Gas-Phase Molecular Clustering of TIP4P and SPC/E Water Models from Higher-Order Virial Coefficients†

Kenneth M. Benjamin, Andrew J. Schultz, and David A. Kofke*

Department of Chemical and Biological Engineering, University at Buffalo, The State University of New York, Buffalo, New York 14260-4200

Higher-order virial coefficients (up to B_6) for TIP4P and SPC/E water models are used to characterize molecular clusters (up to hexamers) formed by water at various gas-phase thermodynamic state points between 298 and 773 K. Comparison of cluster statistics with available molecular simulation data for the same models indicates that the virial approach is effective at characterizing the clustering behavior. Significant deviations from experimentally confirmed ab initio results from the literature at 298 K are ascribed to inadequacies in the TIP4P model and to differences in the treatment of "physical" versus "chemical" association in the two approaches. At two conditions where an analysis could be made, the concentration of clusters that are only physically associated was found to be ∼30% of the concentration of those that are chemically associated (hydrogen bonded).

1.0. Introduction

The formation of clusters in gaseous systems is well-known.[1,2] Water clusters, in particular, have been the focus of much investigation.[1] A proper understanding of water clusters is relevant to several areas of scientific research and engineering application:

- pressure−volume−temperature properties of water vapor and its mixtures. The behavior of water clusters in the vapor also adds to our understanding of how water behaves as a liquid and in the form of clathrates;
- atmospheric chemistry of acid rain formation;
- nucleation of water droplets;
- absorption of infrared radiation by water in the atmosphere;
- reactions in supercritical water; and
- isotope fractionation.

The notion of a gas-phase cluster is ambiguous, even for strongly associating molecules such as water. The appropriate definition can, to some extent, depend on the application of interest. For example, absorption of solar energy by atmospheric water has important contributions from vibrational motions performed by the cluster, and this requires that they be in some type of metastable arrangement that is stabilized by hydrogen bonding. On the other hand, the process of nucleation from a vapor is largely concerned with the degree to which molecules are in loose proximity, interacting significantly without necessarily being bound in a tight metastable configuration.[3] The association involved in these two cases is sometimes characterized as "chemical" versus "physical" association. The difference in cluster concentration based on the two definitions can be significant.[4]

Experimental techniques for investigation of clusters are based primarily on spectroscopy,[5−13] and consequently, they probe the behavior of chemically associated clusters. Methods based on computational quantum chemistry[14−16] have a similar focus, and many detailed studies have been performed of metastable water clusters with the aim to develop model potentials that match spectroscopic data. Molecular simulation has also been applied to the study of vapor-phase association, often with a broader view on the nature of clustering.[2,4,17−19] Simulation requires the adoption of an empirically developed intermolecular force field, or potential, and the quality of the potential will limit the quality (or at least the realism) of the results. Simulation studies must also confront the question of the definition of a cluster, and this is both a strength and weakness of such studies. Simulation can distinguish between chemically and physically associated clusters, assuming that a suitable definition has been provided for the two cases. For physical clustering. this definition can be problematic. Unlike chemical clustering, there are no obvious, well-localized, geometric or energetic criteria for the cutoff of physical association. Calculated cluster concentrations can be sensitive to the choice of this definition.

An underutilized approach to the analysis of clustering is based on the virial equation of state. The virial equation of state (VEOS) can describe successfully the pressure−volume−temperature behavior of fluids at low to moderate densities. The mathematical form of the VEOS[20] is

$$\frac{\beta P}{\rho} = 1 + B_2\rho + B_3\rho^2 + B_4\rho^3 + B_5\rho^4 + B_6\rho^5 + ... \quad (1)$$

where P is the pressure, $\rho = N/V$ is the number N of molecules in the volume V, $\beta = (kT)^{-1}$ is the reciprocal temperature in energy units, and B_i represents the ith virial coefficient. Each B_i relates directly to the interactions between exactly i molecules, and consequently, the virial coefficients can be connected to the equilibrium of clusters in a fluid state.[5,20−23] In the virial approach to clustering, the system is treated as an ideal-gas mixture of noninteracting but chemically equilibrated clusters.[21,23] If two or more molecules interact a way that affects the thermodynamic properties, they are accounted for by the corresponding term in the VEOS and, in this analysis, they constitute a cluster. The size of a cluster is defined in terms of the order of the variation of its concentration with density (or pressure). Thus, a "cluster" in this view is amorphous; it is not identified with a specific arrangement of molecules in space. Nevertheless, it does provide a meaningful and nonarbitrary definition of clustering. In this approach, the calculated cluster concentrations describe the prevalence of associated molecules, modulated by their strength of interaction. Clusters are defined in shades of gray, not black and white.

* Corresponding author. Tel.: (716) 645-2911 ext. 2209. Fax: (716) 645-3822. E-mail: kofke@buffalo.edu.

† Part of the special issue honoring Eduardo Glandt.

10.1021/ie051160s CCC: $33.50
Published on Web 01/05/2006

In this context, it is worthwhile to mention an important class of equations of state that are based on a picture in which chemical association leads to an equilibrium distribution of oligomers that interact via physical forces.[24−28] The chemical equilibrium is modeled by postulating values for the equilibrium constants for consecutive association reactions, and the physical interactions between clusters are handled via a standard (e.g., cubic) equation of state written for a mixture of oligomers, sometimes with mixing rules designed to simplify the resulting form. The present work contrasts with such methods in several ways. First, the virial approach makes no assumptions about the association equilibrium constants—instead these values are given directly in terms of the virial coefficients, which in turn follow directly from the postulated intermolecular potential (details follow). Second, in the virial approach, all of the interactions are treated via cluster equilibrium. Interactions between smaller clusters are (in principle) handled by higher-order virial coefficients. This means, for example, that the monomer−dimer equilibrium does not depend on the assumed form for interactions between dimers. Finally, virial coefficients do not (as normally defined) distinguish between chemical and physical interactions—both can contribute to the concentration of a cluster. If association is strong, the chemical effect dominates the value of B_n and the distinction between physical and chemical contributions is not important. In such cases, the clusters identified by the virial approach coincide with the intuitive notion of a cluster as a close, metastable arrangement of molecules, a notion that underlies the association-based equations of state. Then equilibrium constants determined from the virial coefficients can be used to guide the formulation of equilibrium constants used in association-based equations of state.

