

Chapter 1

Introduction

1.1 Importance of Water

Water [3] is the liquid of which seas, lakes and rivers are composed, and which falls as rain and issues from springs. It is raised from the ocean, diffused through the air and poured down upon the soil. It is one of the four ancient 'elements'. In fact, throughout the universe, no other substance has ever received so much attention [4, 5], socially, scientifically, technologically, environmentally. For its widespread occurrence in nature, its unique role in the life sciences, its importance in the chemical industries, and its inestimable influence on human activity, water occupies an irreplaceable position in man's history.

Water is the most abundant compound on earth, existing naturally in the forms of vapor, liquid and solid. Seventy percent of the surface of the planet is covered by oceans. It can be estimated that of the approximately 5.7 thousand million persons on earth there are 60 thousand million gallons of water available for each person, seemingly more than enough to go around.

Living tissue is composed mainly of water. Cells, organs, and organisms are constantly bathed in an aqueous environment. Without water, many chemical reactions could not take place, biological systems would not function and life as we know it would not have originated. However, the detailed, molecular-level role that water plays in the function of a chemical or biological system is still a mystery and a matter of great scientific debate. One of the central problems in biochemistry is to understand the structure and function of proteins at a molecular level [6-9]. Many enzymes catalyze biochemical reactions through water molecules bound to their active sites. Protein synthesis also requires the presence of an ordered water layer. In fact, more and more, it is being recognized [10-14] that the explicit inclusion of water is absolutely essential in the unveiling of many problems in biology and biochemistry. Thus, it would be expected that without a fundamental and detailed knowledge of water, many phenomena in molecular biology would be difficult to understand.

Water is often called the universal solvent [15]. The structure and dynamics of water constitute a cornerstone for understanding all aspects of solution chemistry and electrochemistry [16–18]. A large variety of chemical reactions take place in water or in aqueous mixed solvents. These aqueous media often display amazing solvent effects on the rate and thermodynamic activation parameters of organic reactions such as nucleophilic displacement reactions, acid-base catalyzed processes, *etc.* [19]. The interaction of water with clays, surfactants, and metal surfaces is important in oil recovery, mining, catalysis, corrosion inhibition, and other technologically important concerns [4]. A complete understanding of these mechanisms relies on our knowledge of water.

Water is also a commonly employed liquid for generating a *supercontinuum*, the production of an intense ultrafast broadband ‘white light’ pulse [20]. Self-focusing, self-defocusing, and supercontinuum production are non-linear optical phenomena originating when a strong laser beam propagates through a medium, inducing an intensity dependent change of refractive index. In order to gain an insight into the non-linear response, an understanding of water is essential. Probing into the essence of water, therefore, has a fundamental utility, not only in many branches of pure science, but also in technology, industry and medicine.

Water is not a simple liquid. It possesses strongly polar hydrogen bonds which are responsible for a striking set of physical and chemical properties [2, 4, 5, 21–25]. For instance, the molar volume at 1 atm pressure and 0°C temperature decreases from 19.66 cm³ for ice to 18.0182 cm³ for liquid water. Such contraction on melting is relatively rare among substances. As the liquid is further heated, it continues to contract until a density maximum is achieved at 3.984°C [26]. This anomalous phenomenon disappears at a sufficiently high pressure, where the temperature dependence of the volume becomes ‘normal’. Water at not too high a pressure possesses a negative pressure coefficient of viscosity for temperatures below about 20°C [27]. This means that *increasing* the pressure gives rise to a *lower* viscosity, clearly an anti-intuitive behavior. Its isothermal compressibility declines under atmospheric pressure with increasing temperature above the melting point and reaches a minimum near 46.5°C [28]. The heat capacity of water is much higher than the value expected through contributions from various degrees of freedom using ordinary considerations. Unlike other substances having comparable molecular weight, such as nitrogen, methane, ammonia, hydrogen sulfide *etc.*, water exhibits anomalously high melting, boiling and critical temperatures. Besides these unusual thermal properties, water has an anomalously large dielectric permittivity and a high surface tension. Some of these properties have recently been described in a review by Kunz, *et al.* [29], though most have been known since the time of Röntgen [30] over a century ago. No other molecular liquid with such abnormal properties is known. In fact, according to Luck [31], ‘Life would not be possible on our planet without these anomalous properties of water.’ Supporting Luck’s point of view, a much earlier and very detailed description of how the various properties of water

make it uniquely suitable for supporting biological life was published in 1913 by Henderson [32]. This marvelous 317 page book devotes more than half its space to the way that water and its unique properties — the high heat capacity and latent heat, as well as the density peculiarities, causing ice to float on water and giving rise to the erosion by ice formation in crevices, to name a few — have been essential to the development and sustenance of life forms on earth. Henderson's book is still well worth reading more than eighty years after its publication. Up to now, however, none of the strange properties of water has had a well-accepted explanation. Liquid water remains, therefore, the most important unsolved problem in science.

