 GPU attached, which makes performing ReaxFF MD simulations for larger system sizes and longer time scales possible on

* Corresponding author. Present address: State Key Laboratory of Multiphase Complex Systems, Institute of Process Engineering, Chinese Academy of Sciences, No. 1, Zhongguancun North Second Street, Beijing 100190, PR China.
Tel.: +86 10 82544944; fax: +86 10 82544945.
** Corresponding author. Present address: State Key Laboratory of Multiphase Complex Systems, Institute of Process Engineering, Chinese Academy of Sciences, No. 1, Zhongguancun North Second Street, Beijing 100190, PR China.
Tel.: +86 10 82544945; fax: +86 10 82544945.
E-mail addresses: xxia@ipe.ac.cn (X. Li), lguo@ipe.ac.cn (L. Guo).

http://dx.doi.org/10.1016/j.jmgm.2014.07.002
1093-3263/

a

Timestep	C22H22O	C108H79O4N2S	CO2	C161H145O8NS	…
1	2	0	0	0	…
50000	1	4	1	2	…
100000	2	6	6	1	…
150000	1	2	9	1	…
200000	2	4	11	2	…
250000	2	3	17	1	…
300000	2	4	17	1	…
350000	2	2	19	1	…
400000	2	1	18	1	…
450000	1	2	16	1	…
500000	1	2	17	0	…
550000	2	1	13	1	…
600000	1	2	16	0	…
650000	1	2	17	0	…
700000	2	2	18	0	…
750000	1	1	21	0	…
800000	1	1	20	0	…
850000	2	2	17	1	…
900000	1	1	17	0	…
950000	1	1	19	0	…
1000000	1	0	16	0	…

b

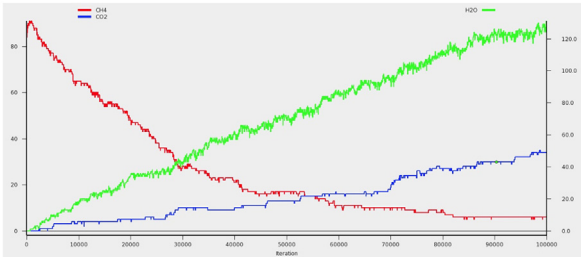


Fig. 1. List of formula-based molecules evolving with time-steps can be obtained by the LAMMPS reax tool and ADF for a ReaxFF MD simulation: (a) a sample text list generated by the LAMMPS reax tool for a ReaxFF MD simulation of a coal model; and (b) a sample graphic representation given by ADF [13].

desktop workstations [12]. The progress that has been made on ReaxFF MD provides a promising approach for studying reaction mechanisms in processes with very complex chemical reactions at an atomic scale, where the reactions sites are difficult to predefine manually.

However, challenges arise in the analysis of ReaxFF MD simulation results for large molecular systems with complex chemical reactions. In addition to three-dimensional (3D) atomic coordinates, the trajectory from a ReaxFF MD simulation is in the form of bond orders without clear bond types between bonded atoms, in contrast to the trajectory obtained from a classical MD simulation. The existing analysis tools for ReaxFF MD simulations can only provide chemical formula-based analysis results that provide a time-step evolution of the system for the number of molecules based on these formulas. As shown in Fig. 1, the reax tool in LAMMPS can only produce a text list file, while the ADF platform can provide similar information in the form of graphs. Because one chemical formula may represent many chemical species in terms of 3D structure of a chemically connected molecular fragment in simulated molecular systems, the existing analysis tools are not capable of distinguishing between the species in the reaction products accurately. More importantly, the information on the chemical reactions in the simulated systems is not available in the analysis results, which is critical in understanding the reaction mechanisms via ReaxFF MD simulation. As a result, the analysis of the trajectory has to be performed manually. Manual analysis is time consuming, yet affordable, for relatively small and simple systems simulated with ReaxFF MD, but this approach is impractical for large systems that involve complex chemical reactions, especially for the investigation of coal pyrolysis. A pyrolysis simulation of a bituminous coal model with 4976 atoms using the ReaxFF MD program revealed that more than 900 reactions might occur at 2000 K within a 250 ps simulation period when the trajectory output interval is 12.5 ps [14]. Even larger numbers of reactions can be expected if the trajectory output interval becomes smaller or for the simulation of larger molecular systems, such as in a pyrolysis simulation of the Liulin coal model with 28,351 atoms [15]. Therefore, VARxMD (Visualization and Analysis of Reactive Molecular Dynamics) was created to provide a new tool specifically designed to facilitate the analysis and visualization of detailed chemical reactions for the trajectory of ReaxFF MD simulation. This paper will present the methods, algorithms and applications of VARxMD.

2. Methods

VARxMD consists of modules that employ a set of algorithms specifically designed for depth analysis and visualization of the trajectories obtained from ReaxFF MD simulations to deliver detailed chemical reactions and other information necessary for revealing the reaction mechanisms of the simulated processes.

Fig. 2 shows the major modules and the data processing flowchart of VARxMD. VARxMD focuses on revealing the reaction information contained in the trajectory of ReaxFF MD simulations by analyzing the internal relationships of atoms and their evolutions in the trajectory. The entity relationship model for the data processing employed in VARxMD is displayed in Fig. 3 to show the relationships of hierarchical atom entities and their connections, on which the chemical reaction networks are generated. The entities consist of atoms, bonds, molecular fragments, unique species, chemical reactions, unique chemical equations, and reaction networks, all of which are derived from the implied information in the simulated trajectory.

The modules in Fig. 2 are designed for data processing during the generation of chemical reactions from the trajectory obtained in ReaxFF MD simulations. The main inputs into VARxMD are the trajectory files with 3D atomic coordinates and bond orders. According to the bond order cutoff values of the bonded atom pairs defined in the configure file, the connectivity of the atoms is determined first. Then, the atoms and bonded atom pairs will be analyzed as connected molecular fragments according to the connectivity of the atoms. The trajectory will be converted from periodic boundary coordinates into normal coordinates when the periodic condition is applied. Next, the bond types between bonded atom pairs in the connected molecular fragments can be recognized based on extra bond order cutoffs or 3D atomic coordinates and connectivity. The 3D structure of each connected molecular fragment is analyzed to determine the species. A species is the unique representation of a fragment that can be indexed by a molecular structure comparison based on InChI [16,17] or canonical SMILES [18]. Examples of the InChI or SMILES codes for some species in VARxMD can be found in Fig. 4. Finally, analyzing the changes of the bonding sites and the consequent changes in the species between two time-steps can generate detailed chemical reactions. The obtained chemical reactions can be represented in 2D chemical structure diagrams and viewed in 3D graphics in VARxMD. Statistics on the number of reactions, specified molecules, etc. can also be performed with VARxMD to facilitate reaction mechanism analysis.

The algorithms of the major modules in Fig. 2 and their implementation strategies are described in the following sections.