Interest in virial-based approaches to clustering analysis has been limited in part by the difficulty of obtaining values for high-order virial coefficients. While this obstacle still remains with regard to values obtained by experiment, computational approaches are becoming feasible. A recent investigation[29,30] has proposed "Mayer sampling" as a new class of methods for obtaining higher-order virial coefficients of various fluids from a model for the intermolecular potential. From these values, one can determine the amounts of different clusters that form, as a function of the thermodynamic state. Such an approach is likely to be more efficient than observations of spontaneous clustering in molecular simulations of the same model, and part of the aim of this work is to show that information from virial coefficients can be used this way.

This paper presents a cluster analysis for water, using (classical) higher-order virial coefficients as a guide. While other investigations[5,22] have explored fluid properties for water from virial coefficients in a manner similar to the one to be outlined here, most have dealt with only the second and third virial coefficients. Here we describe water properties and clustering using up to the sixth virial coefficient. The study focuses on water models TIP4P and SPC/E because molecular simulation data are available about their clustering behavior, and we can use this information to gauge the effectiveness of the virial-clustering approach. In the next section, we briefly review the important aspects of Mayer sampling molecular simulation, including the choice of sampling technique. Then we describe the mathematical formalism for treating chemical association (and the formation of clusters in the gas-phase) via virial coefficients. In Section 4, we examine clustering behavior in the water models for conditions of a saturated vapor, a supercritical vapor, and vapor at ambient conditions. We also consider the behavior of the association equilibrium constants. We conclude in Section 5.

2.0. Virial Coefficients

The virial coefficients used in this study are reported elsewhere,[30] but for completeness we briefly review the methods used in those calculations and present the values here. Evaluation of virial coefficients from a molecular model is based on their expression in the form of cluster integrals.[20] For example, the second virial coefficient is

$$B_2(T) = -\frac{1}{2V}\int\int f_{12}\,\mathrm{d}\mathbf{r}_1\,\mathrm{d}\mathbf{r}_2 \tag{2}$$

and for pairwise-additive potentials, the third virial coefficient is

$$B_3(T) = -\frac{1}{3V}\int\int\int f_{12}f_{13}f_{23}\,\mathrm{d}\mathbf{r}_1\,\mathrm{d}\mathbf{r}_2\,\mathrm{d}\mathbf{r}_3 \tag{3}$$

In the integrals, f_{ij} is referred to as a Mayer function and has the form

$$f_{ij} = [\exp(-\beta u_{ij}) - 1] \tag{4}$$

where u_{ij} is the pair potential between molecules labeled i and j. In expressions such as eqs 2 and 3, one must integrate over all positions, rotations, and internal degrees of freedom available to each molecule. The higher-order coefficients, such as B_4−B_6, are sums of cluster integrals involving a correspondingly larger number of molecules.

The Mayer sampling method[29] was recently introduced for evaluation of the integrals appearing in the expressions for the virial coefficients. It is a Monte Carlo approach that adopts ideas used for the evaluation of free energies.[31,32] Molecular configurations are sampled using importance sampling[33] with weights based on the magnitude of the interactions that are represented in the given cluster integral. Simulation averages give the ratio of the desired cluster integral to that of a known reference integral, as follows,

$$\Gamma(T) = \Gamma_0\frac{\Gamma}{\Gamma_0} = \Gamma_0\frac{\langle\gamma/\pi\rangle_\pi}{\langle\gamma_0/\pi\rangle_\pi} \tag{5}$$

where $\Gamma(T)$ represents a general cluster integral or sum of integrals, $\gamma(\mathbf{r}^n,T)$ is the integrand (or sum of integrands), and $\pi(\mathbf{r}^n,T) = |\gamma|$ is the unnormalized probability distribution that governs sampling. The angle brackets indicate the "ensemble-average" integral over all configuration space. Quantities marked with the subscript "0" correspond to the specified reference system, for which a hard-sphere model provides a suitable choice. More information regarding the method can be found elsewhere.[29,30]

Equation 5 corresponds to a direct-sampling method, as it involves perturbations directly between the target system (which governs sampling) and the reference system. At low temperatures, the water molecules strongly prefer their own energetic wells, and many configurations important to the hard-sphere reference are not sampled and direct sampling fails. This signals that the hard-sphere phase space is no longer a subset of the water phase space,[31,32] but instead that the water's attractive wells are regions of phase-space overlap between the hard spheres and water. Overlap sampling is a desirable alternative to direct sampling in such situations. In this approach, separate simulations are performed of the target and reference systems,

Table 1. Virial Coefficients for the TIP4P Model of Water as Calculated Using the Mayer Sampling Method[30] [a]

T (K)	B_2 (L/mol)	B_3 (L/mol)^2	B_4 (L/mol)^3	B_5 (L/mol)^4	B_6 (L/mol)^5
298	-3.639_3	-93.4_{10}	-8251_{305}	35030_{723}	-377710_{151000}
350	-1.315_2	-4.06_4	-32_3	170_{74}	
370	-0.9752_8	-1.52_1	-5.2_3	47_5	
390	-0.7483_4	-0.605_2	-0.64_3	3.1_2	
410	-0.592_3	-0.2467_8	-0.015_2	1.30_7	-2.6_{21}
430	-0.4797_2	-0.0981_3		0.16_3	-2.1_7
450	-0.3961_2	-0.0341_1	0.0496_6	0.051_2	-0.31_9
470	-0.3328_2	-0.00628_7	0.0308_3	0.0090_4	-0.025_{10}
490	-0.2832_2	0.00589_5	0.0184_2	0.00016_9	-0.0049_3
510	-0.2443_2	0.01067_4	0.01096_7	-0.00155_5	-0.006_2
530	-0.2124_3	0.01211_3	0.00644_4	-0.00149_3	-0.0010_2
550	-0.1865_2	0.01199_2	0.00382_3	-0.00110_2	-0.00011_5
570	-0.1647_1	0.01126_2	0.00226_2	-0.00073_1	-0.00003_2
590	-0.1467_1	0.01025_1	0.001326_9	-0.000514_6	0.00004_2
610	-0.1313_1	0.00920_1	0.000786_6	-0.000337_3	0.00001_1
630	-0.1178_1	0.00819_1	0.000436_4	-0.000229_2	0.00002_1
650	-0.1065_1	0.00728_1	0.000239_3	-0.000154_2	0.000017_5
670	-0.0962_{15}	0.006467_5	0.000112_2	-0.000105_1	0.000019_3
690	-0.0877_1	0.005749_4	0.000039_2	-0.0000723_7	0.000008_2
710	-0.0797_1	0.005123_4	-0.000018_9	-0.0000499_5	0.000007_1
730	-0.0728_1	0.004573_3	-0.000031_1	-0.0000357_4	0.0000037_7
750	-0.0668_1	0.004098_3	-0.000047_1	-0.0000256_3	0.0000035_8

[a] Subscripted digits represent the confidence limits (standard error of the mean) for the rightmost digits of the value.

