Holographic QSAR Models for Estimating Densities of Energetic Materials†

Soo Gyeong Cho,* Eun Mee Goh, and Jeong Kook Kim

Agency for Defense Development, P.O. Box 35-5, Yuseong, Taejon 305-600, Korea
Received January 31, 2001

Keywords: Density, Energetic material, Explosives, HQSAR.

An accurate prediction of explosive performance is of significant importance in finding promising candidates for novel energetic materials, since synthesizing new compounds usually requires a great deal of effort. In the society of explosives researches, it is now generally accepted that the performance of explosives is predicted with a reasonable accuracy, if the heat of formation and density are provided accurately.[1-3] Our preliminary tests with the Cheetah program[2] show that the explosive performance is greatly sensitive to the density values, but somewhat less sensitive to the heat of formation.[3] Based on our previous experiences, the density should be predicted within ±0.03 g/cc and the heat of formation should be estimated within ±5 kcal/mol, if one wants to have a fair accuracy in explosive performance.[3] If one predicts the density value with an error of ±0.05 g/cc and the heat of formation of with an error of ±10 kcal/mol, the predicted performance may also be acceptable in guiding whether new explosive molecules can be worthwhile to pursue synthetically, but in some cases the predicted results have a possibility of being inaccurate. If one of two inputs has a larger error than the criteria mentioned above, the predicted explosive performance may be erroneous, unless two conflicting errors are cancelled by chance.[3]

Predicting the crystal density accurately is one of the most difficult challenges in computational chemistry.[4] Many scientists in various research areas have been attempting to predict the crystal packing patterns as well as crystal densities based on the arrays of 3-D molecular structure,[5-11] but this prediction is still a formidable task and is known to have several huge hurdles in getting the job done in a right fashion. Although some applications have been found in the area of energetic materials,[7,12] this approach may not be performed routinely in explosives modeling at this moment. We also believe that this approach will not be a practical solution since it requires extensive computational works and takes relatively long time. Thus, up till now, many researchers in the research area of energetic materials still utilize the group contribution approach, where the molar volume (including void) is obtained by summing up the volume of each atom or molecular fragment (group).[13] In the society of energetic materials, parameters developed by Stine were most frequently utilized.[14] Stine developed 34 parameters representing specific types of atoms by compiling more than 2000 crystals. Recently, Ammon and Mitchell developed 78 parameters corresponding to each group as well as atom by examining more than 11000 crystal structure data.[15] Although these approaches will eventually faded away when accurate 3-dimensional modeling of crystals is regularly available, they will remain as a workhorse in the arena of explosives modeling.

We wish to present a novel approach in predicting densities of energetic materials by using a holographic quantitative structure-activity relationship (HQSAR) method. This method has recently been developed at the Tripos Inc.[16] HQSAR is a branch of new QSAR techniques, and doesn't require 3D molecular information. In HQSAR, each molecule is divided into structural fragments that are counted in the bins of a fixed length array to form a molecular hologram.[16,17] A number of parameters related to hologram generation effect the HQSAR model. These are hologram length, fragment size, and fragment distinction, where is based on atoms, bonds, connections, hydrogens, and chirality. The bins are occupied by structural descriptors (independent variables) encoding compositional and topological molecular information. Then, QSAR model is generated through PLS regression by deriving a linear regression equation that correlates variation in structural information with variation in property data.

We have carefully selected 449 energetic molecules, the observed density of which are well reported in the ICT thermochemical database. All these molecules were sketched by using the SYBYL program (version 6.4). These compounds have either nitro or azido groups, and are main ingredients of explosives and/or propellants. In order to assess the predictability of the HQSAR models, 49 molecules were left out to utilize as a test set. In choosing compounds for a test set, we carefully selected the compounds the density distribution of which was similar to that of total data set. In addition, the test data included some of well-known explosives such as 2,4,6-trinitrotoluene (TNT), hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX), and octahydro-1,3,5,7-tetranitro-1,3,5,7-tetraazocine (HMX). Four hundred compounds were used to train the HQSAR models, and these comprise a training set. We have performed HQSAR calculations with 400 molecules by varying fragment size and hologram length. All the options for fragment distinction, except chirality, were utilized in this work. Various fragment sizes including 2-6, 3-6, 4-7, and 5-8 were tested.

†Presented in part at the 86th Annual Meeting of the Korean Chemical Society, Gyeongsang National Univ., Jinju, Korea, Oct. 2000.

