Selected Topics in ab Initio Computational Chemistry in Both Very Small and Very Large Chemical Systems

ENRICO CLEMENTI,* GIORGINA CORONGIU, DELEEP BAHATTACHARYA, BRADLEY FEUSTON, DANIEL FRYE, ALEXANDRA PREISKORN, ANTONIO RIZZO,† and WEIGE XUE

IBM Corporation, Center for Scientific and Engineering Computations, Department 48B/MS 428, Neighborhood Road, Kingston, New York 12401

Received October 10, 1990 (Revised Manuscript Received February 23, 1991)

Contents

The group members are as follows: The bottom row, left to right is Dr. Aleksandra Preiskorn, Dr. Giorgina Corongiu, and Dr. Enrico Clementi. The top row, left to right is Dr. Dileep Bhattacharya, Dr. Bradley Feuston, Dr. Weige Xue, Dr. Daniel Frye, and Dr. Antonio Rizzo.

Dr. Clementi, manager of Scientific Engineering Computations—IBM Data Systems Division in Kingston, after his doctoral degree from the University of Pavia, Italy, was a postdoctor with G. Natta (Milan), M. Kasha (Tallahasee), K. S. Pitzer (Berkeley), and R. S. Mulliken (Chicago). He joined the IBM Research Laboratory in San Jose, CA in 1961 and stayed there until 1966; he was a visiting professor at the University of Chicago in 1966, manager of Large Scale Scientific Computations in San Jose, CA from 1967 to 1974, manager of "Dipartimento Calcolo Chimico", Montedison, Novara, Italy from 1974 to 1979. He has been an IBM Fellow since 1968. He won the position of Professor of Theoretical Chemistry by the Italian Ministry of Education in 1975. He was awarded Teresiana gold medal in 1984, named Fellow of the American Physical Society in 1984, named Distinguished Research Professor at Rensselaer Polytechnic Institute, New York in 1986, and received the DIRAC golden medal by the World Association of Theoretical Organic Chemists in 1987. His work, published in about 400 papers and 4 monographs, deals with theory and computer simulations of atoms, small and large molecules, liquids, solutions, macromolecules, microdynamics and fluid dynamics, and computer architecture, particularly parallel computers.

1. Introduction

Computational chemistry encompasses a growing number of fields and corresponding techniques. The traditional techniques for stationary-state quantum chemistry and equilibrium Monte Carlo[1] and Molecular Dynamics[2] are being expanded with Quantum Monte Carlo,[3] Quantum Molecular Dynamics,[4] Microdynamics[5] but also by research and computer programs on data base, interactive animation, artificial intelligence, and chemical knowledge processing.[6]

This very vigorous expansion should, however, not be taken as indication that the problems characteristic in the traditional areas have been overcome. Indeed, some new avenues have been introduced to complement, if not bypass, some of the previous ones.

Presently, we are at the end of the IVth computer generation,[7-9] where we have witnessed 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.

The next computer generation[7-9] is the VIst and, eventually, we shall make use of the products of the much heralded Vth generation. 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 multiprocessing. We already see workstations with the same speed as the old CRAY XMP (single processor) and the disk storage is being addressed by channels working in parallel, with optical fibers of one or more miles in interconnected distances. Slowly, also the Vth computer generation will appear, since artificial intelligence and expert systems are becoming ubiquitous, especially in complex chemical systems, where correlation rather than understanding is all one can obtain.

In the conclusions of this review paper we shall assume today's computational techniques but with the availability of systems with up to 100–500 GFLOPS and capability of retrieving from disk storage hundreds of gigawords at 100–500 Mbytes/s transmission. These predictions are "reasonable" from a technical viewpoint.

†Permanent address: ICQEM del CNR, Via Risorgimento 35, 56100 Pisa, Italy.

0009-2665/91/0791-0679$09.50/0

Clearly, the average computing availability will trail behind these limits, even if one can be confident that everybody will experience much improvement.

We shall limit this review to work carried out mainly at our laboratory, since in this way we can compare amply different techniques which, however, are executed on the same hardware and system software; thus, because of the above common computational basis, it is feasible to extrapolate future performances on different fields. Indeed, it is well realized that comparative performance analysis on different architectures (scalar, vector, coarse-grain parallelism, massive parallelism) is becoming harder and harder.

We shall start by considering one of the first techniques introduced in quantum mechanics[11] 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, and the question we shall pose is on the feasibility to extend it from 2- to 3- or 4-electron systems.

Next, we shall move away from the goal of 0.1 to 10 cm^{-1} (i.e. 0.0000005 to 0.00005 au) accuracy in absolute energy and move to the millihartree accuracy, again using CI, but with a more standard expansion of determinantal functions. The examples analyzed are the isoelectronic series for He (^1S), Li (^2S), Be (^1S), and Ne (^1S). Here we ask the question on how to improve on the accuracy and how to move to second and third row atoms with the present accuracy.

We shall not consider methods like Möller–Plesset perturbations. There is a tendency in today's literature to overclaim the accuracy one obtains using MPx. Indeed, we note that some authors compare, for example, binding energies in a molecule with atomization products, which are far from the limit of the adopted approximation. For example, it is very common to proceed to post-Hartree–Fock corrections before reaching a near-Hartree–Fock limit value (which is very seldom obtained in today's literature). In addition, basis set superposition (BSS) corrections are often neglected and molecular interaction computations with zero BSS error are essentially unknown. In this way ab initio techniques are used "semiempirically" with calibrated basis sets, which we are becoming accustomed to seeing being 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,[10] but where the same obstacles are often present.

Much of chemistry, though, is interested in such large systems where even a MP2 computation would become computationally much too expensive. This is the area where the appropriate techniques are either the Hartree–Fock approximation with correlation corrections by density functionals, or ab initio molecular dynamics, or quantum molecular dynamics. To maintain our discussion on a very concrete level, we shall consider two systems: the first is a 60-carbon atom system, the C_{60} cluster, and the second is liquid water in a simulation of 1000 molecules. Some of these examples have required very extensive time on today's best facilities. Thus, by reasonable extrapolations from today to about five years from now we will discuss a situation that most chemists will start to experience at the end of the century.

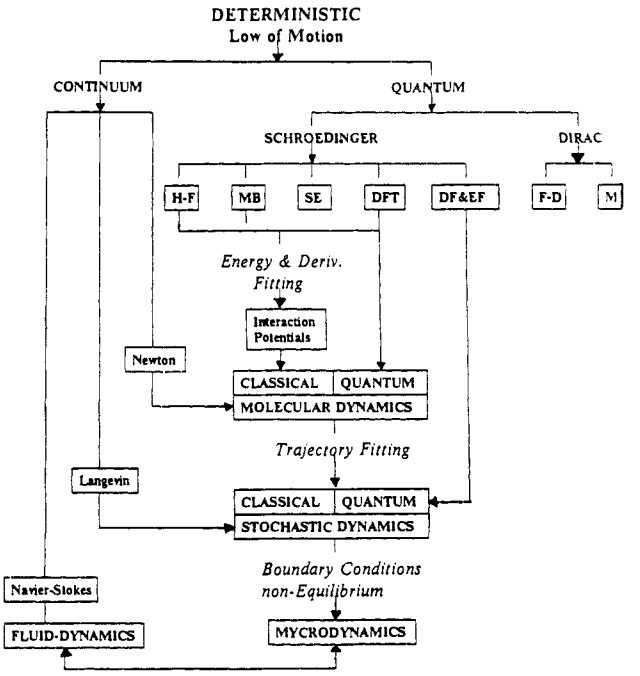

Figure 1. Connectivity between different models in mechanics.

Next, we ask the question on the feasibility of simulations not only for *large systems*, but also for *long time periods*, namely from picoseconds to nanoseconds. A preliminary answer is provided by stochastic dynamics and BPTI in solution is our example.

There are two objectives in this review; the first is to discuss the above models currently used in computational chemistry; the second is to analyze their interdependence. Ideally chemistry needs to have many specialized techniques organized, however, into a global approach, such as to be truly capable of solving the complexity faced daily in a laboratory. In Figure 1 we attempt to convey this idea by connecting a number of models and approximations within an unifying viewpoint. We focus on deterministic mechanics, thus neglecting in this review the statistical approaches like Fokker–Planck and Boltzmann equations, or cellular automata and lattice Boltzmann gas dynamics, all very important to describe fluids and the chemical reaction aspect of fluid dynamics.[7-9] In Figure 1 we stress the evolution of quantum chemistry which deals today mainly with Schrödinger representation; we recall that because of computational difficulties, an equivalent level of development could not be reached in the Dirac representation (either Fock–Dirac or multiconfigurational Fock–Dirac). The different approximations in the figure are indicated as H–F (Hartree–Fock), MB (many body, inclusive of multireference configuration interaction, perturbative approaches, and valence bond), SE (semiempirical, as distinct from ab initio), DF&EF or DFT (density functional theory with and without external fields). The arrows show how models can be linked to models which can deal with more and more complex systems, i.e. they show the flow from small to large, and from zero to short to longer time scales. It is only one part of a larger picture where empirical knowledge, correlations, and artificial intelligence have their role. This broader picture, only partly covered in this review, can be appreciated from the volumes of MOTECC-89[7] and MOTECC-90,[8] and will

be expanded in a forthcoming volume, MOTECC-91.[9]

2. Very Few Electron Systems

The traditional approach to quantum mechanical descriptions of many-electron systems is to optimize linear combinations of one-electron functions. This approach leads then to the description of the correlation in an *implicit* manner, and hence these techniques converge rather slowly, mainly because of the difficulty associated to the representation of the Coulomb hole cusp.

An alternative approach is one where interelectronic coordinates are built *explicitly* into the wave function. Hylleraas did so and calculated the energy and wave function of the helium atom with great success.[11] James and Coolidge[12] extended this method to the hydrogen molecule, and Kolos and Wolniewicz[13] have shown that, to date, it is the most reliable method capable of such accuracy as to even challenge experimental spectroscopic data, "a great triumph of ab initio calculations" to quote R. S. Mulliken.[14] However, due to the numerical complications introduced by the inclusion of interelectronic coordinates in the wave functions, Hylleraas' approach was limited to simple cases such as the helium atom and the hydrogen molecule. We have posed ourselves the question whether with the advent of supercomputers and a more general formulation, the method could be extended from 2- to 3- or 4-electron systems. This more general extension of the Hylleraas method[15-22] consists of multiplying standard many-electron wave functions by powers of the interelectronic 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.

The HCI approach expands the wave function in powers of the interelectronic distances r_{ij} weighted by the original configuration state functions (CSF), Φ_k, i.e.

$$\Psi_{HCI}(1, 2, ..., N) = \sum_{v} \sum_{i<j=1}^{N} \sum_{k=1}^{NCSF} C_{v,k} r_{ij}^{v} \Phi_k(1, 2, ..., N) \quad (1)$$

This expansion in the interelectronic distance is generally assumed to be a power series expansion,[23] and it has been shown by Kutzelnigg[24] that of the possible nonzero v values, $v = 1$ is the most important and our discussion will be limited to the $v = 0$ (normal CI) and $v = 1$ (what we term HCI) terms. Among the different techniques for including the r_{ij} factors leaving the wave function totally antisymmetric, we have chosen the simplest,[25] obtained by restricting in eq 1 the values of v to 0 and 1.

A computer package has been created that performs HCI calculations for many-center, 2- and 3-electron molecular systems[25-27] by using cartesian Gaussian basis sets. Expansion to the most general 4-electron integrals has not yet been done. This package is called HYCOIN (*Hy*lleraas *co*nfiguration *in*teraction)[26] and it has been used successfully to calculate a number of molecular states.[27-33]

The natural system to test our Gaussian code is the ground-state of H_2, two electrons and two protons. Table I summarizes the effect of basis set size for HCI calculations on this state.[28,30,34] The "exact" variational limit for this state was established by Kolos et al.,[35] who used a specialized elliptical coordinate basis set with an explicitly correlated wave function, at −1.174 475 7 hartree. The largest basis set of our calculation, (15s7p2d1f), produces an HCI energy that is less than 0.3 cm^{-1} above this variational limit. If spectroscopic accuracy is defined as having an error of less than ~1 cm^{-1}, Table I indicates that it can be achieved with a basis set of size as small as (13s7p2d).

TABLE I. Effects of the Basis Size in H_2 at an Internuclear Distance of 1.4011 Bohr (Total Energies are Given in Hartrees)

basis set	SCF	conventional CI	hylleraas CI
(13s)	−1.128 532	−1.154 881	−1.168 870
(13s7p)/[13s1p]	−1.133 561	−1.170 495	−1.174 334
(13s7p)/[13s2p]	−1.133 561	−1.171 378	−1.174 380
(13s7p1d)/[13s1p1d]	−1.133 610	−1.171 661	−1.174 399
(13s7p1d)/[13s2p1d]	−1.133 618	−1.172 596	−1.174 456
(13s7p1d)	−1.133 618	−1.173 306	−1.174 467
(13s7p2d)	−1.133 619	−1.173 858	−1.174 473
(14s7p2d1f)	−1.133 622	−1.173 987	−1.174 474
(15s7p2d1f)	−1.133 622	−1.173 987	−1.174 475

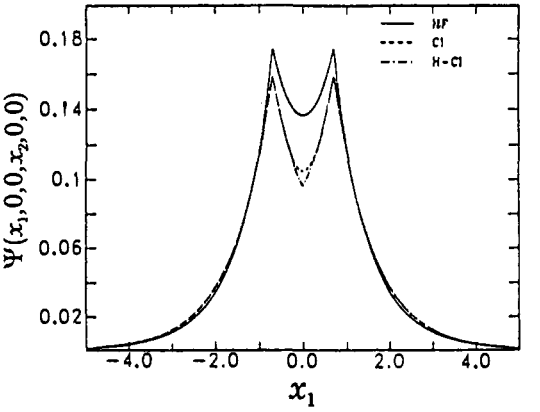

Figure 2. Comparison of the ground-state wave functions, $\Psi(x_1, y_1, z_1, x_2, y_2, z_2)$, of the hydrogen molecule obtained from the Hartree–Fock (HF), the conventional CI, and the HCI calculations. The nuclei are located along the x axis at −0.700 55 and 0.700 55 bohr, whereas one of the two electrons is located at the origin (i.e., x_2, y_2, and $z_2 = 0$).

Compared with conventional CI, the HCI method is much superior in terms of convergence to a given accuracy in energy. For example, addition of the first contracted p Gaussian lowers the conventional CI energy by 0.016 614 hartree, but brings down the HCI energy by only 0.005 464 hartree. Similarly, one d-type function added to the (13s7p)/[13s1p] basis set has an effect of 0.001 16 hartree in conventional CI, compared with 0.000 65 hartree in HCI. All of these seem to be due to the fact that most of the correlation energy has already been picked up in the s-function basis set, 86% with HCI compared to 52% with conventional CI. For example, the HCI energy obtained with (13s7p) contracted to [13s1p], −1.174 334 hartree, is already better than CI employing 5s3p3d2f Slater functions, −1.174 142 hartree.[36] Addition of one d function lowers our calculated energy to −1.174 399 hartree, which is slightly better than a calculation employing uncontracted Gaussians with angular momenta up to g: (12s6p5d4f2g).[37]

One of the most interesting results we have obtained can be seen in the comparison of the HCI wave function with the conventional CI wave function. We see from

TABLE II. Selected ab Initio Calculations for the Ground-State Equilibrium Energy of H_3^+

authors	method[a]	R_{eq}[b]	$E(h)$
Salmon et al.[39] (1973)	CI, 18 SP	1.6500	-1.343 35
Mentch et al.[39] (1981)	random walk	1.6500	-1.3439
Preiskorn et al.[40] (1984)	SCC, 24 CGLO	1.6504	-1.343 422
Burton et al.[41] (1985)	CI, 108 PNO	1.6525	-1.342 72
Meyer et al.[42] (1986)	CI, 104 CGTO	1.6504	-1.343 40
Anderson[43] (1987)	random walk	1.6500	-1.343 76 ± 0.00003
Traynor et al.[44] (1988)	random walk	1.6500	-1.343 87 ± 0.00005
Urdaneta et al.[28] (1988)	HCI, 48 CGTO	1.6504	-1.343 500
Huang et al.[45] (1990)	random walk	1.6500	-1.343 3 ± 0.0005
Alexander et al.[46] (1990)	RTO, 700 GTG	1.6504	-1.343 8220
Frye et al.[32] (1990)	HCI, 138 GTO	1.6499	-1.343 8279

[a] SP = singer polynomial; CGLO = contracted Gaussian lobe orbitals; PNO = pseudonatural orbitals; CGTO = contracted Gaussian-type orbitals; RTO = random-tempered optimization; GTG = Gaussian-type geminal; GTO = Gaussian-type orbitals.