There exist at least ten ice polymorphs [2, 24, 33]: hexagonal ice Ih formed from freezing the liquid at atmospheric pressure, cubic ice Ic formed by vapor phase condensation at low temperature, and a series of more densely packed crystal structures observed under elevated pressures. There are also states of amorphous water (AW), one [34–37], termed l_{da}, LDAW in this book, has a density close to that of ice Ih, while the other, a high density amorphous form [37–40], called h_{da} or HDAW, has a density close to that of some of the high density crystalline forms. These variants indicate that very complicated, often underrated, molecular interactions exist in water.

1.2 Means of Theoretical Study

Scientists, at least since the time of Whiting, in 1884, as described in Chadwell's 1927 review [41], and Röntgen [30], have put forward hypotheses aimed at explaining some of the anomalous properties of water. However, efforts to verify or invalidate these have been hampered by the lack of an accurate theory characterizing all aspects of this substance. All too often also, modern scientific methods, both theoretical and experimental, because of their narrow and specialized form, are not easy to apply in concert to the wide symphony of problems that water presents.

There are essentially two routes along which the theoretical study of water has been carried out. Purely analytical approaches have the advantage of providing a mathematical interpretation of the phenomena studied, sometimes providing formulas that can be used by experimentalists to interpret their data. Such descriptions may be adequate for certain simple liquids. When applied to water, however, they have been, in general, not very helpful, at most being capable of suggesting only a rough, qualitative picture.

More promising are computer simulations, which can be viewed as a sort of experiment. These methods provide a direct route for converting the microscopic or molecular-level information for a system to macroscopic properties of interest and serve as an invaluable complement not only to the analytical or statistical mechanical investigations, but also to laboratory experiments. There are actually five commonly used computer simulation procedures.

1. *Molecular dynamics*, which generates configurations of a system as a function of time by integrating Newton's equations of motion.
2. *Monte Carlo*, which generates configurations of a system by a random sequence of numbers followed by logical operations to accord with the chosen statistical ensemble.
3. *Stochastic dynamics*, which generates trajectories by integrating the stochastic Langevin equations of motion.
4. *Energy minimization*, which searches for a minimum energy configuration by moving along the gradient of the interaction potential through configuration space.
5. *Normal mode analysis*, in which molecular motions are described as a superposition of harmonic vibrations with characteristics determined by the shape of the interaction potential surface near an energy minimum.

Once the intermolecular potential is determined from a reasonable molecular model, one is able to extract useful information from one of these methods, or a combination of them, depending on the problem to be solved.

Normal mode analysis has the advantage that, once the vibrational modes are determined, a number of properties can be computed by analytical procedures. However, this method may sometimes be prohibitive when macromolecules containing more than several hundred atoms are involved, as critically discussed by McCammon and Harvey in their book [8]. More recently, normal mode analysis is having somewhat of a rebirth, at least in the study of bulk liquid water, where very good vibrational density of states for the intermolecular vibrational modes have been obtained [42]. These *quenched* (QNMA) and *instantaneous* (INMA) normal mode analysis methods are briefly described in Section 6.6 of this book.

From a computational point of view, the energy minimization technique falls into the regime of nonlinear optimization problems. Mathematically, any continuous, differentiable function of an independent variable \mathbf{r} can be expanded in a Taylor series about a point \mathbf{r}_0

$$f(\mathbf{r}) = f(\mathbf{r}_0) + (\mathbf{r} - \mathbf{r}_0) \cdot \left(\frac{\partial f}{\partial \mathbf{r}} \right)_{\mathbf{r}_0} + \frac{(\mathbf{r} - \mathbf{r}_0)^2}{2} \cdot \left(\frac{\partial^2 f}{\partial \mathbf{r}^2} \right)_{\mathbf{r}_0} + \dots \quad (1.1)$$

Classification is made according to the highest order derivative considered, *e.g.*, zeroth order (grid search), first order (steepest descent or conjugate gradient), second order (Newton-Raphson), *etc.* This algorithm yields only one local minimum energy configuration that is generally not far from the initial one. It can never reach the global minimum for complex systems with hundreds of degrees of freedom having a large, but unknown, number of local minima. For a detailed description of this method refer to Ref. [43].

