

Linear Relations between Molecular Energies¹

Jane Hylton McCreery and George G. Hall*²

Contribution from the Department of Chemistry, University of Kansas, Lawrence, Kansas 66045. Received May 13, 1974

Abstract: The total energies of a series of molecules, as calculated by Snyder and Basch from *ab initio* wave functions, are analyzed using linear models. In the most accurate of these the energies are reproduced to better than 0.006 hartree. The equations also predict hydrogenation energies in reasonable agreement with experiment. The analysis demonstrates that the concepts of an atom bound in a molecule, heteropolar bond energy, and electronegativity may still be useful in this quantum context and shows their relation to earlier concepts. The value for this type of analysis of calculations on a series of molecules using consistent approximations is stressed.

In a recent book Snyder and Basch³ have published extensive tables giving the results of *ab initio* quantum calculations on a wide variety of molecules. Since the calculations use a uniform set of approximations they open up opportunities of comparing molecules and of analyzing their properties to an extent hitherto impossible. It is highly desirable that these comparisons should isolate the effect of the approximations and concentrate on quantities where they may be expected to cancel.

The Snyder-Basch tables, in particular, list the total energies of molecules as calculated in the Hartree-Fock approximation to double ζ accuracy. The atomic orbitals used for the various atoms are represented by Gaussian functions in a standard way and the calculations are performed in a uniform manner. The results should therefore be strictly comparable. For closed shell molecules the neglect of relativistic corrections and of electron correlation will produce systematic energy differences, but these are likely to remain constant for a particular atom irrespective of the molecule. These total molecular energies are not often observable experimentally though they can frequently be deduced indirectly. Differences between these energies are much more observable as heats of reaction and it is these which will be of greatest interest.

This paper reports an analysis of total energies by using a set of linear models. These models evolve as the analysis proceeds and the energies are explained to greater precision. The form of the models is not prescribed by any quantum mechanical assumptions. It turns out, however, that they lead to interpretations in terms of concepts more familiar in traditional thermochemistry. Thus bond energies are calculated and related to an electronegativity scale which compares well with existing scales.

The major tool used in this paper to establish the models has been multiple regression. The details of this procedure and of the statistical tests which govern it have been given by Ralston and Wilf.⁴ The program itself was the one implemented as part of the biomedical package.⁵

Atoms in Molecules

The simplest model for the total energy of a molecule is that it is the sum of terms characteristic for each of its atoms. For this to be better than a rough approximation it is desirable that each time an atom occurs it should have a similar environment. This can be achieved by restricting the molecules to be considered in the first instance to those closed shell molecules in which the atoms occur with their normal valencies and the bonds have unstrained angles and normal lengths.

The Snyder-Basch tables are restricted to molecules containing the atoms H, B, C, N, O, and F. In practice there are so few containing B atoms that these had to be excluded.

The molecules selected are those listed in Table I. The energy for each is expressed in the simple linear form.

$$-W = hH + cC + nN + oO + fF \quad (1)$$

where h, c, \dots, f are the numbers of these atoms in the molecule and H, C, \dots, F are the atomic energies. Since the energies are all negative it is convenient to work with $-W$ as a positive quantity. The values to be given to the atomic energies are determined by the multiple regression program which fits them to the list of calculated energies. These values, together with their standard errors, are given in Table II. As a standard of comparison the energies³ of the free atoms in their ground states are also given.

The difference between the energy of an atom in a molecule and a free atom may be called an atomic binding energy. By using these the relativistic and correlation errors, which are primarily located in the description of the inner electrons, may be canceled. These binding energies may be interpreted as heats of reaction. Thus, for example, the reaction $\frac{1}{2}\text{H}_2 \rightarrow \text{H}$ has the heat 0.0867 hartree or 54.4 kcal mol⁻¹. This brings the concept very close to that of the heat of atomization. There are some differences, however, since these energies take no account of temperature effects and so are at 0°K. They are also all gas phase results for normal bonding so that, for example, solid graphite could not be taken as the standard for bound carbon.

