

A Semiempirical Formulation for the Study of Molecular Interactions

Serafin Fraga

Laboratory of Atomic and Molecular Structure, Department of Chemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2G2

Received 12 May 1981; accepted 10 November 1981

The coefficients of a $1/R$ expansion (containing long-range electrostatic interaction, medium-range electrostatic-induced dipole attraction, short-range dispersive attraction, and very short overlap interaction terms) have been determined, within a semiempirical approach, from the results of *a priori* calculations. The formulation has been applied to the study of the stable conformations of benzene complexes.

INTRODUCTION

The cost of carrying out large-scale, *a priori* quantum-mechanical calculations for molecular associations is still prohibitive. It seems therefore advisable to find an economical alternative.

The use of the $1/R$ expansion, parametrized on the basis of accurate, *a priori* calculations, may provide such an alternate approach. It is economical, especially when programmed for use in an array processor, and may constitute an efficient tool, when coupled with an appropriate minimum-search technique, for the study of molecular associations.

The only obstacle in such a research program arises in conjunction with the determination of the corresponding expansion coefficients. It is believed that the simple development presented in this work may provide a solution to this problem.

THEORETICAL DESCRIPTION

The interaction between two molecules, A and B, may be approximated by

$$\Delta E = \sum \sum \{ c_{ab}^{(1)}/R_{ab} + c_{ab}^{(4)}/R_{ab}^4 + c_{ab}^{(6)}/R_{ab}^6 + c_{ab}^{(7)}/R_{ab}^7 + c_{ab}^{(8)}/R_{ab}^8 + \dots \}$$

where the summations extend to all the atoms, *a* and *b*, of the two molecules, A and B, respectively, and the expansion coefficients are defined by

$$c_{ab}^{(1)} = Q_a Q_b \quad (1)$$

$$c_{ab}^{(4)} = -(1/2)(\alpha_a Q_b^2 + \alpha_b Q_a^2) \quad (2)$$

$$c_{ab}^{(6)} = -(C_a Q_b^2 + C_b Q_a^2) \quad (3)$$

$$c_{ab}^{(7)} = -(1/2)(B_a Q_b^3 + B_b Q_a^3)$$

$$c_{ab}^{(8)} = -(1/24)(\gamma_a Q_b^4 + \gamma_b Q_a^4)$$

where *Q*, α , γ , *B*, and *C* denote, respectively, the effective electronic charge, dipole polarizability, dipole hyperpolarizability, uniform field quadrupole polarizability, and field gradient quadrupole polarizability of the atom denoted by the corresponding subscript; R_{ab} is the distance between atoms *a* and *b*.

Two difficulties hinder the practical application of this formulation. The values of the charges may be known from independent calculations for the separate molecules and the dipole polarizabilities have been calculated¹ for all the atoms of the Periodic System and their ions, but very limited information is available on the remaining polarizabilities. On the other hand, the above expansion is only valid when the overlap between the density distributions of the two systems is negligible, so that for smaller separations a correcting term must be introduced.

Two different approaches have been proposed and used in practice to circumvent these difficulties. On one hand, semiempirical calculations have been carried out using the expansion

$$\Delta E = \sum \sum \{ c_{ab}^{(1)}/R_{ab} + c_{ab}^{(4)}/R_{ab}^4 + c_{ab}^{(6)}/R_{ab}^6 + c_{ab}^{(12)}/R_{ab}^{12} \} \quad (4)$$

with $c_{ab}^{(1)}$ and $c_{ab}^{(4)}$ defined as given above and $c_{ab}^{(6)}$ approximated (in atomic units) by

