Physical Chemistry

Computer search for energy-consuming organic compounds with desired physicochemical parameters

V. A. Shlyapochnikov, B. B. Smirnov, and A. V. Evtushenko*

N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences,
47 Leninsky prosp., 117913 Moscow, Russian Federation.
Fax: 007 (095) 135 5328. E-mail: 1428@suearn2.bitnet

A new method for computer searching for energy-consuming organic compounds (EOC) with desired physicochemical properties has been developed. The method is based on a specific multilevel system of efficient parametric and "quasi-structural" selection criteria specially oriented to the search for EOC. This method makes it possible to perform the search in the area of practically significant empirical formulae (not less than 15 atoms).

Key words: energy-consuming organic compounds, target parameter; computer search; selection of chemical formulae, generator of structures.

Using computers to search for new compounds with desired properties based on the quantitative modeling of the compounds is one of the most important applied aspects of the general chemical problem of the interrelation between the structure and properties of a compound. Work in this area, including some energy-consuming organic compounds (EOC),[1–7] began rather long ago, and some success has been already achieved. However, in the area of EOC, the main efforts have been directed to the development of computer methods for the calculation of various physicochemical parameters for particular classes of compounds rather than to using computers to search for particular compounds with desired properties. The experimental search for new energy-consuming structures requires great time and material expenses and rarely leads to noticeable results. Therefore, there is a real need to apply computer search methods such as those used, for example, in the area of physiologically active compounds and drugs, to EOC. In this work, a method for a computer search for EOC that takes into account the specific nature of these compounds has been suggested.

Specific nature of EOC as objects of a computer search

The ladder of combinatorial objects studied in detail from axiomatic positions[8] includes two main steps corresponding to two types of chemical combinatorial objects: empirical formulae (EF) and structural formulae (SF). The authors of Ref. 8 pointed out the possibility of including new steps that depend on specific properties of the particular search problem, which can almost always be found in practice. At the corresponding stage of the search, formulae of each type are selected by using these properties. These properties (parameters) will be called *target parameters* (TP).

In this work, the main TP is one of the most important characteristics of EOC, the detonation rate (D). A significant feature of the parameter D is that its value is determined to a great extent by the elemental

---

Translated from *Izvestiya Akademii Nauk. Seriya Khimicheskaya*, No. 9, pp. 2168—2176, September, 1996.

composition of the compound, which makes it possible to select EOC even at the EF level. It should be kept in mind that this specific feature of *D* is not unique and is inherent in many physicochemical parameters, which is evidenced by the great number of additive methods for their calculation (*e.g.*, density (ρ), enthalpy of formation ($\Delta_f H^\circ$), boiling point, *etc.*) based on elemental composition, sometimes with corrections for structural peculiarities of molecules and sometimes without them. This fact allows one to use values of physicochemical parameters as selection criteria (filters) for EOC at all stages of the search.

For clarity, the search area in this work is restricted to C,H,N,O-containing substances. This does not imply that the compound must contain each of these elements, although at least one stoichiometric coefficient (SC) in the EF $C_{x_1}H_{x_2}N_{x_3}O_{x_4}$ should not be equal to zero. In principle, F, Cl, B, Al, S, and other elements can be included with the elements indicated.

Normed compositions. Empirical tetrahedron. "Target parameter—normed composition" diagrams

Each coefficient x_i in the EF $C_{x_1}H_{x_2}N_{x_3}O_{x_4}$ can take any nonnegative integer value. This eliminates the exhaustive selection of EF and makes it necessary to select chemical formulae, which are less informative than EF, at the beginning of the search.

We will use a procedure that has been previously introduced into thermodynamics[9] and metallurgy[10] and has found wide application in materials technology. For this purpose, let us transform the EF $C_{x_1}H_{x_2}N_{x_3}O_{x_4}$ into the form $C_{\gamma_1}H_{\gamma_2}N_{\gamma_3}O_{\gamma_4}$, where

$$\gamma_i = 100 \cdot x_i/\Sigma x_i. \qquad (1)$$

Formulae of this type will be called *normed compositions* (NC). For all NC, the following equations are evident:

$$\gamma_i/\gamma_j = x_i/x_j, \qquad (2)$$

$$0 \le \gamma_i \le 100, \qquad (3)$$

$$\Sigma\gamma_i = 100. \qquad (4)$$

Equation (2) means that the same NC corresponds to any two *similar* EF $C_{x_1}H_{x_2}N_{x_3}O_{x_4}$ and $C_{y_1}H_{y_2}N_{y_3}O_{y_4}$. Equations (1), (3), and (4) mean that each index γ_i represents the relative content of the corresponding element expressed in atomic per cents. Thus, any NC determines an infinite family of similar EF.

It is evident that NC are inappropriate for the exhaustive selection as EF, because they form a set of the same power as an EF set. However, Eqs. (3) and (5) make it possible to use the method mentioned above. They provide a way to represent the whole set of NC to the final geometric object, which, in our case, is a regular tetrahedron 100 in height, hereinafter called an *empirical tetrahedron* (ET).

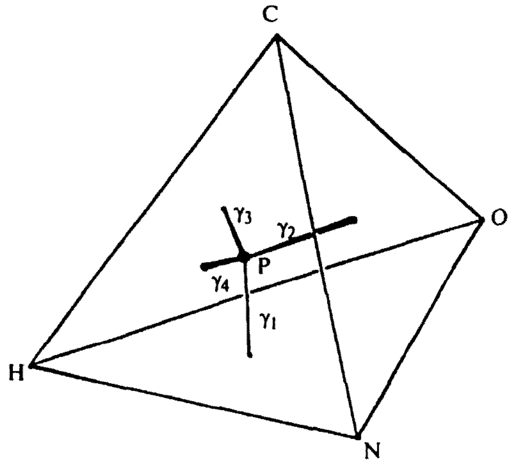


Fig. 1. Empirical tetrahedron. Point *P* corresponds to the normed composition $C_{\gamma_1}H_{\gamma_2}N_{\gamma_3}O_{\gamma_4}$. Reversely: the normed composition $C_{\gamma_1}H_{\gamma_2}N_{\gamma_3}O_{\gamma_4}$ corresponds to the point *P*.