The molecular energies calculated using these values for the atomic energies are given in Table I. The residuals are of approximate magnitude 0.025 thus showing that the linear model of eq 1 successfully explains the major part of the variations. On the other hand a close inspection of the residuals shows that there are some systematic trends which could be explained by a more detailed model.

It is clear that this model is crude from a chemical point of view. Not only does it predict the same energy for all molecules that have the same chemical formula but it also predicts a zero heat of reaction for all reactions involving bound species. A model is needed that takes some account of how the atoms are related to one another in the molecule.

Bond Energies

The next simplest model that can be suggested in the light of the previous one is one in which the energy has contributions depending on the various pairs of nearest neighbors as well as the atomic terms. Such a model is not sound mathematically since, for example, the energy of H₂ would be expressed as $2H + \text{HH}$ with two variables one of which, HH, would occur in this and no other molecule. Furthermore, the value of H can be changed by an arbitrary amount provided that HH and all the other combinations involving H are modified to keep the sums the same. This unsatisfactory feature of the model can be eliminated and

Table I. Total Molecular Energies

Molecules used	S and B energy	Eq 1 and Table II	Eq 2 and Table III	Eq 2 and 3 and Tables III and V
CH ₄	-40.1823	-40.1879	-40.1886	-40.1528
NH ₃	-56.1714	-56.1466	-56.1689	-56.1954
H ₂ O	-76.0035	-75.9882	-75.9981	-76.0113
C ₂ H ₂	-76.7919	-76.8599	-76.8171	-76.7992
HCN	-92.8286	-92.8186	-92.8477	-92.8418
C ₂ H ₄	-78.0054	-78.0319	-78.0038	-77.9680
H ₂ CO	-113.8209	-113.8321	-113.8537	-113.8265
C ₂ H ₆	-79.1981	-79.2039	-79.1905	-79.1368
N ₂ H ₄	-111.1261	-111.1211	-111.1371	-111.1724
H ₂ O ₂	-150.7373	-150.8043	-150.7372	-150.7505
CH ₃ OH	-115.0059	-115.0041	-115.0202	-114.9953
C ₃ H ₄	-115.8203	-115.8759	-115.8190	-115.7832
CH ₃ CO	-151.6595	-151.6761	-151.6689	-151.6417
NH ₂ CN	-147.8437	-147.7931	-147.8467	-147.8684
CH ₃ CN	-131.8674	-131.8345	-131.8496	-131.8258
CHONH ₂	-168.8685	-168.8067	-168.8527	-168.8531
CHOOH	-188.6888	-188.6483	-188.6853	-188.6691
C ₃ H ₄ O	-190.6899	-190.6921	-190.6708	-190.6257
C ₃ O ₂	-263.1575	-263.1643	-263.1492	-263.1306
H ₂	-1.1266	-1.1720	-1.1266	-1.1266
HF	-100.0150	-99.9939	-100.0150	-100.0085
F ₂	-198.6932	-198.8159	-198.6932	-198.6932
CH ₃ F	-139.0203	-139.0099	-139.0247	-138.9925
CH ₂ F ₂	-237.8687	-237.8318	-237.8607	-237.8322
F ₂ CO	-311.5244	-311.4760	-311.5258	-311.5059
CHF ₃	-336.7157	-336.6538	-336.6968	-336.6718
C ₂ F ₄	-473.3331	-473.3197	-473.3480	-473.3267
N ₂ F ₂	-307.5104	-307.5930	-307.5104	-307.5340
FNO	-228.5521	-228.6127	-228.5521	-228.5599
CHOF	-212.6841	-212.6541	-212.6898	-212.6662
N ₂ H ₂	-109.9418	-109.9491	-109.9473	-109.9539

Table II. Energies of Atoms in Molecules (hartrees)

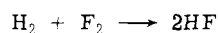
	Energy of bound atom (standard error)	Free atom	Atomic binding energy
H	0.5860 (0.005)	0.4993	0.0867
C	37.8440 (0.010)	37.6812	0.1628
N	54.3886 (0.011)	54.3897	-0.0011
O	74.8162 (0.012)	74.7931	0.0231
F	99.4079 (0.008)	99.3863	0.0216

the comments made earlier interpreted more exactly by using a model in which only the heteropolar neighbors contribute. The homopolar neighbors are taken to contribute to the atomic terms just as in the earlier model. Accordingly the energy relation is

$$-W = hH + \dots + fF + \chi CH + \dots + \phi HF + \dots \quad (2)$$

where χ is the number of carbon-hydrogen bonds in the molecule and CH is their energy contribution. Multiple bonds are counted in terms of the single bonds, e.g., $C\equiv N$ counts as $3CN$. The atomic energies H, \dots, F have to be redetermined.

