INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

H4.SMR/642 - 19

College on Methods and Experimental Techniques in Biophysics

28 September - 23 October 1992

Time Scales and Other problems in Linking Simulations of Simple Chemical Systems to More Complex Ones

E. CLEMENTI

CRS4, Cagliari, Italy

These are preliminary lecture notes, intended only for distribution to participants.

TIME SCALES AND OTHER PROBLEMS IN LINKING SIMULATIONS OF SIMPLE CHEMICAL SYSTEMS TO MORE COMPLEX ONES

Enrico Clementi

Université Louis Pasteur, 3, rue de l'Université, 67084 Strasbourg Cedex, France

Giorgina Corongiu, Dario Estrin, Eduardo Hollauer[*], *Omar G. Stradella*

Centro Ricerca Sviluppo, Studi Superiori Sardegna, P.O. Box 488, 09100 Cagliari, Italy

ABSTRACT

We shall start with very small systems like H_2 and H_3, computed with very accurate methods (Hylleraas-CI) or atomic systems up to Zn with accurate methods (CI), then move to more complex ones, like C_{60} but now with somewhat less accurate methods, specifically Hartree-Fock with density functionals, the latter for the correlation energy but not for the exchange energy. For even more complex tasks like geometry optimization of C_{60}, we have resorted to even simpler and parametrized methods, like local density functionals. Then, we could use quantum mechanics either to provide interaction potentials for classical molecular dynamics or to directly solve dynamical systems, in a quantum molecular dynamics approximation. Having demonstrated that we can use the computational output from small systems as input to larger ones, we discuss in detail *a new model for liquid water, which is born out entirely from ab initio methods*, and nicely links spectroscopic, thermodynamics and other physico-chemical data. Concerning time-scales, we use classical molecular dynamics to determine friction coefficients, and with these we perform stochastic dynamic simulations. The use of simulation results from smaller systems to provide inputs for larger system simulations is the ''global simulation'' approach, which today, with the easily available computers, is becoming more and more feasible. Projections on simulations in the 1996-1998 period are discussed, new computational areas are outlined and a N^4 complexity algorithm is compared to density functional approaches.

---

* On leave of absence from Universidade Federal Fluminense, Niterói, Rio de Janeiro, Brasil.

1. Introduction

It takes a bit less than a picosecond for a water molecule to rotate on itself in liquid water and at room temperature, but it might take microseconds for a target molecule to enter the active site of a protein; thus, both size and time can be vastly different when two or more molecules interact. Moreover, an interaction can bring about either a chemical reaction, or simply change the relative position of two or more molecules and eventually bring about some binding. As it is well known, the maximum of the 1s orbital in the H atom is about 1/2 Å, but when H (^{2}S) reacts with an appropriate photon, it can be excited to higher states which are larger in size, for example by some *fraction of one* Å. However, we can learn much about a protein even if we are limited to 3 - 5 Å resolution. Again the ''different'' space scales of chemistry! It follows that it is not surprising that at one end of the space-time spectrum (small and fast) we are interested in questions which might be rather meaningless at the other end (large and slow). Thus, one would expect that different models are available and that these specialize in yielding answers either at one or at the opposite end of the spectrum. Somewhat paradoxically, ''small and short-lived'' events are not ''simpler'' than those ''large and slow'', since generally for the former we require many details and higher resolution. For example today's quantum chemists are still working at the energy surface of H_3, but the goal is a fraction of cm^{-1} accuracy in the absolute total energy; on the other hand, an *accurate absolute energy* for a protein would be a somewhat useless information. Therefore, in general, ''molecular recognition'' has a different meaning when we study the capture of an electron by H_2 or the docking of a drug onto the active site of a macromolecule. The task of this paper is to recall some attempts in passing from small size to large, from short to long-lived events. The emphasis is in rationalizing and suggesting pragmatic ways to model the entire spectrum of events. Since there is no single theoretical equation available today to model ''any'' aspect of matter and since ''equations which cannot be solved'' shall not be stressed in this paper, our task is ''how to pass from one *working equation* (and model) to another working equation (and model)''.

By and large there are two main avenues of computational methods in chemistry, precisely computational quantum chemistry and computational statistical mechanics. The former includes ground state and excited state properties and corresponding spectra for atoms, molecules and solids. The basic equation is either the Schroedinger equation in its many approximations, or the Fock-Dirac equation and corresponding ''corrections''. The main statistical mechanics approaches are either time independent (Monte Carlo) or time dependent (Molecular Dynamics). The basic equations are either Newton equations or Langevin equations; as it is known, the latter can be considered as a modified form of the former, and it can be used when one is interested in an average over a long time.

A third direction is becoming more and more conspicuous, even if its origin is as old as quantum chemistry: reaction rate determination, scattering processes (atom-electron, atom-atom, atom-molecule, molecule-molecule) and especially computer directed synthesis for organic systems. We expect that this avenue, which presently makes use of ''quantum chemistry and statistical theory'', will (in due time) overshadow the two above mentioned, since it is here that one leaves ''atomic and molecular physics'' to enter into ''computational chemistry''. Indeed, chemistry, as its main objective, must explain and predict ''chemical reactions''. However, in this third avenue, computer simulations must compete with over one century of cumulated and organized experience, which has developed a most accurate and powerful physico-chemical experimentation, today routinely used by tens of thousands of chemists in the academic and industrial world.

There are new approaches in today's computational chemistry taking up their place alongside the traditional ones. Indeed, the traditional techniques for stationary state quantum chemistry and equilibrium molecular dynamics are being expanded with Quantum Monte Carlo,[1] Quantum Molecular Dynamics,[2] Microdynamics[3] but also by research and computer programs on Data Base, Interactive Animation, Artificial Intelligence and Chemical Knowledge Processing[4]. A special comment could be added for graphics and animation, which permeate much of today's computer applications.

Presently, we have left the IVth computer generation, and we are in the VIth generation where we are witnessing not only increased MIPS performance, but especially MFLOPS performance, either because of vector or parallel architectures. From personal computers, to workstations, to mainframes and supercomputers, the advances have been on a very broad front. More relevant: the computer cost has become more reasonable since the introduction of the workstations.

The next computer generation will be the much heralded Vth generation (which has been promised since early 1980!). In the VIst generation we have already announcements of supercomputers with peak performance of about 20-30 GFLOPS; thus, it is expected we shall grow between 100-500 GFLOPS, partly by decreasing the clock speed, mainly because of multi-processing and parallelism. We already see workstations with the same speed as the old CRAY XMP (single processor) and the disk shortage is being addressed by channels working in parallel, with optical fibres and one or more miles in interconnected distances. In this climate, the Vth computer generation is appearing, and artificial intelligence and expert systems are becoming ubiquitous; all this will add pressure to the hardware manufacturers, and prices will once more decrease, while performance will increase.

In this paper we shall consider *today's* computational *techniques*, but we shall already assume the availability of systems with up to 500 GFLOPS and capable of retrieving from disk storage hundreds of Giga words at 100-500 Mbytes/sec

transmission speeds. These predictions are ''reasonable'' in a technical view-point, and also ''reachable to relatively few'' in an economical context. Clearly, the main computing will trail behind these limits, even if everybody will experience much improvement. We shall call ''Nike'' our 500 Gigaflop dream machine (clearly with parallel architecture) and we advise the reader, whenever we refer to Nike, to set his/her clock to about 1996 or 1998. In other words, while reviewing past work, we shall pretend to be five to six years into the future.

We shall start by considering one of the first techniques introduced in quantum mechanics[5] and later adapted within the configuration interaction (CI) framework: Hylleraas CI. The examples will deal with the simplest chemical systems but at the highest accuracy level.

Next, we shall move away from the goal of 0.1 cm^{-1} (i.e. 0.000005 a.u.) accuracy in absolute energy and move to the milli-hartree accuracy, again using CI, but with the standard expansions of determinantal functions. The examples recall computations on iso-electronic series for He (^{1}S), Li (^{2}S), Be (^{1}S), and Ne (^{1}S), and anticipate some results on the Zinc atom (and some of its ions).

We shall not consider by now standard and popular methods like Moller Plesset, MP, perturbations, already too well known. Notice that there is a tendency in today's literature to overclaim the accuracy one has obtained using the MPx techniques. Indeed, we note that most authors are comparing, for example, binding energies in a molecule with atomisation products, which are often far from the limit of the adopted approximation. For example, often a calculation proceeds to post-Hartree-Fock corrections *before* reaching a near Hartree-Fock value (which is seldom obtained in today's literature). In addition, basis set superposition (BSS) corrections are often neglected; indeed, *basis sets with nearly zero BSS error are essentially unknown.* In this way, *ab initio* techniques are used ''semi empirically'' with calibrated basis sets, which we are becoming accustomed to seeing discarded, because of their unreliability, every four to five years! A different situation occurs in the use of MBPT, where the methodological rigor is superior[6]. One of the merits of Nike will be to make routinely feasible the use of truly large basis sets, with little (less than 0.0001 a.u.) basis set superposition error (or with fully ''controlled'' error) and very near the Hartree-Fock limit, when post-Hartree-Fock techniques are selected[7].

Next, we shall discuss a sufficiently large system where even a MP2 computation would become computationally much too expensive. This is the area for *ab initio* molecular dynamics and *quantum* molecular dynamics. We shall consider two systems, one composed of 1000 water molecules in a periodic system, i.e. liquid water, and the second of 60 carbon atoms in the C_{60} cluster.

The main idea of this paper is to show a ''new build-up'' technique, from small size to large, such that the input to a given model can be obtained from the output of a model dealing with smaller scale systems. From atoms to molecules, from

molecules to liquids and solutions, from these to microdynamic and eventually fluid dynamic. We call this approach "Global Simulations"; it is the equivalent of an assembly line.

The examples we have mentioned above have been the subject of recent studies from our department, thus we are in the ideal position to report and discuss the corresponding computer time using the same hardware and system software, namely the IBM-9000/720, which we set at about 1000 times slower than Nike (if the six processors are used in parallel). This will simplify our extrapolations and yield a reasonably reliable forecast for about four to six years from now, namely at about the last stage of the VIst computer generation. We should, however, recognize that a supercomputer of the Nike class (when and if commercially available) will be an expensive tool available to a few in the developed countries only. Let us note that about thirty years after Mulliken's lamentation and plea for more money for computational chemistry,[8] here, we are again on the same subject, essentially for the same reasons.

2. Very Few Electron Systems: H_2 and H_3

The traditional approach to quantum mechanical descriptions of many-electron systems in atoms, molecules, and solids is to optimize linear combinations of one-electron functions, the familiar Hartree-Fock orbitals, namely functions of one electron only (with space and spin components). This type of approach leads then to describe correlation in an implicit way, and hence these techniques converge rather slowly to the true solution of the many-electron time-independent Schroedinger equation.

Instead of the implicit approach, one can argue that the most powerful approach should be one where inter-electronic coordinates are explicitly present into the wavefunction, namely where the one-electron functions, are replaced or extended by many-electron functions. Hylleraas was the first to develop this approach and he used it to calculate the energy and wavefunction of the helium atom with great success[5]. James and Coolidge[9] extended this method to the hydrogen molecule and Kolos and Wolniewicz[10] have shown that, to date, this is the most efficient method available for calculating accurate potential energy curves in the hydrogen molecule. The results of the latter authors even challenged the experimental spectroscopic data, "a great triumph of ab initio calculations" (R.S. Mulliken)[11]. However, due to the numerical complications introduced by the inclusion of inter-electronic coordinates in the wavefunction, Hylleraas approach was limited to simple cases such as the helium atom and later to the hydrogen molecule.

With the advent of supercomputers, a more general extension of the technique of explicit inclusion of the inter-electronic separation into the wavefunction is possible. This consists of multiplying standard many-electron configuration interaction

wavefunctions by powers, ν , of the inter-electronic distance r_{ij}. This approach is completely general, it can be applied to many-center, multi-electron systems of any type. The only limitation, as is generally the case in computational chemistry, is imposed by computational resources and limits in hardware design.

The effect of the r_{ij} terms is to explicitly introduce correlation into the wavefunction. This expansion in the inter-electronic distance is generally assumed to be a power series expansion,[12] thereby limiting ν to non-negative integers. It has been shown[13] that of the possible non-zero ν values, ν=1 is the most important and our discussion will be limited to the ν=0 (normal CI) and ν=1 (what we call Hylleraas CI, or HCI) terms. It has also been shown repeatedly that inclusion of the HCI terms does speed the convergence of configuration expansions and that the HCI method is a viable alternative when high- accuracy calculations are needed.

A computer package with standard cartesian Gaussian basis sets has been created to perform HCI calculations for many-center, two- and three-electron molecular systems[14–16]. Expansion to the most general four-electron integrals has been started. This package is called HYCOIN (Hylleraas Configuration Interaction) and it has been used successfully to calculate a number of molecular states[15–23] for H_2.

The natural place to begin our examination of correlation corrections in small molecular systems is the ground state of H_2, a two electron system[17,19,23]. The ''accurate'' variational limit for this state was established by Kolos *et al.*,[24] at -1.1744757 hartrees by using a specialized elliptical coordinate basis set with an explicitly correlated wavefunction. The largest basis set of our calculation, (15s,7p,2d,1f), produces a HCI energy that is less than 0.3 cm^{-1} above this variational limit. If spectroscopic accuracy is defined as having errors of less than ~ 1 cm^{-1}, our computations show that it can be achieved with a gaussian basis set of size as small as (13s,7p,2d).

The next obvious step towards more complicated molecules is to increase the number of centres from two to three while keeping only two electrons, this means the study of the non-linear H_3^+ molecule, a problem that cannot be tackled with elliptical coordinates. We have recently examined the equilibrium energy and calculated a number of points on the potential energy surfaces of both the ground $^1A_1$ state and first excited $^3\Sigma_u^+$ state of H_3^+.[21,22] The lowest energy obtained by us for the equilateral triangle geometry of the ground state was -1.3438279 hartree at the internuclear distance of 1.6500 bohrs with the use of a (13s,5p,3d) basis set on each site (a total of 138 basis functions). This energy is compared in Table I with other recent ab initio calculations at the equilibrium energy. From Table I we can see that this energy is significantly lower than the previous best published variational calculations of -1.343500 hartree,[17] and is in excellent agreement with the results of the latest quantum Monte Carlo calculations, -1.34387 ± .00005 hartree[31] and -1.3433 ± 0.005 hartree[32]. The energy is also below the result of Alexander *et al.*,[33] who used a

Random-Tempered Optimization method with Gaussian-Type Geminals (also a variational calculation with an explicitly correlation wavefunction), and obtained -1.3438220 hartree at an equilibrium separation of 1.6504 bohrs.