It is known that the sum of the distances from any point of a regular tetrahedron to each of its facets is constant and is equal to the tetrahedron height. Therefore, if the ET vertices correspond to the elements C, H, N, and O, any point of the ET corresponds to one and only one NC (Fig. 1). Each γ_i is equal to the distance from a given point to the facet lying opposite the corresponding vertex. And conversely one and only one point of the ET corresponds to any NC.

Then, using the calculation methods, let us plot "target parameter—normed composition" diagrams (in the given case, *D*—NC) analogous to the constitution diagrams used in materials technology. For this purpose, let us rotate the ET points by a specified increment and calculate the *D* values for the corresponding NC. A known method[11—13] was used for calculating *D*, in terms of which *D* is a continuous function of all SC, ρ, and $\Delta_f H^\circ$. At the same ρ and the same $\Delta_f H^\circ$, the values of *D* determined by this method coincide for any two substances with similar EF, *i.e.*, they are sensitive only to relative, but not to absolute values of SC. For the determination of ρ, the method of calculation of this value based on NC, which gives the same results for similar EF,[14,15] was also used. Although this method is very limitedly appropriate for the evaluation of the density of particular compounds, its accuracy is quite enough for the rough determination of tendencies of changes in ρ depending on the relative content of different elements in the compound, which is required at this stage of the search. Since the calculation of $\Delta_f H^\circ$ based only on elemental composition is impossible in principle, the ET points were calculated at different fixed values of $\Delta_f H^\circ$ from −500 to 1500 kcal kg^{-1} and with an increment of 100 kcal kg^{-1}. The chosen region of the change in $\Delta_f H^\circ$ easily overlaps the region of the real values of this parameter for EOC, and this increment makes it possible to determine the tendency of *D* to change for fixed NC at various $\Delta_f H^\circ$.

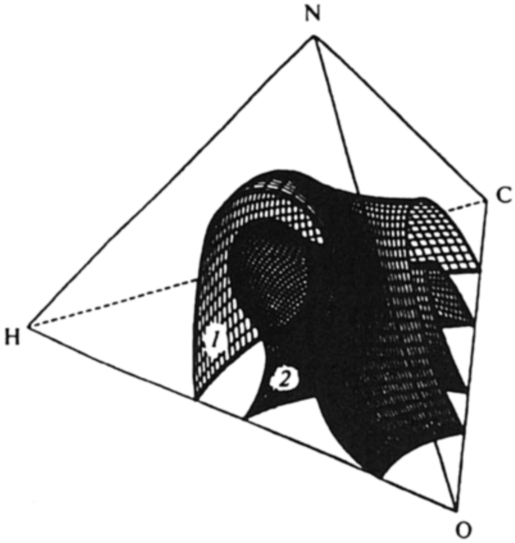


Fig. 2. Surfaces of equal D values at $\Delta_f H^\circ = 0$: D/km s^{-1} = 7 (*1*) and 9 (*2*).

The increment $\Delta\gamma_i = 1$ for all i was accepted for the calculation of the ET points. Larger increments gave too few points for plotting diagrams by interpolation. The expense of computer time increased at smaller increments, and D varied from point to point within the experimental error.

It is evident that the increment $\Delta\gamma_i$ should be chosen in such a way that it divides 100 evenly, thus the number (N_p) of ET points for which TP was calculated is the following[16]

$$N_p = C_{3+100/\Delta\gamma_i}^{100/\Delta\gamma_i} = \frac{(3+100/\Delta\gamma_i)!}{3!(100/\Delta\gamma_i)!} \qquad (5)$$

and at $\Delta\gamma_i = 1$, $N_p = 176851$.

The TP—NC diagrams plotted by the data of these calculations are families of surfaces, each of which is formed by the ET points corresponding to some constant value of TP (Fig. 2). These diagrams demonstrate the general tendencies of the change in TP depending on NC at a fixed value of $\Delta_f H^\circ$.

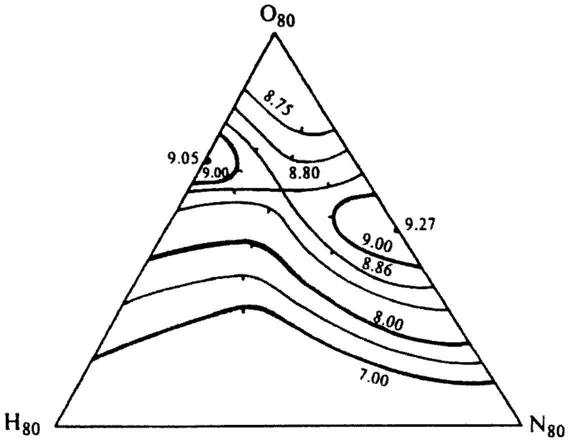


Fig. 3. Lines of equal values of D/km s^{-1} in the ET cross section corresponding to the 20% content of C ($\gamma_1 = 20$) at $\Delta_f H^\circ = 100$ kcal kg^{-1}.