The atomic energies and heteropolar energies that result from fitting (2) to the total molecular energies are listed in Table III. It is noteworthy that all the heteropolar energies are positive showing that these bonds are more stable than the arithmetic mean of the homopolar molecules. Thus a reaction such as



will be predicted to be exothermic. The bound atomic energies are all lower than those in Table II and the atomic binding energies for O and F become negative as well as for N. This is an unchemical feature of the results and may be due to the fact that the Snyder-Basch wave functions are

Table III. Atomic Energies and Heteropolar Bond Energies

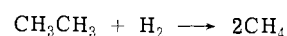
	Bound atom energies (standard error)	C	N	O	F
H	0.5633 (0.001)	0.0301	0.0371	0.0661	0.1051
C	37.8152 (0.001)		0.0238	0.0563	0.0828
N	54.3678 (0.002)			0.0288	0.0408
O	74.7392 (0.003)				
F	99.3466 (0.001)				

Table IV. Hydrogenation Energies (kcal/mol)

Reaction	Predicted from heteropolar energies	Ref 7	
		Calcd	Exptl
CH ₃ CH ₃ + H ₂ → 2CH ₄	-37.8	-23.5	-18.1
CH ₂ CH ₂ + 2H ₂ → 2CH ₄	-75.6	-65.9	-57.2
CHCH + 3H ₂ → 2CH ₄	-113.3	-117.8	-105.4
CH ₃ NH ₂ + H ₂ → CH ₄ + NH ₃	-27.2	-30.9	-25.7
CH ₂ NH + 2H ₂ → CH ₄ + NH ₃	-54.4	-72.1	
HCN + 3H ₂ → CH ₄ + NH ₃	-81.6	-83.4	-76.8
CH ₃ OH + H ₂ → CH ₄ + H ₂ O	-25.0	-32.0	-30.3
CH ₂ O + 2H ₂ → CH ₄ + H ₂ O	-50.0	-64.3	-57.3
CH ₃ F + H ₂ → CH ₄ + HF	-32.9	-27.4	-29.5
NH ₂ NH ₂ + H ₂ → 2NH ₃	-46.6	-50.4	-50.0
NHNH + 2H ₂ → 2NH ₃	-93.1	-90.7	
N ₂ + 3H ₂ → 2NH ₃	-139.7	-47.3	-37.7
NH ₂ OH + H ₂ → NH ₃ + H ₂ O	-46.7	-62.0	
HNO + 2H ₂ → NH ₃ + H ₂ O	-93.4	-114.2	-102.9
NH ₂ F + H ₂ → NH ₃ + HF	-63.6	-71.2	
HOOH + H ₂ → 2H ₂ O	-83.0	-86.3	-86.8
O ₂ + 2H ₂ → 2H ₂ O	-166.0	-107.1	-125.1
F ₂ + H ₂ → 2HF	-131.9	-118.9	-133.8

designed to produce a good energy for the free atom and not necessarily for the bound one. The standard errors of the various terms are much smaller than those in Table II so that the fitting represents a substantial improvement. This is also clear from the predicted energies which are given in Table I. The magnitude of the residuals is now about 0.006.

The increased value of this model of molecular energies is shown not only in the superior fitting but in its predicted heats of reaction. In a reaction such as



the atoms are balanced on both sides of the equation so that the atomic energies are the same. The heat of reaction is entirely due to the heteropolar bond energies and would here be $2(CH)$. Table IV shows the hydrogenation energies for a series of molecules predicted in this way using the energies given in Table III.