From here, the next molecule in complexity is H_3. Now, with three electrons the complexity of the integral calculations goes up dramatically and commensurately, so does the need for increased computational resources. In fact, at this moment, only small basis set calculations for linear H_3 with internuclear distances of 1.75 bohrs are available[16]. The SCF energy for this configuration was -1.5849093 a.u., the CI energy was -1.6217005 a.u., and the HCI energy was -1.6366379 a.u.. This result should be compared with the result of B. Liu[34] who obtained an energy of -1.658743 a.u. at the same saddle point with a CI calculation using a Slater-type function basis set containing orbitals up to 5f. A good computation for H_3 would require the same type of basis set used above for H_2. For example, using the MELD program of Davidson[35] with a (15s,5p,2d)/[11s,5p,2d] Gaussian basis set, we obtained a saddle point energy of -1.658323 a.u., which is 0.000423 a.u. above the best results of B. Liu. This basis would be a more than sufficient starting point for a HCI calculation on H_3 to ensure a few cm^{-1} accuracy in the total energy. The forecast computer need is extrapolated below.

It should be noted that the following timings were found with a very recently implemented HCI code that has not yet had extensive optimization, therefore, the timings and, more importantly the extrapolations, should be taken as indicative rather than concrete predictions, as effort to-date has concentrated on ''correctness'' rather than ''performance''. To begin, a trivial basis set of a single s-type orbital on each centre takes 350 seconds on IBM ES/3090 VF to calculate the integrals. A basis set of 1s1p on each centre (12 orbitals) takes 191 hours on the same system and a basis set of 3s,1p requires approximately 1000 hours; this forecast looks rather bad, but it corresponds to only 1 hour on our Nike. Notice that the number of integrals grows as N^6 where N is the number of basis functions. It is intriguing, and somewhat depressing, to speculate what would be required to perform a H_3 calculation that would be accurate to within 10 cm^{-1} of the ''exact'' value, which is not yet known as it is for H_2.[24] If the lessons learned from H_2 are valid, then a full HCI calculation with a basis set of (13s,7p)/[13s,1p] should provide approximately 10 cm^{-1} accuracy. Scaling from 3s,1p to 13s,7p by N^6 would call for somewhere around 3500 years on a similar system (but only about 10 months on our Nike), not counting the six-index transformation which scales as N^7. This is clearly not yet approachable without fundamental improvements in either, and probably all, theory, implementation, and hardware. Assuming a significant improvement in the numerical analysis and its corresponding code, at best we could reduce the task to 10-20 years on a IBM-9000/720 which further reduces to a few days on Nike. This is not unfeasible especially if a more efficient numerical integration will be found and if integrals smaller than a given threshold are ignored and/or approximated. Indeed the latter task is regularly implemented in

SCF computations where very large savings are well-known and documented.[36] Thus in the second half of this decade we shall finally see some very accurate surface for H_3! In addition, the above computation for H_3 should gain, relatively to the past one, by a superior choice in the selection of the computational grid on the energy surfaces. Today, as in the past, there are experts in obtaining energy surfaces and experts in using them in order to obtain reaction rates and scattering data; hopefully, tomorrow there shall be a few experts in *both* areas.

3. Few Electron Atoms

About thirty years ago, one of us (EC) undertook a systematic study of atomic functions[37] to determine the Hartree-Fock energy of atoms and the corresponding isoelectronic series from 2 to 36 electrons and from Z=2 to Z=36 (including all excited states of the ground state configuration). The main motivation was to be in position to closely estimate the relativistic correction (by perturbation) and with such data and the use of experimental ionization potentials, it was possible to estimate the atomic correlation energy. It is well known that the knowledge of these quantities is essential in the determination of molecular binding. In addition, the quantitative knowledge of the correlation energy was *nearly unexistent*, since there were exceedingly few Hartree-Fock energies and no systematic study.

Only very recently Clementi's old estimates have been re-examined, using modern computers, at the IBM-Kingston laboratory[38] and shortly after by E. Davidson *et al.*[39].

For our task we used a more traditional approach, namely a linear combination of Slater determinantal functions, i.e., the Configuration Interaction approach.

We used ATOMCI, an Atomic Configuration Interaction program based upon the powerful techniques developed in the early seventies by Sasaki,[40] and used by Sasaki and Yoshimine[41,42] to compute correlation energies and electron affinities for the first row atoms. ATOMCI has been recently revised and extensively documented[43,44]. In Davidson *et al.*[39] the same code was used to obtain the correlation energy.

In Table II we report the correlation energy obtained in these studies[38,39]. Table II reports more figures than physically meaningful. These values are not too far from old ones estimated by Clementi[45,46] for some members of each series. Clementi's estimates are close to the calculated values for the 2 and 3-electron series but not for 10-electron isoelectronic series, *especially for high Z*, where they differ by as much as 30%.

For the 2-electron series the correlation energy is accurate to about the fourth decimal figure. As known, very accurate computations of the two electron series are those by Pekeris,[47,48] later often reproduced.

For the 3-electron series the accuracy remains nearly the same as for the 2-electron series, because, as long ago realized, the 1s-2s interpair correlation energy is small[45,49] and thus SDCI yields energies not much different from Full CI.

For the four electrons, the limitations of SDCI start to become apparent and the old estimate of -0.094 a.u.[49,50] for the Be is *nearer* to the exact one than the value of -0.0899 a.u. reported here. Bunge's very accurate calculated value[51] for the correlation energy of Be is -0.0943 a.u.

The steep increase of the correlation energy along the isoelectronic series has been analyzed in detail elsewhere.[46,52] Here we observe that the effect is essentially linear in Z and the 4-electron correlation energy becomes larger than 10-electron correlation energy for Z around 28. As we have previously stressed,[45,46,49] this energy should not be considered as the true correlation energy for the four electrons in the two pairs $1s^2$ and $2s^2$, but rather an example of a strong multiconfigurational electronic structure which is very poorly described by the single determinant approximation. In the limit of high values of Z the ns and np orbitals are energy degenerate. Indeed, a two determinant MCSCF function was sufficient to bring about a constant value for the 4-electron series[46], i.e. -0.052, -0.055, -0.055, -0.056, -0.058, -0.059, -0.062, for Be, B^{+1},C^{+2},N^{+3},O^{+4},F^{+5},Ne^{+6}, respectively.

The computed correlation energy for the 10-electron isoelectric series is about 95% of the estimated total[46], -0.389 a.u. The previous value by Sasaki and Yoshimine[41] for Ne, -03697 a.u., compares nicely with our -0.3706 a.u. Adding a perturbative selection for triple and quadruple excitations from the HF reference state to the lowest two ANO's of the s, p, and d symmetries obtained in the single reference SDCI calculations, yielded a correlation energy of -0.3800 a.u. for Ne (almost 98% of Veillard and Clementi estimate) and -0.941 a.u. for Ar^{+8}. Very recently, Davidson and co-workers tested our work with Slater functions and obtained the improvements[39] shown in Table II. In addition, their computation of the relativistic correction made possible to detect irregularities which were traced to inaccurate values in the ''experimental'' ionization potentials.

Presently, we[53] are analyzing a thirty electron problem, the Zn atom (and a few of its ions) this time with Slater basis sets. The computed non-relativistic total energy of the two-electrons is somewhat better than the one given by Davidson *et al.*[39] (our value is -881.407256 a.u vs -881.407028 a.u.). To us, this brings about very clearly *the painful limitation of quantum chemistry's analytical basis set expansion*: it is exceedingly difficult to obtain a ''limit'' value, even for *very* small problems. This drawback is not limited only to Slater type or gaussian type basis sets. The very popular plain wave expansions used in solid state computations are even more deficient; in present literature this drawback is carefully swept under the rug and overclaims to the contrary are becoming more and more ''accepted in literature''. Of course, in time this point will be settled.

The correlation energy for Zn presently computed is -1.32 a.u., a very preliminary result obtained with a ''first guess'' basis set without much orbital exponent optimization. The computation of the correlation energy with the Coulomb-hole algorithm yields an estimated value of -1.7 a.u. with an error which can be up to 20%. The importance in determining by C.I. computations the Zn correlation energy, is that we know quantitatively very little about *total correlation effects in 3d atoms*. This study is an example of the importance of methods with very high complexity (N^7) which are ''*simply necessary as bench-marks*''. Notice also that, to our knowledge, the Zn computation is the largest computation for accurate and direct determination of the total correlation energy in atoms using ''ab initio'' methods.

Let us now comment on the computation time, fast memory and auxiliary storage needed for these computations in order to extrapolate, for example, to 30 (Zn) and 36 (Kr) electron systems. First let us note that a single computation (15s,11p,9d,8f,5g,4h,3i,2k) for Zn requires about 7 to 8 hours on an IBM-9000/720.

We shall consider S states or even closed shell to make the estimate somewhat simpler. Both the amount of memory required and the time depend strongly upon the number of the two electron integrals computed and stored and upon the number of configurations. While almost 80% of the time is consumed in computing and transforming the integrals for He, approximately 90% of the time is taken by the evaluation of the energy matrix and by the solution of the diagonalization problem for Ne. Approximately the same percentages apply to the amounts of storage needed to store the integrals and the energy expression. As the number of configurations increases, so does the dimension of the CI matrix and the time to diagonalize it. A multi reference, *very accurate CI calculations* for Zinc or Krypton would take approximately days or weeks, respectively, and noticeable amounts of fast and disk memory; large timing on today's computers, but from a few minutes to a few hours for our Nike (parallel) computer.

To conclude this section concerning a few electron systems, it appears that we are very near to obtaine spectroscopic accuracy for H_3, and very high accuracy (over 99.0% of the total correlation energy) for the first half of the periodic table for atomic systems, *but it will require Nike type computers either to routinely perform these computations or to extend these limits*.

4. On Quantum Chemistry Molecular Programs

The two examples dealing with very small systems have already pointed out that it is not feasible to tackle a system of any size with the same approximations and with the same computer application programs. This realization is at the base of much of the computational work in the last thirty to forty years.

In the last few years, faced with the existence of many approximations and even more numerous codes, the computational chemistry community has reacted by introducing "*interfaces*" between two or more application codes aiming at easy and friendly use of the entire spectrum of special features present in the application codes. A second and -in principle- complementary approach is to provide "*platforms*" into which many and different application codes can be implanted and then communicate among themselves via interfaces and procedure routines which constitute the platform. This latter approach is well liked and promulgated especially by hardware vendors, since control of the platform (which in general remains a controlled property) brings about indirect control on the application codes and this in turn of the application software market.

Yet, computational chemistry is still very young and, even if, clearly, it can not avoid the laws of market economy, it still needs an open exchange of computer programs to test and compare methods, accuracy, robustness, applicability, reliability and user friendliness and to expedite the transfer of know-how from the research world to the industrial world and from the advanced Countries to the developing Countries.

In this respect, the Quantum Chemistry Program Exchange in USA (QCPE) and the Computer Physics Communications in UK deserve the most sincere gratitude from the entire scientific community. In this very spirit, one of us (EC) started the MOTECC initiative, which -in a way- formalizes the very liberal posture (in donating and exchanging computer programs) characteristic of the IBM groups in San Jose, California and later in Kingston, New York. Presently, the center of the MOTECC initiative is at CRS4, Italy; the MOTECC-Club, Inc., the non-profit organization registered in the State of New York and the newly founded non-profit organization Club-MOTECC-Europeen, registered in Belgium are an integral part of the initiative.

Among the many codes distributed by MOTECC, we comment below on a new molecular code, MOLECOLE, which attempts to include many aspects -connected via interfaces- of quantum chemistry.

Let us consider point group symmetry. Since the early ab initio quantum chemistry programs the use of symmetry has been recognized as an important tool in chemical calculations. The advantages of this use are related to the improved theoretical interpretation of the molecular structure and the intensive speed-up obtained in many steps of an ab initio calculation. As an example the use of symmetry in the one-electron integral calculation might obtain savings proportional to the square number of basis functions (n^2). Diagonalizations (n^3), integrals calculation, derivatives evaluations (n^4) and integral transformations (n^5) are some other steps where, when possible, the use of symmetry is strongly recommended.

Presently, as part of the continuos development effort in MOLECOLE (previously KGNMOL/IBMOL) one of us (E.H.) has been working to have this option available in this new molecular quantum chemistry package. This implementation will provide a friendly interface for users with no required background on group theory in such a way

that the unique information actually needed is a set of atomic coordinates. This interface will be able to identify and prepare most of the basic information required for the use of symmetry in subsequent steps. Among the positive aspects of this implementation is worthy to mention the wide range of point groups supported by MOLECOLE including not just abelian groups, but also non-abelian groups with or without complex irreducible representations, a feature rarely present in other quantum chemistry molecular packages. The use of subgroups with regular or special orientations is also fully supported.

In MOLECOLE the use of symmetry starts early by the creation of the symmetry adapted linear combination of atomic orbitals (SALCAO) which allows the calculation of orbitals belonging to one of the irreducible representations of the molecular point group and have the property of blocking the representation matrices of any symmetric operator. This work is extended to the symmetry adapted displacements allowing the characterization of normal modes through linear combination of primitive displacements with proper behavior under the effects of symmetry. Use is made of the Rys-King algorithm to generate this set of adapted functions[54]. Among the interesting features present in this step is the possibility of elimination of functions with undesired angular momentum like those obtained with the d of s character, f's of p character and g's with s and d character[55]. In the two electron integral calculation and in the gradient evaluation it is also possible to obtain great benefits from the use of symmetry since all the calculation of symmetry redundant integrals may be done just once through the use of petit-list formalism[56–58].

For the post-SCF calculations, work is in progress[59] exploring the use of symmetry in the numerical integration step required for a regular post-SCF density functional calculation. Other types of post-SCF calculations can be easily adapted to the use of symmetry by using the installed SALCAO's algorithm and the projector technique described in literature[60].

