Molecular Similarity for Small Species: Refining the Isoelectronic Index

Ray Hefferlin* and Myla Thomas Matus†

Physics Department, Southern Adventist University, Collegedale, Tennessee 37315

Received June 30, 2000

This paper tests a chemical-similarity model against properties of gas-phase neutral triatomic and four-atom molecules. The model is a variant of the Diatomics-in-Molecules (DIM) picture, which considers a molecule to be the superposition of all diatomic molecules that could be formed from adjacent atoms in the molecule. The variant is that adjacent atoms are counted as a diatomic molecule only if they are *bonded*. The tests consist of investigating whether molecules with the same number of electrons, computed by the adjacent-DIM model, have data more similar than do molecules selected at random. The tests vindicate the model for the heat of atomization and for the equilibrium constant for formation, they agree with the model with lesser confidence for the entropy and the partition function, and they show that the model fails for the ionization potential. The model applies with most confidence to molecules with the more electronegative atoms from rows 2 or 3 (those of greatest interest in organic chemistry) and with lesser confidence otherwise. For these properties and these molecules, the model passes graphical and statistical tests at least as well as does the traditional isoelectronic model. Thus, this work refines what may be called the isoelectronic index.

1. INTRODUCTION

The nearly desperate search for molecules with properties of interest to toxicology, environmental studies, and pharmaceuticals powers the field of mathematical chemistry. This field seeks correlations of properties with each other and with various indices based on molecular structure; it employs QSPR (quantitative structure−property relationships) and QSAR (quantitative structure−activity relationships); and more recently it has created formula periodic systems of classes of molecules. There is also a search for molecules of interest in astrophysics and in combustion science. It is very useful, particularly in the latter search, to estimate a property of some target molecule as being similar to that of other molecules with the same number of valence electrons. Thus, the total number of valence electrons of the constituent atoms is an index, and this paper investigates a refinement of this isoelectronic index.

This investigation is based on the Adjacent-Diatomics-in-Molecules (Adjacent-DIM) model, in which all diatomic molecules corresponding to the *bonded* atoms in the molecule are superposed. According to this model, the number of valence electrons in acyclic triatomic molecules is $g(C) = (C_1+C_2)+(C_2+C_3) = (C_1+2C_2+C_3)$, where C_2 is the central atom. As an example, FNO has 18 valence electrons and has $g(C) = 7 + 2 \times 5 + 6 = 23$. In a previous paper[1] it was shown that a series of adjacent-DIM isoelectronic molecules has data at least as similar as a series of traditionally isoelectronic species and much more similar than an equal number of molecules chosen at random. It was proposed that this model could join the use of isoelectronic molecules, with constant $n_e = C_1 + C_2 + C_3$, and isovalent molecules, as a tool of chemical similarity.

The evidence presented earlier[1] consists of a graphical display and of statistical analyses of tabulated data for acyclic triatomic molecules. This present study presents additional evidence that the adjacent-DIM model applies to triatomics. It does this with additional graphical analyses of spectroscopic and thermodynamic tabulated data and also of data globally predicted by least-squares fitting and neural-network training. This paper clarifies the triatomic-molecular domain within which the adjacent-DIM model is applicable. Then it presents a statistical analysis of tabulated data that demonstrate the applicability of the model for *four*-atom molecules, which require different $g(C)$ formulas for different structures.

2. HYPOTHESIS

This paper assumes the "Adjacent-Diatomics-in-Molecules" model, in which all diatomic molecules corresponding to adjacent, *bonded*, atoms in the molecule are superposed. This model differs from the Diatomics-in-Molecules (DIM) model wherein diatomic molecules replace *all* atom pairs.

The hypothesis to be tested is that *fixed-row molecules with the same number of electrons, computed by the adjacent-DIM model, have data more similar than do molecules selected at random, just as isoelectronic molecules have data more similar than do molecules chosen at random*. If the test is successful, then a series of isoelectronic adjacent-DIM molecules could be useful for predicting the properties of a target molecule in the series, just as traditionally isoelectronic molecules are useful for predicting the properties of a target molecules.

If the test is successful, a logical problem will emerge. Any given series of "fixed-row molecules with the same number of electrons, computed by the adjacent-DIM model," which by hypothesis has members with data close to some value α, intersects several series of traditionally isoelectronic molecules each of which has members with data close to different values β_1, β_2, β_3, ... If α is equal to any value β_i,

* Corresponding author phone: (423)238-2869; e-mail: hefferln@southern.edu.

† Present address: Texas A&M University, Department of Mathematics, College Place, TX 77843.

10.1021/ci000070p CCC: $20.00
Published on Web 01/09/2001

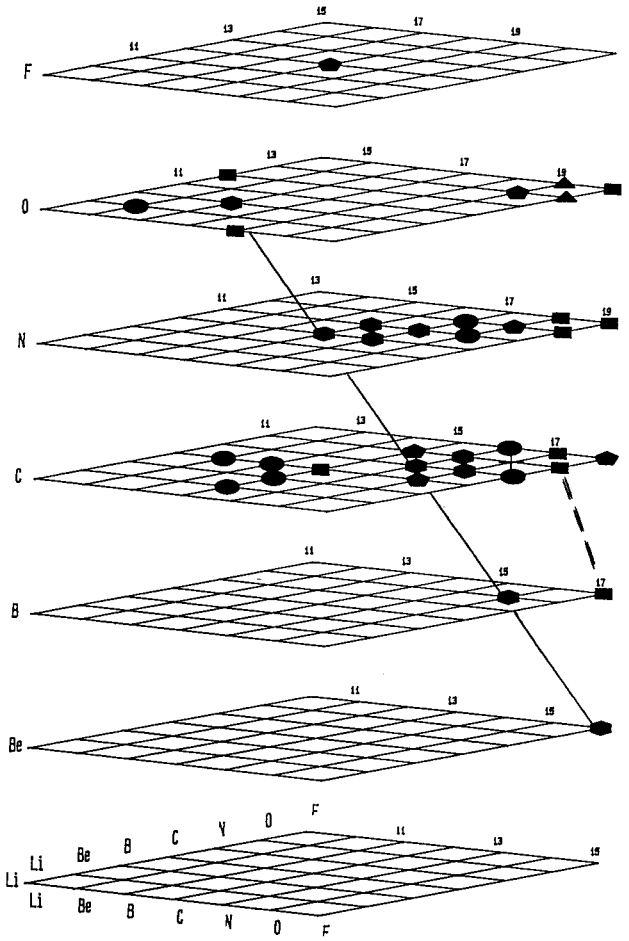

Figure 1. The distribution of triatomic molecules with tabulated data for log Q for $(R_1,R_2,R_3) = (2,2,2)$. The coordinates are C_1 (right front), C_2 (vertical), and C_3 (right rear). The data are separated into equal ranges as follows: + (CO_2 only), from 3.400 to 3.790; ●, from 3.800 to 4.190; ⬢, from 4.200 to 4.590; ⬟, from 4.600 to 4.990; ■, from 5.000 to 5.390; and ▲ (FNO and ONF only), 5.400 to 5.790. Most of the data are above and to the right of a bounding plane containing FBeF, at lower right, and LiON and NOLi, at left on the plane for oxygen-center molecules. On the centerline of this plane is a row of data shown as hexagons connected by a solid line. There is a similar row of data for ΔH_a(*Sauval*) and S_{298}. Isoelectronic molecules lie on tilted parallel planes whose intersections with planes of constant C_2 are indicated by small numbers equal to n_e. Data for several carbon-center molecules push to the left beyond the upper-left lower-right envelope of the other data, suggestive of the stability of (or interest in) such molecules.

then it cannot be equal to any of the others. In other words, a molecule that has an observable close to the average value α of an adjacent-DIM series cannot at the same time have an observable close to the average values βj of all intersecting traditional isoelectronic series.

There is a possible solution to this problem, aside from hiding it behind uncertainties in the data. This solution lies in the possibility that there are domains in the chemical data space where traditional isoelectronic series display the greater constancy of data and other domains where adjacent-DIM isoelectronic series do so. This investigation seeks to establish the existence of these second domains.

3. THEORY

3.1. Definitions. The group space C_1, C_2, C_3, ... C_N enumerates molecules with fixed-period numbers R_1, R_2, R_3, ... R_N. It is a subspace of the periodic system of N-atomic molecules.[2] This paper concerns only main-group molecules from periods $2 \le R_i \le 7$ (where $R_i = 2$ for Li) and groups $1 \le C_i \le 7$. (The IUPAC numbers for these groups are 1, 2, and 13 to 17; they are not used here.)

A molecular notation is adopted that clearly indicates which atom is central, for instance OCO is used instead of CO_2. References to the total numbers of electrons in molecules, n_e, are intended to mean the sum of the valence electrons of the constituent atoms. Thus, for triatomic molecules

$$n_e = C_1 + C_2 + C_3 \tag{1}$$

Isoelectronic molecules have common row numbers and the same value of n_e. If molecules have various row numbers and one common value of n_e, they are called isovalent, homologous, or vertically isoelectronic.

3.2. Adjacent-DIM Series. The quantity $g(C)$ enumerates the number of valence electrons in a model molecule, where C_i is the number of valence electrons of atom i. The two atoms in diatomic molecules are adjacent and bonded; hence, isoelectronic molecules are the same in the traditional and in the adjacent-DIM models. For acyclic triatomic molecules

$$g(C) = (C_1+C_2) + (C_2+C_3) = C_1 + 2C_2 + C_3 \tag{2}$$

where atom 2 is central. To illustrate FBeF, OBO, NBF, BCF, CCO, NCN, BNN, CNC, etc. all have $g(C) = 18$ but have different n_e.