In a series of papers,^{6,7} Pople and his collaborators have reported calculations on a series of molecules somewhat similar to those of Snyder and Basch. The hydrogenations listed in Table IV are those used in one of their papers⁷ to illustrate the accuracy and utility of their calculations of total molecular energies, and we list their values and the experimental values which they quote. They have not analyzed the results in terms of heteropolar energies. For most molecules their energies lie above those of Snyder and Basch showing that they are further from the Hartree-Fock limit. The energy differences are approximately 0.04 for each C, 0.06 for N, 0.1 for O, and 0.12 for F.

Traditionally, in thermochemistry molecular heats of formation have been discussed in terms of bond energies. This is supported by quantum mechanical argument in which the total energy is allocated to atoms and bonds by using localized orbitals.⁸ The analysis above in which the homopolar bond energies do not appear seems to be easier to apply and

Table V. Electronegativities

	H	C	N	O	F
x_a	0	0.02113	0.04594	0.07282	0.09861
Standard errors		0.006	0.006	0.006	0.006
Adjusted x_a	(2.1)	2.52	3.01	3.55	4.06
Pauling	2.1	2.50	3.07	3.50	4.10
Mulliken	2.28	2.63	2.33	3.17	3.91

better related to experiment. It is interesting to note that the approximation of treating heterobond energies as the arithmetic mean of homopolar energies was proposed by Pauling⁹ but later abandoned in favor of the geometric mean.

Electronegativities

The heteropolar energies listed in Table III exhibit some degree of regularity which invites further analysis. The values rise steadily away from the leading diagonal. This suggests a model which depends on differences in atomic contributions. In this model the heteroenergy AB is represented as

$$AB = x_a - x_b \quad (3)$$

where A is the more electronegative atom. To fix the origin of the x variables it is convenient to set $x_H = 0$. The heteroenergies were then all fitted to the values in Table III. The resulting values of these x_a are shown in Table V. These values can be related to the customary electronegativities by giving H the standard value 2.1 and making a scale adjustment to optimize the fit.

$$X_a = 2.1 + x_a/0.0504$$

The resulting electronegativities are given in Table V together with the Pauling⁶ values, as revised by Little and Jones,¹⁰ and the Mulliken values as revised by Pritchard and Skinner.¹¹ The agreement is surprisingly good when the very different arguments used for the different scales are remembered. The Pauling scale, for example, is based on the excess of the heteropolar bond energy over the geometric mean of the homopolar energies and is then related to the square of the electronegativity differences. The present model manages to use purely linear relations.

It is tempting to rationalize the success of this energy formula in terms of molecular orbital concepts just as Pauling has rationalized his in terms of ionic-covalent resonance. The sum over atomic energies is obtained by forcing each localized bonding orbital to give equal weight to the two atoms. The correction to allow for its unequal distribution of charge will have the form

$$\delta W = \delta q_a \frac{\partial W}{\partial q_a} + \delta q_b \frac{\partial W}{\partial q_b}$$

where δq_a is the charge gained by A and for conservation of charge $\delta q_b = -\delta q_a$

$$\delta W = \delta q \left(\frac{\partial W}{\partial q_a} - \frac{\partial W}{\partial q_b} \right) \quad (4)$$

This partial derivative with respect to the charge is easily related to other definitions of electronegativity. Thus if the ionization potential I and electron affinity A for the same valence orbital in the atom are defined by

$$I = W(q - 1) - W(q)$$

$$A = W(q) - W(q + 1)$$

Table VI. Predicted and Calculated Molecular Energies

Molecules with 3-membered rings	Calcd by S-B	Predicted from (2)	Difference kcal mol ⁻¹
C ₃ H ₄	115.7655	115.8190	33.6
CH ₂ N ₂	147.7287	147.7850	35.3
C ₃ H ₆	117.0099	117.0057	-2.6
C ₂ H ₄ NH	132.9726	133.0196	29.5
CH ₂ NHNH	148.8430	148.9858	89.6
CF ₂ N ₂	345.3999	345.4571	35.9
C ₄ H ₆	154.7889	154.8209	20.1
Other molecules			
F-C-F	236.6114	236.6740	39.3
O ₃	224.1905	224.2175	16.9
CH ₃ N=C	131.8400	131.8496	6.0
C=O	112.6763	112.6670	-5.8
CO ₂	187.5377	187.5188	-11.9
CF ₄	435.5554	435.5328	-14.2
N ₂ O ₄	407.8354	407.8076	-17.4
N ₂ O	183.5761	183.5324	-27.4
N ₂	108.8695	108.7355	-84.1