Application of the stochastic dynamics method normally takes considerably less computational effort than the molecular dynamics method. This method is suitable for systems consisting of a very large number of particles such as those involving biologically important molecules. In these simulations, the solvent molecules are not treated explicitly, but rather are treated in an average way by choosing an appropriate friction coefficient ξ . Their influence on the solute can be divided into two parts: a systematic frictional force and a randomly fluctuating force \mathcal{R} . In the Brownian motion limit, solute motions are determined by solutions of Langevin equations

$$\frac{d^2\mathbf{r}_i}{dt^2} = \frac{\mathbf{f}_i + \mathcal{R}}{m_i} - \xi \frac{d\mathbf{r}_i}{dt} \quad (1.2)$$

with \mathbf{f}_i being the external force on particle i of mass m_i . There are excellent sources that provide a detailed review of this subject [44, 45].

The Monte Carlo (MC) method uses random sampling. Interestingly, the use of random sampling methods to investigate a problem in classical mechanics, the equipartition law, is at least a century old, appearing, for example, in a famous 1901 paper [46] by Lord Kelvin, 'Nineteenth century clouds over the dynamical theory of heat and light'. The essence of the modern MC method is to rewrite any multidimensional integral of statistical mechanics

$$F = \int_{\mathbf{r}_1}^{\mathbf{r}_2} f(\mathbf{r}) d\mathbf{r} \quad (1.3)$$

in terms of

$$F = \int_{\mathbf{r}_1}^{\mathbf{r}_2} \frac{f(\mathbf{r})}{\rho(\mathbf{r})} \rho(\mathbf{r}) d\mathbf{r} \simeq \frac{1}{N} \sum_{i=1}^N f(\mathbf{r}_i) / \rho(\mathbf{r}_i), \quad (1.4)$$

where $\rho(\mathbf{r}_i)$ is chosen from a random probability distribution. One might think that a uniform distribution would be sufficient. However, it is known that the probability distributions of statistical ensembles are sharply peaked, so it is more efficient to choose a $\rho(\mathbf{r}_i)$ that corresponds to the function $f(\mathbf{r}_i)$. The technique of concentrating on points in space that make substantial contributions to the integral is called *importance sampling*, a concept introduced by Metropolis, *et al.* [47]. This task is required in order to ensure that the generated sequence of random states at the end of the simulation has the appropriate probability that is not known *a priori*. To implement this, Metropolis, *et al.* [47] set up a Markov chain such that, starting from an initial state Γ_0 , further states can be constructed

$$\Gamma = \{\Gamma_1, \Gamma_2, \dots\},$$

which ultimately follow the thermodynamic equilibrium distribution. This is guaranteed by comparing the transition probability $W(\Gamma_0, \Gamma)$ with a generated random number $R \in [0, 1]$ from a uniform probability distribution. If the transition probability W is less than the random number R , then Γ_0 is counted as a new state,

otherwise Γ is accepted. This process is repeated many times. In every sequence, the state Γ is chosen randomly. When a Monte Carlo simulation is applied to a molecular liquid, the Markov chain matrix consists of combined translational and rotational moves. Other references [44, 45] should be consulted for a more detailed discussion of Monte Carlo methods.

Unlike the MC approach, which is essentially a probabilistic form of computer simulations, molecular dynamics (MD) is a deterministic approach. A fuller description of the molecular dynamics method is given in Chapter 2 of this book. While MC sampling is powerful for analyzing static properties of complex systems at equilibrium, it provides no dynamical information and cannot be applied to nonequilibrium molecular systems. On the other hand, a MD simulation gives results for dynamic properties, and nonequilibrium molecular dynamics (NEMD) is a well established area of study [48, 49].

Among the computer simulation techniques, Monte Carlo and molecular dynamics have provided the most popular and most useful approaches for bulk water. They contain the least number of approximations, can be applied to a variety of experimental conditions, are readily adaptable to many statistical ensembles, and provide detailed molecular-level insights about the liquid structure and other properties that cannot be extracted by other methods.

The most apparent disadvantage of computer simulation approaches is the limitation on sample size. Generally, the simulation cell contains no more than several hundred to several thousand particles. When biological molecules or other complex solutes or interfaces are bathed in a water or water-like solvent [50], an even larger system is often required. With futuristic computer hardware and sophisticated software, it will soon be possible to extend the sample size to millions of particles. Of course, any successful simulation technique requires the development of reasonable potential models, and, in the case of water, these models must intrinsically be rather complicated, greatly reducing the sample sizes that can ordinarily be employed. It is directly because of these computational limitations that water-water potential models have never been properly optimized in past work.

