Chapter 2
Methodology and Tools

2.1 Introduction
Molecular simulation is a useful tool for the study of the properties of a system containing many molecules. Molecular mechanics uses Newtonian mechanics to describe the forces between atoms and molecules. Monte Carlo (MC) and molecular dynamics (MD) are different techniques to approach molecular behaviour that enable to simulate molecular motion. MD resolves Newton’s equation of motion in a step-by-step manner, whereas MC samples configuration in a random manner. In this work, MD was chosen over MC as it was available, but both methods lead to the same results: the ergodic hypothesis states that for a random process, if a sufficient amount of data is gathered at time intervals on a single system (MD) and the data gathered from an ensemble of similar a system (MC), then they have the same statistical properties (McQuarrie, 2000). Moreover, it has been proved that MD and MC lead to similar thermodynamical and conformational results (Jorgensen and TiradoRives, 1996). However, simulation techniques have their limitations, due firstly to the limitation of computer power. Even if computers are improving at a fast rate, computer time has to be balanced with a realistic approach. Simulations are usually run for a few nanoseconds, whereas the timescale to crystallise a polymorphs takes up to four hours (Cashell et al., 2003; Cashell et al., 2004; Ferrari and Davey, 2004). These time scales are not representative of the time scales available in molecular simulation, due to the limitation of today's computer power.

In this chapter, the methodology used for this thesis will be described, by means of an extensive description of the necessary background theory of the simulation algorithm followed by a description of all the tools used. Finally, a description of the tools developed for the analysis of the simulation results will complete this section, leading to a complete understanding of how the simulations have been performed and the resulting data analysed.

2.2 Molecular Dynamics
Molecular dynamics simulations compute the equilibrium and transport properties of a system by solving Newton’s equations of motion (Frenkel and Smit, 2002). The
algorithm used to perform an MD simulation works in steps: in the first step, it reads the parameters specific to that particular run (e.g. temperature, number of steps), the second initializes the system by assigning velocities to the position of the molecules; the algorithm then calculates the forces on each particle in a third step, and the fourth step integrates Newton’s equation of motion. Steps three and four are repeated until the required simulation time is reached. The program then calculates the averages of the calculated quantities (different energies, volume, temperature, pressure...). MD simulates the time evolution of a molecular system, and provides a trajectory of the molecules, which can then be viewed using visualization software, i.e. VMD (Humphrey et al., 1996). An equilibration period is usually required, which allows the system, i.e. the collection of molecules being simulated, to adjust to the conditions of the simulation, i.e. temperature, force field etc. Often the starting arrangement of the molecules in the system is a regular array that is not characteristic of the real system; so a period of simulation is required to allow the system to move from its original state to a state that is at equilibrium with the specific conditions of the simulation. This “equilibration” period is not included in any of the time dependent averaging of the system. The end of this period is characterised by the change from wildly oscillating values and averages of energies, temperature etc, to values that oscillate smoothly around some average value.

The initial coordinates of the atoms are usually provided to the simulation as an input file. The forces governing the atoms present are usually provided as a collection of mathematical functions.

2.2.1 Periodicity
In order to calculate bulk properties of a system using molecular simulation it is necessary to simulate many atoms or molecules. In most cases this is achieved by simulating a number of atoms or molecules in a box. However, this introduces its own problem – the edges of the box provide surfaces, which are clearly not representative of the bulk. In order to overcome this, periodic boundaries are introduced. The same box is replicated in the three dimensions to form an infinite lattice. If one molecule moves in the original box, all its periodic images will move in the same direction. Thus, if a molecule leaves the original box to move in a neighbouring box, one of its images will enter the original box from the opposite face, as show on Figure 2.1. It is then unnecessary to model all of the boxes (which are an infinite number) as a simulation run
using the central box will take into account all periodic images, and not have any surfaces.

Figure 2.1: Schematic representation of the periodic boundaries. The shaded boxes represent the periodic images of the central (clear) simulation box.

2.2.2 Forcefield

The simulation progress is governed by how the atoms move. The forces that describe how these atoms move are gathered into a forcefield: it is the complete list of all the interatomic forces. The interatomic force $F(r)$ is derived from the interatomic energy $U(r)$ through the expression:

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

(2.1)

where $\hat{r} = r^{-1} \cdot r$ is a unit vector along the pair-separation vector $r$.

Hence the interatomic energy can be derived from the force using:

$$ U(r) = \int_{0}^{\infty} F(r') \cdot \kappa dr' $$

(2.2)
where $F(r) = |F(r)|$

For an atomic system, the interatomic forces are sufficient to determine the behaviour of the system. For a molecular system, these forces can be divided into intra- and inter-molecular, depending on whether or not the atoms involved are in the same molecule.

### 2.2.2.1 Intramolecular Interactions

The expressions describing the relationship between atoms from the same molecules can be divided into three categories: the energy terms describing the two body interaction (bonds), the three body interactions (angles) and the four body interactions (dihedrals). These energy terms are described in their simplest form in equations (2.3), (2.4) and (2.5) for the bonds, angle and dihedral terms, respectively, where the subscripts $i$, $j$, $k$ and $l$ denote the specified atoms, $r_{ij}$ the distance between $i$ and $j$, $\theta_{ijk}$ is the angle formed by $ij$ and $k$ and $\phi_{ijkl}$ is the angle between the planes $ijkl$. $k_r$, $k_\theta$ and $k_\phi$ are the force constants described by the chosen force field, and parameters with the subscripted $0$ refer to the geometrical attribute at an equilibrium position and are fixed by the force field. In equations (2.3) to (2.5) these are analogous to the strength of a Hooke’s Law spring. However, in many force field implementations, a more complex form of the dihedral is used, for example see equation (2.20).

\[
U_{\text{bonds}}(r_{ij}) = \frac{1}{2} k_r (r_{ij} - r_0)^2 \tag{2.3}
\]

\[
U_{\text{angles}}(\theta_{ijk}) = \frac{1}{2} k_\theta (\theta_{ijk} - \theta_0)^2 \tag{2.4}
\]

\[
U_{\text{dihedral}}(\phi_{ijkl}) = \frac{1}{2} k_\phi (\phi_{ijkl} - \phi_0)^2 \tag{2.5}
\]

### 2.2.2.2 Intermolecular Interactions

Interactions between atoms in different molecules (i.e. non bonded) are addressed using a combination of short range and long range interactions. Those atoms within the same molecule that are separated by more than four atoms, i.e. not governed by equations
(2.3) to (2.5), also interact this way. The non-bonded interactions are categorized as short-range and long-range interactions.

