the
abdus salam
international centre for theoretical physics

SMR: 1098/15

WORKSHOP ON THE STRUCTURE OF BIOLOGICAL MACROMOLECULES

(16 - 27 March 1998)

"ab Initio Molecular Dynamics of Biological Systems"

presented by:

Paolo CARLONI
International School for Advanced Studies - SISSA
Via Beirut 2 - 4
34123 Trieste
Italy

Unified Approach for Molecular Dynamics and Density-Functional Theory

R. Car
International School for Advanced Studies, Trieste, Italy

and

M. Parrinello
Dipartimento di Fisica Teorica, Università di Trieste, Trieste, Italy, and International School for Advanced Studies, Trieste, Italy

(Received 5 August 1985)

We present a unified scheme that, by combining molecular dynamics and density-functional theory, profoundly extends the range of both concepts. Our approach extends molecular dynamics beyond the usual pair-potential approximation, thereby making possible the simulation of both covalently bonded and metallic systems. In addition it permits the application of density-functional theory to much larger systems than previously feasible. The new technique is demonstrated by the calculation of some static and dynamic properties of crystalline silicon within a self-consistent pseudopotential framework.

PACS numbers: 71.10.+x, 65.50.+m, 71.45.Gm

Electronic structure calculations based on density-functional (DF) theory[1] and finite-temperature computer simulations based on molecular dynamics[2] (MD) have greatly contributed to our understanding of condensed-matter systems. MD calculations are able to predict equilibrium and nonequilibrium properties of condensed systems. However, in all practical applications MD calculations have used empirical interatomic potentials. This approach, while appropriate for systems like the rare gases, may fail for covalent and/or metallic systems. Furthermore, these calculations convey no information about electronic properties. On the other hand, DF calculations have provided an accurate, albeit approximate, description of the chemical bond in a large variety of systems,[1] but are computationally very demanding. This has so far precluded the application of DF schemes to the study of very large and/or disordered systems and to the computation of interatomic forces for MD simulations.

We wish to present here a new method that is able to overcome the above difficulties and to achieve the following results: (i) compute ground-state electronic properties of large and/or disordered systems at the level of state-of-the-art electronic structure calculations; (ii) perform *ab initio* MD simulations where the only assumptions are the validity of classical mechanics to describe ionic motion and the Born-Oppenheimer (BO) approximation to separate nuclear and electronic coordinates.

Following Kohn and Sham[3] (KS) we write the electron density in terms of occupied single-particle orthonormal orbitals: $n(\mathbf{r})=\sum_i|\psi_i(\mathbf{r})|^2$. A point of the BO potential energy surface is given by the minimum with respect to the $\psi_i(\mathbf{r})$ of the energy functional,

$$E[\{\psi_i\},\{R_I\},\{\alpha_\nu\}]=\sum_i\int_\Omega d^3r\,\psi_i^*(\mathbf{r})[-(\hbar^2/2m)\nabla^2]\psi_i(\mathbf{r})+U[n(\mathbf{r}),\{R_I\},\{\alpha_\nu\}]. \quad (1)$$

Here $\{R_I\}$ indicate the nuclear coordinates and $\{\alpha_\nu\}$ are all the possible external constraints imposed on the system, like the volume Ω, the strain $\epsilon_{\mu\nu}$, etc. The functional U contains the internuclear Coulomb repulsion and the effective electronic potential energy, including external nuclear, Hartree, and exchange and correlation contributions.

In the conventional formulation, minimization of the energy functional [Eq. (1)] with respect to the orbitals ψ_i, subject to the orthonormality constraint, leads to the self-consistent KS equations, i.e.,

$$\left\{-\frac{\hbar^2}{2m}\nabla^2+\frac{\delta U}{\delta n(\mathbf{r})}\right\}\psi_i(\mathbf{r})=\epsilon_i\psi_i(\mathbf{r}). \quad (2)$$

The solution of Eq. (2) involves repeated matrix diagonalizations with a computational effort rapidly growing with the size of the problem. Since the whole procedure has to be repeated for any new atomic configuration, the theoretical prediction of the equilibrium geometries, when these are not known from experiment, still remains an unsolved problem in most cases.

We adopt a quite different approach and regard the minimization of the KS functional as a complex optimization problem which can be solved by applying the concept of simulated annealing, recently introduced by Kirkpatrick, Gelatt, and Vecchi.[4] In this approach an objective function $O(\{\beta\})$ is minimized relative to the parameters $\{\beta\}$, by generation of a succession of $\{\beta\}$'s with a Boltzman-type probability distribution $\propto\exp(-O(\{\beta\})/T)$ via a Monte Carlo procedure. For $T\rightarrow 0$ the state of lowest $O(\{\beta\})$ is reached un-

less the system is trapped into some metastable state.

In our case the objective function is the total-energy functional and the variational parameters are the coefficients of the expansion of the KS orbitals in some convenient basis and possibly the ionic positions and/or the $\{\alpha_\nu\}$'s. We found that a simulated annealing strategy based on MD, rather than on the Metropolis Monte Carlo method of Kirkpatrick, Gelatt, and Vecchi,[4] can be applied efficiently to minimize the KS functional. This approach, which may be called "dynamical simulated annealing," not only is useful as a minimization procedure but, as we demonstrate here, it allows also the study of finite temperature properties.

In our method we consider the parameters $\{\psi_i\}$, $\{R_I\}$, $\{\alpha_\nu\}$ in the energy-functional [Eq. (1)] to be dependent on time and introduce the Lagrangean

$$L = \sum_i \tfrac{1}{2}\mu \int_\Omega d^3r\, |\dot{\psi}_i|^2 + \sum_I \tfrac{1}{2} M_I \dot{R}_I^2 + \sum_\nu \tfrac{1}{2}\mu_\nu \dot{\alpha}_\nu^2 - E[\{\psi_i\}, \{R_I\}, \{\alpha_\nu\}], \tag{3}$$

where the ψ_i are subject to the holonomic constraints

$$\int_\Omega d^3r\, \psi_i^*(\mathbf{r},t)\psi_j(\mathbf{r},t) = \delta_{ij}. \tag{4}$$

In Eq. (3) the dot indicates time derivative, M_I are the physical ionic masses, and μ and μ_ν are arbitrary parameters of appropriate units.

The Lagrangean in Eq. (3) generates a dynamics for the parameters $\{\psi_i\}$'s, $\{R_I\}$'s, and $\{\alpha_\nu\}$'s through the equations of motion:

$$\mu\ddot{\psi}_i(\mathbf{r},t) = -\delta E/\delta\psi_i^*(\mathbf{r},t) + \sum_k \Lambda_{ik}\psi_k(\mathbf{r},t), \tag{5a}$$

$$M_I\ddot{R}_I = -\nabla_{R_I} E, \tag{5b}$$

$$\mu_\nu\ddot{\alpha}_\nu = -(\partial E/\partial\alpha_\nu), \tag{5c}$$

where Λ_{ik} are Lagrange multipliers introduced in order to satisfy the constraints in Eq. (4). The ion dynamics in Eqs. (5) may have a real physical meaning, whereas the dynamics associated with the $\{\psi_i\}$'s and the $\{\alpha_\nu\}$'s is fictitious and has to be considered only as a tool to perform the dynamical simulated annealing. Equation (3) defines a potential energy E and a classical kinetic energy K given by

$$K = \sum_i \tfrac{1}{2}\mu \int_\Omega d^3r\, |\dot{\psi}_i|^2 + \sum_I \tfrac{1}{2} M_I \dot{R}_I^2 + \sum_\nu \tfrac{1}{2}\mu_\nu \dot{\alpha}_\nu^2. \tag{6}$$

The equilibrium value $\langle K\rangle$ of the classical kinetic energy can be calculated as the temporal average over the trajectories generated by the equations of motion [Eqs. (5)] and related to the temperature of the system by suitable normalization. By variation of the velocities, i.e., the $\{\dot{\psi}_i\}$'s, $\{\dot{R}_I\}$'s, and $\{\dot{\alpha}_\nu\}$'s, the temperature of the system can be slowly reduced and for $T \to 0$ the equilibrium state of minimal E is reached. At equilibrium $\ddot{\psi}_i = 0$, Eq. (5a) is identical within a unitary transformation to the KS equation [Eq. (2)], and the eigenvalues of the Λ matrix coincide with the occupied KS eigenvalues. Only when these conditions are satisfied does the Lagrangean in Eq. (3) describe a real physical system whose representative point in configurational space lies on the BO surface. For large systems our scheme is more efficient than standard diagonalization techniques.[5] Furthermore, in the present approach, diagonalization, self-consistency, ionic relaxation, and volume and strain relaxation are achieved *simultaneously*. The amount of classical kinetic energy is a measure of the departure of a system from the self-consistent minimum of its total energy.

