Translated from *Uspekhi Khimii*, **55**, 539-554 (1986) U.D.C. 539.194

Computer Chemistry

A.A.Ovchinnikov and A.I.Boldyrev

The state, problems, and prospects of a new branch of chemistry, computer chemistry, associated with the application of computers to the solution of chemical problems is described. The application of computers in quantum chemistry, in solution theory, in the search for new medicinal drugs, in the identification of chemical compounds from spectroscopic data, and in the development of the syntheses of new chemical compounds is examined. The level of the problems solved in the above fields of computer chemistry is demonstrated for specific examples.
The bibliography includes 64 references.

CONTENTS

I. Introduction 272
II. Quantum chemistry 272
III. Modelling of solutions 276
IV. Molecular design of medicinal drugs 277
V. "Artificial chemical intellect" 278
VI. Prospects for the development of computer chemistry 279

I. INTRODUCTION

We are now witnessing a major revolution in natural sciences associated with the wide-scale introduction of computers into all their branches, both experimental and theoretical. This process has taken place particularly vigorously during the last decade, a period during which there has been a considerable growth in the manufacture of fast computers and virtually all those investigators who are interested have gained access to powerful computers. The most important consequence of the introduction of computers into natural sciences has been the appearance of new branches of these sciences. Thus, together with experimental and theoretical physics, computer physics has appeared and in chemistry we have computer chemistry.† The time when we shall be able to speak of computer biology is apparently not distant. Whereas until recently there existed two ways of securing knowledge, either by carrying out experiments or by creating a theory, nowadays one may speak of the appearance of a third way associated with the use of computers.

The study of the physical and chemical properties of molecules, atoms, radicals, and ions in the gas, liquid, and solid phases with the aid of computers is the subject of computer chemistry. Some experience has already been gained in the solution of these problems with the aid of computers and in this review we shall characterise briefly the level of the problems solved at present by the methods of computer chemistry and we shall discuss the advances, problems, and prospects of this branch of science.

II. QUANTUM CHEMISTRY

Computer quantum chemistry constitutes the most developed branch of computer chemistry. Quantum chemists were some of the first to initiate the use of computers in their research; as a result, quantum chemists engaged in calculations are as a rule highly qualified programers and quantum-chemical programs are some of the most complex among scientific and technical programs# together with programs for nuclear physics and programs for the calculation of the trajectories of space devices.

The need to use computers in quantum chemistry is due to the fact that, although the Schrödinger equation has been known for more than fifty years, nevertheless it has been possible to solve it analytically only for the simplest cases—the hydrogen atom and the H_2^+ molecular ion. Successful quantum-chemical calculations involve the use of a whole series of approximations for the solution of the Schrödinger equation. Nevertheless, even the approximate equations obtained for any complex atoms and molecules can be solved only with the aid of computers. The most logical methods in quantum chemistry are the non-empirical (or *ab initio*) methods. This term is understood to embrace a fairly wide set of approximate methods for the solution of the Schrödinger equation for atoms and molecules. It is not fortuitous that the Latin term "*ab initio*" which can be translated as "from the beginning" has found extensive applications in all languages for the designation of these computation methods, because it emphasises that the *ab initio* methods do not use any additional experimental data (ionisation potentials, atomisation energies, etc.) apart from fundamental physical constants (the velocity of light, the Planck constant, and nuclear and electronic charges and masses) and have been devised on the basis of the Schrödinger method by introducing various physically justified approximations. By the successive abandonment of such approximations, it is possible to increase the rigour and accuracy of these calculations.

The simplest version of the *ab initio* approach consists in the solution of the system of Hartree–Fock–Roothaan (HFR) equations:

$$\sum_{\nu=1}^{n} c_{i\nu}(F_{\mu\nu} - \varepsilon_i S_{\mu\nu}) = 0, \text{ where } \mu = 1 \ldots n \ . \qquad (1)$$

In this approach the wave function is used in the form of a Slater determinant and the delocalised molecular basis orbitals (MO) are expanded in terms of one-centre atomic orbitals (AO) and the self-consistent field approximation is employed:

†A series of international journals (Computers and Chemistry, Journal of Computational Chemistry, and Journal of Chemical Information and Computer Sciences), devoted to computer chemistry are already being published.

#For example, the GAUSSIAN-80 program consists of one hundred thousand FORTRAN operators.[1]

$$\Psi^{MO} = \sum_{\mu=1}^{n} c_\mu \varphi_\mu . \qquad (2)$$

At the present time there are already tens of non-empirical programs applying the HFR method. Among foreign programs, these are GAUSSIAN, HONDO, IBMOL, POLYATOM, MOLECULE, DISCO, and others while among Soviet programs mention may be made of SPUSH) Moscow State University, KVANT (Institute of General Chemistry of the USSR Academy of Sciences), and others.

It is at present difficult to estimate what are the largest molecular systems susceptible to calculations by the HFR method. Fairly accurate calculations for molecules of the type C_6F_6 are possible even on minicomputers of the NORD-500 type.[2] Naturally, on large computers the scope for calculations is significantly greater and calculations have already been made for the $Fe[C_5(CH_3)_5]_2$ molecule[3] containing 51 atoms and 176 electrons, a complex of the prophyrin ring with an iron atom and an ammonia molecule adsorbed on it,[4] and a cluster of 13 copper atoms.[5]

In the limiting case where the basis set of AO is complete, it is possible to obtain by this method the exact total Hartree–Fock energy of the molecule, which constitutes a fairly good approximation to the exact non-relativistic value ($\geqslant$99.9% of the exact value). However, the residual error is comparable to the energies of chemical bonds and the HFR method therefore provides a poor description of the dissociation energies of chemical bonds. The example involving the dissociation energy of the F_2 molecule, where the HFR method yields a negative value, is well known. Nevertheless, many molecular characteristics can be obtained with satisfactory accuracy in calculations by the HFR method. The experience gained in calculations for several hundreds of molecules, ions, and radicals has shown that in this approach the calculated equilibrium intermolecular distances as a rule differ from the experimental values by 0.02–0.05 Å, the valence angles differ by 1–3°, the force constants differ by 10–15%, the vibration frequencies differ by 5–10%, the energies of isodesmic reactions, in which the number of bonds is kept constant (for example, $HCCH + H_2 \rightarrow H_2CCH_2$), differ by 5–10 kcal mol^{-1}, the barriers to intramolecular rearrangements differ by 2–5 kcal mol^{-1}, the valence ionisation potentials differ by 1–2 eV, and the dipole moments differ by 0.2–0.5 D.[6–8] These data show that the HFR method is entirely suitable for semi-quantitative estimations of many molecular characteristics. Although the cost of calculations by the HFR method is still high, nevertheless, bearing in mind that the calculations yield a large number of molecular characteristics which are not accessible simultaneously to any one experimental method, such calculations may be justified economically.