**Short Range**

\begin{equation}
U(r) = A \exp(-B_{ij}) \left( \frac{C}{r_{ij}^6} \right)
\end{equation}

which is known as the Buckingham potential, where the exponential term describes the repulsive part of the potential and the \( r^6 \) term, the attractive part. The exponential term of the Buckingham potential can be time consuming to compute, so a mathematically simpler representation of the intermolecular energy would be preferable. The most commonly used model is the Lennard-Jones 12-6 potential (Lennard-Jones, 1931), and is described in Equation 2.7.
\[
U(r) = 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right]
\] 
(2.7)

where \(\sigma\) and \(\varepsilon\) are the contact distance and well depth respectively (as illustrated in Figure 2.2) and \(r\) is the distance between the two atoms \(i\) and \(j\).

The van der Waals type functions asymptotically approach zero at large separations, so it is common to ignore the potential above some cut-off distance (commonly 2.5 or 3 times the collision distance), whereupon the \(U(r)\) function is most commonly assumed to be zero above this cut-off, following Equation (2.8)

\[
U(r) = \begin{cases} 
4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right], & \text{for } r \leq r_{\text{cut}} \\
0, & \text{for } r > r_{\text{cut}} 
\end{cases}
\]

(2.8)

\(\sigma\) and \(\varepsilon\) are computed for unlike atom pairs using the Lorentz-Berthelot mixing rules (Lorentz, 1881; Berthelot, 1898):

\[
\sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j)
\] 
(2.9)

\[
\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}
\] 
(2.10)

The Lennard-Jones expression is often simplified further by calculating beforehand the product of \(\sigma\) by \(\varepsilon\) which results in the expression:

\[
U_{ij}(r) = \left( \frac{A_{ij}}{r_{ij}^{12}} \right) - \left( \frac{B_{ij}}{r_{ij}^{6}} \right)
\] 
(2.11)

Where \(A_{ij} = 4\sigma_{ij}\varepsilon_{ij}^{12}\) and \(B_{ij} = 4\sigma_{ij}\varepsilon_{ij}^{6}\). In this thesis, Equation (2.11) has been used to describe the Lennard-Jones interactions; however, for the purpose of quoting the force
field parameters, the $\sigma$ and $\varepsilon$ of Equation 2.7 are used for ease of interpretation (see Figure 2.2)

The inter-atomic interactions for the simulated system are expressed by the sum of the Lennard-Jones interactions between all atom pairs.

$$U_{ij}^{\text{LJ}} = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right]$$

(2.12)

**Long Range**

The long range interactions are described by the electrostatic interactions, i.e. Coulombic interactions, and the energy of this interaction between two atoms is described by the equation:

$$U(r_{ij}) = \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{r_{ij}}$$

(2.13)

where $r_{ij}$ is the distance between the charged species $i$ and $j$, $q_i$ and $q_j$ are the charges at point $i$ and $j$ and $\varepsilon_0$ is the permittivity of free space.

To incorporate electrostatic interactions in a periodic system, the sum of all of the pair interactions becomes infinite, and therefore impossible to calculate directly. Two schemes are used to avoid this problem. The first one is to introduce a cut-off distance beyond which interactions are neglected. Such a calculation can be very time consuming, even for a powerful computer, because a large cut-off would be required in order not to neglect important contributions to the energy of the system. The alternative is to use the Ewald sum algorithm (Ewald, 1921). This method enables the calculation of the electrostatic contributions of all the atoms in all the periodic images. It divides the space into two: the real space, at short range and the reciprocal space, which extends to infinite values of $r$. The Ewald sum is expressed by:
Chapter 2: Methodology and Tools

\[ U_{\text{Coulomb}} = U_{\text{real}} + U_{\text{reciprocal}} - U_{\text{self}} \]

\[ U = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left( \sum_{|k|=0}^{\infty} q_i q_j \frac{\text{erfc}(\kappa |r_{ij} + n|)}{|r_{ij} + n|} \right) \]

\[ + \frac{1}{\pi L^3} \sum_{k \neq 0} q_i q_j \frac{4\pi}{k^2} \exp \left( -\kappa^2 \right) \cos(k \cdot r_{ij}) \]

\[ - \kappa \sum_{i=1}^{N} q_i^2 \]  

Equation (2.14)

The first term describes the coulombic interactions in the real space, where Gaussian charge distributions are centred on each ion. The second term describes the charge distribution applied to the periodic images in the reciprocal space and in the real space. Because the electrostatic interactions are calculated twice in the real space, a correction is necessary to remove the charge distribution in the real space. This correction term is the third term in Equation 2.14 and is subtracted form the previous two terms. The Ewald method therefore replaces a potentially infinite sum in real space by two finite sums: one in real space and one in reciprocal space, minus the self energy correction.

The accuracy of the long range calculations is dependent on the size of the real and reciprocal spaces, which is based on five variables: the delimitation of the real space, determined by \( r_{\text{cut}} \), the distances that define the reciprocal space, determined by three vectors, \( k_1, k_2, k_3 \), one for each of the three axis directions and the Ewald convergence parameter \( \kappa \). The rate of convergence of both sums is governed by \( \kappa \), and the recommended value is \( 3.2/r_{\text{cut}} \) (Smith et al., 2003).