Along this implementation most of the remarkable features of old versions, like fast calculation of integrals, open-ended programming, orientation towards macromolecular and interaction energy calculations are kept unchanged and compatible with the use of symmetry.

The use of symmetry adapted orbitals, the wide range of groups available and the possibility of elimination of undesired functions place MOLECOLE among the most general programs that can handle symmetry, which can be applied, for example to the new family of compounds involving extended carbon clusters usually presenting high symmetry like icosahedral geometries (C_{60}), D_{5h} symmetries (C_{70}), etc.

This work will be presented in MOTECC-93 in a chapter presenting a short review of the present techniques for the use of symmetry in ab initio methods and a detailed discussion of the present implementation.

Another feature in the MOLECULE program is the possibility of evaluating the energy gradient, necessary to perform geometry optimization. This aspect of the work,

carried out by one of us (O.G.S.), will be also available to researchers as part of the MOTECC-93 package.

The scaling behavior of standard ab initio techniques with system size makes very expensive to perform quantum simulations . The computational effort for solving the HF-SCF equations scales as n^4 (n being the size of the basis set), and all methods including some treatment of the correlation energy scale at least as n^5. Different approaches with better scaling properties have been proposed, i.e. plane wave expansions ($n^2\log(n)$), pseudospectral HF techniques (n^3), and density functional methods (n^3). However, direct comparisons may be dangerous because of the very different prefactors multiplying the scaling term. Modern applications of Density Functional Theory offer at a reasonable cost, the possibility of including correlation effects in the calculation from the start.

One of us (D.E.) has implemented a density functional based program, that uses an auxiliary basis set for the electronic density expansion and treats the exchange correlation term using a least-squares fit to a set of functions, so all necessary integrals are calculated analytically. An implementation of the Car and Parrinello combined Molecular Dynamics - Simulated Annealing is also included. This allows to explore the Born-Oppenheimer surface of a system without the necessity of solving self-consistently the Kohn-Sham equations at each nuclear geometry. In this way considerable savings in computer time can be obtained, but at the cost of some loss in accuracy (small deviations from the BO surface). An additional step is the incorporation of ''a solvent'' using a mixed classical-quantum approach. Effective core potentials for the representation of solvent molecules in the solute hamiltonian should be developed and also a scheme for the polarization of the solvent in the presence of the solute field should be studied. Among our goals are first principle simulations for chemical reactions and biological processes for medium size systems in water.

5. Larger Systems with Quantum Molecular Dynamics

Let us now move to an example of a relatively large molecular system for which we wish to determine the equilibrium geometry and the relative stability of positive and negative ions and the relative energy of excited states. We shall assume no symmetry point group, since the unconstrained geometry determination is one of the main goals.

The study of carbon clusters has attracted attention recently. Large carbon molecules have, however, a long history and were found first not in a laboratory but by quantum chemical computations[61]. In the past few years, both experimental and theoretical methods have been employed to explain the unusual predominance of the 60-atom cluster in the mass spectra obtained with the laser vaporization cluster beam techniques.[62] Experimental evidence has shown that (i) the dominance of C_{60}^+ increases

with longer clustering times, (ii) both C_{60}^- and C_{60} are dominant clusters, (iii) there exists a special binding site for carbon-metal complexes, $C_{60}X$ (X = La, Ca, Sr, Ba), and (iv) the C_{60} cluster is not reactive. These results have led researchers to believe that a single structure is responsible for the experimental observations, with the most probable candidate the truncated icosahedron, the so-called buckminsterfullerene. Having the structure of the soccer ball, this C_{60} configuration has 12 pentagonal and 20 hexagonal faces with each atom identically bonded to 3 atoms.

Previous theoretical investigations of the C_{60} structure have been limited to Hartree-Fock calculations with relatively small basis sets and semi-empirical methods[63,64]. Due to the large number of electrons, the quantum chemical approaches have had difficulty to incorporate all degrees of freedom in the search for the low energy structure. For this reason we have used the Car-Parrinello approximation, which allows complete degrees of freedom while deriving the parametrized interaction potential directly from the electronic ground state.[65] The method, involving both density functional theory and classical molecular dynamic simulation, has been implemented in a study of the structure and dynamics of the C_{60} buckminsterfullerene. The method is presently limited to closed shell ground state functions and energies, namely, it suffers the standard limitations of today's density functional theories.

Our calculation of C_{60}, reported elsewhere in detail,[66] utilized the Bachelet, Hamann, Schluter pseudopotential[67] and plane waves for the wavefunction expansion with an energy cutoff of 35 Ryd. The Perdew- Zunger[68] form of the LDA was used for the exchange-correlation energy. The simulation employed a supercell of 17.5Å and FCC periodic boundary conditions. Approximately 32,000 plane waves were required for each state. The ''mass'', m, was fixed at 500 a.u. and the time step for integrating the equations of motion was 3 a.u.. The simulation required 256MB of memory and 90 seconds per iteration of the IBM-3090/600J. The wall-clock time was reduced by a factor of 5 by running in parallel on 6 processors. To obtain the ground-state structure and achieve equilibration about 300 hours CPU time were used on a dedicated IBM-3090/600J running in parallel.

The initial C_{60} configuration had the structure of a soccer ball with all bonds with the same length as in graphite. Once a self-consistent solution to the Kohn- Sham equations had been achieved, geometry optimization was performed by solving a set of steepest descent equations. The final ground-state structure has two different bond lengths, a short bond, 1.389 Å, on the edge between two neighboring hexagons and a longer bond, 1.448 Å on the edge between an adjacent pentagon and hexagon. A previous quantum mechanical calculation found similar results with short and long bonds of 1.369 Å and 1.453 Å, respectively[63]. The radius of the ball from the present calculation, 3.53 Å, agrees with the experimental estimate[62] of 3.5 Å.

The ground-state structure obtained from the above density functional calculation has been used as input into a Hartree-Fock calculation[69] where (see Table III) the

SCF energy and corresponding correlation corrections, obtained with density functionals, are tabulated not only for ions and excited states, but also for complexes with alkali ions or atoms placed at the centre of the carbon ball. These computations are large on today's standard and have been obtained using a double zeta plus polarization functions. The total number of uncontracted gaussian functions is 1800 which reduces to 960 upon contraction. Since it was decided to test the geometry obtained with the Local Density Functional approximation, the Hartree-Fock computation made no use of symmetry. Notice that the binding energies of Table III are the first literature data on binding of atoms inside fullerene.

The dynamics of the C_{60} cluster was also investigated. The nuclei were first heated by scaling their velocities to achieve the target temperature. Since heating the atoms this way causes the wavefunction to drift from the Born-Oppenheimer surface, periodic steepest descendent quenches for the wavefunction were required to regain the self- consistent solution. Once the velocity scaling ended and a final minimisation for the wavefunction was performed, the atoms were allowed to evolve in time according to the equations of motions. A total of 1.4 ps (20,000 MD steps) were simulated following the velocity scaling. After a period of equilibration the atomic trajectories were analyzed and the phonon density of states determined. The predicted high (1312 cm^{-1}) and low (452 cm^{-1}) frequency A_g modes are expected to be Raman active and should be experimentally observed[70].

Let us now move to large systems, for example, a sample of 1000 molecules of water. Can we simulate this system using quantum mechanics at each time step for energy and forces? Alternatively stated, can we make use of the Car-Parrinello technique for a very broad spectrum of physico-chemical properties? Scaling from 32000 plane waves needed for 60 carbon atoms to the 3000 atoms of the 1000 water molecules, we estimate the need of about 1.5×10^6 plane waves. The corresponding CPU time on an IBM 3090J is conservatively about 90 hours per time step with a storage requirement of 20 Gbytes (this is contrasted to about 1.4 minutes per time step using ab initio potentials of NCC type and a flexible water described below). But 90 hours on a IBM 3090 translates in about 1 minute on our Nike machine, thus totally feasible, but not today! In addition, recent computations with the Car-Parrinello method on the water dimer[71] bring about questions on the present approach, since the computed binding energy and geometry are rather poor on today's standards, despite the parametrization in the pseudo-potential and the density functional. Let us, however recall that some physico-chemical properties can be obtained rather well even with notably poor wavefunctions, and this partly offsets some of the above problems.

6. Large Systems with Classical Molecular Dynamics and Ab Initio Potentials

Recently we have, once more, re–visited[72,73] the task of formulating a new ab initio potential for the gas, liquid and crystal phases of water, which improves the MCY (Matsuoka, Clementi, Yoshimine) two-body potential[74] by including a polarizable water model in place of the computationally expensive three[75] and four–body[76] corrections. The previous preliminary results[72] (limited to a truncated expression of the model) are now replaced with those from MD simulations - below summarized- where the full polarization model[73] with vibrational degrees of freedom is tested. The new potential is designated as NCC, a short form for Niesar-Corongiu-Clementi. It consists of a two-body part and of a polarization term. Previous computations, where three[75,77–79] and four- body[76] corrections were considered, have shown that the many-body corrections are necessary for accurate quantitative predictions. However, the computational time increases dramatically when the two-body MCY potential is extended with three- and four-body corrections, term by term. Therefore we require a computationally less expensive algorithm which takes advantage of the ''large memory'' not available on previous computers.

In the NCC potential we chose an explicit representation of the polarization effects[80] by inducing dipole moments on every interacting molecule. The permanent dipole moments of the water molecules are represented by three point charges per molecule. The polarization on one molecule is hence primarily due to the global point charge distribution of the surrounding matter. These induced dipole moments in turn cause polarization on the other molecules and thus, this effect must also be included in the polarization potential. The induced polarization is, as usual, taken as a linear response to the electric field since the field is not very large.

In the spirit of deriving an ab initio potential to be used for liquid water simulations, all parameters of the NCC potential have been fitted to ab initio calculated data. We refer the interested reader to Ref. 72 where the ab initio calculations of interaction energies for various geometries of 250 water trimer and 350 water dimer configurations are discussed. We use the ab initio interaction energies of trimers of water molecules to fit the many-body parameters, namely in the parameterization of the locations of the induced dipole moments and the point charges, as well as the polarizability and the value q of the point charge. This set of electrostatic parameters was then used -unmodified- in the fit of the two-body potential to ab initio calculated interaction energies of water dimers.

One important characterization of the water dimer potential surface is its absolute minimum, i.e. the most stable water dimer configuration. This binding energy is computed as 5.18 kcal/mole at an intermolecular O-O distance of 2.97Å. The potential around this configuration is very flat. In comparing this data with experimental values, care has to be taken since the available experimental results span a rather broad range, particularly with respect to the binding energy.[81] Recent results[82,83]

seem to settle around 5.4 ± 0.7 kcal/mole; possibly 4.8 ± 0.4 kcal/mole is a more accurate estimate.

The NCC potential is obtained by fitting *ab initio quantum mechanical* interaction energies and makes *no use of empirical parameters*. The high quality of the NCC potential has been demonstrated[84,85] by the agreement with a large number of experiments. We recall the excellent agreement for the pair correlation functions, X-ray and neutron beam scattering, not only at room temperature, but for several temperatures, from hot water (T=361 K) to supercooled water (T=242 K). The NCC potential also yields good agreement with infrared, Raman and inelastic neutron scattering density of states for the translational and rotational motions as well as for the bending and stretching bands in the liquid,[86] in the above temperature range. In addition, the reliability of the NCC potential has been demonstrated not only for liquid water but also for deuterated liquid water[86]. Finally, the NCC potential yields an accurate second virial for the gas (from T=395 K to T=1195 K) and a good density of states in the solid state (ice Ih)[87] at several temperatures. Since, however, the quantum mechanical computations did not include *fully* the electron correlation corrections[72], some errors remain in the potential energy surface (up to a few tenths of a Kcal/mole in the dimer binding energy) and this manifests itself in a slightly too high binding energy with a consequent error in the simulated pressure[84].

In this work we make use of our very extensive data bank on MD computed trajectories, attempting to answer the question ''what is made liquid water of''. The trajectories have been stored, after equilibration, for 12.4 ps for a sample of 512 water molecules at T=242 K, T=268 K, T=284 K, T=328 K, T=361 K and for a sample of 1000 water molecules for T=305 K. All the MD simulations have been carried out with a time step of 0.125 femtoseconds with a cutoff equal to half of the simulation box. The reaction field method has been used to correct the error due to the truncation of the interaction potential and a sixth order predictor corrector to integrate the Newton equation of motion. We refer to Ref. 85 for a more detailed description of the simulation procedure and the interaction potential.

At the quantum mechanical level, liquid water is clearly made of nuclei and electrons, a decomposition which, while correct, is neverthenless not too useful till we describe liquid water with *quantum molecular dynamics* rather than with *classical molecular dynamics*. Thus, we could consider as ''objects'' constituting liquid water the hydrogen (or deuterium) and oxygen ''atoms'', where the quotation marks are added to stress that this description is accurate within the limits of classical molecular dynamics; thus, tunneling and zero order energy correction are neglected. Notice that this decomposition into ''atoms'' is the one adopted in our MD simulations and it yields the trajectories stored in our data bank; unfortunately this description is still too complex to provide in itself a ''simple model'' for liquid water. Attempting a further simplified answer to the question ''what are the species of which liquid water is made of'', we can

consider a decomposition of the liquid into water molecules (either H_2O or D_2O), however *not ''isolated''* molecules of water, *but ''solvated''* molecules of water. The need to consider for the liquid as ''basic particles'' not ''isolated water molecules'' as it is in the gas phase but ''solvated water molecules'' is a direct consequence of the MD findings, as summarized below.

Let us now look at *solvated water molecules* as the *building blocks* of which liquid water is made of. In the following, we shall extract from our trajectory data bank structural and dynamical information on the water molecules enclosed within a given volume, *the solvation shell volume*. To do this, it will be advisable to fine-tune the definition of ''solvation volume''. We can define the hydration shells either with reference to the minima of the pair correlation functions or by focusing attention on the coordination numbers; below we will analyze both alternatives. Later on we shall sharpen the analysis and ask questions on molecular patterns (clusters within the liquid) which have been defined by tracing the hydrogen bond networks.

