Probing the Structural and Electronic Factors Affecting the Adsorption and Reactivity of Alkenes in Acidic Zeolites Using DFT Calculations and Multivariate Statistical Methods

Duangkamol Tantanak,*,† Jumras Limtrakul,*,‡ and Matthew Paul Gleeson‡

Chemistry Department, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand, and Laboratory for Computational and Applied Chemistry, Physical Chemistry Division, Kasetsart University, Bangkok 10900, Thailand

Received February 16, 2005

Quantum mechanical (QM) cluster calculations have been performed on a model of ZSM-5 at DFT and MP2 levels. We investigated how the adsorption energies and the energetics of alkoxide intermediate formation of six different alkene substrates, ethene, propene, 1-butene, cis/trans butene, and isobutene, vary in this zeolite model. An analysis of the DFT geometric, electronic, and energetic parameters of the zeolite−substrate complexes, transition states, and alkoxide intermediates is performed using principal components analysis (PCA) and partial least squares (PLS). These deliver an insight into the correlated changes that occur between molecular structure and energy along the reaction coordinate between the physisorbed and chemisorbed species within the zeolite. To the best of our knowledge, this is the first occasion multivariate techniques such as PCA or PLS have been employed to profile the changes in electronics, distances, and angles in QM calculations of catalytic systems such as zeolites. We find the calculated adsorption and the alkoxide intermediate energies correlate strongly with the absolute charge on the substrate and the length of the substrate double bond. The transition states' energies are not affected by the zeolite framework as modeled, which explains why they correlate strongly with the gas-phase substrate protonation energy. Our cluster results show that for ethene, propene, 1-butene, and isobutene, the relative energetics associated with the formation of the alkoxide intermediate in ZSM-5 follow the same trends as calculations where the effects of the framework are included.

1. INTRODUCTION

The catalytic abilities of zeolites have been seized upon by the chemical and petrochemical industries as an affordable means to perform important, selective reactions. The cracking of hydrocarbons in oil refining, for example, is one area where zeolites have had a significant impact,[1−4] and this has attracted a considerable amount of experimental and theoretical research.[5−48]

There are a large number of reactions catalyzed by zeolites, so the characteristics of their pores that give rise to the catalytic effect have been extensively characterized. Experimental techniques such as infrared/Raman spectroscopy (IR) and nuclear magnetic resonance (NMR) have been used to determine the physical properties of the zeolite pores (active sites), which have shown that their performance can be related to their acidity.[13−15] The medium- and long-range interactions of the zeolite framework will also play an important role in determining not only the acidity[16,17] of zeolites but also the selectivity of reactions because of diffusion effects and steric requirements imposed on reactions occurring within zeolite pores.[18−21] These effects would explain why the similar MFI-like zeolites Theta-1 and ZSM-5 have a different selectivity with respect to the production of isobutene from *n*-butenes.[21] This particular reaction occurs at acidic sites of zeolites where 1-butene, for example, can isomerize to either *cis*-, *trans*-, or isobutene.[22−24] However, the first step in the catalytic process, the diffusion of the zeolite−substrate complex to the site of reaction, can play an important role in determining the overall rate of the reaction.[12]

Quantum mechanical (QM) calculations are often used to model chemical reactions because, through a consideration of the relative energies, geometries, and interactions of the zeolite−substrate complexes, transition states, and products with the zeolite, we can often give insights into chemical processes not achievable by current experimental techniques. In the past few years, a large number of studies on the reactions of small molecules in zeolites have been published, some using relatively small[31−40] and large[41−44] gas-phase clusters, ONIOM models,[45,46] or periodic representations[47,48] of the zeolite lattice.