2.1. Identification of connected molecular fragments

From the perspective of graph theory [19], a molecule can usually be regarded as a labeled undirected graph, whose vertices correspond to the atoms of the molecule and edges correspond to the chemical bonds. A graph is called connected if, given any two vertices in the undirected graph, there exists a path between these two vertices. Accordingly, connected molecular fragments

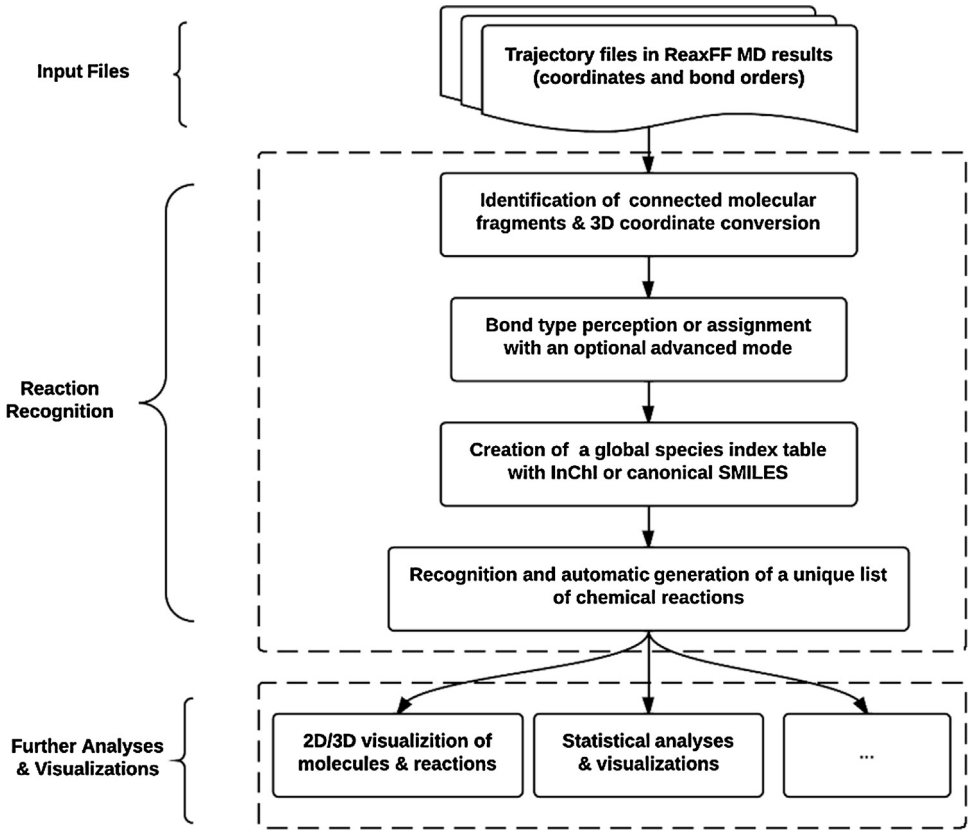


Fig. 2. Overview of the data flowchart and corresponding features in VARxMD for detailed chemical reaction analysis directly from the ReaxFF MD simulation trajectories.

are molecules or fragments with radicals, for which a path exists between any two atoms. A trajectory obtained from a ReaxFF MD simulation contains 3D coordinates of atoms in the simulated molecular system at different time-steps, which represents the evolution of the system with time in terms of number of time-steps, the minimum number is 1. Because of the chemical reactions that occur, the simulated system is a multi-molecule system in which the molecular fragments vary with time. Identifying the connected molecular fragments in such a multi-molecular system is the first task for further recognition of what reactions take place to investigate the chemical reaction mechanism. A molecular system of multi-molecules is an undirected graph, but these molecules are not necessarily connected. Identifying all connected molecular fragments in such a multi-molecular system can be considered as a graph traversal problem in graph theory. Depth-First-Search [20], a general graph traversal algorithm, is used for the connected molecular fragment identification in VARxMD. The connectivity between two bonded atoms is assigned to a molecular fragment if the bond order of the atom pair is greater than the related bond order cutoff value in the default mode. Whenever a new connected molecular fragment is found, this fragment will be identified for indexing later. Traversing atom by atom, each connected molecular fragment will find its constituent atoms and bonds. Consequently, the index from molecular fragment to atoms and bonds, and the reversed index as well can be built. A connected molecular fragment is a fundamental unit in VARxMD for later processing. The molecule or fragment refers to the connected molecular fragment in this paper unless noted otherwise.

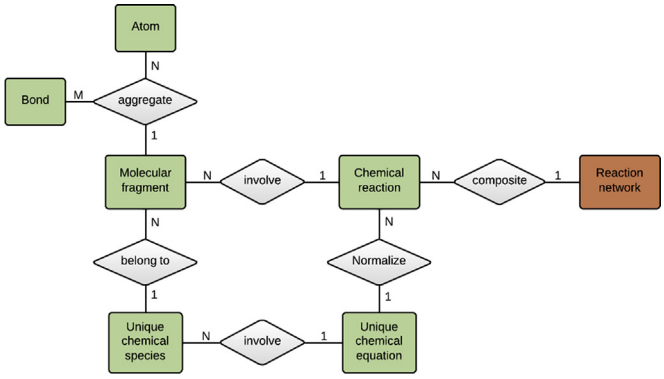
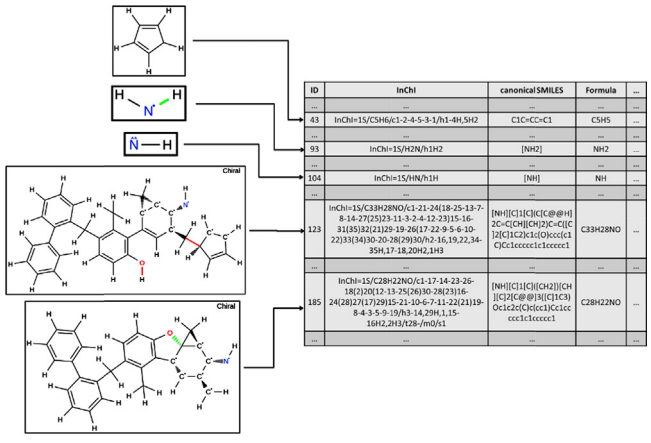


Fig. 3. Entity relationship diagram of bond, atom, molecular fragment, unique chemical species, chemical reaction, unique chemical reaction and reaction network in VARxMD.



ID	InChI	canonical SMILES	Formula	...
...	...	...	...	...
43	InChI=1S/C5H6/c1-2-4-5-3-1/h1-4H,5H2	C1C=CC=C1	C5H5	...
...	...	...	...	...
93	InChI=1S/H2N/h1H2	[NH2]	NH2	...
...	...	...	...	...
104	InChI=1S/HN/h1H	[NH]	NH	...
...	...	...	...	...
123	InChI=1S/C33H28NO/c1-21-24(18-25-13-7-8-14-27(25)23-11-3-2-4-12-23)15-16-31(35)32(21)29-19-26(17-22-9-5-6-10-22)33(34)30-20-28(29)30/h2-16,19,22,34-35H,17-18,20H2,1H3	[NH][C]1[C][C[C@@H]2C=C[CH][CH]2)C=C[[C]2][C]1C2]c1c(O)ccc(c1C)C1ccccc1c1ccccc1	C33H28NO	...
...	...	...	...	...
185	InChI=1S/C28H22NO/c1-17-14-23-26-18(2)20(12-13-25(26)30-28(23)16-24(28)27(17)29)15-21-10-6-7-11-22(21)19-8-4-3-5-9-19/h3-14,29H,1,15-16H2,2H3/t28-/m0/s1	[NH][C]1[C][C[CH2]][CH][C]2[C@@]3][C]1C3)Oc1c2c(C)c(cc1)Cc1ccccc1c1ccccc1	C28H22NO	...
...	...	...	...	...

Fig. 4. Example of indexed chemical species using their InChI or canonical SMILES in the global species index table of VARxMD, taken from the reaction highlighted in Fig. 8.