The second group of computational *ab initio* methods, which have developed particularly vigorously during the last five years, makes it possible in principle to obtain a virtually exact solution of the non-relativistic Schrödinger equation. In these methods account is taken of the correlation energy of the electrons, which is equal to the difference between the Hartree–Fock energy and the exact non-relativistic adiabatic energy of the molecule. A large set of such methods, both methods adapted for the solution of specific problems such as the calculation of ionisation potentials and electron affinities (the method of Green functions, the method of equations of motion, etc.) and more general computational methods for the calculation of the total energy and the wave function (the method of configuration interaction, the multiconfiguration self-consistent field method, and perturbation theory), have now been developed.

The configuration interaction (CI) method is most frequently used to take into account the electron correlation energy in *ab initio* calculations.[9–11] In it the total wave function is formulated as a linear combination of Slater determinants corresponding to different electronic configurations:

$$\Psi_{tot} = \sum_{k=1}^{M} a_{kl}\varphi_k , \qquad (3)$$

where M is the number of configurations taken into account. The calculation reduces to the computer solution of the following system of equations:

$$\Sigma a_{kl}(H_{kl} - E_l\delta_{kl}) = 0 . \qquad (4)$$

In calculations by the CI method several approaches are used for the selection of the configurations to be taken into account. One of the approaches, involving a limited CI expansion (3), is based on the estimation of the coefficients a_{kl} in terms of perturbation theory. Only the determinants whose coefficients a_{kl} are greater than a certain quantity are then left in expansion (3). In this approach it is possible to select the most significant configurations and, using not unduly long expansions, account can be taken of most of the electronic correlation energy.[11] Another approach, in which account is taken of the complete set of configurations of any type, for example singly and doubly excited configurations (CI-SD) or only doubly excited configurations (CI-D), is also widely used.[12] This method of taking into account the correlation energy uses the fact that electrons are mainly correlated in pairs. Indeed, in CI-SD and CI-D calculations it is possible to include 85–95% of the valence correlation. Many programs using this CI method are now available; the first approach is employed, for example, in the SPUSH (Moscow State University) program and the second in the GAUSSIAN-80 program.

Another widely used method whereby the electron correlation energy is taken into account is that of perturbation theory (PT).[12–14] This approach is based on seeking the solution of the exact Schrödinger equation:

$$H\Psi = E\Psi$$

on the basis of the known solutions of the Schrödinger equation

$$H^{(0)}\varphi^{(0)} = E^{(0)}\varphi^{(0)}$$

with a model Hamiltonian $H^{(0)}$ which differs from the exact equation by a small perturbation W. The wave function Ψ and the energy E are sought in the form of the series

$$\Psi = \varphi^{(0)} + \varphi^{(1)} + \varphi^{(2)} + \ldots$$
$$E = E^{(0)} + E^{(1)} + E^{(2)} + \ldots .$$

In quantum-chemical calculations the Hartree–Fock–Roothaan Hamiltonian is chosen as $H^{(0)}$ and the perturbation W is the operator describing the electronic correlation:

$$W = \sum_{i>j} \frac{1}{r_{ij}} - \sum_{i>j} (2J_{ij} - K_{ij}) .$$

For this choice of $H^{(0)}$ and W, $E^{(1)}$, i.e. the first correction to the electronic energy of the molecule, vanishes and the first non-zero contribution comes from the second-order perturbation theory (the energy $E^{(2)}$).

The contributions to the energy by each order of the perturbation theory can be formulated by means of a diagrammatic representation, which has proved extremely useful for the construction of algorithms for the corresponding program. Because of this, programs taking into account fully four PT orders were devised immediately in several laboratories in the world. One such program is the POLYCOR program developed at the Bratislava university (Czechoslovakia); the Gaussian-82 program, developed at the Carnegie-Melon university (USA), also includes the corresponding set of subprograms. It is of interest to discover how rapidly the

members of the PT series diminish. Table 1 presents the corrections to the total energies of the $LiBH_4$ and Li_2BeH_4 molecules arising from the second-, third-, and fourth-order perturbation theories. We may note that a slow decrease of the members of the PT series might have been expected for these molecules owing to the quasidegeneracy of the 2*s* and 2*p* shells of the Be and B atoms. Nevertheless, Table 1 shows that successive members of the PT series decrease rapidly. For the $LiBH_4$ and Li_2BeH_4 molecules, the contribution by the second-order perturbation theory is 90–95% of the valence correlation energy and calculations for the HF, H_2O, and BH molecules have shown[16,17] that the sum of the contributions by the second-, third-, and fourth-order perturbation theories takes into account 98–99% of the valence correlation energy. It must be emphasised that, for molecules having low-lying excited states, the PT series can diverge, as happens, for example, for the C_4 molecule.[18]

Table 1. The contributions to the total energies of the $LiBH_4$ and Li_2BeH_4 molecules by different orders of perturbation theory.[15]*

Molecule	E^{HFR}, a.u.	$E^{(2)}$, a.u.	$E^{(3)}$, a.u.	$E^{(4)}$, a.u.
$LiBH_4$	−34.4535	−0.1379	−0.0211	−0.0068
Li_2BeH_4	−31.8522	−0.0787	−0.0178	−0.0063

*DE + P + D basis set.

Since we are unable to consider other methods for taking into account the electronic correlation energy, we shall deal briefly with the new possibilities provided by *ab initio* methods with allowance for electronic correlation in quantitative calculations of molecular characteristics.

Ionisation potentials. The results of the calculations of ionisation potentials (IP) by the method of Green functions have been summarised[19] and it has been concluded that the maximum error in the calculated IP is between 0.1 and 0.25 eV depending on the basis set used in the calculations. Table 2 presents the calculated experimental IP of the ethylene molecule as a typical example.

Table 2. Calculated* and experimental ionisation potentials (eV) of the ethylene molecule.[20]

Symmetry of MO	Calcn.	Expt.
$1b_{2u}$	10.44	10.51
$1b_{2g}$	13.04	12.85
$2a_g$	14.70	14.66
$1b_{3u}$	16.07	15.87
$1b_{1u}$	19.44	19.23

*The time required to calculate one ionisation potential by the method of Green functions on the IBM 360/91 computer is only 4–5 min.

