J. Phys. Chem. A **1997,** *101,* 3634−3636

COMMENTS

Empirical Relations for the Energy Dependence of Fractional Abundance for Nuclear Substitution Reactions

Tapon Roy

Boehringer Ingelheim Pharmaceuticals, Inc.,
Ridgefield, Connecticut 06877

Received: November 14, 1996

Hierl et al. measure cross sections and product distributions for reactions involving selectively solvated anions with certain neutral molecules.[1] Among the reactions considered is nucleophilic displacement. For nucleophilic displacement, the fractional abundance of the solvated product at collision energy E_T is defined as the cross section for $X^-(H_2O)$ formation divided by the sum of cross sections for X^- and $X^-(H_2O)$ formation at that collision energy. In Figure 8 of their paper, Hierl et al. display the relationship of the energy dependence of fractional abundance $f(E_T)$ for the nucleophilic displacement reaction of $OH^-(H_2O)$ with CH_3Cl.

Hierl et al. use a previously proposed empirical relation[2] to quantify the relationship between fractional abundance and collision energy (E_T):

$$f(E_T) = f^0 \exp(-E_T/E) \qquad (1)$$

The authors state that the linearity of the semilogarithmic plot in their Figure 8 demonstrates the accuracy of the empirical relation. From their plot, one can see that a good linear relationship between the variables is present for collision energies between 0.1 and about 1.1 eV and that this line is extrapolated to the two axes of the plot. This line is based on a fit to the means of replicate measurements at various collision energies. One could also use the variability at each collision energy as a weighting factor in the model or fit the equation to all of the replicate measurements rather than the means. For example, a weighted analysis would be quite effective for the data shown in the Hierl et al. Figure 9. Close inspection of the regression fits to the Figure 8 data reveals that the residuals (observed minus fitted values) exhibit a relatively low magnitude but noticeable periodic behavior, even over the limited range of data on which the line in the Hierl et al. Figure 8 is based (see Figure 1). This nonrandom pattern may indicate the need for modification of the model. Also, the straight line shown in the plot does not quantify the relationship between fractional abundance and collision energy over the entire range of hyperthermal energies considered in the Hierl et al. beam study.

We initially propose an empirical relation of the form

$$f(E_T) = f^0 \exp\left(-\sum_{j=1}^{m}(E_T^j/E_j)\right) \qquad (2)$$

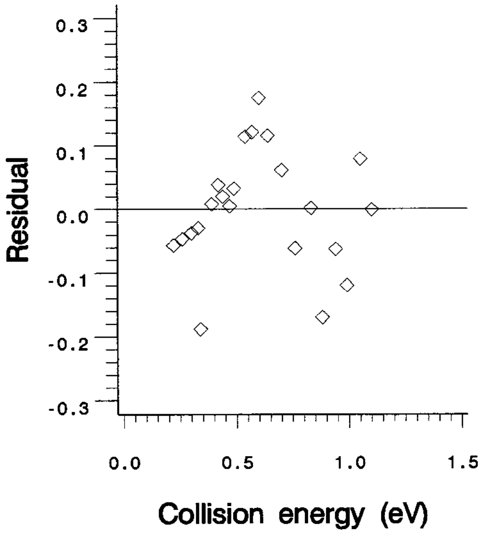

Figure 1. Residual plot from the linearized fit of eq 1 to the data obtained in the range 0.1−1.1 eV from Figure 8 of Hierl et al.,[1] fractional abundance of the solvated product $Cl^-(H_2O)$ formed in the nucleophilic displacement reaction of $OH^-(H_2O)$ with CH_3Cl; residuals vs collision energy.

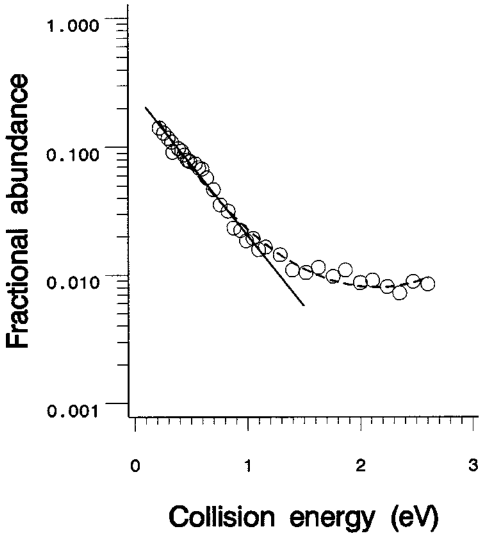

Figure 2. Fit of eq 2, $m = 2$, all data (dashed curve) compared to eq 1 fit, 0.1−1.1 eV data (solid line), extrapolated slightly for visual separation; both equations linearized in the parameters, from Figure 8 of Hierl et al.,[1] fractional abundance of the solvated product $Cl^-(H_2O)$ formed in the nucleophilic displacement reaction of $OH^-(H_2O)$ with CH_3Cl vs collision energy.