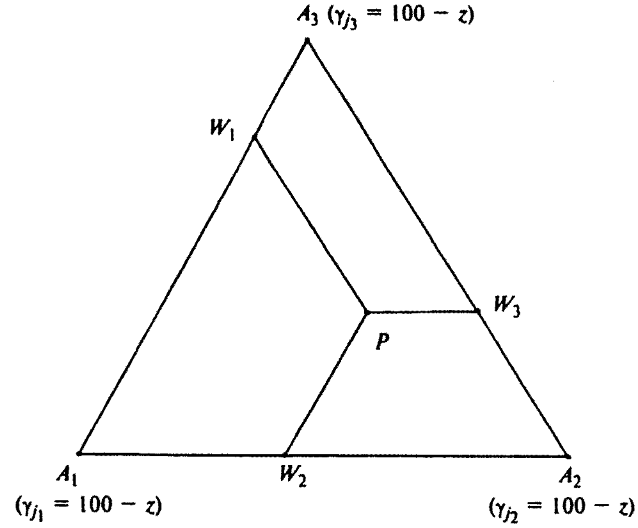


Fig. 4. Determination of coefficients γ_{j_k} (k = 1, 2, 3; j_k = 1, 2, 3, 4) for the NC corresponding to the point P in the ET cross section corresponding to the fixed values of the coefficient $\gamma_i = z$ = const ($i \neq j_1, j_2, j_3$):

$$A_1A_2 = A_2A_3 = A_3A_1 = 100 - z;$$
$$PW_1 \parallel A_2A_3;\ PW_2 \parallel A_3A_1;\ PW_3 \parallel A_1A_2;$$
$$\gamma_{j_1} = A_3W_1;\ \gamma_{j_2} = A_1W_2;\ \gamma_{j_3} = A_2W_3.$$

To determine the particular NC of desired compounds, the cross sections of the ET in the planes corresponding to the constant relative content of one of the elements were plotted (Fig. 3). The edge of the ET, but not its height, is accepted to be 100. These diagrams have the form of topographic maps. Any γ_i for each point of this diagram can be easily determined, as is shown in Fig. 4.

Thus, at a given stage of the search, a computer is used for plotting TP—NC diagrams rather than for selecting particular NC. Diagrams analogous to those presented in Fig. 2 serve as demonstration material, and those presented in Fig. 3 are tools for determining NC of the desired compounds. Working with diagrams of the second type is entirely similar to working with topographic maps.[17] Normed compositions of desired compounds are determined as points of horizontals corresponding to the demanded TP values. Obviously, in the general case, the same TP value corresponds to different NC, which provides considerable freedom of choice for the user. The factors limiting this freedom are manifested only at the next stage of the search (selection of EF).

When the relative elemental composition is expressed not in per cents, but in some other fractions, a different edge of the ET should be used. For example, a stereoscopic image of a unit ET, on which several tens of points correspond to known C,H,N,O-containing EOC, has been obtained previously.[18]

One more fact should be mentioned. At any $\Delta\gamma_i$, the values of TP are calculated, of course, only for rational ET points (at integer $\Delta\bar{\gamma}_i$ for integer points only). When TP—NC diagrams are plotted (see Fig. 3), each horizontal in each cross section of the ET is drawn, according to these calculations, as a continuous curve through *any* points of the cross section plane. The latter circumstance cannot lead to errors in the search, because all TP are described[11–15] by continuous functions of all γ_i, and the set of the points corresponding to these NC is everywhere compact. It is noteworthy that an increase in the number of possible elements composing EOC will not create additional difficulties, although a geometric object with dimensionality higher than 3 will be required for mapping the set of NC.

Empirical formulae

The normed composition corresponding to the desired values of TP determines only the atomic proportions of the elements in the desired compound at a specified value of $\Delta_f H^\circ$. This information is insufficient for determining the possibility of the existence of at least one compound with an appropriate value of $\Delta_f H^\circ$ in terms of the given NC (the TP—NC diagrams are plotted at arbitrary $\Delta_f H^\circ$), and for determining the suitability of the corresponding EF for further processing.

At the stage of the selection of the EF corresponding to the chosen NC, the $\Delta_f H^\circ$ value still cannot be calculated and is specified by convention (more exactly, the $\Delta_f H^\circ$ value at which the NC was selected is used). However, for some TP, for example, for ρ,[14,15] there are calculation methods sensitive to absolute values of SC and more exact than those used at the previous stage. This makes it possible to make parametric filters more "rigid," thus narrowing the ranges of acceptable values of the parameters calculated.

In addition to parameteric filters, "quasi-structural" filters can also be included at this stage. Their action is presented schematically in Fig. 5 using as an example the EF with the fixed sum of SC (SSC; hereinafter the expression SSC is used instead of the "number of atoms in the molecule"), $\Sigma x_i = 12$, and the fixed SC at the carbon atom $x_1 = 2$. It is evident that, unlike NC, the empirical formulae of C,H,N,O-containing compounds correspond not to arbitrary points of the continuous finite ET, but to the nodes of the infinite regular tetrahedral network with a unit distance between any two adjacent nodes. The regular tetrahedral segment with the edge equal to 12 containing the plane corresponding to $x_1 = 2$ (see Fig. 5, *a*) was chosen in our example.

The first and obvious "quasi-structural" filter is the evenness of the sum $\Sigma x_i v_i$, where v_i is the formal valence of the *i*th element (see Fig. 5, *b*).