Electronic transition energies. Either the standard CI method or the CI method based on several initial configurations (CI-SIC) are most often used to calculate the electronic transition energies. As for the ionisation potentials, here the error in the determination of the energies of vertical electronic transitions does not exceed 0.1–0.2 eV and is in fact within the limits of the error of their experimental determination. As an example, Table 3 presents the calculated electronic transition energies for the *trans*-butadiene molecule.

Table 3. The energies of the electronic transitions (eV) in the *trans*-butadiene molecule.[21]

Symmetry of state	Calcn.*	Expt.
^3B_u	3.31	3.2; 3.22; 3.3
^3A_g	4.92	4.93; 4.9; 4.8
^1B_g	6.20	6.22
^1A_u	6.53	6.567 series
^1B_u	6.67	
2^1A_u	6.72	
2^1A_g	7.02	7.06
2^1B_g	7.29	7.328
3^1A_g	7.53	7.481

*Calculation by the CI method taking into account all the valence-excited configurations.

Dissociation energies. It is known that the determination of the dissociation energies of molecules to within the experimental accuracy presents the greatest difficulty for *ab initio* calculations. Table 4 lists the dissociation energies of certain molecules for which the HFR method introduces large errors; however, it is seen that allowance for the correlation energy within the framework of perturbation theory diminishes the errors in the calculated dissociation energies of these molecules to approximately 5 kcal mol^{-1}. These results show that, for small moelcules and especially for radicals, ions, and other chemically active species, whose experimental study involves considerable difficulties, computer calculation can constitute an alternative approach to the determination of their dissociation energies.

Table 4. The calculated and experimental dissociation energies (kcal mol^{-1}) of the H_2, N_2, and NH_3 molecules.[22]

Method*	$H_2 \rightarrow 2H$	$N_2 \rightarrow 2N$	$NH_3 \rightarrow N + 3H$
HFR	83.7	114.8	200.8
HFR + PT(2)	103.7	231.2	290.5
HFR + PT(2-3)	107.2	207.7	288.1
HFR + PT(2-4)	108.1	223.6	292.4
Experiment	109.4	228.4	297.3

*PT(2), PT(2–3), and PT(2–4) are corrections to the correlation energy calculated within the framework of second-, second- and third-, and second, third-, and fourth-order perturbation theory respectively.

Vibration frequencies. At the present level of the accuracy of the calculations of vibration frequencies, one may hope to achieve results with an error of 1–2%. Table 5 lists the fundamental frequencies of the water molecule calculated by the CI method for several configurations.

Table 5. The calculated and experimental fundamental vibration frequencies (cm^{-1})* of the H_2O molecule.[23]

Method of determination	ν_2	ν_1	ν_3
MC-CISD*	1623	3667	3752
Experiment	1595	3657	3756

*The method of configuration interaction based on an initial multiconfiguration function.

Equilibrium geometrical parameters. Allowance for the correlation energy makes it possible to increase the accuracy of the determination of the calculated geometrical parameters to within 0.01 Å for internuclear distances and 0.1° for valence angles. Such accuracy in the above quantities is comparable or even greater than the accuracy of their determination for molecules, radicals, and ions with more than two atoms.

The high accuracy of the *ab initio* calculations of many molecular characteristics has made them a powerful instrument, independent of experiment, for the investigation of molecules and the application of these methods has significantly expanded our ideas. We shall quote only two examples. In chemistry one of the firmly established postulates is that the maximum valence of carbon is four and, in relation to univalent ligands, its maximum coordination number is therefore also four. However, in 1982–1983, Schleyer and co-workers[24,25] predicted by means of *ab initio* calculations the existence of the stable molecules CLi_5 and CLi_6. In their equilibrium configurations these molecules have the structure of a regular trigonal bipyramid and an octahedron respectively. In addition, on the basis of non-empirical calculations, Schleyer and co-workers concluded that the hyperlithium molecules NLi_5, OLi_5, NLi_4, OLi_4, FLi_4, OLi_3, FLi_3, and FLi_2 and also the hypersodium molecules ONa_3, and ONa_4 are stable with respect to unimolecular dissociation.[24–26] The existence of some of these molecules has been confirmed experimentally. Thus the CLi_5 molecule has been detected mass-spectrometrically in the interaction of CCl_4 with lithium atoms in the gas phase. When lithium atoms diffused through a graphite membrane, CLi_6 and CLi_5 were also detected mass-spectrometrically. Mass-spectrometric measurements also led to the detection of OLi_5, OLi_4, ONa_3, and ONa_4 (for further details, see Schleyer[25]). The theoretically predicted discovery of this new class of hyperlithium and hypersodium molecules has extended our ideas concerning the possible valence states and coordination numbers of atoms and molecules.

Non-empirical calculations of the potential surfaces for a large number of structurally non-rigid molecules showed that our ideas about molecular structure are limited. The point is that modern ideas about molecular structure constitute a complex conglomerate of quantum and classical concepts. Whereas the concept of wave functions has been rapidly introduced into chemistry for the description of the electronic subsystem of the molecule as the foundation of quantum mechanics, for the nuclear subsystem the classical description has to a large extent persisted to this day. There are objective reasons for this: in the first place these are the much greater mass of the nuclei and the shape of the potential surfaces (PS) for many molecules, where a deep minimum in the PS corresponds to the equilibrium configuration (in this case one minimum on the PS is sufficient for the description of the molecular structure). For the lowest vibrational levels of such molecules, a quasi-rigid rotator–harmonic oscillator model is satisfactory. This molecular model makes it possible to retain all the classical ideas concerning the length, energy, and force constant of the chemical bond and the subdivision of all interactions into valence and non-valence types. For quasi-rigid molecules, the vibration amplitudes of the nuclei are small compared with the chemical bond lengths and the transition from the classical sphere-rod or sphere-spring model to the quantum model by the introduction of nuclear density functions does not lead to any significant modification of the classical ideas.

Nevertheless it is clear that the quantum nature of nuclei must be manifested in the molecular structure and for many problems it is essential to resort to the nuclear density functions. One can indicate several fields where the introduction of the concept of nuclear density functions is necessary for the description of the structure. In the first place these are highly excited vibrational states and "activated" or "transition" complexes in reactions. Finally, there exists a large class of molecules for which the approach restricted to one minimum in the PS is insufficient even for the lowest vibrational levels. These are the so-called structurally non-rigid or polytopic molecules on the PS of which there are at least several minima having similar energies and separated by low barriers. Such molecules have been known since the 1930s; they include organic molecules in which there is a possibility of hindered rotations about single bonds, amines in which inversion takes place, pentacoordinate MX_5 molecules with pseudorotation, etc. However, in these molecules "movements" with a large amplitude occur between atoms which are not linked by valence bonds and the classical subdivision of interactions into valence and non-valence types is not therefore infringed.