Figure 2 that although the CI wave function agrees almost everywhere with the HCI wave function, it lacks the interelectronic cusp as r_{12} goes to zero.[30] Thus the reason that the CI energy is higher than the HCI energy by $\sim$100 cm^{-1} seems to be this missing of the Coulomb hole cusp. It is difficult to estimate quantitatively the effect of this limitation, but judging from the fact that the (12s6p5d4f2g) calculation just mentioned still leaves more than 16 cm^{-1} unrecovered, one would be surprised if spectroscopic accuracy can be achieved by the conventional CI approach.

The next obvious step toward more complicated molecules is to increase the *number of centers from two to three* while keeping only two electrons, this means the study of the nonlinear H_3^+ molecule, a problem that cannot be tackled by the less general elliptical coordinate methods. We have recently examined both the equilibrium energy and calculated a number of points on the potential energy surfaces of both the ground $^1A'_1$ state and first excited $^3\sum_u^+$ state of H_3^+.[32,33] The lowest energy obtained by us for the equilateral triangle geometry of the ground state was -1.343 8279 hartree at the internuclear distance of 1.6500 bohrs with the use of a 13s5p3d basis set on each site (a total of 138 basis functions). This energy is compared in Table II with other recent ab initio calculations on the equilibrium energy. From Table II we can see that this energy is lower than the previous best published variational calculation of -1.343 500 hartree,[28] and is in excellent agreement with the results of the latest quantum Monte Carlo calculations, -1.343 87 ± 0.000 05 hartree[44] and -1.3433 ± 0.0005 hartree.[45] The energy is also below the soon-to-be published result of Alexander et al.,[46] who used a random-tempered optimization method with Gaussian-type geminals (also a variational calculation with an explicitly correlated wave function), and obtained -1.343 822 0 hartree at an equilibrium separation of 1.6504 bohrs.

From here, we can continue on to 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 juncture, only small basis set calculations have been performed with the HCI method. An example of this is a 3s1p basis set calculation for linear H_3 with internuclear distances of 1.75 b.[27] The SCF energy for this configuration was -1.584 909 3 au, the CI energy was -1.621 700 5 au, and the HCI energy was -1.636 637 9 au. This result should be compared with the result of Liu[36] who obtained an energy of -1.658 743 au at the same saddle point with a CI calculation using a Slater-type function basis set containing orbitals up through 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[47] with a (15s5p2d)/[11s5p2d] Gaussian basis set, we obtained a saddle point energy of -1.658 323 au, which is 0.000 423 au above the best results of Liu. This basis would be a more than sufficient starting point for an HCI calculation on H_3 to ensure a few centimeter^{-1} accuracy in the total energy. Below, we attempt to extrapolate from our CPU timing for this possibility.

It should be noted that the following timings were found with a very recently implemented 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 center takes 359 s on an IBM ES/3090-J/VF to calculate the integrals. A basis set of 1s1p on each center (12 total orbitals) takes 191 h on the same system and a basis set of 3s1p requires approximately 1000 h. Notice that the number of integrals required to be calculated goes 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.[35] If the lessons learned from H_2 are valid, then a full HCI calculation with a basis set of (13s7p)/[13s1p] should provide approximately 10 cm^{-1} accuracy. Scaling from 3s1p to 13s7p by N^6 would call for somewhere around 3500 years on a similar system, 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. This could be reduced further if a more efficient numerical integration is found and if integrals smaller than a given threshold are ignored and/or approximated. Indeed the latter is regularly achieved in SCF computations where savings of a factor of 10^3 are well-known and documented.[7] Most recently a series of papers by Kutzelnigg[48-50] have shown that the Coulomb hole cusp can be "added on" to simpler techniques, like MPx: it is too early to extrapolate from the new result, but the molecular complexity reached in the above recent papers is encouraging, and the above extrapolation clearly points out the need.

3. Few Electron Atoms

Let us now consider a more traditional approach to the correlation energy problem, namely a linear combination of Slater determinantal functions, i.e., the configuration interaction approach, which corresponds to $\upsilon = 0$ in eq 1. We shall start with 2-electron systems, and then expand up to 10 electrons. Thus from one side

TABLE III. Geometrical Basis Sets Employed in the Atomic CI Calculations

system	uncontracted[a]	contracted	α_1[b]	C[b]	j_{min}[b]
He	34,20,16,16	26,20,16,16	0.004 896 1	1.743 580 51	1,1,1,1
Li^{+1}	34,20,16,16	26,20,16,16	0.005 724 8	1.748 162 11	1,3,5,5
Li	27,16,14,14	21,16,14,14	0.001 165 9	2.089 324 30	1,1,1,1
Be^{+2}	34,20,16,16	26,20,16,16	0.006 553 4	1.752 743 70	1,5,7,7
Be^{+1}	27,16,14,14	21,16,14,14	0.006 561 2	2.101 888 21	1,1,1,1
Be	25,16,14,7	21,16,14,7	0.007 434 0	2.097 530 60	1,1,1,1
Ne^{+8}	21,16,9,9	15,10,9,9	0.025 276 3	2.131 648 10	7,7,7,7
Ne^{+7}	27,16,14,14	21,16,14,14	0.017 986 6	2.133 445 96	1,1,1,1
Ne^{+6},Ne	27,22,14,7	21,16,14,7	0.025 276 3[c]	2.131 648 10[c]	1,1,2,1
Ar^{+16}	19,13,10,10	15,11,10,10	0.026 350 4	2.030 666 41	10,10,10,10
Ar^{+15}	27,16,14,14	21,16,14,14	0.035 607 1	2.158 433 47	1,1,1,1
Ar^{+14},Ar^{+8}	22,16,13,13	18,14,13,13	0.026 350 4	2.086 752 00	6,6,6,6
Zn^{+28}	23,17,12,12	16,12,12,12	0.282 646 8	1.917 213 35	8,8,8,8
Zn^{+27}	27,16,14,14	21,16,14,14	0.059 487 4	2.206 749 46	1,1,1,1
Zn^{+26},Zn^{+20}	27,21,14,8	20,16,14,8	0.282 646 8	1.917 213 35	4,4,5,4[d]
Kr^{+34}	23,16,12,12	16,12,12,12	0.219 941 9	1.890 266 87	9,9,9,9
Kr^{+33}	27,16,14,14	21,16,14,14	0.102 722 9	2.238 431 66	1,1,1,1
Kr^{+32},Kr^{+26}	27,20,14,8	20,16,14,8	0.219 941 9	1.890 266 87	5,5,6,5[e]
Xe^{+52}	24,16,12,12	16,12,12,12	0.188 464 4	1.981 823 57	9,9,9,9
Xe^{+51}	27,16,14,14	21,16,14,14	0.194 648 5	2.297 524 88	1,1,1,1
Xe^{+50},Xe^{+44}	28,20,14,8	20,16,14,8	0.188 464 4	1.981 823 57	5,5,6,5[e]

[a] Number of uncontracted and contracted GTO's. k_1, k_2, k_3, k_4 means k_1 functions of symmetry type s, k_2 functions of symmetry type p, k_3 functions in each of the symmetry types d, f, g, h, and k_4 functions of symmetry type i. [b] See text. j_{min} gives the minimum value of the index j. [c] For the i symmetry α_1 = 0.053 880 1 and C = 4.543 923 79. [d] For the i symmetry only even values of j were used. [e] For the i symmetry only odd values of j were used.

TABLE IV. Electronic Correlation Energies and Total Energies in Atomic Units for Some Members of the 2, 3, 4, and 10 Electrons Isoelectronic Series

case	2(^{1}S)	3(^{2}S)	4(^{1}S)	10(^{1}S)
		$E_{tot}-E_{HF}$		
He	−0.041 96			
Li	−0.043 39	−0.045 0		
Be	−0.044 15	−0.047 2	−0.093 9	
Ne	−0.045 51	−0.050 9	−0.179 3	−0.380 0
Ar	−0.045 94	−0.051 9	−0.277 3	−0.394 1
Zn	−0.046 16	−0.052 3	−0.420 0	−0.405 1
Kr	−0.046 22	−0.052 6	−0.490 9	−0.408 1
Xe	−0.046 30	−0.052 8	−0.702 8	−0.412 5
		E_{tot}		
He	−2.903 637 18			
Li	−7.279 807 59	−7.477 772 53		
Be	−13.655 445 22	−14.324 564 89	−14.666 959 80	
Ne	−93.906 617 93	−102.681 953 35	−110.290 278 01	−128.927 112 70
Ar	−312.906 994 89	−346.498 316 00	−377.812 745 63	−506.362 194 29
Zn	−881.407 196 86	−982.223 836 04	−1 079.100 129 78	−1 552.963 473 25
Kr	−1 273.657 240 63	−1 421.586 881 69	−1 564.744 179 18	−2 292.304 761 43
Xe	−2 882.407 308 18	−3 225.674 808 24	−3 561.676 666 35	−5 374.375 579 19

this represents a notable increase in system complexity relative to the HCI approach, but on the other side we should expect some loss in the overall accuracy. The systems considered are atoms and ions, specifically, the neutral ground state of atoms and selected ions of the 2, 3, 4, and 10 electrons isoelectronic series performing single and double CI (SDCI) calculations.[51]

We have used an atomic configuration interaction program ATOMCI[52,53] based upon the powerful techniques developed in the early 1970's by Sasaki,[54] and used by Sasaki and Yoshimine[55,56] to compute correlation energies and electron affinities of the first row atoms.

To decrease the effort in the orbital exponent optimization we have used Gaussian-type orbitals of geometrical basis sets (GBS);[57] we recall that in the geometrical basis set[57] the orbital exponents α_j are related by the equality

$$\alpha_j = \alpha_1 C^{j-1} \qquad (j = 1, ..., n) \qquad (2)$$

and therefore only two parameters need to be optimized for a given set of n elements, i.e. α_1 and C. The same set of exponents were used in orbitals with different angular momenta, and our basis sets included up to i functions.

In Table III we report the final basis sets used to compute the energies of Table IV. In Table III each basis set is characterized by the number of primitive and of contracted functions of s, p, d through h and i type (in all the sets we have used the same number of d, f, g, and h functions). The values of α_1 and C are also given, together with j_{min}, the minimum value for the index j, again for the s, p, d through h and i symmetry types.

In Table IV the correlation energies are reported, together with the total SDCI energies, for selected members of the He ($1s^2$,^{1}S), Li ($1s^22s^1$, ^{2}S), Be ($1s^22s^2$, S) and Ne ($1s^22s^22p^6$, ^{1}S) series, up to the Xe ions. We give one more significant digit than physically meaningful (to allow for roundoff) for the correlation energies, but maintain eight decimal figures, namely the numerical accuracy corresponding to the basis sets of Table III, for the total energies.

TABLE V. Li(^{2}S) Ground State (Contributions from Each Symmetry to the Total and Correlation Energies and Comparison with the Results of Ref 64)

basis	total energy[a]	total energy[b]	correlation energy[b]
s	−7.448 667	−7.448 665 8	−0.015 939 0
sp	−7.473 834	−7.473 800 4	−0.041 073 6
spd	−7.476 760	−7.476 675 6	−0.043 948 8
spdf	−7.477 399	−7.477 381 7	−0.044 655 0
spdfg	−7.477 485	−7.477 624 4	−0.044 897 6
spdfgh		−7.477 725 2	−0.044 998 4
spdfghi		−7.477 772 5	−0.045 045 8

[a] Limits reported for the MCHF calculation of ref 64. [b] This work. See Table III for details on the basis set.

For the 2-electron series, where we perform full CI (FCI) calculations, the correlation energy is accurate to about the fifth decimal figure, but the accuracy of the nonrelativistic total energy is 1 part per 10^6 au for He and goes up to 1 part per 10^9 for Xe^+.[52] As is well known, the most accurate computations on the two electron series are those by Pekeris,[58,59] which have been often reproduced, because He is a test case for many models.[60]

For the 3-electron series the accuracy remains nearly the same as for the 2-electron series, because, as realized long ago, the 1s–2s interpair correlation energy is small[61] and thus SR-SDCI yields energies not much different from FCI. As far as we know, the best results for the Li ground state energy are those by Larsson[62] (−7.478 025 au) and most recently those by Bunge[63] (−7.478 062 4 ± 0.000 007 au). Notice that the total energy for the top member of the series, Li, compares very well to the very recent result of a large scale MCHF calculation by Sundholm and Olsen[64] involving full scale CI within up to 85 orbitals of symmetries ranging from s to g, for a total of 11 514 configuration state functions in $D_{\infty h}$ symmetry (−7.477 773 vs −7.477 485 au). In Table V the contribution to the total energy and to the correlation energy from each symmetry block are reported and compared with the analogous limits for the calculation of ref 64. Our values appear to be very close to Sundholm and Olsen's limits, and the inclusions of several g, h, and i functions seems to be effective in further improving the final result.

For four electrons, the limitations of SR-SDCI being known, MR-SDCI calculations with two-configuration ($(1s)^2(2s)^2$ and $(1s)^2(2p)^2$) MCSCF reference states were performed. For Be our MR-SDCI result for the total nonrelativistic energy (−14.666 960 au) compares well to Bunge's[65] (−14.667 328 au including STO basis set and FCI truncation errors corrections).

The computed SR-SDCI correlation energy for the 10-electron isoelectronic series is about 95% of the estimated total.[61] In an attempt to go beyond SDCI, an approximation which is not sufficiently adequate for the 10-electron isoelectronic series, and thus to include at least partially the contribution of multiple excitations, all triple and quadruple excitations from the HF reference state orbitals to the lowest (i.e. with larger occupation numbers) two atomic natural orbitals of the s, p, and d symmetries obtained in the SR-SDCI calculations were added to the set of singly and doubly excited configurations. These are what we label SR-(SD + tq)CI calculations. The SR-(SD + tq)CI correlation energy for Be (−0.3800 au) is almost 98% of the Veillard and Clementi estimate[61] and is 2.5% better than the SR-SDCI result. On the other end, the SR-SDCI correlation energies were improved by only 0.4% in Ar^+,[8] 0.1% in Zn^+,[20] 0.07% in Kr^+,[26] and 0.03% in Xe^+,[44] respectively. This is a clear indication that for a given electronic configuration, the stronger the central field, i.e. the higher the nuclear charge, the less correlated are the electronic motions.

Let us now comment on the computation time, fast memory, and auxiliary storage needed for these computations in order to extrapolate, for example, to 18- and 36-electron systems (Ar and Kr, respectively). We shall consider S states or even closed shells to make the estimate somewhat simpler. In Table VI we give the pertinent information on the CPU time, the total elapsed time, and the storage requirements for the atomic SR-SDCI calculations on the neutral members of the isoelectronic series taken into account here. Both the amount of memory required and the time depend strongly upon the number of the 2-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. In Table VI we report also a rough estimate of the time and storage requirements which can be extrapolated for some larger systems. The SR-SDCI calculations with an equivalent basis set for argon and krypton would take approximately 25 days and 2.5 years, re-

TABLE VI. CPU Time (on an IBM 3090 V400) and Storage (Mbytes) Requirements for the Atomic SR-SDCI Calculations on the Neutral Members of the Isoelectronic Series Considered Here and Estimates for Some Larger Systems

system	GTO's[a]	CSF's[b]	CPU time	elaps time[c]	2e int[d]	virt mem	disk mem
He	126	1 241	1 300 s	5 760 s	21.0	200	370
Li	107	2 526	855 s	4 320 s	11.5	125	230
Be	100	3 037	1 050 s	7 800 s	8.3	100	160
Ne	100	10 573	4 350 s	19.6 h	8.3	220	410
Ar[e]	124	47 000	21 h[f]	590 h[g]	19.0	700[i]	1 200[i]
Kr[j]	147	215 000	420 h[f]	23 000 h[g]	39.0	3 700[i]	7 000[i]

[a] Number of basis functions. See Table III. [b] Number of configuration state functions. [c] Total elapsed time in seconds (s) or hours (h). [d] Number of computed and stored 2-electron integrals (millions). [e] Rough estimates for an hypothetical SDCI calculation involving 26s, 20p, 17d,f,g,h, and 10i GTO's. [f] Estimated by quadratic extrapolation of the CPU data. 99% and 99.9% of the time would be taken by the generation of the Hamiltonian matrix elements and by its diagonalization for Ar and Kr, respectively. [g] Estimated by assuming a linear relationship between the ratio elapsed time vs CPU time and the number of electrons in the system. [i] Estimated from the number of 2-electron integrals and by a linear fitting of the data for the storage needed to compute and save on disk the energy matrix. [j] Rough estimates for an hypothetical SDCI calculation involving 30s, 24p, 20d,f,g,h, and 13i GTO's.

spectively, and noticeable amounts of fast memory and disc space.

