Efficient and Reliable Calculation of Rice–Ramsperger–Kassel–Marcus Unimolecular Reaction Rate Constants for Biopolymers: Modification of Beyer–Swinehart Algorithm for Degenerate Vibrations

Jeong Hee Moon,[a] Meiling Sun,[b] and Myung Soo Kim[b]

[a] Systemic Proteomics Research Center, KRIBB, Daejeon, Korea
[b] Department of Chemistry, Seoul National University, Seoul, Korea

The Beyer–Swinehart (BS) algorithm, which calculates vibrational state density and sum, was modified for simultaneous treatment of degenerate vibrations. The modified algorithm was used in the grouped-frequency mode of the Rice–Ramsperger–Kassel–Marcus (RRKM) unimolecular reaction rate constant calculation for proteins with relative molecular mass as large as 100,000. Compared to the original BS method, reduction in computation time by a factor of around 3000 was achieved. Even though large systematic errors arising from frequency grouping were observed for state densities and sums, they more or less canceled each other, thus enabling reliable rate constant calculation. The present method is thought to be adequate for efficient and reliable RRKM calculations for any macromolecule in the gas phase such as the molecular ions of proteins, nucleic acids, and carbohydrates generated inside a mass spectrometer. The algorithm can also be used to calculate the internal energy distribution of a macromolecule at thermal equilibrium.   (J Am Soc Mass Spectrom 2007, 18, 1063–1069) © 2007 American Society for Mass Spectrometry

One of the main difficulties in applying the fundamental principles and methods developed in physical sciences to biological systems lies in the very large number of degrees of freedom involved. Efforts are being made in various fields to devise methods to handle this problem. For example, development of more efficient algorithms and computational methods for large molecules is one of the frontiers in quantum chemistry [1–3]. Theoretical treatment of biological reactions is also very challenging, tremendously large number of degrees of freedom being one of the main difficulties.

Unimolecular reactions occurring statistically under the microcanonical condition is described by the Rice–Ramsperger–Kassel–Marcus (RRKM) theory [4–9]. When the molecular rotation is ignored, the RRKM rate constant is expressed as follows:

$$k = \sigma \frac{N^{\ddagger}(E - E_0)}{h\rho(E)} \qquad (1)$$

Here $\rho(E)$ is the vibrational state density of the reactant at the internal energy E, E_0 is the critical energy of the reaction, $N^{\ddagger}(E - E_0)$ is the vibrational state sum from 0 to $E - E_0$ at the transition state (TS), h is the Planck constant, and σ is the reaction path degeneracy, which is usually one for reactions involving large molecules.

Various factors cause uncertainty in a calculated rate constant such as the presence of vibrational anharmonicity and internal rotation [8]. Even with the simplification of treating all the internal degrees of freedom as harmonic vibrations, tremendous difficulties remain. One of these is to obtain the complete sets of vibrational frequencies at the reactant and TS geometries. Unlike for small molecules, attaining a complete frequency set for a molecule with the relative molecular mass (RMM) as large as 10,000 or larger by experiment or by quantum chemical calculation is virtually impossible at the moment. Fortunately, it is known [10–13] from studies on small molecule reactions that accuracy of each vibrational frequency is not critical in RRKM calculation. We previously reported [14] a method to estimate a complete set of vibrational frequencies for proteins with any amino acid sequence using fictitious sets for 20 amino acid residues derived from density functional theory calculations. A similar method may be useful for other biopolymers such as nucleic acids and carbohydrates. The frequencies at TS can be acquired only by quantum chemical calculation. Computations involved are not trivial even for peptides consisting of three simple

Address reprint request to Prof. Myung Soo Kim, Department of Chemistry, Seoul National University, Seoul 151-742, Korea. E-mail: myungsoo@snu.ac.kr

amino acids, as demonstrated by Paizs and Suhai [15]. One of the alternative approaches is to postulate a value for the fictitious entropy of activation at 1000 K ($\Delta S^{\ddagger}$) for the reaction and to derive the TS frequency set by adjusting some frequencies of the reactant to arrive at the postulated value of $\Delta S^{\ddagger}$ [4, 16]. The advantages of the method are that it is straightforward to implement and that the rate constant thus calculated is rather insensitive to details of the method such as the kind and number of modes chosen for adjustment. Technically speaking, E_0 and $\Delta S^{\ddagger}$ are the parameters affecting the magnitude of a calculated rate constant. However, the fact that they are difficult to assign is another cause for uncertainty in RRKM calculation, especially for large-molecule reactions.

Various methods have been developed over the years to calculate the vibrational state density and sum. Direct counting of states [17–19] was attempted from the early days of study together with the computationally less demanding version of using grouped frequencies [20] and approximate methods such as the Whitten–Rabinovitch (WR) semiclassical method [21] and the steepest-descent method [22]. It was observed that all the approximate methods provided erroneous results at low internal energy [4, 20, 23]. These methods became virtually obsolete after the invention of an efficient direct counting algorithm by Beyer and Swinehart (BS) [24].