Algorithm: Unwrap atom coordinates around the periodic boundaries in a fragment to remove overlong bonds for a ReaxFF MD trajectory with periodic boundary conditions (PBC) applied

Input:

A trajectory (3D coordinates with periodic boundary)

Output:

A trajectory (3D coordinates with overlong bonds removed)

Steps:

1) Define the cutoff value of an overlong bond, $L_{cutoff} = \max(L_{bond}) + \varepsilon$;
2) For each overlong bond B_{ij} ($atom_i$, $atom_j$)
3) Consider the fragment as two parts (P_i and P_j) connected by B_{ij}. If B_{ij} is a ring edge in the fragment, P_i and P_j are the same, i.e., the fragment itself, while P_i and P_j are two separated parts in case that B_{ij} is not a ring edge.
4) To improve the performance of unwrapping, evaluate the coordinate translation transformation cost, $C_k = N_{bonds,\,k}$ (k=i, j), for P_i and P_j, where $N_{bonds,k}$ is the total number of bonds in P_i and P_j, and find the lower cost part, of which the coordinates will be transformed.
5) For the lower cost part, taking P_i as an example, get the unwrapped 3D coordinates $\vec{I}(X_i, Y_i, Z_i)$ by transforming each atom $\vec{I}_0(X_{i,0}, Y_{i,0}, Z_{i,0})$ in P_i to that of P_j, $\vec{J}(X_j, Y_j, Z_j)$, using

$$\vec{I} = \vec{I}_0 + \left\lfloor \frac{\vec{J} - \vec{I}_0}{L_{cutoff}} \right\rfloor \cdot |\vec{L}_{boxsize}|$$

Output the wrapped coordinates $\vec{I}(X_i, Y_i, Z_i)$.

Fig. 5. Pseudo code for unwrapping the 3D coordinates of molecular fragments around the boundary of the simulation cell in a ReaxFF MD simulation trajectory when the periodic boundary conditions (PBC) applied.

In addition, it should be pointed out that it is necessary to unwrap the overlong bonding caused by the application of periodic boundary conditions (PBC) for consequent bond perception and visualizations of fragments, reactions or the trajectory. The periodic boundary conditions are frequently used in ReaxFF MD simulations. In a ReaxFF MD trajectory with PBC applied, an unusual overlong bond between two bonded atoms may occur in a connected molecular fragment, because atoms are wrapped around the periodic boundaries in the course of a PBC simulation. The overlong bonds will lead to misidentification of the bond type between the bonded atom pair, which will consequently affect the correctness of the chemical reaction recognition and the 3D chemical structure visualization of the simulated system, including fragments, reactions and trajectory itself. Thus, before bond type perception, the 3D atomic coordinates of the simulation trajectory have to be unwrapped.

Fig. 5 shows the pseudo code of the PBC unwrap algorithm for the ReaxFF MD trajectory in VARxMD. First, the cutoff value of overlong bond length is defined as $L_{cutoff} = \max(L_{bond}) + \varepsilon$, where L_{bond} is the bond length of any bonded atom pair and ε is a redundant value. If the bond length of a bonded atom pair is greater than L_{cutoff}, this bond will be identified as an overlong bond. Next, for each overlong bond $B_{ij}(\text{atom}_i, \text{atom}_j)$ in a connected molecular fragment, the fragment can be considered as two separate parts P_i and P_j connected by B_{ij}. If B_{ij} is a ring edge in the fragment, P_i and P_j are the same. The unwrap can be accomplished by "moving" one part to another, i.e., by coordinate translation transformation of P_i to P_j, or vice versa. To improve the coordinate transformation efficiency, the transformation computing costs for each part, $C_k(k = i, j)$, are evaluated first by the total number of bonds in P_i or P_j, and then the part with lower computing cost is subjected to coordinate translation transformation.

2.2. Perception of bond types

As mentioned above, connected molecular fragments are obtained based only on the connectivity of atom pairs, and their coordinates may have been transformed from PBC coordinates. However, bond types (single, double, triple, or aromatic) between bonded atom pairs are not clear. Assigning bond types for bonded atom pairs in each connected molecular fragment is a necessary and key step for characterizing connections of atom pairs in a chemical species. Many methods for automatic bond type perception from 3D coordinates and element identities have been developed [21–28]. In VARxMD, bond type assignment is currently implemented using the bond order perception function in Open Babel library, an open-source chemical toolbox [29,30]. The algorithms for bond order perception in Open Babel is similar to the techniques proposed by Roger Sayle [23], which is based on bond angles and geometrics to determine the hybridized centers, check for planarity of 5- and 6-membered rings to estimate aromaticity, mark unsaturated atom with a double or triple bond, and consequently match functional groups and aromatic rings. More details on the algorithms for bond order perception in Open Babel can be found in the paper of O'Boyle et al. [29]. The bond type perception of Open Babel was developed for molecular file format conversion applications such as for most of the files in crystallographic information file format (CIF) and for files in non-standard Protein Data Bank format (PDB), where the molecular data are solely available as lists of atoms, without connectivity or bonding information. In VARxMD, the connectivity of atom pairs is determined by the bond order of each atom pair available in the trajectory. Therefore, the recognition step for the connectivity of atom pairs in Open Babel was bypassed.

In addition, an advanced mode for specifying the bond types of atom pairs is provided in VARxMD as an attempt to take advantage of bond order in ReaxFF, which allows users to define and edit a mapping table of bond type and the range of bond order values for atom pairs. This feature can be useful when the relationships of bond type and bond order values are available for all the atoms in a ReaxFF MD simulated system, the bond type identification step using Open Babel in VARxMD will not be used in this case.

2.3. Creation of a global species index

After the recognition of bond types in the default mode or the bond order cutoffs defined in the advanced mode in VARxMD, the 3D structure of each connected molecular fragment is available for further analysis of the chemical reactions between the fragment and other fragments in a ReaxFF MD system. A species is defined as a unique equivalent of the 3D structure of a fragment, so the species is capable of distinguishing the isomers of fragments with the same chemical formula. Each connected molecular fragment should be canonically indexed to its species, as shown in Fig. 4. A proper indexing strategy is very important for fast access to the index and the indexed molecular structure of the fragments, which is critical for lowering the computational cost in the consequent identification and further analysis of the chemical reactions in the simulated trajectory.

In general, a linear notation is an efficient way to represent a molecule as a single-line string of characters. Among the line notations, both SMILES [18] and InChI [16,17] are widely used in various Web-based chemical databases. InChI is a new international standard designed for normalization and canonicalization of structure representation, where a chemical compound can be uniquely identified to facilitate the handling of this compound in computer databases. This standard provides the capability for easier linking of diverse data compilations and unambiguous identification of chemical substances [16,31]. The application of InChI in very large databases, such as PubChem [32] and ChemSpider [33], in which over 100,000,000 compounds are indexed, proves the ability of this linear notation to represent a compound in a completely unequivocal manner. Thus, InChI is adopted in VARxMD as a unique representation of the molecular species.

The canonical SMILES are also used to index the molecular species in VARxMD, which provides not only an alternative to InChI indexing, but also a popular entry for the faster searching

of reactants and products in reactions. As shown in Fig. 4, for each molecular fragment in the simulated system, InChI and canonical SMILES are generated from the 3D structure. Then, a global species index table can be created efficiently, facilitating fast access to species by comparing the InChI or canonical SMILES of the fragment to be indexed directly with the linear notation of other indexed fragments in the global species index table.

The InChI-based or canonical SMILES-based strategy used for the identification of a new fragment in the global index is a way to avoid relatively low-efficiency 3D structure matching of molecular fragments when accessing the global species index, which is broadly used in reaction analysis and statistical analysis in VARxMD. For compatibility with the analysis result of LAMMPS, a molecular formula index is also created by lumping the fragments in the global species index.

2.4. Recognition and generation of chemical reactions

A chemical reaction is a process that leads to the transformation of one or one set of chemical substances to another. The input of chemical reaction analysis in VARxMD is the trajectory file obtained from a ReaxFF MD simulation, consisting of 3D atomic coordinates and bond orders of atoms at different moments or time-steps, sampled as frames at single or multiple time-step intervals. Recognition of chemical reactions in VARxMD is carried out based on the changes in the bonding of an atom pair between two frames in a trajectory, which complies with the classical chemical reaction definition that only involves the position changes of electrons in the forming and breaking of chemical bonds between atoms but not the nuclei. The stereo-transformations within a species are not considered for the moment in order to have a clear representation of the chemical reactions between the two frames.

