Enrico Clementi

Computer Simulations of Complex Chemical Systems: Solvation of DNA and Solvent Effects in Conformational Transitions

As is known, atomic and small molecular systems can be realistically simulated with quantum-mechanical models. In complex chemical systems, however, the natural parameters for a description are not only the electronic density and energy but also entropy, temperature, and time. We have considered as an example for a complex chemical system the structure of water surrounding DNA with counterions. No direct experimental determination of the solvent structure around a single DNA macromolecule is available from experimental data, despite continuous efforts during the last twenty years to obtain both single crystals of DNA and scattering data from solutions. A very detailed representation is now available concerning the structure and interaction energy of water molecules in the first solvation shell or in the "grooves" of the DNA. The data obtained by our computer simulation are in good agreement with indirect data from DNA fibers at different relative humidities and with other indirect evidence. In addition, our simulated results allow us to present a preliminary model for the solvent effects in transition processes between different DNA conformations. The model presented is also in agreement with available experimental data. Finally, our results report the first determination of the position of counterions in DNA at different relative humidities and at room temperature. This application demonstrates the flexibility of the computational approach. The method we have used is very general; namely, it is not limited to a biological system but is valid for any organic or inorganic problem requiring systematic simulation on a class of compounds interacting either as a few molecules or as a large ensemble of molecules.

Introduction

The understanding of natural phenomena is generally obtained by the combined use of experimental data and physical models, the latter usually being expressed in the form of mathematical relationships (equalities and/or inequalities). In an experiment, an assumed known and measurable perturbation is applied to a real system to obtain responses (or signals). The responses are either fitted to phenomenological models or to *ab initio* models, namely those models derived from basic principles and laws. In the former case we obtain an empirical understanding, in the latter an *ab initio* understanding.

More and more over the past thirty to forty years, experiments have been complemented, and at times (but seldom) replaced, by computer simulations. In the simulations a perturbation is applied to an ideal system, implicitly or explicitly defined by the models referred to above. The simulated response represents a "prediction" for a real system only in the measure that the physical model constitutes a realistic representation of the real system.

Any chemical system is characterized by its stability or instability relative to some initial condition. The evolution of a chemical system from initial to final conditions is most naturally represented in a time-space frame; that is, the dynamic characterization is basic to a realistic representation of any chemical system. However, valuable information on a chemical system can be obtained by neglecting the time evolution and by restricting oneself to a static description. Keeping this in mind, reasonable approximations, based on time-scale differences in some specific aspects of the time-dependent evolution, have been proposed. We recall the Born-Oppenheimer hypothesis as a specific example of approximation, and the ergotic hypothesis as a more fundamental relationship

Copyright 1981 by International Business Machines Corporation. Copying is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract may be used without further permission in computer-based and other information-service systems. Permission to *republish* other excerpts should be obtained from the Editor.

between static and dynamic representations. A saving factor for many time-independent approximations is likely to be found partly in the validity of the physical models, but mainly in the existence of a number of experimental techniques involving time averaging (for example, aspects of diffraction experiments) or reaching thermodynamic equilibrium. Under such conditions the static aspect of an intrinsically dynamic event is enhanced and measurable, and time-independent models are meaningful.

Depending on the complexity of a chemical system, namely, the number of degrees of freedom and assumptions on their separability, matter can be described as continuous or as a discrete ensemble of particles. In the latter case, whereas nuclei, electrons, and radiations are the *obvious* "particles" of the system, atoms and molecules (and chemical bonds) can be used as useful operational concepts, especially at the interface between the continuum and the discrete representation of a complex chemical system.

The degree of complexity of a chemical system acts as an operational and conceptual constraint in the formulation of a physical model; it determines the choice between discrete and continuous representations, the statistical description, and the concomitant selection for the equation of motion, classical or quantized. Theoretical chemistry attempts to provide an overlapping set of models to describe a chemical system at *any* degree of complexity. Computational chemistry, a much younger field, attempts both to provide operational techniques for solving such models and to test for the validity of the models by comparing simulated and experimental data.

In the following discussion we shall comment on a set of models valid for increasingly complex chemical systems. The set is characterized by the property that the output of one model constitutes the input of the successive model. Another characterization is that each model is particularly well adapted to represent one aspect of the total system, such as the electronic density of a molecule, the interaction energy between an atom and a molecule (and/or between molecules), the internal energy and free energy of the system at a given temperature T, the most stable structure of a single molecule at $T = 0$ K or the statistical distribution of atoms or molecules at $T \neq 0$ K, and finally the time-dependent aspect of the macrosystem considered as an ensemble of discrete particles or as a continuum.

Linked set of models to simulate complex chemical systems

In Model 1, point-charge nuclei and electrons are assumed to be the "particles" of the system to be described in a stationary state. As a consequence of the above assumptions, we are in the domain of Fermi-Dirac statistics and the appropriate equations of motion are assumed to be those of the time-independent Shrödinger equation [1]. Further, we decouple electronic from nuclear motions and, as a zero approximation, we use the independent one-particle model of Hartree-Fock [2] in the form proposed by C. C. J. Roothaan [3], which can be seen as the natural form for obtaining Mulliken's molecular orbitals via a linear combination of atomic orbitals [4]. Therefore, we assume as a valid approximation the time-independent, nonrelativistic motion of an independent electron immersed in the field due to fixed nuclei and the average field of the remaining electrons. One cannot avoid pointing out that we use a very approximate model. By closer analysis, however, we can identify two types of approximations in this one-electron model. The first (related to electronic density refinements) is *conceptually* rather trivial insofar as the approximation can be improved consistently within the framework of the model. The second type, however, is more basic, since the approximation cannot be removed without a basic alteration of the model's framework. Paradoxically, the first type at times can lead to quantitatively larger disagreement between simulated and experimental data than the second. For example, the inclusion of relativistic corrections to obtain the system's electronic energy can be achieved to a large degree simply by replacing one set of constraints (coupling the spins and the angular moments) with a different set. In this sense, part of the relativistic effects can be considered as an approximation of the first type. Equivalently, if in the Hartree-Fock method we place emphasis on the mathematical technique (self-consistency) rather than on the physical model (one-electron), the Hartree-Fock method leads most naturally to generalized self-consistent-field methods. The multiconfiguration self-consistent-field (MCSCF) model can include a very substantial fraction of the electronic correlation correction.

It has long been known that the Born-Oppenheimer approximation yields quantitatively acceptable binding energies for many molecular ground-state systems. Under the deceptively simple assumption of decoupling the electronic from the nuclear degrees of freedom, one imposes the constraint of limiting the validity of the model to vanishingly small nuclear velocities and to a *single* electronic potential (channel) experienced by the nuclei. These constraints are sufficiently strong to render unlikely a realistic representation of many fine aspects of chemical reactions on any model based on the Born-Oppenheimer approximation. For this reason, we consider the Born-Oppenheimer postulate as an approximation of the second type. Equivalently, the neglect of statistical

degeneracy, implicit in the computation of quantum-mechanical wave functions at nonzero temperatures, is an example of an approximation of the second type. Models based on approximations of the first type essentially assume as a starting point vibrational, rotational, and translational decoupling of the electronic and nuclear motions. Models based on approximations of the second type reintroduce such couplings from the start.