After the development of a method to prepare the complete vibrational frequency set of a protein, we attempted to study mass-dependent change in protein dissociation kinetics [15, 16, 23, 25, 26]. It was found that the computation time needed for RRKM calculation with the BS algorithm increased exponentially with molecular mass, becoming days at RMM of a few tens of thousands. Subsequently, we investigated the performance of the WR method, which can instantly calculate a rate-energy relation. When applied to protein dissociation, the rate constants at low internal energy were found to be larger than the corresponding BS results by orders of magnitude as reported. The fact that some parameters used in calculation proposed by Whitten and Rabinovitch [21] were not adequate for proteins at low internal energy was found to be responsible for the overestimation. Improvement [27] of these parameters using the frequency set obtained in our previous work reduced the error to a factor < 2. Even though the same method of improvement can be applied to other biopolymers, the parameters must be evaluated separately for each class of compounds; that is, lack of universal applicability is the main drawback of this improved WR method.

Grouping vibrations and treating those in a group as degenerate vibrations can reduce the computational demand in state counting. This grouped-frequency method is not in use nowadays because the BS algorithm is fast enough even for fairly large molecules and because the BS algorithm in its current form cannot simultaneously handle degenerate vibrations. As a part of our effort to develop an efficient and reliable method of RRKM calculation that can be used for any biopolymer, a modified BS algorithm that can handle degenerate vibrations simultaneously has been derived in this work. Its utility and performance in the grouped-frequency mode of RRKM calculation have also been tested. The results are reported in this paper.

Algorithm and Computation

A brief description of the original BS algorithm [24] is as follows. Let us express vibrational frequencies and internal energy in cm^{-1} units. Each of these is converted to an integer by dividing with a grain size (δ, in cm^{-1}) and taking the nearest integer. Then, the frequency of the mode i is represented by R_i and the internal energy of $(M-1)\delta$ in cm^{-1} is represented by M. A grain size of 1–10 cm^{-1} is adequate for small molecules. An array T(M) is used that, eventually, becomes the state density calculated at the energy interval of δ. Sum of the first (M + 1) elements corresponds to the state sum up to the internal energy $M\delta$. T(M) is initialized to $(1, 0, 0, 0, 0, \cdots)$ and upgraded for each vibration by the following algorithm:

$$T(M) = T(M) + T(M - R_i) \tag{2}$$

For example, let us consider a system with three vibrational modes with $R_1 = 2$, $R_2 = 3$, and $R_3 = 3$. Use of eq 2 for mode 1 converts the T array into $T_1 = (1, 0, 1, 0, 1, 0, 1, 0, 1, 0, \cdots)$. Its successive use for the modes 2 and 3 converts the latter array first into $T_2 = (1, 0, 1, 1, 1, 1, 2, 1, 2, 2, \cdots)$ and then into $T_3 = (1, 0, 1, 2, 1, 2, 4, 2, 4, 6)$.

As noted earlier, the original BS algorithm treats all the vibrations separately regardless of degeneracy. It is obvious that reduction in computation time may occur if a method to treat degenerate vibrations all at the same time can be devised. In the grouped-frequency direct counting mode [17–19] of RRKM calculation used before the advent of the BS algorithm, this was achieved by taking advantage of the fact that the number of ways to distribute j vibrational quanta into g degenerate vibrations is given by $_gH_j = _{g+j-1}C_j$. To see how the same relation plays in the original BS algorithm, let us inspect the energy density, or the number of state at the internal energy 9 in T_3, or $T_3(10) = 6$, in the preceding example. There are two ways to partition the internal energy 9 into mode 1 and the degenerate modes 2 and 3; three quanta in mode 1 and one quantum in degenerate modes and all the internal energy into the degenerate modes as three quanta. The ways to distribute these quanta, or state degeneracies, in these two cases are $_2H_1 = _2C_1 = 2$ and $_2H_3 = _4C_3 = 4$, respectively, resulting in the total number of states of 6 in agreement with $T_3(10)$ as calculated by the BS algorithm. That is, the algorithm sums the state degeneracies for all single and multiple excitations of the degenerate modes that are allowed at the energy being considered. The algorithm devised in the present work to do the same for a

group(s) of degenerate modes, but all at the same time, is as follows.

Let us use two arrays initialized as $T = (1, 0, 0, 0, \cdots)$ and $AT = (1, 0, 0, 0, \cdots)$. Even though it does not matter at which step a particular group of degenerate modes is treated, let us suppose that they are treated at the beginning. Suppose that this group has frequency R and degeneracy g. At the internal energy specified by M, the maximum number of quanta that can be assigned to this group is as follows:

$$j_{max} = (M - 1)/R \qquad (3)$$

Then, the energy density is represented by the AT array upgraded by the following algorithm.

$$T(M) = AT(M) + \sum_{j=1}^{j_{max}} {}_gH_j \times AT(M - jR) \qquad (4)$$

$$AT(M) = T(M) \qquad (5)$$

This algorithm is equivalent to eq 2 for nondegenerate vibrations and can be used generally.

