APPLICATIONS OF REGULARIZING ALGORITHMS IN STRUCTURAL CHEMISTRY

Kuramshina G. M. , Yagola A. G.

Abstract Regularizing algorithms developed for joint treatment of gas-phase electron diffraction and vibrational spectroscopic data and extended to include systems with large-amplitude oscillatory motion are discussed. The treatment is augmented by the inclusion of microwave rotational constants. The analysis of data from experimental sources is guided by quantum mechanical molecular calculations of molecular geometry and force field. The computed force field matrix can be corrected empirically by using scale factors determined both in Cartesian and internal coordinates.

Key words: inverse problem, nonlinear ill-posed problem, theory of regularization, vibrational spectra, quantum mechanical regularized molecular force field.

AMS Mathematics Subject Classification: 49N45, 65F18, 92E10.

1 Introduction

In previous paper [1] we have described approaches for the molecular force field calculations based on the theory of regularization. The main idea of our approaches is based on an understanding that the ill-posed inverse vibrational problem is undetermined. To find the approximate solutions of such problem we should apply the available additional information on problem (experimental data, model assumptions, etc.) and to formulate certain criteria for choosing the concrete solution. Then on a base of the formalized criteria we can construct the regularizing algorithms or regularizing operators for solving the ill-posed problem and formulate a principle for choosing a unique solution from the set of possible ones. The solution should be close to the given matrix of force constants and should satisfy all a priori assumptions concerning the force field model.

In the present paper we perform results on applications of new algorithms to some practical problems. We also discuss results on development of new physical and mathematical models describing the complicated molecular system as well as some technical details of overcoming difficulties arising in calculations of large molecular systems.

2 Molecular force field models. Regularized quantum mechanical force field

We consider a molecule as a mechanical system of nuclei while all the interactions due to electrons are included in an effective potential function $U(q_1, \ldots, q_n)$, where $q_1, \ldots, q_n$ denote the $n = 3N - 6$ generalized coordinates of the N atomic nuclei of the

molecule. The potential function minimum (with respect to nuclei coordinates) defines the equilibrium geometry of the molecule, and the second derivatives of the potential with respect to nuclei coordinates in the equilibrium

$$f_{ij} = \left.\frac{\partial^2 U}{\partial q_i \partial q_j}\right|_{eq} \qquad (i, j = 1, \ldots, n)$$

constitute a positive defined matrix F determining all the molecular characteristics connected with small vibrations. The choice of generalized coordinates $\{q_i\}$ determines the molecular force field model. The so-called internal coordinated which include the bond stretching, valence bond bending (deformation), out-of-plane bending (determined by changes of dihedral angles) and torsion are the most popular coordinates used in vibrational spectroscopy. From set of internal coordinates one can construct sets of symmetry or local symmetry coordinates, etc. The vibrational frequencies (obtained from IR and Raman spectra) are the main type of experimental information on molecular vibrations. They are connected with the matrix of force constants by the eigenvalue equation

$$GFL = L\Lambda \tag{1}$$

where Λ is a diagonal matrix consisting of the squares of the molecular normal vibration frequencies $\omega_1, \ldots, \omega_n$, $\Lambda = \text{diag}\{\omega_1^2, \ldots, \omega_n^2\}$, and G is the kinetic energy matrix in the momentum representation. L is a matrix of normalized relative amplitudes. We guess that the force field of a molecule does not depend on the masses of the nuclei, and for m isotopic molecular species we have, instead of (1), the system

$$(G_i F)L_i = L_i \Lambda_i, \qquad i = 1, 2, \ldots, m. \tag{2}$$

Rotational-vibrational spectra (Coriolis constants), gas electron diffraction (mean square amplitudes), etc., can be used as an additional information for finding the force constant matrix F.

On a base of formulization of all possible obvious (and not so obvious) model assumptions concerning the character of force fields which are widely used in vibrational spectroscopy we have formulated a principle for choosing a unique solution from the set of solutions [2] in terms of the closeness of the solution to the given matrix of force constants satisfying all a priori assumptions concerning the solution. Inverse vibrational problem is formulated as a problem of finding the so-called normal solution (or normal pseudo-(quasi-)solution in the case of incompatibility of input data) of a nonlinear operator equation (4) [2–4].

The mathematical relation between the molecule vibrational properties ((1)- (2), etc.) and its experimental display can be summarized in the form of a single operator equation

$$AF = \Lambda_\delta \tag{3}$$

where $F \in Z \in R^{n(n+1)/2}$ (Z is a set of possible solutions) is the unknown force constant matrix (real and symmetrical), $\Lambda \in R^m$ represents the set of available experimental data (vibrational frequencies, etc.) determined within δ error level: $\|\Lambda - \Lambda_\delta\| \le \delta$. A is a nonlinear operator which maps matrix F on the Λ.

On a base of formulization of all possible obvious (and not so obvious) model assumptions concerning the character of force fields which are widely used in vibrational spectroscopy we have formulated a principle for choosing a unique solution from the set of solutions [2] in terms of the closeness of the solution to the given matrix of force constants satisfying all a priori assumptions concerning the solution. Inverse vibrational problem is formulated as a problem of finding the so-called normal solution (or normal pseudo-(quasi-)solution in the case of incompatibility of input data) of a nonlinear operator equation (4) [2–4].

The desirable solution is a matrix $F^\alpha \in Z$ that reproduces experimental data within given error level and is the nearest in the Euclidean metrics to some given matrix F^0. All necessary model assumptions (explicit and implicit) concerning the form of force field may be taken into account by the choice of some given *a priori* matrix of force constants F^0 and a preassigned set D of *a priori* constraints on values of the force constants. This set defines a form of matrix F in the framework of the desired force field model (i.e., with specified zero elements, equality of some force constants, etc.). If no *a priori* data constrains the form of solution, then D coincides with the set Z.

It was mentioned in Refs [5,6], that in the Tikhonovв's regularizing procedure, one can increase the stability and accuracy of the calculated solution F^α by using

a) an extended set of experimental data (including, e.g., Coriolis constants, mean square amplitudes, frequencies of isotopomers or related compounds, etc.);

b) an improved choice of the stabilizer matrix F^0;

c) an improved choice of the constraint set D.

As a particularly effective choice of stabilizer, we have proposed [4–7] to use an *ab initio* quantum mechanical F^0 matrix in the regularizing procedure. This leads to the concept of regularized quantum mechanical force field (**RQM FF**), defined as the force constant matrix that is nearest to a corresponding quantum mechanical matrix F^0 and reproduces experimental frequencies within given error level.

The correct choice of constraint set D is also extremely important. Physically stipulated limitations may either decrease the range of acceptable matrices F, or provide criteria for selecting a concrete solution from a set of tolerable ones. An incorrect choice of constraints may lead to increasing incompatibility of the inverse problem, eventually resulting in a pseudosolution having no physical meaning. A set of a priori constraints may arise from several types of limitations on force constant values, e.g. [8,9]:

1. some force constants may be stipulated on a priori grounds to be a zero;

2. some force constants may be stipulated to satisfy inequalities $a_{ij} \leq f_{ij} \leq b_{ij}$, where a_{ij}, b_{ij} are certain known values;

3. some force constants may be stipulated to be equal in a series of related molecules (or conformers);

4. the final solution may be stipulated to conform to Pulay's scaled force constant matrix [10], which may also be considered as a kind of constraint.

In the last (Pulay) approach we can specify the set D as: $D = \{F : F = BF^0B\}$, $B = \mathrm{diag}\{\beta_1, \ldots, \beta_n\}$ (where β_i are the scaling parameters) [11].