4. Quantum Molecular Dynamics

We will not report on small molecules since the corresponding techniques are not much different from those presented above for the neon isoelectronic series, with, however, an increasing emphasis on the need for multiconfiguration. Nor we shall comment on valence bond computations which are reappearing since computers are finally sufficiently fast (as long one does not seek high accuracy). We would like to state that "small molecules" are the objects for *molecular physics* whereas large molecules are those for *quantum chemistry* in particular and *computational chemistry* in general. As done in the previous section, we discuss in some details a practical application, rather than using the more standard approach, namely to provide first a theory and then search for an application.

Below we consider the example of a relatively large molecular system, specifically a carbon cluster, C_{60}, for which we wish to determine the equilibrium geometry and dynamical properties at room temperature. In addition we shall assume that the molecule has no symmetry, since the geometry determination is one of our goals.

As it is known, both experimental and theoretical methods have been employed to explain the predominance of the 60-atom cluster in the mass spectra obtained with the laser vaporization cluster beam technique.[67-76] Experimental evidence has shown that (i) the dominance of C_{60} in the mass spectra increases with longer clustering times, (ii) both C_{60}^+ and C_{60}^- are dominant clusters, (iii) there exist 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.[67] These results have led researchers to believe that a single structure is responsible for the experimental observations with the most probable candidate that of the truncated icosahedron. More recently laser ablation of graphite in an inert gas atmosphere has enabled researchers to accumulate 99% pure C_{60} in thin films for analysis.[76] Subsequently, results of Raman, IR, STM, and NMR experiments have been obtained.[68-70] The existence of a single peak in the NMR data indicates all the carbon atoms are in identical chemical environments.[70] This leaves little doubt that the structure of the most stable isomer of C_{60} is indeed that of the truncated icosahedron, the so-called Buckminsterfullerene (Bf). Having the structure of a soccerball, this C_{60} configuration has 12 pentagonal and 20 hexagonal faces with each atom identically bonded to 3 atoms.

Theoretical investigations of the C_{60} fullerene have included Hückel molecular orbital theory, CNDO, MNDO, density functional theory, and Hartree–Fock calculations as well as molecular dynamics with empirical interactions potentials.[71-78] In this section we shall stress molecular dynamics techniques used to solve problems generally approached with quantum chemical methods. In the next section we will return on quantum chemical techniques. Classical molecular dynamics (MD) and Monte Carlo (MC) simulations with empirical potentials were reported by Ballone and Miliani,[79] but since the parameters for Tersoff's potential[80] are fitted to bulk properties its ability to describe carbon clusters has been questioned.[80-83] On the contrary, a first principles approach by Car and Parrinello (CP), allows complete degrees of freedom while deriving the interaction potential directly from the electronic ground state.[84] As it is known, the method combines density functional theory and classical molecular dynamics simulation, and below we apply it to a study of the structure and dynamics of the C_{60} fullerene.

In principle, density functional theory (DFT) provides a method of investigating ground-state properties from first principle considerations alone, i.e. with no adjustable parameters. Hohenberg and Kohn have shown that the ground-state energy of a many-electron system is uniquely determined by the ground-state density.[85] Practical applications of DFT became feasible with the work of Kohn and Sham (KS).[86] The assumption that the ground-state density for the interacting electron system is the same for a noninteracting system in the presence of some effective potential, leads to a description of the interacting electron system by a set of N 1-electron (KS) equations:

$$(-\tfrac{1}{2}\nabla^2 + V_{\text{eff}}^{\text{KS}})\Psi_i^{\text{KS}} = \epsilon_i^{\text{KS}}\Psi_i^{\text{KS}} \tag{3}$$

and the corresponding energy is

$$E = -\tfrac{1}{2}\sum_\lambda n_\lambda \int d\mathbf{r}\Psi_i + \tfrac{1}{2}\int d\mathbf{r}\int d\mathbf{r}'\frac{\rho(r)\rho(r')}{|r-r'|} + E_{\text{xc}}[\rho] + E_{\text{NN}} \tag{4}$$

where n_λ is the occupation number of the KS state Ψ_λ, H_0 is the bare nuclei Hamiltonian, and the remaining terms are the Coulomb energy, the exchange correlation energy, and the nuclear repulsion energy. The exchange correlation energy can be computed in several ways and in the local density approximation it is fitted to the exchange correlation energy, $\epsilon_{\text{xc}}(\rho)$, for a homogeneous electron gas of density ρ:

$$E_{\text{xc}}[\rho] = \int d\mathbf{r}\rho(r)\epsilon_{\text{xc}}(\rho) \tag{5}$$

The last approximation—in principle not essential to the DFT—makes the DTF an intermediate approach between ab initio and semiempirical.

The density functional approach within the LDA has provided accurate results for crystals, surfaces, and even clusters and molecules. The interested reader should consult the detailed book by Kryachko and Ludeña[87] on this subject. The above formulation has been casted in a form most convenient to compare it with the approach, discussed in the next section, where the Hartree–Fock is complemented by Wigner's correlation corrections.

The CP approach combines density functional theory with the molecular dynamics simulation technique,[84] making possible the determination of the atomic configuration, dynamical properties, and the electronic structure with a single ab initio calculation. In the CP method, a Lagrangian of the form

$$L(\{\dot\Psi_i\},\{\Psi_i\},\{\dot R_I\},\{R_I\}) = (\mu/2)\sum_{i=1}^{N}\int d\mathbf{r}|\dot\Psi_i(\mathbf{r})|^2 + \tfrac{1}{2}\sum_{I=1}^{N_A} M_I\dot R_I^2 - E[\{\Psi_i\},\{R_I\}] - \tfrac{1}{2}\sum_{I=1J=1_{I\neq J}}^{N_A}\sum^{N_A}\frac{Z_I Z_J}{R_{IJ}} + \sum_{i=1j=1}^{N}\sum^{N}\Lambda_{ij}\Big[\int d\mathbf{r}\Psi_i^*(\mathbf{r})\Psi_j(\mathbf{r}) - \delta_{ij}\Big] \tag{6}$$

is constructed for generating equations of motion for the atomic nuclei, $\{R_I\}$, and *for the parameters in the wave function expansion.* The first term in eq 6 is a kinetic energy associated with the electronic wave function with a fictitious mass μ, which can also be considered as a variational scaling factor. The "mass" parameter μ is carefully chosen to ensure that the dynamics of the wave function is decoupled from the real dynamics of the atomic nuclei, $\mu \ll M_I$. In this way a wave function starting at the minimum will remain close to the Born–Oppenheimer surface. The last term in eq 6 requires the wave function to be orthonormalized at each step. The equations of motion

$$\mu\ddot{\Psi}_i(\mathbf{r}) = -\frac{\partial E}{\partial \Psi_i^*} + \sum_{j=1}^{N}\Lambda_{ij}\Psi_j(\mathbf{r}) \tag{7a}$$

$$M_I\ddot{R}_I^q = -\frac{\partial E}{\partial R_I^q} + \sum_J \frac{Z_I Z_J R_{IJ}^q}{R_{IJ}^2 R_{IJ}} \tag{7b}$$

are numerically integrated, generating realistic trajectories for the nuclei where the forces are derived explicitly from the electronic ground state.

Several groups have successfully applied this technique to the study of small metal and semiconducting clusters.[81-86] In particular, Andreoni and Scharf implemented this technique in the study of small carbon clusters, C_2, C_4, and C_{10}.[86] Their calculation utilized nonlocal, angular-momentum-dependent pseudopotentials with a planewave expansion for the Kohn–Sham orbitals.[86,89] The Perdew–Zunger fit to the Ceperley–Alder exchange–correlation energy was used in the local density approximation (LDA).[90,91] The results of their calculation showed that the LDA of density functional theory predicts the same structures as those obtained by an ab initio CI calculation.[92] In addition, results of a MD simulation at T = 200–300 K found that C_4 can have either a ring or linear structure with the corresponding local minima separated by a large barrier (10^4 K). The ground-state structure of C_{10} was found to have two degenerate enantimorphs of D_{5h} symmetry. At low temperatures the dynamical path separating these two minima was investigated. Both C_4 and C_{10} were predicted to have linear structures at high temperatures, as predicted by Pitzer and Clementi long ago.[93]

Let us now discuss in detail the calculation of C_{60}, which utilizes the same nonlocal pseudopotential and exchange–correlation energy mentioned above. A planewave basis set was also used for the wave function expansion with a cutoff of 35 Ry. The simulation employed a supercell with FCC periodic boundary conditions and a lattice constant of 17.5 Å. Approximately 32 000 planewaves were required for each of the 120 electronic states. The "mass", μ, was fixed at 500 au and the time step for integrating the equations of motion was 3 au. The simulation required 256 Mbytes of memory and 90 s per iteration on 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 lowest energy structure, to achieve equilibration, and to study the dynamics about 700 h of a dedicated IBM-3090/600J were required.

The initial C_{60} configuration had the structure similar to that of a soccerball with bond lengths close to that of graphite and bond angles within several degrees of the correct angles of 108° and 120°. Once a self-consist solution to the KS equations had been achieved for the initial configuration, geometry optimization was performed with the conjugate gradient method. The lowest energy Bf structure so obtained has two different bond lengths, a short bond, 1.389 Å, on the edge between two neighboring hexagons and a longer bond, 1.445 Å, on the edge between an adjacent pentagon and hexagon. An earlier quantum chemical calculation found similar results with short and long bonds of 1.369 and 1.453 Å.[74] Bond angles in the pentagonal and hexagonal rings are 108° and 120°, respectively. The radius of the ball from the present calculation, 3.538 Å, also agrees with the experimental estimate of 3.5 Å.

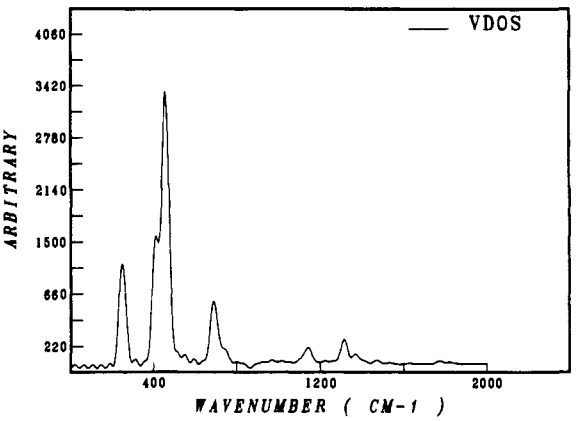

Figure 3. Vibrational density of states (VDOS) of Buckminsterfullerene.

The dynamics of the C_{60} fullerene were also investigated. The nuclei were first heated by scaling their velocities to achieve the target temperature. Since heating the atoms this way caused the wave function to drift from the Born–Oppenheimer surface, periodic minimizations of the wave function parameters were required to regain the self-consistent solution to the Kohn–Sham equations. Once the velocity scaling was completed and a final minimization for the wave function performed, the atoms were allowed to evolve in time, while conserving the total energy. The CP equations ensured that the electron density minimized the quantum mechanical energy at each time step. After a short simulation for equilibration the atomic trajectories were analyzed and the vibrational density of states (VDOS) determined. A total of 1.4 ps (20 000 MD steps) were simulated. The average temperature for the Bf molecule, determined from the kinetic energy, was around 450 K, although the large amplitude oscillation of the symmetric breathing mode caused the temperature to fluctuate between 200 and 800 K.

The VDOS is shown in Figure 3. The two symmetric breathing modes (A_g), corresponding to the expansion and contraction of the 5- and 6-membered rings in and out of phase with each other, were identified in the recent Raman spectra of the C_{60} film at 496 and 1470 cm^{-1}, respectively.[68] A simple analysis of the average radius of Bf molecule as a function of time yields the lower A_g mode at 454 cm^{-1}, corresponding to the largest peak in Figure 3. A similar analysis of the lengths of the two different bond types, allows us to assign the small peak at 1364 cm^{-1} as the higher A_g mode. The normal modes of the C_{60} fullerene have been the object of several theoretical investigations.[94-97] In Table VII,

TABLE VII. Symmetric A_g Breathing Modes (cm^{-1}) for the Buckminsterfullerene

	ν_1	ν_2
present	1364	454
Brendsdal et al.[95]	1409	388
Wu et al.[96]	1627	548
Stanton–Newton[97]	1667	610
expt[157]	1470	496

a comparison of the earlier attempts to determine the symmetric breathing modes as well as the present simulation are made with the experimental results. Overall, the present results compare quite well against the previous theoretical calculations. The lowest mode in the VDOS found in the present simulation is at 246 cm^{-1} which is to be compared to 273 cm^{-1} in the Raman spectrum. In the Raman spectrum, this lowest mode corresponds to the H_g squashing mode.[94] Results of first principles MD simulation of Bf indicate stability up to at least 800 K, the highest temperatures simulated. The calculated VDOS gives frequencies in fair agreement with experiment though consistently yielding lower values than the experimentally observed spectra. One area of difficulty in the present calculation is the attempt to ascertain the VDOS from such a short simulation time (1.4 ps). In addition, subsequent analysis of the dynamical trajectories indicates that not all of the vibrational modes were thermally excited during the equilibriation phase of the simulation. Longer simulation times are required for both the equilibriation and data sampling phases, but would probably yield few new results to justify the additional expense. The present results, reported elsewhere in more detail,[98] indicate that Bf is a very stable isomer of the C_{60} clusters. The close correspondence between the observed and calculated vibrational spectra contributes to the growing body of evidence that the experimentally observed C_{60} cluster is indeed the Buckminsterfullerene molecule.

This concludes our example, which has been given to illustrate that the CP method can, indeed, be used to solve structural and dynamical problems in medium-large molecules. In the Conclusions section we shall discuss the computational cost of the CP method, some of its limitations and give some suggestions for improvements.

5. Large Molecules from Quantum Mechanics

In this section we return on Schrödinger quantum mechanics (see Figure 1), and we consider a relatively large molecule. To facilitate comparison with the previous section, we continue with the carbon cluster C_{60}. Let us recall that the existence of an extended family of "carbon molecules, C_n" was proposed long ago by Pitzer and Clementi[93] using semiempirical quantum chemistry methods. As above noted, after about 20 years, in the last decade "carbon clusters" have become a rather hot subject and the carbon cluster's literature on laboratory data[99-106] and computational experiments[74,78,98,107-124] is becoming larger and larger.