Validity of the modified algorithm was tested for various cases. A software was written to handle relatively complicated cases and state densities calculated with the original and modified algorithms were compared, which turned out to be identical when frequencies used were multiples of the grain size. As a simple example, let us consider a system with six vibrational modes with $R_1 = 2$ and R_2–$R_6 = 3$. The T_3 array obtained by treating R_1, R_2, and R_3 with the original BS is the same as in the previous example, that is, $T_3 = (1, 0, 1, 2, 1, 2, 4, 2, 4, 6, \cdots)$. Successive treatments of R_4–R_6 result in $T_4 = (1, 0, 1, 3, 1, 3, 7, 3, 7, 13, \cdots)$, $T_5 = (1, 0, 1, 4, 1, 4, 11, 4, 11, 24, \cdots)$, and $T_6 = (1, 0, 1, 5, 1, 5, 16, 5, 16, 40, \cdots)$. T_1 obtained by the modified BS is the same as the one by the original BS, that is, $AT_1 = T_1 = (1, 0, 1, 0, 1, 0, 1, 0, 1, 0, \cdots)$. Now, let us treat R_2–R_6 simultaneously using eq 2. For $M \leq 3$, $j_{max} = 0$ and thus $T_2(M) = AT_1(M) = (1, 0, 1)$ in agreement with T_6 derived earlier. Some other elements of $T_2 (=AT_2)$ are as follows. For $M = 4$, $j_{max} = 1$ and thus $T_2(4) = AT_1(4) + {}_5H_1 \times AT_1(1) = 0 + 5 \times 1 = 5$. For $M = 7$, $j_{max} = 2$ and thus $T_2(7) = AT_1(7) + {}_5H_1 \times AT_1(4) + {}_5H_2 \times AT_1(1) = 1 + 0 + 15 = 16$. Finally, for $M = 10$, $j_{max} = 3$ and thus $T_2(10) = AT_1(10) + {}_5H_1 \times AT_1(7) + {}_5H_2 \times AT_1(4) + {}_5H_3 \times AT_1(1) = 0 + 5 + 0 + 35 = 40$. All these are the same as the corresponding elements in T_6 derived by the original BS.

In the grouped-frequency direct counting method used previously [17–19], a complete frequency set, either of the reactant or of TS, was split into several subsets, each consisting of frequencies with similar magnitudes. The representative frequency for each subset was calculated as the geometrical mean of the constituents. These representative frequencies were used to calculate the number of frequency combinations at a given energy by direct count, which was then weighted with the $_gH_j$ factors to obtain the state density. Significant computational economy, without much loss of accuracy, was reported. The same approach, combined with the modified BS algorithm developed in this work, can be especially useful for very large molecules. As the number of elements in a subset increases, however, the number of terms in eq 4 also increases, partially offsetting the economy gained in the previous step. Even though we devised and tested various methods of dividing a frequency set into subsets, this step did not turn out to be critical. The method implemented in the software used in this work is as follows. First, all the frequencies in a set are arranged in increasing order. The first frequency R_1 is taken as an element of the first subset. The arithmetic mean of the elements in this subset $\bar{R}$, which is R_1 at the beginning, is calculated. Then R_2, the second frequency in the set, is taken and $C \times R_2^{0.5}$ is calculated. If $R_2 - R_2 - \bar{R} < C \times R_2^{0.5}$, R_2 is assigned to the same subset as R_1; otherwise, it is assigned to the second subset. The process is continued for all the elements in the set and the number of subsets is counted. Then, the parameter C is adjusted iteratively until the predetermined number of subsets is obtained. Use of the modified BS algorithm for G frequency subsets will be called BS-G.

Actual RRKM calculations must be carried out to test the utility of the present algorithm. One of the prerequisites for such calculations is the complete vibrational frequency set for the reactant, which is not readily available for general macromolecules. Thus, even though the present algorithm has been developed to treat any macromolecule, in this work testing was carried out only for proteins for which a systematic method to construct such a set had been previously established [14]. Details of the software other than those altered for the present purpose are the same as in the previous report [14]. Briefly, various inputs needed for each calculation are fed interactively. These include the amino acid sequence of a protein, the number of protons attached, the critical energy, the entropy of activation, the grain size, and the number of frequency groups to be used. A vibration with 1149 cm^{-1} frequency in the reactant geometry is taken as the reaction coordinate by default even though an option is available to select another. Once the amino acid sequence and the number of protons are specified, the software constructs the complete frequency set for the reactant by using the residue frequencies and the frequencies related to proton motion stored in the software. The reaction coordinate is deleted from this set and frequencies of six low-frequency modes are adjusted using the entropy of activation specified to construct the set at TS. The number of frequencies to be adjusted in this step and their designation are other options in the software. Software was written with the C program language under the Linux environment. The long double precision that can handle numbers as large as 10^{5000} was used in our previous BS software. For molecules with RMM > 20,000, the state density at high internal energy

often exceeded 10^{5000}. To handle such cases, the software has been revised in this work such that a number was represented by two numbers, a real number for the significant coefficient and an integer for the exponent, wherever needed. In principle, this dual-number representation can handle arithmetic calculations for numbers as large as $10^{2,000,000,000}$. Adding two numbers in the original BS algorithm of eq 2 becomes meaningless when one number is much larger than the other. Provision has been made in the new software to avoid such unnecessary calculations. BS calculations using the software thus revised will be denoted BSD. Dual-number representation has also been used in the BS-G software. An AMD Athlon64 3000+ processor (Advanced Micro Devices, Inc., Sunnyvale, CA, USA) was used for the calculations.