In the past twenty years much effort has been devoted to analyzing and extending the first type of model, which has been applied to atomic, molecular, and solid state systems. In this period, a number of advanced computer programs have also been prepared and widely used. We recall as examples the computer programs written by McLean and Yoshimine [5], by McLean, Yoshimine, Bagus, and Liu [6], by Clementi *et al.* for atomic [7] or molecular systems [8], and those by Nesbet [9] (for a review, see [10]). Since solids (like metals and crystals) can be described by imposing periodicity conditions on a molecular model, the development of atomic and molecular models is closely connected to the development of solid state models [10]. In this connection, it is of particular interest in the use of periodic boundary conditions to describe the "molecular" orbitals in polymers. We mention the pioneering work by Andre, partially performed at the IBM Research Laboratory in San Jose, California, his computer program, and some early applications [11]. Today these techniques are being used to simulate aspects of the conduction properties of either synthetic [12] or biological [13] polymers. Before passing to Model 2 we should mention that with these methods and computer programs, atomic, molecular, and solid state physics can be understood, explained, and partially predicted for observables corresponding to expectation values of well-defined single- or many-particle operators. However, this type of "physicists' knowledge" does not necessarily include "chemists' knowledge," just as the latter does not necessarily include "biologists' knowledge" and even less so "social scientists' knowledge." Recently attempts have been made to use Model 1 to rationalize and compute some of the nonobservables that are basic to the conceptual evolution of chemistry: for example, definition and derivation of chemical structural formulae from approximated Schrödinger's wave functions [14]. It is much too early to state whether chemistry will be a chapter of physics, or whether chemistry will contain a few chapters of physics. Possibly Dirac's hurried and enthusiastic viewpoint [15] will be limited, in the long run, to the second alternative.

A chemical system almost invariably is composed of many molecules confined in some volume at a given temperature and pressure. Model 1 is mainly designed to deal with one molecule at zero temperature; however, two or more molecules have been simulated [16]. Temperature effects can be included by considering the population of higher-than-zero vibrational states, but the model soon breaks down at a rather limited complexity level. A very superficial reason for the breakdown—at times presented as the "fundamental" reason—is to be found in the numerical and computational complexity and the associated cost. However, even assuming the availability of extremely fast computers at zero cost, a complex system at finite temperature is described not by one wave function but by a manifold of functions due to the high degeneracy of the system. In such cases a different model is needed. It is noted that only a few molecules are sufficient to reach this situation. We have shown, for example, that seven water molecules constitute a degenerate system [17].

The interactions among many molecules can be obtained by constructing interaction potentials. This phase represents Model 2. In principle, for n molecules, we need two-body, three-body, $\cdots$, n-body potentials, this being the full expansion for the total potential. In Model 2, the fundamental particles are no longer nuclei and electrons (clearly, these are present but at a deeper level, the Model-1 level), but are atoms and molecules. The energy of a given conformation of atoms and molecules is related to the system's temperature through the Boltzmann distribution law. The conformations are modeled as classical distributions of atoms coupled by intermolecular and intramolecular potentials. For not-too-high temperatures, at which a molecule would dissociate, the intramolecular potentials are at first assumed as single-value functions. For the case of multiple dissociation products, one needs multiple-value functions. In our model, the intermolecular potentials are not obtained by fitting experimental data (nor are the phenomenological-type potentials such as those of Lennard-Jones), but directly from Model 1. The two-body potentials are of the form of atom-atom potentials. The *dispersion* correction can easily be added to the two-body potentials when the latter are obtained in the one-electron approximation [18]. The *induction* correction is the main nonadditivity correction needed for three- or many-body systems if the latter are described by two-body potentials. This correction can easily be obtained by perturbation techniques [19]. In Model 2 it is very important to define atom-atom potentials that can be used to describe many different molecules; that is, the potentials must be "transferable" potentials. This task has been accomplished by an analysis that characterizes an "atom" when it is placed in a molecule [14]. This analysis is based on nonobservable quantities like chemical bonds (quantities that are arbitrary in the domain of physics but basic in the domain of

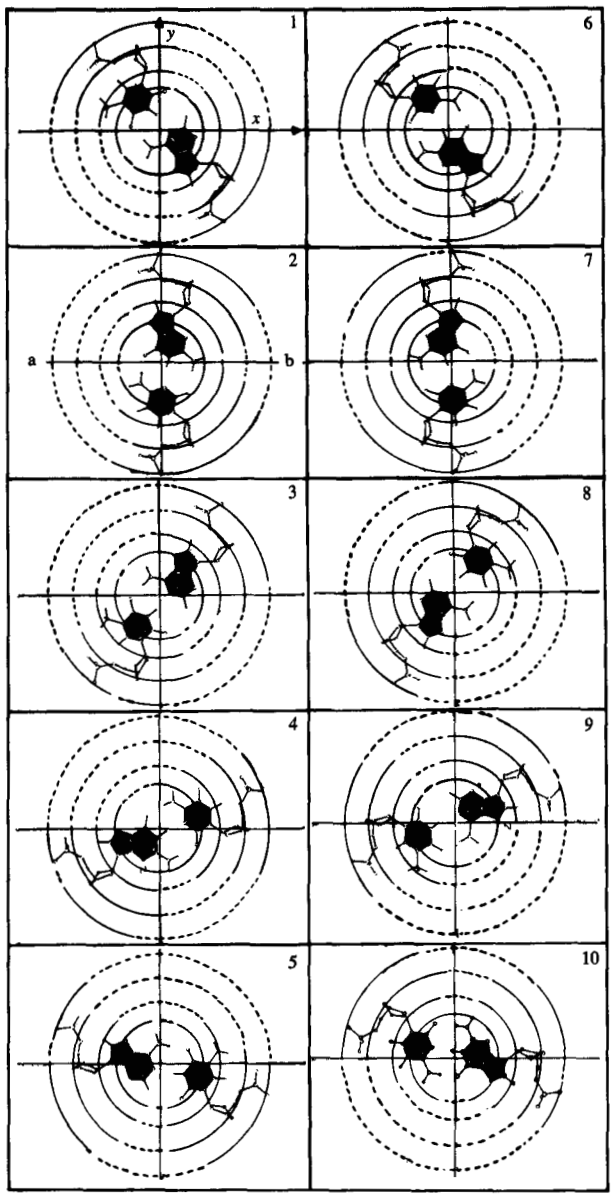

Figure 1 Projection into the *xy* plane of ten (out of twelve) base pairs of the B-DNA fragment; see text.

chemistry). Slowly, over the past ten years, a "library" of transferable atom-atom pair-potentials has been constructed. Today, for example, the simulation of the interaction of ions or water molecules with amino acids, proteins, DNA and RNA bases and base pairs, nucleic acids, or of any amino acid with another amino acid, is a relatively simple task. Before the library existed, such simulations would have been nearly unfeasible.

In Model 3, the transferable potentials are the input parameters simulating static or dynamic properties of an ensemble of atoms or molecules. Static properties can be obtained from time-dependent models, as is known from the ergotic hypothesis. We shall, however, use the time-independent Monte Carlo (MC) technique [20] to simulate static properties, and the time-dependent molecular dynamic technique [21] to simulate dynamic properties. Among the many applications of the MC technique, we recall studies on liquid states and solutions. After an initial period when the use of the MC technique was limited to the study of liquids of rather marginal chemical interest (for example, rare gas liquids), we have witnessed in the last ten years simulations of liquids and solutions of definite chemical interest. Here, the pioneering Monte Carlo study of "liquid" water by Barker *et al.* [22] should be mentioned, stressing in addition the basic role of simulations of rare gas, the necessary prolegomena to simulations of chemical solvents and solutions. In the second part of this decade potentials obtained as described in Model 2 have been used more and more frequently. We recall, for example, the studies on nucleation by Abraham *et al.* [23], and those on coordination numbers of ions in water clusters by Clementi *et al.* [24]. More recently, the MC theory has been used successfully to simulate the free energy of liquid water [25] and to address problems related to liquid-vapor interfaces [26] and hydration in enzymes [27]. In molecular dynamics, both empirical and phenomenological potentials have been used; there is no limitation, however, to the use of quantum-mechanically derived potentials. Work is in progress along such lines.