The Smooth Particle Mesh Ewald (SPME) method is a modification of the standard Ewald method. The main difference from the Ewald method is how SPME calculates the reciprocal space terms: it uses interpolation, and represents the reciprocal space on a three dimensional rectangular grid. This method is a lot faster than the standard Ewald method especially for larger values of \( k \).

2.2.2.3 Dielectric continuum

Another method widely applied in molecular dynamic simulations is the use of a dielectric continuum: it mimics the electric field of a solvent, without explicitly modelling the solvent atoms, resulting in a dramatic decrease in computer time and
without such a large loss of accuracy. The presence of the dielectric field affects the Coulomb algorithm compared to a simulation in vacuum. The dielectric constant ($\varepsilon_r$), sometimes called relative permittivity, is used to determine the ability of an insulator to store electrical energy (static energy). It can be expressed with the equation

\[ \varepsilon_r = \frac{C_x}{C_0} \]  

(2.15)

where $C_0$ is the capacitance between two plates in air or vacuum and $C_x$ is the capacitance between two plates in a medium $x$. In the presence of a dielectric constant, the interaction potential of two particles of opposite charges of equal magnitude and separated by a distance $r$ becomes:

\[ U(r_{ij}) = \frac{e^2}{4\pi \varepsilon_0 \varepsilon(r_{ij}) r_{ij}} \]  

(2.16)

The MD software uses a different approach, which uses a distance dependent dielectric function $\varepsilon(r) = \varepsilon r$, $\varepsilon$ being a constant.

\[ U(r_{ij}) = \frac{1}{4\pi \varepsilon_0 \varepsilon(r_{ij})} \frac{q_i q_j}{r_{ij}} \]  

(2.17)

The dielectric constant $\varepsilon$ is 1 for vacuum, 80.4 for water at ambient temperature (Sears et al., 1982) and 70.7 at 298K for the water model used in this work (Rami Reddy and Berkowitz, 1989).

### 2.2.3 Summary

The total energy of the system calculated for each time step is the sum of all the energies explained in the previous sections. It can be written in the general form as Equation (2.18), and, in the case of this study, as Equation (2.19), which describes the “force-field” used to describe the simulated system.

\[ U_{\text{total}} = U_{\text{bonds}} + U_{\text{angles}} + U_{\text{dihedrals}} + U_{\text{van der Waals}} + U_{\text{elec}} \]  

(2.18)
\[ E_{\text{total}} = \sum_{\text{bonds}} \frac{1}{2} k_i (r_i - r_0)^2 + \sum_{\text{angles}} \frac{1}{2} k_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} \frac{1}{2} k_\phi (\phi - \phi_0)^2 + \sum_{i<j} \left[ \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + q_q q_i \right] \]

(2.19)

The accuracy of these intermolecular interactions will determine the accuracy of the model. In order to completely describe the potential energy of each atom, a full set of parameters describing Equation (2.19) for the system of interest must be provided. These parameters are listed in a force field which is comprised of the force constants and equilibrium constants required to completely define the force field for the system to be simulated.

2.3 System

In order to configure a molecular simulation, several factors must be taken into account prior to start of the simulation run, like the computer used, the software, and what is needed by the software to run a successful MD simulation. The following subsections describe the requirements for the simulations performed in this thesis.

2.3.1 Computer Set Up

Two computer set up were used for the work described in this thesis: the first system comprised of four computers, each equipped with two Pentium® Xeon™ CPU 2.4GHz processors, operating on Linux Red Hat operating system. The software used for the molecular dynamics study on this system was the version 2.16 of DL_POLY (Smith and Forester, 1996). The second system used for larger system is a parallel workstation comprising of 256 processors. The 2.18 version of DL_POLY was used on this workstation, and simulations were run in a parallel manner on a single node of four processors. The molecular dynamics simulation package DL_POLY requires a starting configuration, the inter- and intra-molecular forces parameters, and commands about the chosen simulation run.

2.3.2 Atom Type

The atom type is a set of names given to each atom present in the system in order to label them in terms of their interactions behaviour. The same atom type is used to describe all atoms that interact through the same interatomic force. For clarity, oxygen atoms will generally start with “O”, carbon atoms with “C”, etc. Several atoms can
share the same name if they are equivalent in the molecule, e.g. the two hydrogen atoms of a water molecule. However, the atom type was chosen so that it is possible to differentiate certain equivalent atoms (i.e. H1, H2 and H3 are three equivalent hydrogen atoms in the NH3 group). Figure 2.3 describes the atom type used throughout this study for the glutamic acid molecule.

Figure 2.3: Molecular structure and atom types of L-Glutamic acid. Carbon atoms are gray, oxygen atoms red, nitrogen atom blue and hydrogen atoms white.

### 2.3.3 Force Field

Several force fields are available: one of the most common is the CHARMM (Chemistry at Harvard Macromolecular Mechanics) force field (Brooks et al., 1983), implemented in the software of the same name. It is used mainly for macromolecules like proteins. The CHARMM force field can deal with explicit atoms for small systems, and for bigger systems, the hydrogen atoms are bound to the heavy atom to which they are bonded. This technique consists of considering groups of atoms, e.g. CH3, using specific parameters for the group rather than assigning parameters to each atom individually, and is known as the united atom model. Unfortunately, the function that describes the four body torsional interactions is not compatible with any of the built-in equations that describe these interactions in DL_POLY.

