Tetrahedron Vol. 35, pp. 1419 to 1426
© Pergamon Press Ltd., 1979. Printed in Great Britain

0040-4020/79/0601-1419/$02.00/0

AUTOMATIC ESTIMATION OF HEATS OF ATOMIZATION AND HEATS OF REACTION

JOHANN GASTEIGER
Institut für Organische Chemie, Technische Universität München, D-8046 Garching, West Germany

(*Received in UK 30 November 1978*)

Abstract—Various additivity schemes for estimating heats of atomization have been evaluated to test their usefulness for designing a computer program. The approach is based on the Allen scheme and uses parameters obtained from thermochemical data. In addition, a method has been developed which allows the direct calculation of reaction enthalpies. Only the immediate environment of the bonds involved in the reaction has to be scanned. A program based on this method utilizes thermochemical parameters, too. Computation times with this program are extremely short and independent of the size of the molecules participating in the reaction. The procedure can be applied to a large variety of organic compounds. Examples demonstrate the accuracy of the predicted values.

The initiation and the course of a chemical reaction is determined by many factors. But none are of such fundamental importance as are the thermodynamic aspects of a reaction. Quite often the course of that reaction is taken which leads to the most stable products. This is directly the case in reactions where thermodynamic product control is prevailing. In other cases, the most stable products are obtained after passing through some intermediate structures.

Further, equilibrium constants, K, are directly related to changes in free energy and can be calculated from them:

$$K = e^{-\Delta G/RT}.$$

The changes in free energy during a reaction can be obtained from the heat of reaction, ΔH, and the change in entropy, ΔS.

$$\Delta G = \Delta H - T\Delta S.$$

Hence, heats of reaction can be used, in conjunction with entropy data, to predict whether a particular reaction is likely to give a reasonable yield of a desired product.

A knowledge of heats of reaction is therefore indispensable for a comprehensive treatment of chemical reactivity. As we are designing various models for predicting the reactivity of organic compounds we had to get a hold on heats of reaction. The most rigorous way to test and evaluate a model is to put it into algorithmic form. In designing an algorithm the premises of a model must be clearly defined to obtain a program which is exactly doing what it is supposed to. Further, once a program is written, the model can be applied to a great number of cases with ease and speed to test the validity and range of a model.

We are especially lucky in having developed a computer program which generates reactions by a formal procedure.[1] Bonds and electrons of molecules are rearranged in all mathematically possible manners. Thus, for a given set of molecules all conceivable reactions are obtained. This provides an ideal basis for testing hypotheses and models for the prediction of reactivities. A specific model is put into algorithmic form and converted into a computer program which becomes a subroutine of the main system. The main system then automatically supplies a large variety of reactions which have to be evaluated by the model. A realistic model must serve to extract the chemically feasible reactions from the collection of mathematically possible ones. It is against this background that we were searching for a method to estimate heats of reaction.

The basic approach for calculating heats of reaction is given by eqn (1):

$$\Delta H_r^\circ = \sum \Delta H_f^\circ(\text{products}) - \sum \Delta H_f^\circ(\text{reactants}). \quad (1)$$

This requires a knowledge of the heats of formation of starting materials and products of a reaction.

In principle, heats of formation can be obtained through quantum mechanical calculations. At present, such an approach is feasible for rather small systems only. The limit on the size of the molecules is somewhat depending on the method chosen and the amount of time, effort, and expense one is inclined to spend. But the requirements for storage space, computation times, and careful geometry optimization will place the limit presently at systems with about 15–20 nonhydrogen atoms for *ab initio* techniques, somehow larger for semiempirical methods. In any case, it is still prohibitive to project the use of these methods in our reaction generating system, routinely handling molecules with 30–70 atoms and generating about 4–8 reactions per sec (on an AMDAHL 470 V/6).

The calculated values for the heats of reaction should be accurate to within 1–2 kcal/mole to be chemically useful. Recall that a difference in free energy of about 2 kcal/mole will shift an equilibrium from 90:10 to 10:90 at room temperature, thus completely reversing the ratio of products. Such an accuracy cannot be reliably obtained with present-day quantum mechanical methods in the general case.

Single-determinant molecular orbital methods, which are the most widely used ones, are known to lead to poor values for the total energies of molecules. The use of limited basis sets of various sizes gives results which may be contradictory. But even at the Hartree–Fock limit neglect of electron correlation is a serious problem.

It is the hope when applying single-determinant molecular orbital theory to closed shell systems that there is a cancellation of correlation effects. In studying heats of hydrogenation Pople *et al.*[2] showed that at least a moderately extended basis set was necessary to give reasonable estimates, the mean absolute deviation still being 5.5 kcal/mole. Bond separation energies of isodesmic reactions where starting materials and products have the same number of bonds of the same type are reproduced with slightly better success,[2] the mean absolute deviation being 3.5 kcal/mole.

Semi-empirical quantum mechanical methods have deficiencies of their own. For example, the stability of small rings is overestimated in the CNDO- or MINDO-approach.