In Figure 2, the dashed curve is the fitted relation between fractional abundance and collision energy for all of the Hierl et al. data shown in their Figure 8 using eq 2 with a value of $m = 2$. The solid line shown is the linear fit to the 0.1−1.1 eV data given in their Figure 8 from eq 1, extrapolated slightly for better visual separation. The residual plot for the $m = 2$ case still indicates some nonrandom (periodic) behavior, similar to the pattern seen in Figure 1, but less pronounced. The overall fit is very good; however, notice the tendency toward upward curvature for the dashed curve as collision energy increases in Figure 2. The observed data appear to level off in this region. The extrapolation to thermal energies also exhibits this tendency toward curvature. Increasing m to 6 results in an excellent fit, with more accurate extrapolation to thermal energies, and nearly random-looking residuals, but proceeding to this extent may not be needed as discussed below. To improve the fit further, we propose the following empirical model, which provides direct

S1089-5639(96)03795-4 CCC: $14.00 © 1997 American Chemical Society

TABLE 1: Regression Diagnostics and Comparison of Models

model	residuals			goodness-of-fit[d]	
	normal?[a]	homoscedasticity indicated?[b]	autocorrelation?[c]	(linearized models (1) and (2))	multicollinearity[e]
(1), 0.1−1.1 eV	yes	moderate	moderate	$R^2 = 0.982$	
(2), all data, $m = 2$	yes	slight	moderate	adj $R^2 = 0.984$	no
(3), all data	yes	negligible	slight	excellent; see Figures 3 and 4	

[a] Residuals are normally distributed if the p value for the Shapiro−Wilk test is greater than 0.05. An assumption of regression analysis, for making inferences about parameter estimates, is that the error terms (estimated by the residuals) are normally distributed. [b] Homoscedasticity (equal variances of the residuals across the collision energies) is indicated if the residuals appear uniform across the range of collision energies. An assumption of regression analysis is that the residuals are homoscedastic. [c] Residuals are not autocorrelated if the Durbin−Watson test statistic is near the value 2. An assumption of regression analysis is that the residuals are not autocorrelated. [d] R^2 is the proportion of variation in the data explained by a *linear model with an intercept term*; "adj" means adjusted for the extra term(s) in the model. R^2 ranges from 0 to 1, 1 indicating a perfect fit. [e] Multicollinearity (a linear dependence among independent variables) would indicate that there is some redundancy among the independent variables.

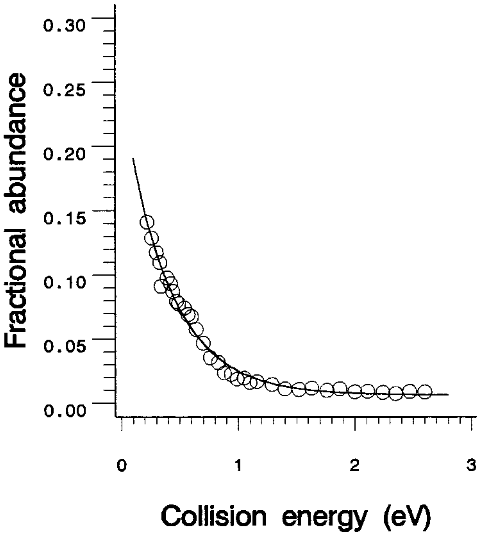

Figure 3. Fit of eq 3, all data, from Figure 8 of Hierl et al.,[1] fractional abundance of the solvated product $Cl^-(H_2O)$ formed in the nucleophilic displacement reaction of $OH^-(H_2O)$ with CH_3Cl vs collision energy.