Non-empirical calculations of the PS for tens of molecules have shown[27] that large-amplitude movements (of the order of the chemical bond length) of molecules are possible also for atoms which are linked by valence bonds from the classical standpoint. The movements of the Li^+, Na^+, and K^+ cations around the CN^-, BH_4^-, and BF_4^- anions,[28–33] rearrangements with intramolecular decomposition of anions in the salts Li_2BeH_4 and Li_2BeF_4,[34,35] and the intramolecular movement of the H^+ cations along the carbon skeletons of carbonium ions[36] and of H^- anions along the beryllium and boron skeletons of hydroberyllates and hydroborates[37] may be quoted as examples. In order to describe such molecules, it is necessary to consider large regions of PS along "non-rigid" coordinates. In the vibrational adsorption spectra the lines are split owing to tunnelling through low barriers[38] and the vibrational states of these molecules must be classified in terms of irreducible representations of either the complete transposition-inversion groups or their subgroups. Curves describing the temperature variation of the thermodynamic functions of structurally non-rigid molecules show inflections and the rigid rotator–harmonic oscillator must not be used to calculate the thermodynamic functions. However, the presence of a number of maxima in the nuclear density function greater than the number of atoms in the molecule, which cannot be described within the framework of classical ideas, must be recognised as the most unusual manifestation of the structural non-rigidity of molecules.

The examples presented show that computer quantum chemistry provides an alternative approach to the study of molecules and makes it possible to discover new classes of molecules, to achieve a deeper understanding of molecular structure, and to obtain numerical values of molecular characteristics.

III. MODELLING OF SOLUTIONS

As mentioned above, the quantum chemistry of gas-phase molecules has already in fact reached the level of computer experiments. However, the majority of chemical reactions take place in solution and allowance for solvation effects constitutes an extremely difficult problem and again new significan progress has been achieved as a result of the employment of computers.

Several approaches have been developed in solution theory. In the most logical approach, the calculation is carried out by one of the quantum-chemical methods described in the previous section for a solute molecule surrounded by the maximum possible number of solvent molecules. This approach is referred to as supermolecular,[39] but it is evident that its realisation involves considerable computational difficulties, although for comparatively simple molecules such calculations have in fact been carried out.

Approximate methods are used in the study of solutions of complex organic and biological molecules: the direct optimisation method,[40] the molecular dynamics method,[41,42] the Monte Carlo method,[43,44] and others. Owing to the small amount of space available in this review, we shall confine ourselves to the consideration of the description of the structure and thermodynamics of solutions by the Monte Carlo method, which has come to be widely employed in recent years. The Monte Carlo (MC) technique is a procedure for the estimation of multiple integrals with the aid of random numbers. The average energy of a liquid system, obtained by averaging the energy values over a canonical *ensemble*, is a multiple integral of this kind. The kinetic energy is known and, for the determination of the potential energy $v(q_1 \ldots q_n)$, it is necessary to calculate its average value:

$$\langle v \rangle = \int v(q_1 \ldots q_n)\, p(q_1 \ldots q_n)\, dq_1 \ldots dq_n \quad , \qquad (5)$$

where

$$p(q_1 \ldots q_n) = \frac{\exp[-v(q_1 \ldots q_n)/kT]}{\int \exp[-v(q_1 \ldots q_n)/kT]\, dq_1 \ldots dq_n} \quad . \qquad (6)$$

The potential of the interaction between n molecules can be represented as follows:

$$V = \sum v(ij) + \sum v(ijk) + \ldots \sum v(ij \ldots n) \quad . \qquad (7)$$

In real calculations only the first sum is left in expansion (7).

The parameters of the pair potential can be determined either by *ab initio* calculations or from experimental data. The first procedure is more effective because *ab initio* calculations make it possible to investigate the interactions of molecules for any orientations, whereas experimental information refers to the equilibrium configurations. Clementi[39] examined 50 classes of atoms (H, C, O, N, and certain other atoms most frequently encountered in organic and biological molecules) as a function of the valence state of the given atom and of its environment. For these classes of atoms, he determined the parameters of the atom–atom potentials from his *ab initio* calculations; these parameters are most frequently used in calculations by the MC method.

The study of the structure and thermodynamics of solutions by the MC method reduces to the random generation of the configurations of the solvent molecules around the solute molecule, which is placed at the origin of coordinates. In order to obtain statistically reliable information, it is necessary to sort through 10^9 configurations, many of which are energetically unfavourable and therefore very improbable. In order to exclude such configurations, one employs the Markov chain technique which selects configurations taking into account the Gibbs distribution. Evidently, the larger the number of solvent molecules considered, the more reliable is the information which can be obtained concerning solvation effects. In the calculations carried out a present, account is taken as a rule of between 100 and 500 water molecules.

One of the first systems investigated by the *ab initio* MC method, was water.[45] These calculations showed that the radial distribution function for the oxygen atoms in water calculated *ab initio* without using empirical data agrees well with the experimental curve (Fig. 1). The average energy of the interaction of the molecule with the remaining molecules is approximately 7 kcal mol^{-1} according to the calculations by the MC method. Apart from pure water, aqueous solutions of LiF,[46,47] and glycine, certain other aminoacids, and other biologically important molecules[39] were investigated. The MC method was applied even in the study of the hydration of the complex DNA molecule; in these calculations a fragment of the DNA double helix was placed in a hydration shell made up of 477 water molecules.[39]

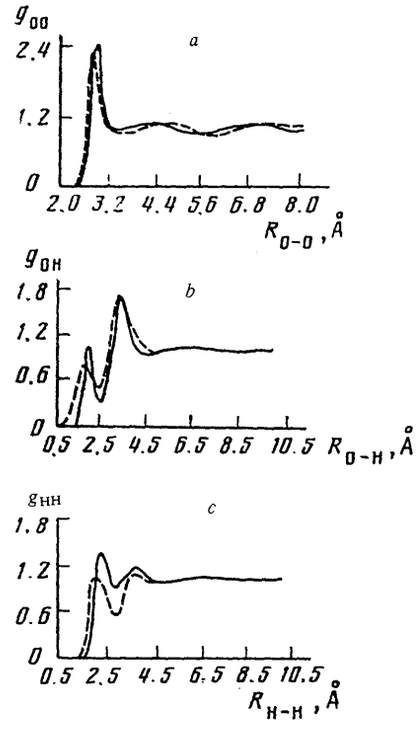


Figure 1. Comparison of the experimental (continuous lines) and calculated by the *ab initio* method (dashed lines) radial distribution functions: *a*) g_{OO}; *b*) g_{OH}; *c*) g_{HH}.[45]