Cyclic triatomic and cyclic tetra-atomic molecules (parallelogram or rhomboid, planar or not) have identical isoelectronic sequences in the traditional and in the adjacent-DIM senses. For other tetra-atomic molecules, $g(C)$ is different for each molecular structure. If the molecule is linear/bent, then

$$g(C) = C_1 + 2C_2 + 2C_3 + C_4 \tag{3}$$

where 1 and 4 are the atoms at the ends. If it is trigonal, planar or not, then

$$g(C) = 3C_1 + C_2 + C_3 + C_4 \tag{4}$$

where 1 is the atom in the center. If its structure is that of a tetrahedron, then

$$g(C) = C_1 + C_2 + C_3 + C_4 \tag{5}$$

independent of the atomic numbering. No kite-shaped molecules were encountered.

3.3. Molecular Stability. Stability against atomization (expansion of the molecule, without change in relative atomic positions, until it consists of atomic fragments) from the ground state of the molecule is measured by the heat or energy of atomization ΔH_a; the larger the value of this property, the greater the stability. Bond lengths, vibration frequencies, force constants, and various anharmonic spectroscopic constants are related to ΔH_a. Collectively they have been called potential hypersurface properties.

Also correlated to stability against atomization is the equilibrium constant for formation, ^pK. It is defined, in the

Table 1. Molecular Coordinates and Symmetries

mol.	symmetry symbol	structure	n_e	groups 1	2	3	4	periods 1	2	3	4	mol.	symmetry symbol	structure	n_e	groups 1	2	3	4	periods 1	2	3	4
Al2C2	Dinfh	linear/bent	14	3	3	4	4	3	3	2	2	ClC2Br	D3h		22	7	4	4	7	3	2	2	4
Al2O2	D2h	alt. paragm	18	3	3	6	6	3	3	2	2	ClF3	C2v	trigonal	28	7	7	7	7	3	2	2	2
Al2S2	Dinfh	linear/bent	18	3	3	6	6	3	3	3	3	ClNCO	Cs	trigonal	22	7	5	4	6	3	2	2	2
AlBO2	Cs	linear/bent	21	3	3	6	6	3	2	2	2	ClNO2	C2v	trigonal	24	7	5	6	6	3	2	2	2
AlBr3	D3h	trigonal	24	3	7	7	7	3	4	4	4	ClO3			25	7	6	6	6	3	2	2	2
AlCl3	D3h	trigonal	24	3	7	7	7	3	3	3	3	Cs2Br2	D2h	alt. paragm	16	1	1	7	7	6	6	4	4
AlF2Cl	C2v	trigonal	24	3	7	7	7	3	2	2	3	Cs2Cl2	D2h	alt. paragm	16	1	1	7	7	6	6	3	3
AlF2O	C2v	trigonal	23	7	7	3	6	2	2	3	2	Cs2F2	D2h	alt. paragm	16	1	1	7	7	6	6	2	2
AlF3	D3h	trigonal	24	3	7	7	7	3	2	2	2	Cs2I2	D2h	alt. paragm	16	1	1	7	7	6	6	5	5
AlFCl2	C2v	trigonal	24	3	7	7	7	3	2	3	3	Cs2O2	D2h		14	1	1	6	6	6	6	2	2
AlI3	D3h	trigonal	24	3	7	7	7	3	5	5	5	CsBO2	Cs		16	1	3	6	6	6	2	2	2
AlO3	*a*	*a*	21	3	6	6	6	3	2	2	2	CsF2	C2v	trigonal	24	4	6	7	7	2	3	2	2
AlOCl2	C2v		23	7	7	3	6	3	3	3	2	CsNO2	C2v	trigonal[b]	18	1	5	6	6	6	2	2	2
As2O2	C2v		22	5	5	6	6	4	4	2	2	FClO2	C2v	trigonal	26	7	7	6	6	2	3	2	2
AsBr3	C3v	trigonal	26	5	7	7	7	4	4	4	4	FNO2	C2v	trigonal	24	7	5	6	6	2	2	2	2
AsCl3	C3v	trigonal	26	5	7	7	7	4	3	3	3	Ga2O2	D2h?		18	3	3	6	6	4	4	2	2
AsF3	C3v	trigonal	26	5	7	7	7	4	2	2	2	GaBr3	D3h	trigonal	24	3	7	7	7	4	4	4	4
AsI3	C3v	trigonal	26	5	7	7	7	4	5	5	5	GaCl3	D3h		24	3	7	7	7	4	3	3	3
AsO3	D3h		23	5	6	6	6	4	2	2	2	GaF3	D3h	trigonal	24	3	7	7	7	4	2	2	2
B2O2	Dinfh	linear/bent	18	3	3	6	6	2	2	2	2	GaI3	D3h	trigonal	24	3	7	7	7	4	5	5	5
B2S2	C2v	trigonal	18	3	3	6	6	2	2	3	3	GaO3			21	3	6	6	6	4	2	2	2
BBeO2	Cs	linear/bent	17	3	2	6	6	2	2	2	2	Ge2O2	D2h	alt. paragm	20	4	4	6	6	4	4	2	2
BBr2Cl	C2v	trigonal	24	3	7	7	7	2	3	3	4	GeBr3	C3v	trigonal[b]	25	4	7	7	7	4	4	4	4
BBr2F	C2v	trigonal	24	3	7	7	7	2	2	4	4	GeCl3	C3v	trigonal	25	4	7	7	7	4	3	3	3
BBr2I	C2v	trigonal	24	3	7	7	7	2	4	4	5	GeF3	C3v	trigonal[b]	25	4	7	7	7	4	2	2	2
BBr3	D3h	trigonal	24	3	7	7	7	2	4	4	4	GeO3	C2v	trigonal[b]	22	4	6	6	6	4	2	2	2
BBrCl2	C2v	trigonal	24	3	7	7	7	2	3	4	4	I2C2	C2		22	7	4	4	7	5	2	2	5
BBrF2	C2v	trigonal	24	3	7	7	7	2	2	2	4	IF3			28	7	7	7	7	5	2	2	2
BCl2I	C2v	trigonal	24	3	7	7	7	2	3	3	5	InBr3	D3h	trigonal	24	3	7	7	7	5	4	4	4
BClI2	C2v	trigonal	24	3	7	7	7	2	3	5	5	InCl3	D3h	trigonal	24	3	7	7	7	5	3	3	3
BCl3	D3h	trigonal	24	3	7	7	7	2	3	3	3	InF3	D3h	trigonal	24	3	7	7	7	5	2	2	2
BClBrI	C2v		24	3	7	7	7	2	3	4	5	InI3	D3h	trigonal	24	3	7	7	7	5	5	5	5
BF2Cl	C2v	trigonal	21	3	7	7	4	2	2	2	2	K2Br2	D2h	alt. paragm	16	1	1	7	7	4	4	4	4
BF3	D3h	trigonal	24	3	7	7	7	2	2	2	2	K2CL2	D2h	alt. paragm	16	1	1	7	7	4	4	3	3
BFBrI	C2v		24	3	7	7	7	2	2	4	5	K2F2	D2h	alt. paragm	16	1	1	7	7	4	4	2	2
BFCl2	C2v	trigonal	24	3	7	7	7	2	2	3	3	K2I2	D2h	alt. paragm	16	1	1	7	7	4	4	5	5
BFClBr	C2v		24	3	7	7	7	2	2	3	4	K2O2	D2h		14	1	1	6	6	4	4	2	2
BFClI	C2v		24	3	7	7	7	2	2	3	5	KBO2	Cs	linear/bent	16	1	3	6	6	4	2	2	2
BI2Br	C2v	trigonal	24	3	7	7	7	2	4	5	5	KNO2	Cs		18	1	5	6	6	4	2	2	2
BI2F	C2v	trigonal	24	3	7	7	7	2	2	5	5	Li2Br2	D2h	alt. paragm	16	1	1	7	7	2	2	4	4
BI3	D3h	trigonal	24	3	7	7	7	2	5	5	5	Li2Cl2	D2h	alt. paragm	16	1	1	7	7	2	2	3	3
BO3			21	3	6	6	6	2	2	2	2	Li2F2	D2h	alt. paragm	16	1	1	7	7	2	2	2	2
BOCl2	C2v		23	7	7	3	6	3	3	2	2	Li2I2	D2h	alt. paragm	16	1	1	7	7	2	2	5	5
BOF2	C2v	trigonal	23	7	7	3	6	2	2	2	2	Li2O2	D2h	alt. paragm	14	1	1	6	6	2	2	2	2
Ba2O2	D2h	alt. paragm	16	2	2	6	6	6	6	2	2	LiBO2	Cs	linear/bent	16	1	3	6	6	2	2	2	2
BaO3			20	2	6	6	6	6	2	2	2	LiNO2	Cs	linear/bent[b]	18	1	5	6	6	2	2	2	2
Be2O2	D2h	alt. paragm	16	2	2	6	6	2	2	2	2	Mg2O2			16	2	2	6	6	3	3	2	2
BeO3			20	2	6	6	6	2	2	2	2	MgO3			20	2	6	6	6	3	2	2	2
Br2C2	D3h		22	7	4	4	7	4	2	2	4	N2CO			20	6	5	4	5	2	2	2	2
BrC2I	D3h		22	7	4	4	7	4	2	2	5	N2F2cis	C2v	linear/bent	24	5	5	7	7	2	2	2	2
BrF3	C2v	trigonal	28	7	7	7	7	4	2	2	2	N2F2tr	C2h	linear/bent	24	5	5	7	7	2	2	2	2
C2Cl2	Dinfh	linear/bent	22	4	4	7	7	2	2	3	3	F4			28	7	7	7	7	2	2	2	2
C2F2	Dinfh	linear/bent	22	4	4	7	7	2	2	2	2	N3F	Cs	trigonal	22	7	5	5	5	3	2	2	2
C2N2	Dinfh	linear/bent	18	4	4	5	5	2	2	2	2	NCl3	C2v	trigonal	26	5	7	7	7	2	3	3	3
C4	Dinfh	linear/bent	16	4	4	4	4	2	2	2	2	NF2Cl	C2v		26	5	7	7	7	2	2	2	3
CBr3	C3v	trigonal	25	4	7	7	7	2	4	4	4	NF3	C3v	trigonal	26	5	7	7	7	2	2	2	2
CCl3	C3v	trigonal	25	4	7	7	7	2	3	3	3	NFCl2	C2v	trigonal	26	5	7	7	7	2	2	3	3
CF2Cl	C2v		25	4	7	7	7	2	2	2	3	NO3	D3h	trigonal	23	5	6	6	6	2	2	2	2
CF3	C3v	trigonal	25	4	7	7	7	2	2	2	2	Na2Br2	D2h	alt. paragm	16	1	1	7	7	3	3	4	4
CFCl2	C2v		25	4	7	7	7	2	2	3	3	Na2Cl2	D2h	alt. paragm	16	1	1	7	7	3	3	3	3
C I3	C3v	trigonal	25	4	7	7	7	2	5	5	5	Na2F2	D2h	alt. paragm	16	1	1	7	7	3	3	2	2
CNF2	C2v	trigonal	23	4	5	7	7	2	2	2	2	Na2I2	D2h	alt. paragm	16	1	1	7	7	3	3	5	5
CO3	C2v	trigonal	22	4	6	6	6	2	2	2	2	Na2O2	D2h	alt. paragm	14	1	1	6	6	3	3	2	2
COCl2	C2v	trigonal	24	7	7	4	6	3	3	2	2	NaBO2	Cs	linear/bent	18	1	3	6	6	3	2	2	2
COF2	C2v	trigonal	24	7	7	4	6	2	2	2	2	NaNO2	Cs		18	1	5	6	6	3	2	2	2
COFCl	Cs	trigonal	24	4	6	7	7	2	2	2	3	P2O2	C2v		22	5	5	6	6	3	3	2	2
CSCl2	C2v	trigonal	24	4	6	7	7	2	3	3	3	P4	Td		20	5	5	5	5	3	3	3	3
CSClBr	Cs	trigonal	24	4	6	7	7	2	3	3	4	PBr3	C3v	trigonal	26	5	7	7	7	3	4	4	4
CSFCl	C2v	trigonal	24	4	6	7	7	2	3	2	3	PCl3	C3v	trigonal	26	5	7	7	7	3	3	3	3
CSeF2	C2v	trigonal	24	4	6	7	7	2	4	2	2	PF2Br	Cs	trigonal	26	5	7	7	7	3	2	2	4