Rejecting quantum mechanical methods, molecular mechanics calculations[3] were considered. This approach was not acceptable to us as only a limited number of classes of compounds has been parameterized. We, however, aimed at a system capable of handling a wide variety of compounds. Further, computation times of force field calculations are still orders of magnitude higher than we wanted to tolerate.

Quantum mechanical and molecular mechanics calculations not having offered a solution to our problem we turned our attention to a completely empirical approach. Experimentally, heats of formation have been determined for a fair amount of compounds.[4,5] Substituting the appropriate values for starting materials and products into eqn (1) would yield values for heats of reactions. But clearly so, only for reactions where the values for all starting materials and products are known. And this requirement amounts to a very severe limitation. In comparison with the number of compounds that exist the number of experimental thermochemical data available is minute.

But thermochemical values can offer a solution! By extracting from the experimental heats of formation parameters for certain structural features and then applying them on the basis of an additivity scheme to other molecular structures predictions for heats of formation can be made. Provided that the parameterization has been made correctly and consistently the accuracy of the predictor system is predominantly dependent on the number of parameters used for a certain class of compounds. On the other hand, too many parameters make the estimation process cumbersome and slow. As our approach is based on such an additivity scheme a brief discussion is necessary.

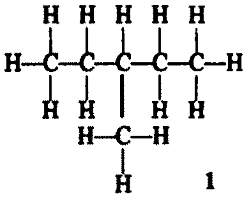
Additivity schemes for estimating heats of atomization

The basic assumption of additivity schemes for estimating heats of formation is that heats of formation can be broken down into structural contributions which are additive and can be transferred from one compound to another. In the following, we will center the discussion on heats of atomization as the values for the structural parameters are often more accessible to direct physical interpretation. The values given for the parameters refer to 298.15° and to the gas phase.

Heats of atomization can be obtained from heats of formation in conjunction with the heats of formation of gaseous atoms (eqn 2):

$$\Delta H^\circ_a(A_aB_b\ldots) = a\Delta H^\circ_f(A(g)) + b\Delta H^\circ_f(B(g)) + \ldots - \Delta H^\circ_f(A_aB_b\ldots). \qquad (2)$$

Quite early, Fajans[6] suggested to estimate heats of atomization by contributions from individual bonds. Since then, many different schemes have been proposed. To find a common basis for comparison we will center our discussion on alkanes. For alkanes, bond additivity asks for two parameters, one for a C–C bond and one for a C–H bond. As has already been recognized by Fajans,[6] bond additivity leads to deviations of several kcal/mole in the alkanes. Also, isomeric alkanes all receive the same value for their heat of atomization. From among the various schemes proposed for estimating heats of atomization of alkanes we decided to choose a scheme which uses four parameters for alkanes as it promised to offer a reasonable compromise between number of parameters necessary and accuracy thus obtained. There exist three different four-parameter schemes[7–9] which were shown[4] to be numerically equivalent. 3-Methylpentane (1) will be used as an illustration for the three different approaches.



1

The experimental heat of atomization of 1 is 1795.82 kcal/mole.[10]

In the Benson scheme[7] groups consisting of polyvalent atoms and their ligands are considered. In the above case, the heat of atomization is given by the contributions of three C–(C)(H)$_3$, two C–(C)$_2$(H)$_2$, and one C–(C)$_3$(H) groups.[12] This gives:[13]

$$3\times 337.39 + 2\times 280.0 + 1\times 223.65 = 1795.82\ \text{kcal/mole}. \qquad (3)$$

In the Laidler scheme,[8] all deviations from simple bond additivity are concentrated into the C–H bonds. In the alkanes we have three values for a C–H bond depending on whether the carbon atom is primary, secondary, or tertiary. In our example, we have nine E(C–H)$_p$, four E(C–H)$_s$, one E(C–H)$_t$ and five E(C–C),[12] resulting in[13]

$$9\times 98.44 + 4\times 97.93 + 1\times 97.44 + 5\times 84.14 = 1795.82\ \text{kcal/mole}. \qquad (4)$$

The Allen schemes[9] uses parameters for individual bonds; for alkanes, B(C–C) and B(C–H). Deviations from simple bond additivity are attributed to interactions of pairs of bonds C–C–C (G(CCC)-parameter[12]) and trios of nonbonded atoms around a central carbon atom C(C)$_3$ (D(CCC)-parameter. Interactions which involve hydrogen atoms are set to zero. The heat of atomization of 1 is obtained from 14 B(C–H), five B(C–C), five G(CCC), and one D(CCC) value:[13]

$$14\times 98.97 + 5\times 80.96 + 5\times 1.10 + 1\times(-0.06) = 1795.82\ \text{kcal/mole}. \qquad (5)$$

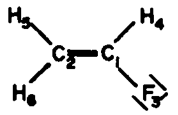
As can be seen, all three schemes are numerically equivalent, and, in this example, reproduce the experimental value fortuitously exactly. Within a set of 49 alkanes the mean absolute deviation was obtained as 0.64 kcal/mole with these parameters.[14] The basic assumptions of the Laidler or Allen scheme—to concentrate deviations from a simple bond additivity scheme into C–H bond parameters, or interactions of nonhydrogen neighbors, respectively—are theoretically unpleasing. But from thermochemical data alone there is no way of assigning the deviations from simple bond additivity to individual bonds or interactions without arbitrary assumptions. There will always be more unknowns than equations as linear dependences occur.

