Theoret. chim. Acta (Berl.) 21, 9—16 (1971)

Small Gaussian Basis Sets in *Ab Initio* Calculations

HELGE JOHANSEN

Department of Physical Chemistry, The University of Copenhagen,
Universitetsparken 5, DK-100 Copenhagen, Denmark

Received June 30, 1970

Ab initio calculations have been carried out on the water molecule, using the IBMOL program, in order to investigate some possible uses and applications of small Gaussian basis sets. Firstly, calculations have been carried out in which emphasis was placed upon an improved description of the valence orbitals and, secondly, a study has been made of the feasibility of performing calculations in which a large basis set is used in the evaluation of the more important integrals and a small basis set for the remaining integrals. The two investigations indicate ways in which the calculated values of certain molecular properties may be improved upon without extending the computation time significantly.

Ab initio-Rechnungen für das H_2O-Molekül mittels des IBMOL-Programms wurden zur Untersuchung kleiner Basissätze ausgeführt. Einerseits wurden Rechnungen mit dem Ziel, eine bessere Beschreibung von Valenzorbitalen zu erreichen, unternommen, und andererseits wurde versucht, durch Verwendung eines großen Basissatzes für wichtige Integrale und eines kleinen für die weniger wichtigen zu einer Ersparnis an Rechenzeit zu kommen. Zwei Vorschläge, molekulare Eigenschaften genauer, aber ohne größeren Aufwand an Rechenzeit, zu bestimmen, wurden gemacht.

Calculs *ab initio* sur la molécule d'eau à l'aide du programme IBMOL afin d'étudier certaines utilisations et applications possibles de petites bases gaussiennes. Tout d'abord des calculs ont été effectués en insistant sur l'amélioration de la description des orbitales de valence, puis une étude a été menée sur la possibilité d'effectuer les calculs avec une grande base uniquement pour les intégrales les plus importantes. Ces deux études montrent comment les valeurs calculées de certaines propriétés moléculaires peuvent être améliorées sans trop augmenter le temps de calcul.

1. Introduction

Large computers and sophisticated programs have now made *ab initio* calculations tractable as tools for investigations on smaller molecules. The fourth power dependence of the number of basis functions in the problem puts, however, natural limits on the size of molecule that may be treated. If, as an example, the molecular size is increased by a factor of three, then the computing time required is about one hundred times larger.

There appear to be four ways of handling the problems concerned with the large molecules:

1) Increase of computer speed and storage capacity.
2) Further improvement of formulae and programs.
3) Optimization of basis sets.
4) New methods altogether.

This investigation has its emphasis on point number 3, and the purpose will be to discuss the use of small basis sets in calculations on small molecules, in order to see if it is at all possible to extract meaningful information from these basis sets, thus making them suitable for calculations on larger molecules.

Table 1. *Partition of s-type basis functions in first-row atoms*

Total number of functions	4	7	11
1*s*-lobe	3	5	7
2*s*-lobe	1	2	4

Two ways will be followed here:

1) Reduction of the number of basis functions for the inner shells.

2) Different sized basis sets in the same calculation.

In addition to these, some contraction effects will be discussed.

When the usual method for finding basis sets is applied, that is, a self-consistent-field Roothaan calculation [13, 14], then the value of most concern is the total energy, and the variation method will minimize this property with respect to the available parameters. But much effort is then spent on the inner electrons, where there is relatively most energy to save, whereas the chemically more important electrons in the valence shell are less accurately described. Table 1 shows how Gaussian basis functions for *s*-orbitals may be shared between the 1*s*- and 2*s*-lobes; from this it is seen that seven basis functions for the 1*s*-lobe have to be carried through all the integrals in order to represent the 2*s*-lobe with only four functions. Therefore, the first part of this investigation is concerned with assigning more functions to the valence orbitals and fewer to the core.

A second reason why large basis sets are required in molecular calculations is that some of the one- and two-centre integrals have to be calculated very accurately. But, as also pointed out by McWeeny [10], it should not be necessary to calculate the small and relatively insensitive three- and four-centre integrals with a large basis set. Thus, using different basis sets for the different integrals, the problem can be reduced to one of only second degree dependence on the size of the larger basis set. The second part of this investigation will, therefore, discuss the use of two basis sets in integral evaluations.