and if W is expanded in a Taylor expansion about the neutral atom value

$$W(q + x) = W(q) + x \frac{\partial W}{\partial q} + \frac{x^2}{2} \frac{\partial^2 W}{\partial q^2}$$

then the Mulliken definition of electronegativity is obtained.

$$\partial W / \partial q = -1/2(I + A) = -X$$

The interpretation of the heterobond term in (2) using (4) would then be that δq has a similar value for all heterobonds.

Total Molecular Energies

The initial purpose in setting up these models of the total molecular energy was to predict the total energies to good accuracy by means of a simple formula. Table I brings together the results that are obtained using the different equations for the original list of molecules. The closeness of the fitting can now be appreciated directly. In particular, the energy values in Table III with eq 2 give the best fitting and the replacement of the heterobond energies by the x_a energies from Table V reduces the accuracy but not to the extent of the simple fitting which uses eq 1 and Table II. It is worth noting that these fittings produce very much smaller deviations than are produced by comparing the total energies as computed by Radom, Hehre, and Pople⁷ with those of Snyder and Basch.³

Not all of the molecules calculated by Snyder and Basch are included in Table I. It is of some interest to examine the calculated energies with those predicted by (2) for a number of other molecules. These are shown in Table VI and fall into two categories. The first is a group of molecules each containing a three-membered ring. The energy differences are similar for most of the molecules and can perhaps be interpreted as strain energy. The second group contains molecules such as C=O and O₃ where the atoms show unusual valencies or molecules showing conjugation such as CO₂ and N₂O₄. One molecule, N₂, is much more stable than our formula predicts. This stability, together with that of O₂, was already demonstrated in Table IV and may be due in part to a peculiarity of the basis functions.

Conclusion

It may be said that one achievement of this paper is the production of formulas which enable the total energies of many molecules to be predicted with much less effort than is required in an *ab initio* calculation. It certainly could be used to check such calculations.

In a more important way it has tried to show that linear models can be successfully applied to quantum mechanical results just as they are to empirical results. This greatly improves the chemical utility of the results since the known errors can be gradually eliminated and quantities produced which relate more directly to experimental measurements.

The paper also demonstrates that, once a sufficiently consistent collection of results is available, quantum calculations are susceptible to chemical explanation and interpretation. This appears to be a most fruitful starting point for the further investigation of many concepts that had seemed to disappear under the sheer mass of numbers produced by the computer.

Acknowledgments. The authors express their appreciation to the University of Kansas for partial support of the computing required for this study.

References and Notes

- (1) This work was supported in part by grants from the National Science Foundation.
- (2) Department of Mathematics, University of Nottingham, Nottingham, England.
- (3) L. C. Snyder and H. Basch, "Molecular Wave Function and Properties," Wiley, New York, N. Y., 1972; *J. Amer. Chem. Soc.*, **91**, 2189 (1969).
- (4) A. Fialston and H. S. Wilf, "Mathematical Methods for Digital Computers," Wiley, New York, N. Y., 1967.
- (5) B. M. D. Biomedical Computer Programs, W. J. Dixon, Ed., University of California Press, Berkeley and Los Angeles, Calif., 1968.
- (6) W. J. Hehre, R. Ditchfield, L. Radom, and J. A. Pople, *J. Amer. Chem. Soc.*, **92**, 4796 (1970).
- (7) L. Radom, W. J. Hehre, and J. A. Pople, *J. Amer. Chem. Soc.*, **93**, 289 (1971).
- (8) G. G. Hall, *Proc. Roy. Soc., Ser. A*, **213**, 113 (1952).
- (9) L. Pauling, "Nature of the Chemical Bond," Cornell University Press, Ithaca, N. Y., 1960.
- (10) E. J. Little and M. M. Jones, *J. Chem. Educ.*, **37**, 231 (1960).
- (11) H. O. Pritchard and H. A. Skinner, *Chem. Rev.*, **55**, 745 (1955).