The generation of reactions

Before we go into the details of our approach for estimating heats of atomization and heats of reaction a brief introduction into the basic logic of our program system seems appropriate.

To manipulate chemical structures by computers a representation for molecules has to be chosen. We decided to deal in the first phase only with the constitution of molecules. The constitutional aspects of molecules are responsible for the bulk of chemical properties. For example, the energies of steric interactions are typically 50–100 times smaller than bond energies. Stereochemistry can be treated as a refinement of a constitutional description.[15] For specifying the constitution of molecules topological representations have been most commonly accepted. From among these we have chosen to represent molecules by lists of atoms and bonds. By also accounting for the free electrons on atoms all valence electrons are taken care of. This is important for the treatment of reactions where shifts between free and bonding electrons can occur. Figure 1 shows the internal representation of vinylfluoride. For each atom the atomic number and the number of free electrons is carried. For each bond the indices of the two atoms involved and the bond order are given.

In our program system[1] reactions are considered as bond breaking and making, and electron shifting processes and are generated accordingly. Starting from input molecules bonds are broken and the fragments joined together in a different manner. Also, bonds are broken and the electrons centered on a single atom (—to become free electrons—), and vice versa. Reaction categories which perform several of these basic operations simultaneously have been formulated. An example is given in Fig. 2 where two bonds are broken and two new ones made. The majority of organic reactions falls into that category.



atoms	1	2	3	4	5	6	7
atomic number	6	6	9	1	1	1	1
free electrons	0	0	6	0	0	0	0

bonds					
1. atom	1	1	1	2	2
2. atom	2	3	4	5	6
bond order	2	1	1	1	1

Fig. 1. Internal representation of vinylfluoride.

By applying this reaction scheme onto the bonds of *n*-hexane indicated in Fig. 3, 3,3-methylpentane, or ethylcyclopropane and methane are obtained as conceivable reaction products.

Applying all reaction categories, and sequences thereof, onto all possible combinations of bonds and free electrons leads to all conceivable reaction products. In the end, all ensembles of molecules with a given empirical formula are obtained.[16,17]

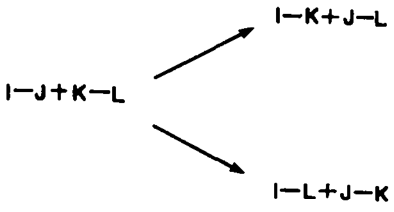


Fig. 2. Reaction category.

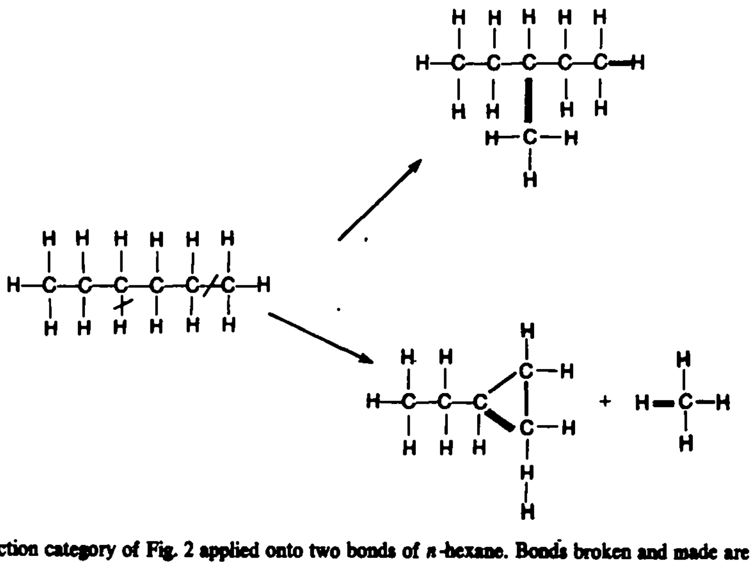


Fig. 3. Reaction category of Fig. 2 applied onto two bonds of *n*-hexane. Bonds broken and made are indicated.

To select the chemically feasible reactions from among the set of all mathematically possible ones various estimations are performed. The calculation of heats of reaction, reported below, is one of the criteria used to determine chemical reactivity.

Automatic estimation of heats of atomization

Having chosen a data structure for representing molecules which gives direct access to the atoms and bonds of molecules we are well prepared for performing the estimation of heats of atomization automatically. All three additivity schemes[7-9] discussed previously can be made a basis for computer programs for the estimation of heats of atomization. To illustrate this, let us return to the example of 3-methylpentane.