The next is the connectedness of the molecular graph (MG), *i.e.*, in the given case, *the formal possibility of designing at least one SF of an individual compound based on the specified EF.* This filter is based on the cyclicity index:[19]

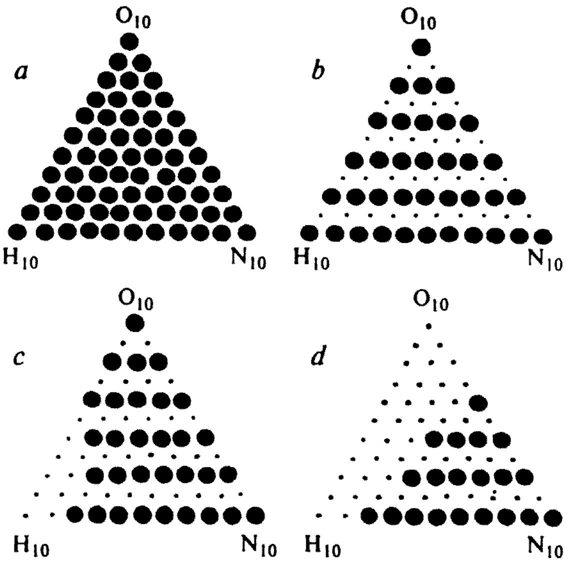
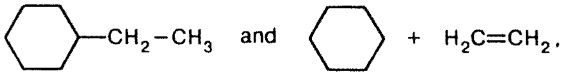


Fig. 5. Subsequent exclusion of empirical formulae (*a*) by evenness (*b*), connectedness (*c*), and active oxygen (*d*) filters at the sum of SC $\Sigma x_i = 12$ and $x_1 = 2$ (points and circles indicate arrested and passed EF, respectively).

$$\Gamma = 1 + \Sigma x_i(v_i - 2)/2, \qquad (6)$$

where Γ is the number of independent cycles of the corresponding SF, which is known in graph theory as the *cyclomatic number.*[20] Multiple bonds also relate formally to cycles: a double bond is a degenerate cycle and a triple bond is two conjugated degenerate cycles. There is an analogy between Eq. (6) and the Euler formula for simple polyhedrons,[21] on the basis of which this filter has appeared.

A necessary condition of connectedness is the nonnegativity of Γ. This means that only at $\Gamma \geq 0$ *can at least one SF of an individual compound be designed* (*connected MG*). Then the number of cycles and (or) multiple bonds is equal to Γ. This, however, does not imply that in the given case all possible EF are exhausted by individual compounds (*i.e.*, by connected MG). Let us consider, for example, the EF C_8H_{16} ($\Gamma = 1$), which can correspond to at least two structures:



The first variant is the SF of an individual compound (connected MG) of the specified EF, and the second structure is two structural formulae of two individual compounds (unconnected MG), for which only the sums of the corresponding SC satisfy the specified EF. In the case of, *e.g.*, the EF C_8H_{20} ($\Gamma = -1$), only variants of the second type are possible, and none of the variants of the first type is possible. These EF are excluded by the connectedness filter (see Fig. 5, *c*) from further consideration.

It is likely that the evenness and connectedness filters do not even exhaust all possible "quasi-structural" restrictions of a formal character that can follow directly from EF. The search for these restrictions is a separate problem of independent interest. However, in addition to formal restrictions of a general character, various "quasi-structural" restrictions following from the specificity of the particular search problem should be taken into account. In the case of C,H,N,O-containing EOC, the "active oxygen" rule is quite appropriate. This means that there must be at least enough N atoms to link them and only them with all O atoms, which is expressed by the inequality $x_3 \geq x_4/2$. The action of this filter is shown in Fig. 5, *d*.

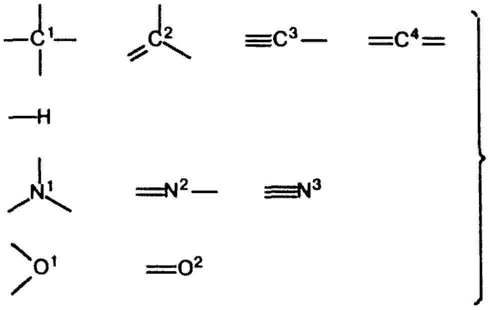
Only the EF that satisfy each of the "quasi-structural" filters reach the parameteric filters. The latter are the intervals of acceptable values of TP calculated on the basis of EF. The usefulness of the inclusion of "quasi-structural" filters in the search procedure at the stage of the selection of EF is demonstrated clearly in Fig. 5.

Expanded empirical formulae

The selection of SF based on expanded empirical formulae is the most important stage of the search (after choosing EF that satisfy the desired values of TP). Several unsuccessful attempts at a direct transition to the design and selection of SF based on the selected EF presented very clearly *a combinatorial explosion*, *i.e.*, a cumulative increase in the number of SF as the SSC increases. To avoid the combinatorial explosion, we used a known method (see, *e.g.*, Refs. 8 and 22) that involves the use of an intermediate stage, at which intermediate formulae are selected between the selection of EF and selection of SF. These formulae must satisfy the following conditions.

1. In terms of the specified EF, the number of intermediate formulae (IF) must be less than that of SF so that the computer search for IF should occur in the appropriate computer time, and the number of IF should not require much time for manual analysis.
2. A number of SF satisfying the preceding condition must correspond to each IF selected.
3. All TP that can be calculated on the basis of NC and EF should also be calculated (and with at least the same or higher accuracy) on the basis of the IF.