It should be stressed that the dynamical simulated annealing technique introduced above is a method of quite general applicability in the context of functional minimization. As such it can be useful in many areas of physics. For instance, it can be applied to the study of classical field theories or to obtain the ground-state energy in Hartree-Fock or configuration interaction schemes. We also observe that, as far as functional minimization is concerned, Newtonian dynamics may be conveniently replaced by Langevin[6] or other types of dynamics.[7]

In order to illustrate how our method works in practice, we present results obtained for the ground-state electronic structure of Si as follows. We have considered a simple cubic supercell containing eight atoms subject to periodic boundary conditions. We have used a local pseudopotential[8] and a local-density approximation to the exact exchange and correlation functional.[9] The single-particle orbitals for the valence electrons have been expanded in plane waves with an energy cutoff of 8 Ry, which amounts to including 437 plane waves at the Γ point. For simplicity, only the Γ point of the Brillouin zone (BZ) of the supercell has been considered in the evaluation of the energy functional.[10] This leads to a total of 16×437 complex electronic variational parameters, since sixteen is the number of doubly occupied KS levels. A simulated annealing run is illustrated in Fig. 1. The lattice parameter was allowed to vary while the ions were kept in their perfect diamond arrangement. The total energy, the lattice parameter, and the eigenvalues of the matrix of the Lagrangean multipliers are plotted as functions of the simulation "time." The initial conditions for the electronic orbitals were fixed by filling the lowest available plane-wave states and giving a Maxwellian distribution of velocities to the components of the fields. The value of μ was chosen to be 1 a.u. The mass μ_Ω associated with variation in the volume was taken to be 10^{-5} a.u. The Verlet algo-

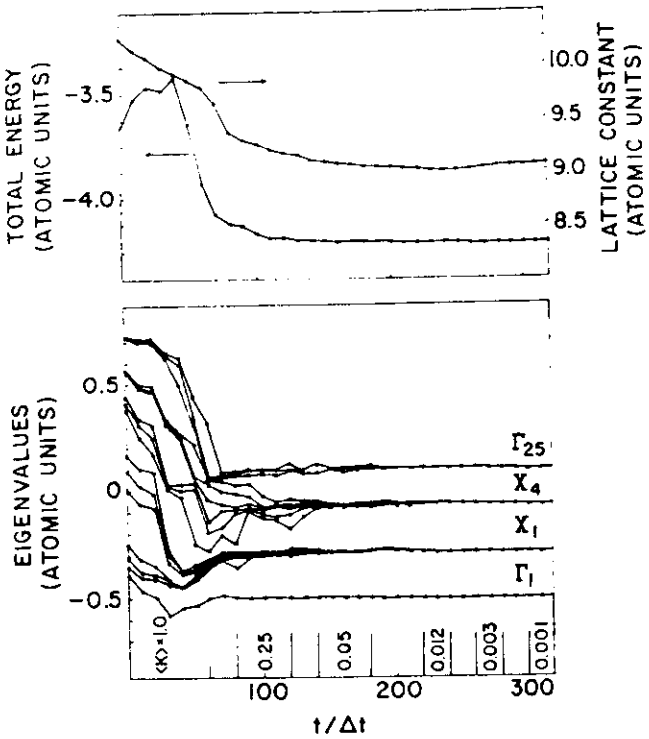

FIG. 1. Evolution of total energy per atom, lattice constant, and eigenvalues of the Λ matrix, during a typical dynamical annealing run. The partial averages of the classical kinetic energy K during each subsection of the run are indicated in the lower part of the picture. For $K \to 0$ the eigenvalues of the Λ matrix tend to the KS eigenvalues. The various multiplets are labeled according to the symmetry of the diamond lattice.

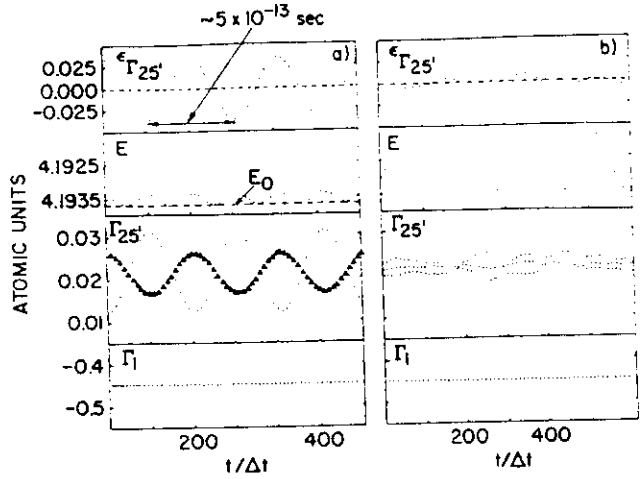

FIG. 2. From top to bottom, temporal evolution of average atomic displacement along $\epsilon_{\Gamma_{25'}}$, potential energy per atom, and $\Gamma_{25'}$ and Γ_1 multiplets for two different MD runs. The lattice constant was taken to be equal to the experimental value of 10.26 a.u.; Δt and μ were taken to be 10 and 300 a.u., respectively. The dashed line in the second panel from the top indicates the $T=0$ ground-state energy. The triangles indicate a doubly degenerate level.

rithm[11] with a time step of 0.1 a.u. was used and the values of the Lagrange multipliers were determined by the method of Ryckaert, Ciccotti, and Berendsen.[12] After some initial equilibration the temperature was progressively reduced to very small values. A satisfactory degree of convergence is seen to be achieved after ~ 200 time steps, when our results agree within numerical errors with those of a conventional self-consistent calculation for the same model.[13]

We can consider now a situation in which the ions, to which we associate their actual physical masses, are allowed to move at a given temperature, while the kinetic energy of the electronic variational parameters remains equal to zero. In this case the electrons are at any time in their ground state and the ions move under the action of BO forces. This can be achieved either by conveniently reoptimizing the electronic variational parameters or by realizing a metastable situation in which the kinetic energy associated with the ψ_i's remains always very small compared to the typical variations of the potential energy of the system. This is equivalent to giving the BO surface a finite thickness proportional to the temperature associated with the ψ_i's. If this temperature remains very small, the ion dynamics generated via Eqs. (5) provides a good representation of the actual dynamics of a physical system.

In Fig. 2 we report the results of two different sets of calculations in which we have performed dynamical simulations for our model. In these calculations it was not necessary to reoptimize the electronic variational parameters at each point along the trajectory, since the thickness of the BO surface never exceeded the value of 7×10^{-6} a.u. per atom, a rather small fraction of the potential energy variation. This is perfectly adequate to represent the interionic forces in this particular case. If the thickness of the BO surface were too large, not only would the forces be incorrectly estimated but also they might depend upon the path along which a given point of the potential energy surface is approached. On the other hand, for very small thicknesses the low velocities of the electronic variational parameters might lead to intolerably long relaxation times. In such a case a compromise would be necessary. In Fig. 2(a) the atoms were initially displaced from their perfect lattice position according to the eigenmode $\epsilon_{\Gamma_{25'}}$, corresponding to the optical phonon mode at the Γ point of the diamond lattice. The system undergoes slightly anharmonic oscillations whose frequency is 20 THz, in perfect agreement with the results of a static frozen-phonon calculation for the same model,[13] showing that the thickness of the BO surface was adequate. In Fig. 2 (a) we also report how the effect of the ionic oscillatory motion is reflected in some electronic properties. Notice that the threefold degenerate

topmost $\Gamma_{25'}$ level splits, in perfect phase with the ionic motion, into a singlet and a doublet, whereas the low-lying Γ_1 state is hardly affected by the ionic motion. These results are contrasted with those reported in Fig. 2(b). Here the ions were first randomly displaced from their equilibrium position and a simulated annealing was performed in order to bring the electrons in the corresponding ground state. The ions were then allowed to move. After some equilibration the average kinetic energy associated with ionic motion had a value corresponding to ~ 250 K and the behavior of the system was as illustrated in Fig. 2(b). The projection of the ionic displacement along the $\epsilon_{\Gamma_{25'}}$ eigenvector and the electronic properties do not show any apparent correlation. The degeneracy of the $\Gamma_{25'}$ one-electron eigenstate is completely lifted by thermal disorder, while the Γ_1 state still remains hardly affected by the ionic motion.

The calculations presented here can all be performed on a VAX-like minicomputer. Access to supercomputers can make possible the simulation of larger systems and more realistic models. Because of the simplicity of Newton's equations the computer code can be fully vectorized with not much effort. However the main advantages of the present approach lie in its ability to perform a global minimization of the energy DF and, more importantly, in offering a convenient and, in principle, exact tool for studying finite temperature effects and dynamical properties.