Table 1. (Continued)

mol.	symmetry symbol	structure	n_e	groups 1	2	3	4	periods 1	2	3	4	mol.	symmetry symbol	structure	n_e	groups 1	2	3	4	periods 1	2	3	4
PF2Cl	Cs	trigonal	26	5	7	7	7	3	2	2	3	SbBr3	C3v	trigonal	26	5	7	7	7	5	4	4	4
PF2I	Cs	trigonal	26	5	7	7	7	3	2	2	5	SbCl3	C3v	trigonal	26	5	7	7	7	5	3	3	3
PF3	C3v	trigonal	26	5	7	7	7	3	2	2	2	SbF3	C3v	trigonal	26	5	7	7	7	5	2	2	2
PFCl2	Cs		26	5	7	7	7	3	2	3	3	SbI3	C3v	trigonal	26	5	7	7	7	5	5	5	5
PO3	D3h	trigonal[b]	23	5	6	6	6	3	2	2	2	SeO3	D3h	trigonal[b]	24	6	6	6	6	4	2	2	2
Rb2Br2	D2h	alt. paragm	16	1	1	7	7	5	5	4	4	SeOCl2	Cs	trigonal[b]	26	6	6	7	7	4	2	3	3
Rb2Cl2	D2h	alt. paragm	16	1	1	7	7	5	5	3	3	SeOF2	Cs	trigonal[b]	26	6	6	7	7	4	2	2	2
Rb2F2	D2h	alt. paragm	16	1	1	7	7	5	5	2	2	Si2O2	D2h	alt. paragm	20	4	4	6	6	3	3	2	2
Rb2I2	D2h	alt. paragm	16	1	1	7	7	5	5	5	5	SiBr3	C3v	trigonal	25	4	7	7	7	3	4	4	4
Rb2O2	D2h	alt. paragm[b]	14	1	1	6	6	5	5	2	2	SiCl3	C3v	trigonal	25	4	7	7	7	3	3	3	3
RbBO2	Cs		16	1	3	6	6	5	2	2	2	SiF3	C3v	trigonal	25	4	7	7	7	3	2	2	2
RbNO2			18	1	5	6	6	5	2	2	2	SiI3	C3v	trigonal	25	4	7	7	7	3	5	5	5
S2Cl2	C2	linear/bent	26	6	6	7	7	3	3	3	3	SiO3	C2v	trigonal[b]	22	4	6	6	6	3	2	2	2
S2F2	C2	linear/bent	26	6	6	7	7	3	3	2	2	SiOF2	C2v	trigonal	24	4	6	7	7	3	2	2	2
S2F2	Cs	trigonal	26	6	6	7	7	3	3	2	2	Sn2O2	D2h	alt. paragm	20	4	4	6	6	5	5	2	2
S2O2	C2v	trigonal	24	6	6	6	6	3	3	2	2	SnBr3	C3v	trigonal[b]	25	4	7	7	7	5	4	4	4
S4			24	6	6	6	6	3	3	3	3	SnCl3	C3v	trigonal[b]	25	4	7	7	7	5	3	3	3
SF3	C2v	trigonal	27	6	7	7	7	3	2	2	2	Sr2O2	D2h	alt. paragm	16	2	2	6	6	5	5	2	2
SO3	D3h	trigonal	24	6	6	6	6	3	2	2	2	Tl2Cl2	D2h		20	3	3	7	7	6	6	3	3
SOBr2			26	6	6	7	7	3	2	4	4	Tl2F2	D2h		20	3	3	7	7	6	6	2	2
SOCl2			26	6	6	7	7	3	2	3	3	TlBO2	C2v	trigonal	18	3	3	6	6	6	2	2	2
SOF2	Cs	trigonal	26	6	6	7	7	3	2	2	2	XeO3	C3v	trigonal	26	8	6	6	6	5	2	2	2
Sb2O2	C2v	trigonal[b]	22	5	5	6	6	5	5	2	2	XeOF2	C2v	trigonal	28	8	6	7	7	5	2	2	2

[a] The authors will be grateful for any information about missing symmetries and structures. [b] Structure determined from similar molecules with same symmetry symbol.

representative case of triatomic molecules, by

$$\log {}^{p}K = [A][B][C]/[ABC] \quad (6)$$

where the quantities in brackets are partial pressures, in Pascals, of atoms A, B, and C. This temperature-dependent quantity is related to the energy of atomization by

$$\log {}^{p}K = -(5040/kT)\Delta H_a + 3.41405 + \log (Q_A Q_B Q_C/Q_{ABC}) + 2.5(\log T) + 1.5(\log \mu) \quad (7)$$

where the base of logarithms is 10, μ is the reduced mass, Q is the partition function, and T is the temperature in K. The first numerical constant is the conversion factor from electron-volts to Joules, divided by Boltzmann's constant and the conversion from common to natural logarithms. The second is a combination of physical constants and the reduced mass conversion of units. The 2.5 and the 1.5 result from exponents in the nonlogarithmic form of the equation. The molecule with the most negative equilibrium constant is the most stable against atomization.

Stability against excitation from the ground state of the molecule will be greater, the temperature and statistical weights being equal, if there are fewer low-lying excited states; hence the molecule with the smallest internal partition functions has the greatest stability. There are thermodynamic relations between the partition function and the entropy which imply that the molecule with the greatest stability against excitation will, for a given temperature, have the smallest entropy value.

Stability against ionization of the molecule from its ground state is obviously measured by the ionization potential *IP*. The molecule most stable against ionization will have the highest value of *IP*.

4. DATA AND ASSOCIATED ERRORS

The investigations reported in this paper depend exclusively on critical-table data. Total entropies at 298 K (hereafter indicated by S_{298}) for triatomic molecules were gleaned from the JANAF Tables of 1971 to 1982.[3] The average of the 95% confidence-limit errors which are given for many of the data is approximately l%. The same data for tetra-atomic molecules were obtained from the JANAF Tables of 1985.[4] Internal entropies at an arbitrarily chosen 1000 K [indicated by $S_{1000}(internal)$] also came from those tables and the use of a conversion formula.[1] When results for both entropies are described, the generic symbol *S* is used. The use of difference sources (in this case, different editions of JANAF) reflects the fact that the work was done in several stages.