Results and Discussion

Efficiency and reliability of the grouped-frequency mode of RRKM calculation using the modified BS algorithm was tested. Because the purpose of the test was to evaluate the performance of the modified algorithm compared to the original one, several factors that could affect the reliability of actual rate-energy relation outcome were ignored. These included vibrational anharmonicity and the presence of internal rotations. Similarly, reliability of the complete frequency sets at the reactant and TS geometries used in the calculation was not an issue. In this work, results from the test will be presented using fictitious multimers (P-mers) of angiotensin I (DRVYIHPFHL, RMM of 1295.7), which will be denoted as AI-P, as examples. With p = 1–80, their RMMs lie in the range of 1300–102,000. We also tested with fictitious multimers of other peptides such as melittin, ubiquitin, and ACTH. The results were essentially the same. Even though we could not test for other types of biopolymers for the reason mentioned in the previous section, materials used for the test would not be an issue either for the present purpose. The influence of some factors had to be carefully investigated. These included the grain size, the number of frequency groups, the critical energy (E_0), and the entropy of activation ($\Delta S^{\ddagger}$). Evaluation of the present algorithm as a function of E_0 was needed because the internal energy needed to result in the same rate constant increased rapidly with E_0 and thus computation time increased exponentially.

Even though the grain size might similarly affect the calculations with the original and modified BS algorithms, its influence was tested with the original algorithm such that rate-energy relations to be used as benchmarks in subsequent tests could be established. Initially, we also hoped that use of a grain size > 1–10 cm^{-1} typically used for small molecules [4] might turn out to be adequate for large molecules and thus reduce computation time. Tests were made for AI-P (p = 1–8, 10, and 20) using the grain size of 1, 10, 25, 35, and 50 cm^{-1}; $E_0 = 0.5$ eV; and $\Delta S^{\ddagger} = 0$ eu (1 eu = 4.184 J K^{-1}
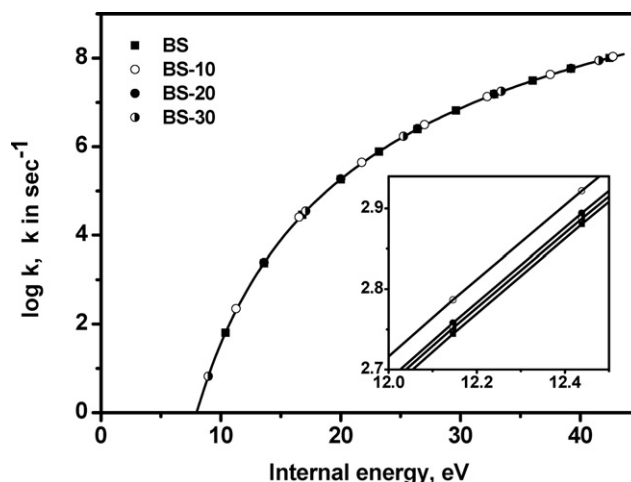

Figure 1. RRKM rate-energy relations for dissociation of 7-mer of angiotensin I, AI-7, calculated with E_0 of 0.5 eV and $\Delta S^{\ddagger}$ of 0 eu using the original BS (■), BS-10 (○), BS-20 (●), and BS-30 (◑) algorithms.

mol^{-1}). Compared to the rate-energy relations obtained with the grain size of 1 cm^{-1}, which were taken as references, the error in each datum decreased steadily with the increase of rate constant. Quoting the errors at the rate constant of 10^{-3} s^{-1}, these were around 6, 22, 75, and 140% for the grain sizes of 10, 25, 35, and 50 cm^{-1}, respectively, with only marginal dependency on molecular mass. Because one of our goals was to develop a grouped-frequency mode of RRKM calculation, which could reproduce the results from the use of the original BS algorithm, it was decided to adopt the grain size of 10 cm^{-1}. This is the most frequently used grain size for small molecules. Computational economy with the use of a larger grain size hoped for large molecules was not realized.

In the next test, we compared the rate-energy relations obtained by the grouped-frequency modified BS calculations, BS-G, with those by the original BS algorithm. Results from calculations done for dissociation of AI-P occurring with $\Delta S^{\ddagger} = 0$ eu and $E_0 = 0.5$ eV using the subgroup number (G) of 10, 20, and 30 will be presented first. To extend the test to as large a molecular mass as possible, we first compared the rate-energy relations obtained by the original and dual-number (BSD) BS algorithms for RMM < 20,000 and confirmed that BSD was working properly. That is, the results from the two algorithms were identical. Thus, we calculated a rate-energy relation either with BS or with BSD, whichever was more convenient, and compared the results with those calculated by BS-G (G = 10, 20, and 30). A typical case (AI-7, $E_0 = 0.5$ eV, and $\Delta S^{\ddagger} = 0$ eu) is shown in Figure 1. All the BS-G results reproduced the BS result rather well. Minor differences can be seen in the magnified rate-energy relations shown as an inset in the figure. Differences grew larger at a smaller rate constant. For the results at around $k = 10^{-3}$ s^{-1} (not shown in Figure 1), the rate constants calcu-

lated with BS-10, BS-20, and BS-30 deviated from the BS results by 20, 6, and −0.5%, respectively. Summarizing the results calculated for all the proteins studied in this work, typical errors with BS-10, BS-20, and BS-30 were around 20, 10, and 2%, respectively. Occasionally, larger errors were encountered with BS-10 and BS-20, for example, 50% for AI with BS-10 and 70% for AI-6 with BS-20.