We benefited from discussions with A. Baldereschi, P. Carnevali, A. Nobile, S. T. Pantelides, A. Selloni, E. Tosatti, and A. R. Williams. Special thanks are due to S. Baroni for precious suggestions and valuable help. This work has been supported by the Gruppo Nazionale di Struttura della Materia del Consiglio Nazionale delle Richerche and by the Ministero della Pubblica Istruzione.

[1]See, for instance, *Theory of the Inhomogeneous Electron Gas*, edited by S. Lundqvist and N. M. March (Plenum, New York, 1983).

[2]A. Rahman, in *Correlation Functions and Quasiparticle Interactions in Condensed Matter*, edited by J. Woods Hailey, NATO Advanced Study Series Vol. 35 (Plenum, New York, 1977).

[3]W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[4]S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, Science **220**, 671 (1983).

[5]For instance, if the N occupied single-particle orbitals are expanded into M plane waves ($M >> N$), a standard diagonalization requires $O(M^3)$ operations, whereas Eq. (5a) requires both $O(NM \ln M)$ and $O(N^2M)$ operations.

[6]P. Carnevali and A. Selloni, private communication.

[7]P. J. Rossky, J. D. Doll, and H. L. Friedman, J. Chem. Phys. **69**, 4628 (1978); C. Pangali, M. Rao, and B. J. Berne, Chem. Phys. Lett. **55**, 413 (1978).

[8]J. A. Appelbaum and D. R. Hamann, Phys. Rev. B **8**, 1777 (1973).

[9]J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[10]The sampling of the BZ and the pseudopotential used in this paper are not very accurate and should be replaced with better ones in more realistic calculations.

[11]L. Verlet, Phys. Rev. **159**, 98 (1967).

[12]J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, J. Comput. Phys. **23**, 327 (1977).

[13]When accurate pseudopotentials are used together with an accurate sampling of the BZ, local-density calculations agree very well with experiment [see M. T. Yin and M. L. Cohen, Phys. Rev. B **26**, 3259 (1982)].

Inhomogeneous Electron Gas*

P. HOHENBERG†
École Normale Superieure, Paris, France
AND
W. KOHN‡
École Normale Superieure, Paris, France and Faculté des Sciences, Orsay, France
and
University of California at San Diego, La Jolla, California

(Received 18 June 1964)

This paper deals with the ground state of an interacting electron gas in an external potential $v(\mathbf{r})$. It is proved that there exists a universal functional of the density, $F[n(\mathbf{r})]$, independent of $v(\mathbf{r})$, such that the expression $E\equiv\int v(\mathbf{r})n(\mathbf{r})d\mathbf{r}+F[n(\mathbf{r})]$ has as its minimum value the correct ground-state energy associated with $v(\mathbf{r})$. The functional $F[n(\mathbf{r})]$ is then discussed for two situations: (1) $n(\mathbf{r})=n_0+\tilde{n}(\mathbf{r})$, $\tilde{n}/n_0<<1$, and (2) $n(\mathbf{r})=\varphi(\mathbf{r}/r_0)$ with φ arbitrary and $r_0\to\infty$. In both cases F can be expressed entirely in terms of the correlation energy and linear and higher order electronic polarizabilities of a uniform electron gas. This approach also sheds some light on generalized Thomas-Fermi methods and their limitations. Some new extensions of these methods are presented.

INTRODUCTION

DURING the last decade there has been considerable progress in understanding the properties of a homogeneous interacting electron gas.[1] The point of view has been, in general, to regard the electrons as similar to a collection of noninteracting particles with the important additional concept of collective excitations.

On the other hand, there has been in existence since the 1920's a different approach, represented by the Thomas-Fermi method[2] and its refinements, in which the electronic density $n(\mathbf{r})$ plays a central role and in which the system of electrons is pictured more like a classical liquid. This approach has been useful, up to now, for simple though crude descriptions of inhomogeneous systems like atoms and impurities in metals.

Lately there have been also some important advances along this second line of approach, such as the work of Kompaneets and Pavlovskii,[3] Kirzhnits,[4] Lewis,[5] Baraff and Borowitz,[6] Baraff,[7] and DuBois and Kivelson.[8] The present paper represents a contribution in the same area.

In Part I, we develop an exact formal variational principle for the ground-state energy, in which the density $n(\mathbf{r})$ is the variable function. Into this principle enters a universal functional $F[n(\mathbf{r})]$, which applies to all electronic systems in their ground state no matter what the external potential is. The main objective of theoretical considerations is a description of this functional. Once known, it is relatively easy to determine the ground-state energy in a given external potential.

In Part II, we obtain an expression for $F[n]$ when n deviates only slightly from uniformity, i.e., $n(\mathbf{r})=n_0+\tilde{n}(\mathbf{r})$, with $\tilde{n}/n_0\to 0$. In this case $F[n]$ is entirely expressible in terms of the exact ground-state energy and the exact electronic polarizability $\alpha(q)$ of a *uniform* electron gas. This procedure will describe correctly the long-range Friedel charge oscillations[9] set up by a localized perturbation. All previous refinements of the Thomas-Fermi method have failed to include these.

In Part III we consider the case of a slowly varying, but *not* necessarily almost constant density, $n(\mathbf{r})=\varphi(\mathbf{r}/r_0)$, $r_0\to\infty$. For this case we derive an expansion of $F[n]$ in successive orders of r_0^{-1} or, equivalently of the gradient operator ∇ acting on $n(\mathbf{r})$. The expansion coefficients are again expressible in terms of the exact ground-state energy and the exact linear, quadratic, etc., electric response functions of a uniform electron gas to an external potential $v(\mathbf{r})$. In this way we recover, quite simply, all previously developed refinements of the Thomas-Fermi method and are able to carry them somewhat further. Comparison of this case with the nearly uniform one, discussed in Part II, also reveals why the gradient expansion is intrinsically incapable of properly describing the Friedel oscillations or the radial oscillations of the electronic density in an atom which reflect the electronic shell structure. A partial summation of the gradient expansion can be carried out (Sec. III.4), but its usefulness has not yet been tested.

I. EXACT GENERAL FORMULATION

1. The Density as Basic Variable

We shall be considering a collection of an arbitrary number of electrons, enclosed in a large box and moving

* Supported in part by the U. S. Office of Naval Research.
† NATO Post Doctoral Fellow.
‡ Guggenheim Fellow.
[1] For a review see, for example, D. Pines, *Elementary Excitations in Solids* (W. A. Benjamin Inc., New York, 1963).
[2] For a review of work up to 1956, see N. H. March, Advan. Phys. 6, 1 (1957).
[3] A. S. Kompaneets and E. S. Pavlovskii, Zh. Eksperim. i. Teor. Fiz. 31, 427 (1956) [English transl.: Soviet Phys.—JETP 4, 328 (1957)].
[4] D. A. Kirzhnits, Zh. Eksperim. i. Teor. Fiz. 32, 115 (1957) [English transl.: Soviet Phys.—JETP 5, 64 (1957)].
[5] H. W. Lewis, Phys. Rev. 111, 1554 (1958).
[6] G. A. Baraff and S. Borowitz, Phys. Rev. 121, 1704 (1961).
[7] G. A. Baraff, Phys. Rev. 123, 2087 (1961).
[8] D. F. Du Bois and M. G. Kivelson, Phys. Rev. 127, 1182 (1962).
[9] J. Friedel, Phil. Mag. 43, 153 (1952).

under the influence of an external potential $v(\mathbf{r})$ and the mutual Coulomb repulsion. The Hamiltonian has the form

$$H = T + V + U, \tag{1}$$

where[10]

$$T \equiv -\frac{1}{2}\int \nabla\psi^*(\mathbf{r})\nabla\psi(\mathbf{r})d\mathbf{r}, \tag{2}$$

$$V \equiv \int v(\mathbf{r})\psi^*(\mathbf{r})\psi(\mathbf{r})d\mathbf{r}, \tag{3}$$

$$U = \frac{1}{2}\int \frac{1}{|\mathbf{r}-\mathbf{r}'|}\psi^*(\mathbf{r})\psi^*(\mathbf{r}')\psi(\mathbf{r}')\psi(\mathbf{r})d\mathbf{r}d\mathbf{r}'. \tag{4}$$

We shall in all that follows assume for simplicity that we are only dealing with situations in which the ground state is nondegenerate. We denote the electronic density in the ground state Ψ by

$$n(\mathbf{r}) \equiv (\Psi, \psi^*(\mathbf{r})\psi(\mathbf{r})\Psi), \tag{5}$$

which is clearly a functional of $v(\mathbf{r})$.