In this case the following strict mathematical formulation of the inverse scaling problem has been proposed [11]: the problem of finding scaling factors on the base of experimental data is treated as an operator equation similar to (4):

$$AF(\beta) = \Lambda_\delta \tag{4}$$

where β are scaling factors. Let the following norms in the Euclidean space be introduced:

$$\|F\| = \left(\sum_{ij}^{n} f_{ij}^2\right)^{1/2}, \quad \|\Lambda\| = \left(\sum_{k}^{l} \lambda_k^2 \rho_k\right)^{1/2},$$

where $\rho_k > 0$ are certain positive weights; f_{ij} are the elements of matrix F; $\lambda_k(k = 1, \ldots, m)$ are the components of Λ.

Since problem (4) is also ill-posed, it should be regularized. We formulate the problem as a requirement to find a solution of (4), $F_{n,\delta}$, that is nearest (by Euclidean norm) to the quantum mechanical matrix F^0, satisfies experimental data within a given error level δ ($\|A(F(\beta)) - \Lambda_\delta\| \leq \delta$) and has a special form proposed by Pulay. If we consider this problem taking into account its possible incompatibility (within harmonic model), we come to the following formulation:
to find

$$F_{n,\delta} = \arg\min \left\|F - F^0\right\|$$

where

$$F \in \{F : F \in D = \{F : F = B^{1/2}F^0B^{1/2}\}, \|AF - \Lambda_\delta\| \leq \mu + \delta\}. \tag{5}$$

Here B is a diagonal matrix of scaling factors β_1, and μ is a measure of incompatibility of the problem [3]. It may arise due to the possible anharmonicity of experimental frequencies or the crudeness of the chosen model.

Such a solution can be determined by minimization of the Tikhonov functional

$$M^\alpha(\beta) = M^\alpha[F] = \|A_hF - \Lambda_\delta\|^2 + \alpha\left\|F - F^0\right\|^2 \tag{6}$$

where $F = F(\beta)$, and regularization parameter α is chosen in accordance with the generalized discrepancy principle [12–15].

3 Correction of quantum mechanical force field in Cartesian coordinates

As a rule, the limitations on the values of force constants of polyatomic molecules cannot be strictly proved. In this situation, the numerical quantum mechanical results on molecular force fields can provide useful guidance in choosing realistic force field models for different types of molecules. The simplicity of the scaling procedure made it extremely popular in recent years. Results of quantum mechanical calculations have demonstrated that the scale factors of many molecular fragments (determined within a given level of quantum-mechanical method) are approximately constant in a wide range of similar molecules. In the most cases scaling factors of a large number of molecules obtained for different sets of coordinates and quantum-mechanical methods allow to approximate experimental frequencies with a reasonable degree of accuracy.

Initially the regularized scaling procedure have been developed [11, 16] for the force fields defined in the internal or symmetry (local symmetry) coordinate systems. In the course of spectroscopic and structural research, introduction of the complete system of internal coordinates is the most tedious and time-consuming procedure, especially for

the large molecular systems. From quantum chemistry we usually obtain force constant matrix in Cartesian coordinates. Therefore we have proposed the procedure [17] to scale ab initio force field matrix in Cartesian coordinates and to avoid introducing internal coordinates on a stage of correction of quantum mechanical force constant matrix.

Within this approach the scaling is still given by Eqs. (4, 5); however, we are not assuming the matrix B to be diagonal. Force field matrix in Cartesian coordinates is not automatically independent of the molecular position and orientation as in a case of using internal coordinates. Physically meaningful force constant matrix should therefore satisfy a number of constraints that would eliminate translational and vibrational degrees of freedom in the expression for the potential energy.

Let the force field matrix in Cartesian coordinates is represented as an array of 3×3 submatrices corresponding to each atom:

$$F = \begin{pmatrix} f_{(11)} & f_{(12)} & \cdots & f_{(1N)} \\ f_{(21)} & f_{(22)} & \cdots & f_{(2N)} \\ \cdots & \cdots & \cdots & \cdots \\ f_{(N1)} & f_{(N2)} & \cdots & f_{(NN)} \end{pmatrix} \tag{7}$$

where N is the number of atoms in a molecule. Independence of potential energy of the translations and rotation of a molecule as a whole leads to the following requirements which were introduced in [17]:

$$\sum_{i=1}^{N} f_{(ij)} = 0, \quad \sum_{i=1}^{N} V_i f_{(ij)} = 0, \quad j = 1, 2, \ldots, N \tag{8}$$

where 3×3 submatrices V_i are defined as

$$V_i = \begin{pmatrix} 0 & -Z_i^0 & Y_i^0 \\ Z_i^0 & 0 & -X_i^0 \\ -Y_i^0 & X_i^0 & 0 \end{pmatrix}$$

and X_i^0, Y_i^0, Z_i^0 are Cartesian components of the i-th atom equilibrium position.

Constraints (8) reduce the rank of matrix F to $3N - 6$ (or $3N - 5$ for linear molecules), thus leaving only vibrational degrees of freedom.

When scaling procedure is applied to the matrix F in Cartesian coordinates, we may assume that a priori matrix F^0 satisfies the requirements (8). However, this does not necessarily mean that the scaled matrix also satisfies these requirements. To ensure that scaled matrix also contains only vibrational degrees of freedom, the scale matrix B should also satisfy certain conditions as it was shown in [17]:

1. Matrix B consists of the 3×3 unit submatrices multiplied by certain factors β_{ij} $(i, j = 1, \ldots, N)$ (similarly to force field matrix in Cartesian coordinates):

$$B = \begin{pmatrix} \beta_{11}E & \beta_{12}E & \cdots & \beta_{1N}E \\ \beta_{21}E & \beta_{22}E & \cdots & \beta_{2N}E \\ \cdots & \cdots & \cdots & \cdots \\ \beta_{N1}E & \beta_{N2}E & \cdots & \beta_{NN}E \end{pmatrix}$$

2. The factors β_{ij} are subject to the following constraints:

$$\beta_{ij} = \beta_{ji},$$

$$\sum_{i=1}^{N} \beta_{1i} = \sum_{i=1}^{N} \beta_{2i} = \cdots = \sum_{i=1}^{N} \beta_{Ni} = S = const. \tag{9}$$

Conditions (8) permit matrix B to be diagonal only for all β_{ii} equal to each other. If any extra constraints exist due to the symmetry or model assumptions, they should be used in addition to (8). In general, matrix B consists of $N(N-1)/2+1$ independent parameters, since all diagonal elements may be represented as

$$\beta_{ii} = S - \sum_{j \neq i} \beta_{ij}.$$

On this way one come to the formulation of inverse vibrational problem in a form (4) where a set of a priori constraints D on the molecular force field includes conditions (8). The solution (a set of scaling factors) can be found by minimization of functional (6). Additionally, a set D can include the constraints such as equality of some off-diagonal factors to zero, in-pair equalities of factors, symmetry constraints etc.

4   Generalized inverse structural problem

The accumulation of data on molecular constants helps one to predict spectra and other properties of compounds not yet investigated and assists the development of physical models in a theory of molecular structure. On this way a method has been proposed to integrate gas phase electron diffraction, vibrational spectroscopy and modern theoretical techniques into coherent computational environment [18].

The numerical framework underlying this method essentially represented the first example of an effective solution for the general inverse problem of structural chemistry (IPSC) by imposing suitable constraints for feasible structure and force field "space" introduced on the basis of theory of non-linear ill-posed problems.