Heats of atomization $\Delta H_a(Gurvich)$ were taken from the eight-volume USSR thermodynamic tables.[5] The average of the 95% confidence-limit errors associated with individual experimental values is 2.63%. Ionization potentials (*IP*) were taken from Gurvich et al.[6] The average of the 95% confidence-limit errors is 4.23%.

Common-log total partition functions at l kK (log *Q*) and common-log equilibrium constants at 1 kK (log ${}^{p}K$) were obtained from Sauval and Tatum.[7] They were computed by them for triatomic molecules at the same time as those for diatomic molecules. The estimate of the errors of the "best" values of log *Q* is near 0.01, i.e., from 0.13% to 0.28%. The errors associated with his log *Q* have since been verified independently or even lowered.[8] Heats of atomization, from the literature and needed for these computations, were also obtained from Sauval and Tatum. The *estimated* 95% confidence-limit errors for $\Delta H_a(Sauval)$ are 10% and in a few cases 15%. Of course $\Delta H_a(Gurvich)$ and $\Delta H_a(Sauval)$ describe the same molecular behavior. When results from both sources are described, the generic symbol ΔH_a is used.

The triatomic-molecular data can be found in ref 8. The tetra-atomic tabulated data are presented in Tables 1 and 2. The tetra-atomic molecular structures were found in Krasnov[9] and in the JANAF tables.[4] The molecules were used in the

Table 2. Molecular Properties

species	ΔHa (*Gurvich*)	±	S_{298}	*IP*	±	species	ΔHa (*Gurvich*)	±	S_{298}	*IP*	±	species	ΔHa (*Gurvich*)	±	S_{298}	*IP*	±
Al2C2	1517.062	60				CO3						NF2Cl					
Al2O2	1550	60	280.98			COCl2	1415.015	2	283.796			NF3	828.628	2	260.773		
Al2S2	1100	20				COF2	1749.42	6	258.887			NFCl2				9.34	0.05
AlBO2			269.567			COFCl			277.012			NO3			252.619		
AlBr3	1067.115	6	349.441			CSCl2						Na2Br2	914.372	6	348.949		
AlCl3	1268.506	6	314.494	7.8		CSClBr						Na2Cl2	1015.766	6	325.432		
AlF2Cl	1597.516	6	297.839			CSFCl						Na2F2	1199.076	6	287.366		
AlF2O	1828.679	60	292.705	11.6		CSeF2						Na2I2	780.852	6			
AlF3	1765.389	6	276.691			ClC2Br						Na2O2	827	20			
AlFCl2	1432.861	6	311.403			ClF3	511.946	6	281.6	9.7		NaBO2			287.248		
AlI3	834.835	6	373.616	13.2		ClNCO				11.1		NaNO2	1233.771	6			
AlO3				9.7		ClNO2	1066.118	2	272.187			P2O2				12	
AlOCl2	1563.369	60				ClO3						P4	1196.012	2	279.992		
As2O2						Cs2Br2	939.874	20				PBr3			348.234		
AsBr3						Cs2Cl2	1036.268	20	383.445			PCl3	960.725	2	311.682		
AsCl3						Cs2F2	1197.578	20	352.319			PF2Br					
AsF3				9.98	0.05	Cs2I2	818.354	20				PF2Cl	1320	20		9.67	0.05
AsI3				11.55		Cs2O2	890	20				PF2I				11.7	
AsO3						CsBO2	1815.333	20				PF3	1499.127	2	273.064		
B2O2	2077.566	20	242.597			CsF2						PFCl2	1140	20			
B2S2	1534.57	60				CsNO2	1247.991	6				PO3					
BBeO2			265.396			FClO2				9.71		Rb2Br2	934.23	20			
BBr2Cl			321.889			FNO2	649.175	20	260.25			Rb2Cl2	1018.624	20			
BBr2F			309.994			Ga2O2						Rb2F2	1168.934	20			
BBr2I						GaBr3						Rb2I2	805.71	20			
BBr3	1096.769	6	324.307			GaCl3	1059.518	6		9.5		Rb2O2	867	20			
BBrCl2			310.4	9.1		GaF3	1420.716	20				RbBO2	1811.675	20			
BBrF2			286.387			GaI3						RbNO2	1229.471	20			
BCl2I				8.8		GaO3				15.6		S2Cl2			327.215		
BClI2						Ge2O2						S2F2			294.082		
BCl3	1322.334	6	290.169	8.64		GeBr3						S2F2			292.827		
BClBrI						GeCl3	1047.897	60				S2O2					
BF2Cl	1719.856	6	274.83			GeF3	1350.667	20				S4	960.14	6	310.646		
BF3	1924.966	6	254.355			GeO3						SF3	1006.61	20	286.174		
BFBrI						I2C2				11.1		SO3	1405.287	0.6	256.769		
BFCl2	1517.934	6	287.675			IF3						SOBr2					
BFClBr				9.25	0.04	InBr3						SOCl2					
BFClI				9.3		InCl3	975.28	6				SOF2	1256.118	60	279.132		
BI2Br						InF3	1327.178	60				Sb2O2					
BI2F						InI3						SbBr3					
BI3	856.006	6	348.721			K2Br2	936.628	6	376.23			SbCl3					
BO3				11.5		K2CL2	1031.022	6	352.864			SbF3				9.03	0.05
BOCl2	1406.023	60		9.5		K2F2	1189.332	6	319.955			SbI3					
BOF2	1791.333	60	267.842	10.2		K2I2	807.108	6	395.772			SeO3					
Ba2O2						K2O2	858.348	20				SeOCl2					
BaO3				9.9		KBO2	1808.579	20	297.408			SeOF2					
Be2O2	1518.072	6	247.611			KNO2	1242.011	6				Si2O2					
BeO3				9.8		Li2Br2	1030.316	6	314.534			SiBr3	1002.9	6	351.774		
Br2C2						Li2Cl2	1149.71	6	288.767			SiCl3	1199.527	20	318.189		
BrC2I						Li2F2	1401.02	6	258.627			SiF3	1670.492	60	282.381		
BrF3	594.562	6	292.397			Li2I2	887.796	6	330.628			SiI3	779.156	20	378.305		
C2Cl2	1464.922	60	272.027	9.4		Li2O2	1084.036	20	273.562			SiO3					
C2F2	1723.92	20	244.06	10.8		LiBO2	1862.545	6	274.727			SiOF2			271.247		
C2N2	2056.852	2	241.565			LiNO2	1319.871	6				Sn2O2				10.72	
C4	1819.74	60	228.322			Mg2O2						SnBr3					
CBr3	807.706	20				MgO3				10.1		SnCl3					
CCl3	989.164	6	296.833	11		N2CO						Sr2O2					
CF2Cl	1258.316	20				N2F2	1029.126	6	259.81			Tl2Cl2	862.04	6		9.4	
CF3	1412.202	6	265.082			N2F2tr	1023.586	6	262.66			Tl2F2	1027.35	6			
CFCl2	1131.274	20				N2F2tr	1029.186	6		12.9		TlBO2					
CI3	622.674	60				N3F				9.7		XeO3	207.693	60			
CNF2						NCl3						XeOF2				13	

analysis if at least one datum for at least one property and the structure are known.

5. GRAPHICAL ANALYSES FOR TRIATOMIC MOLECULES

5.1. Dependence upon the Three Atomic Group Numbers, Using Tabulated Data for ΔH_a(*Gurvich*), ΔH_a(*Sauval*), log *Q*, S_{298}, and *IP*. The similarities of acyclic main-group triatomic molecules were first seen when tabulated data for molecular properties were plotted in (C_1,C_2,C_3) coordinates for each (R_1,R_2,R_3). Figure 1 shows such a plot for $R_1 = R_2 = R_3 = 2$. The symbols for the points were chosen to represent data in equally large ranges of log *Q*. The data exist only for a fraction of the possible positions

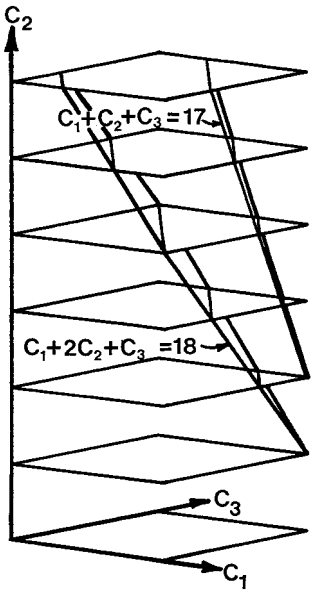

Figure 2. Schematic of the (C_1,C_2,C_3) coordinate system of Figure 1 showing the bounding plane $g(C) = 18$ and also the plane on which are found molecules with 17 valence electrons. Reprinted from *J. Mol. Struct.* vol. 382, Cavanaugh, R.; Marsa, R.; Robertson, R.; Hefferlin, R. Adjacent DIM Isoelectronic Molecules and Chemical Similarity among Triatomics, p 140, (2000), with permission from Elsevier Science.

in the coordinate system; furthermore, they are not randomly distributed but tend to be clustered in $g(C) \geq 18$ (e.g., to the right of OBO, NCN, FBeF). The figure shows that there exists a line of molecules (indicated by hexagons) with similar values of log Q for $g(C) = 18$. Actually, there is a *plane*, in and out of the figure and containing this line, on which all the molecules have data represented by hexagons except for LiON, off the symmetry plane at $|C_1 - C_2| = |C_3 - C_2| = 2$. The spatial relationships of all these planes and lines are shown in Figure 2. Very similar comments apply to plots for S_{298}[1] and ΔH_a(*Sauval*).[10] These lines of molecules are unexpected, as they are not traditional isoelectronic sequences.

