In the case that the parameters to describe the force field, such as bond angles and charges, cannot be added to the library of a molecular dynamics (MD) simulation, self-development of the force field should be considered by performing quantum mechanics calculations and/or utilizing an automatic parameter generation tool. However, these techniques are not suitable for macromolecules with a large number of atoms. Typically, the force field of an oligomer containing three unit structures (a unit at both ends and a repeating unit at the center) is calculated and converted to polymer form (both ends + central part × n). Considering this, we recently developed the program o2p, which is a semi-automated program designed to set up the force field for polymers with repeating structures. However, it is difficult to apply this method to macromolecules with complex repeating structures. Thus, in this project, we developed PolyParGen, a new open-source automatic force field generation program for Gromacs that can relatively easily and reliably simulate the MD of complex macromolecules. The proposed program (1) divides the structure of the polymer into substructures with a number of atoms within the limit of the handling size for the automatic parameter generation tool program; then, (2) acquire the parameters for each divided substructure, and finally, (3) combine the parameters of these substructures to obtain the parameters for the whole polymer. By automating these processes, it is possible to acquire a parameter of a polymer having complicated structures. This program was evaluated by simulating the polymers P3EHT and F-P3EHT in chloroform. In agreement with previous reports, fluorination was found to cause F-P3EHT to adopt an extended structure, thereby indicating the effectiveness of the proposed program.

Keywords: Molecular dynamics simulation, Macromolecules, Parameters, Gromacs, Polymer
structure of the polymer such that the size does not exceed the limit on the number of atoms that can be treated by the automatic parameter generation tool (2. Split small parts). In this step, the program is designed to partition the structure while making sure to avoid the midpoint of the double bond and not destroy the ring structure. The third step entails acquiring the parameters of the partial structure by using the automatic parameter generation tool (3. Make molecular parameter). The fourth step is the process of combining parameter files to build all of the polymer parameters (4. Combine each part). As this step entails combining multiple files of the partial structures, a considerably large number of characters, which have been specifically allocated to each structure via the automatic parameter generation tool, must be substituted to enable the development of a unified name for the entire polymer molecule. It is also necessary to detect and delete the duplication of the atoms at the connecting points. In addition, in order to extend the applicability of the MD simulation to complex macromolecules without regular repeating structures, we aimed to develop a parameter acquisition program (polymer format -> partial structure -> polymer form), i.e., PolyParGen, that is suitable for Gromacs, an open-source MD simulation package [5]. In conclusion, we developed the new program, PolyParGen, which can generate the force field for the polymers with complex repeating structures, by utilizing the open tools of the automatic parameter generation for small molecules.

### 2 DEVELOPMENT ENVIRONMENT

The proposed program includes a web application that was created as a user interface by using PHP [6], HTML, and JavaScript, whereas Python was used as the platform to develop the main program [7]. Additionally, LigParGen, which is employed as an automatic parameter generation tool, is executed on the same server. With the proposed program, when a user can retrieve a set of the MD simulation files for the registered molecule after registering the macromolecular structure files. However, this retrieval process may be time expensive, as computational time is dependent on the size and number of atoms. Thus, to ensure that the simulation does not overload the system, we tentatively set an upper limit for the processing time on our server open to the public.

### 3 SYSTEM OVERVIEW

#### 3.1 Division of the polymer into partial structures

The input polymer structure is divided into partial structures with a size that is within the limit for an input to LigParGen (i.e., approximately 150 atoms). The hydrogen atoms are added to the dangling bond at the cut edges of the divided structures (Figure 2 (a)). Note that, in the process of polymer structure division, the double and triple bonds, in addition to the ring structures, should not be cut in order to preserve the parameter type of the structures (Figure 2 (b)); this is because each parameter is dependent on the type of the molecular structure. It should also be noted that the divided partial structures are allowed to overlap so that the parameters can be combined in a subsequent process (Figure 2 (c)).

#### 3.2 Automatic parameter generation (LigParGen)

As previously mentioned, the LigParGen program was used to obtain the parameters of the partial structures. Generally, LigParGen is a program with a Web interface, and requires that each parameter be manually obtained for each structure. However, because the proposed program creates several dozens of partial structures, it is preferable to automate the LigParGen process to obtain the parameters. For this reason, LigParGen was installed and executed on an on-site server, which was also responsible for sustaining the load of other modules, in consideration of the load, execution time, and stability of the entire system.

![Figure 1. Procedural flow of the process to prepare a set of files required to simulate polymer MD.](image-url)
This module combines the parameters of the partial structures obtained via the LigParGen program to create an itp file for the whole polymer molecule. The procedure is explained using the parameterization of poly (3-ethylhexylthiophene) (P3EHT) 30mer as an example. The polymer chain of P3EHT 30mer (1022 atoms, Figure 3 (a)) was divided into 28 partial structures by PolyParGen according to the procedure of Figure 1 2. The partial structures were shown in Figure 3 (b). Then, the parameters of the partial structures were obtained by LigParGen according to the process shown in Figure 1 3. Note that, some of the atoms at the connecting points between two adjacent partial structures were allowed to have overlap atoms so that the parameters can be combined in a subsequent process.