Table I. General data.^a

Atom	Class ^b	Q	$\alpha(\text{\AA}^3)$	Group	Position ^c	Remarks
H	23	.060	.42	SH		
	2	.204	.36	CH, CH ₂		
	3	.205	.36	CH ₃		
	16	.253	.34	CH		ring C
	1	.266	.33	NH ₂		
	4	.404	.29	COOH		
C	6	-.608	3.60	CH ₃		
	21	-.503	3.20	CH ₂		α to S
	7	-.383	2.75	CH ₂		
	17	-.218	2.20		r	
	8	-.135	2.00	CH		
	28	-.090	1.90		r	α to CH ₃
	14	-.078	1.90		r	fully substituted or next to N
	19	-.032	1.80		r	junction, geminal to C
	18	.008	1.70	COH	(r)	alcoholic or phenolic
	25	.012	1.70		r	geminal to N
	20	.187	1.45		r	junction, geminal to N
	24	.310	1.30		r	α to NH ₂
	26	.423	1.20	CO	r	
	5	.511	1.10	COOH		
N	13	-.630	2.35	CONH ₂		
	11	-.554	2.10	NH ₂		
	15	-.473	1.90		r	with H
	12	-.317	1.50		r	without H
O	9	-.539	1.45	OH		in COOH
	10	-.409	1.25	CO		in COOH
	27	-.380	1.20	CO		attached to ring
S	22	.123	3.15	SH		

^a The corresponding values (ref. 3) for H and O in water are $Q = 0.311$, $\alpha = 0.32 \text{ \AA}^3$ and $Q = -0.622$, $\alpha = 1.65 \text{ \AA}^3$, respectively.

^b The class indices are those of Clementi and co-workers (ref. 3).

^c The "r" in this column indicates that the atoms belong to a ring.

$$c_{ab}^{(6)} = -(3/2)\alpha_a\alpha_b[(\alpha_a/n_a)^{1/2} + (\alpha_b/n_b)^{1/2}] \quad (5)$$

where n_a , n_b denote the corresponding effective numbers of valence electrons. The last term in the above expansion represents a correction included in order to account for the overlap repulsion and its coefficient, $c_{ab}^{(12)}$, is usually determined in a semiempirical way.²

On the other hand, Clementi and co-workers³ have used the expansion

$$\Delta E = \sum \sum \{A_{ab}/R_{ab} + B_{ab}/R_{ab}^6 + C_{ab}/R_{ab}^{12}\} \quad (6)$$

for the study of the interaction of water with a number of biomolecules, with coefficients determined by the fitting of accurate results from a

priori calculations. This approach offers the advantage of taking into account the environment of the atoms, a fact which is reflected in the definition of classes of atoms, as listed in Table I.

The work of Clementi and co-workers³ should now be extended in order to obtain the parameters for the interaction between atoms of any class, but the cost could be prohibitive. A simple way out of this dilemma may be provided by the procedure described below.

METHOD OF CALCULATION

The objective is to determine the coefficients of eq. (4) for the interaction of any two classes of

atoms. Preference is given to this equation because of its completeness: it contains the long-range electrostatic interaction (R^{-1}), the medium-range electrostatic-induced dipole attraction (R^{-4}), the short-range dispersive attraction (R^{-6}), and the very short overlap interaction (R^{-12}).

The method consists of the following steps:

(a) *Adoption of the values for Q , α , and n .* Average values of Q have been determined from the values found by Clementi and co-workers³ in their extensive calculations for biomolecules.

The polarizabilities have been obtained by interpolation, as corresponding to the above charges, from the values available for the corresponding atoms and their ions.

The effective numbers of electrons have been defined as $n = Z - Q$, where Z denotes the nuclear charge of the atom under consideration.

The corresponding values of Q and α are presented in Table I.

(b) *Fitting of the interaction with the atoms of water.* The interaction of an atom of a given class with an atom (hydrogen or oxygen) of water is calculated separately for thirty different separations (starting at 2 Å, with increments of 0.25 Å), using the Clementi coefficients in eq. (6).

These values are then fitted by eq. (4), with coefficients $c_{ab}^{(1)}$, $c_{ab}^{(4)}$, and $c_{ab}^{(6)}$ evaluated according to eqs. (1), (2), and (5) with the values of Q , α , and n discussed in (a), so that the optimum value of $c_{ab}^{(12)}$ may be determined by a least-squares procedure.