Ab Initio Calculations on Large Molecules Using Molecular Fragments. Unrestricted Hartree-Fock Calculations on the Low-Lying States of Formaldehyde and Its Radical Ions¹

Timothy D. Davis,^{2a,b} Gerald M. Maggiora,^{2c} and Ralph E. Christoffersen*^{2a}

Contribution from the Departments of Chemistry and Biochemistry, University of Kansas, Lawrence, Kansas 66045. Received February 25, 1974

Abstract: This paper describes the extension of the *ab initio* molecular fragment procedure to open-shell systems using the unrestricted Hartree-Fock (UHF) approach. A method of obtaining initial charge density and bond order matrices is also presented. Calculations are performed for the $^3A_1(\text{GS})$, $^1,^3A_2(n \rightarrow \pi^*)$ and $^3A_1(\pi \rightarrow \pi^*)$ states of formaldehyde, for the $^2B_1(\text{GS})$ and $^2B_2(n \rightarrow \pi^*)$ states of the anion, and for the $^2B_2(\text{GS})$ and $^2B_2(\pi \rightarrow n)$ states of the cation. A number of properties of these states are presented and analyzed, including the dependence of the total energy on the out-of-plane bending angle, Franck-Condon transitions, molecular orbital structure, and several other molecular properties. Except for the $^3A_1(\pi \rightarrow \pi^*)$ state, where a very shallow potential curve makes definitive conclusions difficult, the geometric predictions are excellent, with predicted equilibrium angles within a few per cent of other calculated or measured values. For all states examined, where comparisons were possible, the correct ordering of states was obtained. For cases where quantitative agreement was not obtained, analysis of the charge distribution and other properties allows assessment of the basis set adequacies and deficiencies and indicates how improvements can be made.

Over the past several years a large number of closed-shell molecular systems have been investigated by the *ab initio* molecular fragment procedure.⁷ This procedure, based on the Hall-Roothaan equations,³ has been found to be applicable to a wide range of problems concerning small molecules such as cumulenes⁴ and larger molecules such as acetylcholine⁵ and glycine polypeptides.⁶ The procedure is currently being extended to handle molecules containing second-row atoms.⁸ Based on the success of the closed-shell studies, it seemed appropriate to examine the suitability of the molecular fragment procedure for the study of open-shell systems. An extension of the method to open-shell states of large molecules would provide another source of new chemical and spectroscopic information which would be quite useful in understanding the chemistry of various neutral and charged species as well as various excited states. The current paucity of information of this nature for large molecular systems underscores further the need for such techniques.

The single determinantal unrestricted Hartree-Fock (UHF) model⁹ used in the current studies has already been applied using the molecular fragment procedure in a preliminary investigation.¹⁰ This paper deals with the applica-

tion of the molecular fragment procedure to an examination of the ground and excited states of formaldehyde and its cation and anion radicals (H_2CO , H_2CO^+ , H_2CO^-). In order to study the usefulness of this procedure, a number of molecular properties were calculated, such as transition energies, excited state geometries, barriers to inversion, molecular orbital (MO) orderings, charge distributions, dipole moments, and Hellmann-Feynman field strengths at the nuclei. A comparison of the computed values and the available experimental data for the wide range of properties listed above provides valuable information concerning the suitability of the method for describing certain properties of open-shell systems and also suggests possible changes to the procedure which would improve the calculation of these properties. Since the UHF wave function is usually not an eigenfunction of the spin-squared operator (S^2), the effect of annihilating the spin component of next higher multiplicity on the various computed properties is also examined.

I. Method and Analysis

The basis set consists of normalized, nonorthogonal, floating spherical Gaussian orbitals (FSGO) defined as follows