We shall now show that conversely $v(\mathbf{r})$ is a unique functional of $n(\mathbf{r})$, apart from a trivial additive constant. The proof proceeds by *reductio ad absurdum*. Assume that another potential $v'(\mathbf{r})$, with ground state Ψ' gives rise to the *same* density $n(\mathbf{r})$. Now clearly [unless $v'(\mathbf{r}) - v(\mathbf{r}) = \text{const}$] Ψ' cannot be equal to Ψ since they satisfy different Schrödinger equations. Hence, if we denote the Hamiltonian and ground-state energies associated with Ψ and Ψ' by H, H' and E, E', we have by the minimal property of the ground state,

$$E' = (\Psi', H'\Psi') < (\Psi, H'\Psi) = (\Psi, (H + V' - V)\Psi),$$

so that

$$E' < E + \int [v'(\mathbf{r}) - v(\mathbf{r})]n(\mathbf{r})d\mathbf{r}. \tag{6}$$

Interchanging primed and unprimed quantities, we find in exactly the same way that

$$E < E' + \int [v(\mathbf{r}) - v'(\mathbf{r})]n(\mathbf{r})d\mathbf{r}. \tag{7}$$

Addition of (6) and (7) leads to the inconsistency

$$E + E' < E + E'. \tag{8}$$

Thus $v(\mathbf{r})$ is (to within a constant) a unique functional of $n(\mathbf{r})$; since, in turn, $v(\mathbf{r})$ fixes H we see that the full many-particle ground state is a unique functional of $n(\mathbf{r})$.

2. The Variational Principle

Since Ψ is a functional of $n(\mathbf{r})$, so is evidently the kinetic and interaction energy. We therefore define

$$F[n(\mathbf{r})] \equiv (\Psi, (T+U)\Psi), \tag{9}$$

where $F[n]$ is a universal functional, valid for any number of particles[11] and *any* external potential. This functional plays a central role in the present paper.

With its aid we define, for a given potential $v(\mathbf{r})$, the energy functional

$$E_v[n] \equiv \int v(\mathbf{r})n(\mathbf{r})d\mathbf{r} + F[n]. \tag{10}$$

Clearly, for the correct $n(\mathbf{r})$, $E_v[n]$ equals the ground-state energy E.

We shall now show that $E_v[n]$ assumes its minimum value for the correct $n(\mathbf{r})$, if the admissible functions are restricted by the condition

$$N[n] \equiv \int n(\mathbf{r})d\mathbf{r} = N. \tag{11}$$

It is well known that for a system of N particles, the energy functional of Ψ'

$$\mathcal{E}_v[\Psi'] \equiv (\Psi', V\Psi') + (\Psi', (T+U)\Psi') \tag{12}$$

has a minimum at the correct ground state Ψ, relative to arbitrary variations of Ψ' in which the number of particles is kept constant. In particular, let Ψ' be the ground state associated with a different external potential $v'(\mathbf{r})$. Then, by (12) and (9)

$$\begin{aligned}\mathcal{E}_v[\Psi'] &= \int v(\mathbf{r})n'(\mathbf{r})d\mathbf{r} + F[n'],\\ &> \mathcal{E}_v[\Psi] = \int v(\mathbf{r})n(\mathbf{r})d\mathbf{r} + F[n].\end{aligned} \tag{13}$$

Thus the minimal property of (10) is established relative to all density functions $n'(\mathbf{r})$ associated with some other external potential $v'(\mathbf{r})$.[12]

If $F[n]$ were a known and sufficiently simple functional of n, the problem of determining the ground-state energy and density in a given external potential would be rather easy since it requires merely the minimization of a functional of the three-dimensional density function. The major part of the complexities of the many-electron problems are associated with the determination of the universal functional $F[n]$.

3. Transformation of the Functional $F[n]$

Because of the long range of the Coulomb interaction, it is for most purposes convenient to separate out from

---

[10] Atomic units are used.

[11] This is obvious since the number of particles is itself a simple functional of $n(\mathbf{r})$.

[12] We cannot prove whether an arbitrary positive density distribution $n'(\mathbf{r})$, which satisfies the condition $\int n'(\mathbf{r})d\mathbf{r} = \text{integer}$, can be realized by *some* external potential $v'(\mathbf{r})$. Clearly, to first order in $\bar{n}(\mathbf{r})$, any distribution of the form $n'(\mathbf{r}) = n_0 + \bar{n}(\mathbf{r})$ can be so realized and we believe that in fact all, except some pathological distributions, can be realized.

Nonempirical Calculations of a Hydrated RNA Duplex

Jürg Hutter, Paolo Carloni, and Michele Parrinello

Contribution from the Max-Planck-Institut für Festkörperforschung, D-70569 Stuttgart, Germany, IBM Research Division, Zurich Research Laboratory, Säumerstrasse 4, CH-8803 Rüschlikon, Switzerland, and Department of Chemistry, University of Florence, Via G. Capponi, 7, I-50121 Florence, Italy

JOURNAL OF THE AMERICAN CHEMICAL SOCIETY®

Reprinted from
Volume 118, Number 36, Pages 8710–8712

Nonempirical Calculations of a Hydrated RNA Duplex

Jürg Hutter,[†] Paolo Carloni,[‡] and Michele Parrinello*,[†]

Contribution from the Max-Planck-Institut für Festkörperforschung, D-70569 Stuttgart, Germany, IBM Research Division, Zurich Research Laboratory, Säumerstrasse 4, CH-8803 Rüschlikon, Switzerland, and Department of Chemistry, University of Florence, Via G. Capponi, 7, I-50121 Florence, Italy

Received April 12, 1996. Revised Manuscript Received June 25, 1996[⊗]

Abstract: We have performed density functional theory based ab initio calculations on the crystal structure of sodium guanylyl-3′-5′-cytidine (GpC) nonahydrate. Our calculations are in good agreement with the experimentally determined X-ray structure. This is one of the first attempts to model ab initio nucleic acids in laboratory-realizable conditions. Comparison is also made with empirical force field based structure calculations.

I. Introduction

Molecular dynamics simulations based on effective potentials have been crucial in understanding the properties of a large variety of biological systems.[1-5] This approach derives its strength from its suitability to study very large systems and to follow their evolution on a relatively long time scale. However, this approach is not devoid of problems. For instance, effective potentials present some difficulties in describing the structure of nucleic acids,[6] so much so that very recently an ad hoc reparametrization of the effective potentials to fit nucleic acids properties explicitly has been attempted.[7] However, only very limited results are yet available on the overall performance of these new potentials. Furthermore, it is becoming increasingly clear that there is the need to transcend the effective potential approach if one wants to study biological processes that involve a change of the chemical bond such as enzymatic reactions.[8] These are better and more reliably described by ab initio quantum-chemical approaches.

Owing to the size of the biological molecule, the quantum-chemical calculations have been confined to the study of fragments in vacuum.[9-21] This, however, is far from being actually relevant because water and the environment are known to play a crucial role in determining the structure, dynamics, and function of proteins and nucleic acids.[1,5] Nevertheless, progress in ab initio molecular dynamics combined with the power of parallel computing has dramatically increased the size of systems currently accessible. Keeping future applications to biochemical processes in mind, it is important to investigate the accuracy of ab initio methods to describe biologically relevant processes in as realistic an environment as possible. To this end we have studied the structure of sodium guanylyl-3′-5′-cytidine (GpC) nonahydrate, which has been determined by single-crystal X-ray diffraction.[22,24]

This structure is favorable in many respects: it has a large but manageable number of atoms (368), and yet it contains all the basic ingredients concerning the stability of the nucleic acid helix. It is a small segment of right-handed, antiparallel double-helical RNA, with Watson−Crick base pairing (Figure 1a). Therefore, it contains both the base−base and the base−sugar backbone interactions. Furthermore, it is fully hydrated, thus allowing a study of the hydration process, and it contains the counterions (Figure 1b). The combination of all these elements has never been investigated by fully ab initio methods. This also provides a stringent test of the ability of ab initio methods to describe nucleic acids.

II. Computational Section

Our calculations are performed within the framework of density functional theory and use the generalized gradient approximation. The

[†] Max-Planck-Institut für Festkörperforschung.

[‡] University of Florence. Present address: IBM Research Division, Zurich Research Laboratory.

[⊗] Abstract published in *Advance ACS Abstracts,* August 15, 1996.

(1) McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, 1987.

(2) Brooks, C. L., III; Karplus, M.; Petitt, B. M. In *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics*; Prigodinen, I., Rice, S. Eds.; Wiley Series on Advances in Chemical Physics, Vol. 71; Wiley: New York, 1988.

(3) van Gunsteren, W. F.; Berendsen, H. J. C. *Angew. Chem., Int. Ed. Engl.* **1990**, *298*, 992−1023.

(4) Kollman, P. A.; Merz, K. M., Jr. *Acc. Chem. Res.* **1990**, *23*, 246−252.

(5) Beveridge, D. L.; Swaminathan, S.; Ravishanker, G.; Withka, J. M.; Srinivasan, J.; Prevost, C.; Louise-May, S.; Langley, D. R.; DiCapua, F. M.; Bolton, P. H. In *Water and Biological Macromolecules*; Westhof, E., Ed.; MacMillan: London, 1993; p 165.