Scanning through all carbon atoms of 3-methylpentane and establishing their first neighbor atoms allows us to find all groups of the Benson scheme[7] present in the molecule. This procedure can be easily performed with our data structure. The various groups are recognized easily and with speed. Looking up the values for the group parameters in tables contained in the program and adding them up gives already the heat of atomization of the molecule.

In an approach based on the Laidler scheme[8] all bonds of a molecule have to be scanned. In addition, for C–H bonds the first neighbors of the carbon atom have to be searched to establish whether it is a primary, secondary, or tertiary C–H bond. Again, searching lists of parameters and adding the values gives the heat of atomization.

Basing the estimation on the Allen scheme[9] involves again searching all bonds to get access to the B-parameters. From the first neighbors of a bond the presence of the substructures typical for the G- and D-parameters can be deduced. After recognition of the structural features, lists where the appropriate values are contained are searched. The heat of atomization is estimated by summation of the parameters found.

In summary, all three additivity schemes can easily be made a basis for efficient computer programs for estimating heats of atomization. This is primarily the result of our data structure for representing molecules. The explicit listing of all atoms and bonds of a molecule gives fast access to the substructures on which the parameters of the three additivity schemes are based. Thus, programming the three approaches[7-9] becomes more or less straightforward.

To account for our final choice on one of these three schemes our objectives have to be defined in more detail. From the outset we were aiming at a computer program capable of estimating heats of atomization for a large variety of compounds. Therefore, situations have also to be taken into consideration where not all parameters of a given class of compounds necessary for either one of the three schemes are included in the program. The reason could be that insufficient thermochemical data are available for this class of compounds which prevents parameterization. In these cases we still wanted to get at least a rough estimate of the heats of atomization. This could be accomplished by then taking a lower-order approximation, i.e. by resorting to simple bond additivity. The basis for such an approach should nearly always be given as for practicably any type of bond at least an approximate estimate of the corresponding bond energy seems possible, be it through chemical intuition. The design of the program for estimating heats of atomization should therefore provide for an automatic switching to a simple bond additivity scheme if not sufficient parameters are available.

To make a decision which one of the three schemes to choose as a basis for the computer program further evaluation of these schemes under these premises is necessary.

The Benson method[7] has several disadvantages when taken as a basis for a computer program with the intended properties. For many classes of molecules composite group parameters can be determined only. It is the usual practice[7] to split these composite terms up into individual group parameters by making additional assumptions. But this increases the number of groups which have to be generated, included in the lists, and scanned in a program run. Thus, the number of parameters is larger than actually required by the accuracy of the approach (approximately by a factor of two). This increases storage requirements and computation times. In addition, the first member in each series of compounds (CH_4 with the alkanes, CH_3OH with the alcohols, CH_3Cl, NH_3, etc.) constitute separate groups which also have to be included in a program which is generally applicable, again increasing the number of parameters. Most severely, however, there is no straightforward way to jump into a lower-level approximation, in particular, into an additivity scheme based on contributions of bonds only. If the parameter for a certain group is not contained in a list in the program, this group and the one associated with it through a bond made or broken would have to be deleted. Then, for this bond an entirely different program based on simple bond additivity would have to be accessed. Further, for other applications, an estimate for the dissociation energy of a bond is desirable. For this problem the group method of Benson is unsuitable.

The Laidler method[8] is basically a modified bond additivity scheme. Deviations from additivity are condensed into different parameters for the various C–H bonds. Therefore, a switch to simple bond additivity should be easy to accomplish: If the parameter for a specific C–H bond is not available some standard mean value for a C–H bond can be taken. This reduces the accuracy of the approach to the one of simple bond additivity. The fact that all deviations are pressed into the C–H bond parameters must introduce errors when bond dissociation energies are desired. As in the Benson scheme, there is the disadvantage that for a sizable number of molecules parameters of their own have to be carried. This is the case for the C–H bonds of the first members of a class of compounds (e.g. CH_4, CH_3Cl, CH_3OH) and for bonds in diatomic and many triatomic molecules (e.g. CO, CO_2).

In the Allen scheme[9] we have again a modified bond additivity scheme. Deviations from simple bond additivity are taken care of by bond and atom interaction parameters. But the predominating values are introduced by the B-parameters which correspond to bond parameters. The dominance of the B-parameters is shown by comparing the four optimized parameters[13] for alkanes (values in kcal/mole): $B(C–C) = 80.96$, $B(C–H) = 98.97$, $G(CCC) = 1.10$, $D(CCC) = -0.06$. That the values of the B-parameters are close to bond energies can be seen from a comparison with the optimized values of a two-parameter scheme for alkanes[13] (values in kcal/mole): $E(C–C) = 82.97$, $E(C–H) = 98.67$. The close similarity in the values of the bond energies and of the

B-parameters justifies the use of the B-values as estimates for bond dissociation energies. Switching to simple bond additivity can be performed in a straightforward manner by only using the B-parameters for the calculation of heats of atomization. Here, the Allen scheme offers similar advantages as the Laidler scheme. But there is an additional point which makes the Allen approach for our purposes superior to the Laidler—and Benson—scheme. In the Allen method the first members of a series of compounds (CH_4, CH_3Cl, CH_3OH, NH_3, etc.) do not need parameters of their own but can be treated with the standard B-parameters. Thus, fewer parameters are needed than with the Laidler or Benson approach resulting in smaller lists and shorter search times.