We also performed calculations using other values of E_0 and $\Delta S^\ddagger$. The rate constant error in BS-30 was insensitive to E_0. On the other hand, it was found to increase with $\Delta S^\ddagger$. With $\Delta S^\ddagger = 10$ eu, which is a typical value for dissociation occurring by loose TS, the BS-30 rate constants were smaller than the corresponding BS data by around 50%. Dependency of the error on RMM was very weak. We also performed BS-60 calculations to see whether use of a larger number of groups would reduce the error. Even though the error was smaller than that in BS-30 at small RMM, they became nearly comparable for AI-30.

To investigate the nature of the preceding errors, we compared state densities and sums calculated by BS and BS-30 for AI-P (p = 1–30). It was found that BS-30 underestimated these data and the difference grew larger for larger RMM values, from a factor of 1.1 for the monomer to a factor of 30 for the 30-mer. With $\Delta S^\ddagger = 0$ eu, these systematic errors in state densities and sums were nearly the same and compensated each other, resulting in 10% error in the final rate constant. Error compensation became poorer with larger $\Delta S^\ddagger$, resulting in larger error in the rate constant. One can think of two origins for this error, one arising from erroneous counting and the other from frequency grouping. As mentioned earlier, the modified algorithm accurately reproduced state densities calculated by BS when frequencies were multiples of the grain size. Otherwise, state densities calculated by the original and modified algorithms were often different. To see the extent of the counting error, we calculated state densities of AI-P using the modified algorithm without frequency grouping. Because of the method used to prepare the frequency set, these systems have many degenerate vibrations. The counting error in state density thus estimated was only a few percent. Then, very large errors in state densities and sums calculated by BS-30 must originate from frequency grouping. An explanation for the increase of this error with $\Delta S^\ddagger$ is as follows. With small $\Delta S^\ddagger$, most of the frequencies of a protein at the reactant geometry and at TS are similar, which results in similar frequency groups and thus similar grouping errors in state densities and sums. As $\Delta S^\ddagger$ becomes larger, these groups become less similar, resulting in larger rate constant error.

Finally, we compared the computation time needed for AI-P (p = 1–10, 20, 30, 40, 50, 60, 70, and 80) with BS, BSD, BS-10, BS-20, and BS-30. The amount of internal energy needed for dissociation with a particular magnitude of rate constant increases almost in proportion to RMM [28]. Thus, for a fair comparison of computation
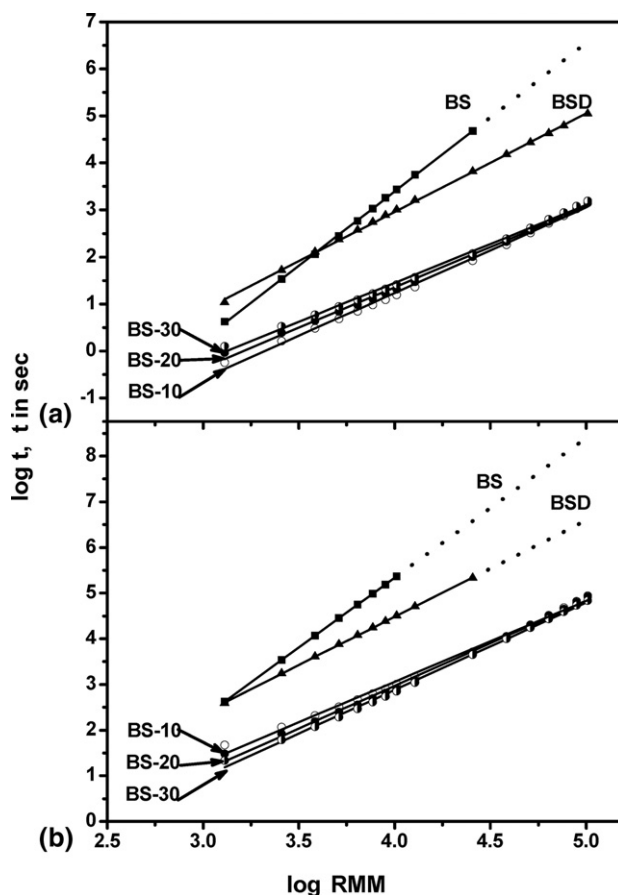

Figure 2. Log-log plots of the computation time versus RMM data for dissociation of multimers of angiotensin I, AI-P (p = 1–10, 20, 30, 40, 50, 60, 70, and 80), calculated with $\Delta S^\ddagger$ of 0 eu and E_0 of (**a**) 0.5 and (**b**) 2.0 eV using the original BS (■), BSD (▲), BS-10 (○), BS-20 (●), and BS-30 (◑) algorithms. For a fair comparison of computation times for molecules with different mass and dissociations occurring with different E_0, the E_{max}/E_Z ratio was set to (**a**) 0.15 and (**b**) 1.3 such that rate constants up to around 10^8 s^{-1} were calculated for all cases.