The calculations are then repeated affecting the polarizability values by independent factors, until the best fitting (on the basis of the absolute value of the maximum error) is obtained. This procedure means that the coefficients $c_{ab}^{(4)}$ and $c_{ab}^{(6)}$ are evaluated as

$$c_{ab}^{(4)} = -(1/2)(f_a \alpha_a Q_b^2 + f_b \alpha_b Q_a^2)$$

$$c_{ab}^{(6)} = -(3/2)f_a \alpha_a f_b \alpha_b / [(f_a \alpha_a / n_a)^{1/2} + (f_b \alpha_b / n_b)^{1/2}]$$

where f_a , f_b are optimization factors to be determined.

(c) *Semiempirical interpretation of the coefficients $c_{ab}^{(12)}$.* The next step consists of trying to obtain, from these results, values that might be ascribed to the individual atoms. An approximation consists of assuming that the coefficient $c_{ab}^{(12)}$ may be expressed

$$c_{ab}^{(12)} = c_a^{(12)} c_b^{(12)}$$

in terms of two pseudocoeficients, $c_a^{(12)}$ and $c_b^{(12)}$,

ascribed to the individual atoms.² Within the context of this work, use of this assumption implies that the parameters $c_{X(c)}^{(12)}$ (i.e., for class c of atom X) may be evaluated from the coefficients $c_{X(c)H(w)}^{(12)}$ and $c_{X(c)O(w)}^{(12)}$ if the respective values of $c_{H(w)}^{(12)}$ and $c_{O(w)}^{(12)}$ are known (i.e., if they may be estimated).

The determination of $c_{H(w)}^{(12)}$ and $c_{O(w)}^{(12)}$ has been carried in the following way, which simultaneously provides a test for the accuracy of the fit. The interaction of O(w) with O(9), which is the hydroxylic oxygen in the carboxyl group, represents a case where the two interacting atoms are very similar: i.e., it may be assumed that $c_{O(w)}^{(12)} \approx c_{O(9)}^{(12)}$. The value, $c_{O(9)}^{(12)} = 552 \text{ kcal}^{1/2} \text{ Å}^6 \text{ mol}^{-1/2}$, obtained under this assumption, compares very well with the value of $534 \text{ kcal}^{1/2} \text{ Å}^6 \text{ mol}^{-1/2}$, obtained by Minicozzi and Bradley² in optimization calculations, which reproduce the experimental results for the dimer of formic acid. In addition, from the value for $c_{O(10)O(w)}^{(12)}$, and using $c_{O(w)}^{(12)} = 552 \text{ kcal}^{1/2} \text{ Å}^6 \text{ mol}^{-1/2}$, one obtains $c_{O(10)}^{(12)} = 473 \text{ kcal}^{1/2} \text{ Å}^6 \text{ mol}^{-1/2}$, which again compares very well with the value $496 \text{ kcal}^{1/2} \text{ Å}^6 \text{ mol}^{-1/2}$, obtained by Minicozzi and Bradley.²

The above results seem to indicate that the results of Minicozzi and Bradley are consistent with those of Clementi,³ to a degree which can be assessed in the following way. From the value for $c_{O(9)O(w)}^{(12)}$, and approximating $c_{O(9)}^{(12)}$ by the value $534 \text{ kcal}^{1/2} \text{ Å}^6 \text{ mol}^{-1/2}$ of Minicozzi and Bradley, one obtains $c_{O(w)}^{(12)} = 570 \text{ kcal}^{1/2} \text{ Å}^6 \text{ mol}^{-1/2}$; similarly, from the value for $c_{O(10)O(w)}^{(12)}$, and approximating $c_{O(10)}^{(12)}$ by the value $496 \text{ kcal}^{1/2} \text{ Å}^6 \text{ mol}^{-1/2}$ of Minicozzi and Bradley,² one obtains $c_{O(w)}^{(12)} = 526 \text{ kcal}^{1/2} \text{ Å}^6 \text{ mol}^{-1/2}$.