Table 2. Virial Coefficients for the SPC/E Model of Water as Calculated Using the Mayer Sampling Method.[29,30] [a]

T (K)	B_2 (L/mol)	B_3 (L/mol)^2	B_4 (L/mol)^3	B_5 (L/mol)^4	B_6 (L/mol)^5
373	-1.8049_4	-10.29_8	-243_{20}	577_{14}	15000_{4500}
423	-0.8865_4	-1.045_2	-2.85_{10}	15.17_{22}	$16.7_{2.6}$
450	-0.65288	-0.347906	-0.2559	2.2434_{332}	$-0.464_{0.23}$
473	-0.5189_1	-0.1373_3	0.0335_{10}	0.524_9	-0.472_{43}
500	-0.40816	-0.041058	0.049757	0.0954_{21}	-0.102_{13}
523	-0.33954_4	-0.00874_6	0.03369_{40}	0.0213_7	-0.028_2
550	-0.2788	0.006226	0.018926	0.0022_2	-0.00856_{68}
573	-0.23914_{10}	0.01076_1	0.01121_{30}	-0.00084_{11}	-0.0017_3
600	-0.20244	0.012047	0.006097	-0.001122_{54}	-0.00059_{11}
623	-0.17715_4	0.01165_1	0.00358_3	-0.00092_2	-0.00021_6
650	-0.15302	0.010582	0.00195	-0.000635_{18}	-0.00005_1
673	-0.13599_2	0.009530_6	0.00115_1	-0.000398_{10}	0.000028_{25}
700	-0.11919	0.008318	0.000636	-0.000258_7	0.0000084_{70}
723	-0.10714_{20}	0.007380_3	0.000377_5	-0.000184_6	0.000021_{10}
750	-0.09493	0.006392	0.000188	-0.000121_3	0.0000065_{17}
773	-0.08596_5	0.005674_3	0.000095_2	-0.000082_2	0.000010_2

[a] Subscripted digits represent the confidence limits (standard error of the mean) for the rightmost digits of the value.

and in each, a perturbation average is performed into their region of overlap. One defines an overlap function to represent mathematically only those regions important to both hard spheres and water,

$$\gamma_{OS} = \frac{|\gamma_0||\gamma|}{\alpha|\gamma_0| + |\gamma|} \tag{6}$$

where γ_{OS} is the overlap function and α is an optimization parameter. The desired cluster integral is given as the ratio of averages taken in the two systems

$$\Gamma(T) = \Gamma_0 \frac{\{\langle\gamma/\pi\rangle_\pi\}/\{\langle\gamma_{OS}/\pi\rangle_\pi\}}{\{\langle\gamma_0/\pi_0\rangle_{\pi_0}\}/\{\langle\gamma_{OS}/\pi_0\rangle_{\pi_0}\}} \tag{7}$$

where $\pi_0 = |\gamma_0|$. The parameter α and the amount of sampling performed for each system can be selected to optimize the calculation.[30]

Values of the virial coefficients calculated as described above and used in the cluster analyses are presented in Tables 1 and 2 for the TIP4P and SPC/E models, respectively.

3.0. Chemical Association via Virial Coefficients: Gas-Phase Clusters

Woolley[20,21] developed an approach mathematically relating virial coefficients to cluster composition in an ideal-gas mixture of clusters. The ideal-gas equation of state for a mixture of j-numbered clusters is

$$PV = kT\sum N_j \tag{8}$$

where N_j is the number of clusters containing j molecules. Also,

$$\sum jN_j = N \tag{9}$$

where N is the total number of molecules in the system. For the equilibrium reaction for the formation of a j-cluster of species X,

$$jX \leftrightarrow X_j \tag{10}$$

one can define an ideal-gas equilibrium constant

$$K_j = \frac{P_j}{(P_1)^j} \tag{11}$$

where P_j is the partial pressure of a j-cluster species, which can be calculated via

$$P_j = \frac{N_j kT}{V} \tag{12}$$

These partial pressures can be summed to yield the total pressure:

$$P = P_1 + P_2 + P_3 + P_4 + P_5 + P_6 + ... \tag{13}$$

Combination of eqs 11 and 13 yields

$$P = P_1 + K_2 {P_1}^2 + K_3 {P_1}^3 + K_4 {P_1}^4 + K_5 {P_1}^5 + K_6 {P_1}^6 + ... \tag{14}$$

This series can be inverted to produce a virial series for P_1 in powers of the total pressure, which then can be used in this result from Woolley,[21]

$$\frac{PV}{NkT} = \frac{P}{P_1}\frac{dP_1}{dP} \tag{15}$$

to obtain an expression which is a virial series in total pressure, P. The individual terms of this virial expression can be compared to the pressure expansion form of the virial equation of state to connect the various equilibrium constants for cluster formation (K_j) to specific virial coefficients (B_j):

$$B_2 = -K_2 kT \tag{16}$$

$$B_3 = (-2K_3 + 4{K_2}^2)(kT)^2 \tag{17}$$

$$B_4 = (-20{K_2}^3 + 18K_2K_3 - 3K_4)(kT)^3 \tag{18}$$

$$B_5 = (112{K_2}^4 + 18{K_3}^2 - 144{K_2}^2K_3 + 32K_2K_4 - 4K_5)(kT)^4 \tag{19}$$

$$B_6 = (-672{K_2}^5 + 1120{K_2}^3K_3 - 315K_2{K_3}^2 - 280{K_2}^4K_3 + 60K_3K_4 + 50K_2K_5 - 5K_6)(kT)^5 \tag{20}$$

These expressions can be inverted to provide K_2−K_6 as functions of B_2−B_6.