Another very commonly used force field is the GROMOS force field (van Gunsteren et al., 1996) implemented in the GROMOS package. It is used for biomolecular systems, like peptides (Peter et al., 2000), proteins (Stocker and van Gunsteren, 2000), sugars (Mark et al., 1994) and nucleotides (Bonvin et al., 1998), and more recently on lipids...
The OPLS (Optimized Potential for Liquid Simulations) force field was first designed to model protein systems (Jorgensen and Tirado-Rives, 1988) using the AMBER united atom force field for the bonded interactions. It was later developed to simulate organic liquids (Jorgensen et al., 1996). The DREIDING force field (Mayo et al., 1990) uses general force constants and geometry parameters based on the hybridization of the atoms rather than on the individual force constants and geometric parameters that depend on the particular combination of atoms involved in the bonds, angles and dihedrals.

The force field chosen for the glutamic acid molecules in this work was the AMBER (Assisted Model Building and Energy Refinement) force field (Cornell et al., 1995) derived from an earlier version developed by Weiner et al. (Weiner et al., 1984). It was designed to simulate the structures, conformational energies and interaction energies of proteins and nucleic acids. It performs well with carboxylic acids (Cornell et al., 1995), and the equations for the intermolecular interactions are mostly compatible with DL_POLY. A description of the intermolecular forces parameters can be found in Appendix Error! Reference source not found.. The atom centred charges were derived quantum mechanically in this force field, but for the peptide form of the amino acids. The set of charges used in this work is described in Table 2.1 and an explanation on how these charges were derived can be found in section 2.3.5. The van der Waals parameters are described with a Lennard-Jones 12-6 potential, with parameters described in Table 2.1 following the atom type given in section 2.3.2.
Chapter 2: Methodology and Tools

Table 2.1: Lennard-Jones parameters for each atom type from the AMBER force field and their charges

<table>
<thead>
<tr>
<th>Atom Name</th>
<th>( \sigma (\text{Å}) )</th>
<th>( \varepsilon (\text{kcal.mol}^{-1}) )</th>
<th>Zwitterion Charge ( q )</th>
</tr>
</thead>
<tbody>
<tr>
<td>N3</td>
<td>3.341</td>
<td>0.170</td>
<td>-0.300</td>
</tr>
<tr>
<td>C1</td>
<td>3.400</td>
<td>0.086</td>
<td>0.700</td>
</tr>
<tr>
<td>C2</td>
<td>3.400</td>
<td>0.086</td>
<td>0.465</td>
</tr>
<tr>
<td>C3</td>
<td>3.400</td>
<td>1.094</td>
<td>0.310</td>
</tr>
<tr>
<td>C4</td>
<td>3.400</td>
<td>1.094</td>
<td>-0.879</td>
</tr>
<tr>
<td>C5</td>
<td>3.400</td>
<td>1.094</td>
<td>-0.879</td>
</tr>
<tr>
<td>O1</td>
<td>2.970</td>
<td>0.210</td>
<td>-0.553</td>
</tr>
<tr>
<td>O2</td>
<td>2.970</td>
<td>0.210</td>
<td>-0.800</td>
</tr>
<tr>
<td>HO</td>
<td>3.067</td>
<td>0.210</td>
<td>-0.573</td>
</tr>
<tr>
<td>HO</td>
<td>0.0</td>
<td>0.0</td>
<td>0.459</td>
</tr>
<tr>
<td>H1</td>
<td>1.069</td>
<td>0.016</td>
<td>0.33</td>
</tr>
<tr>
<td>H2</td>
<td>1.069</td>
<td>0.016</td>
<td>0.33</td>
</tr>
<tr>
<td>H3</td>
<td>1.069</td>
<td>0.016</td>
<td>0.33</td>
</tr>
<tr>
<td>HC</td>
<td>2.649</td>
<td>0.016</td>
<td>0.465</td>
</tr>
<tr>
<td>HP</td>
<td>1.960</td>
<td>0.016</td>
<td>0.0</td>
</tr>
</tbody>
</table>

The expressions used to describe the intermolecular interactions in AMBER are similar to those described in equations (2.3) and (2.4). The expression for the dihedral energy, however, is different to equation (2.5) and is described by:

\[
U_{ijkl} = \frac{V_n}{2} \left[ 1 + \cos(n\phi_{ijkl} - \phi_0) \right] 
\]  

(2.20)

with \( V_n \) being the torsional potential (magnitude of torsion in kcal/mol) and \( n \) the periodicity of the torsion angle.

So that the total energy is written as:

\[
E_{\text{total}} = \sum_{\text{bonds}} \frac{1}{2} k_b (r_j - r_0)^2 + \sum_{\text{angles}} \frac{1}{2} k_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} \frac{V_n}{2} \left[ 1 + \cos(n\phi_{ijkl} - \phi_0) \right] + \sum_{i<j} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}
\]  

(2.21)

In order to respect electroneutrality, in some systems where glutamic acid does not have a neutral global charge, ions had to be included in the system. \( \text{Na}^+ \) and \( \text{Cl}^- \) parameters where taken from Smith and Dang (1994) and Jorgensen et. al. (1987), respectively, and are shown in Table 2.2. Other ions were required for specific simulations in order to
approximate the pH of the solutions, hence H$_3$O$^+$ molecules were added to the simulation. The potentials are taken from Rodwell et al. (1981) and Kusaka et al. (1998) for the hydronium cation, and are shown in Table 2.2.