The plots for (R_1,R_2,R_3) > (2,2,2) have far fewer data than shown in Figure 1. There are enough, however, to hint that the hypothesis is confirmed for molecules which have (R_1,R_2,R_3) = (2,3,2) and are close to the symmetry plane of the space (Figure 3).

In the following summary of this section, each conclusion is provided with a boldface item number (not identifying a molecular structure). This scheme will be adhered to in later sections and is collated in Table 3.

The data for (item **1**) ΔH_a(*Gurvich*), (item **2**) ΔHa(*Sauval*), (**3**) log Q, and (**4**) S_{298} support the hypothesis. This support is under the conditions that (**5**) $R_1 = R_2 = R_3 = 2$, that (**6**) the molecules have $g(C) = 18$, and that (**7**) they lie within $|C_1-C_2|$ and $|C_3-C_2| \leq 1$. Further, (item **1** again) the data for ΔH_a somewhat support the hypothesis if (**5** widened) (R_1,R_2,R_3) has the higher value (2,3,2), if (**6** widened) the molecules have $g(C) \geq 18$, and if (special case of **7**) they lie at C_1-$C_2 = C_3$-$C_2 = 0$. Finally, (**8**) data for *IP* do not support the hypothesis.

5.2. Dependence on the Total Number of Valence Electrons, Using Tabulated Data for ΔH_a(*Sauval*), log ^pK, and log *Q*. Figure 4 shows tabulated data plotted against an

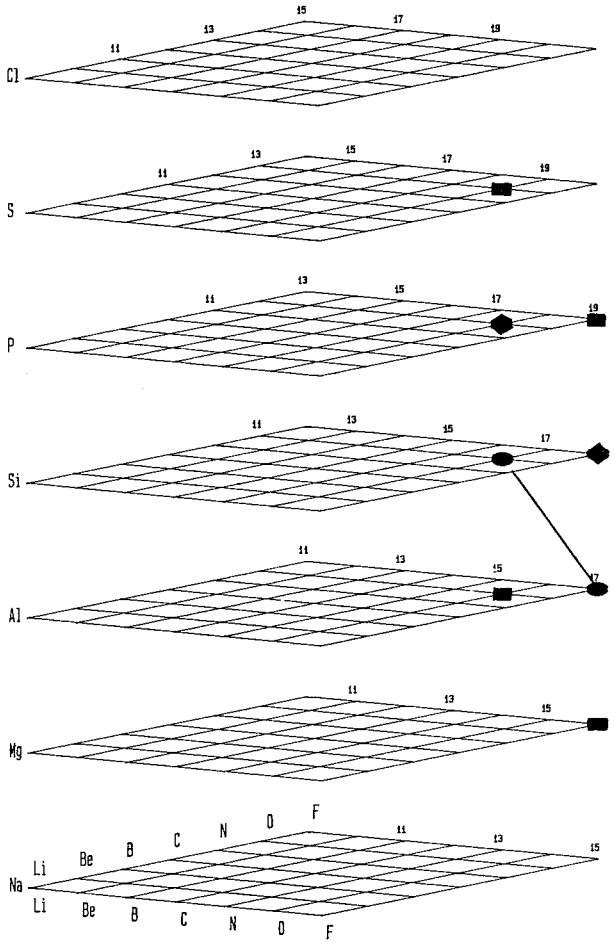

Figure 3. Tabulated data for ΔH_a(*Sauval*) shown in the same coordinates as Figure 1 but for (R_1,R_2,R_3) = (2,3,2). The data are separated into ranges as follows: ■, from 9.50 to 10.99; ⬢, from 11.00 to 12.49; and ●, from 12.50 to 13.99 eV. There are eight data, and all of them confirm the hypothesis. Specifically, there are four pairs of data, each with a value of $g(C)$ falling into one range (FMgF and OAlO, FAlF and OSiO as indicated by the solid line, FSiF and OPO, and FPF and OSO). The similar graph for S_{298} has four same-$g(C)$ pairs, one of which (FSiF and OPO) has members in one range value and another of which (FMgF and OAlO) has members in adjacent ranges. The graph for log Q has four same-$g(C)$ pairs, none of which has members in one range but three of which (FMgF and OAlO, FSiF and OPO, and FPF and OSO) have members belonging in adjacent ranges.

independent variable which is piecewise directly proportional to the numbers of valence electrons, n_e, in the triatomic molecules. Similar plots for ΔH_a(*Gurvich*), S_{1000}(*internal*), and *IP* can be found in ref 8. These plots were originally intended to test periodicity by showing that the peak of each swarm of points lies at the same total number of molecular-atom valence electrons, but upon closer analysis they reveal consistent discrepancies.

The points in each swarm pertaining to the dicarbides, dinitrides, dioxides, and difluorides of series of atoms can be connected in cases where there are enough data. For example, in (R_1,R_2,R_3) = (2,2,2), the dinitrides are NBeN, NBN, NCN, N_3, and NON. (Data for all properties do not exist for all of them.) The connecting lines seem to show that the most stable molecules of such series lie at progressively increasing n_e values.

After deleting data for those molecules that have asymmetric structures, e.g., NNO, those for which the structures

Table 3. Molecular Properties and the Conditions under Which They Confirm the Similarity of Data in Adjacent-DIM Isoelectronic Series[a]

section	item(s)	property(ies)	condition(s)	item(s)	confidence
5.1	**1**, **2**a	ΔH_a	f(R) = 8: $(R_1,R_2,R_3) = (2,2,2)$	**5**	
	3	log *Q*	g(C) = 18	**6**	high, Figure 1
	4	*S*	$\lvert C_1-C_2\rvert$ and $\lvert C_3-C_2\rvert \le 1$	**7**	
5.2	**1**, **2**	ΔH_a	f(R) = 12: $(R_1,R_2,R_3) = (2,3,2)$	**5** widened	
	9	log ^pK	g(C) ≥ 18	**6** widened	high, Figure 4
			$\lvert C_1-C_2\rvert$ and $\lvert C_3-C_2\rvert = 0$	**7** special	
5.3	**1**, **2**	ΔH_a	f(R) = 8: $(R_1,R_2,R_3) = (2,2,2)$	**5**	lower, data
	4	*S*	$\lvert C_1-C_2\rvert$ and $\lvert C_3-C_2\rvert \le 1$	**7**	from networks
5.4	**1**, **2**	ΔH_a	f(R) = 8: $(R_1,R_2,R_3) = (2,2,2)$	**5**	lower, data
			g(C) ≥ 23, $C_2 \ge 4$	**6** restricted	from networks
5.4	**1**, **2**	ΔH_a	f(R) increasing	**5**	lower, data
			ranges of n_e & g(C) decreasing	**6** restricted	from networks
5.4	**1**, **2**	ΔH_a	18 ≤ f(R) ≤ 30	**5**	lower, data
			g(C) ≥ 25	**6** restricted	from networks
6.1	**1**, **2**	ΔH_a	f(R) = 8 or all f(R) together	**5** much wider	high,
			g(C) = 18 and 24	**6**	statistics
6.1	**4**, **10**a	*S*	f(R) = 8 or all f(R) together	**5** much wider	high,
			g(C) = 18 and 24	**6**	statistics
			$\lvert C_1-C_2\rvert$ and $\lvert C3-C2\rvert \le 1$	**7**	
6.1	**3**	log *Q*	all scaled f(R) together	**5** much wider	high,
			g(C) = 18	**6**	statistics
6.2	**1**, **2**	ΔH_a	all scaled f(R) together	**5** much wider	high,
	4, **10**	*S*			statistics
5, 6	**8**	*IP*	does not conform to hypothesis		high

[a] Heat of atomization data are not separated by author; entropy data are not identified by temperature or by their inclusion of translational motion. Items **5** to **7** pertain to the conditions in column 4.

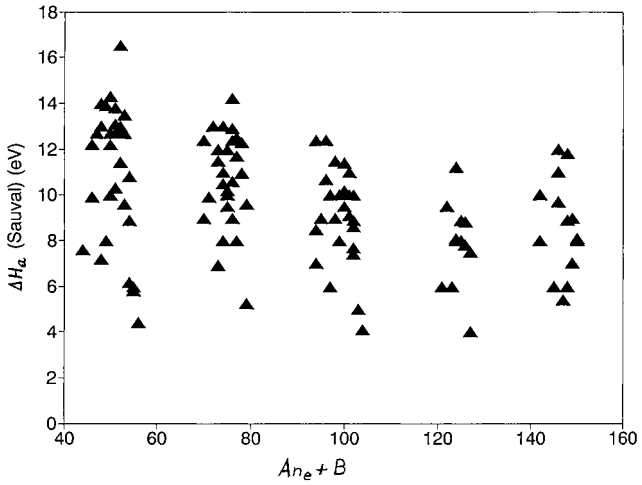

Figure 4. ΔH_a(*Sauval*) plotted on a number line which is piecewise directly proportional to n_e where the swarms of symbols are found in ref 8. The most stable molecules and their abscissae in the five swarms of symbols are the following: OCO, 52; SCO, 76; SCS, 100; FSrF, 124; and FBaF, 148. All of these molecules, with atoms from different combinations of period numbers have 16 valence electrons. This phenomenon is evidence of the periodicity of triatomic molecular data. A surprising feature, described in the text and shown in the next figures, emerges when the symbols for certain classes of molecules, such as dioxides, are connected.

are unknown, and those for which no local extremum for the data exists, there remain four pertinent cases where it is possible to identify the most stable molecule.