VARxMD recognizes complex chemical reactions by taking advantage of the feature in the ReaxFF MD trajectory that only the connectivity of atoms changes with time-step, whereas the ID and type of an atom are static. Fig. 6 shows the algorithm scheme for generating a unique list of chemical reactions during the period of two sampled frames in a trajectory, where each rounded rectangle represents data sets and each rectangle represents the actions applied to the relevant data sets. First, VARxMD generates S_1, a set of formed bonds and broken bonds according to the change of the connectivity of atoms between two frames at two different moments, t_1 and t_2, in a trajectory. A bond is formed if it exists at t_2 but not at t_1. A bond is broken if it exists at t_2 but not at t_1. Then, VARxMD generates S_2, the elementary reactions or basic reactions for the same fragments with bond forming or bond breaking. The time-step interval is defined as $k = \Delta t/\textit{time-step} = (t_2 - t_1)/\textit{time-step}$, where $k \geq 1$. If $k = 1$, the reaction is most likely an elementary reaction. For each elementary reaction, a maximum of three molecular fragments can be involved. If $k > 1$, the reaction may be an elementary reaction or may involve more than one elementary reaction, defined as a basic reaction. Finally, VARxMD generates chemical equations by scaling the number of involved reactant or product fragments by comparing the species within the fragments and merging the same reactions together. Thus, a unique list of chemical reactions between two frames in a trajectory can be generated. This chemical reaction recognition algorithm can be used to recognize chemical reactions between any two moments (two frames) in a trajectory.

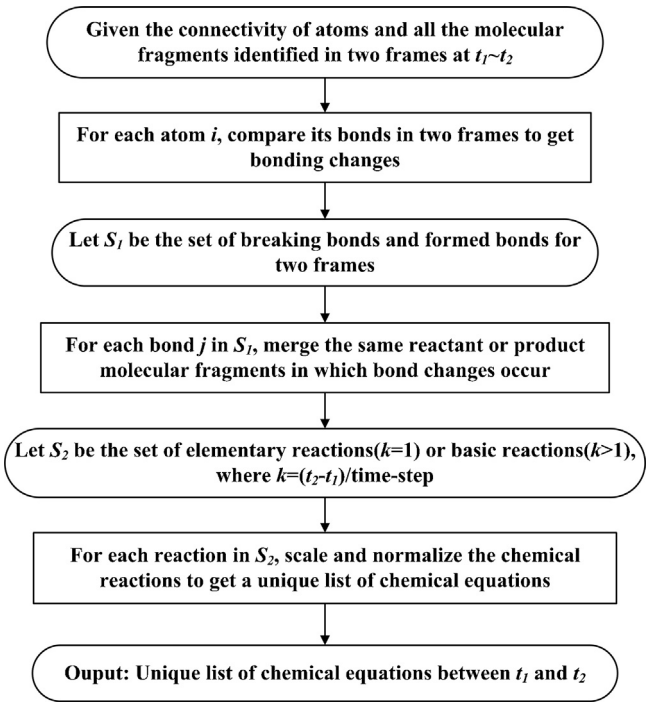


Fig. 6. Algorithm scheme in VARxMD for the generation of a unique list of chemical reactions between two frames for a ReaxFF MD trajectory, where each rounded rectangle represents data sets and each rectangle represents the actions applied to the relevant data sets.

3. Results and discussion

3.1. Automated reaction generation and visualization

With the implementation of algorithms employed in the modules of VARxMD (Fig. 2), for the first time, chemical reactions can be generated automatically and directly from the trajectory obtained in the ReaxFF MD simulations based on the cheminformatics analysis. Given any two frames in a trajectory, a unique list of chemical reactions can be generated. For a long simulation, given an output interval in terms of the number of time-steps, the list of chemical reactions with time (at any given output interval) can be obtained, which will significantly facilitate the chemical reaction mechanism investigation using ReaxFF molecular dynamics simulations, particularly for complex molecular systems.

As an example, the chemical reactions in pyrolysis simulations of a coal model with 13,498 atoms were generated using VARxMD. The detailed steps and results were recorded in a video file with written descriptions that are provided as supplementary of this paper. Fig. 7 shows the 3D view of the entire system for this model. Fig. 8 is a snapshot of VARxMD that shows the list of chemical reactions obtained by VARxMD with an output interval output of 50,000 time-steps. For each row in the list, the chemical equation, chemical equation ID, the time-step ID, and the number of the reactions are displayed. The unique list of reactions is automatically generated based on the species index. As an example, the species index of the reactants and products for the highlighted reaction can be found in Fig. 4.

Most importantly, VARxMD features the automated generation of chemical reactions identified from two frames in the trajectory of a ReaxFF MD simulation, which is not available in other analysis tools such as the reax tool in LAMMPS and ADF for ReaxFF MD. To allow chemical reactions analyzed conveniently and explicitly, reactants and products of a chemical reaction are rendered together with the 2D/3D views. Formed bonds and bonds to be broken are rendered and highlighted with different colors such that the reaction sites of the chemical reaction can be obtained easily both from the 3D and the 2D view.

Fig. 9 shows the highlighted chemical reaction of Fig. 8 in 2D chemical structure diagrams. The bonds in red are those to be broken in the reactants, and green for those to be formed in the

Table 1
Reactions observed directly by VARxMD at 1800 K during 162.5–175 ps in ReaxFF MD pyrolysis simulation of the coal model with 13,498 atoms in Fig. 7 (bonds to be broken in reactants are labeled in red, and formed bonds in products are labeled in green).