One distinct benefit of smaller clusters is the reduced computational requirement, which allows us to consider multiple chemisorbed and physisorbed substrates to acidic zeolites. In this way, we can attain a detailed insight into the physical characteristics of the interactions of the different substrates at the most crucial part of the zeolite. In this paper, we discuss the application of gas-phase cluster calculations to investigate the reactivity and adsorption of a range of different alkene substrates (ethene, propene, and four butene

* Authors to whom correspondence should be addressed. E-mail: fscijrl@ku.ac.th (J.L.); ktduangk@kmitl.ac.th (D.T.).
† King Mongkut's Institute of Technology Ladkrabang.
‡ Kasetsart University.

10.1021/ci0500583 CCC: $30.25
Published on Web 06/29/2005

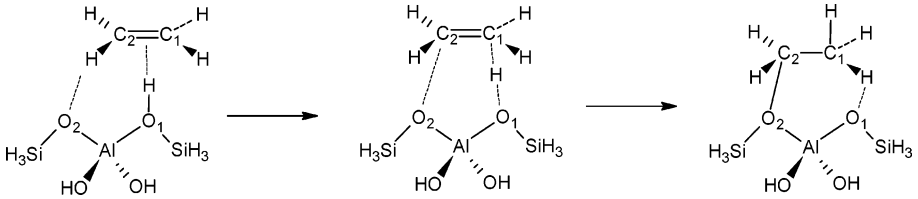

Figure 1. Schematic illustrating the formation of the alkoxide intermediate from the physisorbed zeolite−substrate complex via the transition state for ethene.

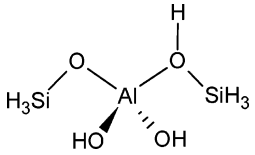

Figure 2. Zeolite cluster used in this study. Dangling H atoms are fixed in order to maintain them in their ZSM-5 geometry.

isomers) at the zeolite ZSM-5. The reaction to be investigated in this study is shown in Figure 1, which is the first step required for the interconversion of butene isomers. An alkene binds to the zeolite via two distinct interactions, (1) between the substrate C1 and the zeolite's acidic proton and (2) between the substrate C2 and the zeolite O2 atoms. The reaction proceeds along these two coordinates by the transfer of a proton to the C1 of the alkene with a subsequent nucleophilic attack by the zeolite O2 to form the alkoxide intermediate.

To extract as much information as possible from our QM calculations on six different alkenes, we employ principal components analysis (PCA) and partial least-squares regression[49−53] (PLS), widely used statistical techniques in quantitative structure−activity relationships (QSARs)[54−56] and comparative molecular-field analysis.[57−59] PCA is a statistical method for reducing the amount of data to be analyzed by exploiting the correlated nature of the variables within a dataset. Linear combinations of the correlated variables are chosen such that the majority of the variance of the original data can be described by a few orthogonal components. PLS also relies on linear combinations of the original variables but differs in that it employs a least-squares regression to relate the extracted components to a response, in this case, energies. Furthermore, in PLS, the linear combinations of variables are chosen to maximize the correlation between the extracted components and the response. These new components encode the same information as that contained in the many original variables but can be applied without the typical drawbacks associated with many correlated descriptors in multiple linear regression. A good review of how these methods can be applied to chemical problems is given by Wold et al.[60]

These are the preliminary results on ZSM-5 and are a prelude to further work that assesses the effect of medium- and long-range electronic and steric effects of different zeolite frameworks using the ONIOM methodology.[61]

2. COMPUTATIONAL PROCEDURES

ZSM-5 was modeled as a 3T cluster consisting of $Si_2AlO_4H_9$ (Figure 2) with the dangling bonds satisfied using hydrogens atoms. The six substrates to be investigated, ethene, propene, 1-butene, *cis*-butene, *trans*-butene, and isobutene, were added to the 3T cluster. With the exception of the acidic proton, all hydrogen atoms in the cluster were frozen to maintain the system in the same geometry as that in ZSM-5.[39] The resulting system was fully optimized at the B3LYP/6-31G(d,p) level, corresponding to the physisorbed, zeolite−substrate complex. The transition states and alkoxide intermediates were subsequently obtained at the same level of theory. All stationary points were confirmed as such using frequency calculations. Minima showed all positive vibrational frequencies with the transition states having just one negative frequency. Finally, single-point energies of all optimized stationary points were performed at the MP2/6-31++G(d,p) level.