For all calculations, the test molecule employed was H_2O, as this molecule is so well know from other calculations. It is, however, a little unfortunate that it is only a three-centre problem. The two ideas presented here have, of course, been used many times in connection with semi-empirical calculations, that is, in core approximations and in zero-differential-overlap approximations, and they may be thought of as applications of experiences in these fields; whereas most other investigations have used accurate calculations to justify empirical methods. It should be noted that some *ab initio* calculations [1, 5, 6, 8] of a similar nature have been reported during the course of this work.

2. Method

The program written by Roos, Salez, Veillard and Clementi [12] has been used for the construction of small contracted basis sets, and the version 2 of IBMOL [3] (QCPE 92) in the molecular calculations. A small modification of the latter program was, however, necessary for the calculations having two basis sets.

There are, by now, several different nomenclatures for identifying basis sets. In this paper, round brackets, (), will be used for uncontracted basis sets, and angular brackets, $\langle\,\rangle$, for contracted sets. Further, the usual notation of this kind is extended in order to describe more accurately how the contractions have been carried out. Thus $(\mathrm{X}/a, b, \ldots)$ describes an uncontracted basis set for atom X with a s-type functions, b p-type function etc., and $\langle \mathrm{X}/a_1, a_2, \ldots; b_1, b_2, \ldots; \ldots\rangle$ describes a contracted set for atom X with a_1 functions in the first s-contraction, a_2 in the second etc., b_1 functions in the first p-contraction, b_2 in the second, and so on. The usual notation for contracted functions in this nomenclature is $\langle \mathrm{X}/a', b', \ldots\rangle$, with a' as the number of s-contractions, b' the number of p-contractions etc. It should be noted that the nomenclature introduced here is an unnecessary complication for most other applications.

For each case in this investigation a potential curve is found through the calculation of five points. A fourth degree polynomial is used to find the extremum value, and the orbital energies are calculated by interpolation to this point. The force constant is found from the coefficient of the second order term in a fourth degree power series expansion around the minimum.

3. Results and Discussion

3.1 Basis Sets

Two kinds of basis set have been constructed for the first row atoms C, N, O and F. The first set has the form $\langle \mathrm{X}/3,1; 2\rangle$, and is a contraction of the set (X/4,2) which, in turn, is found by normal self-consistent-field calculation [12]; the contraction coefficients have been optimized in the contracted set. The set (X/4,2) has recently also been reported elsewhere [18]. It is probably reasonable to say that $\langle \mathrm{X}/3,1; 2\rangle$ is the simplest basis set which might be expected to yield worthwhile results. The second set has been constructed by a least squares technique, where, for the 1s- and 2s-lobes, two Gaussian orbitals have been fitted to single term Slater orbitals [4] and, in the case of the 2p-orbitals a linear combination of four Slater orbitals [2] has been fitted by two Gaussians. The resulting functions were then contracted and the contraction coefficients optimized [12] to produce a basis set of the form $\langle \mathrm{X}/2, 2; 2\rangle$. The only difference between the two final basis sets is, therefore, the different distributions of s-functions among core and valence electrons.

The results, in terms of total energy and orbital energies, are shown in Table 2, where accurate calculations performed by Clementi [2] are also reported for comparison. The second set poorly describes the 1s-core and, therefore, as might be expected, the 1s orbital energy and the total energy are very high. However, the energies for the valence electrons are significantly better with $\langle \mathrm{X}/2,2; 2\rangle$ than with $\langle \mathrm{X}/3,1; 2\rangle$, and $\langle \mathrm{X}/2,2; 2\rangle$ shows even better results than the uncontracted basis set (X/4,2), where e.g. (C/4,2) gives the figures $E = -37.3556$, $\varepsilon_{1s} = -11.244$, $\varepsilon_{2s} = -0.676$ and $\varepsilon_{2p} = -0.374$. The primary purpose, namely to transfer accuracy from the description of the core to the description of the valence electrons, is, therefore, possible as far as atomic orbital energies are concerned.

Table 2. *Total energy and orbital energies for calculations on* C, N, O *and* F. *All energies are in atomic units*

Basis set		E	ε_{1s}	ε_{2s}	ε_{2p}
C(^3P)	⟨C/3,1; 2⟩	−37.2918	−11.186	−0.660	−0.372
	⟨C/2,2; 2⟩	−36.2070	−10.929	−0.709	−0.405
	Ref. a	−37.6886	−11.326	−0.706	−0.433
N(^4S)	⟨N/3,1; 2⟩	−53.7772	−15.453	−0.873	−0.475
	⟨N/2,2; 2⟩	−52.2705	−15.132	−0.960	−0.536
	Ref. a	−54.4009	−15.629	−0.945	−0.568
O(^3P)	⟨O/3,1; 2⟩	−73.8607	−20.475	−1.145	−0.497
	⟨O/2,2; 2⟩	−71.8412	−20.078	−1.280	−0.597
	Ref. a	−74.8094	−20.669	−1.244	−0.632
F(^2P)	⟨F/3,1; 2⟩	−98.0362	−26.128	−1.425	−0.532
	⟨F/2,2; 2⟩	−95.4028	−25.688	−1.633	−0.694
	Ref. a	−99.4093	−26.383	−1.573	−0.730

Ref. a: E. Clementi [2], where an STO basis set with 6 *s*-type and 4 *p*-type functions was used.