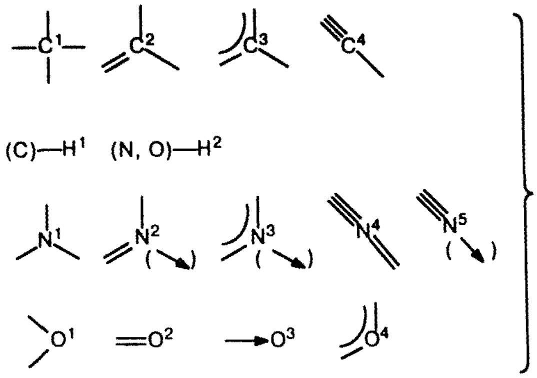
Expanded empirical formulae[23] (EEF) turned out to be the most appropriate type of IF for EOC. The concept of EEF was based to a considerable extent on a previously suggested method[24] for the calculation of ρ based on an additional classification of atoms of each element by type depending on the types of bonds formed by these atoms with adjacent atoms and on the structural fragments containing these atoms. Each type of atom was put in correspondence with a certain contribution to the molar volume[24] (a similar approach was described in a known monograph[25]). For C, H, N, and O atoms, the following types were determined:[24]



(7)

An expanded empirical formula is an EF in which the distribution of the atoms of each element by type is reflected. For example, for the EF C_4H_8, for cyclobutane (**1**) and $H_3C—HC=CH—CH_3$ (**2**) molecules, the following EEF were determined by the given set of atomic types (7): $C^1{}_4H_8$ (**1**) and $C^1{}_2C^2{}_2H_8$ (**2**). Here, in the case of EEF **1**, cyclobutane is composed of four methylene fragments, each of which is formed by a C^1 type atom and two H atoms. In the case of EEF **2**, two C^1 type atoms and six H atoms form two methyl groups, while two C^2 atoms and two H atoms form an ethylene unit.

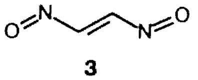
The development of methods for the calculation of ρ and Δ_fH^0 adapted to the computer search for EOC[14,15,26] required the modernization of the set of atomic types (7). The minimum average absolute errors, the maximum correlation coefficients, and the best parameters for error distribution for ρ and Δ_fH^0 were achieved when C, H, N, and O atoms were grouped as the following types:



(8)

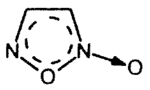
Unlike the initial set (7), in the modernized set of atomic types (8), atoms included in aromatic cycles (C^3, N^3, O^4) are separated; semipolar N→O bonds with O^3 type atoms are taken into consideration, and H atoms linked with N and O atoms are distinguished as H^2. C^4 atoms forming allene structures are excluded from consideration because they generally do not occur in EOC.

Let us describe various structural isomers using the modernized set of atomic types (8) with the EF $C_2H_2N_2O_2$ as an example. Three SF of the given EF, which correspond to three EEF (indicated near the corresponding SF), are shown below.

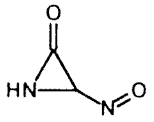


3

$C^2_2H^1_2N^2_2O^2_2$

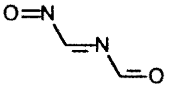


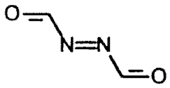
$C^3_2H^1_2N^3_2O^3_1O^4_1$

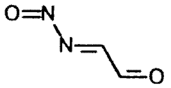


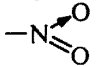
$C^1_1C^2_1H^1_1H^2_1N^1_1N^2_1$

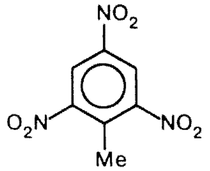
Now, using EEF **3** as an example, let us show that in the general case some array of SF corresponds to each expanded empirical formula. Three more SF corresponding to the EEF $C^2_2H^1_2N^2_2O^2_2$ (in addition to that indicated in the previous example) are presented below.



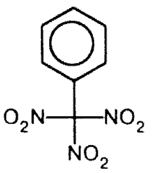


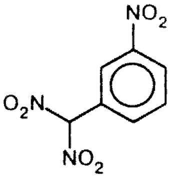


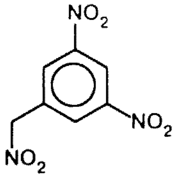
Thus, EEF make it possible to shorten the exhaustive search for combinatorial objects in the "selection of EF—selection of SF" region (the most "narrow point" of the whole search procedure) due both to their relatively small number compared to SF (remember that we deal with orders of magnitude in both cases) and to the fact that they contain direct indications of allowable (and often necessary) structural fragments of the desired SF. For example, in the case of the EF $C_7H_5N_3O_6$ (trotyl and its structural isomers), the EEF $C^1_1C^3_6H^1_5N^2_3O^2_3O^3_3$ corresponding to D = 7 km s^{-1} immediately points to the existence of a benzene ring (C^3_6) and three nitro groups ($-NO_2$, $-N^2_1O^2_1O^3_1$, or ) as necessary fragments, and methyl, mono-, di-, and trinitromethyl groups as acceptable ones. Some of the SF designed on a computer taking into account these restrictions are presented below.



Trotyl







In this case, all "structural possibilities" of EF at the specified D value are reduced to the arrangement of the same three substituents on the toluene skeleton and are exhausted by 15 variants. The working time of the program, which is a structure generator (SG), including the displaying of the results, is comparable with the duration of the initial data entry from the keyboard.

When restrictions similar to those presented above were not specified, the SG in our computer did not reach the design of the SF of trotyl (and other known EOC based on the corresponding EF), and no evidence of approaching the design of the desired SF were observed.

For any fixed EF, the number of expanded empirical formulae (N_F) in terms of the specified set of atomic types is determined exactly, and for the EF $C_{x_1}H_{x_2}N_{x_3}O_{x_4}$ it is the following:

$$N_F = C^{x_1}_{x_1+3}C^{x_2}_{x_2+1}C^{x_3}_{x_3+4}C^{x_4}_{x_4+3}. \tag{9}$$

It can be easily seen that each of the cofactors in the right part of Eq. (9) is analogous to the right part of Eq. (5). The exhaustive search for EEF is performed in the same way as that for NC. Like EF, EEF pass through "quasi-structural" and parameteric filters. The compatibility of atomic types plays the role of the "quasi-structural" filter. Let us clarify this using the EF $C_2H_5NO_2$, which has passed *a fortiori* through the evenness, connectedness, and "active oxygen" filters and has at least one SF, as an example.