$$K_2 = -\beta B_2 \tag{21}$$

$$K_3 = \frac{1}{2}\beta^2(4{B_2}^2 - B_3) \tag{22}$$

$$K_4 = \frac{1}{3}\beta^3(9B_2B_3 - 16{B_2}^3 - B_4) \tag{23}$$

$$K_5 = \frac{1}{24}\beta^4(-360{B_2}^2B_3 + 400{B_2}^4 + 64B_2B_4 + 27{B_3}^2 - 6B_5) \tag{24}$$

$$K_6 = \frac{1}{10}\beta^5(720{B_2}^3B_3 - 576{B_2}^5 - 135B_2{B_3}^2 - 160{B_2}^2B_4 + 20B_3B_4 + 25B_2B_5 - 2B_6) \tag{25}$$

Once the equilibrium constants have been obtained, one can determine the cluster distribution using standard reaction equilibria calculations.[34]

4.0. Results and Discussion

4.1. Pressure−Volume−Temperature Behavior of Water. The values of the virial coefficients for TIP4P and SPC/E water used in this investigation are reported in Tables 1 and 2. We cannot expect the virial equation to describe clustering behavior any better than it gives the equation of state itself, so it is helpful first to examine the range of validity of the virial equation as given in eq 1. This information is presented in Figures 1 and 2.

Figure 1 shows curves of $(Z - 1)$ [where $Z = PV/NRT$ is the compressibility factor] versus reduced density predicted by the 6th-order virial equation of state for TIP4P water (VEOS6-TIP4P) for reduced temperatures between 0.96 and 1.29. The curves from VEOS6-TIP4P are compared with molecular simulation data for TIP4P water.[18] For the two highest temperatures, VEOS6-TIP4P provides excellent agreement up to reduced densities ($\rho_r = \rho/\rho_c$) of 0.35, while at the lowest temperature, the virial equation begins to fail at about ρ_r equal

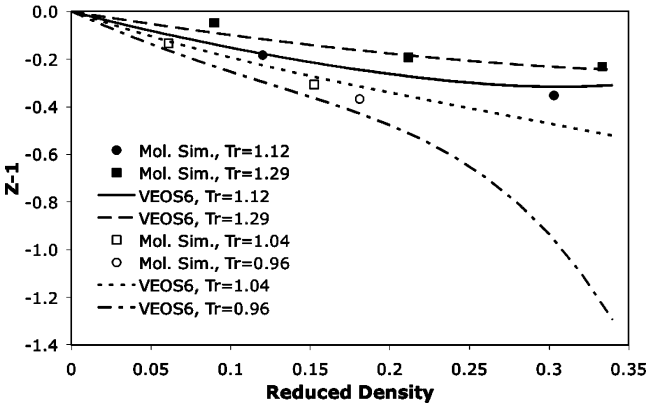

Figure 1. Deviation from ideality as given by VEOS6-TIP4P for TIP4P water. Plotted is $(Z - 1)$ vs reduced density ρ/ρ_c, where Z is the compressibility factor. Lines are EOS predictions and points are molecular simulation data[18] at the following reduced temperatures T/T_c: ■ 1.29, ● 1.12, □ 1.04, ○ 0.96.

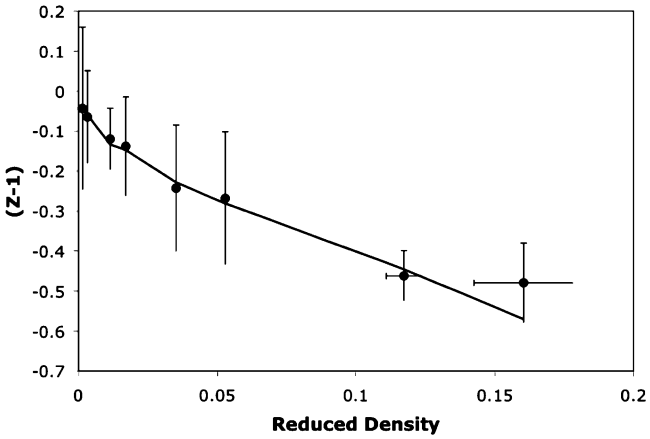

Figure 2. Deviation from nonideality predicted by VEOS6−SPC/E along the coexistence curve for SPC/E water. Plotted is $(Z - 1)$ vs reduced density ρ/ρ_c, where Z is the compressibility factor. Line is EOS predictions, and points are molecular simulation data.[35]

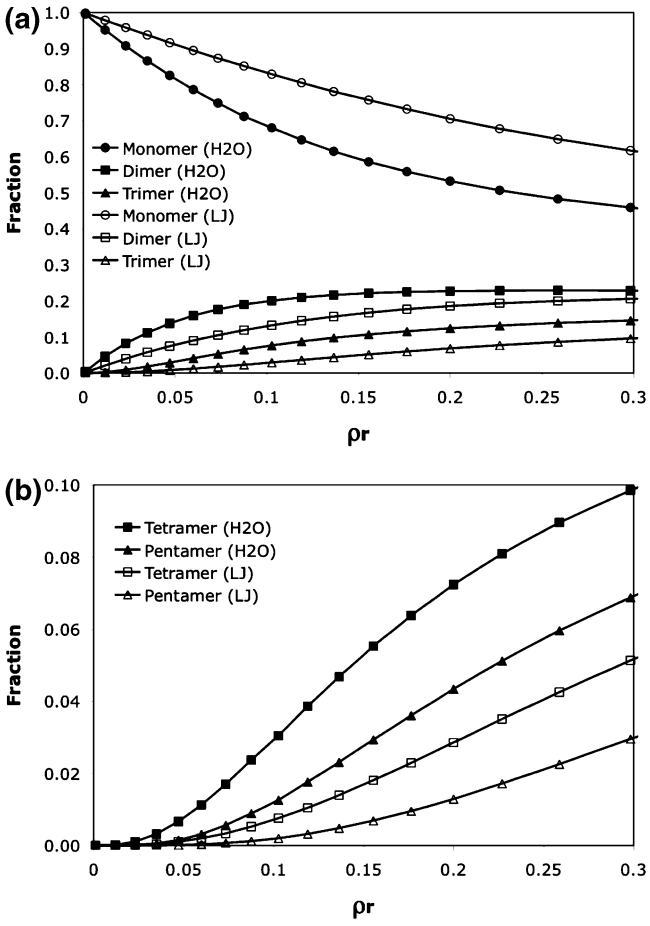

Figure 3. Fraction of water molecules in clusters for TIP4P water and Lennard-Jones fluid (LJ) at $T_r = 1.05$: (a) ● monomer (H_2O), ○ monomer (LJ), ■ dimer (H_2O), □ dimer (LJ), ▲ trimer (H_2O), △ trimer (LJ); (b) ■ tetramer (H_2O), □ tetramer (LJ), ▲ pentamer (H_2O), △ pentamer (LJ).

to 0.2. These results describe the range for which we can usefully interpret information about clustering for the TIP4P model.