ΔH_a(*Sauval*) has peaks at OCO and FBF (n_e = 16 and 17), respectively, as shown in Figures 5 and 6 and in Table 4, and at OSiO and FAlF (also at n_e = 16 and 17). Log ^pK has two-molecule minima that include the same n_e values. The named molecules all have $g(C) = 20$. Table 4 shows that the minima for log *Q* fail to fall at the same values of n_e for two values of (R_1,R_2,R_3) but do so for $(R_1,R_2,R_3) = (2,5,2)$.

The conclusions drawn now will continue the item numbering introduced at the end of section 5.1. The swarm graphs evidence that (**1**) ΔH_a(*Sauval*) and (**9**) log ^pK have similar (in this case, extremal) values in isoelectronic pairs according to the adjacent-DIM model and not according to the traditional model. This evidence exists for molecules that (**6** widened) have $g(C) = 20$ and (**5** widened) lie in $(R_1,R_2,R_3) = (2,2,2)$ and (2,3,2). Further, the molecules used to obtain these conclusions are symmetric, which ensures item (**7**). In other words, the swarm graphs for ΔH_a give some support to the hypothesis that isoelectronic adjacent-DIM molecules tend to have the similar (in this case, extremal) observables. The swarm graph for log *Q* [in a counterexample to item (**3**)] does not support the hypothesis except in one of the three cases. The data for ionization potential (**8**) show that it does not conform to the hypothesis.

It should be recalled that the hypothesis concerns *fixed-row* molecules. The fact that similar conclusions can be drawn from molecules with *different* rows indicates that the hypothesis may apply to isovalent as well as to isoelectronic adjacent-DIM molecules.

5.3. Dependence upon Atomic Group Numbers, Using Neural-Network Globally Predicted Data for ΔH_a(*Gurvich*), ΔH_a(*Sauval*), S_{298}, log *Q*, and *IP*. Figures 1 and 3, and others like them, are very sparsely populated with data. To reveal whatever trends they might have, neural networks were trained with the data to make predictions of molecular data at all positions. These global predictions were plotted in the coordinates of Figures 1 and 3 on the central plane.

The plots, shown in Figures 7 and 8, demonstrate (with the caveat that the data have been smoothed by a network) that adjacent-DIM isoelectronic sequences often contain molecular data for (**1**) ΔH_a(*Gurvich*), (**2**) ΔH_a(*Sauval*), and (**4**) S_{298} that have similar values. The plots were made only for $R_1 = R_2 = R_3 = 2$ (conforming to **5**), and the data are placed only on the center plane (conforming to **7**). The most

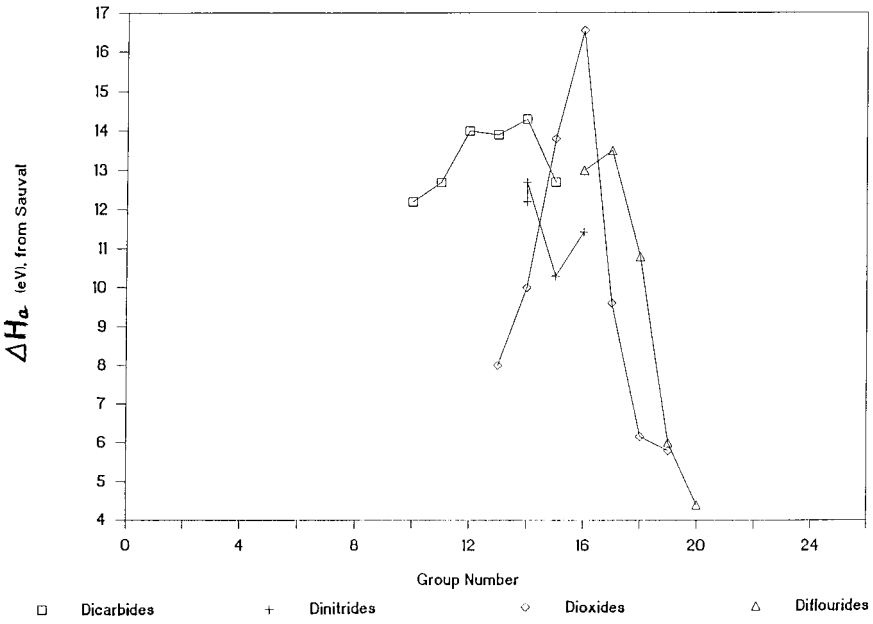

Figure 5. The symbols in the first swarm of Figure 4, i.e., for $(R_l,R_2,R_3) = (2,2,2)$. Classes of molecules are connected together. For dicarbides, the maximum lies at $n_e = 14$. For dinitrides, it is impossible to know where the maximum value lies. For dioxides, the maximum is at $n_e = 16$. For difluorides, the maximum is at $n_e = 17$.

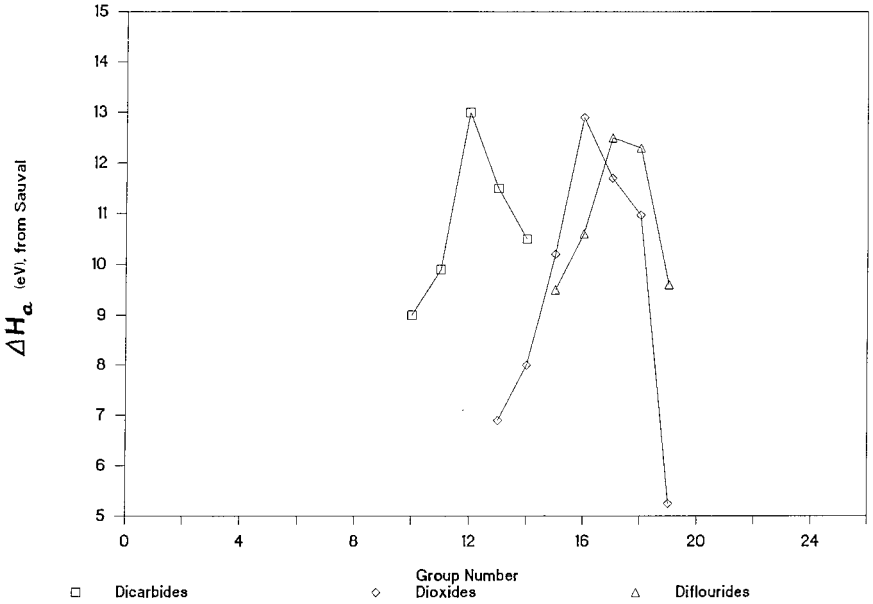

Figure 6. The same as Figure 5 except for molecules with $(R_l,R_2,R_3) = (2,3,2)$. The central atoms now are from row 3.

Table 4. Locations of Most Stable Dioxides and Difluorides in the First (2,2,2) and Second (2,3,2) Swarms of Figure 5 and Others

		n_e			
property	(R1,R2,R3)	16	17	18	g(C)
ΔH_a	(2,2,2)	OCO	FBF		20
	(2,3,2)	OSiO	FAlF		20
log pK	(2,2,2)	OCO	FBF		20
	(2,3,2)	OSiO	FAlF		20
log Q	(2,2,2)	OCO		FCF	20,22
	(2,3,2)	OSiO		FSiF	20,22
	(2,5,2)	OSnO	FInF		20

striking features in the plots for log Q and IP are vertical series of molecules with data with similar magnitudes and hence with the same symbols (in disagreement with **3** and in agreement with **8**, respectively).

5.4. Dependence upon Atomic Group Numbers, Using Neural-Network Globally Predicted Data for ΔH_a(*Gurvich*) and S_{1000}(*internal*). The triatomic-molecular periodic-system coordinates R_l,C_1,R_2,C_2,R_3,C_3 described in section 3 can be collapsed into three-dimensional coordinates for visualization (with the creation of redundancies). A very useful set of collapsed coordinates is $f(R) = R_1R_2 + R_2R_3$, $n_e = C_1 + C_2 + C_3$, and C_2.[8] Global neural-network predictions were plotted on n_e,C_2 coordinates for each value of the coordinate $f(R)$. Figure 9 shows the data for S_{1000}(*internal*) for molecules with $f(R) = 8$, that is $(R_l,R_2,R_3) = (2,2,2)$. The figure illustrates the value of global predictions made by a neural network trained on very few data.