3.2. *Calculations on* H_2O *with Small Contracted Basis Sets*

The two basis sets ⟨O/3,1; 2⟩ and ⟨O/2,2; 2⟩ have been used in test calculations on H_2O, and the results are shown, together with the accurate calculation by Moskowitz and Harrison [11], in Table 3. The function for hydrogen, ⟨H/2⟩, is the one presented by Stewart [17]. As may be seen, the total energy and the orbital energies show the same trends as in the atomic calculations. The total energy for the ⟨O/2,2; 2⟩ calculation is worse, by the same two atomic units, as in the calculation on oxygen, and the four outer orbitals, the valence orbitals, again show significant improvement. In the case of molecules it is, therefore, also possible to improve on the valence orbital energies at the expense of total energy and the 1*s*-core.

In Fig. 1, the potential curve for bond angle variation in calculations with the two basis sets are compared with the (O/9,5) (H/3,1) calculations performed by Moskowitz and Harrison [11]. From Fig. 1 and Table 3 it is seen that both ⟨O/3,1; 2⟩ and ⟨O/2,2; 2⟩ give equilibrium bond angles and force constants in

Table 3. *Total energy, equilibrium distance, equilibrium angle, force constant and orbital energies for different calculations on* H_2O. *All numbers except the angles are in atomic units*

Basis set	E	OH$_{eq}$[a]	HOH$_{eq}$[b]	k[c]	$1a_1$	$2a_1$	$1b_1$	$3a_1$	$1b_2$
⟨O/3, 1; 2⟩ ⟨H/2⟩	−75.0197	1.94	107°.3	1.07	−20.44	−1.27	−0.61	−0.47	−0.42
⟨O/2, 2; 2⟩ ⟨H/2⟩	−72.9671	2.06	107°.9	0.92	−19.94	−1.33	−0.63	−0.53	−0.48
(O/9, 5, 2) (H/3, 2)[d]	−76.0421	1.81[e]	108°.4[e]	1.24[e]	−20.5516	−1.3478	−0.7144	−0.5779	−0.5015

[a] The experimental value is 1.81 [16].

[b] The experimental value is 104°.5 [16].

[c] The experimental value is 1.07 a. u. = 16.7 m dyn/Å [16].

[d] Ref. [11].

[e] These results were found with (O/9,5) (H/3,1).

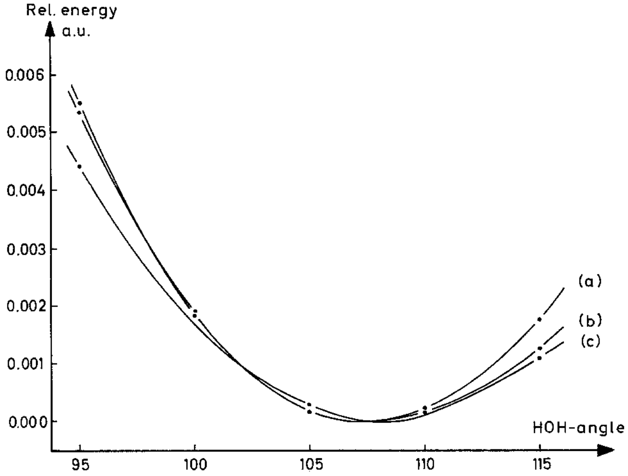

Fig. 1. *a*–*c* Potential curves for angular variations. a ⟨O/3, 1; 2⟩, b ⟨O/2, 2; 2⟩ and c (O/9,5) (H/3,1) from Ref. [11]. The curves have been translated in energy to give the same minimum value