Tested with precise diffraction and spectroscopic frequencies data on benzene, the developed treatment was aimed at avoiding ill-conditioning in attainment of physically plausible solution for structure and dynamics of systems undergoing small deformation of nuclear frame as compared with overall molecular dimensions. The numerical scheme permitted calculations to be consistently carried out in terms of equilibrium molecular geometry and force field. This integrated procedure is provided a mutually augmentative relationship between various methods so that the effect of systematic errors inherent in each of the individual techniques in isolation could be revealed so offering an excellent opportunity for critical comparison of data from various methods.

More complicated procedure has been proposed [19] for solving the generalized inverse structural problem (GISP) in the case of joint treatment of the experimental data obtained by different physical methods (vibrational spectroscopy, electron diffraction (ED) data and microwave (MW) spectroscopy).

Implementation of an integrated procedure for joint refinement of the force field and equilibrium geometry is based on the dynamic model of a molecule. Based on the general approximation of small vibrations, the model has been extended to include cubic and quartic anharmonic potential terms for proper description of the large-amplitude motion. Within this approach, a molecule is described using a set of equilibrium geometry parameters R, and a molecular force field F represented by the matrices of quadratic, cubic and possibly quartic force constants defined in the framework of a certain nonlinear system of internal coordinates. Both parameter sets R and F can be considered as finite-dimensional vectors.

The model is used to predict experimentally measured values, such as vibrational frequencies ω, electron diffraction intensity $M(s)$, rotational constants A, B, C obtained from microwave molecular spectra, etc. All of these values are functions of geometric (R) and force field (F) parameters. With experimental data and parameters represented as elements of normalized finite-dimensional spaces, we can formulate the problem of simultaneous refinement of the force field and equilibrium geometry of the molecule as a system of non-linear equations

$$\begin{cases} \omega(F, R) = \omega_{exp}, \\ M(s, F, R) = M_{exp}(s), \\ \{A, B, C\}(R, F) = \{A, B, C\}_{exp} \end{cases} \tag{10}$$

on a set of predefined constraints $F \in D_F$, $R \in D_R$. This system can be extended to include additional experimental evidence when available (for example, data for isotopic species of a molecule sharing the same force field and equilibrium geometry).

Due to experimental errors, lack of experimental data and model limitations, this system of equations (that can be also treated as a finite-dimensional non-linear operator equation) usually fails to define unique solution, often proves to be incompatible and does not provide stability with respect to the errors of input data. To avoid these unfavorable features characteristic to the ill-posed problems, it is necessary to implement a regularizing algorithm for its solution. This approach based on optimization of the Tikhonov's functional in the next form:

$$\begin{aligned} M^{\alpha}(F, R) = \|\omega(F, R) - \omega_{exp}\|^2 + \|M(s, F, R) - M_{exp}(s)\|^2 + \\ + \|\{A, B, C\}(F, R) - \{A, B, C\}_{exp}\|^2 + \alpha\{\|F - F^0\|^2 + \|R - R^0\|^2\} \end{aligned} \tag{11}$$

where in the last ("stabilizer") term F^0 and R^0 represent parameters of *ab initio* force field and equilibrium geometry, respectively. With the appropriate choice of regularization parameter α (that depends on the experimental errors characterized by some numerical parameter δ), it proves possible to obtain approximations converging to a normal pseudosolution of the system (10) when experimental errors tend to zero [2]. These approximations are obtained as extremals $\{F_\alpha, R_\alpha\}$ of functional (12).

The complexity of molecular models used in the analysis strongly depends on the availability and quality of the experimental data. Mostly, the molecular models are based on the assumption of small harmonic vibrations. In some cases of solving GISP within the scaling approach, it is necessary to include the cubic part of the force field [19]. In order to get a set of more reliable cubic force constants it is necessary to improve empirically the *ab initio* values by means of the Pulay harmonic scale factors.

Two schemes of cubic constant scaling are were used in our investigations [19,20]. The *ab initio* quadratic force constant f_{ij}^0 defined in internal coordinates were scaled as follows:

$$f_{ij}(\text{scaled}) = f_{ij}^0 \beta_i^{1/2} \beta_j^{1/2}$$

where β_i and β_j are the harmonic scale factors. Then, similarly, the cubic constants scaling mode can be formulated [20] as

$$f_{ijk}(\text{scaled}) = f_{ijk}^0 \beta_i^{1/2} \beta_j^{1/2} \beta_k^{1/2}$$

or, alternatively,

$$f_{ijk}(\text{scaled}) = f_{ijk}^0 \beta_i^{1/3} \beta_j^{1/3} \beta_k^{1/3}$$

where f_{ijk}^0 are the unscaled theoretical cubic constants. Both scaling schemes result in reduction of the dimension of vibrational problem and, correspondingly, in determination of a much smaller number of parameters. The examples of the applying the last procedure to different molecular systems including those with large amplitude motion are given in [21–23].

5   Molecular systems with large-amplitude motion

New perspectives were open when the previously developed integrated algorithm for joint treatment of gas-phase electron diffraction and vibrational spectroscopic data has been extended to include systems with large-amplitude oscillatory motion [19–23].

This motion, often being the central feature of molecular structure, comprises such phenomena as, for example, hindered internal rotation, angle bending, pseudorotation in cyclic molecules, inversion, ring puckering, etc. statement is different from earlier attempts to yield solution for large-amplitude motion dynamics introduced the one-dimensional approach when molecular vibrations were assumed to be frozen and only a single large-amplitude degree of freedom plus overall rotation were considered. The dynamic models defined in this manner have included two rigid parts (referred to as top and frame) allowed to rotate with respect to each other around a fixed axis. Nevertheless, these very simple models, proved to be rather successful in attaining quantitative agreement between calculated and observed vibrational spectra for molecules with internal rotation, inversion and angle bending.

The main problems in diffraction analysis of molecules with large-amplitude motion are complexity of such motion and difficulties connected with separation between the large- and small-amplitude motions. In our investigations [21] the approach has been realized based on the integrated analysis: the expansion of potential energy is used to describe non-linear vibrational distortions of a molecule. This approach with the expansion of the potential energy function in terms of normal coordinates is best suited for the integrated analysis provided this expansion meets the requirement to describe non-linear vibrational distortions of a molecule. In pursuit of numerical technique to execute various vibrational averages the perturbation theory including terms through the third-order was chosen. As in [18], the analysis of data from experimental sources is guided by quantum mechanical molecular geometry and force field optimization results and, additionally, extended with a number of advances. The suitable scale factors

can be introduced and adjusted by the least-squares refinement to ensure the best fit between calculated and observed spectroscopic or diffraction evidence. It allowed to moderate the effect of systematic errors in the computed force constants matrix. The treatment is augmented by the inclusion of microwave rotational constants which can yield substantial support to the structural analysis. Centrifugal distortion corrections to interatomic distances caused by the rotational motion of a molecule are included. These corrections would eliminate the source of trouble in the analysis of diffraction data collected at elevated temperatures. Finally, the standard deviations of the parameters determined and the corresponding correlation coefficients can be estimated.

An elementary treatment of separability for large-amplitude motion(s) ignores rotation-vibration coupling, that exists between the large-amplitude and other vibrational motions or between two or more large-amplitude motions and usually is based on exploiting Fourier expansion for large-amplitude motion potential combined with an approximate expression for the kinetic energy containing a constant reduced mass. The average over large-amplitude motion is carried out by a classical distribution function with a constant pre-exponential factor. However, this function is not correct unless the large-amplitude motion is truly rectilinear which generally is not the case. More realistic representation is based on the assumption that the atoms take curvilinear paths in a large-amplitude motion.