Using an average value, $c_{O(w)}^{(12)} = 548 \text{ kcal}^{1/2} \text{ Å}^6 \text{ mol}^{-1/2}$, leads back to values $c_{O(9)}^{(12)} = 555 \text{ kcal}^{1/2} \text{ Å}^6 \text{ mol}^{-1/2}$ and $c_{O(10)}^{(12)} = 476 \text{ kcal}^{1/2} \text{ Å}^6 \text{ mol}^{-1/2}$, in better agreement with the results of Minicozzi and Bradley than the tentative values obtained at the beginning of this discussion.

Then, with the value $c_{O(9)}^{(12)} = 555 \text{ kcal}^{1/2} \text{ Å}^6 \text{ mol}^{-1/2}$, one obtains $c_{H(w)}^{(12)} = 5.47 \text{ kcal}^{1/2} \text{ Å}^6 \text{ mol}^{-1/2}$ from the value of $c_{O(9)H(w)}^{(12)}$, and therefore one can now proceed to the evaluation of all the remaining $c_{X(c)}^{(12)}$. The values, presented in Table II, are those that give a minimum repulsion contribution at very short distances.⁴

PRACTICAL APPLICATIONS

The interaction energy between two molecules, A and B , may then be evaluated as

Table II. Final values^{a,b} of the parameters $f_{X(c)}$ and $c_{X(c)}^{(12)}$.

Atom	Class	$f_{X(c)}$	$c_{X(c)}^{(12)}$	Atom	Class	$f_{X(c)}$	$c_{X(c)}^{(12)}$
H	23	.85	69.3	C	18		186
	2		162		25		439
	3		22.3		20	1.00	1646
	16		131		24	2.60	982
	1	.35	1.28	N	26		385
	4	.95	5.53		5		667
C	6		536		13		1111
	21	.10	21.2		11	.15	245
	7	.10	508		15	.95	998
	17		948		12		362
	8		362	O	9	.21	555
	28		72		10		476
	14	.10	234		27	.80	268
	19	.10	13.4	S	22	.35	1035

^a The values of $c_{X(c)}^{(12)}$ are given in $\text{kcal}^{1/2} \text{\AA}^6 \text{mol}^{-1/2}$.

^b When not given explicitly, the optimum value of $f_{X(c)}$ lies between 0.0 and 0.1. The latter value has been adopted in actual calculations.

$$\begin{aligned} \Delta E = & 0.13894168 \times 10^4 Q_a Q_b / R_{ab} \\ & - 0.69470838 \times 10^3 (f_a \alpha_a Q_b^2 + f_b \alpha_b Q_a^2) / R_{ab}^4 \\ & - 0.15160732 \times 10^4 f_a f_b \alpha_a \alpha_b / \\ & \times [(f_a \alpha_a / n_a)^{1/2} + (f_b \alpha_b / n_b)^{1/2}] R_{ab}^6 \\ & + 4.184 c_a^{(12)} c_b^{(12)} / R_{ab}^{12} \end{aligned}$$

where a and b now include the class designations of the interacting atoms and f denotes the corresponding factor. The interaction energy will be given in kJ (abs.) when α is given in \AA^3 , R_{ab} in \AA , and $c^{(12)}$ in $\text{kcal}^{1/2} \text{\AA}^6 \text{mol}^{-1/2}$.

This expression has an average character, due to the fact that the Clementi coefficients represent averages obtained from the fitting of interactions in many different molecules.