times for the dissociation of different multimers occurring with a specified E_0 value, adjustment of the maximum internal energy (E_{max}) for each multimer was needed such that the maximum value for the rate constant was nearly the same regardless of RMM. The zero-point energy (E_Z) of each multimer also increases with RMM. It turned out that keeping the E_{max}/E_Z ratio constant regardless of RMM resulted in similar values of the maximum rate constants calculated. The results to be presented below were obtained with the E_{max}/E_Z ratio of 0.15, 0.47, and 1.3 for calculations with E_0 of 0.5, 1.0, and 2.0 eV, respectively, which generated the maximum rate constant of around 10^8 s^{-1} for all cases. Log-log plots of the computation time versus RMM data for dissociation of AI-P calculated with $\Delta S^\ddagger$ of 0 eu and E_0 of 0.5 and 2.0 eV using BS, BSD, BS-10, BS-20, and BS-30 are shown in Figure 2. For dissociation with $E_0 = 0.5$ eV, the BS algorithm could handle only up to AI-20 (RMM = 25,572) because a number density $> 10^{5000}$ appeared at larger RMM, whereas calculation

was possible all the way up to AI-80 (RMM = 102,234) with BSD. Because the maximum internal energies used in calculations were larger with E_0 = 2.0 eV, BS could handle only up to AI-8 (RMM = 10,240). Even though there would be no mass limit with BSD, BSD calculation with E_0 = 2.0 eV was done only up to AI-20 because of tremendous computational demand. The computation time for AI-20 in this case was 2.1×10^5 s (2.4 days). Because BSD uses the dual-number representation for some numbers, one would have expected to observe longer computation time than in BS. This was indeed the case in the low-mass region of Figure 2a calculated with E_0 = 0.5 eV. As RMM increased, however, the computation time for BSD increased less rapidly than that for BS, and BSD became more efficient than BS at RMM larger than a few thousand. Situation was similar for the E_0 = 2.0 eV case shown in Figure 2b. Here, use of larger maximum internal energies than in the E_0 = 0.5 eV case made BSD more efficient than BS even for the monomer. Treating exponents as integers and eliminating unnecessary calculations must have been responsible for the high efficiency of BSD.

All the BS-G methods were more efficient than BS and BSD. It was another surprise to note that the computation times for BS-10, BS-20, and BS-30 were rather similar. Thus, BS-30—which shows noticeably smaller error in rate constants than BS-10 and BS-20 in all the cases investigated—is the method of choice. Computational economy arising from the use of a smaller number of subsets in BS-10 must have been offset by the need to compute more terms in eq 4. Good linear relations observed in the log-log plots for all the cases shown in the figure help to estimate the computation time even when actual computation could not be done. In the E_0 = 0.5 eV case for RMM of 100,000, the computation times with BS, BSD, and BS-30 would be 3.53×10^6, 1.12×10^5, and 1.27×10^3 s, respectively. BS-30 would be faster than BS by a factor of 2800. Corresponding data in the case of E_0 = 2.0 eV would be 2.47×10^8, 3.91×10^6, and 6.77×10^4 s, respectively. Here BS-30 would be faster than BS by a factor of 3600. Although 6.77×10^4 s, or 19 h, of computation needed with BS-30 at RMM = 100,000 is long, it is tolerable compared to 8 yr, which would have been needed with BS. Computation time with BS-30 in the above cases is roughly proportional to $(E_0)^3(RMM)^{1.7}$.

When the state density data over a wide internal energy range are available, the internal energy distribution at a specified temperature can be calculated by using the usual Boltzmann distribution, $P(E) \propto \rho(E)\exp(-E/kT)$. Even though large systematic errors are present in the state densities calculated by the grouped-frequency method, this does not prevent reliable calculation of an internal energy distribution because the error is nearly the same regardless of the internal energy. The internal energy distributions for AI and AI-80 at 500 K calculated by BS-30 are shown in Figure 3. The average internal energies of AI and AI-80 at this temperature are 6.29 and 505.9 eV with the full
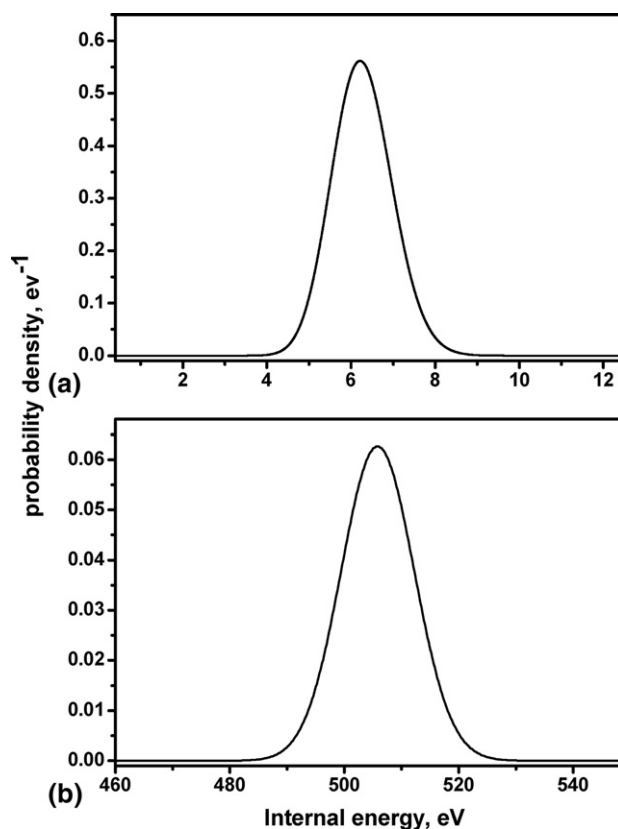