For large-size molecules, there exists an alternative approach to use more complicated multi-dimensional models accounting for various vibrational interactions. It was generally assumed that a molecule may involve a single large-amplitude degree of freedom represented by a structural parameter selected as an independent variable. The other structural parameters, rather than remaining fixed, were allowed to vary as certain functions of the chosen variable.

Since the complexity of large-amplitude motion prohibits a comprehensive quantum mechanical treatment for polyatomic molecules, various approximating models are required to interpret diffraction and spectroscopic data. In conventional electron diffraction studies of systems which provide the main concern in this study the ambition of reliable barrier heights determination in large measure seems illusive so it was thought tempting to follow the well-known simplification of the general vibrational problem which results from the use of the normal coordinates yielding a simple solution for the harmonic approximation involving all degrees of freedom. Thus a model of normal coordinates can be considered as a suitable technique for molecules exerting large-amplitude oscillatory motion if the set of anharmonicity parameters is not becoming intractably large and an appropriate numerical treatment for direct proceeding from assumed model properties to diffraction and spectroscopic observables is applied.

In the **GF** matrix treatment of harmonic force field calculations by equation (1) it is customary to define $3N - 6$ linearized internal displacement coordinates. In a case of molecules with large amplitude motion the potential function must be set up in terms of the true instantaneous values of internal coordinates rather than their projections on the equilibrium positions. Such an approach is essentially an attempt to include non-linear effects into description of vibrational distortions of a molecule avoiding introducing any additional adjustable parameters.

The curvilinear distance corrections have been introduced advantageously in our

study [18] where a simplified and computationally easy procedure of the non-linear transformation from any kind of internal coordinates to normal coordinates was described. It was found that this effect lead to a marked decrease of average internuclear distances and finally resulted in higher values for the equilibrium distances. As expected, the curvilinear effect was found to be most conspicuously pronounced for the $C - H$ equilibrium distance owing to fairly large protonic excursions in the benzene molecule.

One more important factor significantly affecting non-linear vibrational distortions of a molecule is related to dynamic or intrinsic (over and above that from nonlinear coordinate transformations) anharmonicity. This is because the potential energy curve has skewed rather than symmetrical minima or these may be symmetrically deformed by quartic or higher even-order terms. Second, when too little is known about the potential function the Morse-like oscillator approach for covalent bonds may give helpful insight into the anharmonic portion of bond-stretching vibration. This idea has been used profitably in our work [18] for appropriate calculation of anharmonic bond distance corrections.

In the structural analysis the rotational motion of a molecule induces the mean positions of the component atoms to shift away from the center of gravity of the molecule (centrifugal stretching or centrifugal distortion effect). A novel expression more suitable for the purposes of the present analysis and providing fairly accurate estimates for distance stretching corrections have been proposed in our publication [18].

The quantum mechanical Hamiltonian for a vibrating rotor can be written as follows

$$\hat{H} = \frac{1}{2}\sum_s (\hat{P}_s^2 + \omega_s^2 \hat{Q}_s^2) + \frac{1}{2}\hat{\mathbf{M}}^* \boldsymbol{\sigma} \hat{\mathbf{M}} \tag{12}$$

where Q_s is a normal coordinate, $\hat{P}_s$ - the momentum conjugate to Q_s, ω_s - the s-th normal frequency, $\mathbf{M}$ - a vector of the rotational angular momentum. The elements of the inverse matrix $\boldsymbol{\sigma}$ are defined as

$$\sigma_{\alpha\beta}^{-1} = I_{\alpha\beta}(Q) - \sum_{klm} \zeta_{kl}^{\alpha} \zeta_{km}^{\beta} Q_l Q_m$$

where ζ are the Coriolis coupling constants, $\mathbf{I}(Q)$ is the inertia tensor. Here the subscript s denotes the s-th vibrational mode. The Hamiltonian (12) neglecting anharmonicity terms and terms involving mixed rotational and vibrational angular momenta essentially amounts to partial separability of rotational motion from molecular vibrations.

Since the inertia tensor of a molecule depends on normal coordinates, the rotational energy of a molecule can be expanded in terms of these coordinates. Then, neglecting in this expansion all terms other than the linear one, the average over rotational motion can be carried out to yield the effective potential energy function as

$$U_{eff}(Q) = \frac{1}{2}\sum_s \left[\omega_s^2 Q_s^2 - \frac{1}{2}\sum_\alpha \frac{\langle M_\alpha^2 \rangle}{I_{\alpha\alpha}^2} \frac{\partial I_{\alpha\alpha}}{\partial Q_s}\bigg|_{eq} Q_s + \dots \right] \tag{13}$$

where $\alpha = \{x, y, z\}$ are the Cartesian coordinates, $I_{\alpha\alpha}$ stands for the diagonal elements of the inertia tensor, M_α denotes the Cartesian components of a rotational momentum,

and the derivatives in the second term on the right hand side should be taken at the equilibrium configuration.

Solution of the vibrational problem with the potential function (13) would yield non-zero equilibrium values for the normal coordinates. If it is presumed that the spacings of the rotational energy levels are sufficiently small as compared with kT (k is the Boltzmann constant), then, except for the case of fairly low temperatures, the classical law of equipartition of energy can be used to calculate the average value of the angular momentum leading to

$$\left\langle M_\alpha^2 \right\rangle = KTI_{\alpha\alpha}^{(eq)} \tag{14}$$

where diagonal components of the inertia tensor are taken at the equilibrium configuration. Then it can be shown that the following equation holds with a good approximation for the mean value of the normal coordinate

$$\left\langle Q_s \right\rangle_{rot} = \frac{kT}{2\omega_s^2} \sum_\alpha \frac{1}{I_{\alpha\alpha}^{(eq)}} \left. \frac{\partial I_{\alpha\alpha}}{\partial Q_s} \right|_{eq}. \tag{15}$$

The centrifugal stretch correction is given by transformation of all $\left\langle Q_s \right\rangle_{rot}$ values into the average shifts of individual interatomic distances. Since averaging over all possible rotations can be supposed not to cause the distortion which alters the molecular symmetry, the average values given by Eq. (4) for non-totally symmetric species must all vanish.

7 Numerical calculations

The general procedure for numerical calculations developed on the basis of methods described above is realized in our computer program which was outlined in previous paper [1] in the schematic diagram of data processing design. The input information consists of spectroscopic frequencies of vibration, electron diffraction intensities and microwave rotational constants. These data are fed into a series of sub-programs leading ultimately to a regularized least-squares determination of geometrical and force field parameters. As a prerequisite of the analysis the *ab initio* optimization results are included to produce plausible trial model for molecular geometry and force field and suitable constraining stabilizer. Further restrictions, if necessary, are guessed or transferred from related molecules. The sequence of operations in a compiled package is organized in such a manner that output data from one sub-program serve as input data for the next. All expressions used in electron diffraction, vibrational and microwave spectroscopy analyses are computed as rigorously as is presently possible. Provision is also made for the possibility of improving methods of calculation in any block of the program.

To start operation with the present software package description of the internal coordinates should be given and the Cartesian coordinates and masses of all constituent atoms be introduced. To calculate molecular scattering intensities, vibrational frequencies and rotational constants several consecutive steps are necessary to follow.

1. Given the availability of the harmonic force field, the vibrational problem is formulated and solved. The harmonic eigenvalues and eigenvectors are found. The non-linear transformation is made from the internal coordinates to the normal coordinates Q.

2. The vibrational Hamiltonian of a molecule is set up as a conventional series expansion in terms of Q (this expansion is limited by terms of third-order in Q).