As it is well known, the position of the first minimum of the oxygen-oxygen pair correlation function, $g_{OO}(r,T)$, can be used to define the value of the first solvation shell radius; equivalently, the second (third) minimum defines the second (third) solvation shell radius. In Fig. 1 (inset a) the $g_{OO}(r,T)$ are reported. Notice that i) the first minimum, $r_{min}(T)$, is *temperature dependent* (the higher the temperature, the larger the radius) and ii) that for T= 361 K the radius is about 3.43 Å. The second solvation shell has a minimum at R=5.61 Å, which is temperature independent. Notice also the existence of isosbestic points, particularly the one at about 2.8 Å and the one at about 3.8 Å (we shall return on these feature below).

In inset b of Fig. 1, we report the OH pair correlation functions, $g_{OH}(r,T)$. Notice that the first peak, centered around 1 Å, corresponds to the intra-molecular and the others to the inter-molecular correlation functions. Details on the intra-molecular O-H pairs are given elsewhere[3]. For $g_{OH}(r,T)$, the position of the first minimum is *almost temperature independent* and it occurs at R~ 2.45 Å.

As known, the integral from 0 to r of the $g_{\alpha\beta}(r,T)$ pair-correlation function defines the coordination number, $N_{\alpha\beta}(r,T)$ for a sphere of radius r. By choosing an oxygen atom as the sphere center, we can ask how many additional oxygen atoms can be found within a sphere of a radius r; equivalently, selecting again an oxygen atom at the origin, we can count the hydrogen atoms within the same sphere, this time integrating $g_{OH}(r,T)$. In Fig. 1, we report the coordination numbers N_{OO} (inset c) and N_{OH} (inset d) of the OO and OH pairs as a function of the temperature. We note that the same value of a coordination number is reached at very well defined values of R *independent from the temperature* (isosbestic point). Notice also that the isosbestic points for the coordination numbers *do not coincide with the isosbestic points found in the corresponding pair correlation functions*. From insert c of Fig. 1 we learn that for N_{OO} the first two isosbestic points occur at R=3.24 Å and R=4.50 Å, respectively. The first isosbestic point falls

in the neighborhood of the first minimum of $g_{OO}(r,T)$; the second isosbestic point occurs near the maximum of the second peak of $g_{OO}(r,T)$.

At R=3.24 Å and R=4.50 Å the values of $N_{OO}(r)$ are ~ 4.15 and ~ 12.0. The growth of the N_{OO} functions is relatively linear for high temperature, but has step-function behavior for low temperatures; this observation suggests the existence in the liquid of a *non-homogeneous "local" density*, which becomes, however, more and more homogeneous as the temperature increases.

By inspecting the growth of $N_{OH}(r,T)$ we notice that the first isosbestic point occurs at R=~ 2.52 Å with N_{OH}=1.02× 4= 4.08 OH pairs. (We recall that the total number of OH pairs, occurring within a sphere of radius R, is given by the N_{OH} value multiplied by four, since for any two water molecules, there are four *intermolecular* OH pairs, which might or might not be hydrogen bonds).

In Fig. 2 (inset a) we report $N_{OO}(r)$ and $N_{OH}(r)$ as function of T, using as shell radius the value of the first minimum of $g_{OO}(r,T)$ and $g_{OH}(r,T)$. From the figure we learn that the N_{OO} coordination number is a bit more than 4 in the range 240 - 305 K. Above the latter temperature value, the coordination number increases sharply, pointing to a structural reorganization in the liquid. It is interesting to notice that above room temperature, the coordination number indicates sizable probability of *five* water molecules coordinating a central one (rather than four). By integrating up to the first minimum of $g_{OH}(r,T)$, i.e. R= ~ 2.45 Å, we obtain, for the temperatures here considered, the N_{OH} values shown in the a inset of Fig. 2. We note that, as temperature increases, the two curves do not grow in a parallel way. Indeed, at high temperature, one oxygen atom is coordinated with 5 oxygen atoms and 3.80 hydrogen atoms, whereas at low temperature one oxygen atom is coordinated with 4 oxygen and 4 hydrogen atoms.

If liquid water would have a regular structure, for example with water tetrahedrally coordinated (via hydrogen bonds), then each oxygen would experience four oxygen and four hydrogen atoms. Thus, the above N_{OO} and N_{OH} values suggest that we should look for irregular structures or for other hydrogen-bonded aggregates, for example dimers, trimers, tetramers, pentamers and higher complexes.

In the past decades, much attention has been given to the structure of liquid water and on the role of the hydrogen bond (for recent literature, see for example Refs. 88-94). These papers advance a number of "non-unique" definitions for the hydrogen bond. According to some,[88] geometric constraints are assumed, but arbitrariness is introduced in the "h-bond" values of the distances and angles defining the mutual orientation between water molecules. For others[89–94] the hydrogen bond should be defined in terms of interaction energy between water molecules; but this definition too has its own shortcomings, since arbitrary choices are introduced in the definition of the interaction energy and in its cutoff. All those papers (Refs. 88-94), explicitly or implicitly, assume that each water molecule is tetra-coordinated (and deviation from this coordination is considered an anomaly); in addition, the coordination is obtained with reference to

oxygen atoms only, and this is assumed to constitute also a valid identification of hydrogen bonds.

However, from inset a of Fig. 2, we learn that we must distinguish between O-O and O-H interactions. Indeed, as already pointed out, the two coordination numbers do not grow in a parallel way upon temperature variation. In other words, we can not assume that if a O-O pair satisfies some distance conditions there is *necessarily* a hydrogen atom in between, forming a hydrogen bond. To refine the above reasoning with a more *quantitative* description, we first define a radius more suitable for our analysis and then we look more closely at the number of oxygen and hydrogen atoms within the shell volume.

Here, we now chose as shell *radius* the *R* value of the *first* N_{OO} *isosbestic point*, R = ~ 3.24 Å, which is somewhat larger than the radius of the first g_{OO} minimum for low temperatures and somewhat smaller for high temperatures. Being N_{OO} (R=3.24) *temperature independent, it eliminates one variable from our analysis.* We recall that at R=~ 3.24 Å, the N_{OO} coordination number is 4.15. This done, let us ask the question: considering a water molecule, how does its solvation number vary within 12 picoseconds (the time length of our simulation). In other words, *how many oxygen atoms can we detect within this solvation volume as function of time*, i.e. from the beginning to the end of our MD simulation. In inset b of Fig. 2 we consider a random molecule from our sample (for the simulation at T=305K) and we follow the N_{OO} coordination for a few picoseconds: at the beginning (t=0) it is surrounded by four neighbors, then by 3, by 2, by 3 and again by 4, all this within the first picosecond; as the simulation time proceeds, from a minimum of 1 to a maximum of 7 oxygen atoms surround that one taken as the solvation sphere center. In general, the ''lifetime'' (i.e. the time length with the *same* coordination value) of these ''clusters'' varies from a small fraction of ps to about 1 ps. Equivalent histograms can be obtained not only considering the oxygen atoms, but also the OH pairs. We recall that an ''OH pair'' does not necessarily imply a ''hydrogen bond''.

Let us now generalize and consider, for a given temperature, the total population of the water molecules surrounded by 2, 3, 4, etc. oxygen atoms within the solvation radius of 3.24 Å, and let us repeat this analysis for all the temperatures we have simulated. The resulting data (normalized) are reported in Fig. 3 (inset a). We can immediately notice that, for all the temperatures, the water molecules tetra-coordinated are the most abundant, but by increasing the temperature, this population decreases with a concomitant increase of water molecules with both lower and higher coordination numbers. This is equivalent to say that upon heating, the liquid, because of increasing diffusion, becomes more disordered with simultaneous presence of high density regions (coordination higher than 4) and low density regions (coordination 2 and 3).

From the value, 4.15, of N_{OO} coordination number at R=3.24 Å (for all the temperatures), one could conclude that each water is tetra-coordinated and thus propose a

continuous model for liquid water, valid for all temperatures; this conclusion, however, would not be in agreement with the data of Fig. 3. The conclusion that the population of the tetra-coordinated water molecules is the most abundant, could be a consequence of the value of R selected to define the hydration volume. By repeating the same kind of analysis, but with a different radius for the shell, specifically the one of the first minimum in the $g_{OO}(r,T)$, one obtains the data reported in inset b of Fig. 3. In this figure we notice that, except for T ~ 361 K, still the tetra-coordination is the most abundant. However, passing from low to high temperature, the coordinations higher and lower than four do no follow anymore the same patterns as in inset a of Fig. 3: for example, above room temperature the penta- and hexa-coordinations start to be more and more abundant.

We now consider the time stability of the cluster distributions, defined as the average number of consecutive time steps the same water molecules belong to a given cluster. This is reported in Fig. 4 for the various temperatures and coordination numbers for the OH clusters. The tetra-coordinated waters not only are the most abundant, but also the most long living. In addition, we learn that by increasing the temperature the lifetime decreases as a consequence of the thermal motion. Notice also that for the tetrahedrally coordinated molecules, as temperature approaches the supercooled values, the life time is large but still much shorter than in ice. We have not reported the decay diagrams for the OO clusters, since those present the same trends given in Fig. 4.

Each water molecule is both a ''solvent molecule'' and a ''solute molecule''; at the same time, each water molecule is part of a network. Indeed, we can view the solvation shell as a time averaged representation, resulting by considering a given water molecule as fixed in space with its neighboring molecules librating around it; the librations are, however, motions which are transmitted through the network.

If we consider only a small solvation volume, for example the first hydration shell, then all what we can see are ''librations'' of the solvating water molecules, hydrogen-bonded to the central water molecule. However, if we consider a larger volume, corresponding to water molecules strongly hydrogen-bonded, then we would see that the ''libration'' of one molecule is coupled to the libration of a nearest molecule, with the coupling being stronger along the networked pathway. In this larger volume the *network representation becomes the characterizing feature*, whereas in a small volume, *the solvation is the main aspect*. Strongly coupled librations extend over a few water molecules and appear as localized ''collective motions'' over this limited pathway. Clearly, there are many pathways, each one with its ''localized motions''. The motions are not in phase and thus have strong positive or negative interferences. These interferences build up (or destroy) micro-regions of ''ordered material'' and in so doing create new patterns, which modify the network, some of the patterns can be interpreted in terms of cyclic structures. In the following we shall attempt to show that this is the model derived from our MD trajectories.

We now focus our attention on specific pathways formed by hydrogen bridges, i.e. ''structures'' which can be obtained by following a specific path once some pre-defined rules and constraints are selected.

At each time step in our simulation, each water molecule belongs to a ''cyclic structure'', which is defined by the following rules: Starting with one of the two OH bonds, one moves to the nearest hydrogen bridged water, then to the next one, satisfying the condition that among different pathways one selects the one which will close the pathway on the starting water molecule. Notice that ''by construction'' we search for ''cyclic structures'' purely on geometrical grounds, neglecting both energetics and life-time conditions (for this reason, we have used the old notation ''hydrogen bridge'' rather than ''hydrogen bond''). In the above selection rule, the operational definition for the existence of a hydrogen bridge is that a hydrogen atom must exist between two oxygen atoms, with a maximum intermolecular distance of 2.45 Å, the distance of the first minimum in the g_{OH}.

With the above definition, however, one might generate large polygons, which often will contain small polygons as substructures. To our knowledge, two algorithms have been proposed in literature to analyze the network topology in terms of polygon statistics. One algorithm[89] focus attention on ''non-short circuited'' polygons, and the other[92] on ''primitive'' polygons. For details between the two algorithms we refer to Refs. 89 and 92.

In the search and count of the polygons, we adopted two definitions, the first one counts *all* possible closed circuits, eliminating however those large polygons which contain all the atoms of small ones. The second definition, coincides with the ''primitive'' polygon definition of Ref. 92. We call the two ''ways of counting'' N_r and N_p (for redundant, r, and primitive, p) polygons, respectively. Notice that, while, the N_r definition is somewhat redundant, the N_p is not unique. Notice in addition that, when we apply both definitions to water molecules with configuration as in ice Ih, limiting our analysis up to polygons with dimension 8 (octamers) we obtain hexagons and octagons with the N_r definition; whereas with the N_p definition we find only hexagons.

In Figure 5 we report the number of cyclic structures $(H_2O)_n$ with n=3, 4, etc. relative to the number of water molecules, N_W, in the sample. The top inset of Fig. 5 refers to the ''redundant'' definition and the bottom inset to the ''primitive'' definition. By construction, the number of trimers and tetramers coincide in the two definitions. In the top inset of Fig. 5 one can notice that for n=3 and 4, the ratio of N_i/N_W increases with temperature, but from n=5 to n=8, the opposite holds; in addition, for n=8, at low temperature, the ratio is larger than 1. Both observations are related to the fact that the water molecules of a given $(H_2O)_n$ cyclic polymer can participate to more than one polygon.

We notice that the trimers have the largest O-O distances, followed by the tetramers and then by the larger polygons. Whereas for the trimers the distance remains almost constant for all the temperatures, the O-O distance of the larger polymers have a

strong temperature dependence, which approaches the value of ice. Comparing these distances with those of the clusters in gas phase[95], we find for the trimers in the liquid an average O-O distance of 2.90-2.92 Å, whereas in the gas phase the distance is of 2.96 Å. Larger differences are found for the other clusters: in the gas phase the distances are in the range 2.95-2.97 Å, in the liquid phase 2.77-2.87 Å.

In conclusion, a molecule of water in the liquid can be considered as the node (site) for one or more hydrogen bonds; our model yields about four (a bit less for high temperature) hydrogen atoms in the first solvation shell surrounding a central oxygen atom. However, our model predicts from five (high temperature) to four (low temperature) oxygen atoms coordinating a central one in the first shell. These are average distributions; a closer analysis yields a distribution of two-, three-, etc., up to seven atoms coordinated to the central one. The most abundant are the tetra-coordinated and the penta-coordinated structures: at low temperature the former are the most abundant, but at high temperature, the latter become equally important or even somewhat predominant, depending on the specific choice of the solvation shell radius. If, however, we focus attention on the O-H pairs (rather than the O-O pairs), then the trimers appear to be as important as the pentamers. From our trajectories we learn that the population of the tetra-coordinated water molecules approaches the total population as the temperature approaches the lowest supercooled temperature: this indicates that at the lowest supercooled temperature liquid water is a fully connected network. The significance of these findings is enhanced by our finding that the above structures persist for time scales much longer than water vibrational frequencies and vary from a few hundreds to a few tenths of a picosecond, longer for the tetra-coordinated structures, shorter for the others. There are also indications that the decay time requires a mechanism involving two processes, a short (with high population) and a slow one (with low population).