<table>
<thead>
<tr>
<th>ion names</th>
<th>$\sigma$ (Å)</th>
<th>$\epsilon$ (kcal.mol$^{-1}$)</th>
<th>$q$ zwitterion ($e$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Na$^+$</td>
<td>2.350</td>
<td>0.130</td>
<td>1</td>
</tr>
<tr>
<td>Cl$^-$</td>
<td>4.417</td>
<td>0.118</td>
<td>-1</td>
</tr>
<tr>
<td>O (H$_3$O$^+$)</td>
<td>2.9009</td>
<td>0.274</td>
<td>-0.248</td>
</tr>
<tr>
<td>H (H$_2$O)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.416</td>
</tr>
</tbody>
</table>

Table 2.2: Lennard-Jones parameters for the ions used and their charges (see text for references)

### 2.3.4 Water Model

Many computational studies have been undertaken to investigate water-water interactions in order to simulate properties either as a pure species or as a solvent. Most of the special properties of water are due to the ability of water to form hydrogen bonds with other water molecules. Experimentally, a molecular level investigation of water has been performed using X-ray diffraction (Narten, 1971) and neutron scattering (Neilson and Enderby, 1986; Soper and Phillips, 1986). Many different potentials have been derived, using either a rigid or flexible molecular structure. Some of the most common water models include TIP3P (transferable intermolecular potential 3P) (Jorgensen et al., 1983), TIP4P (Jorgensen et al., 1983), SPC (simple point charge) (Berendsen et al., 1981) and SPC/E (extended simple point charge) (Berendsen et al., 1987). They all describe a 3 point charge potential, apart from TIP4P, which uses 4 sites for the electrostatic interactions. They all describe a rigid pair potential created using different Lennard-Jones and Coulombic terms, along with a slightly different geometry that leads to different bulk water properties. A summary of the water model properties gathered from a wide source of molecular simulation is presented in Table 2.3 and Table 2.4. The first table describes the physical properties and the second describes the water-water interactions via the RDF peak position. All the parameters are compared with experimental value. Although TIP4P/2005 gives the closest physical properties, the inter-atomic correlation was not discussed in details (Abascal and Vega, 2005). Moreover, the parameters of this water model were published in December 2005, well after this work had started.
Chapter 2: Methodology and Tools

<table>
<thead>
<tr>
<th>Model</th>
<th>Ref.</th>
<th>(\rho) (g/cm(^3))</th>
<th>(C_p) (cal.mol(^{-1}).K(^{-1}))</th>
<th>(\Delta H) (kcal.mol(^{-1}))</th>
<th>(\varepsilon)</th>
<th>(D) (x 10(^9) m(^2)/s)</th>
<th>(E_{pot}) (kJ/mol)</th>
<th>(I_h) (K)</th>
<th>(T_{md}) (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPC</td>
<td>(Spoel et al., 1998)</td>
<td>0.988</td>
<td></td>
<td></td>
<td>63</td>
<td>4.2</td>
<td>-42.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPC</td>
<td>(Jorgensen and Tirado-Rives, 2005)</td>
<td>0.985</td>
<td>20</td>
<td>10.74</td>
<td>60</td>
<td>3.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPC</td>
<td>(Mark and Nilsson, 2001)</td>
<td>0.988</td>
<td></td>
<td></td>
<td>4.2</td>
<td>-40.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPC</td>
<td>(Vega et al., 2005)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TIP3P</td>
<td>(Spoel et al., 1998)</td>
<td>1.001</td>
<td></td>
<td></td>
<td>86</td>
<td>5.4</td>
<td>-40.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TIP3P</td>
<td>(Jorgensen and Tirado-Rives, 2005)</td>
<td>1.002</td>
<td>20</td>
<td>10.41</td>
<td>88</td>
<td>5.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TIP3P</td>
<td>(Mark and Nilsson, 2001)</td>
<td>0.998</td>
<td></td>
<td></td>
<td>5.6</td>
<td>-39.0</td>
<td></td>
<td></td>
<td>-15</td>
</tr>
<tr>
<td>TIP3P</td>
<td>(Vega et al., 2005)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TIP3P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TIP4P</td>
<td>(Spoel et al., 1998)</td>
<td>0.999</td>
<td></td>
<td></td>
<td>52</td>
<td>3.7</td>
<td>-41.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TIP4P</td>
<td>(Jorgensen and Tirado-Rives, 2005)</td>
<td>1.001</td>
<td>20</td>
<td>10.65</td>
<td>60</td>
<td>3.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TIP4P</td>
<td>(Vega et al., 2005)</td>
<td>0.998</td>
<td></td>
<td></td>
<td>232</td>
<td>252.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TIP4P/2005</td>
<td>(Abascal and Vega, 2005)</td>
<td>0.9979</td>
<td>18.9</td>
<td>10.89</td>
<td>60</td>
<td>2.08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPC/E</td>
<td>(Spoel et al., 1998)</td>
<td>1.008</td>
<td></td>
<td></td>
<td>81</td>
<td>2.7</td>
<td>-47.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPC/E</td>
<td>(Mark and Nilsson, 2001)</td>
<td>0.998</td>
<td></td>
<td></td>
<td>2.8</td>
<td>-45.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPC/E</td>
<td>(Vega et al., 2005)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPC/E</td>
<td>(Rami Reddy and Berkowitz, 1989)</td>
<td></td>
<td></td>
<td></td>
<td>70.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expt</td>
<td>(Schmidt, 1969)</td>
<td>0.9965</td>
<td>18</td>
<td>10.51</td>
<td>78.5</td>
<td>2.27</td>
<td>-41.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>(Dorsey, 1940)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>(Dorsey, 1940)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>(Weast, 1977)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>(Krynicki et al., 1978)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>(Postma, 1985)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.3: Physical properties of the different water models. The second and successive columns represent the values for the density \(\rho\), the heat capacity \(C_p\), the heat of vaporization \(\Delta H\), the dielectric constant \(\varepsilon\), the self-diffusion coefficient \(D\), the potential energy \(E_{pot}\), the melting temp of ice \(I_h\), and the temperature of max density \(T_{md}\).
Table 2.4: Oxygen-oxygen pair distribution functions for different water models. $r_1$ and $r_2$ are the position of the first and second peak of the distribution function, and $g(r_1)$ and $g(r_2)$ are the height of the peaks.