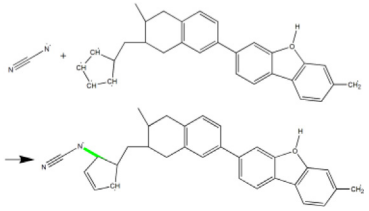
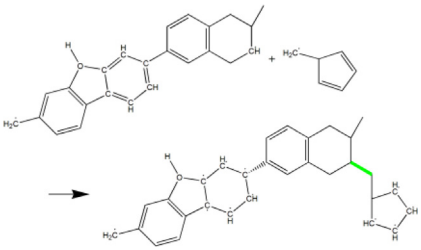
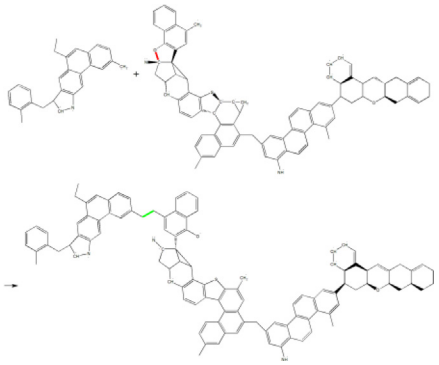
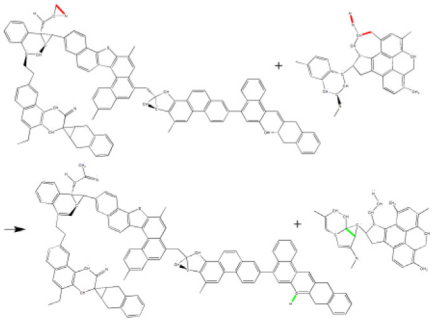
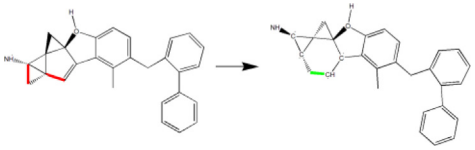
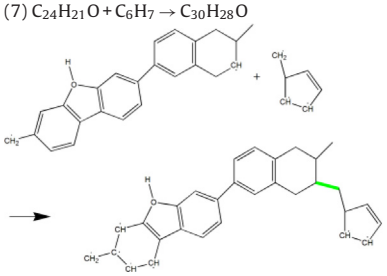
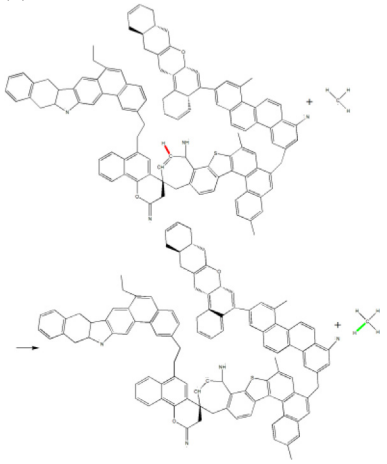
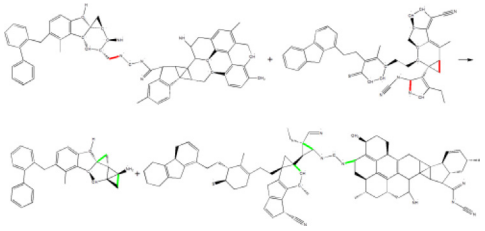
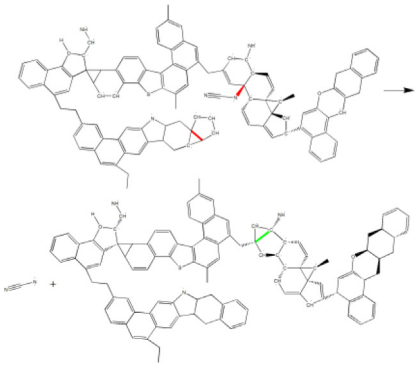
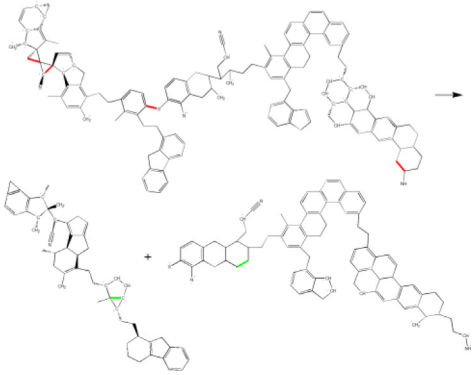
Chemical reactions
(1) $C_{30}H_{28}O + CN_2 \rightarrow C_{31}H_{28}ON_2$ 
(2) $C_{24}H_{21}O + C_6H_7 \rightarrow C_{30}H_{28}O$ 
(3) $C_{27}H_{24}N + C_{81}H_{54}O_2N_2S \rightarrow C_{108}H_{78}O_2N_3S$ 
(4) $C_{108}H_{77}O_2N_2S + C_{35}H_{30}ON_2 \rightarrow C_{108}H_{78}O_2N_2S + C_{35}H_{29}ON_2$ 
(6) $C_{28}H_{23}ON \rightarrow C_{28}H_{23}ON$ 
(7) $C_{24}H_{21}O + C_6H_7 \rightarrow C_{30}H_{28}O$ 
(8) $C_{108}H_{78}O_2N_4S + CH_3 \rightarrow CH_4 + C_{108}H_{77}O_2N_4S$ 
(9) $C_{63}H_{50}O_2N_5 + C_{46}H_{39}N_4S \rightarrow C_{28}H_{23}ON + C_{81}H_{66}ON_8S$ 

Table 1 (*Continued*)

(5) $C_{109}H_{78}O_2N_5S \rightarrow C_{108}H_{78}O_2N_3S + CN_2$	(10) $C_{135}H_{118}N_4S \rightarrow C_{53}H_{43}N + C_{82}H_{75}N_3S$
	

products in the reaction, respectively. Fig. 10 is a screenshot of VARxMD for 3D visualization of the same reaction in Fig. 8, where the bonds to be broken and formed bonds are also be highlighted for the identification of reaction sites. Table 1 lists the reactions observed directly by VARxMD at 1800 K during 162.5–175 ps for the system in Fig. 7, to demonstrate the features of VARxMD for labeling the bonding sites in a reaction. The feature of VARxMD for highlighting the bonding sites explicitly in a reaction can be helpful to further improve the analysis efficiency for chemical reaction mechanisms exploration using ReaxFF MD simulations.

3.2. Applications in pyrolysis mechanism study of complex systems

VARxMD has been used in a number of reaction analyses of complex systems. With the capability of generating chemical reactions, it becomes practical to explore reaction mechanisms of complex systems using ReaxFF MD simulations. The role of VARxMD has been critical for revealing the reaction mechanisms of the complex hypothetical bituminous coal model with 4976 atoms [14] and the Liulin coal models [15] from the recent work of the authors' group. The Liulin coal model containing 28,351 atoms is the second largest coal model used in ReaxFF MD simulations. With the aid of VARxMD, the evolving trend of products and the reactions in coal pyrolysis were investigated using ReaxFF MD simulations. With the species identified and 2D chemical structure diagram generated, it was found that the major pyrolysates with temperature obtained in a 250 ps simulation are in reasonable agreement with the Py-GC/MS experiments. By analyzing the chemical reactions, the main radicals generated at the early stage confirm that pyrolysis of coal is a radical driven process, which agrees well with the fundamental steps of coal pyrolysis [34,35]. The generation and consumption of $H_3C^{\bullet}$ and $HO^{\bullet}$, and their connection with generation of CH_4 and H_2O in pyrolysis product were discussed. The possible pathways for conversion of 6-membered aromatics ring into 5- or 7-membered rings were observed [36]. Without VARxMD, the understanding of these insightful mechanisms can hardly be obtained with other available analysis tools for ReaxFF MD simulations or experimentally.

VARxMD has also been used in pyrolysis investigation of HDPE (high-density polyethylene) with ReaxFF MD [37]. Detailed reaction mechanisms and generation pathways of primary gas molecules were obtained by using VARxMD. Major reactions