(6) McConnell, K. J.; Nirmala, R.; Young, M. A.; Ravishankar, G.; Beveridge, D. L. *J. Am. Chem. Soc.* **1994**, *116*, 4461−4462.

(7) MacKerell, A. D., Jr.; Wiórkiwicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946−11975.

(8) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; J. Wiley and Sons: New York, 1991.

(9) Clementi, E.; Mehl, J.; von Niessen, W. *J. Chem. Phys.* **1971**, *54*, 508.

(10) Del Bene, J. *J. Mol. Struct.: THEOCHEM* **1985**, *124*, 201.

(11) Hobza, P.; Sandorfy, C. *J. Am. Chem. Soc.* **1987**, *109*, 1302.

(12) Aida, M. *J. Comput. Chem.* **1988**, *9*, 362.

(13) Dive, G.; Dehareng, D.; Ghuysen, J. M. *Theor. Chim. Acta* **1993**, *85*, 409.

(14) Hrouda, V.; Florian, J.; Hobza, P. *J. Phys. Chem.* **1993**, *97*, 1542.

(15) Colson, A. O.; Besler, B.; Sevilla, M. D. *J. Phys. Chem.* **1993**, *97*, 13852.

(16) Gould, I. R.; Kollman, P. A. *J. Am. Chem. Soc.* **1994**, *116*, 2493.

(17) Nagata, C.; Aida, M. *J. Mol. Struct.: THEOCHEM* **1988**, *179*, 451.

(18) Aida, M. *J. Theor. Biol.* **1988**, *130*, 327.

(19) Aida, M.; Nagata, C. *Chem. Phys. Lett.* **1982**, *86*, 44.

(20) Hobza, P.; Sponer, J.; Polasek, M. *J. Am. Chem. Soc.* **1995**, *117*, 792−798.

(21) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.

(22) Rosemberg, J. M.; Seeman, N. C.; Day, R. O.; Rich, A. *J. Mol. Biol.* **1976**, *104*, 145−167.

(23) Hydration patterns fairly consistent with experiment have been obtained also by Monte Carlo calculations on the RNA duplex in aqueous solution (Subramanian, P. S.; Pitchumani, S.; Beveridge, D. L.; Berman, H. *Biopolymers* **1990**, *29*, 771−783).

(24) X-ray structures have been determined also for other two salts of GpC: with ammonium (Aggarwal, A.; Islam, A. S.; Kuroda, R.; Sanderson, M. R.; Neidle, S. *Acta Crystallogr.* **1983**, *B39*, 98) and with calcium (Hingerty, B.; Subramanian, E.; Stellman, S. D.; Sato, T.; Broyde, S. B.; Langridge, R. *Acta Crystallogr.* **1976**, *B32*, 2998).

S0002-7863(96)01220-6 CCC: $12.00

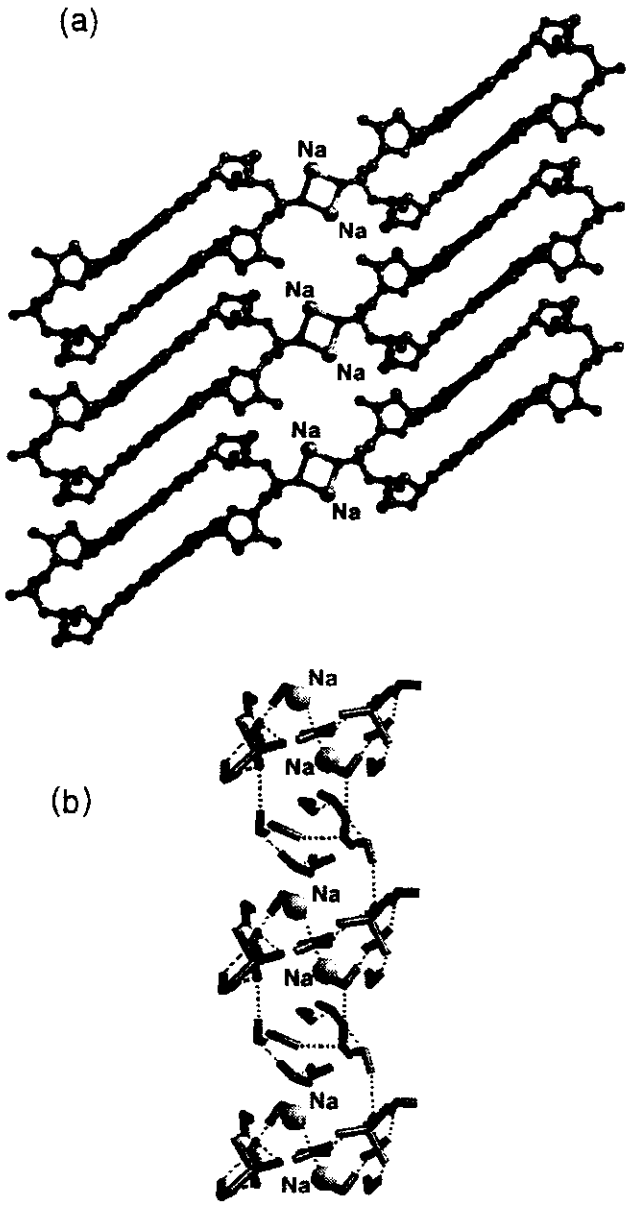

Figure 1. (a) Side view of the GpC crystal structure, which exhibits the stacking of the Watson–Crick base pairs. The water molecules and the hydrogen atoms are not shown. The two RNA fragments are held together by the sodium counterions shown in the picture. The interstice is filled with water. The structure of this water channel is shown in part b.

local functional for correlation in the Perdew–Zunger parameterization was used together with Becke's gradient-corrected exchange functional.[26] We treat explicitly only the 1160 valence electrons. The interaction between valence electrons and ionic cores is described by supersoft pseudopotentials of the Vanderbilt type.[27] The Kohn–Sham orbitals are expanded in plane waves up to an energy cutoff of 24 Ry, resulting in 45477 degrees of freedom per state. This scheme has been tested elsewhere.[30,31] We have used the CPMD[32] code, and optimized the structure using a combination of DIIS for electronic minimization and a Newton–Raphson method for ionic relaxation. This program

(25) Perdew, J. P.; Zunger, A. *Phys. Rev. B* **1981**, *23*, 5048.
(26) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.
(27) Vanderbilt, D. *Phys. Rev. B* **1990**, *41*, 7892.
(28) Sponer, J.; Florian, J.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **1996**, *13*, 827.
(29) Florian, J.; Baumruk, V.; Strajbl, M.; Bednárová, S. J. *J. Phys. Chem.* **1996**, *100*, 1559.
(30) Laasonen, K.; Sprik, M.; Parrinello, M.; Car, R. *J. Chem. Phys.* **1993**, *99*, 9081.
(31) Tuckerman, M.; Laasonen, K.; Sprik, M.; Parrinello, M. *J. Chem. Phys.* **1995**, *103*, 150–161.

Table 1. Root Mean Square Deviations (Å) (Heavy Atoms) with Respect of the X-ray Structure for the Elements Contained in the Unit Cell of GpC, Namely the Two RNA Duplexes, the 16 Water Oxygens, and the 4 Sodium Ions

	RNA	water	sodium
ab initio calculations	0.40	0.31	0.37
force field calculations	1.01	0.93	0.87

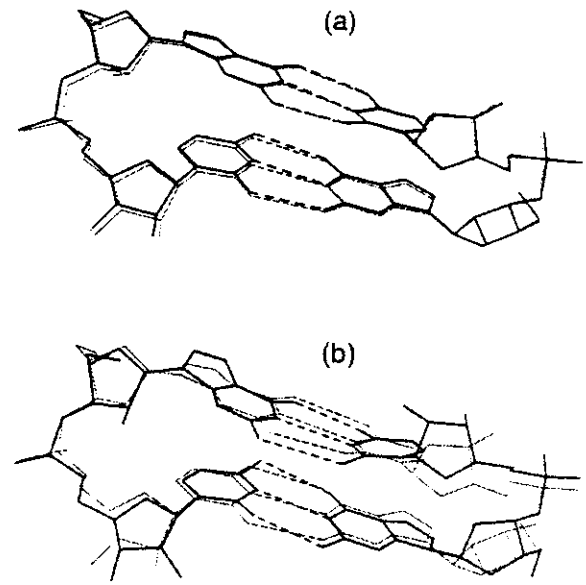

Figure 2. Comparison of the X-ray structure of the GpC duplex (thin lines) with (a) the ab initio structure (thick lines) and (b) the force field structure (thick lines).

uses periodic boundary conditions; the Coulomb interactions are evaluated by the Ewald sum method. No symmetry restriction was imposed on the calculation. We stopped the relaxation when the root-mean-square value of the force was less than 10^{-3} au. We estimate that the resulting uncertainty in the position is less than the experimental error.