Mark and Nilsson (2001) performed molecular dynamics simulations using the three different water models, i.e. SPC, SPC/E and TIP3P, and compared the resulting structures with the experiment of Soper and Phillips (1986) and Mills (1973). They found that of these three models, SPC/E gives the best fit to the experiment overall: it gives the best bulk water dynamics and structure, with the closest agreement with experiment for the OW-OW interaction (Soper and Phillips, 1986). All three water models gave a higher diffusion coefficient than the experimental value (Mills, 1973), but SPC/E is the closest of the three. The potential energy, however, is best reproduced by the TIP3P and SPC water models. Guillot wrote a review on the different models of water and compare them to experimental values (Guillot, 2002). He found that density, diffusivity, critical temperature and structure were well described by the SPC/E parameters. For these reasons, the SPC/E water model and standard combining rules were used to describe the interactions between water molecules and glutamic acid molecules.

In the SPC/E model, the OW-HW bond length is fixed at 1 Å and the HW-HW distance at 1.63298 Å, so that the HW-OW-HW angle is 109°. The charges used are described in
Table 2.5. The melting temperature of ice using SPC/E water is 215K (Vega et al., 2005).

<table>
<thead>
<tr>
<th>atom names</th>
<th>$\sigma$ (Å)</th>
<th>$\varepsilon$ (kcal.mol$^{-1}$)</th>
<th>$q$ zwitterion ($e$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OW (SPC/E)</td>
<td>3.166</td>
<td>0.155</td>
<td>-0.848</td>
</tr>
<tr>
<td>HW (SPC/E)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.424</td>
</tr>
</tbody>
</table>

Table 2.5: Lennard-Jones parameters for the SPC/E water model

In order to verify the accuracy of the water model and the correct set up of the simulations, pure water was simulated using a simulation box of 512 SPC/E water molecules. The RDF calculated from this simulation was compared to data obtained from neutron diffraction experiment (Soper and Phillips, 1986). The RDF from the simulation compares favourably with the experimental data (Figure 2.4 and Table 2.4) and confirms that SPC/E is a very acceptable model to use.

Figure 2.4: Pure water RDF using SPC/E water model (dashed line) compared to experiment (plain line) (Soper and Phillips, 1986)
2.3.5 Charges

The electrostatic potentials described by Equation 2.13 require the electronic point charges for each atom present. For the water molecules and small ions, the point charges have been provided in Table 2.2 and Table 2.5. For the glutamic acid molecule, however, the AMBER force field does not provide electronic charges for single molecules of the amino acid but gives the charges for amino acids included in a peptide chain, which is not the case in the work performed in this thesis. A study of the -1 isoelectric form of glutamic acid provides a set of charges (Jorgensen and Tirado-Rives, 1988) which includes a common set of charges for both deprotonated carboxylic acid ends, and a global charge for the methylene groups without considering the carbon and hydrogen atoms separately. Study of propionic acid, which is the side chain of glutamic acid (SanRoman-Zimbron and Ortega-Blake, 1997), provides explicit charges for each atom. These two charge sets were then combined to give a complete set of charges for the zwitterionic form of glutamic acid with a global neutral charge. The charges of the CH\textsubscript{2} and COOH groups of propionic acid were applied to the side chain of glutamic acid (CH\textsubscript{2}-CH\textsubscript{2}-COOH). The common part of the amino acid charges of the -1 form were applied to the COO\textsuperscript{-}, NH\textsubscript{3}\textsuperscript{+} and CH groups. The resulting global charge of the molecule being -0.038, that number was divided by six and the result subtracted equally to the three atoms on each of the methylene groups, and the numbers rounded so that to obtain an exact global charge of zero. The resulting set of charges for the glutamic acid molecule can be found in Table 2.1.

2.4 DL_POLY

DL_POLY (Smith et al., 2002) is a molecular dynamics simulation package for which the open source code is provided free under license to academic institutions. DL_POLY can be use to study a variety of systems, which include biomolecules (Chau, 2006), glasses (Palin et al., 2002), interfaces (Bordat et al., 2006; Martin et al., 2006; Spagnoli et al., 2006), liquids (Pereira et al., 2001; Zoranic et al., 2007) and sol-gel solutions (Pereira et al., 2002). It can be run either on a single processor or in parallel over a number of processors.

2.4.1 Input Files

In order to run a successful simulation on DL_POLY (Smith et al., 2002), the correct data must be provided. These data are in the form of 'input' files, which consist of three text files: the ‘CONFIG’ file, the ‘CONTROL’ file and the ‘FIELD’ file.
2.4.1.1 **CONFIG file**

The CONFIG file is a text file that lists the coordinates in three dimensions of all the atoms present in the simulation box. It also gives the coordinates of the simulation box, and a code describing the periodicity of the simulation box. An annotated example of a CONFIG file can be found in Appendix Error! Reference source not found.

2.4.1.2 **FIELD file**

All the force field parameters described in sections 2.2.2 are specified in the FIELD file. The FIELD file is an input file describing the number of molecules, the number of atoms per molecule, the molecular mass and mass of each atom. The name of the atoms must match the atom names and appear in the same order as those written in the CONFIG file. It then lists the bond, angle and dihedral interactions for all the bonds and angles present in the system. It continues by describing all the van der Waals terms between all possible combinations of atom pairs. See Appendix Error! Reference source not found. for a more detailed description of the file organisation.

2.4.1.3 **CONTROL file**

The CONTROL file is the command file: it lists all the protocols to be followed in a particular simulation:

(i) **Timestep**: length of a single simulation step in picoseconds (ps), usually \(10^{-3}\) ps.

(ii) **Steps**: number of timesteps the simulation is to be run for.

(iii) **Ensembles.** In this work the NVT and NPT ensemble have been used. NVT is a simulation run with a constant Number of molecules, at constant Volume and constant Temperature. The NPT ensemble uses a constant Number of molecules, constant Pressure and constant Temperature.