Figure 2 shows curves of $(Z - 1)$ versus reduced density predicted by VEOS6-SPC/E, the 6th-order virial equation of state for SPC/E water, along the coexistence curve. Here, the VEOS6-SPC/E predictions are compared to the molecular simulation data of Boulougouris et al.[35] The agreement between VEOS6-SPC/E and the simulation confirms that it is acceptable to use VEOS6-SPC/E in the chemical association calculations for clustering up to reduced densities of ~0.15.

4.2. Molecular Clustering of Water. We can gauge the importance of chemical versus physical association by examining the clustering behavior of water against a fluid model that does not have specific association interactions. To this end, parts a and b of Figure 3 show the cluster compositions for a Lennard-Jones (LJ) fluid and TIP4P water at a reduced temperature (defined in terms of the critical point for each model) of 1.05. It is evident from this figure that the amount of clustering found in water is substantially higher than that found in the LJ fluid, reflecting the importance of the hydrogen-bonding interactions. In this regard, one should note that the approach fails entirely in situations where the second virial coefficient is positive. Adaptations such as those used in the association equations of state—in which physical interactions (or at least repulsions) get special treatment—are needed to handle these more-general cases.

4.2.1. Supercritical Conditions (TIP4P). Chemistry in supercritical water is an area of research that is concerned greatly with the cluster composition of water. The extent of water clustering can affect the solvation of various reactants, transition states, and products during elementary reactions.[36] Also, different water clusters (dimers, trimers, etc.) could become potential reactants and provide alternative reaction pathways. In addition, the clustering is a function of the thermodynamic state (temperature and pressure) of the water.[4,17,18] A thorough knowledge of the cluster composition is helpful in understanding the structure of the local molecular environment, deducing the fundamental reaction mechanism and energetics, and selecting appropriate reaction conditions.

Table 3. Comparison of Average Cluster Sizes, $\langle S \rangle$, Predicted by Monte Carlo Simulation[18] and from the Virial-Clustering Approach

state point	T_r	ρ_r	$\langle S \rangle^{18}$	$\langle S \rangle$ (virial-clustering w/ VEOS6)	error (%)
A	0.96	0.18	1.82	1.45	21
B	1.04	0.06	1.15	1.14	0.9
C	1.04	0.15	1.39	1.33	4.0
D	1.12	0.12	1.23	1.21	1.6
E	1.29	0.09	1.12	1.10	1.8
F	1.29	0.34	1.48	1.35	8.7

Figure 3a shows that, for TIP4P water, there is nonnegligible formation of dimers and trimers in the vapor phase at supercritical conditions at reduced densities <0.3. There is an observable amount of tetramers and pentamers in Figure 3b, but these are found at a much lower concentration than the smaller clusters (one should note in examining the figures that the quantities plotted are not the cluster fractions but are the fraction of atoms in a cluster of a given size; this measure gives greater weight to the larger clusters). This behavior is in agreement with spectroscopic observations of clustering in real water at the same reduced temperature.[6]

Kalinichev and Churakov[18] probed the sizes of clusters in near-critical and supercritical water using the TIP4P water model and Monte Carlo simulation. In their investigation, the authors calculated average cluster sizes at different thermodynamic state points by employing hybrid geometric−energetic criteria for hydrogen bonding. Table 3 compares the predicted average cluster sizes based on VEOS6-TIP4P with the average cluster sizes reported from the Monte Carlo simulation study. Comparison is made only for conditions where the VEOS gives a good description of the TIP4P equation of state (Figure 1). With the exception of the (lowest-temperature) point A, all of the average cluster sizes predicted from VEOS6-TIP4P agree within 9% of the Kalinichev and Churakov values taken from Monte Carlo simulations. The level of agreement is excellent over a wide range of reduced temperatures and densities.

4.2.2. Saturated Vapor Conditions (SPC/E). Next we consider states located along the vapor−liquid coexistence curve. Johansson et al.[37] performed Gibbs-Ensemble Monte Carlo (GEMC) simulations of the SPC/E model for water and analyzed clusters that formed in the vapor phase. They considered clusters defined according to two different criteria:

1. A purely distance-based criterion between oxygen atoms, which indicates the presence of a physical or chemical (hydrogen-bonded) cluster.

2. A combined distance and energy criterion. In this case, there is a minimum distance required between an oxygen atom of one molecule and a hydrogen atom of another. In addition, there is a required maximum energy (intermolecular energy) between an oxygen atom of one molecule and a hydrogen atom of another. Satisfying these criteria indicates the presence of a hydrogen-bonded cluster.

The results of Johansson et al.[37] provide a suitable set of data for comparison to the virial-based approach to cluster charac-