The crystal structure of sodium GpC nonahydrate, as determined in ref 22, contains four molecules per monoclinic unit cell. The space group is $C2$, with cell dimensions $a = 21.460$ Å, $b = 16.927$ Å, $c = 9.332$ Å, and $\beta = 90.54°$. The cell parameters were not optimized, and β was set to 90° for computational convenience. The coordinates were taken from the Cambridge Data Base.[33] The hydrogen atom positions, not resolved in the X-ray structure, were given in an arbitrary way, respecting only the constraint of standard bond angles and bond lengths. This yielded a rather unlikely initial configuration in which the water dipoles were pointing in the same direction. This has been meant to be a test of the ability of our scheme to generate spontaneously a hydrogen bond network (Figure 1b). As we shall demonstrate below, this test has been successful. In this respect the ab initio approach appears to be more robust than the effective potential, which at times has difficulties dealing with high-energy starting configurations.[34]

III. Results

The root-mean-square deviations from experiment are shown in Table 1 for the RNA moiety. The water oxygens and the sodium counterions are compared with the results of a standard force field model.[35]

(32) CPMD version 2.5, written by Jürg Hutter, Max-Planck-Institut für Festkörperforschung, Stuttgart, 1995, with the help of the group for numerically intensive computations of IBM Research Laboratory Zurich and the Abteilung Parrinello of MPI Stuttgart.
(33) Allen, F. H.; Davies, J. E.; Galloy, J. J.; Johnson, O.; Kennard, O.; Macrae, C. F.; Mitchell, E. M.; Mitchell, G. F.; Smith, J. M.; Watson, D. G. *J. Chem. Inf. Comp. Sci.* **1991**, *31*, 187–204.
(34) The atomic coordinates of our optimized structure are available by anonymous ftp (address: parrix1.mpi-stuttgart.mpg.de, directory pub/outgoing).

Table 2. Backbone Torsion Angles for the Two RNA Duplexes

9	$\chi(5')$	$\xi(5')$	α	β	γ	δ	ϵ	$\chi(3')$	$\xi(3')$
X-ray	13	89	211	292	285	184	50	32	77
ab initio	14(2)	85(1)	210(5)	298(6)	281(4)	193(4)	49(1)	37(4)	81(4)
force field	15(20)	115(4)	213(10)	285(8)	238(17)	184(6)	98(28)	51(13)	105(5)

[a] For the ab initio and force field structures, the average values of the four GpC moieties are reported. Standard deviations are also reported in parentheses. For the definitions of the torsional angles, see ref 36.

We do not compare our results with the very recent parametrization in ref 7 because more extensive tests are necessary to establish its transferability.

We now examine the various structural building blocks of the crystal in more detail. First, and most important, we reproduce with good accuracy the Watson–Crick hydrogen bond distances (see Figure 2) and the conformation of the bases. A measure of the planarity of the nucleobase rings is given by the maximum deviation with respect to the ideal value of 180°. We find that this value is 5° for the ab initio and 10° for the force field-based structure.

These small but significant deviations between ab initio and force field structures are due to the well-known difficulty for force fields to ensure the planarity of the aromatic rings.[1] This is one effect of electronic origin, which in the force field method is mimicked by the so-called "improper" torsional forces.[1]

Another important feature of the conformation of two complementary nucleobases is their noncoplanarity.[36] The value of the "propeller twist" angle[28,36] of the X-ray structure of GpC is 9°. In the case of our optimized structure, we find values between 7° and 9°. For the force field structure, the values are larger, between 10° and 17°.

The other important building block is the sugar–phosphate backbone, whose torsional angles are crucial in determining the RNA structure. As the corresponding torsional barriers are very small (typically of the order of 1 kcal/mol),[29] they are rather difficult to model with effective potentials. The results are shown in Table 2. We note that they are again in very good agreement with experiment and appear to give a better description than do molecular models.

Finally, the hydrogen bond network obtained in their relaxation process is identical to that postulated by the X-ray crystallographers.[22,23] The water molecules form hydrogen bonds with themselves as well as with the RNA moiety. This can be considered a real prediction because, as stressed above, protons are invisible to X-rays.

IN our calculations we find that some of the nucleobase amino groups are distorted: the hydrogens are out of the plane (maximum displacement: 17° for a guanine) to form hydrogen bonds with a neighboring water. Again, this effect is missed by the effective potential.

Our calculations also reproduce well the highly symmetric octahedral solvation structure of the sodium ions, which is more distorted in the effective potential calculations.

IV. Concluding Remarks

In conclusion, we have shown that it is now possible to perform ab initio simulations of molecules whose size approaches that relevant for biologically interesting systems. The quality of the results is very high even in the case of nucleic acids, which have traditionally been very difficult to model.[6] The ab initio modeling automatically includes all the physical and chemical effects that are so difficult to mimic in effective potential simulations, such as polarization effects, many-body forces, and the rigidity of aromatic rings. For instance, in our simulations the electronic structure of water is modified by the local environment, giving different properties to different water molecules.

Another important advantage to the ab initio method is that no painstaking parametrization is needed to extend the domain of applicability of the theory. We know it works for water,[30] water solutions,[31] and peptide bonds.[37] We have shown here its validity for the study of nucleic acids. It is expected that it will work with similar accuracy in a very large variety of biological environments, and have considerable predicting power. In particular, one can reliably model bioinorganic molecules, which contain transition metal ions, such as metalloproteins and metal-based drug–DNA adducts.[38]

Furthermore, and this is at odds with most effective potentials and ab initio codes, the Coulombic forces are calculated using the Ewald summation procedure. We believe that this is very important for periodic systems and that the use of cutoff in the Coulombic interactions introduces spurious effects.[39]

Of course, these kinds of simulations are several orders of magnitude more demanding than those based on effective potentials. However, progress in computer architecture and in the algorithms used gives us confidence that this gap can be narrowed in a short span of time and that a new dimension can be added to the simulation of biochemical processes.

Acknowledgment. We thank Pietro Ballone for many useful discussions. Marta Ferraroni is acknowledged for the use of the XPLOR program at EMBL, Hamburg, Germany. We also thank the Maui High Performance Computing Center (MHPCC) for a generous allocation of computing time on their IBM SP2 computer.

JA9612209

(35) For our classical simulations, we have used the XPLOR 3.1 package running on a Silicon Graphics Iris. The standard united atom CHARMM force field, available in the package, was used for the GpC and the sodium ions.[40] For water, we have used the TIP3P model.[41] The unit cell was duplicated along the c direction, in order to be able to use a larger nonbonding interaction cutoff. A cutoff of 8 Å was used. The crystal packing interactions are calculated in XPLOR with the periodic image convention. The geometry optimizations were performed with the conjugate gradient method up to an energy gradient of less than 1×10^{-3} kcal/mol for all the atoms.

(36) Saenger, W. *Principles of Nucleic Acid Structure*; Springer: New York, 1984.

(37) Parrinello, M.; Carloni, P.; Andreoni, W. Paper in preparation.

(38) Carloni, P.; Andreoni, W. Submitted for publication.

(39) Saito, M. *J. Chem. Phys.* **1994**, *101*, 4055–4061.

(40) Brooks, B.; Bruccoleri, R.; Olafson, B.; States, D.; Swaminthan, S.; Karplus, M. *J. Comp. Chem.* **1983**, *4*, 187–217.

(41) Jorgensen, W.; Chandrasekar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926–935.

Self-Assembled Peptide Nanotubes from First Principles

Paolo Carloni and Wanda Andreoni
IBM Research Division, Zurich Research Laboratory, 8803 Rüschlikon, Switzerland

Michele Parrinello
Max-Planck-Institut für Festkörperforschung, 70569 Stuttgart, Germany
(Received 21 March 1997)

Nanotubular structures made up of polypeptides have recently become available. We present a characterization of the structural and electronic properties of one such system within density functional theory with calculations using gradient-corrected exchange-correlation functionals. Comparison with data on natural amino acids, and with the predictions of empirical models, is drawn from the structural features. A large gap in the low-energy electronic excitation spectrum is predicted, and the presence of extended as well as localized states near the gap is found. [S0031-9007(97)03727-7]

PACS numbers: 81.05.Zx, 81.05.Rm

Hollow tubular structures have a variety of potential applications in biochemistry, chemistry, and materials science, ranging from catalysis to separation technology, from optoelectronics to the construction of drug delivery vehicles [1,2].

A very interesting approach to tailor-made nanostructures is the process of self-assembling that mimics what naturally occurs in living matter and is thus able to provide materials with especially high conformational stability and unique functional properties [2]. Along these lines, well-characterized crystal structures exhibiting parallel open channels have been synthesized. Their building block is a cyclic polypeptide subunit consisting of an even number of alternating *D*- and *L*-amino acid residues, which stack to produce β-sheet-like tubular structures [3] (see Fig. 1). These regular arrangements do not occur in nature and are a consequence of the presence of *D*-amino acids which, as is well known, are not present in biopolymers, except for rare cases such as that of antibiotic gramicidin A.