Below we use Hartree–Fock functions; we recall that in the Hartree–Fock model one describes the N interacting electrons by a set of N 1-electron (HF) equations:

$$(-\tfrac{1}{2}\nabla^2 + H_{\text{eff}}^{\text{HF}})\Psi_i^{\text{HF}} = \epsilon_i^{\text{HF}}\Psi_i^{\text{HF}} \qquad (8)$$

and the corresponding energy is

$$E = 2\sum_k H_k + \sum_{kl}(2J_{kl} - K_{kl}) + f[2\sum_m H_m + f\sum_{mn}(2aJ_{mn} - bK_{mn}) + 2\sum_{km}(2J_{km} - K_{km})] + E_{\text{NN}} \qquad (9)$$

where the indices k and l refer to orbitals of closed shells and m and n are open shell orbitals; the constants a and b are the vector coupling coefficients, f is the fractional occupation of the shell, and J and K are the Coulomb and the exchange energies. By comparing eqs 4 and 9, it is evident why often today's DFT is restricted to closed shell systems. By defining the correlation energy of the N electron and M nuclei system as the difference between the exact non relativistic energy and the Hartree–Fock energy of eq 9, we can write

$$E(\text{exact}) = E(\text{Hartree–Fock}) + E(\text{corr}) \qquad (10)$$

and, as known since the early 1930's,[125,126] the latter can be approximated as

$$E(\text{corr}) = \int d\tau \rho \epsilon_c(\rho) \qquad (11)$$

where $\epsilon_c(\rho)$ is the functional of the Hartree–Fock density, assumed to be proportional to the correlation energy of the homogeneous electron gas of density ρ. By including E(corr) in eq 9, we have the Coulomb–Hartree–Fock (C–HF) approximation proposed by Clementi,[127] and eq 8 is now

$$(-\tfrac{1}{2}\nabla^2 + H_{\text{eff}}^{\text{C-HF}})\Psi_i^{\text{C-HF}} = \epsilon_i^{\text{C-HF}}\Psi_i^{\text{C-HF}} \qquad (12)$$

A computer program to obtain C–HF functions for closed and open shell atoms has been available since long.[128] It is not difficult to derive rules for the correlation energy in terms of interacting electron pairs i and j, triplets i, j, and k, etc., and this leads for example to the Clementi–Chakravorty (Cl–Ch) electron pair approximation[129] based on the expansion

$$E(\text{corr}) = \sum \eta_{ij} + \sum \eta_{ijk} + \ldots$$

Let us now return to the carbon cluster. Following the QMD simulation[98] reported in the previous section, we did consider the electronic structure of the C_{60} neutral cluster. Very preliminary computations were performed with a (9,5) Gaussian basis set of geometrical-type[57] contracted to single-ζ [(9,5)/(2,1)]. In subsequent computations, however, we have used the (9,5) basis set of van Duijneveldt[130] contracted to a double-ζ [(9,5)/(4,2)]. The SCF-MO closed shell energies resulting from these computations are −37.767 and −37.845 au per carbon atom, respectively, which can be compared with the value of −37.819 au from Lüthi and Almlöf[74] (as expected, better than our single-ζ, but inferior to our double-ζ). Since the correlation energy corrections are very important, but because accurate many-body computations are very expensive we have used the Clementi–Chakravorty electron pair approximation,[126,129] known to yield reliable estimates of the correlation energy for many molecules at minimal computational cost. From the (Cl–Ch) model the computed total correlation energy using the [(9,5)/(2,1)] basis set in C_{60} is −13.109 au; for the [(9,5)/(4,2)] basis set we compute the value of −13.090 au, indicating that the correlation energy is indeed important in the C_{60} cluster. This way the cohesive energy per carbon atom in C_{60} is 139.74 kcal/mol and 124.31 kcal/mol for the [(9,5)/(4,2)] and the [(9,5)/(2,1)] basis sets, respectively. These preliminary computations yield only estimates of the binding

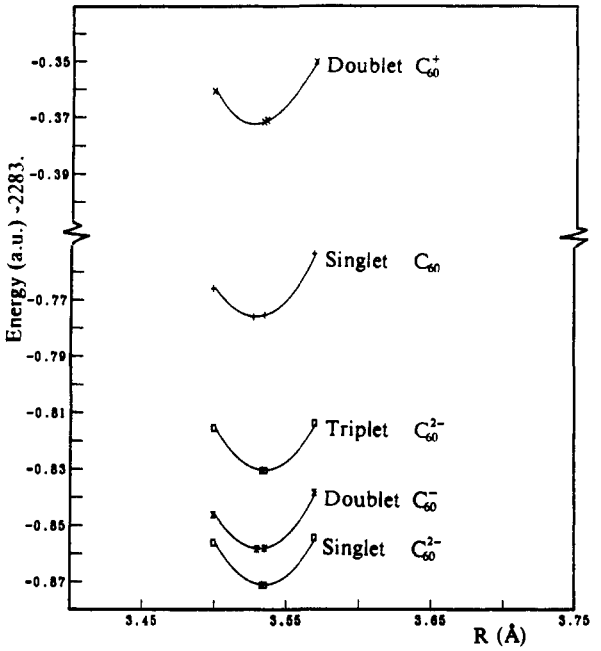

Figure 4. Energy surfaces for three different geometries for double-ζ basis set in C_{60}.

energy since the basis set is too small. Let us now expand the scope of our investigation and deal with two topics: first an analysis of the relative stability of C_{60} and its ions, both positive and negative, then the inclusion of atoms, like Li and Ca, into the C_{60} cavity. In the following, for C_{60} we will use a geometry were the icosahedral symmetry is exact, rather than the "near" symmetry obtained from the molecular dynamics simulation.[98]

Since in the following we will discuss C_{60} spectroscopic quantities, like ionization potentials, electron affinities, and other excitation energies we must, first of all, test the sensitivity of these quantities to geometry variations. In order to reduce the computational cost, we have assumed that the double/single bond length ratio (given by the quantum molecular dynamics simulation) remains constant, while for the value of R, the sphere radius enclosing C_{60}, we have considered R = 3.50 Å, R = 3.536 Å, and R = 3.57 Å. From Figure 4 it appears that for the different states of the C_{60} clusters the minima are not exactly the same, but nearly so. Our conclusion, obtained by performing 15 double-ζ basis set computations (three for each of the five states considered), should retain its over-all validity, even by expanding the basis set. From Figure 2 we predict for the three different geometries, an electron affinity of 2.46–2.75 eV, to be compared with an experimental value of 2.8 eV.[131] The computed values of the ionization potential remains, however, far from the experimental value: it is either ~ 11.0 or 9.2 eV, by including or not the correlation correction. The average elapsed time for one of the above computation was about 15 h on an IBM-3090/600J used in parallel.

As it is known, a double-ζ basis set, even of good quality, but without polarization is rather unreliable for energy determination aiming at about 5–10 kcal/mol accuracy; thus we have added a 3d polarization function to the [(9,5)/(4,2)] basis set for each carbon atom. Being now the computations are much more expensive, we limited our work to one geometry (R = 3.57 Å); this

TABLE VIII. Comparison of SCF and Total Energies for the C_{60} Clusters Using the [(9,5,1)/(4,2,1)] Basis Set and for the C_{60}–Li Complexes

cluster[a]	energies, au		cohesive energies, kcal/mol
C_{60} (s)	E(SCF)	-2271.7596	111.36
	E(SCF + CC)	-2284.8453	150.89
	E(SCF + B)	-2285.0280	145.76
C_{60}^{2-} (s)	E(SCF)	-2271.7442	111.20
	E(SCF + CC)	-2284.9176	151.65
	E(SCF + B)	-2285.0981	146.49
C_{60}^{2-} (t)	E(SCF)	-2271.7481	111.24
	E(SCF + CC)	-2284.8775	151.23
	E(SCF + B)	-2285.0995	146.51
C_{60}^{-} (d)	E(SCF)	-2271.8093	111.88
	E(SCF + CC)	-2284.9168	151.64
	E(SCF + B)	-2285.1196	146.72
C_{60}^{+} (d)	E(SCF)	-2271.4518	108.14
	E(SCF + CC)	-2284.4715	146.98
	E(SCF + B)	-2284.6799	142.12
complex[a]	energies, au		complex binding energies: ΔE, kcal/mol
C_{60}–Li^{+} (s)	E(SCF)	-2278.9995	-2.37
	E(SCF + B)	-2292.3325	-14.27
C_{60}–Li (d)	E(SCF)	-2279.1735	11.43
	E(SCF + B)	-2292.5470	-20.00

[a] The notation s, d, and t refers to singlet, doublet, and triplet.

should be acceptable *for preliminary results* since the double-ζ computations show nearly parallel potential energy displacements for the different electronic states and configurations (see Figure 4).

We have considered C_{60} (singlet), C_{60}^{2-} (singlet), C_{60}^{2-} (triplet), C_{60}^{-} (doublet), and C_{60}^{+} (doublet). We have used the Cl–Ch electron pair density approximation[129] to estimate the correlation energy, since it is rather reliable[129] and *computationally very fast* (about 2 min of CPU time, in the IBM-3090/400 with vector features, for the C_{60} cluster with the double-ζ basis set, and assuming no symmetry). As an alternative and as a test we have made use of the Becke density functional algorithm,[132] which is of the type of eq 11 and has been proven to be quite reliable even if notable more expensive in terms of CPU time. The two algorithms to compute the total correlation energy agree to within 98%. In Table VIII (top) we report the SCF energy, E(SCF) and the total energy obtained by adding the correlation energy correction either with the Cl–Ch or the Becke algorithms, the latter are denoted as E-(SCF+CC) and E(SCF+B), respectively. The SCF energy for the states and configurations in the order listed above are -2271.7596, -2271.7442, -2271.7481, -2271.8096, and -2271.4518 au, respectively. In Table VIII we report also the SCF binding energies per carbon atom, which are -111.36, -111.20, -111.24, -111.88, and -108.14 kcal/mol, respectively, to be compared with those corrected by adding the electronic correlation effect, given Table VIII. Two main features stand out: firstly the ionization potential decreases from 10.98 eV (double-ζ) to 10.17 eV and secondly the energy separation between the C_{60}^{-} (in the doublet state) and C_{60}^{2-} (in the singlet state) is reduced to 0.02 eV from the previous value of 0.44 eV (obtained with the double-ζ without polarization). As a consequence, the new computed value for the electron affinity is reduced from 2.75 to 1.97 eV. Expanding the basis set even further and on the basis of pilot computations reported elsewhere[78]

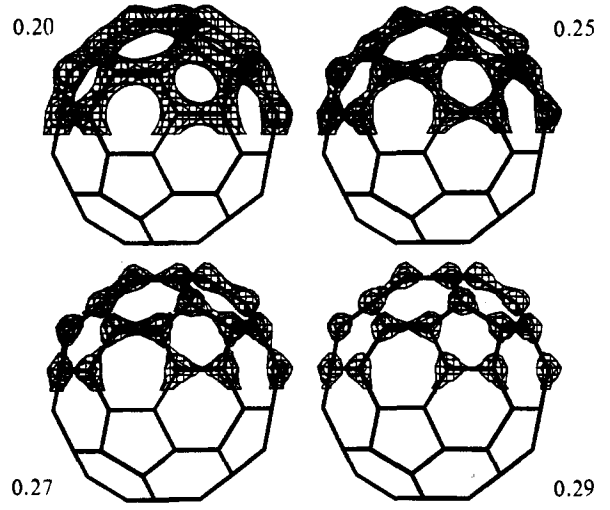

Figure 5. Four electronic densities (0.20, 0.25, 0.27, and 0.29 isocontours) for a fraction of the C_{60} (top half only) illustrating the C=C double and C—C single bonds.

we conclude that the Hartree–Fock limit value for the ionization potential is 7.95 ± 0.05 eV, rather than 8.38 eV obtained with the [(9,5,1)/(4,2,1)] basis set. Notice that the correlation effects can only increase this value; notice also that the ionization potential value has not been corrected for zero-point energy. This may suggest that the C-C to C=C bond length ratio obtained from the quantum molecular dynamics simulation needs some refinement.

From these computations on the C_{60} carbon clusters at our geometry, we would like to conclude that, at 0 K, the ionization potential, uncorrected for the zero-point energy, is in the range 9.0 ± 0.5 eV, and that the electron affinity is in the range 2.2 ± 0.3 eV. Therefore, the computed ionization potential and electron affinity are in substantial agreement with the experimental data,[133] keeping in mind the approximations adopted. In addition, the above computations on the C_{60} cluster confirm the value of R = 3.536 Å, obtained with the Car-Parrinello simulation[98] (we obtained a value of R = 3.528 Å for the ground state and about equal values for a few different electronic configurations); both values are near to the experimental one[105] of R = 3.55 Å. We have also offered preliminary predictions on the electronic spectra for a few electronic states. Finally, we have presented error bars on the computed data, providing the reader with the necessary information to assess the reliability of our computations. In Figure 5 we report the electronic density for a fraction of C_{60} obtained from the (DZ + P) computation.

We shall now discuss preliminary computations on a C_{60} cluster, at R = 3.57 Å, with, at its center, either a Li neutral atom or its positive ion. Since the Cl–Ch approximation is not reliable for intermolecular interactions, we will use the Becke algorithm only. In addition we recall that for molecular complexes the binding energy can be easily overestimated because of basis set superposition errors. In these computations we have used either the double-ζ [(9,5)/(4,2)] basis set for the carbon atoms or the double-ζ plus polarization; for lithium (either neutral atom or positive ion) we have used a [(9,1)/(4,1)] basis set with diffuse s and p Gaussian functions, with orbital exponents equal to (0.07, 0.03) and (0.1) for the two most diffuse s and for the p functions, respectively, which yields for Li (^{2}S) a SCF energy of −7.432 14 au. The computation using the double-ζ plus 3d polarization (DZ + P) brings about an interaction for the C_{60}–Li complex of +11.43 kcal/mol at the SCF level and −20.00 kcal/mol by including Becke's correlation correction (see Table VIII). The Mulliken population analysis for the lithium atom yields an ionized lithium with a positive charge value of 1.23, namely $Li^{+1.23}$, with the 1.23 electrons distributed onto the cluster. We safely predict a stable C_{60}–Li complex.

In a second computation, we have included a Li^+ (^{1}S) ion at the C_{60} cluster center. With the (DZ + P) basis set our computation yields an attraction of 2.37 kcal/mol at the SCF level, and 14.27 kcal/mol by including Becke's correlation correction (see Table VIII). Notice that the Mulliken population analysis yields a net

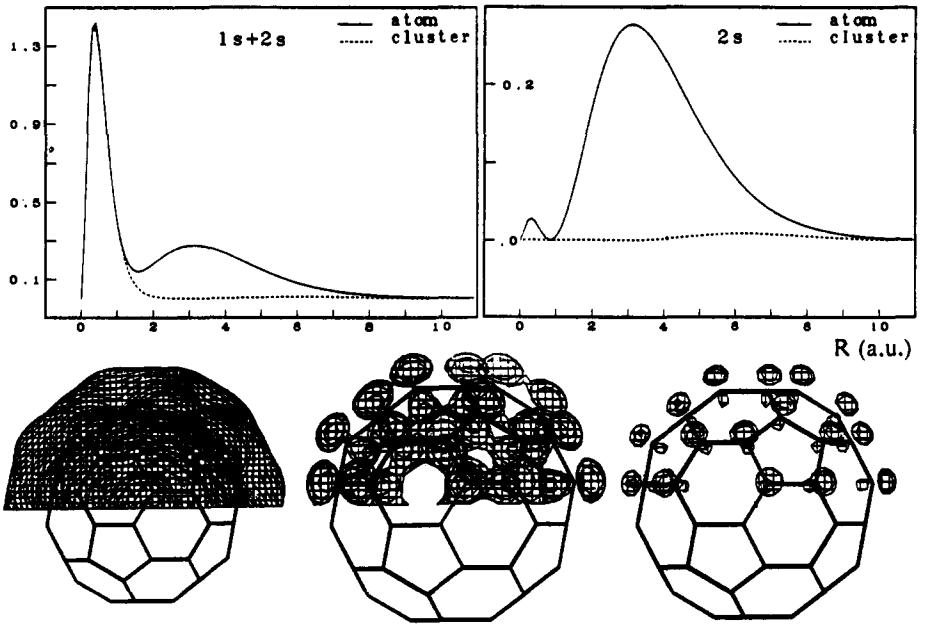

Figure 6. C_{60}–Li complex. Top: orbital densities for the 1s and 2s (left) and for the 2s electron of lithium in the cluster or in the free atom (right). Bottom: Electronic density at three isodensity contours (0.0001, 0.001, 0.002).

charge of 1.22, namely $Li^{+1.22}$, nearly the same value as in the previous computation. In Figure 6 we display the electronic density of the orbital where the electron from the lithium atom has migrated.