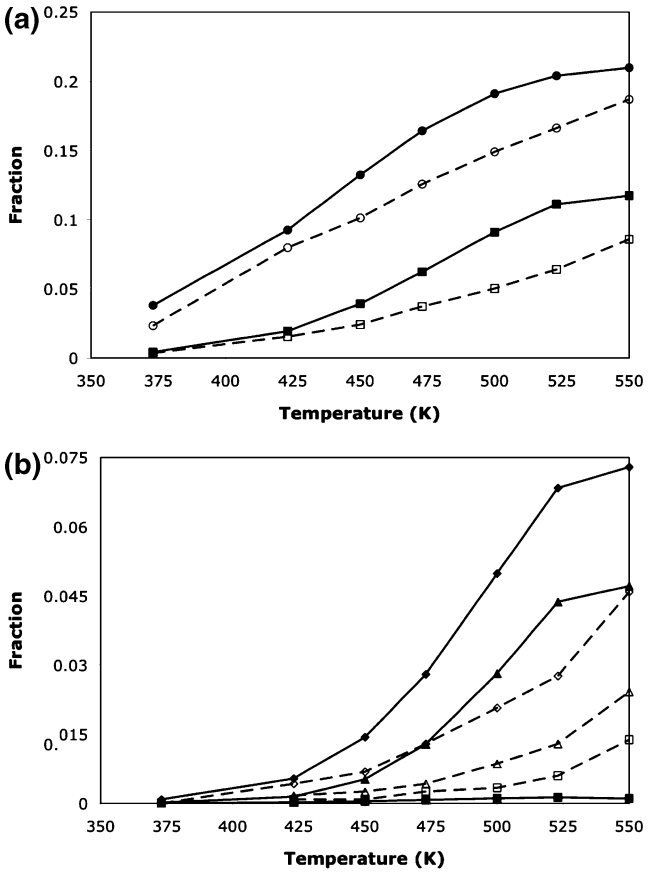

Figure 4. Fraction of water molecules in clusters in the vapor phase along the coexistence curve of SPC/E water, GEMC simulations[37] (H-bonded clusters) vs virial-clustering approach: (a) ● dimer−virial, ○ dimer−GEMC, ■ trimer−virial, □ trimer−GEMC; (b) ◆ tetramer−virial, ◇ tetramer−GEMC, ▲ pentamer−virial, △ pentamer−GEMC, ■ hexamer−virial, □ hexamer−GEMC.

terization. Parts a and b of Figure 4 show the comparison, with the virial results including coefficients up to B_6. The plots show the fraction of molecules that are present in a cluster of size n, for n from 2 (dimer) to 6 (hexamer). Results are shown over the range of conditions for which the VEOS is effective, as determined in Figure 2. The cluster fractions for $n = 2-5$ given by the virial approach are consistently greater than those counted by GEMC simulations, while the fraction for $n = 6$ from the virial approach is less than that counted by GEMC simulation. The overestimation by the virial approach for $n = 2-5$ can be attributed to differences in the way clusters are defined here and by Johansson et al. The detailed data reported by Johansson et al. consider only hydrogen-bonded clusters (definition 2, above) and, thus, exclude clusters that are physically associated but not hydrogen bonded. At higher temperatures, hydrogen-bonding association is diminished relative to physical association, and the GEMC simulations will overlook even more clusters that are counted in the virial approach. Accordingly, the discrepancy between the GEMC and the virial results increases with increasing temperature. (In both approaches, the cluster fraction still increases because the density is increasing with temperature along the saturation line). Johansson et al. report that the number of physical clusters they observe is consistently larger than the number of hydrogen-bonded clusters, which would tend to bring the simulation data more in line with the virial characterization. They do not provide detailed data that would permit us to perform the quantitative comparison, but from data given in a related figure, it would seem that inclusion of the physical clusters raises the total number of clusters by 30% at the highest temperature considered, which is consistent with the discrepancy from our results.

Table 4. Comparison of Water Cluster Concentrations in Saturated Air at 298 K with $P^{vap}(H_2O) = 0.031$ atm−Computational Quantum Chemistry[16] vs Virial Coefficient Method

cluster	cluster concentration from ab initio/DFT (clusters/cm^3)	cluster concentration from virial coefficient (clusters/cm^3)
dimer	9.0E+14	3.5E+15
trimer	2.6E+12	8.7E+13
tetramer	5.8E+11	6.0E+12
pentamer	3.5E+10	2.3E+11
hexamer	NA	3.3E+11

4.2.3. Ambient Conditions. As mentioned previously, another area of research heavily interested in water clusters is atmospheric chemistry. It has long been speculated that water clusters may be abundant in atmospheric systems and may play important roles in physical and chemical processes contributing to global warming.[1,2] Dunn et al.[16] attempted to quantify the concentrations of water clusters up to pentamers in the atmosphere via computational quantum chemistry, which necessarily treats only chemically associated clusters. Table 4 compares the results of Dunn et al. with cluster concentrations calculated from TIP4P virial coefficients for water vapor in saturated air at 298 K. The dimer and pentamer concentrations given by the two approaches differ by a factor of 4−6, and trimer and tetramer concentrations both differ by >1 order of magnitude. The DFT (density functional theory) results of Dunn et al. reproduce well the experimentally measured dimer concentrations at low temperatures. There are two reasons for the discrepancy with our results. First, the magnitude of the TIP4P second virial coefficient is ∼3× larger than that of real water,[38] and this error translates to a tripling of the dimer concentration for the TIP4P model; there is also a significant effect on the higher-order clusters. Dividing the predicted dimer concentration of 3.5×10^{15} clusters/cm^3 by this factor of 3 yields a dimer concentration of 1.17×10^{15} clusters/cm^3 and accounts for most of the disagreement between the virial-cluster and DFT results. The rest of the discrepancy with Dunn et al. dimer results must be ascribed to the difference between chemical and physical association, with the conclusion that the number of water molecules that are only physically associated is ∼30% of the number that are chemically associated by hydrogen bonding.

Given that the TIP4P second virial coefficient deviates significantly from the value for real water, we cannot put too much stock in the results for the concentrations of the higher-order clusters. The results, however, should be internally consistent and correct for the TIP4P model. Thus, they might be examined to demonstrate the trend in the cluster concentration—we see that each higher-order cluster is ∼10−40× less concentrated—and they may be of interest in subsequent work comparing to molecular simulation data taken for TIP4P or another water model.

4.2.4. Association Equilibrium Constant. The association-based equations of state discussed in the Introduction postulate a form for the association equilibrium constants as a function of temperature and cluster size. It may be helpful to the formulation of those models to examine the equilibrium constants given by the virial coefficients. The equilibrium is usually given for consecutive association reactions, rather than the j-mer formation reaction of eq 10. Thus we consider equilibrium constants for this reaction sequence

$$X_{j-1} + X \rightarrow X_j \tag{26}$$

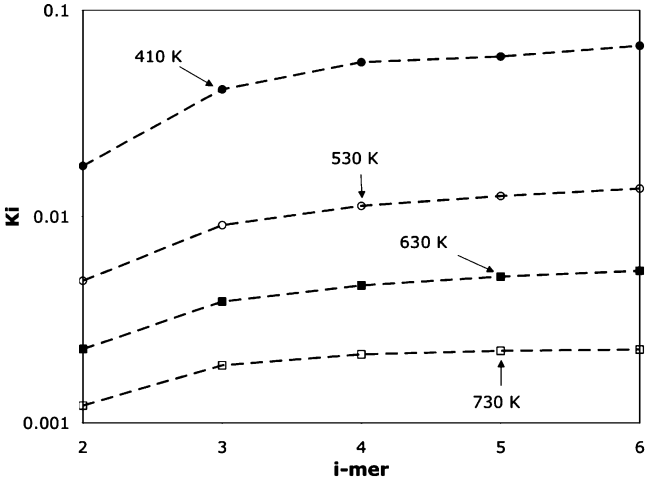

Figure 5. Chemical association equilibrium constants for consecutive association reactions: ● 410 K, ○ 530 K, ■ 630 K, □ 730 K.