Solvation of DNA

From Hartree-Fock simulations it is known that an oxygen atom ground state is the ^3P state ($1s^2 2s^2 2p^4$) [28]. Inclusions of correlation corrections [29] and relativistic effects [30] confirm this "prediction." Starting from two hydrogen atoms and an oxygen atom, a stable molecule can be formed in the Hartree-Fock approximation with an O–H distance of 0.97 Å (in this paper, distances and bond lengths are given in angstroms, the traditional crystallographic unit: 1 Å = 0.1 nm) and an H–O–H bond angle of 106°. This molecule can be simulated more accurately by including electronic correlation effects; its bond angle becomes smaller (104.5°) and the bond length becomes marginally shorter [31]. Thus, predictions by simulations on the structure and stability of the water molecules can be obtained rather standardly from Model 1. Two such molecules can interact, and in the Hartree-Fock limit the interaction has been computed and expressed in a rather simple analytical form. The dispersion correction has been obtained by perturbation theory [18] and by configuration-interaction techniques [32] and is in remarkable agreement with experimental data. When three such molecules interact, the two-body approximation is only

partially valid, but the nonadditivity error can easily be accounted for quantitatively [19]. At finite temperature (T = 298 K), a Monte Carlo computation (at the two-body-potential level) has been performed on an ensemble of these molecules assuming periodic boundary conditions. This system, if probed by x-rays and neutron beams, yields a scattering diffraction pattern which also has been obtained by simulation [32]. The experimental diffraction pattern and the computer-simulated pattern are in notable agreement. The computational effort has been extended by considering the interaction of water molecules with ions, *e.g.*, Li^+, Na^+, K^+, Be^{++}, Mg^{++}, Ca^{++}, Zn^{++}, NH_4^+, F^-, Cl^-, or zwitterions of biological interest [14, 32]. The resulting potentials (obtained as described in Model 2) have been used in MC computations for simulating solutions consisting of one or two of the above ions as solute species and an ensemble of water molecules as the solvent. Presently, work is in progress to include three-body correction in the simulation of liquid water, as was previously done for ion-water clusters [14]. This refinement is expected to bring scattering intensity simulations to current limits of experimental accuracy. In addition, we expect to obtain a reasonable value for the simulated pressure of the liquid. (At present, the simulated value is far from the experimental one, as is clearly to be expected from a potential representation that includes only two-body quantum-mechanically derived interactions.)

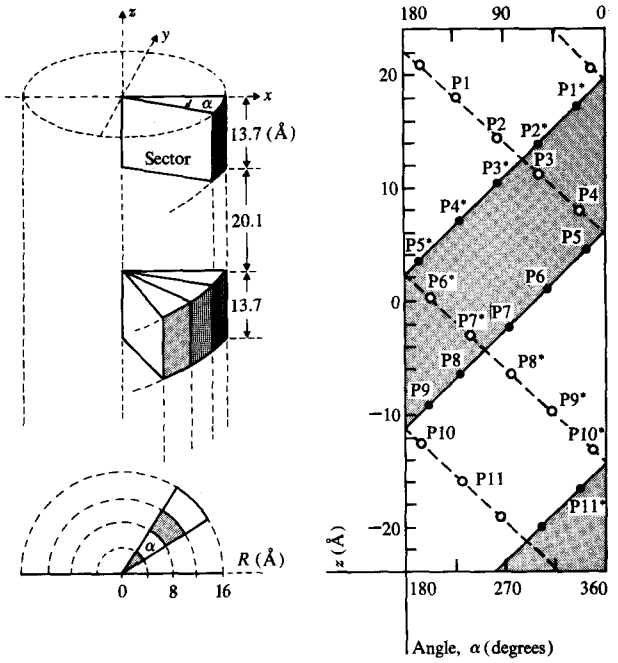

Figure 2 Analyses of grooves: sectors and subvolumes (left), major (solid lines) and minor (dashed lines) grooves (right); the bottom scale is seen from the $-y$ axis, the upper scale from the y axis.

Therefore, starting only from the assumption that atoms or ions have a given number of electrons and a point-charge nucleus, with the above three-level model, simulations have first characterized the electronic structure of the atoms. These are then used to simulate either a molecule (water) or ion-water and water-water complexes in the gas phase at zero temperature, as well as liquid water and solutions at finite temperatures, all in remarkable agreement with experimental data.

The above simulations are, however, still far from the chemical complexity of everyday chemistry. In the following discussion we shall present a new set of simulations on more complex systems. Let us now assume a given conformational structure for the DNA double helix (*e.g.*, the B conformation). As is known, single crystals of B-DNA have not been obtained, and the structure of this important nucleic acid is to date mainly a model, which is, however, in agreement with somewhat crude diffraction patterns obtained from fibers (rather than crystals) and with a large amount of indirect but very strong evidence. As is experimentally known, water solutions of B-DNA contain counterions like Li^+, Na^+, K^+, Mg^{++} or Ca^{++}. In addition, no high-resolution diffraction pattern of B-DNA in solution has been experimentally reported or interpreted. However, the structure of liquid-water-solvating B-DNA has been recently simulated using the atom-atom pair potential library and Monte Carlo techniques [14]. Unfortunately, B-DNA without counterions is unstable in solution. In this work, we report the solvent structure for B-DNA with counterions. Over the last twenty years, a large number of studies attempting to determine the structure of water around DNA in solution have been published. Because of experimental difficulties, however, no direct experiment yielding an unambiguous answer has been presented [33].

The DNA fragment considered in the Monte Carlo simulation is composed of 760 atoms and consists of 12 base pairs, 22 phosphate groups, and 24 sugar units. It is known that a full turn of the B-DNA helix consists of only ten base pairs and of the corresponding sugar-phosphate units. We have added two additional units to improve the representation at the fragment boundaries. In Fig. 1 the ten base pairs and the sugar-phosphate units are projected into the xy plane. The z axis is the main axis of the double helix (hence, the base pairs are stacked one above the other along the z direction; the base pair indicated as 1 is at the top, while that indicated as 10 is at the bottom). In the simulation the water molecules and the DNA fragment are enclosed in a cylinder of radius R = 14.5 Å (it is