3. The average normal coordinate moments are expressed in terms of the potential function coefficients entering the Hamiltonian of the previous item the necessary equations being derived by the first-order perturbation theory.

4. The derivatives of the inertia tensor with respect to normal coordinates are evaluated. Eigenvalues of the inverse inertia tensor the elements of which are averaged over normal coordinates then yield the rotational constants.

5. The normal coordinate moments are transformed into those of internal coordinates. From the latter the cumulants are calculated.

6. The reduced molecular intensity curve is calculated (the definition and explicit expressions of all relevant quantities and further calculation details is performed in [18]).

Finally, the Tikhonov functional is generated and comparison between trial model and experiment is made.

The procedures 1-6 are recycled until the best self-consistent structure and dynamic molecular model that is in agreement with all input gas phase data. If necessary, the program allows to treat data from any individual experimental technique. In spite of the fact that even primary application of the present software package has met with encouraging success, this package is not considered as a final issue and improvements of several blocks of the program aiming at reducing the difficulty and cost of the analysis are permanently under way.

Fast growing computational resources and numerical methods leads to the great advantage of modern methods of quantum chemistry for the solving many problems of structural chemistry in application to the large molecular systems such as biological objects, polymers, giant aggregates etc. But there are existed obvious severe limitations of using purely *ab initio* methods for the analysis of molecular systems consisting of a few hundred atoms. This situation and stimulates the development of special approaches for the describing the physicochemical properties such as molecule geometry, vibrational frequencies, and thermodynamic functions in cases where such systems are organized from separate smaller size units. The most successful approaches for their analysis are as a rule based on the joint use of theoretical results (e.g. *ab initio* or density functional theory (DFT)) data obtained for these chosen unites) with some empirical approach (e.g. molecular mechanics) and in many cases result in good descriptions of investigated systems.

The main approach used in molecular force field calculations is based on the so-called transferability of force constants in a series of related compounds. The schematic

illustration of such model is performed in Fig.1. Suppose we have calculated regularized force constant matrices (RQMFF) of two molecules: *trans*- $CF_3CH_2CH_2Cl$ and $CCl_3CH_2CH_2Cl$, which were found as the nearest to corresponding matrices optimized at the same model chemistry (at one level of theory, e.g. within the 2-nd order perturbation theory MP2/6-31G**). We can construct the matrix F of $CF_3CH_2CCl_3$ by joining the regularized force constants of two substituted methyl groups (CF_3 and CCl_3) and CH_2-group.
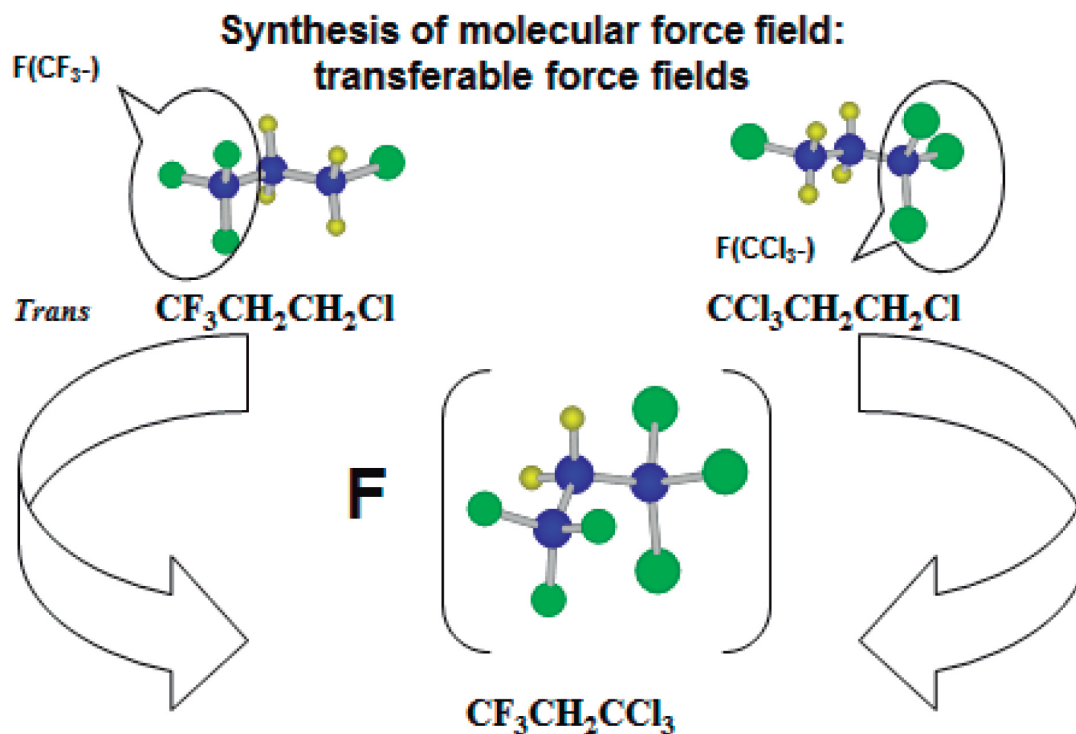
Figure 1: Scheme of synthesis of the unknown $CF_3CH_2CCl_3$ force constant matrix from previously calculated regularized force constant matrices of related *trans-* $CF_3CH_2CH_2Cl$ and $CCl_3CH_2CH_2Cl$ molecules.

Similar approach has been used for the prediction of vibrational spectrum of fullerene C_{240} [24] (Fig. 2).

The molecular geometry of C_{240} was optimized at the B3LYP/6-31G level of theory. The equilibrium configuration of the icosahedral (I_h) symmetry molecule C_{240} is completely defined by five bond lengths (Fig. 3).

Calculation of vibrational spectrum of this molecule has been performed with the list of internal coordinates included 360 bond-stretching and 720 bond angles coordinates. Altogether 1080 redundant coordinates were introduced, only 714 of them being independent. Internal coordinates were optimized automatically with the help of special utility in the SPECTRUM program package. The list of 90 different force constants of C_{240} was composed from regularized quantum mechanical force fields of fullerene, C_{60}, and corannulene, $C_{20}H_{10}$, (obtained with B3LYP/6-31G** results for
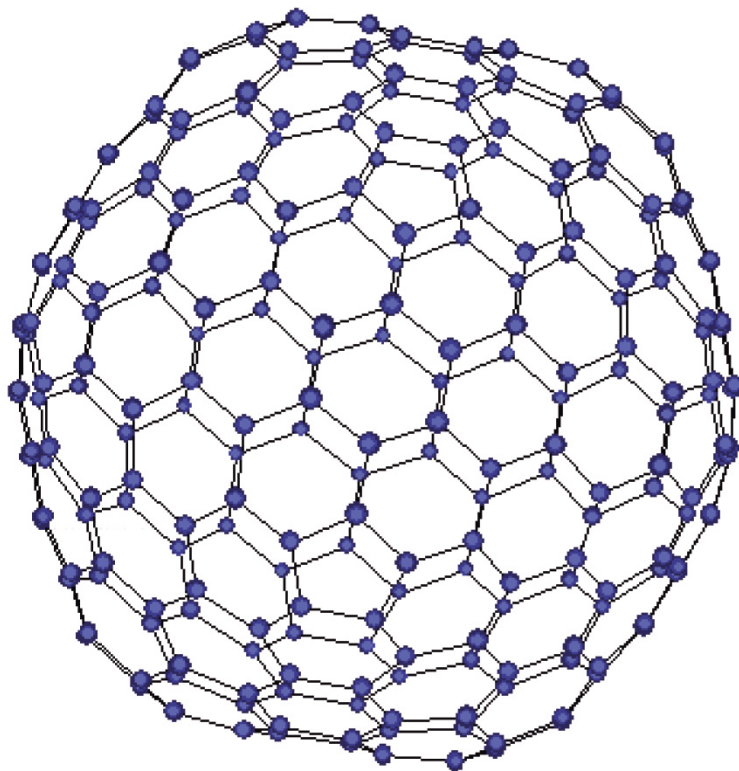
Figure 2: Schematic illustration of C_{240} molecule.

both model molecules) and was extended by certain model assumptions on intraball forces on a base of theoretical results.