Figure 3. The internal energy distributions at 500 K for (**a**) angiotensin I, AI-1 and (**b**) its fictitious 80-mer, AI-80, calculated by BS-30.

widths at half maximum of 1.66 and 14.8 eV, respectively. The average internal energy at a specified temperature increases in proportion to mass, whereas the standard deviation increases only in proportion to the square root of mass, as expected. This leads to an interesting conclusion that a macromolecule at thermal equilibrium is nearly monoenergetic. The internal energy distribution and the rate-energy data that can be calculated by the present method can be used for the investigation of the stability and decomposition of large m/z molecular ions generated by techniques such as matrix-assisted laser desorption ionization (MALDI) [29–31] and electrospray ionization (ESI) [32] in mass spectrometry, which occur under a quasi-thermal condition.

Conclusion

A modified BS algorithm for simultaneous treatment of degenerate vibrations was derived and its performance in the grouped-frequency mode of RRKM calculation was evaluated using proteins as test samples. An enormous reduction in computation time was achieved at large molecular mass, a part of which was attributed to the use of smaller number of frequencies than in the original BS algorithm and the other part to the improvements made in the software for efficient handling of

very large numbers. When 30 frequency groups were used, the new method reproduced the rate constants calculated by the original BS method within a factor of 2. For proteins, almost instantaneous and reliable (also within a factor of 2) RRKM calculation can be done irrespective of molecular mass using the improved Whitten–Rabinovitch (IWR) method reported previously [27]. Even though the IWR method itself is probably universal, a prerequisite is that some parameters used in the calculations must be evaluated separately for each class of compounds as we did for proteins. Such an effort is not needed for the grouped-frequency BS method developed in this work. That is, the present method is a genuinely universal method applicable to any macromolecule—either biological or synthetic—as far as complete vibrational frequency sets at the reactant and TS geometries are available. Even with the tremendous computational economy achieved by frequency grouping, rate constant calculation for very large molecules (RMM > 100,000) at high internal energy may still require a lot of computation time. Within the present method, using a larger grain size, at the sacrifice of reliability, may be the only way to further reduce the computation time. Computation time is inversely proportional to the grain size.