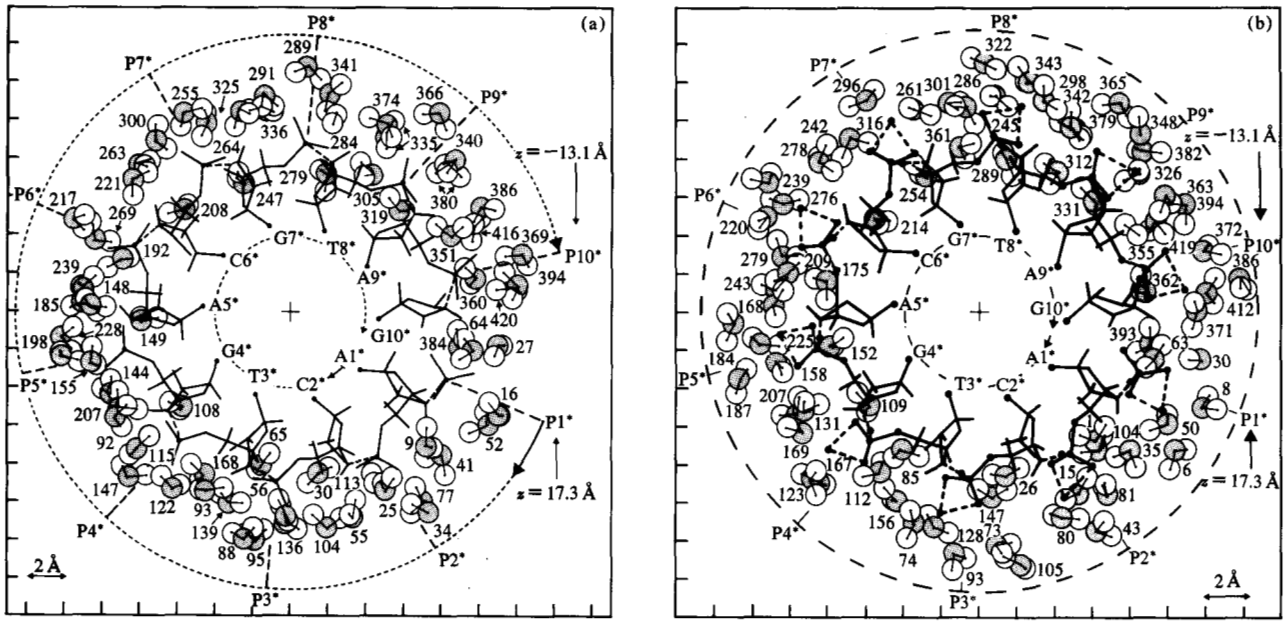

Figure 3 Water molecules solvating (a) the phosphate groups of the h* strand of B-DNA and (b) the sodium phosphate groups of the h* strand of Na^+-B-DNA. The bases are A = adenine, G = guanine, C = cytosine, and T = thyamine.

noted that the most external atoms in DNA correspond to $R = 10.2$ Å). *Two* different analyses have been performed on the millions of water configurations generated in the Monte Carlo process. In the first, the distributions and energies are averaged over a volume defined by two coaxial cylinders of radius r and $r + dr$ (with limiting values of 0 and 14.5 Å). In the second analysis, we follow the water distribution along the major and minor grooves, by considering subvolumes defined by sectors delimited by the groove heights (13.7 and 20.1 Å, respectively), measured at the phosphorous atoms, and by an arc defined by two angle values α and $\alpha + d\alpha$. These representations are indicated in Fig. 2, where we also report the projections onto the xz plane of the phosphorous atom positions along the two strands h and h* (the presence or absence of an asterisk differentiates the two strands) and the values of α for a full rotation (360°). The shaded area refers to a portion of the minor groove.

In Figs. 3(a) and (b) we present details of the structure of water solvating at $T = 300$ K, phosphate groups in B-DNA, and phosphate groups in Na^+-B-DNA [34]. Figures 4(a) and (b) present the probability distributions for the water molecules solvating B-DNA either with or without counterions (Na^+). In these figures we have distinguished between the hydrogen and oxygen atoms of water. The number of water molecules enclosed in a volume from r to $r + dr$ are reported for the entire solvent sample (*total distribution*), for the water molecules in the first solvation shell (*bound distribution*), or for the remaining water molecules (namely the difference from the total obtained by subtracting the bound water molecules or the *remaining* or *difference distribution*). The intensity peaks for hydrogen atoms are identified by lower-case letters and those for oxygen atoms by upper-case letters. By analyzing the solvation at the DNA sites (see bottom inserts of Figs. 4(a) and (b), each peak of the total distribution can be resolved [35, 36]. By adding counterions to DNA, the main effect for the solvent is the *ion-compression-induced effect*. That is, water molecules are packed more densely around DNA (and in the major and minor grooves) than when counterions are absent. This effect can be seen very clearly by comparing the shift of the intensity peaks to smaller R values in Na^+-B-DNA relative to B-DNA [compare, for example, peaks U and V in Figs. 4(a) and (b)]. This structural variation in the solvent mirrors the field variation due to the presence of the counterion. This can also be seen in Figs. 5(a) and (b), where we present the interaction energy for the water molecules enclosed in the volume between r and $r + dr$. Again, we partition (as in Fig. 4) the total water-solute interactions (total distribution) between strongly bound water molecules (bound distribution) and those not belonging to the first solvation shell (difference distribution). In addition, the inserts to the right provide details of the interaction with specific groups. The main difference in the interaction energy is its drastic increase when counterions are added; this increase is group-selective.

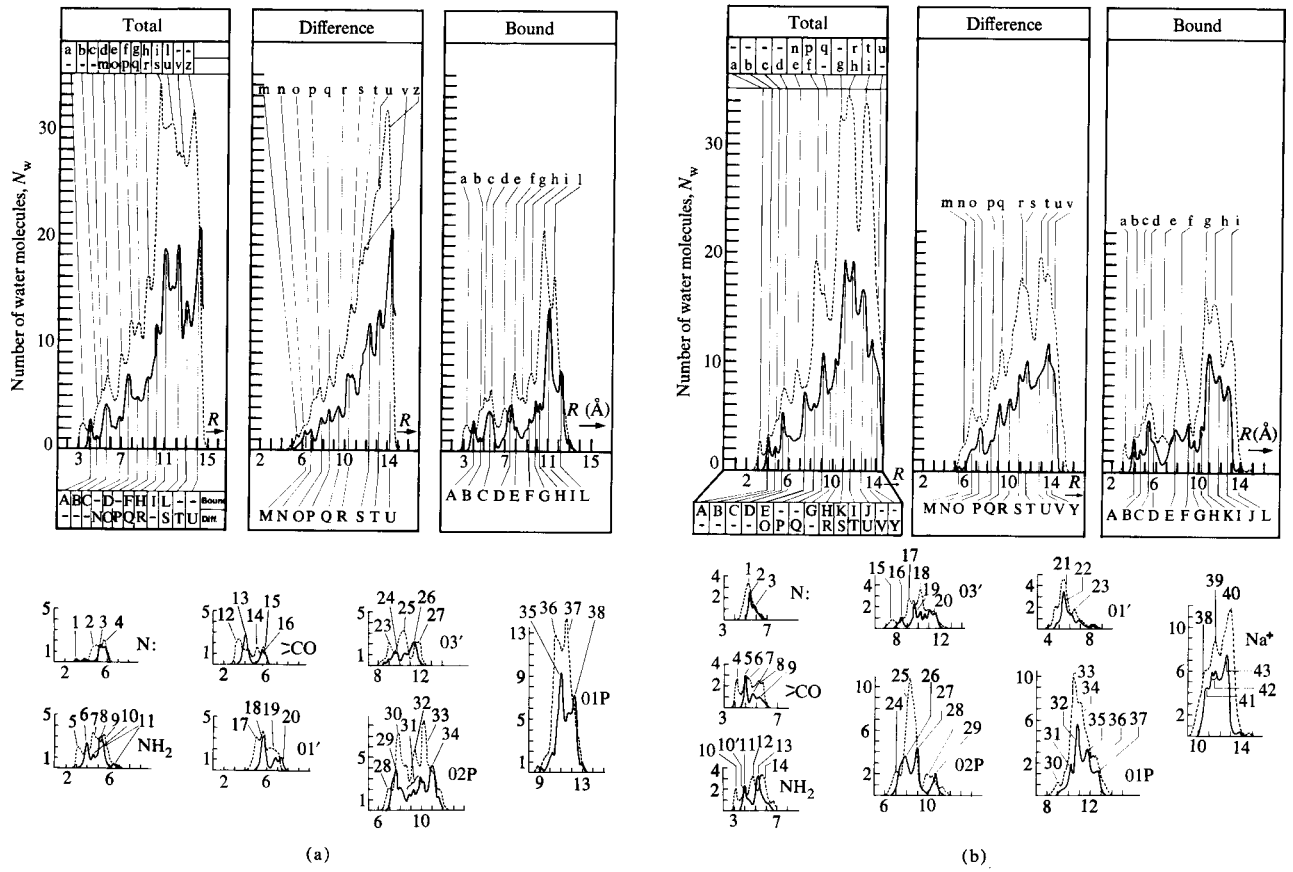

Figure 4 Probability distributions (N_w) for hydrogen and oxygen atoms in the water molecules as a function of R (Å) at 300 K in (a) solvated B-DNA and (b) solvated Na^+-B-DNA. In each case 447 water molecules are used in the simulation. The three inserts at the top of each figure refer to the total, difference (or remainder), and bound distributions; see text.