A set of experimental data that includes high-resolution x-ray crystallography, cryoelectron microscopy, electron diffraction, and Fourier-transform infrared spectroscopy have led, with the aid of molecular modeling, to a qualitative interpretation of the structure. Quantitative structural refinement, however, is necessary to determine the steric and chemical selectivity of these channels [4] and therefore to assess their potential practical use.

Classical molecular dynamics simulations can be and have long been used to refine polypeptide structures [5]. However, for this particular polypeptide the accuracy of such refinement is questionable. In fact, the force fields that are normally used have been parametrized on the naturally occurring *L*-amino acids, whereas *D*-amino acids are also present in this case. Moreover, the complexity of the architecture, for which covalent as well as hydrogen bonds are equally crucial, and the importance of the specific amino acid side chains for self-assembling and functionality, make an *ab initio* approach to this class of materials highly appealing. First-principles geometry optimization has not yet gone beyond isolated monopeptides and dipeptides [6]. The electronic structure of idealized

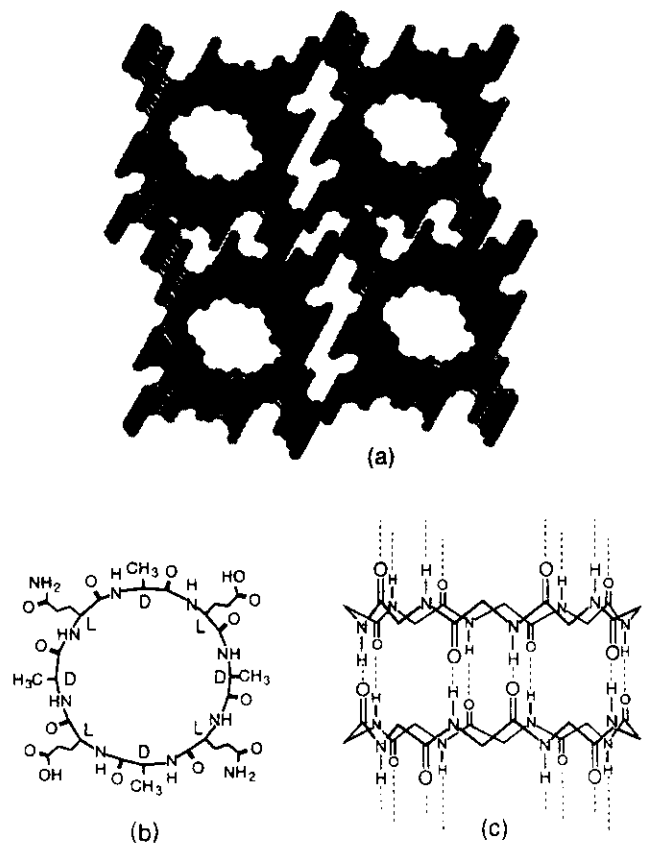

FIG. 1. *Cyclo*[-(*D*-Ala-Glu-*D*-Ala-Gln)$_2$] structure. (a) View of the ordered arrangement of the nanotubes. The internal diameter is 9.5 Å. Hydrogens are omitted. Spheres of increasing darkness represent C, N, and O atoms, respectively. (b) 2D representation of one individual subunit and (c) of the H-bond interacting subunits at an average distance of 4.8 Å, from Ref. [3].

0031-9007/97/79(4)/761(4)$10.00

polypeptide chains has been studied with density functional theory (DFT) based calculations in the local density approximation, using small basis sets for the electron wave functions [7].

We present here a nonempirical determination of one of these organic nanotubular structures, namely, the octapeptide *cyclo*[-(*D*-Ala-Glu-*D*-Ala-Gln)$_2$] [8] (Fig. 1) in the solid phase [3]. Our calculations are based on DFT with a gradient-corrected exchange-correlation functional [9], and use efficient plane-wave-based schemes for the optimization of this relatively complex structure [10]. In particular, the angular-momentum-dependent pseudopotentials were derived using the Troullier-Martins procedure [11], wave-function optimization was performed with the direct inversion in iterative subspace (DIIS) method [12], and geometry optimization with combined DIIS-Newton-Raphson methods.

The accuracy of our approach has already been proved for biologically relevant systems [13]. However, the present work is the first application to a real-life polypeptide.

Comparison with the results of empirical potentials is drawn, and the differences between them and the DFT structure are evaluated in terms of the database of naturally occurring polypeptides. Details of the backbone structural pattern as well as the structural links between side chains are determined. Our results show that, electronically, the system is a large-gap semiconductor with impuritylike states at the band edges.

Octapeptide *cyclo*[-(*D*-Ala-Glu-*D*-Ala-Gln)$_2$] crystallizes in a triclinic lattice ($a = 9.5$ Å, $b = c = 15.1$ Å, $\alpha = 90°$, $\beta = \gamma = 99°$) [3]. The two peptide subunits, which are related by local C_2 symmetry, correspond to a total of 112 heavy atoms per unit cell and 100 hydrogen atoms. The basic characteristics are illustrated in the 2D scheme in Fig. 1: the chemical composition of a subunit [Fig. 1(b)], the antiparallel stacking, and the hydrogen-bond network [Fig. 1(c)].

In our optimization, all the atomic positions are relaxed in the periodically repeated cell with lattice parameters fixed to the experimental values. For the starting configuration, we adopted the model obtained by Ghadiri *et al.* [3,14] and use this as a basis of discussion. The overall initial conformation is preserved, namely, the backbone torsional angles remain close to the original value [15], although acquiring a 10° dispersion. The backbone amine and carbonyl groups remain nearly parallel to the tube axis, and the entire postulated intratubular hydrogen-bond network is confirmed. However, relaxation brought significant changes both in the rigid part of the structure—the backbone—and in the side chains. An analogous minimization carried out with classical potentials well established for proteins (CHARMM) [16] yields a geometry fairly similar to ours. However, some unexpected distortions of the structure are present, notably that of the peptide unit. In the model of Gadhiri *et al.* the OCNH torsion angle of the peptide unit is nearly 180°, as in the case of amino acids [17]. Deviations are as large as 35° in the CHARMM force-field derived structure, and are at most 10° in DFT.

Figure 2 shows how some basic features of the DFT structure nicely fit into the expected range according to the amino acid database [18]. Similar overall agreement is found with the CHARMM force field, but not with the original model. This emphasized the discrepancies between *L*- and *D*-amino acids. However, the difference between the C_α-C and N-C_α bonds in *L*- and *D*-amino acids turns out to be minor in the DFT structure (0.01 Å on average), and the C-N bonds connecting *L*- and *D*-amino acids become close to those occurring in natural amino acids (average difference 0.01 Å).

Novel information is obtained for the intertubular interaction, namely, concerning the way the side chains interact. This is the more flexible part of the structure, where discrepancies between empirical and *ab initio* determinations are expected to be larger. DFT reduces the repulsion between the Ala residues, which leads to a significant decrease of the intertubular region. The interaction between Gln and Glu residues is attractive. Beyond the formation of H bonds between the Gln groups already postulated in Ref. [3], we find that H bonds are also formed between the Glu side chains. In addition, Gln and Glu interact via cross-link intratubular H bonds. Figure 3 shows the local structure associated with the formation of this H-bond network, which enhances the stability of the surfaces.

Considering the electronic properties of this material, a few questions arise as to the possibility of electrical conductivity, the effects of doping, and the existence

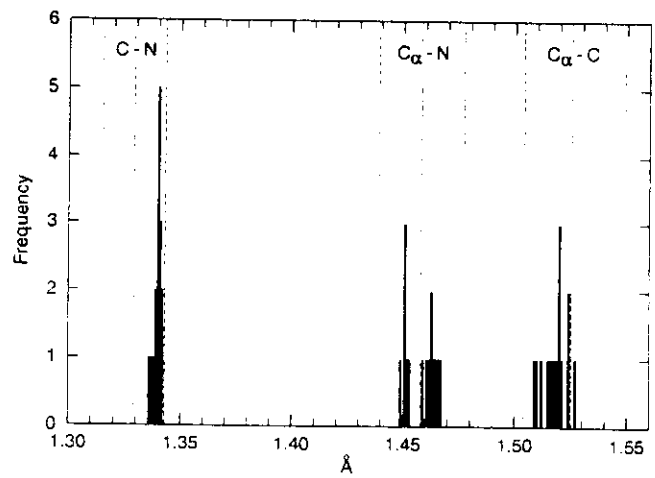

FIG. 2. Calculated distribution of key bond lengths in the backbone of the peptide subunits (solid line), compared to the range of experimental values in the amino acid database (dashed) (from Ref. [18]). Both mean value and standard deviation are reported for the latter. The same type of agreement is found for the bond angles.