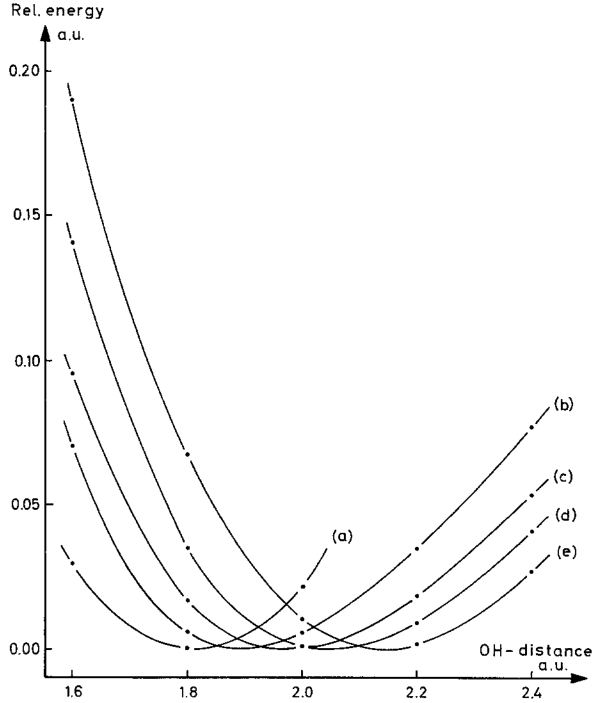

Fig. 2. *a*–*e* Potential curves for distance variation. a (O/9,5) (H/3,1) from Ref. [11], b (O/4,2) (H/2); c ⟨O/7,4; 7⟩ ⟨H/6⟩ (A in Table 4); d ⟨O/2,2; 2⟩ ⟨H/2⟩ (*D* in Table 4); e calculation with all one-electron integrals and some two-electron integrals from *A* and the other two-electron integrals from *D* (*B* in Table 4). The curves have been translated in energy to give the same minimum value

close agreement with the accurate calculation; the $\langle O/2{,}2;2\rangle$ set is, however, slightly superior.

From Fig. 2 and in Table 3 it is seen, on the other hand, that neither one of the two sets is very reasonable to use for the determination of equilibrium bond lengths, whereas the force constants are quite well reproduced. It is a general trend that small and very contracted basis sets give poor and consistently too large bond lengths. A calculation on CO with the basis $\langle X/2{,}2;2\rangle$ gave, as an example, an equilibrium bond length which was about 0.4 atomic units longer than the one found in the accurate calculation by McLean and Yoshimine [9]; the shapes of the potential curves were, however, very similar.

The error in the OH bond length is about 0.25 a. u. with $\langle O/2{,}2;2\rangle$ when the contracted hydrogen function $\langle H/2\rangle$ is used; but if the uncontracted (H/2) is applied with $\langle O/2{,}2;2\rangle$, the error is only 0.15 a. u., which demonstrates the well known difference between hydrogen functions in atoms and molecules. If the contractions are taken away altogether, the basis set of the form (O/4,2) (H/2) improves the equilibrium distance to 1.90 a. u. This indicates that basis functions should not be too strongly contracted if accurate bond lengths are sought.

3.3 Use of Different Basis Sets for Different Integrals in Calculations on H_2O

A second way to reduce the number of integrals is to use different sized basis sets in integral calculations. The (O/11,7) Huzinaga, Coiro, Rusconi and Clementi basis set (presented by Salez and Veillard [15]) was contracted to $\langle O/7{,}4;7\rangle$ (A), and a further basis set of the form $\langle O/2{,}2;2\rangle$ (D) was constructed by a least squares fit for each of the three contractions. The set found this way is thus very similar to the one reported in Section 3.1. The basis set for hydrogen was, in calculation A, $\langle H/6\rangle$ from Huzinaga [7] and, in calculation D, the previously mentioned $\langle H/2\rangle$. Two mixtures of the sets A and D were tried:

B) All one-electron integrals and the two-electron integrals of the form $\int\int \chi_r(1)\,\chi_r(1)\,\frac{1}{r_{12}}\,\chi_s(2)\,\chi_s(2)\,d\tau_1 d\tau_2$ calculated with A and the other two-electron integrals with D (where $\chi_r(i)$ is a contracted Gaussian type orbital on centre r for the electron i).

C) All one-electron integrals calculated with A and all two-electron integrals with D.