The energy required to protonate the substrate in the gas phase was calculated for the purpose of comparison to the cluster calculations. This was taken as the difference between the optimized protonated and unprotonated forms of the substrate at the B3LYP/6-31G(d,p) level.

Adsorption energies of the substrates were calculated by taking the difference in energy between the zeolite−substrate complex and the isolated zeolite and the isolated substrate. A basis-set superposition error correction was not considered because others have shown the correction to be negligible for similar clusters and substrates.[43]

All QM calculations performed in this study used Gaussian 98 on Intel P4 computers running Linux at Kasetsart University.

2.1. Multivariate Data Analysis. PCA and PLS models were generated to ascertain how the structural and electronic parameters of the species on the reaction coordinate change from substrate to substrate and also how these changes affect the energy of the system (See Supporting Information Tables S1−S9). All models generated used 26 QM-derived geometric and electronic descriptors obtained from the optimized zeolite−substrate complex (ZSC), transition states (TS), and alkoxide intermediates (ALK). These descriptors consisted of the key bond distances and angles that change during the course of the reaction as well as the Mulliken charges found on the heavy atoms of the zeolite and substrate and the acidic proton being transferred between them (See Supporting Information Table S10 for further details). Energies for each substrate were expressed relative to the zeolite−substrate complex in kcal/mol because it is not possible to make a direct comparison between the energies of the different substrates because of the differing stoichiometries of the clusters. All descriptors were mean centered and scaled to unit variance because the numerical values of the descriptors vary significantly. This gives each variable the same opportunity of influencing the PCA/PLS models. Components were added to a model if they passed SIMCA's internal cross-validation procedures.[62]

The strengths of the correlation in the PCA and PLS models are reported using the coefficient of determination

(r^2) (eq 1), where n is the number of observations, σ_{obs} is the standard deviation of observed Y values, and the root-mean-square error (rmse) is given in eq 2. Here, Y_{obs} is the observed response and Y_{pred} the response predicted from the model.

$$r_o^2 = 1 - \frac{n}{n-1}\left(\frac{\text{rmse}^2}{\sigma_{\text{measured}}^2}\right) \quad (1)$$

$$\text{rmse} = \sqrt{\frac{\sum(Y_{obs} - Y_{pred})^2}{n}} \quad (2)$$

The greater the number of variables compared to observations in a model, the more likely it is that a strong chance correlation may be found. To determine the significance of the PLS models generated, we assessed the latter in cross validation and the former in cross validation and using *Y*-randomization trials. The first method involves a variation on the leave-one-out procedure (LOO), where $^1/_7$ of the observations are left out of the model building process and their energy is subsequently predicted by the reduced-observation model.[62] This is repeated until every observation has been left out and repredicted at least once. A cross-validated q^2 can then be calculated, where Y_{pred} is replaced by $Y_{pred(LOO)}$, the predicted response of the observations from the model where they were left out. The closer the r^2 value to the q^2 value, the greater the confidence in the model one can have.

Y-randomization trials were also employed to test the statistical significance of the PLS models. The *Y* values for each of the 22 observations were randomly mixed up 100 times and PLS models reconstructed. If a model is significant, none of the randomly generated examples should approach the performance of the true model.