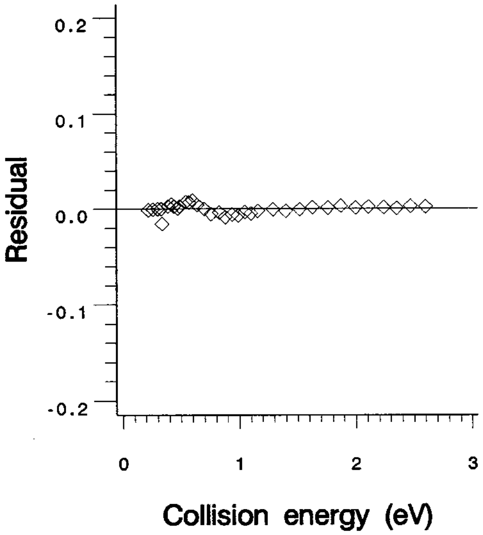

Figure 4. Residual plot from the fit of eq 3 to all of the data from Figure 8 of Hierl et al.,[1] fractional abundance of the solvated product $Cl^-(H_2O)$ formed in the nucleophilic displacement reaction of $OH^-(H_2O)$ with CH_3Cl; residuals vs collision energy.

estimation of f^0, extrapolates very accurately to thermal energies and properly models the asymptotic behavior of the data as collision energy increases:

$$f(E_T) = f^0 + \theta_1(1 - \exp(-\theta_2 E_T)) \tag{3}$$

Figures 3 and 4 display the fit of the data and the residual plot for model (3). As expected, this model provides an excellent fit. The residuals still display some nonrandom behavior, but they are of such small magnitude that further refinement of the model is not necessary.

In Table 1, regression diagnostics, including results of tests for residual normality, homoscedasticity, and autocorrelation, are presented, along with goodness-of-fit and multicollinearity results for the regression models considered in Figures 2 and 3. These are the usual measures needed for evaluating any regression analysis.[3,4] From Table 1, it is evident that model (3) provides the best fit, since all of the data are used, the regression assumptions are satisfied, and the fitted curve and residual plots exemplify the goodness of fit. Model (2) with $m = 2$ also provides a good fit, though it does have the curvature problem discussed previously.

Since relations 1−3 are empirical, the additional parameters introduced by using (2) or (3) would not obscure the physical meaning of the relationship. The parameter f^0 is still the abundance estimated at zero collision energy. The difference in using (2) with $m > 1$ is that the f^0 parameter is estimated from an effectively polynomial model, since in practice, equations like (1) and (2) are linearized (in the parameters) by taking the natural logarithm of both sides, as they are intrinsically linear.[3] Thus, the estimate of f^0 is obtained as the natural antilogarithm of the estimated intercept parameter in the linearized model, and the E_j parameter estimates are the negative reciprocals of the parameters associated with the E_T^j variables. The E_j would take the role of E in the model (1) analysis and would govern the curvature of the model. It is important to remember that the usual assumptions about the residuals must hold on the natural logarithm scale for the linearized versions of models (1) and (2), not the original scale. Polynomial models can be useful in determining empirical relationships, but certain issues need to be considered. Adding independent variables to a model will allow one to fit practically any smooth curve, but there can be the problem of overfitting to a particular set of data that would lack generality. Also, one must be cautious to use goodness-of-fit measures like the *adjusted* R^2, rather than the usual R^2 when judging the utility of higher order terms in the model. Further, it is critical to avoid excessive extrapolation of polynomial models to regions beyond the range of the data used. Model (3) is estimated directly using nonlinear regression; the estimate of f^0 along with its standard error and confidence interval can be obtained using any nonlinear regression program, and the estimates of θ_1 and θ_2 determine the curvature of the model. Models (2) and (3) can easily incorporate weighting factors based on replicate measurements, as discussed earlier for model (1), and can be fit without subjectively restricting the range of data as might be necessary when using model (1). The f^0 values estimated from model (2), with $m = 2$, and from model (3) are 0.32 and 0.24 (standard error = 0.01), respectively, using all of the data (obtained by interpolation) in comparison to the value of 0.25 obtained by Hierl et al. on the “linear portion” of the data from their Figure 8.

Thus, by generalizing model (1), as is done in model (2), or by using a new empirical relation, described by model (3), beam data from bimolecular nuclear substitution reactions can be used to extrapolate to thermal energy, allowing comparison to the fractional abundance obtained by selected-ion flow tubes (SIFT) and other techniques.

References and Notes

(1) Hierl, P. M.; Paulson, J. F.; Henchman, M. J. *J. Phys. Chem.* **1995**, *99*, 15655.

(2) Hierl, P. M.; Paulson, J. F. *J. Chem. Phys.* **1984**, *80*, 4890.

(3) Draper, N. R.; Smith, H. *Applied Regression Analysis*, 2nd ed.; John Wiley & Sons: New York, 1981.

(4) Bates, D. M.; Watts, D. G. *Nonlinear Regression Analysis and Its Applications*; John Wiley & Sons: New York, 1988.