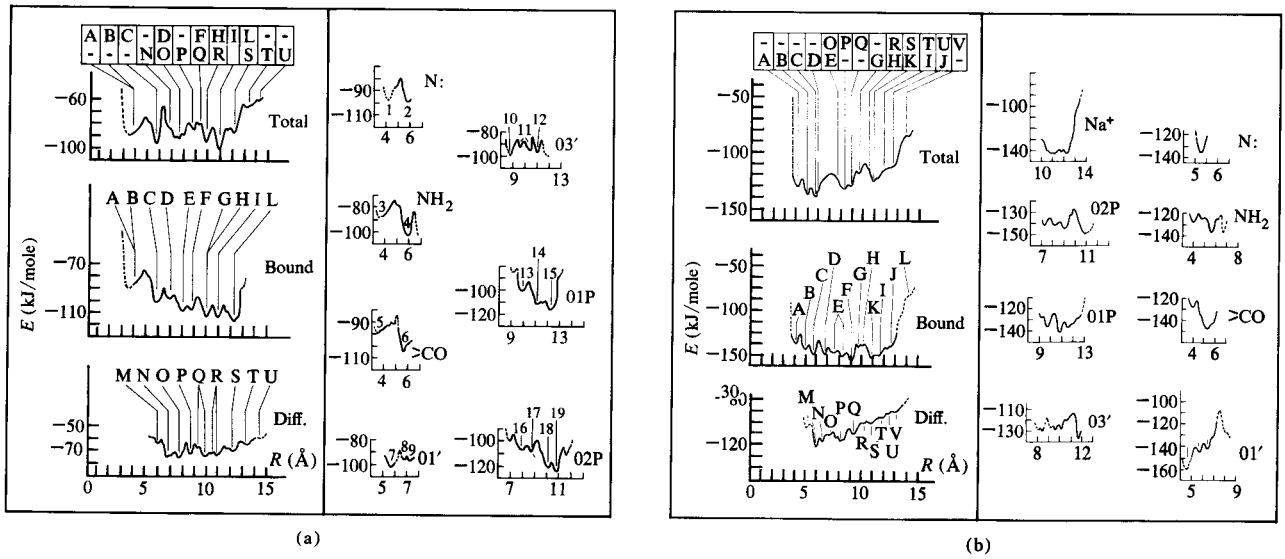

Figure 5 Energy distributions (in kJ/mole) *versus* R (in Å) at 300 K for the water molecules solvating (a) B-DNA and (b) Na^+-B-DNA. The inserts to the right refer to specific site distributions.

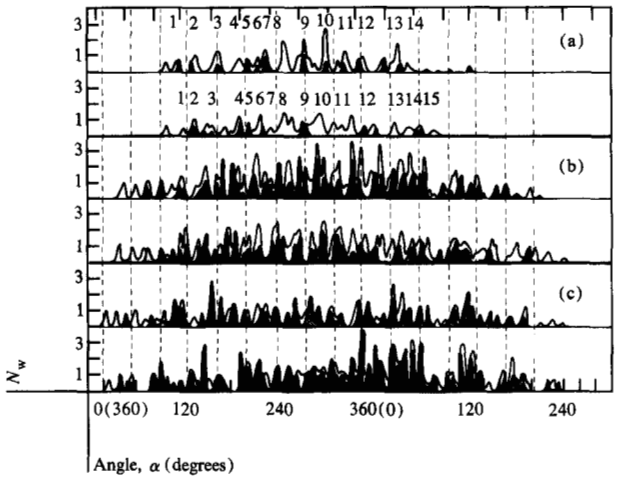

Figure 6 Probability distributions for water molecules in Na^+-B-DNA and B-DNA major grooves. Both first-solvation-shell and groove water molecules are considered. (a) top: Na^+-B-DNA (N_w = 47.56, E = −118.48 kJ/mole; bottom: B-DNA (39.99, −81.40); 4 Å < R < 8 Å. (b) as in (a) but top: (142.15, −116.88); bottom: (121.26, −87.74); 8 Å < R < 12 Å. (c) as in (a) but top: (89.52, −96.02); bottom: (117.12, −67.79); 12 Å < R < 14.5 Å.

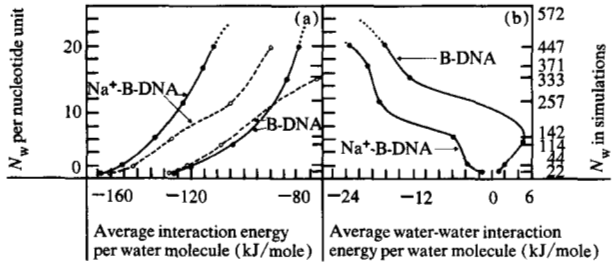

Figure 7 Average interaction energies per water molecule (in kJ/mole) during absorption: (a) total interactions (solid lines), water-B-DNA interactions (dashed lines); (b) water-water interactions.

The water in the grooves is highly structured. Along a given strand (h or h*) we find *intra*-phosphate water molecules bridging water molecules bound to two successive phosphate groups [*e.g.*, P3* and P4* (on h* strand) or P3 and P4 (on h strand) in Fig. 2]. Across two strands we find *inter*-phosphate filaments of hydrogen-bound water molecules connecting two phosphate groups (*e.g.*, P5 and P11* in Fig. 2). This network adds stability to the DNA conformation. The network's exact structure is determined by the DNA field. By variations in the counterion position (concomitant to conformational transitions in DNA) or type (for example, substitution by Na^+ for K^+), the field changes and the network structure changes accordingly. Therefore, in considering the stability of the DNA conformation, the entire solute-solvent system must be considered. In Fig. 6 we report the probability intensity distributions along the major groove. The shaded areas refer to water molecules that are not in the first solvation shell of the DNA atoms facing the major groove; the increments in the probability intensity (nonshaded areas) refer to water molecules in the first solvation shell. A detailed analysis of the peaks allows us to establish the existence of intra- and inter-phosphate hydrogen-bound water filaments, as defined above. It is noted that the existence of highly structural water was postulated years ago [37]; however, no direct structural determination has previously been reported.

The early studies [33] of the water structure for DNA were often done by adding differing amounts of water to DNA fibers (absorption at different relative humidities) or by removing water from them (desorption). These experiments can be simulated by using variable numbers of solvent molecules in the Monte Carlo simulation. In Fig. 7, we report the primary results of the simulated interaction energy using 22, 44, 142, 257, 371, or 447 water molecules to solvate Na^+-B-DNA and 22, 44, 114, 333, or 447 water molecules to solvate B-DNA. All computations were performed at a simulated temperature of 300 K. In the figure we present the energy data either relative to the number of water molecules N_w selected in the simulation or as the number of water molecules per nucleotide unit. The top insert reports the total interaction energy, *i.e.*, the sum of the water-water and water-DNA interaction energies. These data, when replotted as a function of the relative humidity, yield a sigmoidal shape characteristic of absorption-desorption isotherms, in agreement with experimental findings [33]. Note the stabilization (increase in interaction energy) due to the presence of counterions and the different behavior of the water-water interaction energy due to the presence or absence of counterions. Detailed analysis [35, 36] indicates that at low humidity a water molecule in Na^+-B-DNA is affected by two opposing orientational forces (one from the phosphate group, the other from the counterion), thus allowing for water-water conformations that are weakly attractive. In B-DNA, only one force is predominant (that from the phosphate group), compelling the water molecules to assume water-water conformations that are weakly repulsive. Clearly, as the relative humidity increases, the water-water interaction becomes more and more important; for infinite dilution, it tends to the simulated limit of about −36 kJ/mole. The region of intermediate relative humidity (about six to twelve water molecules per nucleo-

tide unit) is very interesting, since it corresponds to large water-water interaction-energy variations for relatively small relative-humidity variations.