The electronic charges to be used are those given in Table I, corrected so that the corresponding total charges of the system considered are what they should be. This correction may be introduced in the following way: Denoting by Q_e , Q_p , and Q_n the values predicted using the effective charges from Table I for the total charge and for the contributions of the positive and negative charges, respectively, and by Q the true total charge, the charge Q_i of the i th atom may be transformed according to

$$Q'_i = Q_i \{1 + c Q_i / |Q_i|\}$$

where the correction factor c is given by

$$c = (Q - Q_e) / (Q_p - Q_n)$$

The above expression for ΔE has been programmed, coupled with a steepest-descent procedure for the search of the minimum energy, for the study of the stable conformations of molecular associations. This program,⁵ which only requires as input the geometries of the interacting molecules, runs on a Floating Point Systems array processor AP-190L, hosted by the Amdahl 470V/7 computer of the University of Alberta Computing Centre. An automatic search of the configuration space may be first carried out, with a posterior refinement of the results for the conformations of interest. The graphical representations of the complex are obtained directly in a Tektronix T-4015-1 terminal, adapted with a hard copier.

As the parametrization of the $1/R$ expansion has been performed on the basis of the results of Clementi for the interaction of biomolecules (especially amino acids) with water, it is expected that the present formulation should be appropriate for the evaluation of the interactions between such systems. In any case, a number of trial calculations have been performed for a variety of simple systems, with satisfactory results,⁶ but for

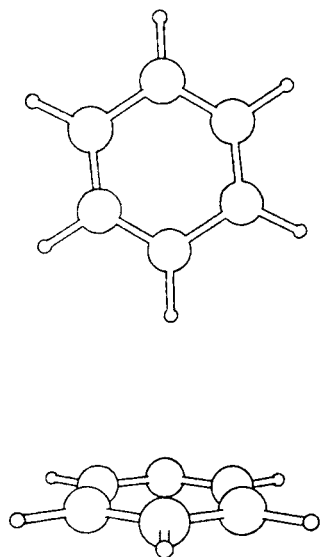


Figure 1. Most stable conformation of the benzene dimer ($\Delta E = -6.3$ kJ).

simplicity only the results for benzene are reported here as an illustrative example. (Benzene has been considered a good test system, not being similar or related to the biomolecules studied by Clementi.³)

The search through the configuration space has yielded only two stable conformations, represented in Figures 1 and 2, with stabilization energies of -6.3 and -4.6 kJ, respectively. It is interesting to note that the most stable conformation is related to that observed⁷ in crystalline benzene, where pairs of molecules fit together like six-toothed bevel gear wheels, with axes at almost right angles. Such a structure will not be stable in gas phase, because of the symmetry of the interaction, but one can easily visualize how the crystalline conformation will result (from the first conformation mentioned above) by rotation of one of the wheels under the influence of the other molecules.

CONCLUDING REMARKS

Calculations of molecular interactions, based on the use of $1/R$ expansions, are simple and inexpensive and will yield satisfactory results if the expansion coefficients have been determined on the basis of results from accurate calculations, as in the work of Clementi.³

Within that approach, the determination of the coefficients for the evaluation of the interaction between any two classes of atoms will require, however, a considerable amount of computer time and it may be that they will not be available for a long time, if ever.

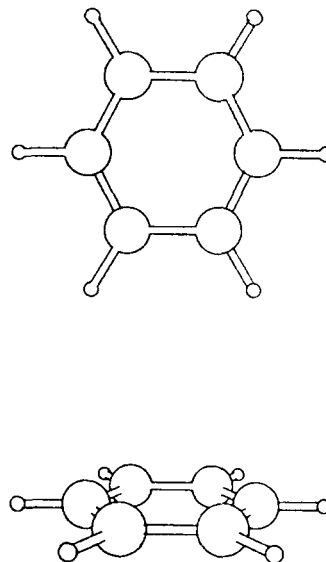


Figure 2. Another stable conformation of the benzene dimer ($\Delta E = -4.6$ kJ).

The alternative is a semiempirical approach, such as the one discussed here, whose goodness should be judged on the basis of numerical results.

This work represents a first attempt in this direction and therefore it may be of interest to highlight its deficiencies and the possible improvements that might be considered in future developments.