Table 1. Comparison of HQSAR Results Due to the Use of the Different Hologram Lengths and Fragment Lengths

Run No.	Hologram Length	Fragment Size	Training set r^2	Training set q^2	Test Set r^2	Number of Components	
1	353	2-6	0.767 (0.070)[a]	0.648 (0.086)	0.672 (0.071)	9	
2	401	2-6	**0.941** (0.036)	**0.747** (0.074)	**0.819** (0.064)	23	**Model I**
3	455	2-6	**0.927** (0.040)	**0.696** (0.081)	**0.782** (0.064)	18	**Model II**
4	353	3-6	0.748 (0.073)	0.613 (0.090)	0.648 (0.072)	9	
5	401	3-6	**0.936** (0.037)	**0.690** (0.082)	**0.794** (0.067)	23	**Model III**
6	455	3-6	**0.934** (0.038)	**0.693** (0.082)	**0.733** (0.071)	20	**Model IV**
7	353	4-7	0.750 (0.073)	0.503 (0.103)	0.486 (0.072)	11	
8	401	4-7	0.791 (0.061)	0.553 (0.097)	0.607 (0.076)	12	
9	455	4-7	0.770 (0.070)	0.539 (0.099)	0.585 (0.080)	11	
10	353	5-8	0.695 (0.080)	0.447 (0.108)	0.465 (0.079)	11	
11	401	5-8	0.844 (0.058)	0.423 (0.112)	0.511 (0.117)	18	
12	455	5-8	0.762 (0.071)	0.412 (0.112)	0.547 (0.084)	13	

[a]Values in parentheses are average errors.

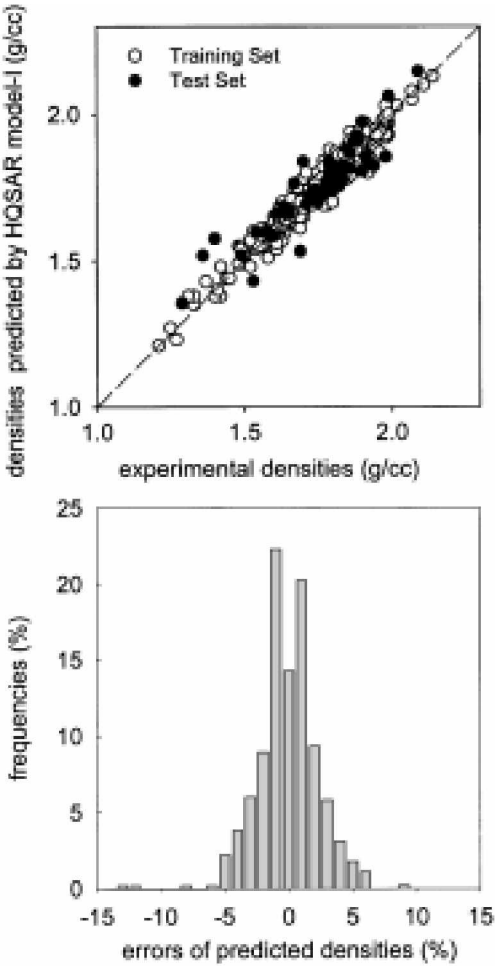

Figure 1. Plot of densities predicted by HQSAR **Model-I** against experimental densities (up) and histogram of predicted errors (down).

In addition, the hologram lengths up to 455 were employed. Our HQSAR results are summarized in Table 1. According to our results, several models, *i.e.* hologram lengths 401 and 455 at the molecular fragment of 2-6, and hologram lengths 401 and 455 at the molecular fragment of 3-6, provided reasonably good results. We assigned these models from **Model I** to **IV**. Calculated density values from these models against experimental density values are plotted in Figure 1. We have examined the values of r^2 and q^2 (q^2 is also called as the cross-validated r^2, whose value usually indicates how good the model is in prediction.[18]) to assess the quality of these HQSAR models. The r^2 values of these 4 models from our HQSAR results are 0.93-0.94, which appears to be reasonably good, and the q^2 values of these models are estimated to be 0.75-0.69, which shows our HQSAR models have good predictability. Based on these values, the **Model I** is considered as the best model among the options we had investigated. We also have utilized a test set in order to see the predictability of the models. According to the **Model I**, the r^2 values of the test set are calculated to be 0.82, which appears to be a reasonable value, although it is slightly lower than that of the training set.

As mentioned previously, group contribution approach has been widely used to predict the densities of highly energetic materials. To evaluate the usefulness of our HQSAR models critically, all the density values of the compounds were calculated by using the Stine's method and the Ammon's method again, and compared those density values with our HQSAR predicted values. This comparison will provide the usefulness of the HQSAR models. Both group contribution approaches have two sets of parameters depending upon the fitting schemes, *i.e.* linear and nonlinear. Both approaches are known to yield quite similar results.[14,15] which is also confirmed in our studies. Since the nonlinear fitting is slightly better than the linear fitting, our discussion will be proceeded only with the results of nonlinear fittings. The results from the Stine's method and the Ammon's method were depicted in Figure 2 and 3, respectively. As it is shown in Figures 2 and 3, in some energetic materials, the predictions based on the group contribution approaches had large deviations. The r^2 values of Stine's and Ammon's methods were low, *i.e.* 0.736 and 0.751, respectively, and were somewhat disappointing. The densities predicted by HQSAR models are far superior to these group contribution approaches, which are quite ubiquitous in the society of energetic materials. Another advantage of the HQSAR method is ultra fast, once the HQSAR model has been established. We currently continue to elaborate our HQSAR models to