The DNA macromolecular structure has been determined for a number of different conformations such as the well-known A, B, C, etc., conformations [38]. Simulations can answer the important questions concerning the stabilization of these conformations in solution, at different relative humidities, with different counterions, or with temperature variations. Preliminary data are reported in Fig. 8, where we consider the following cases: 1) stabilization due to the solvent, considering the A → B transition at constant temperature, Li^+ counterions, and different relative humidities; 2) the same transition with Na^+ counterions; 3) considering an A → B transition by allowing relative-humidity variations and by changing the counterions from Li^+ to Na^+; and 4) as in 3) but changing the counterions from Na^+ to Li^+. These preliminary results, now being refined, predict that an increase in relative humidity in Na^+-B-DNA or in Li^+-B-DNA would stabilize the B-conformation (in the A → B transition process). This stabilization is obtained more efficiently by substituting Na^+ for Li^+ but hindered by substituting Li^+ for Na^+. The critical humidity range is around seven to nine water molecules per nucleotide unit, where the transition crossover occurs. These theoretical predictions, confirmed by a number of experimental data [38], are now being refined.

In the above simulations the counterions have been held rigidly at the minimum-energy position, as determined from quantum-mechanical computations for the interaction energy surfaces between ions and a sugar-phosphate-sugar fragment. However, in nucleic acids the counterions experience the entire field of the nucleic acid and of the surrounding medium. Therefore, a more realistic (and also more demanding) simulation has been performed, allowing not only the water molecules but also the counterions to find the most probable positions (at a given temperature) in the field of DNA (kept rigid) and of the remaining water molecules and ions. In this new Monte Carlo simulation [39] we have imposed periodic boundary conditions for the water molecules and the ions; a B-DNA fragment of ten base pairs (with the corresponding sugar and phosphate units and representing a full B-DNA turn) is periodically repeated three times. Thus, in this new simulation each water molecule and each counterion experiences the field of three B-DNA full turns (yielding a B-DNA fragment of 2280 atoms). The simulation temperature is 300 K. Within each B-DNA full turn we have simulated 20, 40, 140, 180, 220, 240, 380 or 400 water molecules, representing the range from very low to very high relative humidity.

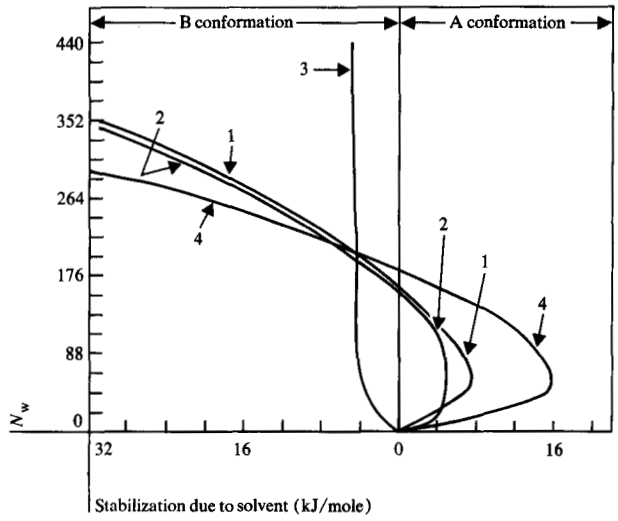

Figure 8 Solvent stabilizations at constant temperature (300 K) and different humidities for the A → B transition in DNA. Curve 1: Na^+-A-DNA → Na^+-B-DNA, Curve 2: Li^+-A-DNA → Li^+-B-DNA, Curve 3: Na^+-A-DNA → Li^+-B-DNA, and Curve 4: Li^+-A-DNA → Na^+-B-DNA. See text.

We have determined [39] that the twenty Na^+ counterions (present on each B-DNA turn) are subdivided into two groups arranged on two helices, the first one of radius R_1, larger than the R value of the phosphorous atoms in B-DNA; the second helix is of radius $R_2 << R_1$. A projection of the two helices onto the xy plane yields a two-ring structure. Thus, in the following we shall talk either of outer and inner helices or of outer and inner rings. The outer ring is almost exactly circular (*i.e.*, each ion of the ring has about the same R value), but the inner ring is noncircular, since it is dependent on the base-pair sequence (as expected, considering the very specific interactions between ions such as Li^+, Na^+, K^+, Mg^{++}, and Ca^{++} and the bases or base pairs).

In Fig. 9 we present the probability distributions for the ions projected onto the xy plane for the simulation in which 400 water molecules are selected for one B-DNA turn. (The positions of the counterions in nucleic acids in solution were previously unknown.) This new double-helix structure minimizes ion-ion repulsion (by depleting the region outside the PO_4^- groups of ions) and maximizes the counterion attractions to the PO_4^- groups and to the bases. The *vertical* separation between two successive ions (on a given helix) is dependent on the base sequence because it is also the xy position for the ions in the inner helix. Different counterions have different attractions to a base and to the phosphate groups; for different monovalent counterions (Li^+, Na^+, K^+), the ion-ion repulsion is nearly constant (excluding short ion-ion distances).

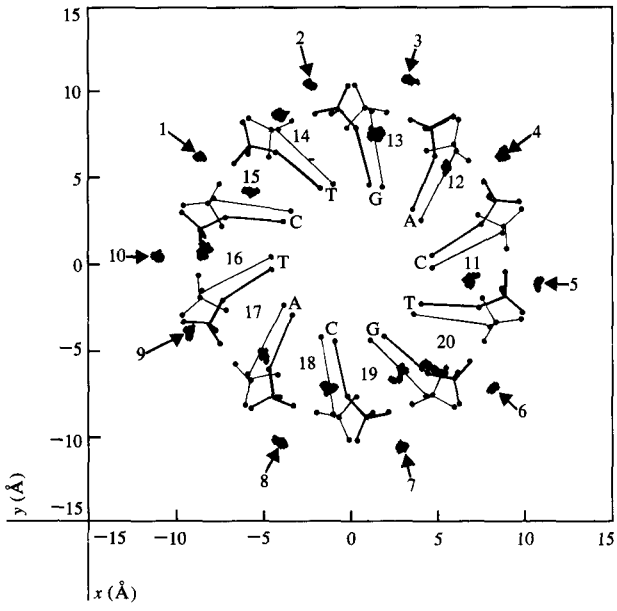

Figure 9 Projection onto the xy plane of the Na^+ counterion probability distribution (at 300 K and in the presence of 400 water molecules per B-DNA turn). The numbers shown correspond to individual counterions. The bases of the h strand are explicitly indicated (A = adenine, G = guanine, C = cytosine, T = thyamine). See Fig. 1 for the base pairs and Fig. 3 for part of the phosphate group.