Notice that the ionization potential for Na and K are 5.14 eV and 4.34 eV, respectively, thus the corresponding complexes with C_{60} are expected to be stable. With the (DZ) basis set, the ionization potential for the C_{60}–Li complex is 5.41 eV by using Koopmans' theorem, and 5.20 eV by computing it from the SCF energy difference between the C_{60}–Li^+ and the C_{60}–Li clusters. The corresponding values from the (DZ + P) computations are 4.74 eV and 4.96 eV, respectively. By including also the correlation correction, we obtain an ionization potential of 6.32 eV for the (DZ) and 5.83 for the (DZ + P). Thus, we would like to predict that the ionization potential of metal–cluster complexes are near to those of the metal atom, especially for the case of alkaline atoms. Preliminary computations on the C_{60}–Na complex seems to confirm this point. Preliminary computations on the C_{60}–Ca complex yield a small repulsion, which, however, we attribute to our choice of a cluster with a too small radius.

Work is in progress to refine these computations, expanding the basis set and performing a more extensive geometry optimization. This study is computationally very demanding, and it was made possible only because we used the parallel version of our code.[126] Inclusion of symmetry would have saved considerable computer time.

Computationally, three features have been essential for this computation, namely (1) the use of coarse-grain parallelism,[126] (2) a special feature of our quantum mechanics code, KGNMOL, called ADD feature,[126] and (3) the availability of a number of density functionals to be used after the Hartree–Fock computation.[126] Let us explain in some details. Parallelism proved to be essential in the formation of the Fock matrix from the very extended 1- and 2-electron integral files, which averaged 96 Gbytes. Indeed, the bottleneck of large molecular computations is the large I/O time; N-way parallelism reduces this time N-folds. The ADD feature allows us to retain and reuse all the 2-electron integrals, for example of the C_{60} cluster, in the computation of the C_{60}–Li complex and one needs to compute only the 2-electron integrals for the added basis set of lithium and those of the basis set of lithium and the one for C_{60}. Typically, on an IBM-9000/720 the CPU time for the 2-electron integrals of C_{60} is 96 h, but for the C_{60}–Li complex is only 9 h with the ADD feature. The Cl–Ch density functional for the cluster required about 2 min of computer time; it yields correlation corrections within a few percent of the values obtained with more computer-intensive density functionals, thus, is an ideal and very inexpensive addition for any Hartree–Fock computation. We recall that the basis set for the C_{60} complex is 1812 primitive functions contracted to 967.

In the Conclusions section we shall extrapolate on the use of the techniques here adopted for the study of the C_{60} cluster, to larger molecules.

6. Molecular Dynamics with ab Initio Potentials

Sixty atoms, like in the carbon cluster, or even 348, as in the computation for 12 base pairs in DNA,[126] fall much short of most chemistry, indeed much of biochemistry and macromolecular chemistry. Thus we have to change once more approach and this brings us back to molecular dynamics simulations, but classical rather than quantum (see Figure 1). Here we count atoms, not electrons, and with an algorithm which grows with the second power of the number of atoms rather than the third (DFT) or higher powers (HF etc.) in the number of electrons. We assume point-like atoms with a mass subject to Newton's equation of motion expressed by some effective potential representing the interaction between pairs of atoms, the latter obtained either from experimental data or from ab initio computations. Below, we will discuss simulations with an ab initio derived potential.

However, the square power dependence on the number of atoms is good only for pair-wise additive interactions. Unfortunately, the classical many-body aspect is very important and cannot be neglected, but this would bring back higher power dependence. In this section we present a technique which grows as a square power of the number of atoms but recovers the main fraction of the many-body correction. In addition, the example which deals with water molecules and hydrogen bonds, brings us into a chemical system of biological interest and complexity.

There are many simulations on liquid water, but the one which we will describe is somewhat unusual. The novel features of this simulations are (a) the larger than usual number of particles, (b) the merging of two ab initio potentials, the n-body Nieser–Corongiu–Clementi (NCC) potential[134,135] (an ab initio two body with addition of polarization) and of an ab initio *flexible model* to account for vibrations,[136] (c) a discussion on a very large number of experimental data in the gas and liquid phases, all nicely reproduced.

Let us start with a brief discussion on this new potential.[135] In general, the potential energy between two flexible water molecules can be written as a sum of two contributions arising from intermolecular and intramolecular motions:

$$V(r_{\alpha\beta},q_i) = V_{\text{inter}}(r_{\alpha\beta};a_k) + V_{\text{intra}}(q_i) \qquad (13)$$

where $r_{\alpha\beta}$'s are the intermolecular atomic distances and q_i's the internal coordinates (as often done,[136] we neglect the finer coupling between the n interacting molecules, i.e. we treat a_k as constants).

The intermolecular potential here discussed is the NCC ab initio potential,[134,135] which consists of two parts:

$$V_{\text{NCC}} = \sum_{i,j<l} [V_{\text{two-body}}(i,j)] + V_{\text{pol}} \qquad (14)$$

where the pairwise additive part of the potential is very similar to another ab initio water–water potentials, known as MCY and the second part of V_{NCC} is a polarization term. Previous computations, where three-[137,138-140] and four-body[141] corrections were considered, have shown that the many-body corrections are necessary for accurate quantitative predictions.

In the NCC potential we chose an explicit representation of the polarization effects[142,143] 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

TABLE IX. Comparison of the Most Stable Configurations of the Water Dimer as Predicted by the NCC, the MCYL, and the NCC-Vib Potentials (The Atoms Belonging to the Two Water Molecules Have Been Labeled H1, H2, O5, and H3, H4, O6, Respectively)

properties	NCC	MCYL	NCC-vib
E_{min}, kcal/mol	-5.18	-5.94	-5.15
$R_{O_5-O_6}$, Å	2.97	2.87	2.97
α, deg	~4	~4	~1
β, deg	~35	~37	~28
$R_{O_5-H_1}$, Å	(0.9572)[a]	0.9564	0.9520 (0.9500)
$R_{O_5-H_2}$, Å	(0.9572)[a]	0.9665	0.9818 (0.9823)
$R_{O_6-H_3}$, $R_{O_6-H_4}$, Å	(0.9572)[a]	(0.9610)	0.9596 (0.9595)
$\angle H_1O_5H_2$, deg	(104.52)[a]	103.94	103.60
$\angle H_3O_6H_4$, deg	(104.52)[a]	104.46	104.34

[a] Experimental gas-phase geometry.[145]

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:

$$\mu_{i_\lambda}^{ind} = \tilde{\alpha}_{i_\lambda} \mathbf{E}_{i_\lambda}^{tot} \tag{15}$$

where $\mu_{i_\lambda}^{ind}$ is the λ-th induced moment on molecule i and the polarizability $\tilde{\alpha}_{i_\lambda}$ is assumed to be a static property of molecule i. By expressing $E_{i_\lambda}^{tot}$ on molecule i in function of the point charges and induced dipole moments arising from the surrounding molecules, we obtain

$$\mu_{i_\lambda}^{ind} = \tilde{\alpha}_{I_\lambda}(\mathbf{E}_{i_\lambda}^{q} + \sum_{k \neq i}^{N_m} \sum_{\upsilon}^{N_d} \tilde{T}_{i_\lambda k_\upsilon} \mu_{k_\upsilon}^{ind}) \qquad i = 1, ..., N_m; \lambda = 1, ..., N_d \tag{16}$$

where $\mathbf{E}_{i_\lambda}^{q}$ is the electric field generated by the charges q on molecule i at the λ position, N_m the number of molecules in the system, N_d the number of induced dipole moments on one molecule, and $\mathbf{T}$ the dipole–dipole matrix.

By using the above definitions, the induction energy is given by[143]

$$V_{pol} = -\frac{1}{2} \sum_{i}^{N_m} \sum_{\lambda}^{N_d} \mu_{i_\lambda}^{ind} \mathbf{E}_{i_\lambda}^{q} \tag{17}$$

In the spirit of deriving a nonempirical 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 refs 134 and 135 for additional details.

The second term of our potential, see eq 13, is again an ab initio potential for the intramolecular motions, obtained from many-body perturbation theory calculation by Bartlett, Shavitt, and Purvis.[144] The potential is expressed, up to quartic terms, in function of the three internal coordinates of water, the changes in the OH bond lengths $\delta_i = R_i - R_e$ and the HOH bond angle $\delta_3 = R_e(\theta - \theta_e)$.

Let us start by considering the water dimer. Some characteristics of the most stable dimer configuration predicted by the NCC, NCC-vib, and the MCYL potential are given in Table IX. The three potentials predict that the most stable water dimer is of the open form with a nearly linear hydrogen bond.

In the following we will discuss the results of a molecular dynamics simulation on 1000 water molecules carried out for 10.4 ps after equilibriation. The details

TABLE X. Comparison of the Geometry for the Water Molecule (Distances are Given in Å, Angles in Degrees)

	experimental		MD simulation	
	gas phase (IR)[a]	liquid phase[b]	MCYL	NCC-vib
$\langle R_{OH} \rangle$	0.9572	0.966 ± 0.006	0.975	0.978
$\langle R_{HH} \rangle$	1.514	1.510 ± 0.005	1.530	1.509
$\langle \Delta R_{OH} \rangle_{rms}$		0.095 ± 0.005	0.023	0.024
$\langle \Delta R_{HH} \rangle_{rms}$		0.090 ± 0.02	0.050	0.051
$\langle \angle HOH \rangle$	104.52	(102.8)[c]	103.5	101.1

[a] See ref 145. [b] See ref 148. [c] Calculated from $\langle R_{OH} \rangle$ and $\langle R_{HH} \rangle$.

of this simulation are given elsewhere.[146] Let us start by commenting on the temperature, energy, and pressure, namely static properties. In MD simulations the average kinetic energy, and therefore the temperature of the system, always fluctuates. In our MD simulation, the temperature was given as a target parameter and repeated normalizations of the kinetic energy were performed during the equilibriation period. The total potential energy, relative to infinitely separated water molecules, was -10.72 kcal/mol at the beginning of data collection and it varied by less than 0.006 kcal/mol during the whole simulation. Including the inter- and intramolecular zero order corrections, the above value falls in the range -9.9 kcal/mol to be compared with the experimental vaporization energy of -9.98 kcal/mol.[147] The absolute temperature of our system, 305.5 K, was determined by averaging the kinetic energies of all the atoms over the whole data collection simulation period. From our simulation we obtain the average value of the pressure $P = -2160 \pm 364$ atm, whereas it is $P = -1180 \pm 470$ atm for the rigid model.[135] We recall that water is a very noncompressible liquid and therefore a minimal deviation in the density away from experiments brings about very large variation in the simulated pressure.

Let us now compare the geometry of a water molecule in the gas[145] and in the liquid phases. There are substantial changes in bond angle and bond length between the gaseous and liquid state in our simulation. These changes are not built into the potential but reflect the collective interactions in the liquid state. The average OH bond length in the liquid is computed as 0.978 Å, about 0.020 Å longer than the gaseous value, and is in good agreement with the "experimental" value[148] of 0.966 Å. It should be noted that the "experimental" value is not a measured quantity, but rather a parameter adjusted to give a best fit in the "disentangling" of the experimental neutron scattering data. The simulated HOH bond angle, 101.1°, is found to be narrower then the experimental value of 102.8°. Also given in Table X are the root-mean-square changes, $\langle \Delta r \rangle_{rms}$, of the intramolecular OH and HH distances in the liquid; the simulated results are smaller than the experimental values, and of the same order as computed in the MCYL[136] simulation. Due mainly to the elongation of the OH bond, and the narrowing of the bond angle, the average dipole moment of water as determined by the potential is 3.11 D, larger than the rigid water model which yields a value of 2.8 D.[135] We recall that the total dipole moment for each molecule of water can change at each time step in the simulation and is recomputed with a self-consistent field procedure.

Let us comment on the radial distribution functions (RDF), compared in Figure 7 with the $g_{OO}(r)$, $g_{OH}(r)$,

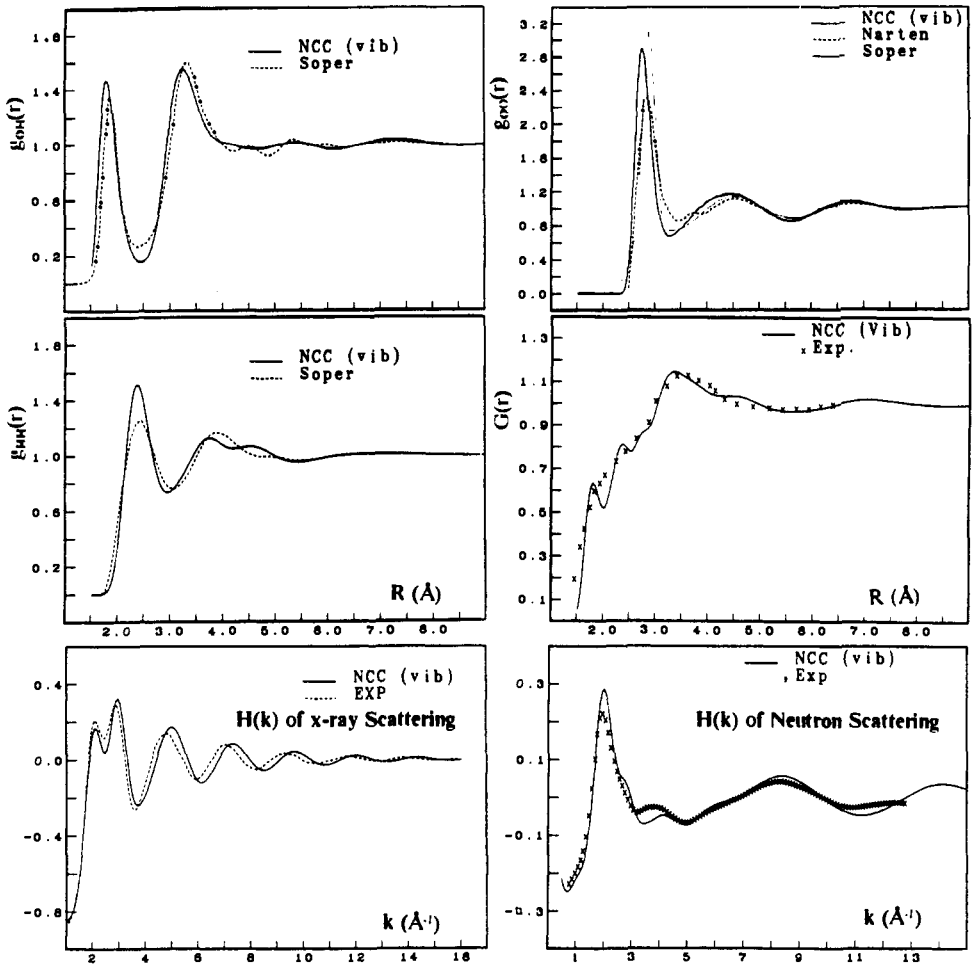

Figure 7. Top: g_{OH} and g_{OO}. Middle: g_{HO} and Dore' composition. Bottom: X-ray and neutron structure functions for liquid water.

and $g_{HH}(r)$, obtained from X-ray scattering by Narten,[149] from a time-of-flight neutron diffraction experiment by Soper,[150] and from neutron data by Dore.[151] By comparing the pair correlation functions of Figure 7 with those obtained with the rigid model potential (Figure 3 of ref 135) we notice that the inclusion of vibrations have changed the height of the first peak of the $g_{OO}(r)$ which now is more pronounced, an indication that the water molecules in the first hydration shell are more localized, namely the hydrogen bond is stronger and the amplitudes of librations smaller. Being this peak shifted toward smaller distances, the pressure of the system is noticeably affected as previously reported. The difficulties in separating the three correlation functions in experimental data are well recognized.[152] It has been suggested[151] that a more appropriate way to compare computed vs experimental pair correlation functions would be to compare the weighted sum of the three functions with the neutron data of Dore.[151] From Figure 7 the remarkable agreement between the simulated and experimental data is evident.

Since the radial distribution functions are indirect experimental data, it is desirable to calculate the X-ray and neutron scattering intensities from our simulated radial distribution functions, and compare directly the simulated with the original experimental data. In Figure 7 the computed (from our $g_{\alpha\beta}(r)$ pair correlation functions) and the experimentally determined[149] X-ray structure functions are plotted in function of k. The agreement is good. In the bottom insert of Figure 7 we compare the computed total structure functions of the neutron scattering with the most recent experimental results of Thiessen and Narten.[148] The calculated $H(k)$ curve in Figure 7 is obtained by using the MD simulated results of $g_{\alpha\beta}(r)$ as shown in Figure 7 and the intramolecular parameters given in Table X. In the figure the agreement between the simulated and experimental $H(k)$ is very good.