The 714 normal vibrations of C_{240} are distributed by irreducible representations as

$$7A_g + 5A_u + 16F_{1g} + 18F_{1u} + 17F_{2g} + 19F_{2u} + 24G_g + 24G_u + 31H_g + 29H_u$$

The symmetry properties allow one to reduce the complete force constant matrix of C_{240} into 10 blocks with orders varying from 6 to 48 in redundant symmetry coordinates. Symmetry coordinates were run by means of the SYMM program included into the SPECTRUM program package. The vibrational density plots (a distribution of calculated frequencies by a wavenumber scale) for fullerene C_{240} are presented in Fig. 4. There are two plots for vibrations with different inversion symmetry (g and u), one referring to the total number of frequencies active in the Raman spectrum, while the other to frequencies active in the infrared absorption spectrum.

These frequencies can be used for the calculation of the thermodynamic functions of C_{240} in the wide temperature region.

The similar additive model is successful in a case of molecular systems such as supramolecules, polymers, biological molecules, etc. But in a case of large molecules, the introduction of internal coordinates is a rather tedious procedure due to the high
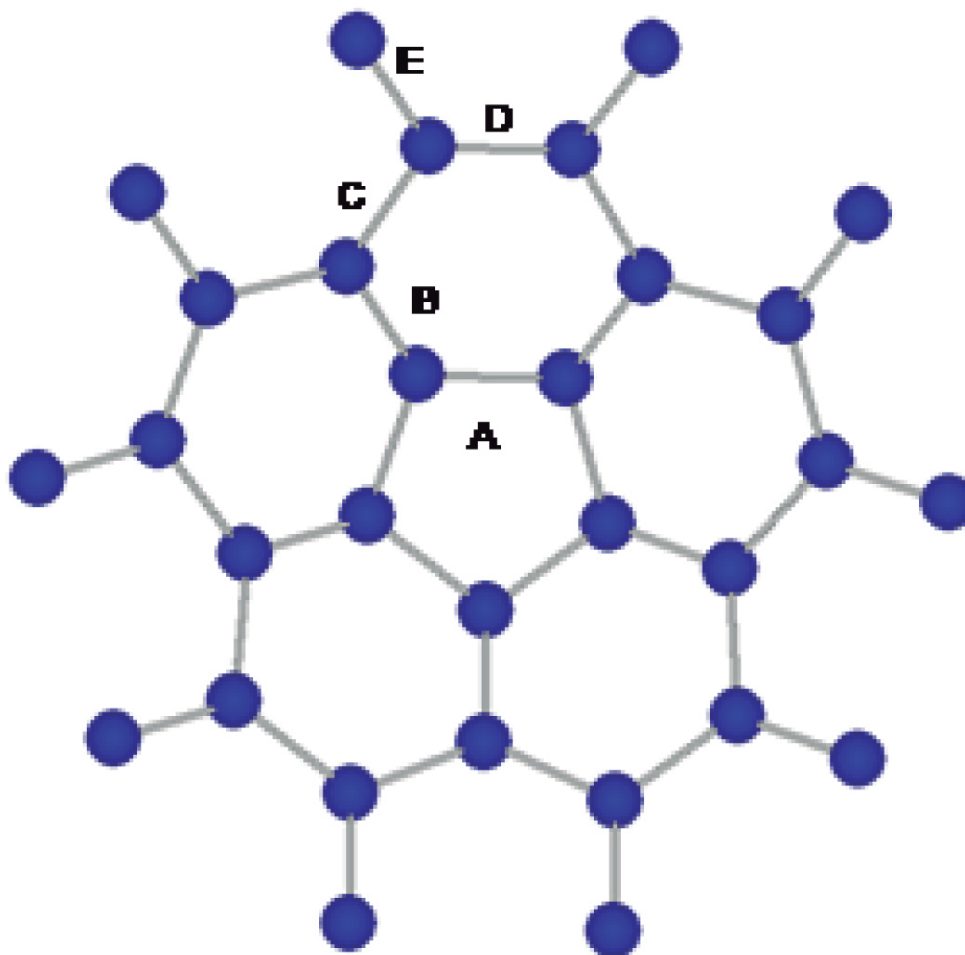
Figure 3: Five types of bond lengths in C_{240}.

dimension of arising mathematical problems and can be a source of possible erroneous calculations. As it was discussed earlier, the dimension of inverse vibrational problem can be significantly diminished in a case of working in a space of Cartesian coordinates which are obviously more convenient in the case of large molecules. The aforesaid procedure of solving inverse vibrational problem expressed by scaling parameters in Cartesian coordinates is a promising procedure which allows to get reasonable results for large molecular systems. Guanine (Fig. 5) is a purine base, one of the four main nucleobases in the DNA and RNA. Due to intermolecular hydrogen bonding between four molecules guanine can form G-quartets (Fig. 5) which are involved in sequences called G-quadruplex structures (Fig. 5). The basic principle in calculations vibrational spectra and force fields of large molecules is finding bridge between the simple molecules and more complex ones.
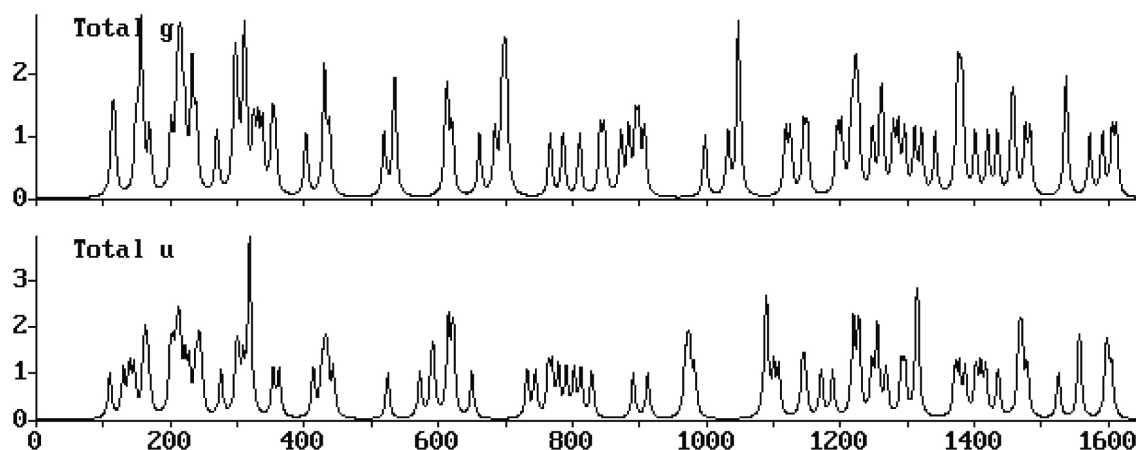
Figure 4: Vibrational state density for C_{240} molecule in the infrared absorption region between 100 and 1650 cm $^{-1}$.