The influence of solvation on the tautomerism of 4-hydroxypyridine and 4-pyridone has been investigated within the framework of the MC method.[48] In these calculations, the Clementi potentials were used and account was taken of 110 water molecules; it was found that the solvation energy of

4-pyridone is greater by ~5 kcal mol^{-1} than that of 4-hydroxypyridine, which agrees with the experimental value (~5 kcal mol^{-1}). Naturally, one must not overestimate such good agreement between the calculated and experimental quantities because 110 water molecules is not a very large number for a reliable modelling of the solvation of these molecules. We may note that the MC method permits the investigation of the influence of temperature on the structure and thermodynamics of solutions.[44]

A very interesting aspect of solution theory involves the combination of the supermolecular and Monte Carlo methods, in which the geometrical structure of the solvation shell is determined by the MC method, while a quantum-chemical calculation makes it possible to determine the influence of solvation on the electronic structure of the solute molecule, which influences in its turn the electronic structure of the solvent molecules.

IV. MOLECULAR DESIGN OF MEDICINAL DRUGS

The major advances in medicine achieved during the last decade are to a large extent due to new highly effective medicinal drugs. However, in order to introduce new drugs into medical practice, it is necessary to synthesise and test the physiological activities of approximately 20000 chemical compounds, which requires up to 7–10 years of work by investigators and a large financial outlay. The use of the computer-aided molecular design of new chemical compounds makes it possible to overcome these difficulties and leads to great possibilities in the purposeful search for medicinal preparations.

The essential feature of the new approach is that the physiological action of a medicinal drug is related to its stereospecific interaction with a receptor, which can be hydrophobic or of the type involving hydrogen bonding between the active fragment of the drug molecule (pharmacophor) and the biologically important centre of the receptor. At the molecular level, the problem therefore reduces to isolating the active and toxic fragments of the drug molecule and to the determination (by considering both the short-range and long-range environment) of the optimum chemical composition and structure of the molecule such that the useful interaction between the pharmacophor and the receptor is most favourable, while toxic fragments are as far as possible absent. The isolation of such active and toxic fragments can be achieved by comparing the useful biological activities and toxicities of a large number of compounds in which individual fragments are substituted in succession.

After this stage of the research has been carried out, the problem of increasing the physiological activity and of decreasing the toxicity of the preparation can be solved in several ways. The most logical procedure involves a direct non-empirical quantum-chemical calculation of the electronic structures of a set of different derivatives of the initial molecule with the receptor model. The results of such calculations can be compared with, for example, the strength of the hydrogen bond between the active centre of the preparation and the receptor, which might serve as a basis for the selection of molecules with the highest physiological activity. However, such an approach nevertheless requires much computational effort in the case of complex biologically important molecules and therefore so far other simpler procedures are used.

The structure of the receptor for biologically active compounds of a given series as well as the mechanism of their interaction are usually unknown *a priori* and must be determined together with the identification of the pharmacophor molecule and the determination of its optimum steric and electronic structures. An essential condition for the solution of this problem is the availability of extensive experimental data reflecting the relation between the structures and biological activities of the chemical compounds investigated. The molecules of biologically active substances are conformationally labile and an important feature of such investigations must therefore be the determination of biologically active conformations in which there is a possibility of interaction with the receptor. The selection of the total number of low-energy conformations of each molecule investigated can be carried out both by semiempirical quantum-chemical methods and on the basis of model force fields, whose parameters are determined either empirically or from non-empirical calculations for simple molecules. In the field under consideration, the method of atom–atom potentials, which makes it possible at present to investigate fairly complex molecules (for example, those of oligopeptides), has come to be most widely used among the theoretical methods of conformational analysis.

Biologically important molecular conformations contain information about the topography of the receptor and can therefore be used in the purposeful search for new compounds with a high affinity for the receptor. Apart from the steric structure of the pharmacophor, one can therefore optimise also its electronic structure.

One of such methods reduces to the calculation of the electrostatic potential surface of an isolated molecule on the assumption that the electrostatic potential reflects the ability of the active centre of the therapeutic model to form a hydrogen bond. This approach requires a preliminary quantum-chemical calculation, which, however, is carried out only for an isolated molecule. Programs for the electrostatic potential have now been developed and are included in the sets of programs for both non-empirical and semiempirical (of the CNDO, MINDO/3, MINDO, etc.) methods of calculation.

Thus the problem of the analysis of the biological activity of the drug molecule reduces to the determination of the most stable molecular conformation in which the atoms in the active centre are accessible to interaction and the environment is complementary with respect to the corresponding receptor atoms. The relevant programs have been developed and are already widely used for the solution of specific problems. We shall quote below several examples of the successful purposeful search for medicinal drugs with the aid of the methods described above.

One of the known examples is provided by a series of theoretical studies which led to a preparation analogous to somatostatin in its physiological activity but remaining active for a longer period. We may recall that somatostatin increases the susceptibility to insulin in patients suffering from diabetes, but natural somatostatin molecules retain their physiological activity in the human body for only a short time. A group of investigators led by Weber[49] analysed a series of somatostatin conformations by the molecular modelling method and identified the active component of the molecule with the aid of computer graphics. Next, nine aminoacids were removed from somatostatin in a computer experiment and it was thus possible to stabilise the structure, which was in Weber's view most complementary to the corresponding receptor atoms. When the molecule for which the calculations had been made was synthesised in the laboratory and its effect on animals was tested, it was found that the new preparation retained its activity in the animal organism 40 times longer, whilst retaining the physiological activity of the natural hormone.

Another example involves the theoretical study of cardiac glucosides used in medicine to treat heart failure. In therapeutic doses the glucosides increase the strength of the contractions of the myocardium, reduce the frequency of the

contractions of the heart, and diminish the excitability of the myocardium. A major disadvantage of the cardiac glucosides employed hitherto is their high toxicity, to the overcoming of which investigators have devoted much effort. A complete theoretical conformational analysis of 46 cardiosteroids with known biological activity and a wide variety of monosaccharide residues in the corresponding monoglycosides has been carried[50,51] by the method of atom–atom potential functions. These calculations led to the determination of the biologically active conformations of the cardiac glycosides, which made it possible to understand the steric structure of the digitalis receptor and the mechanism of the glycoside–receptor interaction and also it permitted the prediction of a series of configurations of drug molecules which should exhibit an enhanced affinity for the receptor. These studies may serve as a basis for the purposeful synthesis of new highly effective preparations.

The advances achieved by means of computer molecular design in the development of new effective medicinal drugs give rise to the hope that the use of computers will contribute also to the development of antitumour preparations.