(1) The expansion adopted for this work was chosen because it may offer the possibility of extrapolation to new classes of atoms and provides the basis for further improvements (such as the semiempirical inclusion of the dispersion contribution).

At this moment, as the determination of the expansion coefficients has been based on the results of SCF calculations, the expansions presented in this work are most appropriate for those cases in which the contribution of the dispersion energy is not appreciable.

(2) The main deficiency in the formulation, both at the development and the application levels, arises from the values used for the effective charges.

The dependence⁸ of the values of the atomic charges on the population analysis adopted could be removed through a semiempirical scaling,* as carried out for the polarizabilities. This is a possibility to be considered in future work, especially when data (such as those of Clementi³) will be

* The author would like to thank one of the referees for this suggestion.

available for the complete series of atomic classes. At this moment, however, such an approach would hinder the possibility of extrapolation of the present results to other classes of atoms. It is for this reason that values related to those obtained by Clementi by the well established Mulliken analysis have been used in this work.

The selection of the values used in the development was conditioned by the average character of the Clementi expansion, insofar as they were obtained from the results for a number of molecules (in spite of the concept of atom class, as charges for a given class vary, even if slightly from molecule to molecule). Consequently, average values were used in the development and have been proposed for the actual calculations.

There is, of course, another possibility. It is implicit in the work of Clementi that the actual values of the charges, as calculated independently for each molecule, may be used in his expansions. The same approach may be adopted in this case, whenever such values are available, but the use of average charges is proposed here in order to avoid the dependence on the availability of any data other than the molecular geometries of the interacting molecules.

(3) The correct values of the scaling factors for the polarizabilities, as given in Table II, may be used in actual calculations, although it is proposed that the value 0.1 be used instead of zero, as argued below. Those correction factors have been obtained using the average values of the effective charges, but in actual calculations the latter are modified in order to reproduce the correct total charge of the system. The scaling factors should be modified appropriately, but that would imply a reevaluation of the interaction parameters for each

pair of molecules to be studied. The alternative is to leave them unchanged, except when they are zero, for in such a case the change in the charges will affect only the electrostatic interaction. The value 0.1, chosen arbitrarily, has led to satisfactory numerical results. In this connection it is worthwhile to mention that a semiempirical modification of the present formulation with view to account for the electrostatic-induced dipole and the dispersion contributions may be achieved through an appropriate increase of the values of the scaling factors.

References

1. S. Fraga, J. Karwowski, and K. M. S. Saxena, *Handbook of Atomic Data*, Elsevier, Amsterdam, 1980; S. Fraga and J. Muszynska, *Atoms in External Fields*, Elsevier, Amsterdam, 1981.
2. See, e.g., the work of W. P. Minicozzi and D. F. Bradley, *J. Comput. Phys.*, **4**, 118 (1969), as a representative example of the use of this type of expansion.
3. See, e.g., E. Clementi, *Computational Aspects for Large Chemical Systems*, in *Lectures Notes in Chemistry*, Springer-Verlag, Berlin, 1980, and references therein.
4. Except for classes 2, 3, and 16, for which the values listed are those obtained from the results for $c_{X(c)H(w)}^{(12)}$. The value adopted for class 5 represents an average of the values for classes 17 and 26.
5. This program, which is available on request, may be easily adapted to be run in any other computer.
6. Hydrocarbons, alcohols and phenol, carboxylic acids, amines, and amides were among the systems tested. In addition, the method has been applied to the study of the conformations of the complexes of cyclobutadiene and formaldehyde with various systems in inert matrices.
7. E. G. Cox, D. W. J. Cruickshank, and J. A. S. Smith, *Proc. R. Soc. Ser. A.*, **247**, 1 (1958).
8. See, e.g., V. H. Smith, *Phys. Scr.*, **15**, 147 (1977); and K. B. Wiberg, *J. Comput. Chem.*, **2**, 53 (1981).