Equilibrium constants for this reaction are presented in Figure 5, for *j* from 2 to 6 for TIP4P water. The association equilibrium constant increases with decreasing temperature, as would be expected. The most notable feature of the curves is the apparent leveling off, approaching a constant value. Some association-based equations of state employ an equilibrium constant that is independent of *j*, and this result gives some credence to that choice, at least up to the hexamer. It is certainly possible that the equilibrium constants decrease for larger oligomers.

5.0. Conclusions

Knowledge of the virial coefficients can provide information about the nature of physically bound clusters in a vapor phase. Indeed, the virial-based approach provides what is perhaps the least arbitrary definition of what constitutes a physically bound cluster. Cluster statistics given by this treatment compare very well with detailed observations of clustering performed in Monte Carlo simulations. The virial-cluster approach invokes no explicit configuration-based definition for a cluster, and yet it is in reasonable agreement with simulations that do apply explicit criteria. This comparison lends support to the use of the virial coefficients as a means to characterize clustering behavior in a vapor.

In two of the state conditions studied here, data were sufficient to separately infer concentration of physically and chemically bound clusters. In those cases, the data indicate that the concentration of physically bound clusters is ~30% of the concentration of those that are chemically bound. Certainly this outcome must be state-dependent, so we would not wish to generalize this result.

A significant limitation in the present study is the use of relatively simple water models that were developed with an emphasis on characterizing the behavior of liquid water. It would be of interest in further work to consider better models for water and attempt to develop results that can be considered true to the behavior of real water. Such models should give some account of multibody contributions to the potential and, perhaps, quantum effects. Polarizable models seem to be an effective treatment for multibody interactions in water. Progress in this direction will require extension of the Mayer sampling calculation method to handle these more sophisticated models and the related phenomena.

Note. After submission of our manuscript, we learned of the work of Chen et al.[39] who investigated atmospheric water clustering using histogram-reweighting Monte Carlo (HRMC) methods. Chen et al. calculated water-cluster compositions for various empirical water force fields (including TIP4P) for temperatures and pressures corresponding to elevations ranging from sea level to 15 km. Since we also used the TIP4P water model in our atmospheric-cluster calculations, it is appropriate to compare the virial-clustering approach to the HRMC simulations of Chen et al. We have done this and found that the results from the two methods are in good agreement. Using the virial-clustering approach, we were able to reproduce Figure 8 of Chen et al. for dimer and trimer compositions. For tetramer and pentamer compositions, the virial-clustering approach agrees with the HRMC up to ~11 km. Beyond that elevation, the tetramer and pentamer compositions predicted by the virial-clustering approach are lower than those predicted by HRMC. At all elevations, the virial-clustering approach yields larger hexamer mole fractions than the HRMC approach. There are two main reasons for these differences in cluster compositions. One reason is the difference in cluster definitions between the two approaches. Chen et al. relies on a distance-based criterion for their cluster statistics.[39] As mentioned previously, the virial-clustering approach employs no such criterion and, therefore, captures a larger number of physical clusters, including those which are associated over longer distances. This would explain the smaller composition of hexamers predicted by the HRMC approach relative to that predicted by the virial-clustering approach. In addition, some of the differences between the predicted tetramer, pentamer, and hexamer compositions from the virial-clustering method may result from the increasing error in the values of B_4, B_5, and B_6 at low temperatures ($T \leq 298$ K; see Table 1).

Acknowledgment

We thank Dr. Jayant K. Singh for helpful comments and discussions regarding the use and interpretation of the higher-order virial coefficients for water. Funding for this research was provided by grant CTS-0414439 from the U.S. National Science Foundation.

Literature Cited

(1) Liu, K.; Cruzan, J. D.; Saykally, R. J. Water Clusters. *Science* **1996**, *271*, 929.

(2) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.* **1982**, *76*, 637.

(3) Oxtoby, D. W. Homogeneous nucleation: Theory and experiment. *J. Phys.: Condens. Matter* **1992**, *4*, 7627. tenWolde, P. R.; Frenkel, D. Enhancement of protein crystal nucleation by critical density fluctuations. *Science* **1997**, *277*, 1975.

(4) Mountain, R. D. Voids and clusters in expanded water. *J. Chem. Phys.* **1999**, *110*, 2109.

(5) Curtiss, L. A.; Blander, M. Thermodynamic Properties of Gas-Phase Hydrogen-Bonded Complexes. *Chem. Rev.* **1988**, *88*, 827.

(6) Tassaing, T.; Danten, Y.; Besnard, M. Supercritical water: Local order and molecular dynamics. *Pure Appl. Chem*. **2004**, *76*, 133.

(7) Keutsch, F. N.; Cruzan, J. D.; Saykally, R. J. The water trimer. *Chem. Rev.* **2003**, *103*, 2533.

(8) Brown, M. G.; Viant, M. R.; McLaughlin, R. P.; Keoshian, C. J.; Michael, E.; Cruzan, J. D.; Saykally, R. J.; van der Avoird, A. Quantitative characterization of the water trimer torsional manifold by tetrahertz laser spectroscopy and theoretical analysis. II. $(H_2O)(3)$. *J. Chem. Phys.* **1999**, *111*, 7789.

(9) Buck, U.; Ettischer, I.; Melzer, M.; Buch, V.; Sadlej, J. Structure and spectra of three-dimensional $(H_2O)(n)$ clusters, $n = 8, 9, 10$. *Phys. Rev. Lett.* **1998**, *80*, 2578.

(10) Huisken, F.; Kaloudis, M.; Kulcke, A. Infrared spectroscopy of small size-selected water clusters. *J. Chem. Phys*. **1996**, *104*, 17.

(11) Pugliano, N.; Saykally, R. J. Measurement of the Nu-8 Intermolecular Vibration of $(D_2O)_2$ by Tunable Far Infrared-Laser Spectroscopy. *J. Chem. Phys.* **1992**, *96*, 1832.