V. "ARTIFICIAL CHEMICAL INTELLECT"

As computers came to be used on an increasingly wide scale, it became understood that the computer can compete with man even in fields which have been traditionally regarded as intellectual. It is possible to identify two fields in chemistry where one can already speak in a certain sense of the creation of an "artificial chemical intellect". Firstly, these are problems associated with the recognition of the structure of a molecule from the available experimental spectroscopic data. Secondly, these are problems concerning the development of a strategy for the synthesis of complex organic molecules.

The development of chemistry and especially the application of physical methods to the study of molecular structure has led to an information explosion during the last decades. It is frequently much simpler to synthesise a compound than to identify it. One of the ways of solving the complex problem of the identification of a new compound consists in the recognition of a chemical compound by comparing its spectrum with one available in the literature. However, since information about the spectra of more than 100000 organic and inorganic compounds has already accumulated in the world and the amount of this information is increasing, it follows that realistically such a comparison can be made effectively only with the aid of a computer. Such computer recognition systems for chemical compounds have already been developed by Koptyug and co-workers and have been used for a number of years in the Siberian Division of the USSR Academy of Sciences.[52, 53]

In another approach to the identification of compounds from spectra, the molecule is represented as a set of separate structural fragments such as aromatic rings, methyl groups, amino-groups, etc. which have characteristic spectral features. In each specific spectrum of the compound to be identified one isolates features which are regarded as characteristic. After such characteristic spectral features have been found, each is brought into correspondence with a set of competing structural fragments. Next, for any permissible set of fragments, all the possible molecular structures satisfying the overall formula and the requirements of the classical theory of valence are constructed. Among all the structures thus found, a further selection is now made on the basis of spectroscopic parameters, which are calculated, for example, in the case of infrared spectra, from the force constants of the fragments. Information about the experimental spectroscopic parameters of the fragments is stored in the computer—in a library of standard fragments. The closeness of the calculated spectrum of a particular structure to the experimental spectrum of the molecule to be identified provides the basis for a conclusion concerning the correctness of the recognition of its structure. Evidently, when several spectroscopic methods, such as mass spectrometry and infrared, Raman, NMR, and ultraviolet spectroscopy are used in the identification process, the reliability of the determination of the molecular structure increases significantly. Sets of programs for carrying out such computer analysis have been developed under Gribov's supervision[54,55] and have already been used successfully for several years.

Apart from the problem of the recognition of molecules from the spectra, computers are being introduced into another field of "chemical thinking", namely the development of a strategy in synthesis. So far several trends have been established in the computer synthesis of organic compounds.

One of them is based on the determination of the possible ways of synthesising the required compound on the basis of thoroughly investigated organic reactions. The corresponding programs have been developed in several laboratories. One of the most popular programs, LHASA, was developed at Harvard University.[56] It consists of approximately 50000 FORTRAN operators and its data bank contains information about approximately 1000 reactions. In developing a strategy for synthesis, the LHASA program as it were solves the converse problem, i.e. it begins with the molecule which is to be synthesised and then, by "breaking" the bonds in it via a series of reactions, it arrives at comparatively simple initial compounds which are either commercially available or can be readily obtained in the laboratory.

Another trend involves a formal logical approach based on the classical theory of valence, which regards a reaction as an intermolecular or intramolecular redistribution of bonds (cyclic electron transfer). In this case the ways of obtaining a given compound can be determined from its structure by the methods of graph theory and the combinatorial theory (for further details, see Zefirov and co-workers[57,58]). The advantage of this approach consists in the fact that it makes it possible to generate on a computer processes which have no direct analogues among the known reactions. The program FLAMINGO has been developed on this unique logical basis under Zefirov's supervision and has been designed to find all possible single-state processes with cyclic electron transfer in both closed and open systems taking into account specified conditons.

The application of this program to the computer synthesis of tricyclo[2.1.0.$0^{2,5}$]pentane has been described.[58] All single-stage many-centre reactions with cyclic electron transfer directed towards the synthesis of this compound were found with the aid of a computer. The following initial data were specified: (1) the graph representing the skeleton of tricyclo[2.1.0.$0^{2,5}$]pentane with numbered atoms (Fig.2); (2) information about the molecular symmetry; (3) information necessary to select the possible reactions. The computer sorted through all the possible processes involving the redistribution of bonds between the carbon atoms on the basis of the initial graph and predicted 6 three-centre, 12 four-centre, and 11 five-centre reactions with cyclic electron transfer. On the basis of special criteria, 19 reactions were then rejected and the remaining 10 (Fig.2) were recommended as possible ways of synthesising tricyclo[2.1.0.$0^{2,5}$]pentane. Some of them have not so far been applied; we may note that, since the publication of the above communication, a reaction prediced by the computer has already been carried out under laboratory conditions.[58]

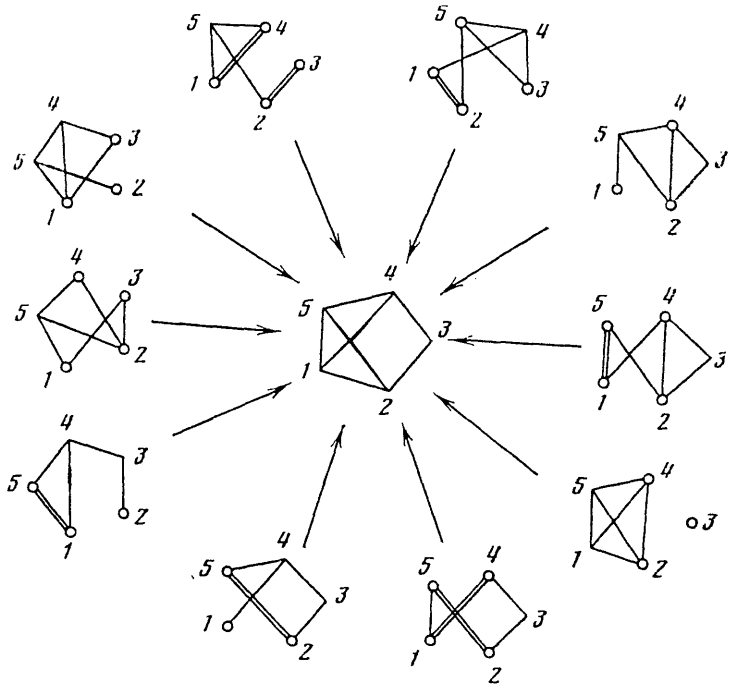
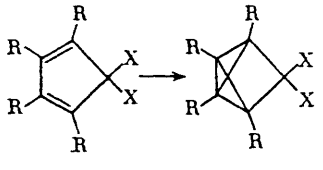


Figure 2. Schematic representation of the possible ways of constructing the tricyclo[2.1.0.$0^{2,5}$]pentane system found with the aid of a computer.