Therefore, different counterions will have different R_1 and R_2 values. In Fig. 10(a) we report the projection onto the xy plane of the probability distributions for the counterions, the position of the phosphate groups, and the molecular plane of the base pairs. The double helix for the phosphate is explicitly reported. In Fig. 10(b) we report the same data as in Fig. 10(a), but here we have drawn the counterions as belonging to a double-helix-like structure of ions. Notice how one of the two helices of ions H (in correspondence to the h strand) extends farther from the DNA main axis (z axis) than the second helix H* (in correspondence to the h* strand). Whereas the ions of the inner helix H* are coordinated to two phosphates of one strand and to the bases, the ions of the outer helix are coordinated mainly to the phosphates of the other strand. The ions in the inner helix are associated with the strand with the sequence (phosphate, sugar, CH_2) or the 5′-3′ strand (the h strand), whereas the ions in the outer helix are associated with the strand with the sequence (phosphate, CH_2, sugar) or the 3′-5′ strand (the h* strand). Thus, the two strands of the DNA double helix, when considered with the connected counterions, are much less similar to one another than hitherto assumed. One would expect that this differentiation will play a notable role in the unwinding mechanism of DNA. The above-mentioned results are for 300 K, a given relative humidity (about 16 water molecules per nucleotide unit), and an ionic concentration of one Na^+ counterion per phosphate group. (The findings for Li^+, K^+, Mg^{++}, and Ca^{++} will be reported elsewhere.) By considering a larger and/or smaller number of counterions, preliminary simulations indicate that ion addition or removal is very selective relative to the two ionic distributions because of the asymmetric double-helix-like structure of the ions.

The small charge transfer computed for the $Na^+PO_4^-$ groups is different from the charge transfer computed for the Na^+-base systems. The charge-transfer differential, the unequal ionic population in the two rings, and the asymmetry of the rings bring about a net charge difference between the inner and outer helices. Long-wavelength torsional and compressional modes are present in the DNA double helix [40] since such periodic motions drag the counterions along. Thus, the field associated with the two rings of ions and the just-mentioned electrical potential differences are expected to be oscillatory. Therefore, we can propose a mechanism for the recognition of the base sequence: The inner helix, being base-pair-dependent and strongly coupled to the outer ring (the ion-ion interaction is a very-long-range interaction), produces a field which depends on the base-pair sequence of the nucleic acid. This base-sequence recognition extends considerably *outside* the helix (large R values) and can structure the water molecules even far away from DNA. In addition, because of the strong couplings between the ions in each of the two rings, the recognition mechanism is also very efficient along the z direction.

An appealing but tentative speculation can be advanced, even if it is based on grounds that are still rather weak. The small charge transfer from the phosphate and from the sugar group to the bases recently reported (but obtained with a somewhat deficient basis set [41] opens once more the question of whether DNA should be considered not as an insulator but rather as a weak semiconductor. The two-ring structure, however, will enhance the charge transfer, and as a consequence, one might expect some semiconducting character in nucleic acids. Thus, it appears that biological systems might provide ideas and models for new electronic devices. The need to include more and more solid state techniques in the study of specific biological problems is also apparent. We note that the need to account for perturbations by the solvent media of the band structure of organic polymers was anticipated over ten years ago [42]. Today, we can restudy the problem with superior techniques and more

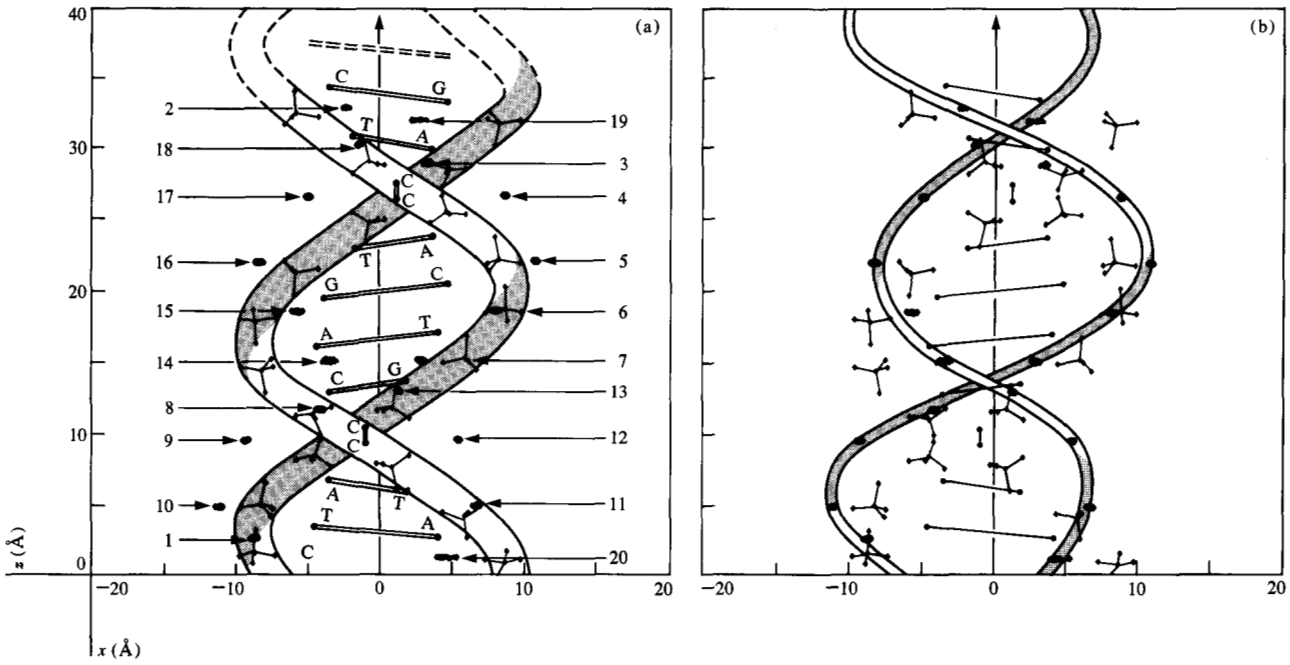

Figure 10 (a) Helix projections (h, h*) on the xz plane of the Na^+ counterions reported in Fig. 9. (Only the phosphate ions for B-DNA are shown; see Fig. 9.) (b) As in (a) but helices (H, H*) are projected for the counterions.

powerful computers, and are thus in a position to simulate more meaningful chemical systems.

It is noted that the simulation techniques reported here can be extended by adding other molecules to the solvent-solute system (not considered in this work). For example, of interest are chemical compounds with carcinogenic activity. These, if present as intercalations in DNA, not only directly affect DNA in the vicinity of the intercalation site but strongly perturb the solvent, thus influencing the quasi-A or quasi-B conformation of the nucleic acids.

Conclusions

Chemistry deals with a large number of mutually interacting molecules confined in some volume at a given pressure and temperature. The field of *ab initio* computational chemistry, when distinct from the field of *ab initio* computational molecular physics, is notably young; in fact, it is nearly at its beginning. One can only wonder whether this type of computer simulation will lead to marginally useful computations or to an unparalleled rigor and unforeseen predictive power in the field of chemistry.

In reporting our recent simulations on the solvation of DNA with or without counterions, and on our model for conformational transitions, we have emphasized two aspects: the need for an interdisciplinary approach, where atomic and molecular quantum mechanics need to be complemented with statistical mechanics in order to efficiently deal with large chemical systems, and the long time span required for the evolution of this type of research. Indeed, the atomic computations referred to in this work and marking the needed first step were started in the early 1960s. At that time, the atomic computations were performed on an IBM 704. The computations on B-DNA and Na^+-B-DNA reported here were performed partly on an IBM System 370/3033 and partly on the faster IBM System 370/3081 (recently announced).

Three different evolutions are crucial to ensure the growth of *ab initio* computational chemistry: evolution in computer performance, reliability, and accessibility; evolution in the algorithms for the specialized fields of atomic and molecular quantum mechanics on one side and statistical mechanics on the other; and finally, evolution in modeling and computer programs where energy and electronic density are considered as well as temperature, statistics, and entropy. Work is now in progress to sharpen algorithms needed to compute the entropy and to introduce the time parameter via molecular dynamics techniques.