Therefore, the Allen scheme has been made the basis for the development of a computer program for the estimation of heats of atomization. The procedure, as discussed above, of scanning bonds to find access to B-parameters, and of searching the direct vicinity of bonds for the substructures corresponding to the G- and D-parameters has been put into algorithms. Thus a computer program for estimating heats of atomization on the basis of thermochemical parameters has been implemented.

Automatic estimation of heats of reaction

Being able to calculate heats of atomization we are now also in the position of estimating heats of reaction. For, heats of reaction can be calculated as the differences in the heats of atomization of starting materials and products. This is shown by eqn (6) which is obtained from eqns (1) and (2).

$$\Delta H^\circ_r = \sum \Delta H^\circ_a(\text{reactants}) - \sum \Delta H^\circ_a(\text{products}). \quad (6)$$

Scanning all bonds of both starting materials and products of a reaction gives access to the number and types of B-parameters of the Allen scheme. The immediate neighbors of the bonds allow us to recognize the substructures associated with G- and D-parameters.

The procedure shall be illustrated with the isomerization of *n*-hexane to 3-methylpentane (Fig. 2). This reaction is known from experimental data to be exothermic by 1.06 kcal/mole. Recall that a simple bond additivity scheme (using two parameters, one for a C–H and one for a C–C bond) would give a value of zero for the heat of this reaction. For the starting material, *n*-hexane, the heat of atomization is evaluated according to:

$$5\,B(C\text{–}C) + 14\,B(C\text{–}H) + 4\,G(CCC) = 1794.78\ \text{kcal/mole}. \quad (7)$$

The heat of atomization of 3-methylpentane has been estimated in eqn (5) to amount to 1795.82 kcal/mole. Thus, a value of $1794.78 - 1795.82 = -1.04$ kcal/mole is calculated for the heat of reaction. This result is in excellent agreement with experiment.

This approach is conceptually simple and straightforward. But our formal treatment of reactions allows a more immediate calculation of heats of reaction. The reaction generators point to the bonds broken and made in a reaction (see marked bonds in Fig. 2). The structural changes inflicted by a reaction occur only in the immediate vicinity of these bonds. And only these structural changes have influence on the heat of reaction. Those parts of the starting materials and products further away from the reaction site conserve their structure. Therefore, these parts have the same contributions to the heat of atomization of both reactants and products and can be neglected.

Then, the heat of reaction is directly calculated as the difference in the structural contributions P to the heats of atomization around the bonds broken and made (eqn (8)).

$$\Delta H^\circ_r = \sum P\,(\text{bonds broken}) - \sum P\,(\text{bonds made}). \quad (8)$$

This approach has the effect that only the immediate vicinity of the reaction site has to be scanned. It shall be illustrated again with the isomerization of *n*-hexane to 3-methylpentane.

Considering the bonds which are broken in *n*-hexane gives the B(C–C) and B(C–H) parameters. In addition, we see that the C–C bond interacts with another C–C bond. Thus, the substructure corresponding to the G(CCC) parameter is encountered. The G(CCC) parameter has to be taken into account a second time as the carbon atom of the C–H bond is the central atom of another C–C–C substructure. This gives:

$$\begin{aligned} P\,(\text{bonds broken}) &= (B(C\text{–}C) + G(CCC)) + (B(C\text{–}H) + G(CCC)) \\ &= 80.96 + 1.10 + 98.97 + 1.10 = 182.13\ \text{kcal/mole}. \end{aligned}$$

In the reaction giving 3-methylpentane a C–C bond and a C–H bond is made. This leads to the B(C–C) and B(C–H) parameter. Further, it is found that the C–C bond interacts twice with another C–C bond and that one carbon is the central atom of a third C–C–C arrangement. Therefore, three G(CCC) parameters have to be taken into account. And finally, the C–C bond is part of the substructure of three carbon atoms bonded to a central one ($C(C)_3$). This gives access to one D(CCC) parameter.

$$\begin{aligned} P\,(\text{bonds made}) &= (B(C\text{–}C) + 3\,G(CCC) + D(CCC)) + B(C\text{–}H) \\ &= 80.96 + 3.30 - 0.06 + 98.97 = 183.17\ \text{kcal/mole}. \end{aligned}$$

Thus, an estimate of $182.13 - 183.17 = -1.04$ kcal/mole is obtained for the heat of reaction. As required, this value is in agreement with the estimate obtained by calculating the heats of atomization of the entire molecules. The more direct approach is vastly superior. Only a small number of substructures have to be established and their corresponding parameters searched in lists. In the example of the isomerization of *n*-hexane the number of parameters necessary to estimate the heat of reaction could be reduced from 48 to 10. With larger molecules the reduction in the number of parameters is even greater. This results in remarkably shorter computation times. Further, as only the reaction site has to be scanned, computation times are independent of the size of the molecules involved in the reaction. This is an extraordinary advantage not encountered in other approaches to the estimation of heats of reaction. Both in quantum mechanical calculations and with molecular mechanics methods computation times increase drastically with the number of atoms in the molecules.