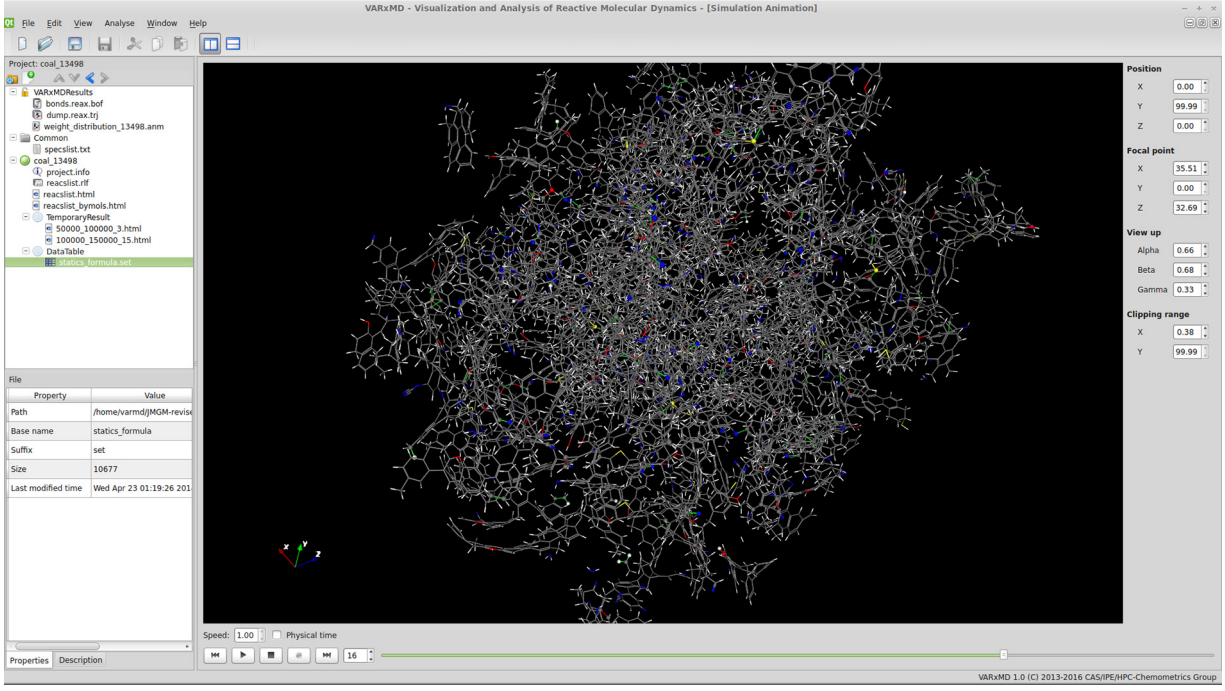


Fig. 7. VARxMD snapshot of three-dimensional visualization of the entire system of a coal molecular model simulated with ReaxFF MD.

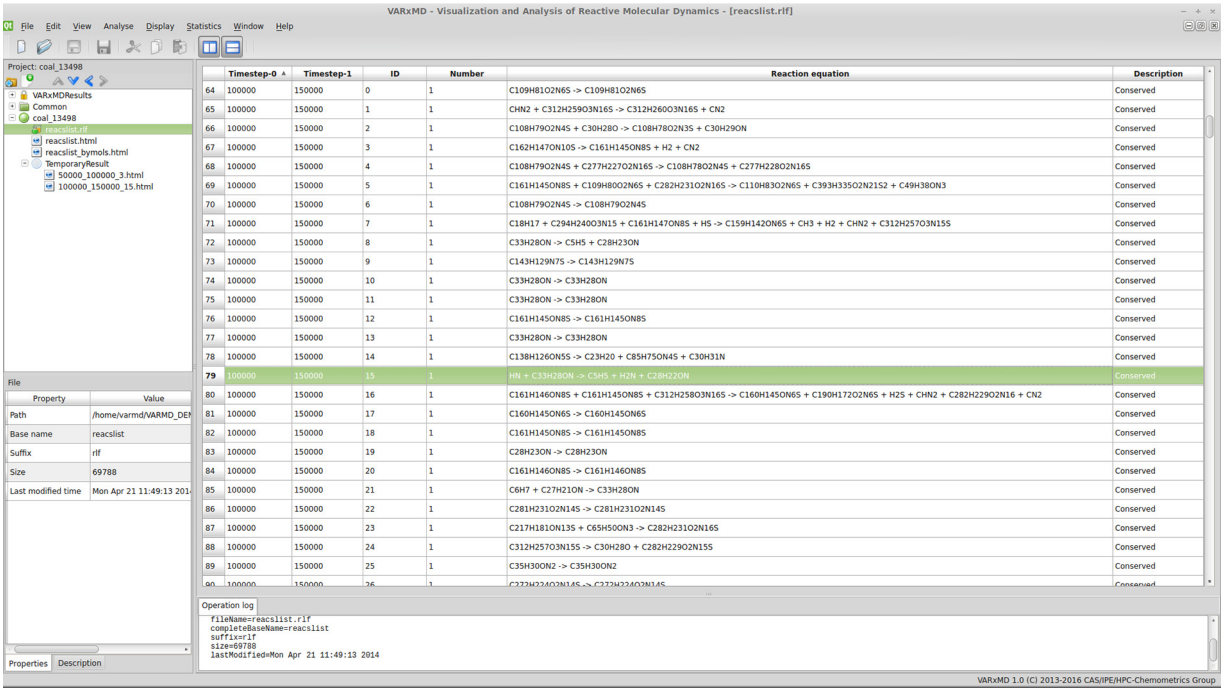


Fig. 8. VARxMD snapshot of the generated unique list of chemical reactions occur in the pyrolysis process simulated using ReaxFF MD for the coal model in Fig. 7.

including initiation and termination reactions, β-scissions, H-abstractions, intra-molecular H-shift, and generation paths of main gas products were observed. Other reactions such as dehydrogenation, H_2 decomposition, branching, cyclization, and detailed pathways of intra-molecular branching and (1, 2)-H-shift reactions were obtained. The reaction mechanisms based on the simulation trajectory analysis are in broad agreement with the literature [38].

The reasonable results obtained in applications of VARxMD above validates that VARxMD is very helpful for a systematic understanding of detailed chemical reactions in pyrolytic process for complex molecular systems. Moreover, VARxMD is a general tool for chemical reaction analysis of ReaxFF MD simulations and can be used in other applications.

3.3. Features of VARxMD from a perspective of software

Analysis of chemical reactions from the trajectory of ReaxFF MD simulation shares the common grounds in cheminformatics for chemical structure processing [39,40]. For example, the overlong bond problem occurs both in conventional MD and in ReaxFF MD simulations when the periodic boundary conditions are applied. The overlong bonding between atom pairs around the PBC cell boundary only affects the proper visualization of a trajectory for the conventional MD. But for ReaxFF MD simulations, the overlong bonds can also lead to wrong perception of bond type, which is critical for correct generation of chemical reactions in VARxMD. The unwrapping of the atom coordinates in a PBC applied trajectory is a necessary step for chemical reaction analysis in VARxMD.

With the algorithms for chemical reaction analysis implemented, VARxMD is promising to become a software tool for ReaxFF MD simulations in the future. However, further improvements of VARxMD should be considered, for example, the visualization performance, perception of dynamic bonds and building of standard benchmark datasets.

3D graphics in VARxMD was implemented by taking advantage of the execution pipeline in VTK [41], which is further improved via the VTK 5.10 release, in which the chemical structure visualization code developed by David Lonie was incorporated [42]. The function of 2D chemical structure diagrams for molecular fragments in VARxMD is currently available based on the open-source cheminformatics library of Indigo Toolkit [43], which is very helpful for faster implementation of 2D chemical structure diagram generation and validation of the functions in VARxMD particularly designed to facilitate reaction mechanism analysis in ReaxFF MD simulations. One challenge for further improving VARxMD is the computational performance of 2D structure diagram generation for molecular fragments using the Indigo Toolkit [43]. It is relatively slow, especially for large molecular fragments in coal models. This

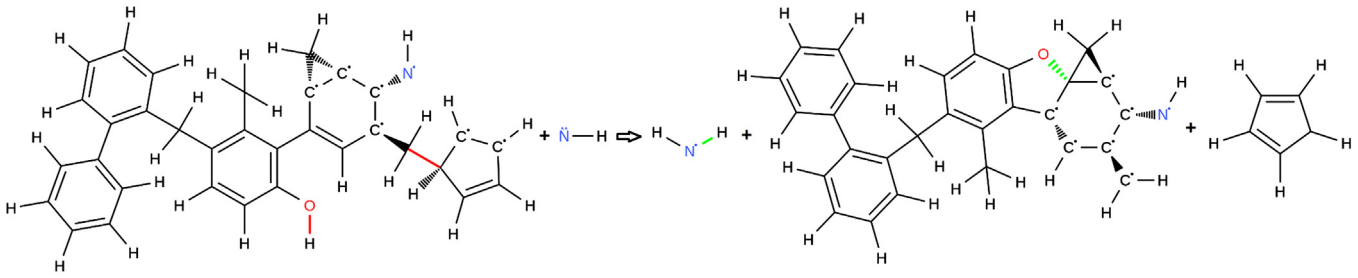


Fig. 9. VARxMD snapshot of the highlighted chemical reaction in Fig. 8 viewed in 2D chemical structure diagrams, where the bonding sites are labeled (the bonds to be broken are in red, and the formed bonds are in green). (For interpretation of the references to color in text, the reader is referred to the web version of this article.)