The results are reported in Table 4 and in Fig. 2. If we look at the orbital energies, it is seen that going from D, to C, to B changes the valence orbitals so that they are almost identical to the ones found in A. The total energy for calculation B is quite good, but for the case C this quantity is even below the Hartree-Fock limit. This is, of course, not reasonable, and the calamity arises because the forces in the molecule are out of balance, that is, the kinetic energies are calculated with a basis set which is different from the one used for the repulsion energies; with other words the virial theorem is strongly violated. The equilibrium bond distances are again found with rather large values, and it is worth noticing that the distance from calculation A, $\langle O/7{,}4;7\rangle$ $\langle H/6\rangle$, is worse than the distance found from (O/4,2) (H/2); this again stresses the problems, associated with overcontraction, encountered in bond length calculations. The equilibrium distances

Table 4. *Total energy, equilibrium distance, force constant and orbital energies for different calculations on* H_2O. *The basis set in D is a least squares fit to A; calculation C has all two-electron integrals from D and all one-electron integrals from A; calculation B is as C, but some two-electron integrals are included from A (see text). All numbers are in atomic units*

Basis set	E	OH_{eq}	k	$1a_1$	$2a_1$	$1b_1$	$3a_1$	$1b_2$
A. ⟨O/7, 4; 7⟩ ⟨H/6⟩	−75.8083	1.97	0.90	−20.42	−1.28	−0.66	−0.55	−0.52
B.	−75.6625	2.14	0.95	−20.47	−1.17	−0.66	−0.55	−0.52
C.	−76.1833	2.09	0.97	−20.54	−1.17	−0.69	−0.58	−0.56
D.⟨O/2, 2; 2⟩ ⟨H/2⟩	−72.9622	2.05	0.91	−19.66	−1.26	−0.60	−0.49	−0.45

in calculations *B* and *C* both fall outside the range set by *A* and *D* (see Table 4 and Fig. 2), and a mixed basis set method of this kind should, therefore, not be used to calculate this property. On the other hand, it turns out that the force constants are calculated with reasonable accuracy, and they seem to be relatively insensitive to the method chosen.

4. Conclusion

The idea of simplifying the core in order to extend the "valence" basis set gives good results in terms of orbital energies. The total energy is, of course, increased but the exact value of this property is not often needed and, indeed, is usually not found by approximate methods. However, accurate orbital energies may be calculated with a mixed basis, as exemplified by *B* in Sect. 3.3. The force constants are reasonably well reproduced in all the calculations, but there seems to be very little likelihood of extracting good bond lengths from small and strongly contracted basis sets. The final conclusion is, therefore, that basis sets should be chosen according to the properties required, and that the small basis sets may be used for certain properties.

Acknowledgements. The author wishes to acknowledge the use of both the program for atomic calculations and the IBMOL program, and is grateful to the Northern Europe University Computing Center at The Technical University of Denmark for the use of the IBM 7094 computer. Further, it is a pleasure to thank Dr. J. P. Dahl for his continued interest in this work.

References

1. Bendazzoli, G. L., Bernardi, F., Palmieri, P.: To be published.
2. Clementi, E.: IBM Res. Develop. **9**, 2 (1965).
3. — Davis, D. R.: J. comput. Physics **2**, 223 (1967).
4. — Raimondi, D. L.: J. chem. Physics **38**, 2686 (1963).
5. Cook, D. B., Palmieri, P.: Molecular Physics **17**, 271 (1969).
6. — — Chem. Physics Letters **3**, 219 (1969).
7. Huzinaga, S.: J. chem. Physics **42**, 1293 (1965).
8. Klessinger, M.: Theoret. chim. Acta (Berl.) **15**, 353 (1969).
9. McLean, A. D., Yoshimine, M.: IBM Res. Develop. **12**, 206 (1968).
10. McWeeny, R.: Symposia of the Faraday Society **2**, 55 (1968).
11. Moskowitz, J. W., Harrison, M. C.: J. chem. Physics **43**, 3550 (1965).

12. Roos, B., Salez, C., Veillard, A., Clementi, E.: IBM Research Report, RJ 518 (#10901) August 12, 1968.
13. Roothaan, C. C. J.: Rev. mod. Physics **23**, 69 (1951).
14. — Rev. mod. Physics **32**, 179 (1960).
15. Salez, C., Veillard, A.: Theoret. chim. Acta (Berl.) **11**, 441 (1968).
16. Smith, S., Linnett, J. W.: Trans. Faraday Soc. **52**, 891 (1956).
17. Stewart, R. F.: J. chem. Physics **50**, 2485 (1969).
18. Whitman, D. R., Hornback, C. J.: J. chem. Physics **51**, 398 (1969).

Dr. Helge Johansen
Department of Physical Chemistry
The University of Copenhagen
Universitetsparken 5
2100 Copenhagen, Denmark