However, the main aim of a molecular dynamics simulation is to obtain dynamic properties; indeed, most of the static properties, previously discussed, could have been obtained with Metropolis Monte Carlo. Let us start with the translational and rotational velocity autocorrelation functions. Figure 8 (top left) shows the velocity autocorrelation functions (VACF), for the oxygen and hydrogen atoms and for the center of mass (CM) of the water molecule. These functions are normalized to $\langle \vec{v}^2(0) \rangle$.

The Fourier transform of the velocity autocorrelation functions is called the spectral density. Our results obtained by fast Fourier transform[153] of VACF are presented graphically in Figure 8 (top right). Because of their high frequencies, the band centered at 1756 cm^{-1} is undoubtedly the intramolecular bending mode, while those at 3626 and 3694 cm^{-1} must be associated with intramolecular OH bond stretching. Comparing with the classical harmonic motions of the isolated molecules, which are obtained from the intramolecular potential we used and which are also shown in Figure 8, we find that going from gas to liquid phase, there are an up shift of 71 cm^{-1} in the bending frequency and down shifts of 220 cm^{-1} and 261 cm^{-1} in the stretching

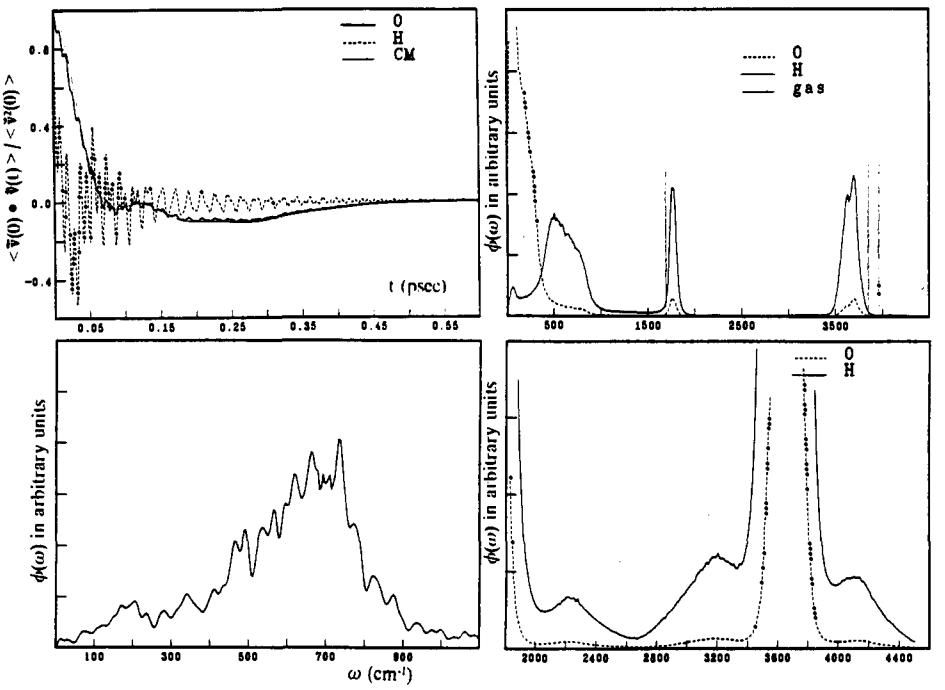

Figure 8. Top left: Velocity autocorrelation functions for oxygen and hydrogen atoms. Bottom left: Simulated infrared spectrum. Top right: Spectra density for the liquid and the gas. Bottom right: Detail of spectra density.

frequencies. These shifts are all in good agreement with the experimental IR and Raman results of 50, 167, and 266 cm^{-1}, respectively.[154]

The bending mode band (ν_2) is found at $\sim$1638 cm^{-1} in ref 157, and at $\sim$1650 cm^{-1} in refs 155 and 156. The position of this band in our simulation results (at 1756 cm^{-1}) is overestimated because the intramolecular potential overestimates this position for the gas phase (1685 cm^{-1} computed, versus 1595 cm^{-1} experimental),[155,156] but both, computed and experimental, agree in the shift, for the liquid phase, toward higher frequencies. In the Raman spectrum[157] is also detected a combination band at higher frequencies than ν_2, at $\sim$2130 cm^{-1}. This band, also present in our simulated spectrum in the region $\sim$2230 cm^{-1} (see the bottom right insert of Figure 8, which enlarges the region 1800–4400 cm^{-1}), has been associated to a combination band ($\nu_2 + \nu_L$) where L is a librational mode. It should be noted that the positions of the ν_2 and $\nu_2 + \nu_L$ differ by $\sim$500 cm^{-1}, as found experimentally.[157]

The bands at 3626 and 3694 cm^{-1} must be associated with the symmetric (ν_1) and asymmetric (ν_3) stretching modes, respectively. The resolution into ν_1 and ν_3 is not well resolved in the Raman spectra,[157] but a large band is observed at $\sim$3400 cm^{-1}. A small band is observed[157] at $\sim$3230 cm^{-1} which is attributed to a bending-mode overtone ($2\nu_2$) and its Fermi resonance with the stretching modes. Interestingly, we find a low intensity band at the same frequency, as it is evident in the bottom right insert of Figure 8. Also present in the same figure is an additional high frequency band at $\sim$4180 cm^{-1} also predicted in the MCYL spectrum, and later confirmed by neutron-scattering experiments.[158] This band has been interpreted[158] as a combination band arising from simultaneous excitations of the stretch vibrational mode and a breaking of the adjacent hydrogen bond.

Since the center of mass (c.m.) of the water molecule is very close to the oxygen atom, the drastic intensity difference between the $\phi(\omega)$'s of hydrogen and oxygen in Figure 8 (top right) allows us to identify immediately that the broad band centered around 500 cm^{-1} is due mainly to the rotational motions of the molecules, whereas the band centered around 65 cm^{-1} and the large shoulder centered around 190 cm^{-1} arise from the hindered translational motions. In the experimental IR spectra,[154] there is also a prominent shoulder near 193 cm^{-1} to be identified with the motions of oxygen atoms as Figure 8 (top right inset) shows. The peak (not fully reported) centered at 65 cm^{-1} in our simulated spectral density should correspond to the narrow band which appears at $\sim$60 cm^{-1} in the Raman and inelastic neutron scattering spectra.[154]

Experimentally,[154] there is a broad and intense band in the IR spectrum near 700 cm^{-1} and extends from 300 cm^{-1} to above 900 cm^{-1}. This can certainly be identified with the motions represented by the broad band extending from 300 to $\sim$1100 cm^{-1} in our $\phi(\omega)$. However, as done for the NCC potential,[135] if we compute this range of the spectrum not from the Fourier transform of the velocity pair correlation function, but from the Fourier transform of the single dipole moment correlation function, we obtain the spectrum presented in the bottom left insert of Figure 8, where the band, is clearly centered around 700 cm^{-1}, in very good agreement with the experimental findings.[154] This concludes the simulations carried out to test the NCC potential; we note that the above list of predictions is, to our knowledge, the most comprehensive one in literature to test the reliability of a potential—with the same model and parameters—in simulations of water in the gas and liquid phases.

Similar and more standard MD simulations for liquid water, but without the intermolecular flexibility, have been reported from a periodic system of 512 water molecules, at 310 K, by using a sixth-order Gear-predictor corrector algorithm, Ewald-sums, a cutoff radius of 12.4 Å, a time step of 0.5 fs, and a simulation length

of 32 ps after equilibriation. The computer time for this simulation is 53 h for equilibriation and 34 h for the collection of statistics on an IBM ES/3090-600J running in parallel with 6 processors. The entire calculation takes about 40 Mbytes of main memory. The simulation with the flexible model takes 4 times longer, due to the smaller time step. Notice that by increasing the sample from 512 to 1000 molecules, the execution time increases by a factor of about 8.

The by now spontaneous question relates to the feasibility to simulate liquids and/or solutions for hundreds or even thousands of picoseconds, possibly without a proportional increase of the already very substantial amount of computer time. This problem is discussed in the next section.

7. Nanosecond Simulations with Stochastic Dynamic

The dynamic fluctuations of proteins around their average conformations play an important role in many biological processes such as enzyme activity, macromolecular recognition, and complex formations. In the MD approach,[159-159,163] given the interaction force field, the dynamic evolution of the protein atoms are determined by solving Newton's equations of motion (see Figure 1). 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.[161,164]

In this section we shall concentrate on the effect of the solvent molecule on protein dynamics.[165,167] 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 microseconds 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[168-174] wherein the dynamic evolution of the biomolecule is studied in a 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.

The computational techniques for stochastic dynamic simulations have evolved along three different pathways. One successful approach, due to Ermak and McCammon,[169] is based on the assumption that the mass of the Brownian particles are sufficiently large so that the momentum relaxation phenomena can be ignored and the dynamic trajectories can be obtained by solving a Fokker–Planck equation in a configuration space. The algorithm has been employed for investigating the bending and twisting dynamics of short linear DNAs, the folding dynamics of two α helices connected by a flexible polypeptide, and the diffusion of a substrate on an active site of an enzyme.[170,171] The second approach is due to Ciccotti et al. and is based on generalized brownian dynamics wherein the coupled Langevin equations are solved.[172] Due to the complexity inherent in the accurate determination of the memory functions, the application of this method has been restricted to rather simple systems. The third approach, which we use below, is due to Allen[173] and van Gunsteren and Berendsen.[174] In this approach of the stochastic dynamic simulation, 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. The equation of motion then is written as

$$m_i\dot{\nu}_i = -m_i\gamma_i\nu_i + F_i + R_i(t) \tag{18}$$

$$\langle R_i(0)R_j(t)\rangle = 2m_i\gamma_i k_b T_{\text{ref}}\delta_{ij}\delta(t) \tag{19}$$

$$W(R_i) = [2\pi\langle R_i^2\rangle]^{-1/2}\exp\left[-\frac{R_i^2}{2\langle R_i^2\rangle}\right] \tag{20}$$

$$\langle R_i\rangle = 0 \tag{21}$$

$$\langle \nu_i(0)R_j(t)\rangle = 0 \tag{22}$$

$$\langle F_i(0)R_j(t)\rangle = 0 \tag{23}$$

Here $\langle ... \rangle$ denotes the averaging over an equilibrium ensemble, k_b is the Boltzmann constant, T_f is the reference temperature, and $W(R_i)$ is the Gaussian probability distribution function of the random force, and γ_i are the friction coefficients.

In this section 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.[161,175-179] 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.[175] Smith et al. analyzed the incoherent neutron scattering in a BPTI molecule.[179] 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.[166] Cusak et al. measured the time-of-flight spectra and the generalized density of states in the BPTI protein.[177-180] Clementi et al. conducted two 52-ps molecular dynamic simulations[161] 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.,[181] where the hydrogen bond is represented by an explicit term. (Presently, we are recomputing, this time with an ab initio potential.)

The results of the stochastic simulations are compared with the molecular dynamic simulation of BPTI in water conducted by Clementi et al.,[161] where the dynamic evolution of 892 atoms of BPTI and 2676 water molecules was studied for 52 ps.

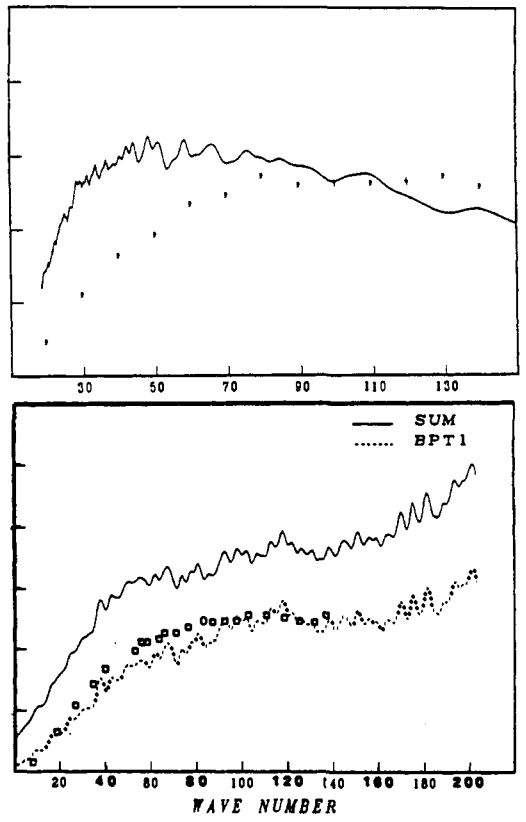

Figure 9. Comparison between experimental and simulated density of states for BPTI: top, SDS simulation (full line); experiment (crosses); bottom, MD simulation, BPTI contribution (dotted line), sum of BPTI and D_2O contribution (full line), and experiment (squares).

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 use 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 × 120.0 × 120.0 Å^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 s. The molecular dynamic simulation of the BPTI protein in 2676 water molecules took 48 s per time step, while for the molecular dynamic simulation in vacuo the CPU time per time step was 2.69 s. Thus the stochastic dynamic simulation was approximately 16 times faster than the molecular dynamic simulation in solution. In the stochastic dynamic simulation the average RMS deviation was around 0.8 Å which is considerably less than the one found in the molecular dynamic simulation[161] in vacuo (2.7 Å) but close to the RMS deviation for the molecular dynamic simulation in solution[161] (~1.2 Å).

Let us now compare the MD and SDS trajectories by analyzing the incoherent neutron scattering spectra and the generalized density of states. In Figure 9 (top inset) we have shown the generalized density of states computed from the 8192 configurations collected during 32-ps stochastic dynamics simulations. The neutron-scattering intensities are calculated from the SDS trajectories in the frame work of the classical time correlation functions.[161] The corresponding molecular dynamics results[161] are also shown in Figure 9 (bottom inset). Our stochastic dynamic simulation are in reasonable agreement with the molecular dynamics results. Both the molecular and stochastic simulation results are consistent with the earlier theoretical analysis based on the normal mode method.[179,180] The shape of the curve is also consistent with the available experimental scattering intensities from the powder sample.[177] 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.[180] who argue that small conformation modifications of protein due changes in surface interaction upon addition of water modifies the mode spectrum such that there is a net shift from higher to lower frequencies.

TABLE XI. Ground-State Energies (au) for the Hartree–Fock and Fock–Dirac with and without Breit Corrections

atom	H-F	F-D	F-D + B
Ne	-128.547 10	-128.691 93	-128.675 29
Ar	-526.817 51	-528.683 79	-528.551 47
Kr	-2 752.055 0	-2 788.862 2	-2 787.436 2
Xe	-7 232.138 4	-7 446.903 0	-7 441.133 1

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 1 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.[7-9] 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 was sketched in Figure 1. 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. Conclusions

All the computations above neglect relativistic effects. Whereas this might not be too inconsistent for an accurate computation on H_2 or H_3 aiming at 1–10 cm^{-1} accuracy, it clearly becomes a more and more questionable approximation when one deals with molecules and with atoms past neon and aims at millihartree accuracy even in relative energies. It has been noted long ago[182] that relativistic effects scale as the second to third power of the atomic number (Z^2 and Z^3) while correlation effects scale with the number of electrons (Z for neutral atoms) and are taken over by the former (the relativistic ones) at about Z = 12. This "caveat" will

likely be acknowledged more and more in the coming years. In Table XI we report some recent results obtained in our laboratory for atomic computations which include the Breit term into the Fock–Dirac Hamiltonian.[183] Notice that the Hartree–Fock and Fock–Dirac energies differ in the first decimal in atomic units for Ne but for Xe are already into the hundreds of atomic units. The same observation holds for the Fock–Dirac with and without Breit correction: from a few hundredths of atomic units for Ne to a few atomic units for Xe. These values should make us a bit less optimistic on assuming cancellations of relativistic corrections when comparing atoms with molecules and we seek chemical accuracy, namely 0.001 au. If we compare a Hartree–Fock computation with a Fock–Dirac computation for a given molecule and with the "same" basis set we note that the latter will require about 5–10 times more CPU time than the former. There are three main reasons for this. Firstly, the kinetic balance[184] brings about an effective increase in the starting basis set; then the SCF part is more complex in the Fock–Dirac formalism, and finally if one starts with cartesian Gaussian functions they need to be transformed into spherical functions. In conclusion, presently relativistic computations remain problematic mainly because we are still searching for refinements of the methods, but once this problem is solved, these calculations, even if 5–10 times more expensive than equivalent nonrelativistic computations, will become *standard and common* because of the increased performance gain of the VIth generation relative to the IVth.