The analysis of the hydrogen bonded networks brings about a search for closed circuit hydrogen bonded substructures (polygons). This analysis must contend with non uniqueness in the polygon definition. In turn, different definitions bring about largely different substructure population. From this work and also from previous findings available in literature[88–92] we can identify triangular, quadrilateral, etc. polygons; the pentagonal structures are the most abundant. From our work we learn that, with exception of the triangular structures, the O-O distance is temperature dependent and tends to the gas phase values for high temperatures, whereas, at low temperatures it approaches values found in ice. The computed lifetime values are very close for the different polygons and are of the order of a few hundreds of a picosecond. Notice that in the lifetime analysis of solvation complexes reported above each molecule of water contributes to the statistics; however, for the polygons this is not the case and as a consequence our statistics can be used only for preliminary predictions.

A second conclusion relates to the methodology. Starting with ab initio quantum chemistry we have developed more and more refined interaction potentials[73–79]. These

are our input for MD simulations, which accurately predict a vast number of data obtainable also from laboratory experiments. *But our simulations yield also a vast library of trajectories, experimentally not attainable*: this library is the input needed to obtain ''a model'' -even if preliminary- of the most interesting liquid on earth.

7. The Time Scale Problem

Molecular dynamic computations are now extensively used to refine the experimental X-ray or NMR structure and to calculate the free energy differences which are essential to a correct evaluation of binding equilibria and the changes introduced by site-specific mutagenesis. The success of MD simulations largely depends on the accuracy of the interaction force fields, and the ab initio quantum mechanical calculations have played a critical role in the determination of these interaction force fields[96,97].

In this section we shall concentrate on the effect of the solvent molecule on protein dynamics.[98,99] Due to the complexity of the force field needed to describe the interaction between the protein atoms, protein-water molecules and water-water molecules and due to the fact that the time steps used in the MD simulations are of the order of a few femtoseconds, in general, the total duration of the molecular simulation lies in the range of 10-200 picoseconds. To describe the protein dynamics in the nanoseconds and microsecond time domain and beyond, we have to seek new simulation strategies which can transcend the limitations inherent in the molecular dynamic approach. One possible alternative is the stochastic dynamic simulation (SDS methods)[100–106] wherein the dynamic evolution of the biomolecule is studied in hierarchy of Langevin equations, on a time scale which gets progressively coarser. In SDS the effect of the dynamic modes not considered explicitly are taken into account by suitably defining friction and random forces. Following Allen[104] and van Gunsteren and Berendsen,[105] the random force R, is presumed to have a white-noise character and thus the frictional force depends on the instantaneous velocity of the particle involved.

Below we shall discuss one particular application, a study of the dynamic motion in a small protein, the bovine pancreatic trypsin inhibitor. In recent years several experimental, theoretical and molecular dynamic studies on the structure and the dynamics of the BPTI protein have been reported[96,106–111]. Three 25-ps molecular dynamics studies by van Gunsteren and Karplus detailed the dynamics of the BPTI in vacuo, in a nonpolar solvent, and in a crystalline environment[106]. Smith et al. analysed the incoherent neutron scattering in a BPTI molecule,[110] Levitt and Sharon conducted a 300-ps molecular dynamic simulation and studied the way the BPTI protein changes the properties of the surrounding water molecules.[111] Cusak et al. measured the time-of-flight spectra and the generalized density of states in the BPTI

protein.[108–110,112] Clementi et al. conducted two 52-ps molecular dynamic simulations[96] of the BPTI protein in vacuo and in solution and calculated the density of states and time-of-flight spectra and compared them with the experimental and theoretical results obtained in the framework of the normal mode analysis. In the following, we discuss a stochastic dynamic simulation of the BPTI protein. To calculate the interaction force between protein atoms we use the potential force field by Weiner *et al.*, [113] where the hydrogen bond is represented by an explicit term.

The results of the stochastic simulations are compared with the molecular dynamic simulation of BPTI in water[96] where the dynamic evolution of 892 atoms of BPTI and 2676 water molecules was studied for 52ps.

To initiate the stochastic dynamics simulation of the BPTI protein, first we have to determine the friction coefficient γ_i of 892 atoms constituting the BPTI protein. In our study, we used the molecular dynamic trajectories to estimate the friction coefficients. First we calculate the diffusion coefficient of each of the 58 residues constituting the BPTI molecule and then define the friction coefficient of each atom, assuming that the atomic diffusion coefficient is equal to the one of the corresponding amino-acid residue.

The stochastic dynamics simulation was conducted in a cube of 120.0 x 120.0 x 120.0 $\AA^3$. The temperature was set to 300 K and the time step of the simulation was 0.5 fs. The 8192 configurations were collected during a 32 ps simulation. On an IBM-3090 computer, the CPU time per time step for the stochastic dynamic simulation was approximately 2.95 sec. The molecular dynamic simulation of the BPTI protein in 2676 water molecules took 48 sec. per time step, while for the molecular dynamic simulation in vacuo the CPU time per time step was 2.69 sec. Thus, the stochastic dynamic simulation was approximately 16 times faster than the molecular simulation in solution. In the stochastic dynamic simulation the average RMS deviation was around 0.8Å which is considerably less than the one found in the molecular dynamic simulation[96] in vacuo (2.7 Å) but close to the RMS deviation for the molecular dynamic simulation in solution[96] (~ 1.2 Å).

When we compared the MD and SDS trajectories by analyzing the incoherent neutron scattering spectra and the generalized density of states, we found that our stochastic dynamic simulation results are in reasonable agreement with the molecular dynamics results. The shape of the curve is also consistent with the available experimental scattering intensities from the powder sample[108]. In the stochastic simulation the protein atoms experience the viscous damping due to the presence of solvent atoms. This seems to enhance the low frequency modes in the density of states spectra. This is in agreement with the results predicted by Smith *et al.*[112] who argue that small conformation modifications of protein due to changes in surface interaction upon addition of water modifies the mode spectrum such that there is a net shift form higher to lower frequencies.

Our simulations indicate that aspects of the internal motions of the BPTI protein can be obtained by using the stochastic dynamic simulation techniques which are one order of magnitude faster than the molecular dynamics method. Here, it would be appropriate to remark that the methodology employed in the present SDS simulation is a realization of the simulation strategy alluded in the ''global simulation technique''[114,115]. One of the characteristic features of the global simulation technique is the existence of an inherent connectivity in the different computation models describing the natural processes at different time scales, and this is sketched in Fig. 6. Thus in the case of protein dynamics, the determination of the potential force field from the quantum mechanics methods constitutes the first step in the global simulation strategy, the second step is performing the molecular dynamics simulation and the determination of the friction coefficients from the molecular dynamical simulation, marks the beginning of the third stage of the global simulation approach.

8. Global Simulation

Ideally we demand from computational chemistry the ability to realistically simulate ''any'' problem at ''any'' desired accuracy. Well, this sounds a bit like ''Alice in Wonderland'', but this - however hard and far - is exactly the goal of computational chemistry. Today there are many methods and within each method several approaches or techniques. Today the computer power is notable and easily available, especially if compared with the past. Thus, it is finally possible to realistically attempt to experiment with the ''global simulation approach''. The main idea is that any problem, no matter how complex, can be subdivided into soluble subproblems and the output of one subproblem is all we need (i.e. ''the full input'') to solve the next subproblem and so on. Thus we could realize the equivalent of an ''assembly line'', solve any problem and increase our productivity. Clearly, today we are only at the beginning, a bit like where England was at the beginning of the first industrial revolution. But knowledge processing, artificial intelligence, expert systems will help computational chemistry in the same way as production lines, which nowadays are equipped with robots or even ''intelligent robots'' and not only ''blue collar workers''. For this reason in this work, we have often referred to the 1996-1998 period and to the Nike computer.

The road map of the assembly line for chemistry and material sciences is given in Fig. 6. It is part of a more complex map, which one of us (E.C.) has described elsewhere[37,115]. From ab initio (Hartree-Fock, HF, or many body, MB) or from semi-empirical, SE, or density functional theories, DFT, we can compute interaction energies, fit the data to appropriate interaction potentials and use these in classical molecular dynamics. Alternatively, we can use quantum mechanics at each time step of the molecular dynamics simulation.

The first attempt in this direction is the one undertaken by one of us (E.C.) and co-workers in 1978 in a study of the papaine reaction. There, we realized the need for two scale of motions ''atomic'' motion for the protein and reagent far from the reaction site and ''electronic'' motions needed to describe bond breaking and bond formation[116]. The former ''macro-motion'' was obtained with a 1978 molecular dynamics approach (for the study of proteins) the latter ''micro-motion'' was obtained using a quantum chemistry ab initio program. The computations was however extremely heavy for the 1978 computers. A subsequent physically similar but more refined approach is the one by Car and Parrinello[65] yielding what we call ''quantum molecular dynamics''. Presently, the Car-Parrinello is essentially a semi-empirical method, since it is empirically parametrized both in the density functional and in the pseudo-potential. Also the ''fictitious mass'' is an empirical parameter; hopefully, in time, the method will become ''truly ab initio''.

In Fig. 6 we have assumed an expansion and growth for the Fock-Dirac based methods equivalent to the one we have already experienced for the Hartree-Fock based methods. We think, that this new development will be basic for the understanding of such fenomena as catalysis, materials with special optical, magnetic or electronic properties, new drugs, etc. The bottom third of the periodic table is an unacceptable limitation! A very approximate rule of thumb, approximates to a factor of 16 the slow down of a Fock-Dirac relative to a Hartree-Fock computation. This factor arises because of the increase in the basis set due to the kinetic balance approximation, and partly because of the overall heavier algorithm.

In Fig. 6 we have also explicitly indicated the third avenue mentioned in the introduction, i.e. progress in reaction rate computations. The new laws controlling the air quality should be noted in this regard. *An ab initio flame*, the simplest one like O_2 with H_2, is within the feasibility of today's methods, software and hardware.

Finally, in Fig. 6 we have shown the links between the scientific methods and other methodologies indicated as A.I. (artificial intelligence), N.N. (neural nets), factor analysis, etc. In Ref. 115, we have shown that a reasonable starting point for annealing in protein folding can be provided by a N.N. search. Of course, there is a most notable difference between ''starting point'' and ''solving the problem''. Concerning the impact of the Vth generation concept on chemistry, we suggest to look at Ref. 117.

We have discussed how to connect molecular dynamics to stochastic dynamics: clearly much remains to be done, to truly deal with the different time scales needed in chemistry. We recall that the connection to fluid dynamics is obtained via microdynamics but also by using cellular automata techniques : MOTECC-91 contains chapters on these topics and MOTECC-Club, Inc. is distributing the corresponding computer programs.

9. Conclusions

In this work we have reported timings obtained on different areas of computational chemistry, all using computers with essentially the same speed, either an IBM-3090/600J or an IBM-9000/720, both with vector features. Then, we have extrapolated to more complex systems obtaining ''horrifying forecast'' which, however, were nicely divided by 5000, the factor between the IBM-9000/720 and Nike our ''dream computer of 1996-1998''. Perhaps we can look at computational chemistry in a more even-handed way by focusing on the complexity of various methods and techniques.

Let us consider a system of N electrons described by M_c contracted functions and M_P primitive gaussian functions. A semi-empirical package, like AMPAC, grows as N^2, thus one can perform medium size computations (for examples 3 Guanine-Cytosine base pairs, GC, i.e. 87 atoms and 408 electrons) very confortably on today's workstations. Certainly here we do not need neither Nike nor an IBM-9000. However, a Hartree-Fock type computation with its M_P^4 complexity becomes a bit more demanding. Indeed, for a (9,5/4,2) double zeta basis set, M_c=630, and the value of M_P=1548 brings about 2×10^{10} integrals over primitive functions. (Actually many integrals are vanishingly small and only about 4×10^7 are above a small threshold; a $M_P^{3.08}$ rather than $M_P^{4.0}$ dependence. The computer time is of the order of 10 hours on a dedicated IBM-9000/720 used in parallel or about one hundreth of an hour on our Nike!

In between the N^2 and the N_P^4 complexity, there is the $\sim N_P^3$ complexity of the density functional approaches with gaussian basis sets. Naively, one might assume that the 3 GC base pair computation above mentioned should required about (10/1548) hours, i.e. a fraction of a second on the dedicated IBM-9000/720. Unfortunately this is not the case: in the density functional computation there is a prefactor, S, much larger than unity. Thus, the scaling is not from 1548^4 to 1548^3 but to $(S\times1548)^3$. It is for this reason that density functionals are often used in connection with pseudopotentials to reduce N and thus M_P.

This is only one of the limitations of the density functionals, a more fundamental one is in the ''incorrect'' ratio between the exchange and the Coulomb energy and in the inability to satisfy the virial theorem. We recall that all the vector coupling coefficient theoretical development is simply ignored in today's density functional approaches, particularly those where the exchange and the correlation energies (both the potential and the kinetic contributions) are lumped together. This approximation, acceptable in the 1950's (the beginning of the $X\alpha$ approximation) somewhat less acceptable in the mid 1960's (the beginning of Kohn-Sham applications), has little reasons to survive in the 1990's. It seems abundantly clear that a much superior physical description is obtained by solving the Hartree-Fock problem correctly and by adding to it the correlation energy obtained by density functionals. There is a simple variant: one can start either with a Hartree-Fock or with a very short multi-reference set which ensures a correct reference state[118–120].

Alternatively, one can construct one-electron functions from an effective potential which sums to the standard Hartree-Fock potential a correlation term; the latter can be either one of the density functionals available in literature (with or without gradient corrections) or some functional form related to the Coulomb hole, for example terms of the type $\exp[-f(\rho)r_{ij}^2/r_{ij}]$. Previously, $f(\rho)$ has been parametrized in a rather cumbersome way not explicitly as a function of the density. Notice also that the exponential term brings about matrix elements for which the analytical solution is known. Such Coulomb hole terms should correct each one of the Hartree-Fock Coulomb terms with a weight given by the standard Hartree-Fock vector coupling coefficients. This technique yields reliable values for the ground and the excited state[37,121,122] while retaining a N^4 complexity.