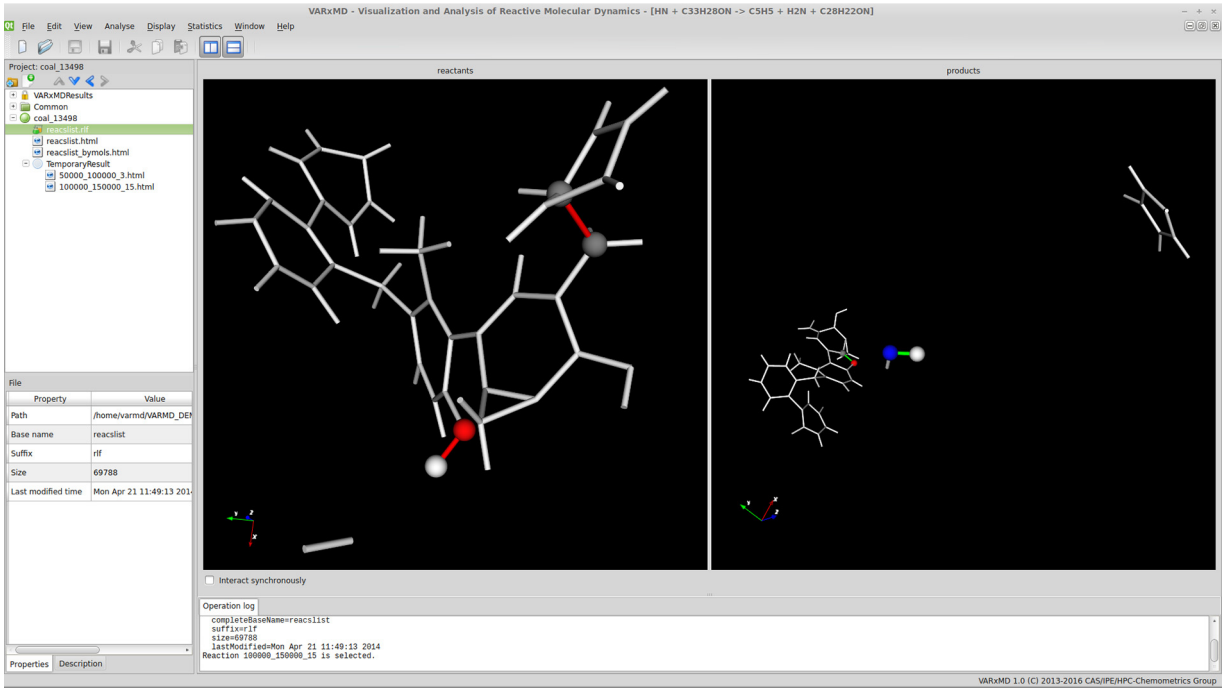


Fig. 10. Three-dimensional visualization in VARxMD with bonding sites labeled for the highlighted chemical reaction of Fig. 8 (bonds to be broken are labeled in red, and formed bonds are labeled in green).

can be avoided in current VARxMD analysis by limiting the atom scale of molecular fragments or rendering time. Probably the RDKit [44] or other libraries/toolkits for chemical structure processing can be used to replace the Indigo Toolkit in VARxMD for the generation of 2D molecular structure diagram, which needs further study. Multithreaded or GPU-based acceleration strategies may be also considered in parallel bond type perception for different fragments in future work.

The proper assessment of dynamic bonding in ReaxFF MD simulations is also challenging. Although ReaxFF MD simulations allow a smooth/continuous dependence of bond order on the distance for each of the single, double, and triple bond types, and the transition of the total bond order from a completely non-bonded interaction to a full triple-bonded state can proceed smoothly [8], some unreasonable fragments can be generated in the bond type perception during the reaction analysis. This is most likely because the evolving intermediates of fragment formed during the ReaxFF MD simulations may not be properly processed by the bond type perception algorithms employed in VARxMD. Further investigations are needed to solve this problem. Moreover, the refinement of bond order cutoffs is highly needed for the bond type perception, particularly in the advanced mode, which is challenging and more efforts are in the making. Benchmarking of the advanced mode is complicated undertaking.

Statistical functions are available for the analyzed species and can be grouped by functional groups, chemical formula, etc. By lumping species, the number of molecules for fragments with the same molecular formula generated in VARxMD can be compared with that generated via the reax tool in LAMMPS. Benchmarks for an n-dodecane system and several coal models show that the number of molecules generated by VARxMD is the same as that generated by LAMMPS, indicating that VARxMD was validated to some extent. The species generated in VARxMD is discriminated by 3D chemical structure of a fragment, and the number of species is obviously larger than that of molecules. However, this function in VARxMD cannot be compared with LAMMPS because it is not available in LAMMPS. Because of the lack of reaction analysis capability in LAMMPS, ADF and Materials Studio [45], benchmarks of the chemical reactions obtained with VARxMD is not readily available. The chemical reaction generation in VARxMD is guaranteed by the algorithms employed.

VARxMD is a new tool for reaction analysis and can be used in any molecular systems simulated using ReaxFF MD including biological molecular models. In addition, the strategy and algorithms employed in VARxMD make it applicable in other reactive force field based MD simulations. The input for reaction analysis in VARxMD is the evolution of bond connectivity of atom pairs in the simulation trajectory. The reactions between two frames of the trajectory are determined by the changes of the bond connectivity. Thus, for simulations that can provide the evolution of bond connectivity of atom pairs, the strategy of VARxMD should be applicable with minor coding modifications.

4. Conclusion

The algorithms implemented in VARxMD are dedicated to chemical reaction analysis and visualization in ReaxFF MD simulations. Detailed chemical reactions can be automatically generated directly from the 3D coordinates and bond orders of the atoms in a trajectory by analysis of the depth relationships among the atoms, bonds, fragments, species and reactions. VARxMD is the first tool allowing for uncovering the detailed chemical reactions in ReaxFF MD simulation trajectories and can be particularly useful for large systems with complex chemical reactions that are difficult to access manually. VARxMD has been used in ReaxFF MD simulations of coal and HDPE proving to be useful in revealing their pyrolysis reaction mechanisms.

Competing interest

The authors declare no competing financial interest.

Acknowledgments

This work was co-supported by the funding granted from the National Natural Science Foundation of China (21073195, 21373227, and 21103196) and the grant of China's State Key Laboratory of Multiphase Complex Systems (MPCS-2012-A-05).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm.2014.07.002.