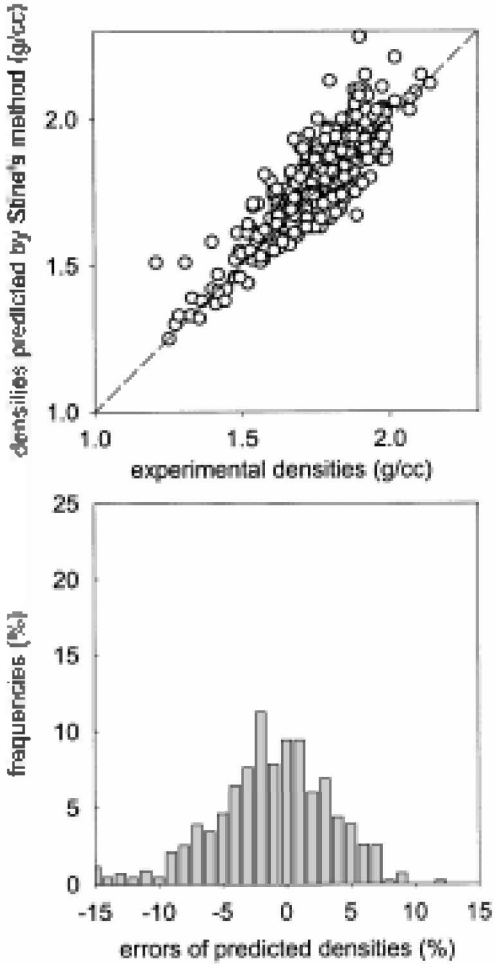

Figure 2. Plot of densities predicted by Stine's method against experimental densities (up) and histogram of predicted errors (down).

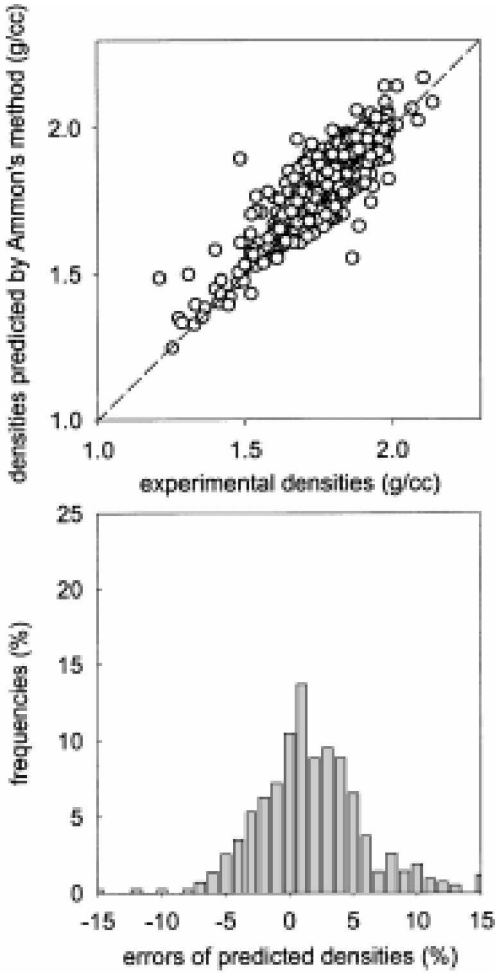

Figure 3. Plot of densities predicted by Ammon's method against experimental densities (up) and histogram of predicted errors (down).

increase the accuracy in the prediction, and believe this method to find itself also in the modeling of energetic molecules.

Acknowledgment. The authors thank Ms. Su Yeon Kim and T&J Tech for the help to use HQSAR program.