Acknowledgments

References

1. Urbanc, B.; Borreguero, J. M.; Cruz, L.; Stanley, H. E. Ab Initio Discrete Molecular Dynamics Approach to Protein Folding and Aggregation. *Methods Enzymol.* **2006,** *412,* 314–338.
2. Hardin, C.; Pogorelov, T. V.; Luthey-Schulten, Z. Ab Initio Protein Structure Prediction. *Curr. Opin. Struc. Biol.* **2002,** *12,* 176–181.
3. Agostini, F. P.; Soares-Pinto, D. D. O.; Moret, M. A.; Osthoff, C.; Pascutti, P. G. Generalized Simulated Annealing Applied to Protein Folding Studies. *J. Comput. Chem.* **2006,** *27,* 1142–1155.
4. Holbrook, K. A.; Pilling, M. J.; Robertson, S. H. *Unimolecular Reactions,* 2nd ed.; Wiley: Chichester, UK; 1996.
5. Marcus, R. A.; Rice, O. K. The Kinetics of the Recombination of Methyl Radicals and Iodine Atoms. *J. Phys. Colloid. Chem.* **1951,** *55,* 894–908.
6. Rosenstock, H. M.; Wallenstein, M. B.; Wahrhaftig, A. L.; Eyring, H. Absolute Rate Theory for Isolated Systems and the Mass Spectra of Polyatomic Molecules. *Proc. Natl. Acad. Sci. U.S.A.* **1952,** *38,* 667–678.
7. Steinfeld, J. I.; Francisco, J. S.; Hase, W. L. Chemical Kinetics and Dynamics; Blackwell Scientific: Oxford, UK, 1990; pp 342-401
8. Gilbert, R. G.; Smith, S. C. Theory of Unimolecular and Recombination Reactions; Oxford Univ. Press: Oxford, UK, 1990; pp 136–211
9. Baer, T.; Hase, W. L. Unimolecular Reaction Dynamics: Theory and Experiments; Oxford Univ. Press: Oxford, UK, 1996; pp 171–211
10. Lifshitz, C. Recent Developments in Applications of RRKM-QET. *Adv. Mass Spectrom.* **1989,** *11A,* 713–729.
11. Rosenstock, H. M.; Stockbauer, R.; Parr, A. C. Kinetic Shift in Chlorobenzene Ion Fragmentation and the Heat of Formation of the Phenyl Ion. *J. Chem. Phys.* **1979,** *71,* 3708–3714.
12. Rosenstock, H. M.; Stockbauer, R.; Parr, A. C. Photoelectron-Photoion Coincidence Study of the Bromobenzene Ion. *J. Chem. Phys.* **1980,** *73,* 773–777.
13. Lifshitz, C. Time-Resolved Appearance Energies, Breakdown Graphs, and Mass Spectra: The Elusive "Kinetic Shift." *Mass Spectrom. Rev.* **1982,** *1,* 309–348.
14. Moon, J. H.; Oh. J. Y.; Kim, M. S. A Systematic and Efficient Method to Estimate the Vibrational Frequencies of Linear Peptide and Protein Ions with Any Amino Acid Sequence for the Calculation of Rice–Ramsperger–Kassel–Marcus Rate Constant. *J. Am. Soc. Mass Spectrom.* **2006,** *17,* 1749–1757.
15. Paizs, B.; Suhai, S. Combined Quantum Chemical and RRKM Modeling of the Main Fragmentation Pathways of Protonated GGG. II. Formation of b₂, y₁, and y₂ Ions. *Rapid Commun. Mass Spectrom.* **2002,** *16,* 375–389.
16. Hu, Y.; Hadas, B.; Davidovitz, M.; Balta, B.; Lifshitz, C. Does IVR Take Place Prior to Peptide Ion Dissociation? *J. Phys. Chem. A* **2003,** *107,* 6507–6514.
17. Current, J. H.; Rabinovitch, B. S. Decomposition of Chemically Activated Ethyl-d₃ Radicals. Primary Intramolecular Kinetic Isotope Effect in a Nonequilibrium System. *J. Chem. Phys.* **1963,** *38,* 783–795.
18. Schlag, E. W.; Sandsmark, R. A. Computation of Statistical Complexions as Applied to Unimolecular Reactions. *J. Chem. Phys.* **1962,** *37,* 168–171.
19. Wilde, K. A. Anharmonicity in Vibrational State Sums. *J. Chem. Phys.* **1964,** *41,* 448–451.
20. Robinson, P. J.; Holbrook, K. A. Unimolecular Reactions; Wiley–Interscience: New York, 1972; pp 125–126
21. Whitten, G. Z.; Rabinovitch, B. S. Accurate and Facile Approximation for Vibrational Energy-Level Sums. *J. Chem. Phys.* **1963,** *38,* 2466–2473.
22. Morse, P. M.; Feshbach, H. *Methods of Theoretical Physics,* Part I; McGraw-Hill: New York, 1953; pp 434-443.
23. Derrick, P. J.; Lloyd, P. M.; Christie, J. R. *Physical Chemistry of Ion Reactions in Advanced Mass Spectrometry*, vol. 13; Wiley: Chichester, UK, 1995; pp 25-52.
24. Beyer, T.; Swinehart, D. F. Algorithm 448 Number of Multiply Restricted Partitions [A1]. *Commun. ACM* **1973,** *16,* 379–379.
25. Griffin, L. L.; McAdoo, D. J. The Effect of Ion Size on Rate of Dissociation RRKM Calculations on Model Large Polypeptide Ions. *J. Am. Soc. Mass Spectrom.* **1993,** *4,* 11–15.
26. Laskin, J.; Bailey, T. H.; Futrell, J. H. Fragmentation Energetics for Angiotensin II and Its Analogs from Time- and Energy-Resolved Surface-Induced Dissociation Studies. *Int. J. Mass Spectrom.* **2004,** *234,* 89–99.
27. Sun, M.; Moon, J. H.; Kim, M. S. Improved Whitten–Rabinovitch Approximation for the Rice–Ramsperger–Kassel–Marcus Calculation of Unimolecular Reaction Rate Constants for Proteins. *J. Phys. Chem. B* **2007,** *111,* 2747–2751.
28. Oh, J. Y.; Moon, J. H.; Kim, M. S. Tandem Time-of-Flight Mass Spectrometer for Photodissociation of Biopolymer Ions Generated by Matrix-Assisted Laser Desorption Ionization (MALDI-TOF-PD-TOF) Using a Linear-Plus-Quadratic Potential Reflectron. *J. Am. Soc. Mass Spectrom.* **2004,** *15,* 1248–1259.
29. Hillenkamp, F.; Karas, M.; Beavis, R. C.; Chait, B. T. Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Biopolymers. *Anal. Chem.* **1991,** *63,* 1193A–1202A.
30. Zenobi, R.; Knochenmuss, R. Ion Formation in MALDI Mass Spectrometry. *Mass Spectrom. Rev.* **1998,** *17,* 337.
31. Karas, M.; Bachmann, D.; Bahr, U.; Hillenkamp, F. Matrix-Assisted Ultraviolet Laser Desorption of Non-Volatile Compounds. *Int. J. Mass Spectrom. Ion Process.* **1987,** *78,* 53–68.
32. Cole, R. B. Electrospray Ionization Mass Spectrometry; Wiley–Interscience: New York, 1997; pp 3-65