ACKNOWLEDGEMENT

Support from ''Regione Autonoma della Sardegna'' is kindly acknowledged. The authors whish to thank Drs. Murgia, Paddeu, Paglieri and Rizzo for making available the unpublished results on the Zinc atom.

REFERENCES

1. M.H. Kalos Phys. Rev. **128**, 1791 (1962); D.M. Ceperley, B.J. Alder, Science **231**, 555 (1986); J. Anderson, J. Chem. Phys. **85**, 2839 (1987), and references therein.
2. *Modern Techniques in Computational Chemistry: MOTECC-90*, E. Clementi, Ed., ESCOM, Leiden, 1990. See for example the chapter by M. Parrinello, pp 731.
3. D.K. Bhattacharya, E. Clementi, *Modern Techniques in Computational Chemistry: MOTECC-91*, E. Clementi, Ed., ESCOM, Leiden, 1991, pp 939.
4. A.M. De Callatay, *Natural and Artificial Intelligence*, North Holland, Amsterdam, 1986.
5. E.A. Hylleraas, Z. Physik **54**, 347 (1929).
6. R. Bartlett, Ann. Rev. Phys. Chem., **32**, 359 (1981); M. Urban, I. Cernusak, V. Kellö, J. Noga, *Methods in Computational Chemistry*, S. Wilson, Ed., Plenum, New York, 1987.
7. *Note by EC*: During the past thirty years of computational quantum chemistry, many approximations have been proposed, often with much overclaiming (to me the birth-day of computational chemistry is the Bulder, Colorado meeting in 1959, where the first systematic ''computer results'' for atoms and molecules were presented (see the special volume of Rev. Mod. Phys. **32** (1960)) and where the *ab initio* trend emerged, well differentiated from semi-empirical approaches). The overclaims were of two kinds: on the accuracy of the proposal and on its applicability to ''chemistry''. The first over-claim is -all in all- understandable. The second reveals an inability to extrapolate into time a process which is N^m dependent. From the early 1950 and especially in the 1960 it was clear that the computer industry could essentially deliver a new computer every four years, three times as fast (at least) and essentially at the same price. This type of trend clearly favors a N^2 or N^3 complexity algorithm and places a hard burden on a N^4 algorithm; it simply wipes out any N^7, or higher complexity, algorithm, unless N is either very small (today less than 50 electrons), or exceedly small (2 to 4 electrons) for N^8 and higher complexity. A second line of progress was in the algorithm. Unfortunately, there is no careful study on the speed up due to algorithms, since most of the published ''timing'' are for statistically insignificant samplings. My view point was that we must ''hold the line'' at the N^4 complexity even if this complexity is -admittedly- an upper limit, but an attainable one. This leaves correlation corrections *necessarily* to techniques that are no-more complex than N^4. Others assumed that N^4 is too high a limit and held their ''maximum'' at N^3; notable among them is the strong voice of solid state physics (Prof. J.C. Slater would most likely be among them). Still others assumed the N^2 as a ''maximum complexity'' and in this group we have most of the ''semi-empirical approaches''. Finally, others placed their trust in N^5–N^7 techniques, most useful, indeed essential, especially to test simpler approaches, but not really ''chemistry oriented''.

   In 1960 we computed a SCF-LCAO-MO function for C_2 using symmetry (Gazz. Chim. It. **91**, 717 (1961)); in 1990 we computed a double-zeta + polarization SCF

function for C_{60} without use of symmetry (G. Corongiu and E. Clementi, Int. J. Quantum Chem. **42**, 1185 (1992)). The elapsed time for C_{60} was ~ 24 hours (running in parellel mode on a IBM-9000/720). Our new computer code can compute the same function using I_h symmetry in a few hours. For C_{60} we have used a IV generation computer (a IBM-9000/720); thus with our supercomputer Nike, we could compute C_{60} without symmetry in minutes, and with symmetry in seconds: this points out strongly that a N^4 strategy with use of density functionals, limited to correlation effects, *was and is very reasonable*.

8. R.S Mulliken, *Spectroscopy, Molecular Orbitals, and Chemical Bonding*, Nobel Lecture, 1966.
9. H.M. James, A.S. Coolidge, J. Chem. Phys. **1**, 825 (1933).
10. W. Kolos, L. Wolniewicz, J. Chem. Phys. **41**, 3663 (1964); *ibid.* **43**, 2429 (1965); Phys. Rev. Lett. **20**, 243 (1968).
11. R.S. Mulliken, W.C. Ermler, *Diatomic Molecules: Results of Ab Initio Calculations*, Academic, New York, 1977.
12. H.A. Bethe, E.E. Salpeter, *Quantum Mechanics of One-and Two-Electron Atoms*, Plenum, Rosetta, New York, 1977.
13. W. Kutzelnigg, Theor. Chim. Acta **68**, 445 (1985).
14. A. Largo-Cabrerizo, E. Clementi, J. Comp. Chem. **8**, 1191 (1987).
15. D. Frye, A. Preiskorn, G.C. Lie, E. Clementi, *Modern Techniques in Computational Chemistry: MOTECC-90*, E. Clementi, Ed., ESCOM, Leiden, 1990, pp 535.
16. D. Frye, A. Preiskorn, E. Clementi, J. Comp. Chem. **12**, 560 (1991).
17. C. Urdaneta, A. Largo-Cabrerizo, J. Lievin, G.C. Lie, E. Clementi, J. Chem. Phys. **88**, 2091 (1988).
18. D. Frye, G.C. Lie, E. Clementi, J. Chem. Phys. **91**, 2366 (1989).
19. D. Frye, G.C. Lie, E. Clementi, J. Chem. Phys. **91**, 2369 (1989).
20. A. Preiskorn, G.C. Lie, D. Frye, E. Clementi, J. Chem. Phys. **92**, 4941 (1990).
21. A. Preiskorn, G.C. Lie, D. Frye, E. Clementi, J. Chem. Phys. **92**, 4948 (1990).
22. A. Preiskorn, D. Frye, E. Clementi, J. Chem. Phys., **94**, 7204 (1991).
23. D. Frye, G.C. Lie, E. Clementi, *HCI Calculations on the Potential Energy Curve for the* $X\,^1\Sigma_g^+$ *State of* H_2; IBM Technical Report KGN 179; Kingston, 1989.
24. W. Kolos, K. Szalewicz, H. Monkhorst, J. Chem. Phys. **84**, 3259 (1986).
25. L. Salmon, R. Poshusta, J. Chem. Phys. **59**, 3497 (1973).
26. F. Mentch, J. Anderson, J. Chem. Phys. **74**, 6307 (1981).
27. A. Preiskorn, W. Woznicki, Mol. Phys. **52**, 1291 (1984).
28. P.G. Burton, E. Von N. -Felsobuki, G. Doherty, M. Hamilton, Mol. Phys. **55**, 527 (1985).
29. W. Meyer, P.G. Burton, P. Botschwina, Pbu, J. Chem. Phys. **84**, 891 (1986).
30. J. Anderson, J. Chem. Phys. **86**, 2839 (1987).
31. C.A. Traynor, J. Anderson, Chem. Phys. Lett. **147**, 389 (1988).

32. S. Huang, Z. Sun, W.A. Lester, Jr. J. Chem. Phys. **92**, 597 (1990).
33. S.A. Alexander, H. Monkhorst, R. Roeland, K. Szalewicz (private communication).
34. B. Liu, J. Chem. Phys, **80**, 581 (1984).
35. E. Davidson, *Modern Techniques in Computational Chemistry: MOTECC-90*, E. Clementi, Ed., ESCOM, Leiden, 1990, pp 553.
36. E. Clementi, G. Corongiu, S. Chakravorty, *Modern Techniques in Computational Chemistry: MOTECC-90*, E. Clementi, Ed., ESCOM, Leiden, 1990, pp 346.
37. E. Clementi, Int. J. Quantum Chem. **42**, 527 (1992).
38. A. Rizzo, E. Clementi, M. Sekiya, Chem. Phys. Lett. **177**, 477 (1991).
39. V.M. Umar, C.F. Fischer, E.R. Davidson, S.A. Hagstrom, S.J. Chakravorty, Phys. Rev. (1991).
40. F. Sasaki, Int. J. Quantum Chem. **8**, 605 (1974).
41. F. Sasaki, M. Yoshimine, Phys. Rev. A **9**, 17 (1974).
42. F. Sasaki, M. Yoshimine, Phys. Rev. A **9**, 26 (1974).
43. F. Sasaki, M. Sekiya, T. Noro, K. Ohtsuki, T. Osanai, *Modern Techniques in Computational Chemistry: MOTECC-90*, E. Clementi, Ed., ESCOM, Leiden, (1990), pp 181.
44. M. Sekiya, T. Noro, K. Ohtsuki, F. Sasaki, A. Rizzo, E. Clementi, *Modern Techniques in Computational Chemistry: MOTECC-90, Input/Output Documentation.* Special publication by the Department of Scientific/Engineering Computations, IBM Kingston, 1990.
45. E. Clementi, J. Chem. Phys. **38**, 2248 (1963).
46. E. Clementi, A. Veillard, J. Chem. Phys. **44**, 3050 (1966).
47. C.L. Pekeris, Phys. Rev. **112**, 1649 (1958); *ibid.* **115**, 1216 (1959).
48. T. Kinoshita, Phys. Rev. **108**, 1490 (1957).
49. A. Veillard, E. Clementi, J. Chem. Phys. **49**, 2415 (1968).
50. D. Feller, C.M. Boyle, E. Davidson, J. Chem. Phys. **86**, 3424 (1987).
51. C.F. Bunge, Phys. Rev. A **14**, 1965 (1976).
52. H. Hartmann, E. Clementi, Phys. Rev. **133**, 1295 (1964).
53. G. Murgia, G. Paddeu, L. Paglieri, A. Rizzo, E. Clementi (in preparation).
54. J. Rys, PhD. Thesis, University of New York at Buffalo, 1978.
55. E. Hollauer, M. Dupuis, *Elimination of undesired functions in the Rys-King algorithm to obtain SALCAO*, unpublished reports.
56. M. Dupuis, H.F. King, Int. J. Quantum Chem **11**, 613 (1977).
57. M. Dupuis, H.F. King, J. Chem. Phys. **68**, 3998 (1978).
58. P. Lazzeretti, R. Zanasi, E. Rossi, Chem. Phys. Letters **75**, 392 (1980).
59. D. Hoffmann, E. Hollauer (in progress).
60. J. Almlof, P.R. Taylor, *Advanced Theories and Computational Approachs to the Electronic Structure of Molecules*, C.E. Dykstra, Ed., Reidel, Dordrecht, 1984.
61. K.S. Pitzer, E. Clementi, J. Am. Chem. Soc. **81**, 4477 (1959).

62. J.R. Health, S.C. O'Brien, R.F. Curl, H.W. Kroto, R.E. Smalley, Com. on Cond. Matt. Phys. **13**, 119 (1987).
63. H.P. Luthi, J. Almlof, Chem. Phys. Lett. **135**, 357 (1987).
64. M.D. Newton, R.E. Stanton, J. Am. Chem. Soc. **108**, 2469 (1986).
65. R. Car, M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
66. B.P. Feuston, E. Clementi, W. Andreoni, M. Parrinello, Phys. Rev. Letters (1991).
67. G.B. Bachelet, D.R. Hamann, M. Schulter, Phys. Rev. B **26**, 4199 (1982).
68. J.P. Perdew, A. Zunger, Phys. Rev. B **23**, 5048 (1981).
69. G. Corongiu, E. Clementi, Int. J. Quantum Chem. **42**, 1185 (1992).
70. E. Brendals, B.N. Cyvin, J. Brunvoli, S.J. Cyvin, Spectro. Lett. **21**, 313 (1988).
71. K. Laasonen, F. Csajka, M. Parrinello, Chem. Phys. Letters **194**, 172 (1992).
72. a) U. Niesar, G. Corongiu, M.-J. Huang, M. Dupuis, E. Clementi, Int. J. Quantum Chem. Symp. **23**, 421 (1989); b) U. Niesar, G. Corongiu, M.-J. Huang, M. Dupuis, E. Clementi, IBM Technical Report KGN-191, Kingston, 1989.
73. U. Niesar, G. Corongiu, E. Clementi, G.R. Kneller, D.K. Bhattacharya, J. Phys. Chem. **94**, 7949 (1990).
74. O. Matsuoka, E. Clementi, M. Yoshimine, J. Chem. Phys. **64**, 1351 (1976).
75. E. Clementi, G. Corongiu, Int. J. Quantum Chem. Symp. **10**, 31 (1983).
76. J.H. Detrich, G. Corongiu, E. Clementi, Chem. Phys. Lett. **112**, 426 (1984).
77. M. Wojcik, E. Clementi, J. Chem. Phys. **84**, 5970 (1986).
78. M. Wojcik, E. Clementi, J. Chem. Phys. **85**, 3544 (1986).
79. M. Wojcik, E. Clementi, J. Chem. Phys. **85**, 6085 (1986).
80. J.O. Hirschfelder, C.F. Curtiss, R.B. Bird, *Molecular Theory of Gases and Liquids*, John Wiley, New York, 1954.
81. P. Obza, R. Zahradnik, Chem. Rev. **88**, 871 (1988).
82. T.R. Dyke, K.M. Mack, J.S. Muenter, J. Chem. Phys. **66**, 498 (1977).
83. L.A. Curtiss, D.J. Frurip, M.L. Blander, J. Chem. Phys. **64**, 1120 (1982).
84. G. Corongiu, Int. J. Quantum Chem. **42**, 1209 (1992).
85. G. Corongiu and E. Clementi, J. Chem. Phys. (in press).
86. G. Corongiu and E. Clementi, (to be published).
87. F. Sciortino and G. Corongiu, (to be published).
88. M. Mezei and D. L. Beveridge, J. Chem. Phys. **74**, 662 (1981).
89. A. Raman and F. H. Stillinger, J. Am. Chem. Soc **95**, 7943 (1973).
90. F. Stillinger, Science **209**, 451 (1980).
91. F. H. Stillinger and T. A. Weber, J. Phys. Chem. **87**, 2833 (1983).
92. R. J. Speedy, J. D. Madura, and W. L. Jorgensen, J. Phys. Chem. **91**, 909 (1987).
93. F. Sciortino and S. Fornili, J. Chem. Phys. **90**, 2786 (1989).
94. P. Mausbach, J. Schnitker, and A. Geiger, J. Tech. Phys. **28**, 67 (1987).
95. H. Kistenmacher, G. C. Lie, H. Popkie, and E. Clementi, J. Chem. Phys. **61**, 546 (1974).
96. M. Aida, G. Corongiu, E. Clementi, Int. J. Quantum Chem. **42**, 1353 (1992).