The field in which the "artificial chemical intellect" is developing is nowadays undergoing a vigorous growth and industrial firms interested in this problem already participate in the creation of similar programs. The computer programs described in this Section apparently simulate the thinking process of a research chemist, which in fact makes it possible to classify the operation of such programs as "artificial chemical thinking". An advantage of the computer over the human brain is that it can memorise an enormous amount of information and can manipulate it fairly rapidly.

VI. PROSPECTS FOR THE DEVELOPMENT OF COMPUTER CHEMISTRY

Within the framework of a single review, it is at present already difficult to deal even briefly with all the fields of the rapidly developing new science—computer chemistry, which has arisen at the junction of chemistry, physics, mathematics, and computer technology. For example, apart from the applications considered above, one can point out also other important applications of computer chemistry such as the solution of the problem of optimising chemical technology with the aid of a computer and the solution of dynamic problems in the theory of elementary chemical processes. Evidently, as computers become more available to chemists, new fields of computer chemistry will arise.

The development of computer chemistry is in many respects determined by the advances in computer technology, which have been very great during recent years. Supercomputers capable of performing 100–200 million operations per second exist and, towards the end of the 1990s, the creation of supercomputers capable of 1–10 billion operations per second are expected. At the same time the size of both the operating memory of computers and the capacity of peripheral storage devices has been increasing. Discs with a capacity of one gigabyte have already been created and work on the creation of videodiscs is being completed.

The use of these new possibilities makes it possible to solve fundamentally new problems in computer chemistry, to raise the level of accuracy in the problems already dealt with, and to extend the range of applications. In quantum chemistry this involves taking into account the relativistic effects in molecules and calculations for large systems taking into account electron correlation; in principle, one may expect the appearance of dialogue programs, which would enable the chemist, before synthesising any one compound, to investigate theoretically its energy stability and the properties of its molecules. In solution theory, the supermolecular approach will become available for large organic molecules as the computer technology is improved and in methods of the Monte Carlo type it will be possible to take into account tens of thousands of solvent molecules, which will permit the investigation of bimolecular reactions in solution. In studies on molecular design, it will be possible to consider a large number of medicinal drug molecules and to take into account tens of their parameters in identifying the active centres, which will significantly increase the reliability of the prediction of the activities of new preparations. In studies on the "artificial chemical intellect", the data banks

will increase to tens of thousands of reactions, which will enable us to increase fundamentally the effectiveness of the synthetic procedures being developed; in problems on molecular structure recognition, the increase in the number of spectra in the data bank will also raise the reliability of the identification of new compounds. The development of chemistry is nowadays no longer imaginable without computers.

---o0o---

During the preparation of the review for the press, a number of other review articles, devoted to various branches of computer chemistry were published.[59-64] Clementi[59] described new results obtained in the *ab initio* modelling of water (three- and four-particle interactions, and also corrections associated with zero-point vibrations were taken into account), the hydration of a DNA fragment (30 pairs of bases surrounded by 1500 water molecules), and the transport of Na^+ and K^+ ions through membrane channels. The results of the application of supercomputers to the calculation of the structure of the solvated electron and the structures of liquid sulphur, amorphous nickel, and silicon in crystalline, liquid, and low-temperature amorphous states have been discussed in a review.[60] Other communications[61-64] have been devoted to the prospects for the creation and use of supercomputers in chemistry, the computational aspects of the determination of crystal structures, and non-empirical quantum-chemical methods.