$CH_3-CH_2-NO_2$ (EEF $C^1_2H^1_5N^2_1O^2_1O^3_1$).

There are other EEF appropriate for the design of SF; however, there are some EEF that involve incompatible combinations of atomic types. For example, the EEF $C^2_2H^2_5N^1_1O^4_2$. Here the O^4 type atoms in aromatic heterocycles have no corresponding C^3 and N^3 atoms (see set (8)), there are not enough N^1, N^2, and O^1 atoms for the arrangement of all H^2 type atoms, and the C^2 type atoms are unsuitable. The compatibility filter in the EEF selection program acts similarly to the "active oxygen" filter in the EF selection program, excluding EEF with a deficiency or excess of mutually complementary atoms from further consideration. Expanded empirical formulae that have passed through the compatibility filter undergo the action of parametric filters. The latter work in the same manner as at the preceding stages and transmit only the formulae for which calculated values of all TP lie within the specified intervals. The difference is that at this stage, unlike preceding stages, $\Delta_fH°$, which now is one of the TP, can also be calculated, *i.e.*, the number of parametric filters increases. In addition, the accuracy of the calculation of ρ and, hence, that of D, increases. Thus, the efficiency of the selection of chemical formulae increases from stage to stage in parallel with the progressive increase in their information content.

The selection of EEF eliminates from consideration the majority of the SF of the specified EF. None of the SF is designed in the explicit form at this stage, and if the EEF is eliminated by any of the parametric filters, *the corresponding SF will not appear further anywhere in any form*, which prevents the combinatorial explosion.

As shown above using the EF $C_7H_5N_3O_6$ as an example, the selection of EEF results in the formation of a small number of formulae (in the ideal case, one formula), which contain in the explicit form such detailed information about the desired SF that the work of the SG begins immediately from the design of the latter. No forbidden fragments are included in the initial data for the SG.

A computer search for C,H,N,O-containing EOC with a required value of the detonation rate

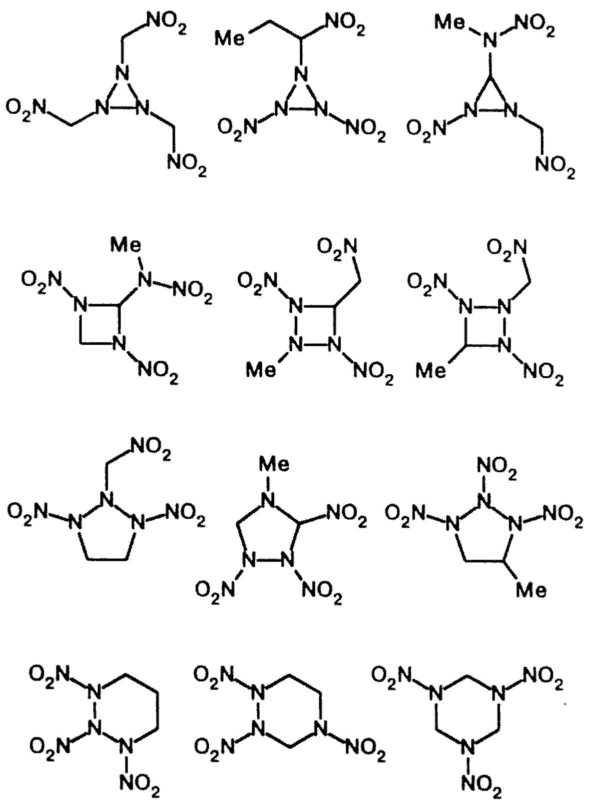
We have not discussed the reliability of the method suggested. Final conclusions can only be drawn from the results of experimental studies of new EOC found by this method. First one must be convinced that *the already known EOC are not missed* at any stage of the search, at least in terms of the actually achieved TP values. Let us show this using the following as an example.

Let us specify the value $D = 9\pm0.2$ km s^{-1}. At $\Delta_f H^o = 100$ kcal kg^{-1} on a diagram similar to that presented in Fig. 3 and corresponding to x_1 = const = 14%, let us choose the NC $C_{14}H_{28.(6)}N_{28.(6)}O_{28.(6)}$, which satisfies the specified range of D values. Among the EF approximately corresponding to this NC and passed through "quasi-structural" and parameteric filters, we choose $C_3H_6N_6O_6$ (hexogen and its structural isomers). Among the EEF of this EF passed through the compatibility filter, the one whose value is closest to $D = 9$ km s^{-1}, is $C^1{}_3H^1{}_6N^1{}_3N^2{}_3O^2{}_3O^3{}_3$, which corresponds to $D = 8.9$ km s^{-1}.

Based on the set of atomic types (8), let us determine the structural restrictions on SF following from this EEF.

1. The nine atoms $N^2{}_3O^2{}_3O^3{}_3$ indicate that there must be three nitro groups that can be linked to C^1 and N^1 atoms.
2. In the absence of H^2 and O^1 atoms, the three N^1 atoms can form bonds only with one another, with C^1 atoms, and with nitro groups.
3. The H^1 atoms can be linked only with C^1 atoms.
4. The C^1 atoms can form bonds only with one another, with H^1 and N^1 atoms, and with nitro groups.
5. If the nitro group is considered to be a univalent atom, the connectedness criterion (see Eq. (7)) is $\Gamma = 1$. This implies the necessary existence of one and only one cycle formed by C^1 and N^1 atoms and containing from 3 to 3 + 3 = 6 atoms, or of one and only one double C=C, N=N, or C=N bond involving the same atoms. The latter, however, is excluded, because the atoms of the types indicated do not form multiple bonds.