1.3 Historical Review

A most interesting historical review of 'Computer simulations in statistical mechanics' has been presented from a first-hand perspective by W. W. Wood [51]. Here, we can only briefly provide some of the high points of this history, primarily as they concern water.

Over forty years have passed since the first Monte Carlo computer simulation of a liquid was performed at the Los Alamos National Laboratories in the United States [47]. The method itself, in its modern form, was developed during the late 1940's [52, 53]. The first representations of molecular liquids were highly idealized within the framework of hard disks or hard spheres, until in 1957 the Lennard-Jones potential

was introduced into MC simulations by Wood and Parker [54]. This improvement in molecular modeling opened the way for comparing results generated from computer simulations with actual laboratory data for liquids consisting of atomic particles or near spherically symmetric molecules. The first successful attempt to carry out a molecular dynamics simulation was accomplished by Alder and Wainwright [55, 56] for a hard sphere system in which the particles move at constant velocities until perfectly elastic collisions take place. This achievement was later extended to MD systems containing Lennard-Jones particles [57, 58] and to diatomic molecules [59]. An extensive review of MD computations on nonaqueous liquid systems has recently been published by Kessler, *et al.* [60].

Even though the early MD computations were purely classical, as are most current ones, a good representation of the P, Q and R branches ($\Delta J = -1, 0, +1$) in the pure rotational spectrum of a rigid diatomic molecule in a liquid argon bath, and quantum-mechanical-appearing selection rules for Raman and infrared transitions, could be obtained through the appropriate Fourier transforms [61, 62]. This, of course, is a consequence of the conservation of angular momentum. This old result, in addition to more recent ones [63, 64], calls into question statements in the literature [65, 66] that blame some of the deficiencies of classical computational results on the absence of quantum effects. In fact, Bader and Berne [67] have recently shown that the purely classical and the purely quantum mechanical descriptions of a harmonic system, but not a mixed treatment, give equivalent results, not only for the correlation functions, but also for the absorption band shapes. These authors conclude their paper by stating that, even when anharmonicities are present, it is likely that a purely classical study of vibrational spectra and relaxation will give an accurate representation of the quantum mechanical result. While quantum effects or at least quasi-quantum effects are certainly important and necessary for phenomena such as proton transfer or electron transfer [68–74] aided by tunneling, or certain discrete-level vibrational relaxation events in the liquid [75], the cause of many remaining flaws in classical MD calculations of water very likely still arise from deficiencies in the choice of intermolecular potentials [76–79].

Though written earlier than any Monte Carlo or molecular dynamics computation on water, the final chapter of the book by Eisenberg and Kauzmann [2] gives a good comparison of most early theoretical liquid water models. Full computer simulations on liquid water were initiated in 1969 by Barker and Watts [80] using a MC approach. This study was closely followed by the celebrated MD work of Rahman and Stillinger [81]. Both of these pioneering simulations used rigid five-site nonpolarizable models for the water molecules. These studies as well as other early computational efforts on water focused attention on pairwise additive interactions, treating the effects of the condensed phase with a *mean field approximation* that brings in the concept of *effective potentials* [82–99].

Further progress in these directions relied on the development of more realistic water models. For instance, the central force (CF-F) models [100, 101] provided the

first example in which intramolecular flexibility in the water molecules was explicitly taken into account. In intramolecularly flexible models, the O-H bond lengths and H-O-H bond angle are allowed to vary in the vicinity of their equilibrium values, following harmonic or preferably anharmonic laws. With these flexible models [63, 102, 103], it is possible to investigate vibrational frequency shifts and geometrical deformations of the water molecule in condensed phases. Such data, compared with other types, are relatively straightforward to obtain experimentally and to interpret. They are most interesting from the point of view of effects of the surroundings, not only in aqueous solutions and near interfaces of chemical and biological importance, but also in bulk water because the fluctuating perturbations around any given water molecule can shed light on the forces present. For these reasons, flexibly bonded models of water in computational studies will certainly gain more attention in the future as computers become more powerful.