Figure 10 shows (item **1**) ΔH_a(*Gurvich*) for (item **5**) $(R_l,R_2,R_3) = (2,2,2)$ plotted on the same axes. The figure shows a region where the contours have a slope of about

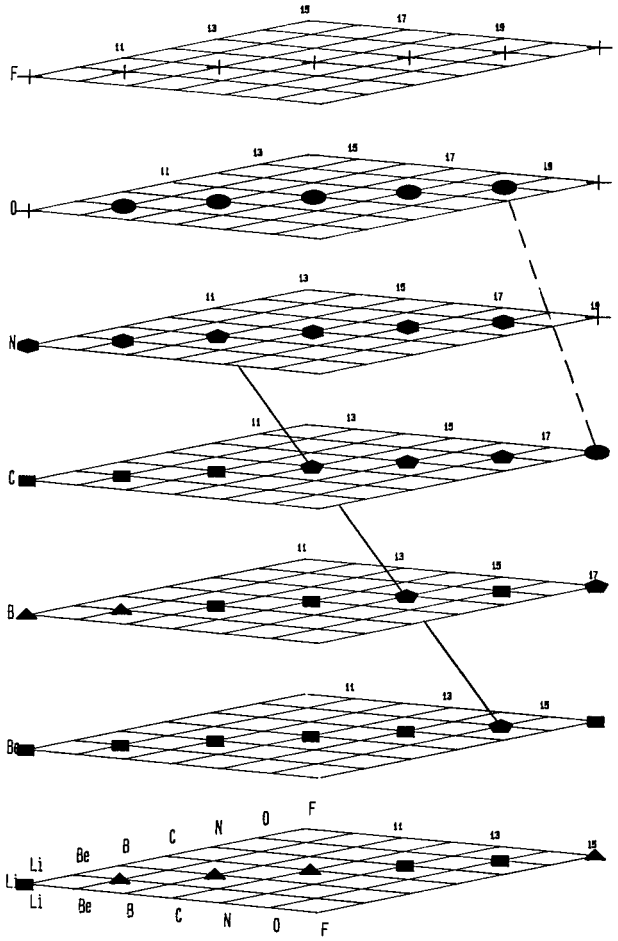

Figure 7. Same as Figure 1 except for globally predicted data of ΔH_a(*Gurvich*) on the center plane. The data are separated into equal-range ranges as follows: +, from 390.00 to 792.99; ●, from 793.00 to 973.99; ⬢, from 974.00 to 1118.99; ⬟, from 1119.00 to 1276.99; ■, from 1277.00 to 1428.90; ▲, from 1429.00 to 1780.99 kJ/mol. Data for molecules with Li or F as the central atom must be ignored. When this is done, there remains one four-molecule series with constant $g(C)$ having data in the same range, indicated by the solid line, and several shorter ones. There are some two-molecule series with constant n_e having data in the same range, one indicated with a dashed line.

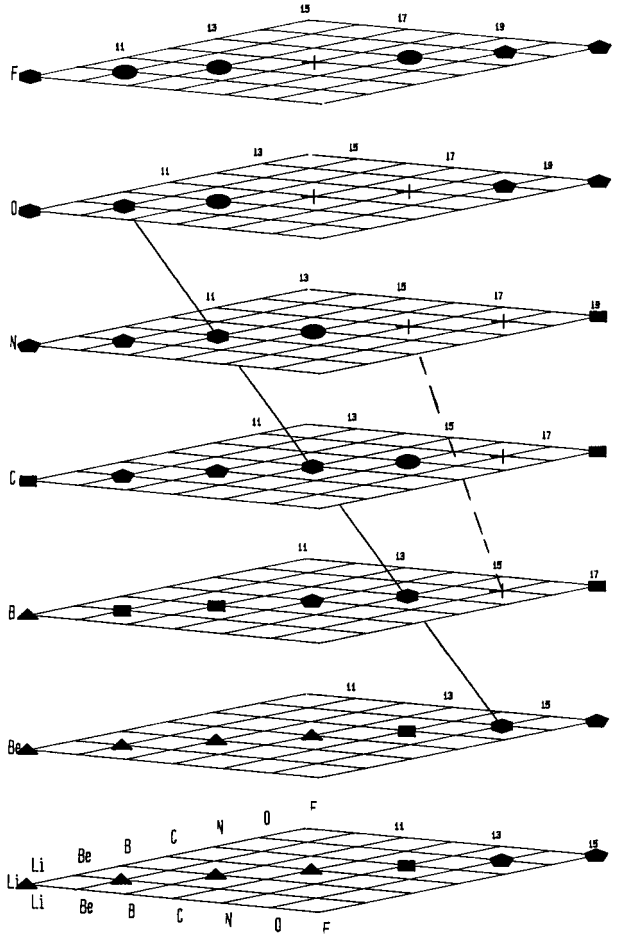

Figure 8. The same as Figure 7 except for S_{298}. The data are separated into equal-range ranges as follows: +, from 23.980 to 25.5508; ●, from 25.5509 to 26.5526; ⬢, from 26.5527 to 27.4507; ⬟, from 27.4508 to 29.4661; ■, from 29.4662 to 31.9229; and ▲, 31.9300 to 39.2800 J/(K·mol). There is one five-molecule series with constant $g(C)$ having data in the same range, indicated by the solid line, and several shorter ones. There are some two-molecule series with constant n_e having data in the same range, one indicated with a dashed line.

−45° degrees on the n_e,C_2 plane, i.e., the region (restricted case of item **6**) $n_e + C_2 = g(C) \geq 23$ and $C_2 \geq 4$. When $f(R)$ increases, the region shrinks until when $18 \leq f(R) \leq 30$, the model applies in the very limited region $g(C) \geq 25$ (restricted case of item **6**).

6. STATISTICAL ANALYSES

6.1. Statistical Analysis for Acyclic Triatomic Molecules Using ΔH_a(*Sauval*), S_{000}(*internal*), log *Q*, and *IP*. A statistical analysis was performed[1] to test the hypothesis, to 95% confidence, that the variances of adjacent-DIM-molecular data are significantly less than variances of data for equal numbers of molecules selected at random, using the χ-squared test. For ΔH_a (*Sauval*), when $(R_1,R_2,R_3) = (2,2,2)$, the adjacent-DIM molecular series $g(C) = 18$ and 24 passes the test, supporting item (**2**) under conditions (**5**) and (**6**). For S_{1000}(*internal*), item (**10**), the series $g(C) = 18$ and 22 pass if $|C_1\text{-}C_2|$ and $|C_3\text{-}C_2| \leq 1$, supporting items (**6**) and (**7**). However, for the same properties, when all (R_1,R_2,R_3) are scaled and lumped together, the results are essentially the same, which supports item (**5** widened even more). For log *Q*, item (**3**), with all (R_1,R_2,R_3) scaled and lumped together (**5** widened as above), only the series g(C) = 18 (**6**) passes the test. For *IP*, whether (R_1,R_2,R_3) is limited to (2,2,2) or all (R_1,R_2,R_3) are lumped together, no series passes, thus supporting item (**8**).

6.2. Statistical Analysis for Tetra-Atomic Molecules, Using Tabulated Data for ΔH_a(*Gurvich*), S_{298}, and *IP*. A statistical analysis[12] was performed to test the hypothesis, to 90% confidence, that the variances of tetra-atomic adjacent-DIM molecular series are significantly less than the variance of the population (where the population consists of all available data except those in the series being tested). The analysis was also performed for molecules in isoelectronic sequences to provide a comparison for the results of the analysis.

Table 5 shows that for ΔH_a(*Gurvich*), S_{298}, and *IP*, the majority of the adjacent-DIM molecular series have smaller variance than the population. However, according to the *t* test, few of these differences in variance are statistically significant. The results for traditional isoelectronic sequences are similar.

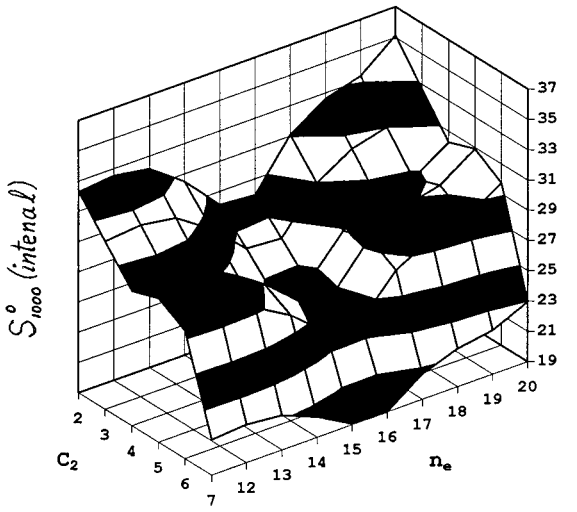

Figure 9. Neural-network predicted S_{1000}(*internal*), in J/(K·mol), plotted against n_e and C_2 for (R_1,R_2,R_3)=(2,2,2), i.e., for $f(R) = 8$. The figure shows a deep valley of stability at $n_e = 16$ (an isoelectronic sequence), which includes the very stable molecules OCO, FBeF, ONF, and others. It shows a small hanging valley for $C_2 = 4$, i.e., for molecules whose central atom is carbon.

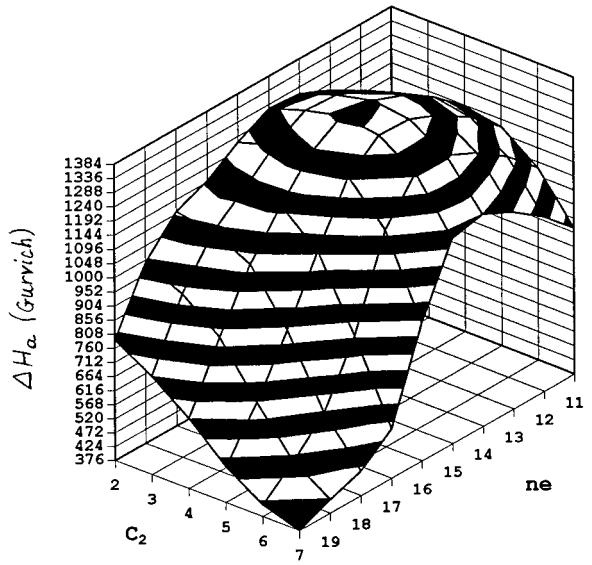

Figure 10. Same as Figure 9 except for global predictions of ΔH_a(*Gurvich*) in kJ/mol. The region of immediate interest; with contours oriented at close to a 45-degree angle with respect to the coordinates, is closest to the reader and bounded by $n_e + C_2 = g(C) \geq 23$ and $C_2 \geq 4$. On this boundary lie CFN (at $n_e = 16$ and $C_2 = 7$, impossible), NOO and FOC (at $n_e = 17$ and $C_2 = 6$), FNO (at $n_e = 18$ and $C_2 = 5$), FCNe (at $n_e = 19$ and $C_2 = 4$, possible only under special conditions), and NeCNe (at $n_e = 20$ and $C_2 = 4$, also possible only under special conditions). Other molecules at intersections of mesh lines can be inferred from n_e and C_2; some of them are shown explicitly shown in the similar Figure 8 of ref 11.