The method is a predictor system calculating heats of reaction for a large number and wide variety of reactions based on few fundamental parameters. Values of heats of reaction can also be predicted for those reactions where no experimental data are available. The approach could be further sophisticated by incorporating more parameters for a given class of compounds. For example, for alkanes parameters for gauche 1.4- and for 1.5-interactions could be included. This would enhance the accuracy of the calculated values. At present, the greater simplicity of the approach outlined here is preferred as the accuracy of the results is satisfactory.

Outline of the algorithm

As already mentioned, our program system works on a topological representation of molecules. Each atom and bond of a molecule is directly accessible. The procedure for calculating the enthalpies has this representation at its disposal. If all bonds of a molecule are scanned heats of atomization are obtained. For calculating heats of reaction the procedure has to be called twice, with the starting materials and with the products. In the direct calculation of heats of reaction only the bonds broken and made in the reaction are considered. These bonds have been marked in the reaction generating part of the program system.

Each bond leads to the corresponding B-parameter contained in lists. A total of 90 different bond types for molecules made up of C, H, O, N, S, P, Si, F, Cl, Br and I atoms is considered in these lists, thus allowing us to calculate heats of atomization and heats of reaction for a large variety of compounds.

Next, the environment of the marked bonds is scanned and the substructures consisting of two adjacent bonds and trios of atoms bonded to a central atom are generated. These substructures identify the G- and D-parameters, respectively. Searching lists gives access to the values of these parameters. By adding all parameters found the estimate for the heat of atomization or heat of reaction is obtained.

Working with the B-parameters only, corresponds already to an approach based on simple additivity of bond parameters. If no additional parameters are accessible at least the accuracy of a simple bond additivity scheme is achieved. Thus, the switch between additivity of bond parameters and a higher order approximation is performed rather smoothly. For many classes of organic compounds G- and D-parameters have been determined.[4,14,18] These compounds include alkanes, alkenes, alkynes, haloalkanes, alcohols, ethers, thiols, thioethers, carbonyl- and carboxyl-compounds, peroxides, amines, nitriles.

For cyclic structures corrections for ring strain energies and resonance energies have to be calculated. A simple model has been developed which allows us to perform this task automatically, too.[19] This approach of evaluating energies of cyclic structures has been converted to a subroutine of the main program.

The program has been written in PL/1 and has been implemented on an IBM 360/91, an AMDAHL 470 V6 and a Telefunken TR 440. The program is extremely fast; on an AMDAHL 470 V/6 the calculation of the enthalpy for one reaction takes about 5–20 msec.

RESULTS

The accuracy of the values obtained for heats of atomization and heats of reaction is determined by the quality of the parameters and the validity of an Allen-type approach for a particular class of compounds. To obtain parameters of general utility an optimization of the parameters (e.g. by multi-linear regression with a least squares criterion) based on a larger body of thermochemical data has to be performed. It is the lack of sufficient experimental data for various classes of compounds which may prevent both the determination of appropriate parameters and the comparison of the estimated values with experimental ones. An Allen-type approach will become inaccurate when additional steric, polar, or conjugative interactions not absorbed in the parameters are present.

For those classes of compounds where sufficient experimental data for parameterization and comparison are available usually excellent results are obtained. For a collection of 49 alkanes the mean absolute deviation was 0.64 kcal/mole, for 38 alkenes 0.65 kcal/mole, for 12 chloroalkanes 0.51 kcal/mole, and for 11 bromoalkanes 0.37 kcal/mole.[14] In many cases these deviations are smaller than the error in the experimental data.

The examples given in the following tables were chosen as to allow extensive comparison with experimental data. This objective has put restraints on the types of reactions to be selected. Therefore, sometimes rather simple reactions are given. The success of our approach in those cases where comparison with experimental data can be made also gives credibility to our estimated values for heats of atomization and heats of reaction where the lack of experimental data prevents a comparison.

The mean absolute deviation of 0.57 kcal/mole obtained for the isomerization reactions of *n*-heptane (Table 1) is representative for the deviations usually encountered with alkanes.