The first computer simulation of a liquid water-type substance having non-pair-additive effects caused by the 'molecular' polarizability was performed on a one-site model by Barnes and coworkers [104]. Since then, there has been an increased interest in developing new polarizable water models [105–108]. The most recent progress [64, 106, 109, 110] has concerned itself with both intramolecular bond flexibility and rapidly responsive electronic polarization. These more advanced models are of particular importance for studying water in perturbed states or over extended thermodynamic ranges, where the built-in rigidity and thus the lack of an appropriate molecular response in the earlier models can give misleading results.

Past computer simulations on perturbed liquids other than water have been very numerous [60]. A few of these are mentioned here: binary mixtures [111, 112], liquid/vapor interfaces [113–123], fluids in narrow channels [124–137], chemical reactions in condensed phases [138–146], a nonpolar liquid perturbed by a polar solute [147], and molecular liquids in intense external fields [148–150]. The systems studied for water are more varied. For instance, the general problem of the crystal-liquid interface has been addressed by Laird and Haymet [151]. Articles regarding MD simulations of aqueous electrolyte solutions [152–154] have been described frequently. A comprehensive description of both bulk and field-perturbed water with a summary of relevant work up to 1993 has been reviewed by Zhu and coworkers [77].

As computers grow more powerful, especially as supercomputers and massively parallel data processors are developed, more realistic MD studies of larger systems, including biologically interesting aqueous solutions of macromolecules, become feasible [155–157]. Pioneered by using a small globular protein, namely, bovine pancreatic trypsin inhibitor (BPTI) [158–160], the early simulations in this area treated the isolated biomolecule without explicit solvent surroundings. Later modifications included water and neighboring proteins, as might be found in a real system [161]. As the most powerful and reliable means for studying the behavior of a many-particle system, MD simulations on liquid water and aqueous chemical and

biological systems will progress rapidly in the next decade because of improvements in the following aspects:

1. More precise non-pair-additive interaction models that include both intramolecular bond flexibility and polarization.
2. More careful attention [162] paid to the non-hydrogen-bonded interactions that are prominent in certain dense forms of ice.
3. Larger sample sizes leading to a more appropriate treatment of long-range electrostatic forces and boundary conditions that are particularly important for dielectric characteristics and long-range collective phenomena.
4. Quantum effects for all systems, particularly those involving proton tunneling or electronic distortions, such as occur in the presence of strong laser fields.

1.4 Scope of This Book

This book is intended to present a relatively complete description of all molecular-level aspects of water. Background material is provided on various theoretical methods, including molecular dynamics (MD), Monte Carlo (MC) and statistical mechanics. Experimental and analytical studies of water are discussed as well.

There has been an extensive literature aimed at providing a thorough overview of computational studies of many-body problems [44, 45, 155, 163–166]. These techniques as they apply to biological systems have also been reviewed [8, 167–169]. However, except for a relatively brief review article [77], there has been no recent detailed survey concentrating specifically on water. We hope that this book will partially fill the gap between the rapidly developing techniques and the relatively few reviews.

Following this Introduction, the remainder of the book is structured into four general parts: Part One, which is composed of Chapter 2 and Chapter 3, is devoted to a description of both computational and purely theoretical methods for many-body systems from the microscopic and macroscopic points of view. A description mainly of bulk phase liquid water is the task of Part Two, which is composed of Chapters 4, 5 and 6. Whereas Chapter 4 summarizes useful data obtained from experiments, not only for the bulk phase, but also for water at certain interfaces, the motive of Chapters 5 and 6 is to proceed towards a molecular-level understanding of water. Chapter 5 covers the basic principles of theoretical or computational modeling of water, while Chapter 6 presents some representative computer simulation results for a variety of bulk water models and discusses the possibility of improving these models. The third part of the book, Chapters 7–10, concerns water in perturbed states. Applications of computational methods to these problems are described in this part. For example, Chapter 7 provides insights for aqueous

solutions, while Chapter 8 reviews recent progress in computer studies of water at macroscopic interfaces. Chapter 9 then describes applications of these methods to more complicated problems, such as biological systems, and Chapter 10 discusses molecular dynamics calculations used to describe water undergoing nonequilibrium events, such as adiabatic Couette flow, the behavior of the water solvent following a sudden change of the solute, and water in the presence of intense fields. One purpose of the fourth and last part of the book is to attempt to look down some future paths for water studies, both experimental and theoretical. Chapter 11, for example, briefly describes what the effect of massively parallel data processing might be on water MD studies over the next decade or so, while the last chapter of the book covers some almost forgotten history of water research from the deep past, then, for the near future, suggests some experiments and describes some modifications of existing computational models for water that seem to be needed for a better approach towards an ultimate resolution of this most important problem.