Let us now consider the techniques discussed in this work starting with the H_3 surface, aiming at a 1.0–10.0 cm^{-1} accuracy and with the HCI technique. From the discussion in section 2, it appears that—assuming one has improved the numerical analyses and the code—a reasonably accurate surface can be obtained with a 50–100 GFLOPS system. Keeping in mind the recent performance announcements for example by the Japanese computer industry, there is reason to assume that this problem will have an accurate HCI solution before the end of this century, namely from the VIth generation computers. However, one could consider alternatives, particularly quantum Monte Carlo techniques[3] or geminal techniques.[46] An important merit of the HCI is having helped our understanding for the Coulomb hole cusp; this aspect likely will be exploited in years to come.

Essentially the same conclusion holds for atomic computations with the CI type technique: systems with a 50 GFLOPS performance will be most useful. There is, however, an apparent need to expand the ATOMCI code and corresponding technique either in the direction of multireference CI or of MCSCF followed by extended CI. In addition, present efforts by Prof. Sasaki and co-workers to parallelize the code are another most welcome step. However one could also look at alternatives like "direct" techniques,[185] thus using many more configurations and at the same time circumventing the I/O bottleneck. But it is unlikely that we will see many computations with 50–100 electrons and with an *absolute* accuracy of 0.001 au in this century. True, for much of chemistry, relative accuracy is what it is needed, but is equally true that the "relative accuracy requirement" has been often abused by quitting a computation as soon agreement with experimental data has been obtained.

Let us now move to medium and medium-large molecules with 200–2000 electrons; this is the field for either DFT or for the approach we have used with C_{60}, namely Hartree–Fock and correlation energy by density functionals. The main problem is the old one for quantum chemistry: too small memories, and/or insufficient bandwidth from intermediate storage devices to random access memory.

Let us now move to even larger systems, 1000 molecules of water. This larger than usual number of water molecules is of interest mainly for a proper treatment of collective effects, like the dielectric constant and sound wave simulations, where there are large fluctuations. Can we simulate this system using quantum mechanics at each time step for energy and forces? Alternatively stated, we ask whether one can use the Car–Parrinello technique.[84] Scaling from 32 000 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 h per time step with a storage requirement of 20 Gbytes. This is contrasted to about 1.4 min per time step when ab initio polarization potentials of NCC type and a flexible water are used. While a single Car–Parrinello time step might be reduced to less than 1 h on a large VIth generation computer, the 10^5 required time steps seem to make this task infeasible within this century, assuming a requirement for comparable accuracy. Notice that the number of plane waves could be considerably smaller by using a mixed basis, namely plane waves and Gaussian orbitals localized on the nuclei (work is in progress in our laboratory along this line).

We would like to forecast that direct use of quantum mechanics will become more and more diffuse in molecular dynamics computations. Considering the analysis made above however, it seems that one will have to merge energies and forces obtained quantum mechanically with energies and forces derived from ab initio potentials. This will be particularly important for the study of the active site in protein and enzymes. Indeed, only in the neighborhood of the active site one will follow the reaction at the quantum mechanical level while all the remaining interactions would be treated at the classical level, using ab initio derived potentials. A precursor of this approach is a study of the proton transfer in papain.[186]

Let us now move from MD to SDS. In this review we were mainly concerned with the compatibility of the stochastic simulation results with those from molecular dynamics simulations. However the real promise of the stochastic dynamic simulation lies in its ability to describe the large scale motion in biomolecules. To achieve this goal we have to further reduce the degrees of freedom. Instead of the atomic representation employed in the present study we can rewrite the Langevin equations in an extended atomic representation or in a residue representation. In the residue representation the dynamics of the BPTI protein would be studied by computing the evolution of the 58 residues constituting the protein. This method which we call the extended stochastic dynamic method, being orders of magnitude faster than currently used molecular dynamic simula-

tions, can be applied to study the cooperative phenomena giving rise to the protein folding process. However, progress in this direction requires the development of new tools, the most basic among these being the development of a new set of potential describing the interaction between the residues. On a large time scale as the protein unfolds, the distribution of the solvent molecule is likely to change and one might need to recalculate the distribution of friction coefficients. Thus a long SDS run can be punctuated by the short molecular dynamics run wherein the new set of friction coefficients are determined taking into account the deformation of the protein surface as a consequence of the unfolding process.

The present simulation and the above comments allow us to extrapolate, even if in a very approximate manner, the speed one can obtain from the extended stochastic dynamic simulation of the BPTI protein. Given that the CPU time for calculating the interaction forces among protein atoms (or residues) scales as the square of the number of the protein atoms (or residues) and assuming that the potential needed for describing the interaction among residues requires M times more CPU time than one needed in atomic representation, the net speed up from the stochastic dynamic simulation to extended stochastic dynamic simulation is proportional to $(1/M) \times (892/58)$.[2] Thus for $M = 4$, the speed up is around 59, while for $M = 10$ the speed up is more than 20. In comparison to a molecular dynamic simulation wherein the solvent molecules are explicitly considered, the gain in the residue representation would be between 944 and 320. Thus we have proven that in the framework of the extended stochastic dynamic simulation the study of protein dynamics in the nanoseconds time scale will soon become feasible.

The common denominator in our conclusions is that if we rely too much on the increased performance offered by the computer industry, then computational chemistry will only inch forward at best. The main avenue lies in alternative methods and new techniques. Thus quantum Monte Carlo, ab initio derived polarization potentials, Car–Parrinello and its extension (as outlined above), MBPT, geminals, and direct techniques seem to represent the future of quantum chemistry. Techniques designed to move away from basis set expansion, toward numerical methods, should be examined with rigor, both to escape or reduce the BSS error and the N^4 dependency. At the other side of the spectrum, semiempirical methods, density functional theory, including Wigner-type correction[125,126] added to the Hartree–Fock, should be also pursued with vigor. Finally quantum chemistry should strive to merge more deeply into chemistry and computational chemistry; this will require much more attention than presently given to data base, expert systems, interactive graphics and animation, artificial intelligence and parallelism, both coarse and massive.

In this context we note that SDS appears to have some characterization ideally suited for a link with neuronal network (NN) simulation studies, presently probed for an eventual solution to biological problems like molecular recognition and protein folding. At our laboratory we have recently concluded an NN simulation on BPTI and we are now using the NN output as starting guess (input) for our SDS program. The dimensions of Figure 1 are thus expanding further and further.

We summarize our conclusions in one recommendation: "look into new techniques, but connect old ones."

References

(1) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 230.
(2) Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1957**, *27*, 1208.
(3) Kalos, M. H. *Phys. Rev.* **1962**, *128*, 1791. Ceperley, D. M.; Alder, B. J. *Science* **1986**, *231*, 555. Anderson, J. *J. Chem. Phys.* **1987**, *85*, 2839, and references cited therein.
(4) Parrinello, M. In *Modern Techniques in Computational Chemistry: MOTECC-90*; Clementi, E., Ed.; ESCOM: Leiden, 1990; pp 731–744.
(5) Bhattacharya, D. K.; Clementi, E. In *Modern Techniques in Computational Chemistry: MOTECC-90*; Clementi, E., Ed.; ESCOM: Leiden, 1990; pp 889–918.
(6) De Callatay, A. M. *Natural and Artificial Intelligence*; North-Holland: Amsterdam, 1986.
(7) Clementi, E. *MOTECC-89*. Clementi, E., Ed.; ESCOM Publisher: Leiden, 1989; Chapter 1.
(8) Clementi, E. *MOTECC-90*. Clementi, E., Ed.; ESCOM Publisher: Leiden, 1990; Chapter 1.
(9) Clementi, E. *MOTECC-91*. Clementi, E., Ed.; ESCOM Publisher: Leiden, 1991; Chapter 1.
(10) Bartlett, R. J. *Ann. Rev. Phys. Chem.* **1981**, *32*, 359. Urban, M.; Cernusak, I.; Kellö, V.; Noga, J. In *Methods in Computational Chemistry*; Wilson, S., Ed.; Plenum: New York, 1987.
(11) Hylleraas, E. A. *Z. Phys.* **1929**, *54*, 347.
(12) James, H. M.; Coolidge, A. S. *J. Chem. Phys.* **1933**, *1*, 825.
(13) Kolos, W.; Wolniewicz, L. *J. Chem. Phys.* **1964**, *41*, 3663; **1965**, *43*, 2429; *Phys. Rev. Lett.* **1968**, *20*, 243.
(14) Mulliken, R. S.; Ermler, W. C. *Diatomic Molecules: Results of Ab Initio Calculations*; Academic: New York, 1977.
(15) Sims, J. S.; Hagström, S. *Phys. Rev. A* **1971**, *4*, 908.
(16) Sims, J. S.; Hagström, S. *J. Chem. Phys.* **1971**, *55*, 4699.
(17) Preiskorn, A.; Zurawski, B. *Int. J. Quantum Chem.* **1985**, *27*, 641.
(18) Ohrn, Y.; Nordling, J. *J. Chem. Phys.* **1963**, *39*, 1864.
(19) Sims, J. S.; Hagström, S.; Rumble, J. R. *Phys. Rev. A* **1976**, *13*, 242.
(20) Sims, J. S.; Hagström, S.; Munch, D.; Bunge, C. *Phys. Rev. A* **1976**, *13*, 560.
(21) Sims, J. S.; Hagström, S.; Rumble, J. R. *Int. J. Quantum Chem.* **1976**, *10*, 853.
(22) Clary, D. C.; Handy, N. C. *Phys. Rev. A* **1976**, *14*, 1607.
(23) Bethe, H. A.; Salpeter, E. E. *Quantum Mechanics of One- and Two-Electron Atoms*; Plenum/Rosetta: New York, 1977.
(24) Kutzelnigg, W. *Theor. Chim. Acta* **1985**, *68*, 445.
(25) Frye, D.; Preiskorn, A.; Lie, G. C.; Clementi, E. In *Modern Techniques in Computational Chemistry: MOTECC-90*; Clementi, E., Ed.; ESCOM: Leiden, 1990; pp 235–275.
(26) Frye, D.; Lie, G. C.; Clementi, E. HYCOIN: Documentation and User's Guide. IBM Technical Report KGN-180; IBM: Kingston, 1989.
(27) Frye, D.; Preiskorn, A.; Clementi, E. *J. Comp. Chem.*, submitted for publication.
(28) Urdaneta, C.; Largo-Cabreizo, A.; Lievin, J.; Lie, G. C.; Clementi, E. *J. Chem. Phys.* **1988**, *88*, 2091.
(29) Frye, D.; Lie, G. C.; Clementi, E. *J. Chem. Phys.* **1989**, *91*, 2366.
(30) Frye, D.; Lie, G. C.; Clementi, E. *J. Chem. Phys.* **1989**, *91*, 2369.
(31) Preiskorn, A.; Lie, G. C.; Frye, D.; Clementi, E. *J. Chem. Phys.* **1990**, *92*, 4941.
(32) Frye, D.; Preiskorn, A.; Lie, G. C.; Clementi, E. *J. Chem. Phys.* **1990**, *92*, 4948.
(33) Preiskorn, A.; Frye, D.; Clementi, E. *J. Chem. Phys.*, submitted for publication.
(34) Frye, D.; Lie, G. C.; Clementi, E. HCl Calculations on the Potential Energy Curve for the $\chi^1\Sigma_g^+$ State of H_2. IBM Technical Report KGN-179; IBM: Kingston, 1989.
(35) Kolos, W.; Szalewicz, K.; Monkhorst, H. *J. Chem. Phys.* **1986**, *84*, 3259.
(36) Liu, B. *J. Chem. Phys.* **1984**, *80*, 581.
(37) Feller, D. Private communication.
(38) Salmon, L.; Poshusta, R. D. *J. Chem. Phys.* **1973**, *59*, 3497.
(39) Mentch, F.; Anderson, J. B. *J. Chem. Phys.* **1981**, *74*, 6307.
(40) Preiskorn, A.; Woznicki, W. *Mol. Phys.* **1984**, *52*, 1291.
(41) Burton, P. G.; Von N.-Felsobuki, E.; Doherty, G.; Hamilton, M. *Mol. Phys.* **1985**, *55*, 527.
(42) Meyer, W.; Botschwina, P.; Burton, P. *J. Chem. Phys.* **1986**, *84*, 891.
(43) Anderson, J. B. *J. Chem. Phys.* **1987**, *86*, 2839.