(iv) **Temperature**: temperature at which the simulation needs to be run. The temperature is regulated by a thermostat, which is a heat bath coupled to the system. It is an algorithm that periodically scales the kinetic energy. This period is set by the user in order to obtain the desired average temperature. The Nosé-Hoover (Nose, 1984; Hoover, 1985) algorithm was used in the work described in this thesis, unless otherwise stated. The relaxation constant was set at 0.5ps.

(v) **Pressure**: sets the pressure for constant pressure ensemble (NPT). The pressure is regulated by a barostat, which adds a friction coefficient to the equation of
motion in order to obtain the desired average pressure by adjusting the size and shape of the simulation cell. It uses the Melchionna modification of the Nosé-Hoover algorithm (Melchionna et al., 1993) in which the equation of motion involves a Nosé-Hoover thermostat and barostat. The relaxation time for the barostat was set at 1 ps and the pressure at 1 bar for all constant pressure simulations described in this thesis.

(vi) Equilibration: the number of timesteps for which the simulation needs to be equilibrated. During the equilibration period, the data will not be averaged, and the forces will be capped to a maximum, so that the energy of the system does not become too large.

(vii) Restart: command to continue a previous simulation by using the forces and velocities of the atoms from the last timestep specified in the REVCON file.

(viii) Cut-off: a cut-off has to be specified of no more than half the cell size. It determines the separation distance to which the Coulombic forces will be calculated in real space, and also to which distance the Lennard-Jones parameters and the RDF are to be calculated.

(ix) Smooth Particle Mesh Ewald: the user must specify $\kappa$, the Ewald convergence parameter in Å$^{-1}$, $k_1$, $k_2$ and $k_3$, the maximum k-vector index in the x, y and z directions, respectively.

(x) Output selection: one can specify the starting point and the frequency at which a configuration of the whole system needs to be saved in the trajectory file, and also the frequency at which it must write to the output file. A command line specifies if the RDF and the z-density profile need to be written in the appropriate files.

(xi) Shake algorithm: at the end of every timestep, DL_POLY lists all the atoms bonded by constraint, and compares the actual value to the list in the force field. If they do not match, DL_POLY applies a correction vector to the coordinates of the atoms concerned and compares the new length with what is in the force field. The iteration process continues until the coordinates converge (Ryckaert et al., 1977).

An example of a CONTROL file has been added to the Appendix Error! Reference source not found.
2.4.2 Output files
While the simulation is running, DL_POLY saves a record of the important information and parameters in several output files that can be used for subsequent analysis.

2.4.2.1 OUTPUT file
The OUTPUT file is a summary of the simulation and consists of a summary of the simulation specification (from the CONTROL file), the force field specification (from the FIELD file) and a sample of the starting configuration. Secondly it records the simulation progress, where DL_POLY details all the instantaneous energies, temperature, volume, time etc, at a frequency set by the CONTROL file. It also calculates and writes the averages of these data after the equilibration period is completed.

2.4.2.2 HISTORY file
The HISTORY file saves a complete configuration: it writes the coordinates, velocities and forces at a frequency dictated by the CONTROL file. This HISTORY file can be analysed once the simulation is complete.

2.4.2.3 REVCON file
The REVCON file is the configuration of the system as it is when the simulation run finishes. It can be used as a CONFIG file for the next simulation if a restart is needed. It saves the coordinates, velocities and forces of each atom present in the system.

2.4.2.4 REVIVE file
The REVIVE file contains the statistical data necessary for a restart run. It pairs with the REVCON file to achieve a successful restart of the simulation. It is unformatted, therefore impossible to read by the user.

2.4.2.5 STATIS file
The STATIS file is a statistical history file. DL_POLY appends the file at a frequency determined by the CONTROL file. The STATIS file is a list of all the energies found in the output file, and also the mean square displacement, the stress tensor in the x, y and z direction. The values are not tabulated and therefore harder to read compared to the OUTPUT file.
2.4.2.6 RDFDAT File
The radial distribution function (RDF) is saved to a separate file. It gives the pair correlation function for each atom pair possible following the atom type given in the FIELD file. The RDF data is used for the analysis of the simulation (see section 2.5.1).

2.5 Analysis
Once the simulation is complete, the output files are analysed in order to give visual results. The analysis can be carried out using different methods: it can be from simply plotting data, using a code supplied with DL_POLY, or by constructing a program that will carry out calculations on the trajectory file.

2.5.1 RDF & Coordination Number
The distribution function of a liquid phase is an important tool for studying the structure of that liquid. The radial distribution function (RDF), which describes the relative probability of the presence of a specified species at a radius $r$ around the atom of concern is particularly important. Figure 2.5 illustrates the case where three atoms are located at a distance $r$ from the atom of concern, in this case the central atom. The RDF is the Fourier transform of the structure factor obtained from X-ray diffraction and neutron scattering experiments. The probability distribution function $g_{ij}(r)$ can be written (Rossky and Karplus, 1979):

$$g_{ij}(r) = \frac{1}{4\pi \rho_i r^2} \frac{dN_{ij}(r)}{dr}$$  \hspace{1cm} (2.22)

with $N_{ij}(r)$ the average number of $i$ atoms within a sphere of radius $r$ around atom $j$, and $\rho_i$ the density of $i$ atoms in the bulk fluid. The factor $(4\pi \rho_i r^2)^{-1}$ is to ensure that $g_{ij}(r)$ is normalized to one at positions in the bulk far from atom $j$.  

47
An RDF features the probability relative to the bulk of finding an atom at a distance $r$. It is then simple to compare the shape, height and position of the peaks with other RDF from the literature (Mark and Nilsson, 2001), obtained either by molecular dynamics or experimentally (Soper and Phillips, 1986). DL_POLY uses bin sizes of 0.05 Å to the cut-off distance specified in the CONTROL file.