References

[1] A.C.T. van Duin, S. Dasgupta, F. Lorant, W.A. Goddard, ReaxFF: a reactive force field for hydrocarbons, J. Phys. Chem. A 105 (2001) 9396–9409.
[2] K. Chenoweth, S. Cheung, A.C.T. van Duin, W.A. Goddard, E.M. Kober, Simulations on the thermal decomposition of a poly(dimethylsiloxane) polymer using the ReaxFF reactive force field, J. Am. Chem. Soc. 127 (2005) 7192–7202.
[3] D. Bedrov, G.D. Smith, A.C.T. van Duin, Reactions of singly reduced ethylene carbonate in lithium battery electrolytes: a molecular dynamics simulation study using the ReaxFF, J. Phys. Chem. A 116 (2012) 2978–2985.
[4] A.D. Mayernick, M. Batzill, A.C.T. van Duin, M.J. Janik, A reactive force-field (ReaxFF) Monte Carlo study of surface enrichment and step structure on yttria-stabilized zirconia, Surf. Sci. 604 (2010) 1438–1444.
[5] F. Castro-Marcano, A.M. Kamat, M.F. Russo, A.C.T. van Duin, J.P. Mathews, Combustion of an Illinois No. 6 coal char simulated using an atomistic char representation and the ReaxFF reactive force field, Combust. Flame 159 (2012) 1272–1285.
[6] A.K. Rappe, W.A. Goddard, Charge equilibration for molecular-dynamics simulations, J. Phys. Chem. 95 (1991) 3358–3363.
[7] R.G. Parr, Density-functional theory, Chem. Eng. News 68 (1990), 45–45.
[8] M.F. Russo Jr., A.C.T. van Duin, Atomistic-scale simulations of chemical reactions: bridging from quantum chemistry to engineering, Nucl. Instrum. Methods Phys. Res. B 269 (2011) 1549–1554.
[9] H.M. Aktulga, J.C. Fogarty, S.A. Pandit, A.Y. Grama, Parallel reactive molecular dynamics: numerical methods and algorithmic techniques, Parallel Comput. 38 (2012) 245–259.
[10] A. Nakano, R.K. Kalia, K. Nomura, A. Sharma, P. Vashishta, F. Shimojo, A.C.T. van Duin, W.A. Goddard, R. Biswas, D. Srivastava, A divide-and-conquer/cellular-decomposition framework for million-to-billion atom simulations of chemical reactions, Comput. Mater. Sci. 38 (2007) 642–652.
[11] Sandia National Laboratories, LAMMPS. http://lammps.sandia.gov/ (accessed 10.05.14).
[12] M. Zheng, X. Li, L. Guo, Algorithms of GPU-enabled reactive force field (ReaxFF) molecular dynamics, J. Mol. Graph. Model. 41 (2013) 1–11.
[13] SCM, ADF. http://www.scm.com/ReaxFF/ (accessed 10.05.14).
[14] M. Zheng, X. Li, J. Liu, L. Guo, Initial chemical reaction simulation of coal pyrolysis via ReaxFF molecular dynamics, Energy Fuels 27 (2013) 2942–2951.
[15] M. Zheng, X. Li, J. Liu, Z. Wang, X. Gong, L. Guo, W. Song, Pyrolysis of liulin coal simulated by GPU-based ReaxFF MD with cheminformatics analysis, Energy Fuels 28 (2013) 522–534.
[16] S. Heller, A. McNaught, S. Stein, D. Tchekhovskoi, I. Pletnev, InChI – the worldwide chemical structure identifier standard, J. Cheminformatics 5 (2013) 7.
[17] IUPAC, InChI. http://www.iupac.org/home/publications/e-resources/inchi.html (accessed 10.05.14).
[18] D. Weininger, SMILES, a chemical language and information system: 1. Introduction to methodology and encoding rules, J. Chem. Inf. Comput. Sci. 28 (1988) 31–36.
[19] J. Gasteiger, Handbook of Chemoinformatics: From Data to Knowledge in 4 Volumes, Wiley-VCH Verlag GmbH, Weinheim, Berlin, 2008.
[20] T.H. Cormen, C.E. Leiserson, R.L. Rivest, C. Stein, Introduction to Algorithms, second ed., MIT Press and McGraw-Hill, Cambridge, 2001.
[21] Q. Zhang, W. Zhang, Y. Li, J. Wang, L. Zhang, T. Hou, A rule-based algorithm for automatic bond type perception, J. Cheminformatics 4 (2012) 26.
[22] P. Labute, On the perception of molecules from 3D atomic coordinates, J. Chem. Inf. Model. 45 (2005) 215–221.
[23] R. Sayle, PDB: Cruft to Content (Perception of Molecular Connectivity from 3D Coordinates). http://www.daylight.com/meetings/mug01/Sayle/m4xbondage.html (accessed 10.05.14).
[24] A.K. Dehof, A. Rurainski, Q.B.A. Bui, S. Böcker, H.-P. Lenhof, A. Hildebrandt, Automated bond order assignment as an optimization problem, Bioinformatics 27 (2011) 619–625.
[25] M. Froeyen, P. Herdewijn, Correct bond order assignment in a molecular framework using integer linear programming with application to molecules where only non-hydrogen atom coordinates are available, J. Chem. Inf. Model. 45 (2005) 1267–1274.
[26] M. Hendlich, F. Rippmann, G. Barnickel, BALI: automatic assignment of bond and atom types for protein ligands in the Brookhaven Protein Databank, J. Chem. Inf. Comput. Sci. 37 (1997) 774–778.
[27] Y. Zhao, T. Cheng, R. Wang, Automatic perception of organic molecules based on essential structural information, J. Chem. Inf. Model. 47 (2007) 1379–1385.
[28] J. Wang, W. Wang, P.A. Kollman, D.A. Case, Automatic atom type and bond type perception in molecular mechanical calculations, J. Mol. Graph. Model. 25 (2006) 247–260.
[29] N. O'Boyle, M. Banck, C. James, C. Morley, T. Vandermeersch, G. Hutchison, Open Babel: an open chemical toolbox, J. Cheminformatics 3 (2011) 33.
[30] Open Babel. www.openbabel.org (accessed 10.05.14).
[31] S. Bachrach, InChI: a user's perspective, J. Cheminformatics 4 (2012) 34.
[32] NCBI, PubChem. http://pubchem.ncbi.nlm.nih.gov/ (accessed 10.05.14).
[33] RSC, ChemSpider. http://www.chemspider.com/ (accessed 10.05.14).
[34] M.A. Serio, D.G. Hamblen, J.R. Markham, P.R. Solomon, Kinetics of volatile product evolution in coal pyrolysis: experiment and theory, Energy Fuels 1 (1987) 138–152.
[35] P.R. Solomon, T.H. Fletcher, R.J. Pugmire, Progress in coal pyrolysis, Fuel 72 (1993) 587–597.
[36] X. Li, Z. Mo, J. Liu, L. Guo, Revealing chemical reactions of coal pyrolysis with GPU-enabled ReaxFF molecular dynamics and cheminformatics analysis, Mol. Simul. (2014), http://dx.doi.org/10.1080/08927022.2014.913789.
[37] X. Liu, X. Li, J. Liu, Z. Wang, B. Kong, X. Gong, X. Yang, W. Lin, L. Guo, Study of high density polyethylene (HDPE) pyrolysis with reactive molecular dynamics, Polym. Degrad. Stabil. 104 (2014) 62–70.
[38] R. Vinu, L.J. Broadbelt, Unraveling reaction pathways and specifying reaction kinetics for complex systems, Annu. Rev. Chem. Biomol. Eng. 3 (2012) 29–54.
[39] T. Engel, Basic overview of chemoinformatics, J. Chem. Inf. Model. 46 (2006) 2267–2277.
[40] N. O'Boyle, R. Guha, E. Willighagen, S. Adams, J. Alvarsson, J.-C. Bradley, I. Filippov, R. Hanson, M. Hanwell, G. Hutchison, C. James, N. Jeliazkova, A. Lang, K. Langner, D. Lonie, D. Lowe, J. Pansanel, D. Pavlov, O. Spjuth, C. Steinbeck, A. Tenderholt, K. Theisen, P. Murray-Rust, Open data, open source and open standards in chemistry: the Blue Obelisk five years on, J. Cheminformatics 3 (2011) 37.
[41] Kitware, Visualization Toolkit. www.vtk.org (accessed 10.05.14).
[42] D. Lonie, Chemistry Visualization. http://www.kitware.com/source/home/post/44 (accessed 10.05.14).
[43] GGA, Indigo Toolkit. http://www.ggasoftware.com/opensource/indigo (accessed 10.05.14).
[44] Sourceforge, RDKit. www.rdkit.org (accessed 10.05.14).
[45] Accelrys, Materials Studio. http://accelrys.com/products/materials-studio/index.html (accessed 10.05.14).