(44) Traynor, C. A.; Anderson, J. B. *Chem. Phys. Lett.* **1988**, *147*, 389.
(45) Huang, S.; Sun, Z.; Lester, W. A., Jr. *J. Chem. Phys.* **1990**, *92*, 597.
(46) Alexander, S. A.; Monkhorst, H. J.; Roeland, R.; Szalewicz, K. *J. Chem. Phys.* **1990**, *93*, 4230.
(47) Davidson, E. In *Modern Techniques in Computational Chemistry: MOTECC-90*; Clementi, E., Ed.; ESCOM: Leiden, 1990; pp 553–592.
(48) Kutzelnigg, W.; Klopper, W. *J. Chem. Phys.* **1991**, *94*, 1985.
(49) Termath, V.; Klopper, W.; Kutzelnigg, W. *J. Chem. Phys.* **1991**, *94*, 2002.
(50) Klopper, W.; Kutzelnigg, W. *J. Chem. Phys.* **1991**, *94*, 2020.
(51) Rizzo, A.; Clementi, E.; Sekiya, M. *Chem. Phys. Lett.*, in press.
(52) Sasaki, F.; Sekiya, M.; Noro, T.; Ohtsuki, K.; Osanai, T. In *Modern Techniques in Computational Chemistry, MOTECC-90*; Clementi, E., Ed.; ESCOM Science Publisher: Leiden, 1990; p 181.
(53) Sekiya, M.; Noro, T.; Ohtsuki, K.; Sasaki, F.; Rizzo, A.; Clementi, E. In *MOTECC-90: Input/Output Documentation*; a special publication by the Department of Scientific Engineering Computations, Dept 48B/MS 428, IBM: Kingston, 1990; p 37.
(54) Sasaki, F. *Int. J. Quantum Chem.* **1974**, *8*, 605.
(55) Sasaki, F.; Yoshimine, M. *Phys. Rev. A* **1974**, *9*, 17.
(56) Sasaki, F.; Yoshimine, M. *Phys. Rev. A* **1974**, *9*, 26.
(57) Clementi, E.; Corongiu, G. *Chem. Phys. Lett.* **1982**, *90*, 359.
(58) Pekeris, C. L. *Phys. Rev.* **1958**, *112*, 1649.
(59) Pekeris, C. L. *Phys. Rev.* **1959**, *115*, 1216.
(60) Davidson, E. R. *Int. J. Quantum Chem.* **1990**, *37*, 811.
(61) Veillard, A.; Clementi, E. *J. Chem. Phys.* **1968**, *49*, 2415.
(62) Larsson, S. *Phys. Rev.* **1968**, *169*, 49.
(63) Bunge, C. F. Private communication.
(64) Sundholm, D.; Olsen, J. *Phys. Rev.* **1990**, *42*, 2614.
(65) Bunge, C. F. *Phys. Rev. A* **1976**, *14*, 1965.
(66) Clementi, E. *J. Chem. Phys.* **1963**, *38*, 2248.
(67) Health, J. R.; O'Brien, S. C.; Curl, R. F.; Kroto, H. W.; Smalley, R. E. *Com. Cond. Matt. Phys.* **1987**, *13*, 119.
(68) Bethune, D. S.; Meijer, G.; Tang, W. C.; Rosen, H. J.; Golden, W. G.; Seki, H.; Brown, C. A.; de Vries, M. S. *Chem. Phys. Lett.* **1990**, submitted for publication.
(69) Meijer, G.; Bethune, D. S.; Tang, W. C.; Rosen, H. J.; Johnson, R. D.; Wilson, R. J.; Chambliss, D. D.; Golden, W. G.; Seki, H.; de Vries, M. S.; Brown, C. A.; Salem, J. R.; Hunziker, H. E.; Wendt, M. R. *Mat. Res. Soc. Symp.* **1990**, (Boston, MA).
(70) Johnson, R. D.; Meijer, G.; Bethune, D. S. *J. Am. Chem. Soc.* **1990**, *112*, 8983.
(71) Haddon, R. C.; Brus, L. E.; Raghavachari, K. *Chem. Phys. Lett.* **1986**, *125*, 459.
(72) Larsson, S.; Volosov, A.; Rosen, A. *Chem. Phys. Lett.* **1987**, *137*, 501.
(73) Rosen, A.; Wastberg, B. *J. Chem. Phys.* **1989**, *90*, 2525.
(74) Lüthi, H. P.; Almlöf, J. *Chem. Phys. Lett.* **1987**, *135*, 357.
(75) Newton, M. D.; Stanton, R. E. *J. Am. Chem. Soc.* **1986**, *108*, 2469.
(76) Kratschmen, W.; Fostiropoulos, K.; Huffman, D. R. *Chem. Phys. Lett.* **1990**, *170*, 167.
(77) Painter, G. P.; Averill, F. W. *Phys. Rev. B* **1982**, *26*, 1781.
(78) Corongiu, G.; Clementi, E. *Int. J. Quantum Chem.*, in press.
(79) Ballone, P.; Miliani, P. Private communication.
(80) Tersoff, J. *Phys. Rev. Lett.* **1988**, *61*, 2879.
(81) Ballone, P.; Andreoni, W.; Car, R.; Parrinello, M. *Europhys. Lett.* **1989**, *8*, 73.
(82) Ballone, P.; Andreoni, W.; Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1988**, *60*, 271.
(83) Ballone, P.; Andreoni, W. *Phys. Scripta* **1987**, *T19*, 289.
(84) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471.
(85) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864.
(86) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133.
(87) Kryachko, E. S.; Ludeña, E. V. *Energy Density Functional Theory of Many-Electron Systems*; Kluwer Academic Press: Dordrecht, 1990.
(88) Andreoni, W.; Scharf, D. In press.
(89) Bachelet, G. B.; Hamann, D. R.; Schulter, M. *Phys. Rev. B* **1982**, *26*, 4199.
(90) Perdew, J. P.; Zunger, A. *Phys. Rev. B* **1981**, *23*, 5048.
(91) Ceperley, D. M.; Alder, B. J. *Phys. Rev. Lett.* **1980**, *45*, 566.
(92) Raghavachari, K.; Binkley, J. S. *J. Chem. Phys.* **1987**, *87*, 2191.
(93) Pitzer, K. S.; Clementi, E. *J. Am. Chem. Soc.* **1959**, *81*, 4477.
(94) Weeks, D. E.; Harter, W. G. *J. Chem. Phys.* **1989**, *90*, 4744.
(95) Brendsdal, E.; Cyvin, B. N.; Brunvoll, J.; Cyvin, S. J. *Spectrosc. Lett.* **1988**, *21*, 313.
(96) Wu, Z. C.; Jelski, D. A.; George, T. F. *Chem. Lett.* **1987**, *137*, 291.
(97) Stanton, R. E.; Newton, M. D. *J. Phys. Chem.* **1988**, *92*, 2141.
(98) Feuston, B. P.; Clementi, E.; Andreoni, W.; Parrinello, M. *Phys. Rev. Lett.*, in press.
(99) Kroto, H. W.; Heath, J. R.; O'Brien, S. C.; Curl, R. F.; Smalley, R. E. *Nature* **1985**, *318*, 162.
(100) Heath, J. R.; O'Brien, S. C.; Zhang, Q.; Liu, Y.; Curl, R. F.; Kroto, H. W.; Tittel, F. K.; Smalley, R. E. *J. Am. Chem. Soc.* **1985**, *107*, 7779.
(101) Zhang, Q.; O'Brien, S. C.; Heath, J. R.; Liu, Y.; Curl, R. F.; Kroto, H. W.; Smalley, R. E. *J. Phys. Chem.* **1985**, *90*, 528.
(102) Liu, Y.; O'Brien, S. C.; Zhang, Q.; Heath, J. R.; Tittel, F. K.; Curl, R. F.; Kroto, H. W.; Smalley, R. E. *Chem. Phys. Lett.* **1986**, *126*, 215.
(103) O'Keefe, A.; Ross, M. M.; Baronavski, A. P. *Chem. Phys. Lett.* **1986**, *130*, 17.
(104) Heath, J. R.; Curl, R. F.; Smalley, R. E. *J. Chem. Phys.* **1987**, *87*, 4236.
(105) Krätschmer, W.; Lamb, L. D.; Fostiropoulos, K.; Huffman, D. R. *Nature* **1990**, *347*, 354.
(106) Haufler, R. E.; Conceicao, J.; Chibante, L. P. F.; Chai, Y.; Byrne, N. E.; Flanagan, S.; Haley, M. M.; O'Brien, S. C.; Pan, C.; Xiao, Z.; Billups, W. E.; Ciufolini, M. A.; Hauge, R. H.; Margrave, J. L.; Wilson, L. J.; Curl, R. F.; Smalley, R. E. *J. Phys. Chem.*, in press.
(107) Haymet, A. D. J. *Chem. Phys. Lett.* **1985**, *122*, 421.
(108) Satpathy, S. *Chem. Phys. Lett.* **1986**, *130*, 545.
(109) Fowler, P. W.; Woolrich, J. *Chem. Phys. Lett.* **1986**, *127*, 78.
(110) Ozaki, M.; Takahashi, A. *Chem. Phys. Lett.* **1986**, *127*, 242.
(111) Stone, A. J.; Wales, D. J. *Chem. Phys. Lett.* **1986**, *128*, 501.
(112) Schmalz, T. G.; Seitz, W. A.; Klein, D. J.; Hite, G. E. *Chem. Phys. Lett.* **1986**, *130*, 203.
(113) Newton, M. D.; Stanton, R. S. *J. Am. Chem. Soc.* **1986**, *108*, 2469.
(114) Hale, P. D. *J. Am. Chem. Soc.* **1986**, *108*, 6087.
(115) Klein, D. J.; Schmaltz, T. J.; Hite, G. E.; Seitz, W. A. *J. Am. Chem. Soc.* **1986**, *108*, 1301.
(116) Haymet, A. D. J. *J. Am. Chem. Soc.* **1986**, *108*, 319.
(117) Fowler, P. W. *Chem. Phys. Lett.* **1986**, *131*, 444.
(118) Laszlo, I.; Udvardi, L. *Chem. Phys. Lett.* **1987**, *136*, 419.
(119) Shibuya, T. I.; Yoshitani, M. *Chem. Phys. Lett.* **1987**, *137*, 13.
(120) Schulman, J. M.; Disch, R. L.; Miller, M. A.; Peck, R. C. *Chem. Phys. Lett.* **1987**, *141*, 45.
(121) Negri, F.; Orlandi, G.; Zerbetto, F. *Chem. Phys. Lett.* **1988**, *144*, 31.
(122) Schmalz, T. G.; Seitz, W. A.; Klein, D. J.; Hite, G. E. *J. Am. Chem. Soc.* **1988**, *110*, 1113.
(123) Feng, J.; Li, J.; Wang, Z.; Zerner, M. *Int. J. Quantum Chem.* **1990**, *37*, 599.
(124) Disch, R. L.; Schulman, J. *Chem. Phys. Lett.* **1986**, *125*, 465.
(125) Wigner, E. P. *Phys. Rev.* **1934**, *46*, 1002. Wigner, E. P.; Seitz, F. *Phys. Rev.* **1933**, *43*, 804; *Trans. Faraday Soc.* **1938**, *34*, 678.
(126) Clementi, E.; Corongiu, G.; Chakravorty, S. *MOTECC-90*; Clementi, E., Ed.; ESCOM Publisher: Leiden, 1990; Chapter 7.
(127) Clementi, E. *IBM J. Res. Dev.* **1965**, *9*, 2.
(128) Roos, B.; Salez, C.; Veillard, A.; Clementi, E., Atomic Program. IBM Research Report RJ1518; 1968.
(129) Clementi, E.; Chakravorty, S. *J. Chem. Phys.* **1990**, *93*, 2591.
(130) van Duijneveldt, F. B. IBM Tech. Rep. RJ 945; 1971.
(131) Curl, R. F.; Smalley, R. E. *Science* **1988**, *242*, 1017.
(132) Becke, A. D. *J. Chem. Phys.* **1988**, *88*, 1053.
(133) Cox, D. M.; Trevor, D. J.; Reichmann, K. C.; Kaldor, A. *J. Am. Chem. Soc.* **1986**, *108*, 2457.
(134) (a) Niesar, U.; Corongiu, G.; Huang, M.-J.; Dupuis, M.; Clementi, E. *Int. J. Quantum Chem. Symp.* **1989**, *23*, 421. (b) Niesar, U.; Corongiu, G.; Huang, M.-J.; Dupuis, M.; Clementi, E. IBM Technical Report KGN-191; IBM: Kingston, 1989.
(135) Niesar, U.; Corongiu, G.; Clementi, E.; Kneller, G. R.; Bhattacharya, D. K. *J. Phys. Chem.* **1990**, *94*, 7949.
(136) Lie, G. C.; Clementi, E. *Phys. Rev. A* **1986**, *33*, 2679.
(137) Clementi, E.; Corongiu, G. *Int. J. Quantum Chem. Symp.* **1983**, *10*, 31.
(138) Wojcik, M.; Clementi, E. *J. Chem. Phys.* **1986**, *84*, 5970.
(139) Wojcik, M.; Clementi, E. *J. Chem. Phys.* **1986**, *85*, 3544.
(140) Wojcik, M.; Clementi, E. *J. Chem. Phys.* **1986**, *85*, 6085.
(141) Detrich, J. H.; Corongiu, G.; Clementi, E. *Chem. Phys. Lett.* **1984**, *112*, 426.
(142) Hirschfelder, J. O.; Curtiss, C. F.; Bird, R. B. *Molecular Theory of Gases and Liquids*; John Wiley: New York, 1954.
(143) Böttcher, C. F. J. *Theory of Electric Polarization*; Elsevier: Amsterdam, 1973; Vol. 1.
(144) Bartlett, R. J.; Shavitt, I.; Purvis, G. D. *J. Chem. Phys.* **1979**, *71*, 281.
(145) Benedict, W. S.; Gailar, N.; Plyler, E. K. *J. Chem. Phys.* **1956**, *24*, 1139.
(146) Corongiu, G. *Int. J. Quantum Chem.*, submitted for publication.
(147) Dosey, N. E. *Properties of Ordinary Water Substances*; Reinhold: New York, 1940.
(148) Thiessen, W. E.; Narten, A. H. *J. Chem. Phys.* **1982**, *77*, 2656.
(149) Naten, A. H.; Levy, H. A. *J. Chem. Phys.* **1971**, *55*, 2263.

(150) Soper, A. K.; Silver, R. N. *Phys. Rev. Lett.* **1982**, *49*, 471. Soper, A. K.; Phillips, M. G. *Chem. Phys.* **1986**, *107*, 47.
(151) Dore, J. D. *Faraday Discuss. Chem. Soc.* **1978**, *66*, 82.
(152) Chen, S. H.; Texeira, J. *Adv. Chem. Phys.* **1988**, *64*, 1.
(153) Cooley, J. W.; Lewis, P. A. W.; Welch, P. D. *IEEE Trans.* **1965**, *E-12*, 1.
(154) Eisenberg, D.; Kauzmann, W. *The structure and properties of water*; Clarendon: Oxford, 1969.
(155) Herzberg, G. *Molecular Spectra and Molecular Structure, II. Infrared and Raman Spectra of Polyatomic Molecules*; Van Nostrand: Princeton, NJ, 1945.
(156) Walrafen, G. E. *Water: A Comprehensive Treatise*; Franks, F., Ed.; Plenum: New York, 1971, Vol. 1, p 151.
(157) Tatcliff, C. I.; Irish, D. E. *J. Chem. Phys.* **1982**, *86*, 4897.
(158) Toukan, K.; Ricci, M. A.; Chen, S. H.; Loong, C. K.; Pierce, D. L.; Texeira, J. *Phys. Rev.* **1988**, *37*, A2580.
(159) Brooks, C. L.; Karplus, M.; Montgomery Pettit, B. *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics*; John Wiley: New York, 1988.
(160) Clementi, E., Chin, S., Eds. *Structure and Dynamics of Nucleic Acids, Proteins and Membrane*; Plenum: New York, 1986.
(161) Clementi, E.; Corongiu, G.; Aida, M.; Niesar, U.; Kneller, G. R. In *Modern Techniques of Computational Chemistry: MOTECC-90*; ESCOM: Leiden, 1990; Chapter 17.
(162) Clementi, E.; Chin, S. *Biological and Artificial Intelligence Systems*; ESCOM, Leiden, 1988.
(163) Beveridge, D. L., Jorgensen, W. L., Eds. Computer Simulation of Chemical and Biomolecular Systems. *Ann. N.Y. Acad. Sci.* 1986.
(164) Aida, M.; Corongiu, G.; Clementi, E. Submitted for publication.
(165) Levy, R. M.; Sheridan, R.; Keepers, J. W.; Dubey, J. S.; Swaminathan, S.; Karplus, M. *Biophys. J.* **1985**, *48*.
(166) Levitt, M.; Sharon, R. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 7557.
(167) Procacci, P.; Corongiu, G.; Clementi, E. IBM Technical Report KGN-195, 1989.
(168) Bhattacharya, D.; Xue, W.; Clementi, E. Submitted for publication.
(169) Ermak, D. L.; McCammon, J. A. *J. Chem. Phys.* **1978**, *69*, 1352.
(170) Alison, S.; Austin, R.; Hogan, M. *J. Chem. Phys.* **1989**, *90*.
(171) Sharp, K.; Fine, R.; Honig, B. *Science* **1987**, *236*, 1460.
(172) Ciccotti, G. C.; Guardo, E.; Sese, G. *Mol. Phys.* **1982**, *46*, 875.
(173) Allen, M. P. *Mol. Phys.* **1982**, *47*, 599.
(174) van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Simulations* **1988**, *1*, 173.
(175) van Gunsteren, W. F.; Karplus, M. *Biochemistry* **1982**, *21*, 2259.
(176) Smith, J.; Cusack, S.; Pezzeca, U.; Brooks, B.; Karplus, M. *J. Chem. Phys.* **1986**, *88*, 3636.
(177) Cusak, S.; Smith, J.; Tibor, B.; Finney, J. L.; Karplus, M. *J. Mol. Biol.* **1988**, *202*, 903.
(178) Cusak, S.; Smith, J.; Finney, J. L.; Karplus, M.; Treewhella, J. *Physica B* **1986**, *136*, 256.
(179) Smith, J.; Kuczera, K.; Tibor, B.; Doster, W.; Cusak, C.; Karplus, M. *Physica B* **1989**, *156*, 437.
(180) Smith, J.; Cusack, S.; Poole, P.; Finney, J. *J. Biomol. Struct. Dyn.* **1987**, *4*, 583.
(181) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Pweiner, J. *Am. Chem. Soc.* **1984**, *106*, 765.
(182) (a) Clementi, E. *J. Mol. Spec.* **1964**, *12*, 18. (b) Hartmann, H.; Clementi, E. *Phys. Rev.* **1964**, *133A*, 1295.
(183) Mohanty, A.; Clementi, E.; Parpia, F. To be published.
(184) Mohanty, A.; Clementi, E. *J. Chem. Phys.* **1990**, *93*, 1829, and references cited therein.
(185) Roos, B.; Siegbahn, P. E. In *Methods of Electronic Structure Theory*; Schaefer, H. F., III, Miller, W. H., Eds.; Plenum: New York, 1977. Shavitt, I. *Methods of Electronic Structure Theory*; Schaefer, H. F., III, Miller, W. H., Eds.; Plenum: New York, 1977. Davidson, E. *J. Comput. Phys.* **1975**, *17*, 87.
(186) Bolis, G.; Ragazzi, M.; Salvaderi, D.; Ferro, D. R.; Clementi, E. *Int. J. Quantum Chem.* **1978**, *14*, 815.