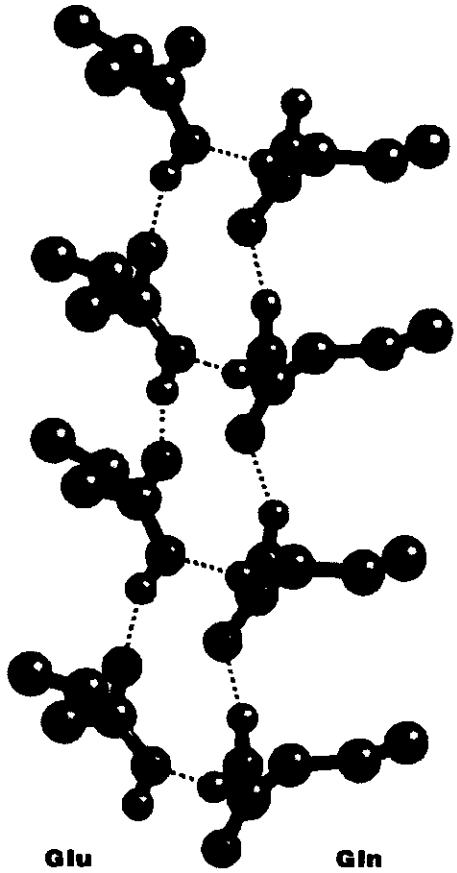

FIG. 3. Local structure of Gln and Glu side chains in the intertubular regions. H bonds are represented by dashed lines.

of localized states. The system is a large-gap (≈4 eV) semiconductor, consistent with the satisfaction of all bonds in the peptide subunits. The charge distribution is illustrated in Fig. 4(a). The overlap between neighboring tubes is minor.

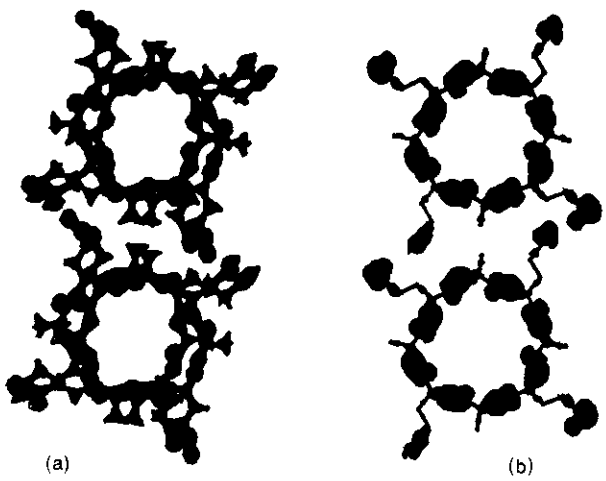

FIG. 4(color). (a) One representative isodensity surface [$0.2e/(\text{a.u.})^3$] of the valence charge distribution. (b) Probability density corresponding to the highest peak of the density of occupied states (see text). The surface value is $0.01e/(\text{a.u.})^3$.

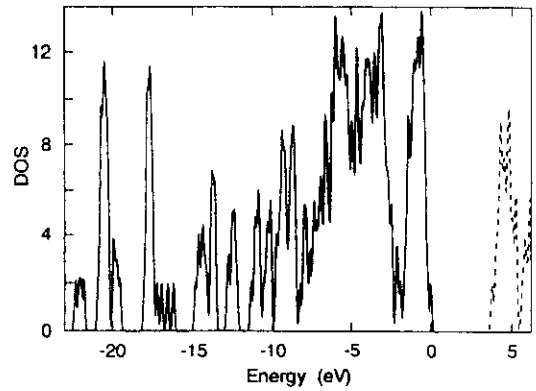

FIG. 5. Electron density of states: occupied (solid) and unoccupied (dashed).

The electron density of states (see Fig. 5) presents two characteristic peaks, both extending ≈2 eV below and above the energy gap. Both correspond to states having π character, mainly extended over the amide group of the backbone, but including contributions from the Glu and Gln groups. The spatial distribution of the peak in the occupied part of the spectrum is illustrated in Fig. 4(b). The top of the occupied spectrum and the bottom of the empty one consist of impuritylike states, namely, orbitals that are highly localized on the side chains (Gln and Glu, respectively). This feature suggests that adding a few electrons by doping will not alter the structure.

In conclusion, our DFT-based calculations have revealed interesting new structural features of self-assembled organic nanotubes and established their electronic properties. The way is now open for *ab initio* investigations of technological applications that require filling the tubes with water or incorporating specific chemicals, and for the use of Car-Parrinello [19] simulations of dynamical processes and thermodynamic stability.

We are grateful to M. R. Ghadiri for providing us with the atomic coordinates of his model, and to M. Ferraroni for providing us with the CHARMM geometry.

[1] See, e.g., S. Iijima, Nature (London) **354**, 56 (1991); A. Harada, J. Li, and M. Kamachi, Nature (London) **364**, 516 (1993); J. M. Schnur, Science **262**, 1669 (1993); A. Thess, R. Lee, P. Nikolaev, H. Dai, P. Petit, J. Robert, C. Xu, Y. H. Lee, S. G. Kim, A. G. Rinzler, D. T. Colbert, G. E. Scuseria, D. Tománek, J. E. Fischer, and R. E. Smalley, Science **273**, 483 (1996); Y. Miyamoto, S. G. Louie, and M. L. Cohen, Phys. Rev. Lett. **76**, 2121 (1996).

[2] J. D. Hartgerink, J. R. Granja, R. A. Milligan, and M. R. Ghadiri, J. Am. Chem. Soc. **118**, 43 (1996), and references therein.

[3] M. R. Ghadiri, J. R. Granja, R. A. Milligan, D. E. McRee, and N. Khazanovich, Nature (London) **366**, 324 (1993).

[4] M. R. Ghadiri, J. R. Granja, and L. Buehler, Nature (London) **369**, 301 (1994).

[5] See, e.g., K. Kirshenbaum and P. Dagget, Biochemistry **34**, 7629 (1995), and references therein.

[6] See, e.g., P. Cieplak, W.D. Cornell, C. Bayly, and P.A. Kollman, J. Comput. Chem. **16**, 1357 (1995); D.S. Maxwell, J. Tirado-Rives, and W.L. Jorgensen, J. Comput. Chem. **16**, 984 (1995).

[7] K. Takeda and K. Shiraishi, J. Phys. Soc. Jpn. **65**, 421 (1996).

[8] The residues present in this system are alanine (Ala), glutamic acid (Glu), and glutamine (Gln). *D* and *L* indicate the two optical isomers.

[9] A. Becke, Phys. Rev. A **38**, 3098 (1988); C. Lee, W. Yang, and R.G. Parr, Phys. Rev. B **37**, 785 (1988).

[10] Valence electron wave functions are considered only at the center of the Brillouin zone and are expanded in plane waves up to 70 Ry.

[11] N. Troullier and J.L. Martins, Phys. Rev. B **46**, 1754 (1992). We are grateful to P. Giannozzi for his help.

[12] P. Pulay, Chem. Phys. Lett. **73**, 393 (1980); J. Hutter, H.P. Lüthi, and M. Parrinello, Comput. Mater. Sci. **2**, 244 (1994).

[13] P. Carloni and W. Andreoni, J. Phys. Chem. **100**, 17797 (1996); J. Hutter, P. Carloni, and M. Parrinello, J. Am. Chem. Soc. **118**, 8710 (1996).

[14] M.R. Ghadiri (private communication).

[15] The conformation of the main chain of a polypeptide is defined by the torsional angles around the C_α-N (ϕ) and C_α-C (ψ) backbone bonds [see G.N. Ramachandran and V. Sassiekharan, Adv. Protein Chem. **28**, 283 (1968)]. The calculated (ϕ, ψ) values are [155(7), −126(7)] and [−145(10), 119(16)] for *D*- and *L*-amino acid residues, respectively.

[16] The classical molecular mechanics calculations were carried out by M. Ferraroni with the XPLOR 3.0 package. The simulation procedure is similar to that used by J.R. Granja and M.R. Ghadiri, J. Am. Chem. Soc. **116**, 10785 (1994). The CHARMM22 force field is from B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983). The crystal unit cell is replicated on the vector cell directions, with a 12 Å cutoff. A dielectric constant $\epsilon = 1.0$ is assumed. The energy minimization is carried out with the conjugate gradients method.

[17] C. Branden and J. Tooze, *Introduction to Protein Structure* (Garland Publishing, Inc., New York, 1991).

[18] R.A. Engh and R. Huber, Acta Crystallogr. A **47**, 392 (1991).

[19] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).