Next, a χ-squared test was performed on the adjacent-DIM molecular series variances, keeping in mind the null hypothesis, i.e., that half of the variances should be smaller than the population variance and half should be larger. This null hypothesis is clearly rejected, as shown by the variances in Table 4, for (**2**) ΔH_a(*Gurvich*), and (**4**) S_{298}; however, the null hypothesis is not rejected for (**8**) *IP*.

The *number* of variances smaller than the population variance is statistically significant for ΔH_a(*Gurvich*) and S_{298}. The tetra-atomic molecules included atoms from various period combinations, thus collaborating the expanded form of item (**5**). The number of variances smaller than the population variance is not statistically significant for (**8**) *IP*.

Table 5. Statistical Analyses of Adjacent-DIM and Traditional Isoelectronic Sequences of Tetra-Atomic Molecules

isoelectronic series				adjacent-DIM molecular series			
property	n_e	$\sigma_{population}$	σ	symm	g(C)	$\sigma_{population}$	σ
ΔH_a	14	19.63	15.58	trigonal	30	18.51	17.67
(*Gurvich*)	16	19.77	18.81		33	18.60	16.83
	18	19.37	18.20		36	18.64	15.48
	20	19.63	11.67	linear/bent	25	17.96	4.70
	22	19.51	11.37		27	17.93	21.35
	23	19.39	13.13	rhomboid	14	18.48	9.57
	24	19.91	17.90		16	18.89	13.78
	25	19.78	16.82				
	26	19.43	19.85				
	28	19.37	15.06				
χ^2	9/10a	2.71b	6.40c		6/7	2.71	3.57
S_{298}	16	5.62	6.92	trigonal	30	6.24	5.48
	18	6.09	4.59		32	6.17	2.75
	21	6.13	1.62		33	6.08	6.52
	22	6.10	3.74		36	6.15	5.91
	23	6.13	4.06		38	6.20	2.38
	24	6.32	5.62		42	6.18	2.32
	25	6.05	6.27	linear/bent	25	6.20	3.05
	26	6.19	5.23		27	6.10	3.60
	28	6.15	2.32		33	6.13	3.74
				rhomboid	16	5.66	6.59
χ^2	7/9	2.71	2.78		8/10	2.71	3.60
IP	20	1.28	0.66	trigonal	29	1.25	0.84
	21	1.14	1.57		30	1.19	1.23
	22	1.27	0.94		36	1.27	1.02
	23	1.26	0.93	linear/bent	32	1.25	0.84
	24	1.19	1.23		33	1.26	0.86
	26	1.28	0.99	rhomboid	20	1.25	0.81
	28	1.21	1.24				
χ^2	4/7	2.71	0.14		5/6	2.71	2.67

[a] Fraction of cases where σ is less than σ population. [b] Expected value for χ^2 test, 90% confidence and one degree of freedom. [c] Observed value pertaining to cases with smaller σ.

7. SUMMARY

The items of evidence presented throughout this paper, and summarized in sections 5.5 and 6.2, constitute an accumulation of indirect proofs for the hypothesis that "fixed-row molecules with the same number of electrons, computed by the adjacent-DIM model, have data more similar than do molecules selected at random". No rigorous proof based on first principles yet exists. Since the boundaries for the phenomenon are much more clearly defined by this paper, search for a rigorous proof may be more attractive.

The work reported here shows that molecules in isoelectronic adjacent-DIM series can, in a certain domain of chemical space (the space R_1, R_2, ... R_N, C_1, C_2, ... C_N) and for certain properties, join isoelectronic and isovalent series as a useful tool for the estimation of data for target molecules. The domain and the properties are summarized in the next two paragraphs.

For acyclic triatomic molecules containing highly electronegative atoms whose outer atoms come from groups differing by at most two units, the tool can be used for heats of atomization ΔH_a and for equilibrium constants log ^pK. It can be used with greatest confidence if all three atomic constituents come from rows 2 or 3 and will lesser confidence

otherwise; these molecules comprise a large fraction of those of interest in organic chemistry. Under the same conditions it also applies to entropies S and to partition functions log Q but with much larger uncertainty.

For tetra-atomic molecules, the tool can be used for ΔH_a and S. Since there are correlations between gas-phase molecular properties such as heat of atomization and equilibrium constant and entropy and partition function, it can be presumed that the tool might be useful also for log Q and log ^pK.

ACKNOWLEDGMENT

We are grateful to Dr. M. W. Chase (National Institute of Standards and Technology), Dr. E. Babaev (Moscow State University), Dr. Mitchell Thiel, and Dr. Norman Peek for insights on the structures of the tetra-atomic molecules. Rick Cavanaugh, Chris Carlson, Brian Hartman, Katie Linderman, and Robert Marsa prepared the graphics in partial fulfillment of their undergraduate researches at SAU.

REFERENCES AND NOTES

(1) Cavanaugh, R.; Marsa, R.; Robertson, J.; Hefferlin, R. Adjacent DIM Isoelectronic Molecules and Chemical Similarity among Triatomics. *J. Mol. Struct.* **1996**, *382*, 137−145. Available by request at hefferln@southern.edu.

(2) Hefferlin, R., Matrix-Product Periodic Systems of Molecules. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 314−317.

(3) Stull, D. R.; Prophet, H. *JANAF Thermochemical Tables*; Office of Standard Reference Data, National Institute of Standards and Technology, U.S. Government Printing Office: Washington, DC, 1971. Chase, M. W.; Curnutt, J. L.; Hu, A. T.; Prophet, H.; Syverud, A. N.; Walker, L. C. JANAF Thermochemical Tables, 1974 supplement. *J. Phys. Chem. Ref. Data*, **1974**, *3*, 311−480. Chase, M. W.; Curnutt, J. L.; Prophet, H.; McDonald, R. A.; Syverud, A. N. JANAF Thermochemical Tables, 1975 supplement. *J. Phys. Chem. Ref. Data* **1975**, *4*, 1−175. Chase, M. W.; Curnutt, J. L.; McDonald, R. A.; Syverud, A. N. JANAF Thermochemical Tables, 1978 supplement. *J. Phys. Chem. Ref. Data* **1978**, *7*, 793−940. Chase, M. W.; Curnutt, J. L.; Downey, J. R.; McDonald, R. A.; Syverud, A. N.; Valenzuela, E. A. JANAF Thermochemical Tables, 1982 supplement. *J. Phys. Chem. Ref. Data* **1982**, *11*, 695−940.

(4) Chase, M. W.; Davies, C. A.; Downey, J. R., Frurip, D. J.; McDonald, R. A.; Syverud, A. N. JANAF Thermochemical Tables, Third Edition, Supplement 1. *J. Phys. Chem. Ref. Data* **1985**, *14*.

(5) *Thermodinamicheskie Svoista Individual'nikh Veschestv*; Gurvich, L. V., Ed.; Nauka: Moscow, 1978; 1979; 1981; 1982; Vols. 1−4.

(6) *Energii Razryva Khimicheskikh Svyazei. Potentialy Ionizatzii i Srodsvo k Electronu*; Kondrat'ev, V. N., Ed.; Nauka: Moscow, 1974.

(7) Sauval, A. J.; Tatum, B. J., A Set of Partition Functions and Equilibrium Constants for 300 Diatomic Molecules of Astrophysical Interest. *Astrophys. J. Suppl.* **1984**, *56*, 193−209.

(8) Carlson, C.; Gilkeson, J.; Linderman, K.; LeBlanc, S.; Hefferlin, R.; Davis, W. B. Estimation of Properties of Triatomic Molecules from Tabulated Data Using Least-Squares Fitting. *Croat. Chem. Acta* **1997**, *70*, 479−508.

(9) *Molekulyarnye Postoyanye Neorganucheskikx Soedinenij*; Khimia: Leningrad, Krasnov, K. S., Ed.; 1979; pp 159−263. An English translation of an earlier edition: Schmorak, J. *Handbook of Molecular Constants of Inorganic Compounds*, *Israel Program for Scientific Translations*; Krasnov, K. S., Ed.; Jerusalem, 1970.

(10) Davis, W. B.; Hefferlin, R. Hazards in Mining Molecular Trends with Neural Networks and Tabulated Data. In *Concepts in Chemistry*; Mohanty, R. K., Ed.; in press.

(11) Wohlers, J.; Laing, W. B.; Hefferlin, R.; Davis, W. B. Least-Squares and Neural-Network Forecasting from Critical Data: Diatomic Molecular r_e and Triatomic ΔH_a. In *Advances in Molecular Similarity*; Carbo-Dorca, R., Mezey, P. G., Eds.; JAI Press: 1998; Vol. 2, pp 265−287.

(12) Thomas, M. Senior thesis, Southern Adventist University, April 30, 1999.

CI000070P