Molecule of guanine consists of 16 atoms and, correspondingly, has 42 normal vibrations. 64-atomic G-quartet is characterized by 186 normal vibrations . Matrix of force constants in internal coordinates of guanine includes 903 elements while the number of force constants in internal coordinates of G-quartet is increased up to 17891. In practice many elements of force constants in internal coordinates corresponding to the interactions of far removed atoms can be fixed as equal to zero. Nevertheless the real number of different force constants remains very large.

In a case of using Cartesian coordinates the number of coordinates is equal to 48 and 192 for guanine and G-quartet, accordingly. The number of scaling factors in Cartesian coordinates for guanine molecule can be reduced to 24 by fixing some off-diagonal elements (corresponding interactions of far removed atoms) equal to zero. The set of 24 regularized scaling factors in Cartesian coordinate for B3LYP/6-31+G** level of theory was determined from solving inverse problem (4) by minimization of (6) using experimental frequencies from [25]. The solution has been obtained for $\alpha =\sim 2.0 \cdot 10^{-3}, \rho =\sim 9.7 \cdot 10^{-2}$, average errors in frequency fitting, ρ , were equal to $= 16.7$ cm $^{-1}$. This set can be used for the correction of theoretical matrices of force constants for G-quartet, stacked structures and G-quadroplex in accordance with a scheme presented in Fig. 5.

Many examples of practical calculations carried out within the schemes described above are presented in Refs [26–33].
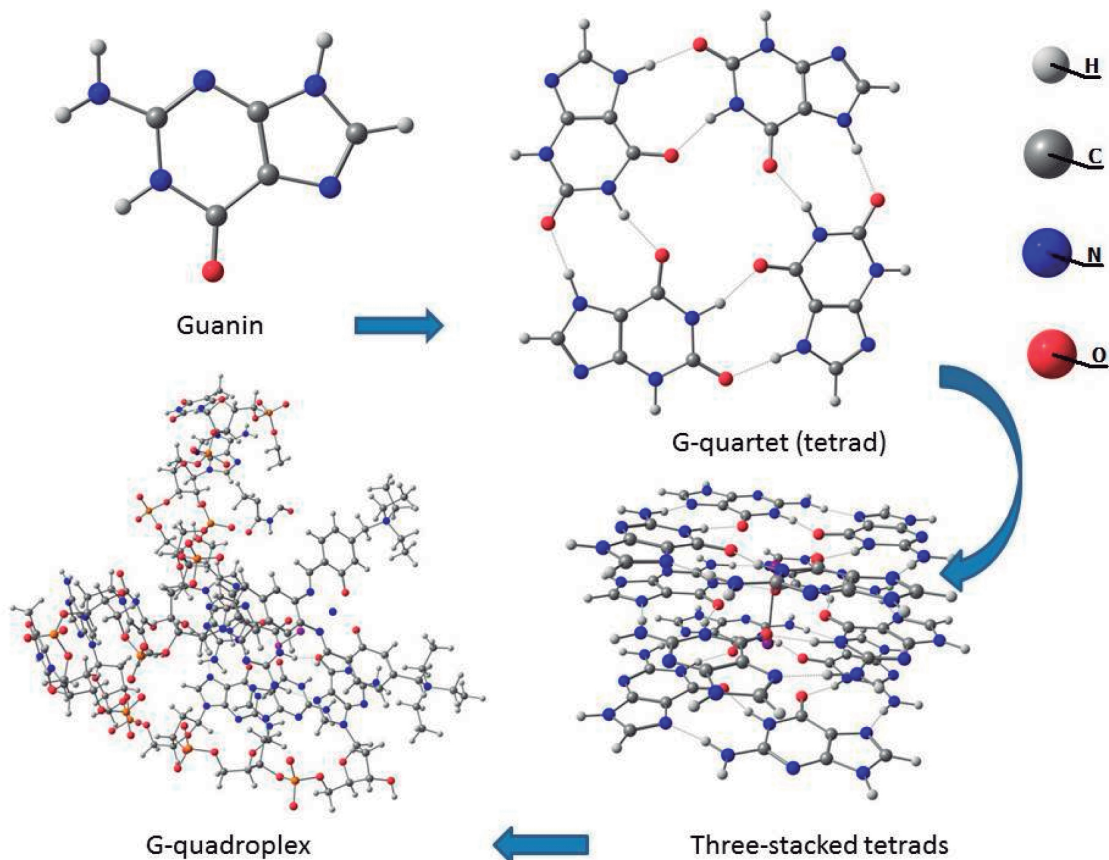
Acknowledgements

Figure 5: Principal scheme of calculations of molecular force fields and vibrational spectra of large molecular systems: guanine and its derivatives.

References

[1] G.M. Kuramshina, A.G. Yagola, *Regularization algorithms for solving nonlinear ill-posed problems of vibrational spectroscopy,*Eurasian Journal of Mathematical and Computer Applications, Vol. 4, no. 4 (2016), pp. 14-36.

[2] I.V. Kochikov, G.M. Kuramshina, I.A. Pentin, A.G. Yagola, *Regularizing algorithm of the inverse vibrational problem solution*, Doklady Akademii Nauk SSSR, Vol. 261, no. 5 (1981), pp. 1104-1106.

[3] I.V. Kochikov, G.M. Kuramshina, A.G. Yagola, *Stable numerical methods of solving certain inverse problems of vibrational spectroscopy*, USSR Computational Mathematics and Mathematical Physics, Vol. 27, no. 11-12 (1987), pp. 33-40.

[4] G.M. Kuramshina, F. Weinhold, I.V. Kochikov, A.G. Yagola, A.G., Yu.A. Pentin, *Joint treatment of ab initio and experimental data in molecular force field calculations with Tikhonov's method of regularization*, Journal of Chemical Physics, Vol. 100, no. 2 (1994), pp. 1414-1424.

[5] G.M. Kuramshina, F. Weinhold, I.V. Kochikov, Yu.A. Pentin, A.G. Yagola, *Regularized force molecule fields based on nonempirical quantum chemical calculations*, Zhurnal Fizicheskoi Khimii, Vol. 68, no. 3 (1994), pp. 401-414.

[6] Yu.A. Pentin, G.M. Kuramshina, *Molecular force fields obtained using the nonempirical stabilizer of the regularizing functional*, Russian Journal of Structural Chemistry, Vol 36, no. 2 (1995), pp. 180-190.

[7] I.V. Kochikov, G.M. Kuramshina, I.A. Pentin, A.G. Yagola, *Regularizing algorithms for molecular-force field calculations*, Journal of Molecular Structure, Vol. 272 (1992), pp. 13-33.

[8] G.M. Kuramshina, A.G. Yagola, *A priori constraints in the force field calculations of polyatomic molecules*, Journal of Structural Chemistry, Vol. 38, no. 2 (1997), pp. 181-194.

[9] G.M. Kuramshina, F.Weinhold, *Constraints on the values of force constants for molecular force field models based on ab initio calculations*, Journal of Molecular Structure, Vol. 410-411 (1997), pp. 457-462.

[10] G. Fogarashi, P. Pulay, *Ab initio calculation of force fields and vibrational spectra*, In: Durig, J.R. (ed.): Vibrational spectra and structure. Elsevier, Amsterdam, Vol. 1, (1985) pp. 125- 217.

[11] I.V. Kochikov, G.M. Kuramshina, A.V. Stepanova, A.G. Yagola, *Regularized method of scaling factors for calculations of molecular force fields*, Moscow University Physics, Bulletine, no. 5(1997), pp. 21-25.