References

1. Persson, P-A.; Holmberg, R.; Lee, J. *Rock Blastings and Explosives Engineering*; CRC Press: Boca Raton, FL, 1993.
2. Fried, L. E.; Howard, W. M.; Souers, P. C. *Cheetah 2.0 User Manual*, Lawrence Livermore National Laboratory Report UCRL-MA-117541 Rev. 5; 1998.
3. Goh, E. M.; Cho, S. G.; Park, B. S. *J. Def. Tech. Res.* **2000**, *6*, 91.
4. (a) Gavezzotti, A. *Acc. Chem. Res.* **1994**, *27*, 309. (b) Verwer, P.; Leusen, F. J. J. Computer Simulation to Predict Possible Crystal Polymorphs. Reviews in *Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 1998; Vol. 12, Chapter 7.
5. (a) Gavezzotti, A.; Filippini, G. *Acta Cryst.* **1993**, *B49*, 868. (b) Gavezzotti, A. *J. Am. Chem. Soc.* **1983**, *105*, 5220. (c) Gavezzotti, A. *J. Am. Chem. Soc.* **1991**, *113*, 4622. (d) Gavezzotti, A. *J. Phys. Chem.* **1991**, *95*, 8943. (e) Gavezzotti, A.; Filippini, G. *J. Phys. Chem.* **1994**, *98*, 4831.
6. (a) Williams, D. E. *Acta Cryst.* **1996**, *A52*, 326. (b) Hall, D.; Williams, D. E. *Acta Cryst.* **1975**, *A31*, 56. (c) Williams, D. E.; Gao, D. *Acta Cryst.* **1998**, *B54*, 41. (d) Williams, D. E.; Gao, D. *Inorg. Chem.* **1997**, *36*, 782.
7. Holden, J. R.; Du, Z.; Ammon, H. L. *J. Comput. Chem.* **1993**, *14*, 422.
8. (a) Perlstein, J. *Chem. Mater.* **1994**, *6*, 319. (b) Perlstein, J. *J. Am. Chem. Soc.* **1991**, *114*, 1955. (c) Perlstein, J. *J. Am. Chem. Soc.* **1994**, *116*, 455. (c) Perlstein, J. *J. Am. Chem. Soc.* **1994**, *116*, 11420. (d) Perlstein, J. *J. Am. Chem. Soc.* **1994**, *116*, 11420.
9. (a) Gibson, K. D.; Scheraga, H. A. *J. Phys. Chem.* **1995**, *99*, 3752. (b) Wawak, R. J.; Gibson, K. D.; Liwo, A.; Scheraga, H. A. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 1743. (c) Wawak, R. J.; Pillardy, J.; Liwo, A.; Gibson, K. D.; Scheraga, H. A. *J. Phys. Chem. A* **1998**, *102*, 2904. (d) Pillardy, J.; Wawak, R. J.; Arnautova, Y. A.; Czaplewski, C.; Scheraga, H. A. *J. Am. Chem. Soc.* **2000**, *122*, 907.
10. (a) Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M. *J. Phys. Chem.* **1996**, *100*, 7352. (b) Price, S. L.; Wibley,

K. S. *J. Phys. Chem. A* **1997**, *101*, 2198. (c) Beyer, T.; Price, S. L. *J. Phys. Chem. B* **2000**, *104*, 2647.
11. Mooji, W. T. M.; van Eijck, B. P.; Kroon, J. *J. Am. Chem. Soc.* **2000**, *122*, 3500.
12. (a) Cromer, D. T.; Ammon, H. L.; Holden, J. R. *A Procedure for Estimating the Crystal Densities of Organic Explosives*, Los Alamos National Laboratory Report LA-11142-MS; Nov. 1987. (b) Sorescu, D. C.; Rice, B. M.; Thompson, D. L. *J. Phys. Chem. A* **1998**, *102*, 6692. (c) Sorescu, D. C.; Rice, B. M.; Thompson, D. L. *J. Phys. Chem. A* **1998**, *102*, 8386. (d) Sorescu, D. C.; Thompson, D. L. *J. Phys. Chem. B* **1999**, *103*, 6774. (e) Lewis, J. P.; Sewell, T. D.; Evans, R. B.; Voth, G. A. *J. Phys. Chem. B* **2000**, *104*, 1009.
13. Tarver, C. M. *J. Chem. Eng. Data* **1979**, *24*, 136.
14. Stine, R. S. *Prediction of Crystal Densities of Organic Explosives by Group Additivity*, Los Alamos National Laboratory Report LA-8920; Aug. 1981.
15. Ammon, H. L.; Mitchell, S. *Propell. Explos. Pyrotech.* **1998**, *23*, 260.
16. Lowis, D. R. *HQSAR. A New, Highly Predictive QSAR Technique*. Tripos Technical Notes; Oct. 1997; Vol. 1, No. 5.
17. (a) Winker, D. A.; Burden, F. R. *Quant. Struct.-Act. Relat.* **1998**, *17*, 224. (b) Seel, M. S.; Turner, D. B.; Willett, P. *Quant. Struct.-Act. Relat.* **1999**, *18*, 245. (c) Tong, W.; Lowis, D. R.; Perkins, R.; Chen, Y.; Welsh, W. J.; Goddette, D. W.; Heritage, T. W.; Sheehan, D. M. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 669. (d) So, S.-S.; Karplus, M. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 243.
18. Cramer, R. D. III; DePriest, S. A.; Patterson, D. E.; Hecht, P. In *3D QSAR in Drug Design: Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, 1993; p 443.