REFERENCES

1. J.S.Binkley, R.A.Whiteside, R.Krishnan, R.Seeger, D.J.DeFrees, H.B.Schlegel, S.Topiol, L.R.Kahn, and J.A.Pople, Quantum Chemistry Program Exchange, 1981, **13**, 406.
2. J.Almlof and K.Faegri, J.Chem.Phys., 1983, **79**, 2284.
3. H.P.Luthi, P.E.M.Siegbahn, J.Almlof, K.Faegri, and A.Heiberg, Chem.Phys.Letters, 1984, 111, 1.
4. M.M.Romer, A.Dedieu, and A.Veillard, Chem.Phys., 1984, **77**, 419.
5. C.Bachmann, J.Demuynck, and A.Veillard, Faraday Symp., Chem.Soc., 1984, (14), 170.
6. H.F.Schaefer, "The Electronic Structure of Atoms and Molecules. A Survey of Rigorous Quantum Mechanical Results", Addison-Wesley Publ., Reading, 1972.
7. P.Carsky and M.Urban, "Ab initio Calculations. Methods and Applications in Chemistry. Lecture Notes on Chemistry", Springer Verlag, Berlin, 1980, Vol.16.
8. S.Fujinaga, "The Molecular Orbital Method" (Translated into Russian), Izd.Mir, Moscow, 1983.
9. W.Meyer, in "Methods of Electronic Structure Theory", (Edited by H.F.Schaefer), Plenum Press, New York, 1977, p.413.
10. I.Shavitt, in "Methods of Electronic Structure Theory", (Edited by H.F.Schaefer), Plenum Press, New York, 1977, 189.
11. I.Shavitt, in "New Horizons of Quantum Chemistry" (Edited by P.-O.Lowdin and B.Pullman), D.Reidel Publ., New York, 1983, p.279.
12. I.Hubac and P.Carsky, Topics Curr.Chem., 1978, **75**, 97.
13. R.J.Bartlett, Ann.Rev.Phys.Chem., 1981, **32**, 359.
14. J.Paldus and J.Cizek, Adv.Quant.Chem., 1975, **9**, 105.
15. V.Kello, M.Urban, and A.I.Boldyrev, Chem.Phys. Letters, 1984, **106**, 455.
16. F.B.Brown, I.Shavitt, and R.Shepard, Chem.Phys. Letters, 1984, **105**, 363.
17. R.J.Bartlett, H.Sekino, and G.D.Purvis, Chem. Phys.Letters, 1983, **98**, 66.
18. R.A.Whiteside, R.Krishann, D.J.DeFrees, J.A.Pople, and P.V.R.Schleyer, Chem.Phys.Letters, 1981, **78**, 538.
19. V.W.Niessen, G.H.F.Diercksen, and L.S.Cederbaum, J.Chem.Phys., 1977, **67**, 4124.
20. V.W.Niessen, G.H.F.Diercksen, L.S.Cederbaum, and W.Domcke, Chem.Phys., 1976, **18**, 469.
21. R.J.Buenker, S.Shih, and S.D.Peyerimboff, Chem. Phys.Letters, 1976, **44**, 385.
22. M.J.Frisch, J.A.Pople, and J.S.Binkley, J.Chem. Phys., 1984, **80**, 3265.
23. G.J.Sexton and N.C.Handy, Mol.Phys., 1984, **51**, 1321.
24. P.v.R.Schleyer, E.-U.Wurthwein, E.Kaufmann, and T.Clark, J.Amer.Chem.Soc., 1983, **105**, 5930.
25. P.V.R.Schleyer, in "New Horizons of Quantum Chemistry (Edited by P.-O.Lowdin and B.Pullman), D.Reidel Publ., New York, 1983, p.95.
26. P.v.R.Schleyer, E.-U.Wurthwein, and J.A.Pople, J.Amer.Chem.Soc., 1982, **104**, 5839.
27. O.P.Charkin and A.I.Boldyrev, "Potentsial'nye Poverkhnosti i Strukturnaya Nezhestkost' Neorganicheskikh Molekul. Itogi Nauki i Tekhniki. Ser.Neorganicheskaya Khimiya" (The Potential Surfaces and the Structural Non-rigidity of Inorganic Molecules. Advances in Science and Engineering. Inorganic Chemistry Series), Izd.VINITI, Moscow, 1980, Vol.8.
28. E.Clementi, H.Kistenmacher, and H.Popkie, J.Chem. Phys., 1973, **58**, 2460.
29. R.Essers, J.Tennyson, and P.E.S.Wormer, Chem. Phys.Letters, 1982, **89**, 223.
30. M.L.Klein, J.D.Goddard, and D.G.Bounds, J.Chem. Phys., 1981, **75**, 3909.
31. P.E.S.Wormer and J.Tennyson, J.Chem.Phys., 1981, **75**, 1245.
32. A.I.Boldyrev, O.P.Charkin, N.G.Rambidi, and V.I.Avdeev, Chem.Phys.Letters, 1976, **44**, 20.
33. V.G.Zakzhevskii, A.I.Boldyrev, and O.P.Charkin, Chem.Phys.Letters, 1980, **73**, 54.
34. A.I.Boldyrev, L.P.Sukhanov, and V.G.Zakzhevskii, Chem.Phys.Letters, 1981, **79**, 421.
35. E.-U.Wurthwein, M.B.Krogh-Jespersen, and P.v.R.Schleyer, Inorg.Chem., 1981, **20**, 3663.
36. K.Raghavachari, R.A.Whiteside, J.A.Pople, and P.v.R.Schleyer, J.Amer.Chem.Soc., 1981, **103**, 5649.
37. Yu.B.Kirillov and A.I.Boldyrev, Dokl.Akad.Nauk SSSR, 1983, **272**, 635.
38. L.Ya.Baranov and A.I.Boldyrev, Chem.Phys.Letters, 1983, **96**, 218.
39. E.Clementi, "Computational Aspects for Large Chemical Systems, Springer Verlag, Berlin, 1980, p.184.
40. T.Langlet, P.Claverie, F.Caron, and T.L.Bocuve, Int.J.Quant.Chem., 1981, **19**, 299.
41. B.J.Alder and T.Wainwright, J.Chem.Phys., 1957, **27**, 1209.
42. P.J.Roossky and M.Karplus, J.Amer.Chem.Soc., 1979, **101**, 1913.
43. W.W.Wood and J.J.Erpenbeck, Ann.Rev.Phys.Chem., 1976, **27**, 319.
44. G.A"Croxton, "The Physics of the Liquid State" (Translated into Russian), Izd.Mir, Moscow, 1978.
45. G.C.Lie, E.Clementi, and M.Yoshimine, J.Chem. Phys., 1976, **64**, 2314.
46. E.Clementi and R.Barsotti, Theor.Chim.Acta, 1976, **43**, 101.

47. J.Fromm, E.Clementi, and R.O.Watts, J.Chem.Phys., 1975, **62**, 1378.
48. B.Ya.Simkin and I.I.Sheikhet, Zhur.Strukt.Khim., 1983, **24**, 80.
49. J.B.Tucker, High Technolog., 1984, (1), 38.
50. I.L.Shamovskii, G.M.Barenboim, and A.A.Ovchinnikov, Bioorg.Khim., 1982, **8**, 1677.
51. I.L.Shamovskii, G.M.Barenboim, and A.A.Ovchinnikov, Bioorg.Khim., 1983, **9**, 1112.
52. G.P.Ul'yanov, A.R.Maslov, B.G.Derendyaev, V.N.Piottukh-Peletskii, and V.A.Koptyug, Vestnik Akad.Nauk SSSR, 1978, (3), 32.
53. B.G"Derendyaev, V.A.Koptyug, K.S.Lebedev, and O.P.Sharapova, Avtometriya, 1979, (4), 3.
54. A.A.Gribov, "Vvedenie v Molekulyarnuyu Spektroskopiyu" (Introduction to Molecular Spectroscopy), Izd.Nauka, Moscow, 1976.
55. M.E.Elyashberg, L.A.Gribov, and V.V.Serov, "Molekulyarnyi Spektral'nyi Analiz i EVM" (Molecular Spectroscopic Analysis and Computers), Izd.Nauka, Moscow, 1980.
56. E.J.Corey and N.F.Feiner, J.Org.Chem., 1980, **45**, 757.
57. N.S.Zefirov, S.S.Trach, and O.S.Chizhov, "Karkasnye i Politsiklicheskie Soedineniya. Molekulyarnyi Dizain na Osnove Izomorfnogo Zameshcheniya. Itogi Nauki Tekhniki. Ser. Organicheskaya Khimiya" (Skeletal and Polycyclic Compounds. Molecular Design Based on Isomorphous Substitution. Advances in Science and Engineering. Organic Chemistry Series), Izd.VINITI, Moscow, 1979, Vol.3.
58. N.S.Zefirov and S.S.Trach, Zhur.Org.Khim., 1981, **17**, 2465.
59. E.Clementi, J.Phys.Chem., 1985, **89**, 4426.
60. R.Pagani, Chem.and Enginer.News, 1985, **63**, 7.
61. J.S.Rollett, J.Mol.Struct., 1985, **130**, 1.
62. G.M.Sheldrick, J.Mol.Struct., 1985, **130**, 9.
63. J.E.Boggs, J.Mol.Struct., 1985, **130**, 32.
64. G.A.Jeffrey, J.Mol.Struct., 1985, **130**, 43.

Institute of Chemical Physics,
USSR Academy of Sciences,
Moscow