[12] A.N.Tikhonov, A.V.Goncharsky, V.V.Stepanov, A.G.Yagola, *Numerical methods for the solution of ill-posed problems*, Kluwer Academic Publishers, Dordrecht, 1995.

[13] A.N.Tikhonov, A.S.Leonov, A.G.Yagola, *Nonlinear ill-posed problems*, V.1, 2. - Chapman and Hall, London, 1998.

[14] A.G. Yagola, *Choice of the regularization parameter using the generalized discrepancy principle*, Doklady Akademii Nauk SSSR, Vol. 245 (1979), pp. 37-39.

[15] A.V. Goncharsky, A.S. Leonov, A.G. Yagola, *The discrepancy principle for solving nonlinear ill-posed problems*, Soviet. Math. Doklady, Vol. 15 (1974), pp. 166-168.

[16] I.V. Kochikov, G.M. Kuramshina., A.V. Stepanova, A.G. Yagola, *Numerical aspects of the calculation of scaling factors from experimental data*, Numerical methods and programming, Vol. 5 (2004), pp. 281-290.

[17] I.V.Kochikov,G.M.Kuramshina,A.V.Stepanova, *New approach for the correction of Ab initio molecular force fields in PЎartesian coordinates*, International Journal of Quantum Chemistry, Vol. 109, no. 1 (2009), pp. 28-33.

[18] I.V. Kochikov, Yu.I. Tarasov, G.M. Kuramshina, V.P. Spiridonov, A.G. Yagola, T.G. Strand, *Regularizing algorithm for determination of equilibrium geometry and harmonic force field of free molecules from joint use of electron diffraction, vibrational spectroscopy and ab initio data with application to benzene*, Journal of Molecular Structure, Vol. 445 (1998), pp. 243-258.

[19] Kochikov, Yu.I. Tarasov, V.P. Spiridonov, G.M. Kuramshina, A.G. Yagola, A.S. Saakjan, M.V. Popik, S. Samdal, *Extension of a regularizing algorithm for the determination of equilibrium geometry and force field of free molecules from joint use of electron diffraction, molecular spectroscopy and ab initio data on systems with large-amplitude oscillatory motion*, Journal of Molecular Structure, Vol. 485-486 (1999), pp. 421-443.

[20] I.V. Kochikov, Yu.I. Tarasov, V.P. Spiridonov, G.M. Kuramshina, A.S. Saakjan, A.G. Yagola, *The use of ab initio anharmonic force fields in experimental studies of equilibrium molecular geometry*, Journal of Molecular Structure, Vol. 550-551 (2000), pp. 429-438.

[21] I.V. Kochikov, Yu. I. Tarasov, V. P. Spiridonov, G. M. Kuramshina, D. W. H. Rankin, A. S.Saakjan, A. G.Yagola, *The equilibrium structure of thiophene by the combined use of electron diffraction, vibrational spectroscopy and microwave spectroscopy guided by theoretical calculations*, Journal of Molecular Structure, Vol. 567-568 (2001), pp. 29-40.

[22] L.S. Khaikin, I.V. Kochikov, D.S. Tihonov, O.E. Grikina, *Analysis of Electron Diffraction Data for Several Symmetric Coordinates of Large-Amplitude Motions in the Case of the 1,3,5-Trinitrobenzene Molecule*, Russian Journal of Physical Chemistry, Vol. 89, no 6 (2015), pp. 1033-1040.

[23] I.V. Kochikov, D.M.Kovtun, Yu. I. Tarasov, *Electron diffraction analysis for the molecules with degenerate large amplitude motions: Intramolecular dynamics in arsenic pentafluoride* , Journal of Molecular Structure, Vol. 1132 (2017), pp. 139-148.

[24] G. Kuramshina, I. Kochikov, A. Yagola, *Stable Numerical Methods of Approaching Quantum Mechanical Molecular Force Fields to Experimental Data*, In: Beilina L. (eds) Applied Inverse Problems. Springer Proceedings in Mathematics & Statistics, vol 48. Springer, New York, NY, Vol. 48, (2013), pp.165-181.

[25] R. Pedro Lopes, M. Paula M. Marques, Rosendo Valero, John Tomkinson, Lu'is A. E. Batista de Carvalho, *Guanine: A Combined Study Using Vibrational Spectroscopy and Theoretical Methods*, Spectroscopy: An International Journal. Vol. 27, no 5-6 (2012), pp. 273вЪ"292.

[26] I.V. Kochikov, A.G. Yagola, G.M. Kuramshina, V.M. Kovba, Yu.A. Pentin, *Force fields and mean amplitudes of vibration of chromium, molybdenum and tungsten oxotetrafluorides*, Journal of Molecular Structure, Vol. 106, no BS (1984), p. 355-360.

[27] I.V. Kochikov, A.G. Yagola, G.M. Kuramshina, V.M. Kovba, Y.A. Pentin, *Calculation of force fields of chromium, molybdenum and tungsten hexafluorides and dioxofluorides by means of Tikhonov's regularization method*, Spectrochimica Acta, Ser. A., Vol. 41, no. 1/2 (1985), pp.185-189.

[28] G.M. Kuramshina, F. Weinhold, Yu.A. Pentin, *Ab initio and regularized force fields of haloethanes:*$CH_3CH_2Cl, CH_3CHCl_2, CH_3CF_2Cl,$ *and* CH_3CFCl_2, Journal of Chemical Physics, Vol. 109, no 9 (1998), pp. 7286-7299.

[29] G.M. Kuramshina, Yu.A. Pentin, *DFT force fields, vibrational spectra and potential functions for hindered internal rotation of* $CF_3CH_2CH_2Cl$ *and* $CCl_3CH_2CH_2Cl$, Journal of Molecular Structure, Vol. 481 (1999), pp. 161-168 (1999).

[30] G.M. Kuramshina, Yu.A. Pentin, A.G. Yagola, *New approaches for molecular conformer force field analysis in combination with ab initio results*, Journal of Molecular Structure, Vol. 509 (1999), pp. 255-263.

[31] I.V. Kochikov, Y.I. Tarasov, V.P. Spiridonov, G.M. Kuramshina, A.S. Saakjan, A.G. Yagola, *The use of ab initio anharmonic force fields in experimental studies of equilibrium molecular geometry*, Journal of Molecular Structure, Vol. 550-551 (2000), pp. 429-438.

[32] M.I.Shilina, V.M.Senyavin, G.M.Kuramshina, V.V.Smirnov, Yu.A.Pentin, *Vibrational spectra, conformational stability, vibrational assignment, and ab initio/DFT calculations for 1-nitropropane*, Struct. Chem., 2003, Vol 14, No. 6, pp. 557-571.

[33] G.M. Kuramshina, O.A. Vakula, N.I. Vakula, A.G. Mazhuga, V.M. Senyavin, Leonid G. Gorb, Jerzy Leszczynski, *Conformations, Vibrational Spectra and Force Field of Methyl -2-(2'-pyridyl)benzimidazole: Experimental data and density functional theory investigation in comparison with 2-(2'-pyridyl)benzimidazole*, Structural Chemistry, Vol. 27, no 1 (2016), pp. 209-219.

Gulnara M. Kuramshina
Department of Physical Chemistry
Faculty of Chemistry
Lomonosov Moscow State University
Moscow 119991, Russia
Fax: (7)(495)932-8846
Email: `kuramshi@phys.chem.msu.ru`

Anatoly G. Yagola
Department of Mathematics
Faculty of Physics
Lomonosov Moscow State University
Moscow 119991, Russia
Fax: (7)(495)932-8820
Email: `yagola@physics.msu.ru`