The coordination number represents the number of atoms present around the atom of interest. It can be obtained through integration of the RDF following Equation (2.23), where $n^\beta_\alpha (r)$ is the average coordination number of $\beta$ atoms around $\alpha$ atoms, $\rho$ is the atomic number density and $C_\beta$ is the concentration of atoms $\beta$. By integrating up to the first minimum, $r_{min}$, the number of $\beta$ atoms surrounding $\alpha$ for the first solvation layer can be calculated.

$$n^\beta_\alpha (r) = 4\pi C_\beta \rho \int_0^{r_{min}} g_{\alpha\beta}(r)r^2dr$$  \hspace{1cm} (2.23)

### 2.5.2 Torsion angles

In order to monitor change in the geometry of a glutamic acid molecule, three torsion angles (dihedrals) were examined for each saved configuration using the code "torsion" (see Appendix Error! Reference source not found.). The three dihedrals were labelled T1, T2 and T3. They were chosen so that the backbone of the molecule was included and also because these dihedrals were also reported in the literature for the glutamic acid crystal structure (Hirokawa, 1955; Bernstein, 1991; Fletcher et al., 1996). The
definitions of these torsion angle are $T_1 = N_3-C_3-C_4-C_5$, $T_2 = C_1-C_3-C_4-C_5$ and $T_3 = C_3-C_4-C_5-C_2$ (see Figure 2.3 for the description of these atoms). In the $\alpha$ crystal, the values of $T_1$, $T_2$ and $T_3$ are 178.4°, 59.1° and 68.8° (Hirayama et al., 1980; Bernstein, 1991; Fletcher et al., 1996), and for the $\beta$ crystal they are -51.0°, -171° and -74.2° (Bernstein, 1991; Fletcher et al., 1996), respectively (Figure 2.6).

![Figure 2.6: the three torsion angles monitored as they appear in the a) $\alpha$ and b) $\beta$ crystals](image)

The calculation of a torsion angle has to be broken down into two steps: first, the calculation of the equation of the planes involved in the dihedral, and then the calculation of the angle formed by these two planes.

Assuming that three points in space $(x_1, y_1, z_1)$, $(x_2, y_2, z_2)$ and $(x_3, y_3, z_3)$ define a plane, the equation of that plane would be: $Ax + By + Cz + D = 0$. $A$, $B$, $C$ and $D$ are given by the determinants of the following matrices:

$$
A = \begin{vmatrix} 1 & y_1 & z_1 \\ 1 & y_2 & z_2 \\ 1 & y_3 & z_3 \end{vmatrix}, \\
B = \begin{vmatrix} x_1 & 1 & z_1 \\ x_2 & 1 & z_2 \\ x_3 & 1 & z_3 \end{vmatrix}, \\
C = \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}, \\
D = \begin{vmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{vmatrix}
$$

Which can be expanded and give:

$A = y_1(z_2-z_3)+y_2(z_3-z_1)+y_3(z_1-z_2)$

$B = z_1(x_2-x_3)+z_2(x_3-x_1)+z_3(x_1-x_2)$

$C = x_1(y_2-y_3)+x_2(y_3-y_1)+x_3(y_1-y_2)$

$D = x_1 (y_2 z_3 - y_3 z_2) + x_2 (y_3 z_1 - y_1 z_3) + x_3 (y_1 z_2 - y_2 z_1)$

49
The sign of \( s = Ax + By + Cz + D \) determines which side the point \((x,y,z)\) lies with respect to the plane. If \( s > 0 \) then the point lies on the same side as the normal \((A,B,C)\). If \( s < 0 \) then it lies on the opposite side, if \( s = 0 \) then the point \((x,y,z)\) lies on the plane. This sign will then determine the sign of the torsion angle.

The dihedral \( \theta \) formed by two planes \( n_1 \) and \( n_2 \) of the equations: \( A_1x + B_1y + C_1z + D_1 = 0 \) and \( A_2x + B_2y + C_2z + D_2 = 0 \), respectively, is calculated by the dot product of the normal vector of each plane \( n_1 \) and \( n_2 \) (Gellert, 1989):

\[
\cos \theta = \hat{n}_1 \cdot \hat{n}_2 = \frac{A_1 A_2 + B_1 B_2 + C_1 C_2}{\sqrt{A_1^2 + B_1^2 + C_1^2} \sqrt{A_2^2 + B_2^2 + C_2^2}} \tag{2.24}
\]

### 2.5.3 Cluster counting
When more than one glutamic acid molecule is present in the simulation box, the solute molecules can interact with each other in such a way that they stay close to each other to form clusters. In this work, two molecules are considered clustered if the distance between two of their atoms is less than a cut off distance. In the case of glutamic acid, two variables can be altered in the process of determining clusters: the choice of the atoms involved, and the choice of the cut-off distance. In order to differentiate different cluster types, three distinct distances were considered: \( C_1-C_1 \), \( C_2-C_2 \) and \( C_4-C_4 \). These three distances were calculated post simulation and molecules were regarded as being a cluster if two or more molecules are located at a distance similar to that of the intramolecular distances of the crystals, allowing a 10% margin. The sizes of all the clusters, if any, were recorded for each saved configuration, as well as the molecules involved in each cluster (see code "cluster" in Appendix Error! Reference source not found.).

### 2.5.4 Density Profile
The density profile, also called z-density, is the number of molecules per unit volume averaged over the x and y directions. DL_POLY uses the same number of bin than previously for the RDF, which is equal to the cut off distance divided by 0.05. The bin width for the density profile is then calculated by dividing the cell dimension in the z direction over the number of bins. The software calculates the number of molecule per
bin and accumulates the data into a histogram that gives the number of atoms per Å³ as a function of the distance along the z dimension.