Scheme 1



Thus, the initial data for the SG includes the EF and the following information about the number of fragments in the structure: $—NO_2$, 3; $—C_3H$, from 0 to 2; $>CH_2$, from 0 to 3; $\geq CH$, from 0 to 1; $—NH_2$, 0; $>NH$, 0; $\geq N$, 3; and a three- to six-membered cycle, 1. Based on these restrictions, the SG designs 134 SF, including 72 with three-membered cycles, 46 with four-membered cycles, 13 with five-membered cycles, and 3 with six-membered cycles. Some of them are presented in Scheme 1. These are all the SF the SG designs with these restrictions. The SF with six-membered cycles include the SF of the known EOC hexogen (see Scheme 1), for which D_{exp} = 8.84 km s^{-1}.

This example, along with the example for trotyl (see above), makes it possible to draw the *preliminary* conclusion that the reliability of the method suggested is sufficient, and to begin the computer search for *new* EOC.[27]

General scheme and specific aspects of the computer search for EOC

Thus, the search for EOC with desired physicochemical properties by the method suggested is performed in four stages according to the scheme presented in Fig. 6. An *active* role of the user at each stage is implied.

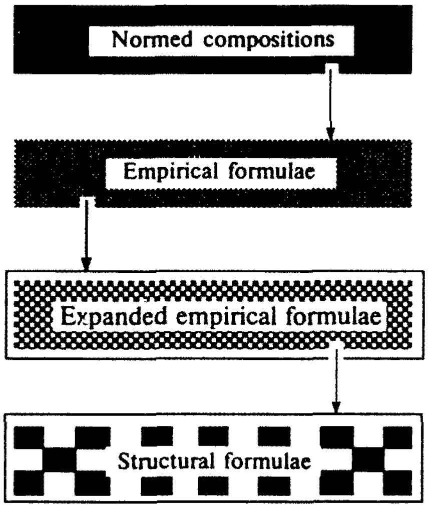


Fig. 6. General scheme of a computer search for EOC with desired physicochemical parameters.

At the first stage (selection of NC), the user provided with the necessary TP—NC diagrams can work without a computer, can use one if necessary for designing lacking programs, or can use a computer to rotate the region of the diagram in which he or she is interested, by a finer increment.

Going to the second stage (selection of EF), the user must determine the upper and lower values of the SSC for the computer search for similar EF in terms of the NC chosen or (if the search is performed without the first stage) to specify the fixed SSC. In any case, the user specifies the range of acceptable values for each TP.

At the third stage, the user specifies EF and ranges of acceptable TP values at the input of the program. At the output, the user obtains the EEF satisfying the specified TP values and, based on these EEF, determines possible and necessary fragments of desired structures. These fragments together with EF form the *initial data*, which make use of the SG efficient and, as has been shown above using trotyl and hexogen as examples, *shorten the work of the SG to the minimum*. Thus, the efficiency of the SG becomes independent of its quality, and it becomes possible to *perform the search almost independently of the number of skeleton atoms in the molecules* of the desired compounds.

It can be concluded that in the case of target parameters that can be calculated on the basis of low-information chemical formulae of the NC and EF types, the initial data that makes use of the SG efficient can be obtained by the selection of formulae whose complexity exceeds slightly that of EF, and by a procedure as simple as that for EF. The selection becomes more rigid due to the fact that not only does the number of TP that can be calculated increase from stage to stage, but also the accuracy of the corresponding calculation methods increases, *i.e.*, the increase in the number of parametric filters is accompanied by an increase in their efficiency. Only formulae that have passed through "quasi-structural" filters reach the parametric filters. In other words, the specifics of EOC in the computer search makes it possible to realize to a considerable extent the very attractive principle of ***negative*** selection.

It should be kept in mind that the results obtained at the final stage of the search according to the scheme suggested (Fig. 6) are not final. They should be processed additionally using refined TP values and the possibility of existence and synthesis. This requires an expert's evaluation and more complicated and expensive calculations (compared to additive calculations) by quantum-chemical, molecular mechanics, and other methods. However, the purpose of this work was to design a general strategy, and was restricted to the provision of methods and software for a computer search for the material for this analysis.

Software for the search procedure

For the design of the "detonation rate—normed composition" diagrams, the selection of EF, and the selection of EEF, we designed (in the same sequence) the GEPARD, IRBIS, and TIGER programs adapted to IBM-compatible computers, and designed procedures and "quasi-structural" and parametric filters of chemical formulae of the corresponding types.

The program-structure generator GIOS was used for the design of structural formulae.*

* * *

The efficiency of the suggested method for the computer search for EOC is determined to a considerable extent by their specific character. It should be emphasized that target parameters of EOC can be estimated quantitatively based on any form of description of the compound: its empirical formula, expanded empirical formula, or structural formula.

This gives grounds to believe that the method suggested can also be used for a computer search for compounds with other functions and useful properties, if these properties can be calculated on the basis of low-information chemical formulae, which reflect only the elemental composition of the substance.

The authors are grateful to Mrs. I. D. Solodukhina for help in designing the graphic material.

References

1. A. R. Martin and H. J. Yallop, *Trans. Faraday Soc.*, 1958, **54**, 257.
2. M. J. Kamlet and H. Hurwitz, *J. Chem. Phys.*, 1968, **48**, 3690.