97. R. Levy, R. Sheridan, J.W. Keepers, J.S. Dubey, S. Swaminathan, M. Karplus, Biophys. J. **48**, 509 (1985).
98. P. Procacci, G. Corongiu, E. Clementi, IBM Technical Report KGN-195, 1989.
99. D. Bhattacharya, W. Xue, E. Clementi, Int. J. Quantum Chem. **42**, 1397 (1992).
100. D.L. Ermak, J.A. McCammon, J. Chem. Phys. **69**, 1352 (1978).
101. S. Alison, R. Austin, M. Hogan, J. Chem. Phys. **90**, (1989).
102. K. Sharp, R. Fine, B. Honig, Science **236**, 1460 (1987).
103. G.C. Ciccotti, E. Guardo, G. Sese, Mol. Phys. **46**, 875 (1982).
104. M.P. Allen, Mol. Phys. **47**, 599 (1982).
105. W.F. van Gunsteren, H.J.C. Berendsen, Mol. Simulations **1**, 173 (1988).
106. W.F. van Gunsteren, M. Karplus, Biochemistry **21**, 2259 (1982).
107. J. Smith, S. Cusak, U. Pezzeca, B. Brooks, M. Karplus, J. Chem. Phys. **88**, 3636 (1986).
108. S. Cusak, J. Smith, B. Tibor, J.L. Finney, M. Karplus, J. Mol. Biol. **202**, 903 (1988).
109. S. Cusak, J. Smith, J.L. Finney, M. Karplus, J. Treewhella, J. Physica B **136**, 256 (1986).
110. J. Smith, K. Kuczera, B. Tibor, W. Doster, C. Cusak, M. Karplus, Physica B **156**, 437 (1989).
111. M. Levitt, R. Sharon, Proc. Natl. Acad. Sci. USA, **85**, 7557 (1988).
112. J. Smith, K. Kuczera, P. Poole, J. Finney, J. Biomol. Struc. Dyn. **4**, 583 (1987).
113. S. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta Jr., J. Pweiner, J. Am. Chem. Soc. **106**, 765 (1984).
114. E. Clementi, Ed., *Modern Techniques in Computational Chemistry, MOTECC-89*, ESCOM, Leiden, Chapter 1, 1989.
115. E. Clementi, Ed., *Modern Techniques in Computational Chemistry, MOTECC-91*, ESCOM, Leiden, Chapter 1, 1991.
116. G. Bolis, M. Ragazzi, D. Salvaderi, D.R. Ferro, E. Clementi, Int. J. Quantum Chem. **14**, 815 (1978).
117. *International Symposia on Neural Information Processing, Natural Language Understanding and AI, Advanced Computing for Life-Science, Fuzzy Systems*, July 12-15, 1992, Kyushu Institute of Technology, Fujiki Printing Co., Iizuka, Fukuoka, Japan.
118. E. Clementi, Proc. Nat. Acad. Sci. USA **69**, 2942 (1972).
119. G.C. Lie, E. Clementi, J. Chem. Phys. **69**, 1275 (1974).
120. G.C. Lie, E. Clementi, J. Chem. Phys. **69**, 1278 (1974).
121. E. Clementi, G. Corongiu, D. Bhattacharya, B. Feuston, D. Frye, A. Preiskorn, A. Rizzo, W. Xue, Chem. Rev. **91**, 679 (1991). In particular, see Eqs. 8 to 12. Another equivalent approach is to introduce a Coulomb hole correction in Eq. 9, as proposed by E. Clementi, IBM J. Res. and Dev. **9**, 2 (1965).
122. S. Chakravorty, E. Clementi, Phys. Rev. A **39**, 2290 (1989).

Table I Selected ab initio calculations for the ground state equilibrium energy of H_3 .

Authors	Method	R (b)	E (h)
Salmon, et al.[25] (1973)	CI, 18 SP	1.6500	-1.34335
Mentch, et al.[26] (1981)	Random Walk	1.6500	-1.3439 ± 0.0002
Preiskorn, et al.[27] (1984)	SCC, 24 CGLO	1.6504	-1.343422
Burton, et al.[28] (1985)	CI, 108 PNO	1.6525	-1.34272
Meyer, et al.[29] (1986)	CI, 104 CGTO	1.6504	-1.34340
Anderson[30] (1986)	Random Walk	1.6500	-1.34376 ± 0.00003
Traynor, et al.[31] (1988)	Random Walk	1.6500	-1.34387 ± 0.00005
Urdaneta, et al.[17] (1988)	HCI, 48 CGTO	1.6504	-1.343500
Huang, et al.[32] (1990)	Random Walk	1.6500	-1.3433 ± 0.0005
Alexander, et al.[33] (1990)	RTO, 700 GTG	1.6504	-1.3438220
Frye, et al.[34] (1990)	HCI, 138 GTO	1.6499	-1.3438279

Note: SP = Singer Polynomial, CGLO = Contracted Gaussian Lobe Orbitals, PNO = Pseudo Natural Orbitals, CGTO = Contracted Gaussian Type Orbitals, RTO = Random Temperd Optimization, GTG = Gaussian Type Geminal, GTO = Gaussian Type Orbitals.

Table I Selected ab initio calculations for the ground state equilibrium energy of H_3 .

Authors	Method	R (b)	E (h)
Salmon, et al.[25] (1973)	CI, 18 SP	1.6500	-1.34335
Mentch, et al.[26] (1981)	Random Walk	1.6500	-1.3439 ± 0.0002
Preiskorn, et al.[27] (1984)	SCC, 24 CGLO	1.6504	-1.343422
Burton, et al.[28] (1985)	CI, 108 PNO	1.6525	-1.34272
Meyer, et al.[29] (1986)	CI, 104 CGTO	1.6504	-1.34340
Anderson[30] (1986)	Random Walk	1.6500	-1.34376 ± 0.00003
Traynor, et al.[31] (1988)	Random Walk	1.6500	-1.34387 ± 0.00005
Urdaneta, et al.[17] (1988)	HCI, 48 CGTO	1.6504	-1.343500
Huang, et al.[32] (1990)	Random Walk	1.6500	-1.3433 ± 0.0005
Alexander, et al.[33] (1990)	RTO, 700 GTG	1.6504	-1.3438220
Frye, et al.[34] (1990)	HCI, 138 GTO	1.6499	-1.3438279

Note: SP = Singer Polynomial, CGLO = Contracted Gaussian Lobe Orbitals, PNO = Pseudo Natural Orbitals, CGTO = Contracted Gaussian Type Orbitals, RTO = Random Temperd Optimization, GTG = Gaussian Type Geminal, GTO = Gaussian Type Orbitals.

Table II Electronic Correlation Energies in a.u. for some of the terms of the 2, 3, 4, 10 and 30 electrons isoelectronic series.

Z	2	3	4	10	30
2	-0.04196				
	-0.04204				
3	-0.04339	-0.0450			
	-0.04350	-------			
4	-0.04415	-0.0472	-0.0899		
	-0.04427	-0.0471	-0.0939		
10	-0.04551	-0.0509	-0.1758	-0.3706	
	-0.04569	-0.0507	-0.1791	-0.3915	
18	-0.04594	-0.0519	-0.2740	-0.3925	
	-0.04612	-0.0519	-0.2770	-------	
30	-0.04622*	-0.0523	-0.4168	-0.4046	-1.32
	-0.04599	-0.0524	-0.4197	-0.404	-----

Note: Top values from Refs. 38 or 53 (for Zn); bottom values from Ref. 39
(*) old value: -0.04616

Table III Top: SCF and Total Energies (in a.u.), Cohesive Energies, C.E. (in Kcal/mole), for the C_{60} clusters using the [(9,5,1)/(4,2,1)] basis set. The notation (s), (d) and (t) refers to singlet, doublet and triplet. Bottom: SCF and Total Energies (in a.u.), Complex Binding Energies, ΔE (in Kcal/mole), for the C_{60}–Li, C_{60}–Na and C_{60}–K complexes, using for the lithium atom the [(6,1)/(4,1)] basis set, for the sodium atom the [(14,8,1,)/(7,3,1)] basis set, for the potassium atom the [(14,9,1)/(8,5,1)] basis set, and for the carbon atom the [(9,5,1)/(4,2,1)], basis set.

	E(SCF)	C.E.	E(SCF+CC)	C.E.	E(SCF+B)	C.E.
C_{60} (s)	-2271.7596	111.36	-2284.8453	150.89	-2285.0280	145.76
C_{60}^{2-} (s)	-2271.7442	111.20	-2284.9176	151.65	-2285.0981	146.49
C_{60}^{2-} (t)	-2271.7481	111.24	-2284.8775	151.23	-2285.0995	146.51
C_{60}^{-} (d)	-2271.8093	111.88	-2284.9168	151.64	-2285.1196	146.72
C_{60}^{+} (d)	-2271.4518	108.14	-2284.4715	146.98	-2284.6799	142.12
	E(SCF)	ΔE			E(SCF+B)	ΔE
C_{60}–Li^{+} (s)	-2278.9995	-2.37			-2292.3325	-14.27
C_{60}–Li (d)	-2279.1735	11.43			-2292.5470	-20.00
C_{60}–Na^{+} (s)	-2433.4422	-7.09			-2447.1430	-20.55
C_{60}–Na (d)	-2433.6149	-1.65			-2447.3570	-18.30
C_{60}–K^{+} (s)	-2870.7555	0.22			-2884.8769	-25.18
C_{60}–K (d)	-2870.9295	-16.80			-2885.0919	-62.29

FIGURE CAPTION

Figure 1. Pair Correlation functions. a): $g_{OO}(r,T)$. b): $g_{OH}(r,T)$. Coordination numbers. c): $N_{OO}(r,T)$. d): $N_{OH}(r,T)$.

Figure 2. a) Coordination numbers. $N_{OO}(T)$ and $N_{OH}(T)$ at different temperatures for $R=R_{min}$ of $g_{OO}(r,T)$ and $g_{OH}(r,T)$. b) Histogram of coordination numbers for one water molecule during 12 ps, at T=305 K.

Figure 3. Distribution of water molecules with coordination number from 2 to 7. Top: O-O coordination number within a sphere with R=3.24 Å. Middle: O-O coordination number within a sphere with $R=R_{min}$ of g_{OO}.

Figure 4. Normalized life time decay for solvated complexes with three-, four-, five- and six-coordination.

Figure 5. Distribution of polygons with 3 to 8 water molecules in the redundant description (top) and in the primitive description (bottom).

Figure 6. The global simulation approach.