In the module of PolyParGen, the first and second partial structures were first combined at the connecting points with overlapping atoms by deleting one of the overlapped atoms and their parameters, and the terminal hydrogen atoms added to prevent the dangling bonds in the partial structures. When two partial structures were combined, partial charges of the overlapping atoms were selected to reduce the total charge of the combined structure (1–2). Then, the combined structure was further combined to the third partial structure and formed the larger structure (1–2-3). This procedure was repeated until the parameters of whole molecule was obtained (Figure 3 (c)). However, the residual partial charges would remain by the combination procedure. It will not be usually canceled but piled up in the net charge of the whole molecule. Thus, it is necessary to adjust the partial charges on the atoms to become zero in the total charge. In this program, the residual charges are equally distributed and subtracted on each atom. Since the number of atoms in the polymer is large, the distributed partial charge on each atom becomes small. In the case of the polymer chain of P3EHT 30mer with 1022 atoms, the total residual charge of −7.0302 was distributed to each atom with 0.0069. This is sufficiently small from the practical point of view. The proposed module also produces the gro file for the itp file.

### 4 APPLICATION EXAMPLES

MD simulations of poly (3-ethylhexylthiophene) (P3EHT) and poly (3-ethylhexyl-4-fluorothiophene) (F-P3EHT) in chloroform were conducted in order to evaluate the proposed program. F-P3EHT is a fluorinated P3EHT that is reported to have an extended structure in chloroform [8]. As another example, MD simulations of polystyrene in THF and water were also conducted. The procedures are described below.

#### 4.1 P3EHT and F-P3EHT

##### 4.1.1 Create initial structure

The structure of P3EHT 4-mers was prepared by using Avogadro, an advanced molecule editor and visualizer [9]. The structure of the 2-mers at the center of 4-mers was duplicated 14 times by using o2p to obtain a structure of 30-mers that was then optimized.
by using the Avogadro editor. Then, F-P3EHT was obtained by using Avogadro to fluorinate the P3EHT 30-mers before structure optimization. The initial structures are shown in Figure 4.

4.1.2 Acquisition of itp and gro files

The structure file obtained in the previous step was used as the input to PolyParGen. As previously mentioned, in the proposed program, the polymers were divided into the partial structures in order to determine their parameters. Then, the partial structures, with their defined parameters, were joined to recreate the whole polymer molecule. Finally, the itp and gro files were downloaded. Note that all processes were automatically performed, and the parameter file for chloroform was separately obtained by LigParGen.

4.1.3 MD simulation

The obtained files were implemented in the 2018 version of Gromacs to perform the 20-ns MD simulation. The simulation conditions were set as suggested by Hua et al. [8]. That is, P3EHT, and F-P3EHT were set in a 14.4-nm cubic box filled with chloroform, and the periodic boundary condition was applied. However, the original parameter values reported by Hua et al. were not used, as the values obtained via PolyParGen and LigParGen were applied in order to validate our program. As can be seen in Figure 5, the MD simulation yielded an extended F-P3EHT structure. Furthermore, it was confirmed that the end-to-end length of F-P3EHT was comparable to that reported in previous literature (Figure 6).

4.2 Polystyrene

4.2.1 Create initial structure

Atactic polystyrene with 150-mers was constructed by using MarvinSketch [10]. The initial structure is shown in Figure 7.

4.2.2 Acquisition of itp and gro files

The obtained structure file was input to PolyParGen, and as with the preceding example, the itp and gro files were downloaded from this program. The parameter file for THF was separately obtained via LigParGen.

4.2.3 MD simulation

As with the preceding example, the obtained files were implemented in the 2018 version of Gromacs to perform the 40-ns MD simulation. The simulation conditions were set as suggested by Morozova et al. [11]. A polystyrene molecule was set in a box with dimensions of $42 \times 12 \times 12$ nm$^3$ containing either THF or water. As is shown in Figure 8, the THF-based MD simulation yielded an extended molecular structure of polystyrene.

5 CONCLUSION

With the proposed program, it becomes possible to easily prepare the files necessary for the MD simulation of macromolecules with complex repeating structures and dendrimer-type large molecules. Moreover, it simplifies the setup process that must precede MD simulation. Furthermore, the effectiveness and accuracy of this program were confirmed through the reproduction of previously reported results. However, there are two limitations that should be considered when using this program. First, to obtain the parameters from the LigParGen program, the net charge of the input molecule must be within $\pm 2$ charges. When the charge of a partial structure that has been divided by using the PolyParGen program exceeds...
± 2 as a result of a treatment implemented by the program, e.g., protonation, it is impossible to prepare the parameter. Therefore, the program is currently only applicable to neutral molecules. The second consideration is related to molecules with large ring structures and/or cross-linked structures. When PolyParGen divides the molecule into various partial structures, cyclic structure interruptions and midpoint breakage of double and triple bonds should be avoided. Therefore, in case that the target macromolecule has a large ring structure, or a condensed ring and/or a cross-linked structure, it is not possible to find an appropriate cutting site that would yield a partial structure with a number of atoms that is allowable for LigParGen input. Future work will focus on modifying the proposed program to overcome these limitations. This program is free and open to all at http://www.polypargen.com/.

**REFERENCES**