---

* Composed and kindly presented by A. P. Pavlova and V. P. Suboch (Institute of Physics, Belorussian Academy of Sciences).[28]

3. M. J. Kamlet and C. Dickinson, *J. Chem. Phys.*, 1968, **48**, 43.
4. M. J. Kamlet and J. E. Abalard, *J. Chem. Phys.*, 1968, **48**, 36.
5. M. J. Kamlet and S. J. Jacobs, *J. Chem. Phys.*, 1968, **48**, 23.
6. I. N. Aizenshtadt, *Fiz. Goren. Vzryva* [*Combustion and Explosion Physics*], 1976, **12**, 754 (in Russian).
7. R. Rothstein and R. Petersen, *Propellants and Explosives*, 1979, No. 4, 56.
8. S. S. Trach and N. S. Zefirov, *Printsipy simmetrii i sistemnosti v khimii* [*Principles of Symmetry and Systematics in Chemistry*], Izd. MGU, Moscow, 1987, 54 (in Russian).
9. J. W. Gibbs, *Trans. Connect. Acad.*, 1873, **2**, 309.
10. N. S. Kurnakov, *Vvedenie v fiziko-khimicheskii analiz* [*Introduction to Physicochemical Analysis*], Izd. AN SSSR, Moscow—Leningrad, 1940, 564 pp. (in Russian).
11. V. I. Pepekin, M. N. Makhov, and Yu. A. Lebedev, *Dokl. Akad. Nauk SSSR*, 1977, **232**, 852 [*Dokl. Chem.*, 1977, **232** (Engl. Transl.)].
12. V. I. Pepekin, N. M. Kuznetsov, and Yu. A. Lebedev, *Dokl. Akad. Nauk SSSR*, 1977, **234**, 105 [*Dokl. Chem.*, 1977, **234** (Engl. Transl.)].
13. V. I. Pepekin and Yu. A. Lebedev, *Dokl. Akad. Nauk SSSR*, 1977, **234**, 1391 [*Dokl. Chem.*, 1977, **234** (Engl. Transl.)].
14. A. V. Evtushenko, B. B. Smirnov, and V. A. Shlyapochnikov, *Izv. Akad. Nauk, Ser. Khim.*, 1993, 685 [*Russ. Chem. Bull.*, 1993, **42**, 654 (Engl. Transl.)].
15. A. V. Evtushenko and B. B. Smirnov, *Izv. Akad. Nauk, Ser. Khim.*, 1994, 22 [*Russ. Chem. Bull.*, 1994, **43**, 13 (Engl. Transl.)].
16. I. I. Ezhov, A. V. Skorokhod, and M. I. Yadrenko, *Elementy kombinatoriki* [*Elements of Combinatorial Analysis*], Nauka, Moscow, 1977 (in Russian).
17. I. A. Bubnov, S. F. Bogatov, S. D. Dubov, A. K. Kalinin, and P. T. Savchenko, *Voennaya topografiya* [*Military Topography*], Voenizdat, Moscow, 1976, 279 pp. (in Russian).
18. J. R. Stine, *J. Energ. Mater.*, 1990, **8**, 41.
19. B. B. Smirnov, T. S. Pivina, and V. A. Shlyapochnikov, *Izv. Akad. Nauk SSSR, Ser. Khim.*, 1984, 2396 [*Bull. Acad. Sci. USSR, Div. Chem. Sci.*, 1984, **33**, 2189 (Engl. Transl.)].
20. V. N. Sachkov, *Vvedenie v kombinatornye metody diskretnoi matematiki* [*Introduction to Combinatorial Methods of Discrete Mathematics*], Nauka, Moscow, 1982, 384 pp. (in Russian).
21. R. Kurant and G. Robbins, *Chto takoe matematika?* [*What is Mathematics?*], Gostekhizdat, Moscow—Leningrad, 1947 (Russ. Transl.).
22. L. M. Masinter, N. S. Sridharan, J. Lederberg, and D. H. Smith, *J. Am. Chem. Soc.*, 1974, **96**, 7702.
23. B. B. Smirnov, V. A. Shlyapochnikov, and V. A. Tartakovsky, *Izv. Akad. Nauk SSSR, Ser. Khim.*, 1988, 1442 [*Bull. Acad. Sci. USSR, Div. Chem. Sci.*, 1988, **37**, 1275 (Engl. Transl.)].
24. A. Immirzi and B. Perini, *Acta Crystallogr.*, 1977, **A33**, 216.
25. V. M. Tatevskii, V. A. Benderskii, and S. S. Yarovoi, *Zakonomernosti i metody rascheta fiziko-khimicheskikh svoistv parafinovykh uglevodorodov* [*Regularities and Methods for Calculation of Physicochemical Properties of Paraffinic Hydrocarbons*], Gostoptekhizdat, Moscow, 1960, 114 pp. (in Russian).
26. A. V. Evtushenko and B. B. Smirnov, *Izv. Akad. Nauk, Ser. Khim.*, 1994, 1389 [*Russ. Chem. Bull.*, 1994, **43**, 1315 (Engl. Transl.)].
27. A. V. Evtushenko, B. B. Smirnov, V. A. Shlyapochnikov, and M. S. Molchanova, *Izv. Akad. Nauk, Ser. Khim.*, 1996, **45**, 330 [*Russ. Chem. Bull.*, 1996, **45**, 315 (Engl. Transl.)].
28. V. P. Suboch, *Sistema programm dlya interpretatsii molekulyarnykh spektrov* [*A System of Programs for Interpretation of Molecular Spectra*], Nauka i Tekhnika, Minsk, 1987, 24 pp. (in Russian).

Received November 16, 1994;
in revised form May 6, 1996