In Table 2 a comparison between experimental and calculated heats of hydrogenation of all olefins from ethylene up to the hexenes is made (*cis*-isomers are disregarded as they introduce extra strain energy over the *trans*-isomers). The mean absolute deviation of 0.38 kcal/mole for these 23 reactions shows the high accuracy of our predictions. Neglecting the last value

Table 1. Isomerization reactions of *n*-heptane. Comparison of experimental and calculated heats of reaction. [a]All values in kcal/mole; [b]Experimental (Exp.) data taken from Ref. 20; [c]The deviation (Dev.) is defined as Dev. = calculated − experimental value

Product	Heat of reaction[a]	
	Exp.[b]	Dev.[c]
2-methylhexane	-1.71	0.67
3-methylhexane	-1.07	0.03
3-ethylpentane	-0.45	-0.59
2,2-dimethylpentane	-4.40	1.34
2,3-dimethylpentane	-1.75	-0.33
2,4-dimethylpentane	-3.41	1.33
3,3-dimethylpentane	-3.28	0.22
2,2,3-trimethylpentane	-4.07	-0.03
	mean absolute deviation	0.57

Table 2. Comparison of experimental and calculated heats of hydrogenation of alkenes. [a]Values in kcal/mole; [b]Experimental (Exp.) values for alkenes taken from Ref. 4, for alkanes from Ref. 20; [c]The deviation (Dev.) is defined as Dev. = calculated − experimental value

Alkene	Heat of hydrogenation[a]	
	Exp.[b]	Dev.[c]
ethylene	-32.49	0.02
propene	-29.90	0.07
1-butene	-29.83	0.00
trans-2-butene	-27.04	-0.15
2-methylpropene	-27.81	-0.36
1-pentene	-29.74	-0.09
trans-2-pentene	-27.14	-0.05
2-methyl-1-butene	-28.00	-0.17
3-methyl-1-butene	-29.94	0.12
2-methyl-2-butene	-26.43	0.89
1-hexene	-30.01	0.19
trans-2-hexene	-27.08	-0.11
trans-3-hexene	-26.95	-0.24
2-methyl-1-pentene	-27.47	-0.71
3-methyl-1-pentene	-29.19	-0.64
4-methyl-1-pentene	-29.41	-0.43
2-methyl-2-pentene	-25.68	0.14
3-methyl-*trans*-2-pentene	-25.93	0.39
4-methyl-*trans*-2-pentene	-26.96	-0.24
2-ethyl-1-butene	-27.63	-0.55
2,3-dimethyl-1-butene	-27.30	-0.88
3,3-dimethyl-1-butene	-29.84	0.01
2,3-dimethyl-2-butene	-26.07	2.18
mean absolute deviation		0.38

which involves a highly substituted alkene reduces the mean deviation of the heats of hydrogenation to a mere 0.29 kcal/mole for 22 reactions. The dependence of the heat of hydrogenation on the type of substitution of the olefin is also reproduced quite well. For ethylene the heat of hydrogenation is 32.5 kcal/mole, for monosubstituted olefins 29.2–30.0, for 1,1-disubstituted 27.3–28.0, for 1,2-disubstituted 27.0–27.1, for trisubstituted 25.7–26.4, and for the tetrasubstituted olefin 2,3-dimethyl-2-butene 26.1 kcal/mole. Only in the last case the estimated value deviates appreciably from the experimental one by 2.18 kcal/mole. Closer scrutiny shows that 1.8 kcal/mole of this deviation are due to an underestimation of the strain in the tetrasubstituted olefin. It should be recalled that all values of Table 2, and, in fact, all heats of hydrogenation of acyclic alkenes are calculated from eight parameters only. The four parameters already mentioned for the alkanes and (values in kcal/mole): $B(H{-}H) = 104.2$, $B(C{=}C) = 142.24$, $G(CC_dC_d) = 3.73$, and $D(CCC_d) = -1.04$.

Table 3. Reaction of propane with bromine. Comparison of experimental and calculated heats of reaction. [a]Values in kcal/mole; [b]Experimental (Exp.) data taken from Ref. 4; [c]The deviation (Dev.) is defined as Dev. = calculated − experimental value

Products	Heat of reaction[a]	
	Exp.[b]	Dev.[c]
1-bromopropane + HBr	-11.42	0.34
2-bromopropane + HBr	-14.42	0.39
ethylbromide + methylbromide	-6.66	-0.02
propene + 2 HBr	5.40	-0.07
mean absolute deviation		0.21

The results given in Table 3 for the reaction of propane with bromine demonstrate the consistency of the approach when several classes of compounds are involved. In the example given, alkanes, alkenes, bromoalkanes, bromine, and hydrogen bromide are covered. The deviations between experimental and calculated values are smaller than the experimental error of the data. The estimated values are sufficiently accurate and reliable to serve as a basis for a thermochemical evaluation of reactions. In the example of Table 3 it is clear that 2-bromopropane and hydrogen bromide are the most stable products in the reaction of propane with one mole of bromine.