References and notes

1. L. D. Landau and E. M. Lifschitz, *Quantum Mechanics*, Pergamon Press, Inc., Elmsford, NY, 1962.
2. See for example D. R. Hartree, *The Calculation of Atomic Structures*, John Wiley & Sons, Inc., New York, 1959.

3. C. C. J. Roothaan, *Rev. Mod. Phys.* **23,** 69 (1951) and **32,** 179 (1960); see also R. Carbo and J. M. Riera, *Lecture Notes in Chemistry*, Vol. 5, Springer-Verlag, New York, 1978.
4. *Selected Papers of Robert S. Mulliken*, D. A. Ramsay and J. Hinze, Eds., University of Chicago Press, Chicago, 1975.
5. A. D. McLean and M. Yoshimine, *IBM J. Res. Develop.* **12,** 206 (1968).
6. See for example A. D. McLean, in *Proceedings of the Conference on Potential Energy Surfaces in Chemistry*, W. A. Lester, Jr., Ed., IBM Research Division Publication RA18, San Jose, CA, 1971. For a recent description of the program, see M. Yoshimine, A. D. McLean, B. Liu, M. Dupius, and P. Bagus, in *Computational Methods in Chemistry*, T. Bargan, Ed., Pergamon Press, Inc., Elmsford, NY, 1980.
7. B. Roos, C. Sales, A. Veillard, and E. Clementi, *Technical Report RJ518*, IBM Research Division laboratory, San Jose, CA, 1968.
8. E. Clementi and D. R. Davis, *J. Comp. Phys.* **1,** 223 (1966); E. Clementi and J. Mehl, *Technical Reports RJ853* and *RJ883*, IBM Research Division laboratory, San Jose, CA, 1971; E. Clementi, E. Ortoleva, and G. Castiglione, *Comp. Phys. Commun.*, in press.
9. R. K. Nesbet, in *Adances in Quantum Chemistry*, P. O. Löwdin, Ed., Vol. 3, Academic Press, Inc., New York, 1967, and references given therein.
10. Paul S. Bagus and Arthur R. Williams, "Electronic Structure Theory," *IBM J. Res. Develop.* **25** (September 1981 issue, in press).
11. Detailed information on programming techniques is given in J. M. Andre, *Technical Report RJ527*, IBM Research Division laboratory, San Jose, CA, 1968; J. M. Andre, *J. Chem. Phys.* **50,** 1536 (1969) and *Comp. Phys. Commun.* **1,** 391 (1970); see also E. Clementi, *J. Chem. Phys.* **54,** 2492 (1971).
12. M. Philpott, P. Grant, K. Syassen, and J. Turlet, *J. Chem. Phys.* **62,** 4229 (1977).
13. J. Ladik, in *Quantum Theory of Polymers*, J. M. Andre, Ed., D. Reidel Publishing Co., Boston, 1978, p. 257.
14. E. Clementi, *Lecture Notes in Chemistry*, Vol. 19, Springer-Verlag, New York, 1980.
15. P. Dirac, *Principles of Quantum Mechanics*, Oxford University Press, Oxford, England, 1958; see Foreword.
16. The SCF-LCAO method was first applied to a molecular complex in E. Clementi, *J. Chem. Phys.* **46,** 3851 (1967) and **47,** 2323, 3837 (1967).
17. H. Kistenmacher, G. C. Lie, H. Popkie, and E. Clementi, *J. Chem. Phys.* **61,** 546 (1974).
18. W. Kolos, *Theoret. Chim. Acta* **51,** 219 (1979).
19. E. Clementi, H. Kistenmacher, W. Kolos, and S. Romano, *Theoret. Chim. Acta* **55,** 257 (1980).
20. N. Metropolis, A. W. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21,** 1078 (1953).
21. B. M. Alder and T. W. Wainwright, *J. Chem. Phys.* **31,** 459 (1959).
22. J. A. Barker and R. O. Watts, *Chem. Phys. Lett.* **3,** 144 (1969); see also J. K. Lee, J. A. Barker, and F. F. Abraham, *J. Chem. Phys.* **58,** 3166 (1973).
23. F. F. Abraham, D. E. Schreiber, and J. A. Barker, *J. Chem. Phys.* **62,** 1958 (1975); M. Mruzik, F. F. Abraham, D. E. Schreiber, and G. M. Pound, *J. Chem. Phys.* **64,** 481 (1976).
24. R. O. Watts, J. Fromm, and E. Clementi, *J. Chem. Phys.* **61,** 2250 (1974) and **62,** 1388 (1975); E. Clementi and R. Barsotti, *Theoret. Chim. Acta* **43,** 101 (1976) and *Chem. Phys. Lett.* **59,** 21 (1978).
25. M. Mezei, S. Swaminathan, and D. L. Beveridge, *J. Amer. Chem. Soc.* **100,** 3255 (1978); S. Romano and K. Singer, *Mol. Phys.* **37,** 1765 (1979).
26. J. Miyazaki, J. A. Barker, and G. M. Pound, *J. Chem. Phys.* **64,** 3364 (1976); F. F. Abraham, *J. Chem. Phys.* **72,** 1412 (1980).
27. E. Clementi, G. Corongiu, B. Jonsson, and S. Romano, *FEBS* **100,** 313 (1979); *J. Chem. Phys.* **72,** 260 (1980); E. Clementi, G. Ranghino, and R. Scordamaglia, *Chem. Phys. Lett.* **49,** 218 (1977); G. Ranghino and E. Clementi, *Gazz. Chim. Ital.* **108,** 157 (1978).
28. E. Clementi and C. Roetti, *Atomic Data and Nuclear Data Tables*, Academic Press, Inc., New York, 1974; see also *IBM J. Res. Develop.* **9,** 2 (1965) and supplement.
29. See for example F. Sasaki and M. Yoshimine, *Phys. Rev.* **A9,** 26 (1974) and references given therein.
30. H. Hartmann and E. Clementi, *Phys. Rev.* **A133,** 1294 (1964).
31. G. H. F. Dierksen, *Theoret. Chim. Acta* **21,** 335 (1971).
32. See for example E. Clementi, *Lecture Notes in Chemistry*, Vol. 2, Springer-Verlag, New York, 1976.
33. J. Texter, *Prog. Biophys. Mol. Biol.* **33,** 83 (1978).
34. The remaining atoms of DNA are not shown in order to simplify the figure.
35. G. Corongiu and E. Clementi, *Biopolymers* **20,** 551 (1981).
36. G. Corongiu and E. Clementi, *Biopolymers* **20,** in press (1981). This work details quantitatively the data reported in the figures and compares our simulation with a number of experimental findings.
37. S. Levin, *J. Theor. Biol.* **17,** 181 (1967).
38. See for example *Stereodynamics of Molecular Systems*, H. Sharma, Ed., Pergamon Press, Inc., Elmsford, NY, 1979.
39. G. Corongiu and E. Clementi, IBM Data Processing Products Group laboratory, Poughkeepsie, NY, unpublished results.
40. L. L. VanZandt, E. W. Prohofsky, and M. Kohli, *Inter. J. Quantum Chem.* **S7,** 35 (1980).
41. J. Ladik and S. Suhau, *Inter. J. Quantum Chem.* **S7,** 180 (1980).
42. E. Clementi, *J. Chem. Phys.* **54,** 2492 (1971).

Received January 13, 1981; revised March 9, 1981

The author is located at the IBM Data Processing Product Group laboratory, Poughkeepsie, New York 12602.