(12) Pugliano, N.; Saykally, R. J. Measurement of Quantum Tunneling between Chiral Isomers of the Cyclic Water Trimer. *Science* **1992**, *257*, 1937.

(13) Keutsch, F. N.; Saykally, R. J. Water clusters: Untangling the mysteries of the liquid, one molecule at a time. *Proc. Nat. Acad. Sci.* **2001**, *98*, 10533.

(14) Mhin, B. J.; Lee, S. J.; Kim, K. S. Water-cluster distribution with respect to pressure and temperature in the gas phase. *Phys. Rev. A.* **1993**, *48*, 3764.

(15) Montero, L. A.; Molina, J.; Fabian, J. Multiple minima hypersurfaces of water clusters for calculations of association energy. *Int. J. Quantum Chem.* **2000**, *79*, 8.

(16) Dunn, M. E.; Pokon, E. K.; Shields, G. C. Thermodynamics of Forming Water Clusters at Various Temperatures and Pressures by Gaussian-2, Gaussian-3, Complete Basis Set-QB3, and Complete Basis Set-ANPO Model Chemistries; Implications for Atmospheric Chemistry. *J. Am. Chem. Soc.* **2004**, *126*, 2647.

(17) Boero, M.; Terakura, K.; Ikeshoji, T.; Liew, C. C.; Parrinello, M. Water at supercritical conditions: A first principles study. *J. Chem. Phys.* **2001**, *115*, 2219.

(18) Kalinichev, A. G.; Churakov, S. V. Thermodynamics and structure of molecular clusters in supercritical water. *Fluid Phase Equilib.* **2001**, 183−184, 271.

(19) Schenter, G. K.; Kathmann, S. M.; Garrett, B. C. Equilibrium Constant for Water Dimerization: Analysis of the Partition Function for a Weakly Bound System. *J. Phys. Chem. A* **2002**, *106*, 1557.

(20) Mason, E. A.; Spurling, T. H. *The Virial Equation of State, The International Encyclopedia of Physical Chemistry and Chemical Physics, Topic 10: The Fluid State*; Pergamon Press Ltd.: Oxford, U.K., 1969; Vol. 2.

(21) Woolley, H. W. The representation of gas properties in terms of molecular clusters. *J. Chem. Phys.* **1953**, *21*, 236.

(22) Kell, G. S.; McLaurin, G. E. Virial Coefficients of Methanol from 150 C to 300 C and Polymerization in the Vapor. *J. Chem. Phys.* **1969**, *51*, 4345.

(23) Stillinger, F. H., Rigorous basis of the Frenkel-Band theory of association equilibrium. *J. Chem. Phys.* **1963**, *38*, 1486. Lee, J. K.; Barker, J. A.; Abraham, F. F., Theory and Monte Carlo simulation of physical clusters in the imperfect vapor. *J. Chem. Phys.* **1973**, *58* (8), 3166.

(24) Prausnitz, J. M.; Lichtenthaler, R. N.; de Azevedo, E. G. *Molecular Thermodynamics of Fluid Phase Equilibria*; PTR Prentice-Hall: Englewood Cliffs, NJ, 1986.

(25) Heidemann, R. A.; Prausnitz, J. M. A van der Waals-type equation of state for fluids with associating molecules. *Proc. Natl. Acad. Sci.* **1976**, *73*, 1773.

(26) Anderko, A. Association and semiempirical equations of state. *J. Chem. Soc., Faraday Trans.* **1990**, *86*, 2823.

(27) Ikonomou, G. D.; Donohue, M. D. Thermodynamics of hydrogen-bonding molecules: The associated perturbed anisotropic chain theory. *AIChE J.* **1986**, *32*, 1716.

(28) Chapman, W. G.; Gubbins, K. E.; Jackson, G.; Radosz, M. SAFT: Equation-of-state solution model for associating fluids. *Fluid Phase Equilib.* **1989**, *52*, 31. Chapman, W. G.; Gubbins, K. E.; Jackson, G.; Radosz, M. New reference equation of state for associating liquids. *Ind. Eng. Chem. Res.* **1990**, *29*, 1709.

(29) Singh, J. K.; Kofke, D. A. Mayer Sampling: Calculation of Cluster Integrals using Free-Energy Perturbation Methods. *Phys. Rev. Lett.* **2004**, *92*, Art. No. 220601. All values of B_5 for the SPC/E model given in this reference have a sign error.

(30) Singh, J. K.; Benjamin, K. M.; Schultz, A. J.; Kofke, D. A. Higher order virial coefficients of water models. *J. Phys. Chem. B* to be submitted.

(31) Kofke, D. A. Free energy methods in molecular simulation. *Fluid Phase Equilib.* **2005**, *228−229C*, 41.

(32) Kofke, D. A.; Frenkel, D. Perspective: Free energies and phase equilibria. *Handbook of Molecular Modeling*; Yip, S., Ed.; Springer: Dordrecht, The Netherlands 2005.

(33) Frenkel, D.; Smit B. *Understanding Molecular Simulation: From Algorithms to Applications*; Academic Press: San Diego, CA, 1996.

(34) Sandler, S. I. *Chemical and Engineering Thermodynamics*; John Wiley & Sons: New York, 1999.

(35) Boulougouris, G. C.; Economou, I. G.; Theodorou, D. N. Engineering a Molecular Model for Water Phase Equilibrium over a Wide Temperature Range. *J. Phys. Chem. B* **1998**, *102*, 1029.

(36) Akiya, N.; Savage, P. E. Roles of water for chemical reactions in high-temperature water. *Chem. Rev.* **2002**, *102*, 2725.

(37) Johansson, E.; Bolton, K.; Ahlstrom, P. Simulations of vapor water clusters at vapor−liquid equilibrium. *J. Chem. Phys.* **2005**, *123*, 024504.

(38) Harvey, A. H.; Lemmon, E. W. Correlation for the second virial coefficient of water. *J. Phys. Chem. Ref. Data*. **2004**, *33*, 369.

(39) Chen, B.; Siepmann, J. I.; Klein, M. L. Simulating Vapor−Liquid Nucleation of Water: A Combined Histogram-Reweighting and Aggregation-Volume-Bias Monte Carlo Investigation for Fixed-Charge and Polarizable Models. *J. Phys. Chem. A* **2005**, *109*, 1137.

Received for review October 18, 2005
Accepted November 22, 2005

IE051160S