Summary. A procedure has been developed which allows the estimation of heats of atomization and heats of reaction. This method is also applicable to molecules and reactions where no specific experimental data are available. Thus, predictions of unknown data can be made. The results are sufficiently reliable and accurate to serve for the prediction of the course of reactions and of chemical equilibria. The wealth of individual data on the enthalpies of molecules and reactions can be reduced to a few fundamental data. It is our hope, that future experimental data allow an extension of this approach to additional classes of compounds and the improvement of the fundamental parameters.

The accuracy of our approach can be made progressively higher, depending on the number of parameters being considered. Comparing the results of our method with molecular mechanics calculations allows to extract the magnitude of those steric interactions not yet parameterized. Thus incorporating our program for calculating heats of atomization into a molecular mechanics program would enhance the insights gained with force field calculations. This is the more appealing as our program is so much faster that the additional computation times are negligible.

A computer program based on our approach has been developed. It rests on a representation of the constitution of molecules only. Computation times are many orders of magnitude smaller than with force field methods or quantum mechanical calculations. In addition, computation times for the estimation of heats of reaction are independent of the size of the molecules participating in the reaction as only the reaction center is considered.

The program can process reactions obtained from a reaction generating system. Thus, an automatic evaluation of each reaction generated is performed. Decisions based on this evaluation have impact on the further course of the program run. This paves the way for the development of a program system with artificial intelligence capabilities.

Acknowledgements—Support of this work by Deutsche Forschungsgemeinschaft is gratefully acknowledged. The author wishes to thank Ms. Erika Diener-Weselsky for programming assistance.

REFERENCES

[1] J. Gasteiger and C. Jochum, *Topics Curr. Chem.* 74, 93 (1978).

[2] W. J. Hehre, R. Ditchfield, L. Radom and J. A. Pople, *J. Am. Chem. Soc.* 92, 4796 (1970).

[3] For reviews see:

[a] E. M. Engler, J. D. Andose and P. v. R. Schleyer, *Ibid.* 95, 8005 (1973); [b] N. L. Allinger, *Adv. Phys. Org. Chem.* 13, 1 (1976).

[4] J. D. Cox and G. Pilcher, *Thermochemistry of Organic and Organometallic Compounds.* Academic Press, London (1970).

[5] D. R. Stull, E. F. Westrum and G. C. Sinke, *The Chemical Thermodynamics of Organic Compounds.* Wiley, New York (1969).

[6] K. Fajans, *Ber. Deut. Chem. Ges.* 53B, 643 (1920); *Z. phys. Chem.* 99, 395 (1921); *Ber. Deut. Chem. Ges.* 55B, 2826 (1922).

[7] S. W. Benson, F. R. Cruickshank, D. M. Golden, G. R. Haugen, H. E. O'Neal, A. S. Rodgers, R. Shaw and R. Walsh, *Chem. Rev.* 69, 279 (1969).

[8] K. J. Laidler, *Can. J. Chem.* 34, 626 (1956).

[9] T. L. Allen, *J. Chem. Phys.* 31, 1039 (1959); A. J. Kalb, A. L. H. Chung and T. L. Allen, *J. Am. Chem. Soc.* 88, 2938 (1966).

[10] Calculated from a value of −39.96 kcal/mole for the heat of formation,[11] with 170.9 kcal/mole for the heat of atomization of graphite, and 52.10 kcal/mole for the heat of atomization of hydrogen.

[11] *Selected Values of Properties of Hydrocarbons and Related Substances*, American Petroleum Institute Research Center, Texas A and M University, College Station (1972).

[12] Conventional notations have been used, e.g. $C{-}(C)_2(H)_2$ represents a C atom with two C and two H atoms as immediate neighbors; $E(C{-}H)_2$ is the enthalpy parameter for a secondary C–H bond. For the Allen scheme instead of using Γ- and Δ-symbols, as originally suggested, G- and D-symbols have been introduced.

[13] All values used here were obtained from a reevaluation[14] of the various additivity schemes.

[14] J. Gasteiger, P. Jacob and U. Strauss, *Tetrahedron* 35, 139 (1979).

[15] J. Blair, J. Gasteiger, C. Gillespie, P. D. Gillespie and I. Ugi, *Ibid.* 30, 1845 (1974).

[16] J. Dugundji and I. Ugi, *Topics Curr. Chem.* 39, 19 (1973).

[17] J. Blair, J. Gasteiger, C. Gillespie, P. D. Gillespie and I. Ugi, *Computer Representation and Manipulation of Chemical Information*, (Edited by W. T. Wipke, S. Heller, R. Feldmann and E. Hyde), pp. 129–145. Wiley, New York (1974).

[18] K. Pihlaja and E. Taskinen, *Physical Methods in Heterocyclic Chemistry*, (Edited by A. R. Katritzky), Vol. 6, pp. 199–246. Academic Press, New York.

[19] J. Gasteiger and O. Dammer, *Tetrahedron* 34, 2939 (1978).

[20] J. W. H. Kao and A. Chung-Phillips, *J. Chem. Phys.* 63, 4143 (1975).