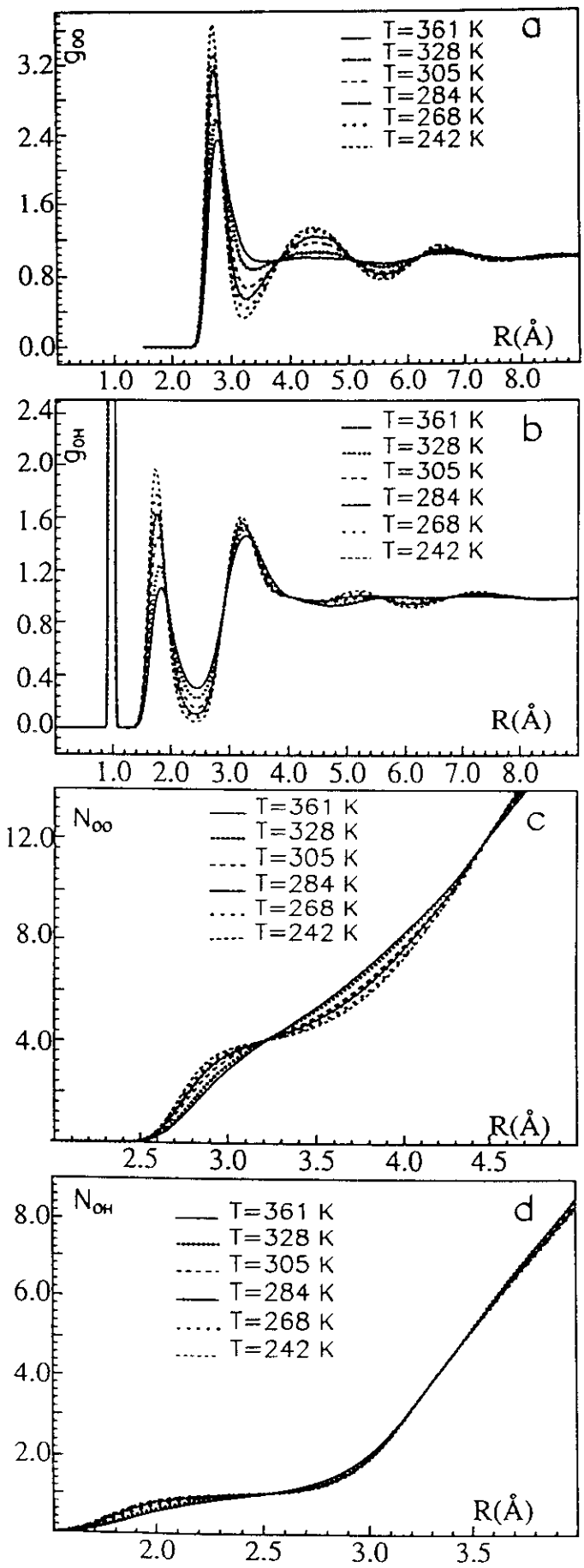


Fig. 1

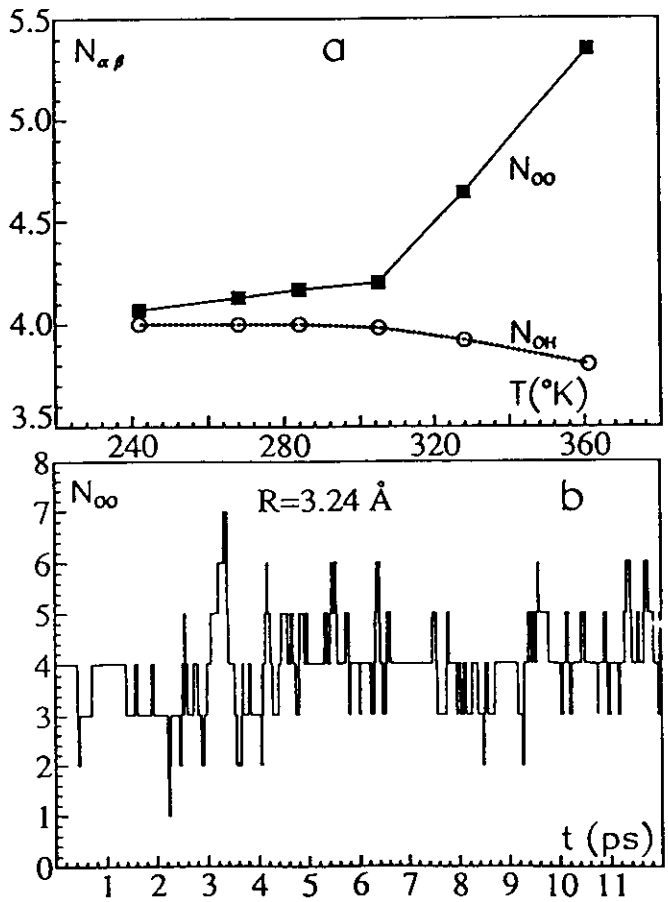

5.5
5.0
4.5
4.0
3.5
$N_{\alpha\beta}$
a
N_{oo}
N_{oH}
T(°K)
240
280
320
360
8
7
6
5
4
3
2
1
0
N_{oo}
R=3.24 Å
b
t (ps)
1 2 3 4 5 6 7 8 9 10 11

Fig 2.

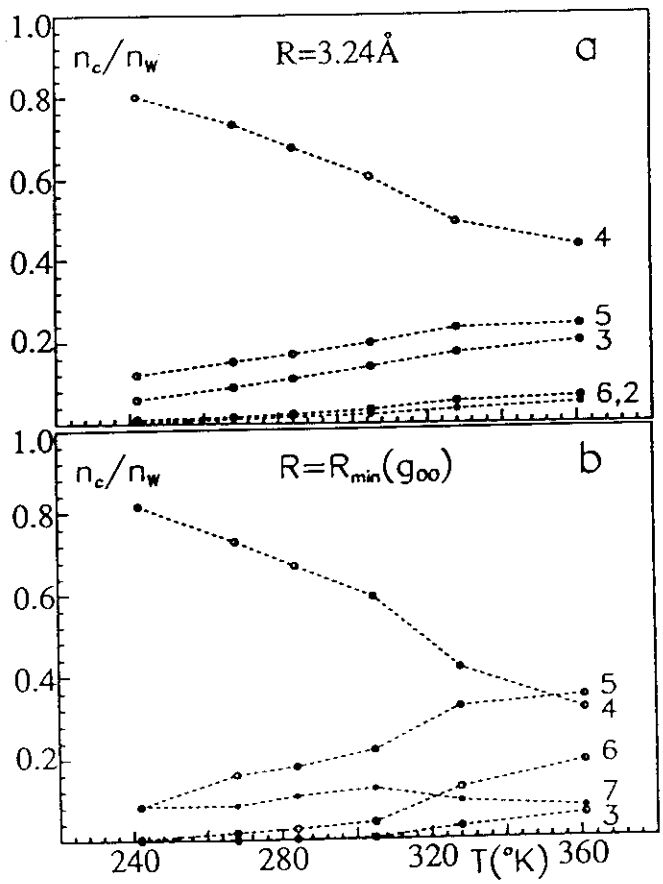

1.0
0.8
0.6
0.4
0.2
n_c/n_w
R=3.24Å
a
4
5
3
6,2
n_c/n_w
$R=R_{min}(g_{oo})$
b
5
4
6
7
3
240
280
320
T(°K)
360

Fig 3

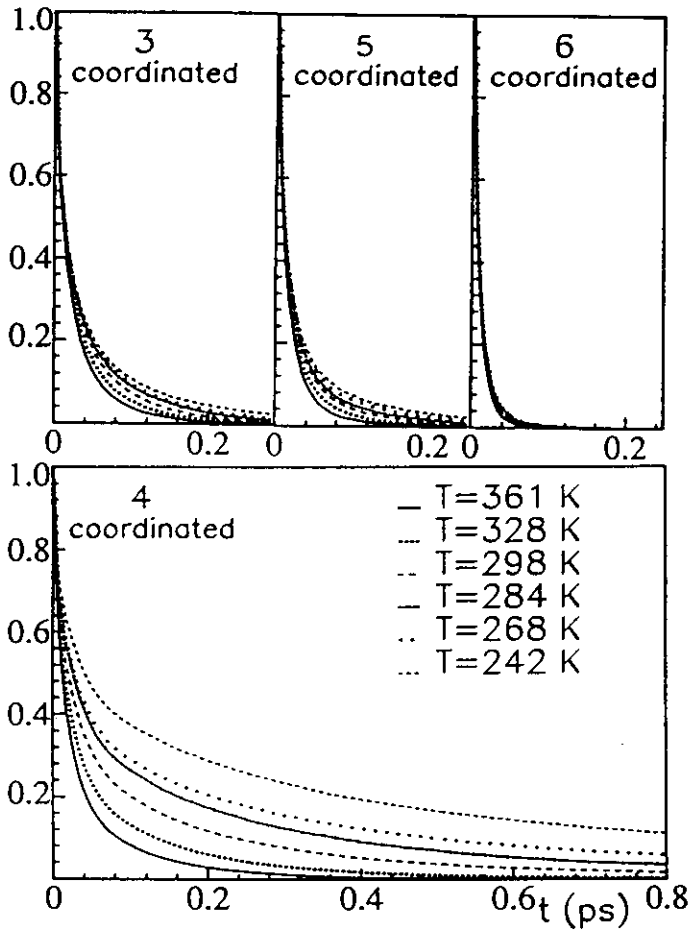

3
coordinated
5
coordinated
6
coordinated
4
coordinated
T=361 K
T=328 K
T=298 K
T=284 K
T=268 K
T=242 K
1.0
0.8
0.6
0.4
0.2
0
0.2
0.4
0.6
0.8
t (ps)

Fig 4

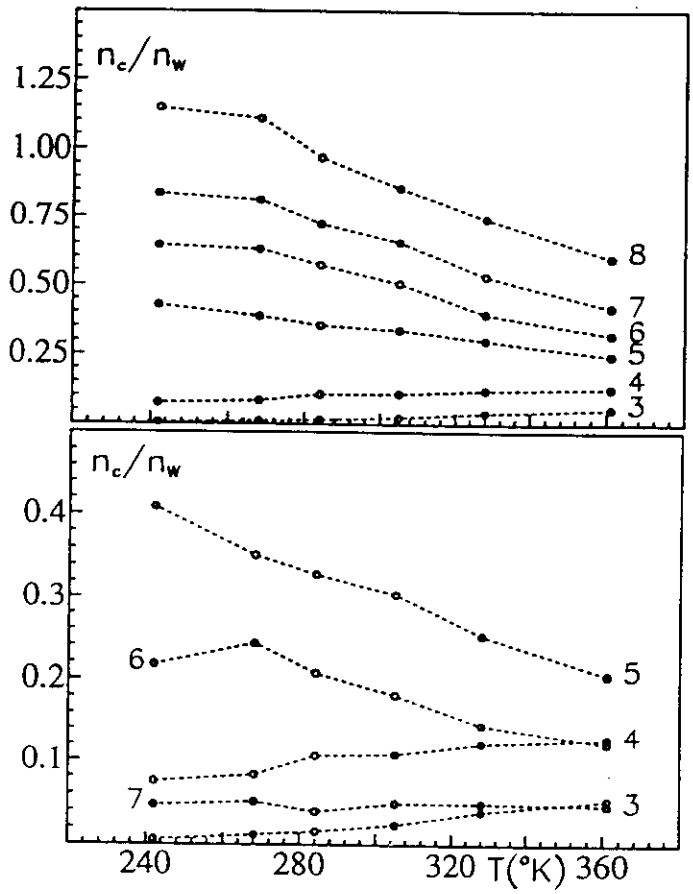
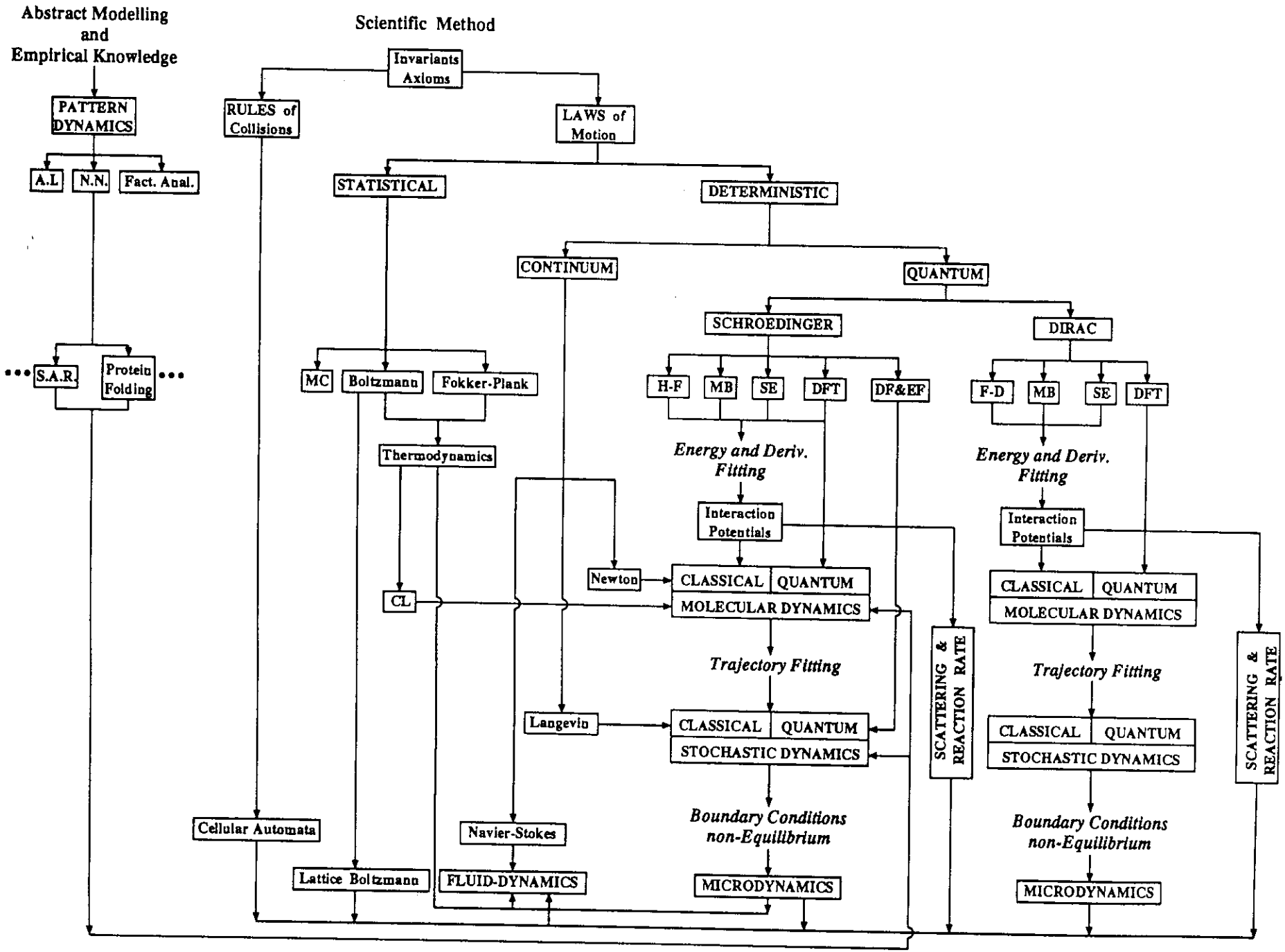

n_c/n_w
1.25
1.00
0.75
0.50
0.25
8
7
6
5
4
3
n_c/n_w
0.4
0.3
0.2
0.1
6
5
4
7
3
240
280
320 T(°K) 360

Fig. 6


Abstract Modelling and Empirical Knowledge
PATTERN DYNAMICS
A.L
N.N.
Fact. Anal.
S.A.R.
Protein Folding
Scientific Method
Invariants Axioms
RULES of Collisions
LAWS of Motion
STATISTICAL
DETERMINISTIC
CONTINUUM
QUANTUM
SCHROEDINGER
DIRAC
MC
Boltzmann
Fokker-Plank
H-F
MB
SE
DFT
DF&EF
F-D
MB
SE
DFT
Thermodynamics
Energy and Deriv. Fitting
Energy and Deriv. Fitting
Interaction Potentials
Interaction Potentials
Newton
CL
CLASSICAL
QUANTUM
MOLECULAR DYNAMICS
CLASSICAL
QUANTUM
MOLECULAR DYNAMICS
Trajectory Fitting
Trajectory Fitting
Langevin
CLASSICAL
QUANTUM
STOCHASTIC DYNAMICS
CLASSICAL
QUANTUM
STOCHASTIC DYNAMICS
SCATTERING & REACTION RATE
SCATTERING & REACTION RATE
Cellular Automata
Navier-Stokes
Boundary Conditions non-Equilibrium
Boundary Conditions non-Equilibrium
Lattice Boltzmann
FLUID-DYNAMICS
MICRODYNAMICS
MICRODYNAMICS