2.2. Univariate Data Analysis. A univariate data analysis was performed by fitting a regression line between the QM descriptors and the energies in question. The strength of the correlation was determined using the square of the Pearson's correlation coefficient (R^2), which is determined to the line of best fit (eq 3), rather than the line $x = y$ for r^2. Here, x and y are the predicted and observed values of the response, n is the number of compounds, and $\bar{x}$ and $\bar{y}$ are the means of the predictions and the observations, respectively.

$$R^2 = \left(\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2\sqrt{\sum(y-\bar{y})^2}}}\right)^2 \quad (3)$$

Analysis of the variance (ANOVA) was used to assess the significance of our univariate analyses. This results in a probability (P) that the correlation between the two variables could have occurred by chance for the number of observations present. A P value less than 0.05 is a commonly used significance cutoff, indicating one can be 95% sure the correlation did not occur by chance.

All PCA and PLS models were built in SIMCA-P10.[63] The univariate data analysis was undertaken in Microsoft Excel.

3. RESULTS AND DISCUSSION

The calculated adsorption energies, the relative energetics associated with the formation of the alkoxide intermediate, and the gas-phase protonation energies of the six substrates at the DFT level are given in Table 1. The corresponding single-point energies at the MP2/6-31++G(d,p) level are given in Table 2. We also report the optimized geometric parameters in the zeolite−substrate complex, transition state, and alkoxide intermediate and the Mulliken charges found on the heavy atoms of the substrate and the zeolite in the zeolite−substrate complex, transition state, and alkoxide intermediate.

Table 1. Adsorption Energy, Activation Energy, Alkoxide Intermediate Energy, and the Energy Required for the Protonation of the Substrate in the Gas Phase at the B3LYP/6-31G(d,p) Level in kcal/mol[a]

substrate	adsorption energy	transition state	alkoxide intermediate	gas-phase protonation energy
ethene	−7.96	25.86	−10.18	−169.18
propene (1)	−9.03	22.78	−4.10	−193.71
propene (2)	−9.03	30.42	−5.22	−181.61
1-butene (1)	−9.02	22.98	−3.67	−195.72
1-butene (2)	−9.02	30.16	−4.48	−185.67
cis-butene	−9.26	26.42	−1.68	−193.29
trans-butene	−9.06	29.45	0.02	−192.11
isobutene	−9.46	16.29	3.58	−206.30

[a] Activation and alkoxide intermediate energies given relative to the zeolite substrate complex. Substrates with a 1 or 2 in parentheses indicate whether protonation has occurred on the primary or secondary carbon.

Table 2. Adsorption Energy, Activation Energy, and Alkoxide Intermediate Energy at the MP2/6-31++G(d,p)//B3LYP/6-31G(d,p) Level in kcal/mol[a]

substrate	adsorption energy	transition state	alkoxide intermediate
ethene	−9.71	30.29	−11.56
propene (1)	−12.45	26.92	−6.70
propene (2)	−12.45	34.36	−6.59
1-butene (1)	−12.80	27.30	−6.11
1-butene (2)	−12.80	34.13	−5.59
cis-butene	−12.98	29.75	−3.44
trans-butene	−13.14	33.00	−3.05
isobutene	−12.92	25.36	−2.03

[a] Activation and alkoxide intermediate energies given relative to the zeolite substrate complex. Substrates with a 1 or 2 in parentheses indicate whether protonation has occurred on the primary or secondary carbon.

The analysis of the results obtained at the B3LYP/6-31G(d,p) level and the corresponding single-point energies at the MP2/6-31++G(d,p) level show that the same energetic trends for the activation and the alkoxide-intermediate energies are predicted by both methods. The results at the latter level do not suggest the former is as much of an overestimate as others where higher correlated MP2 methods have been used.[43] This may be due to our inclusion of higher lying orbitals (++), which are known to better account for negatively charged species.[64] We, therefore, restrict our analysis to the DFT results only.

3.1. Principal Components Analysis. On the reaction coordinate studied here, we see significant changes in the geometries, electronic distribution, and relative energies from substrate to substrate. An analysis of the raw data shows that, moving along the reaction coordinate from the zeolite−substrate complex via the transition state to the alkoxide intermediate, the *r*O1−H, *r*Al−O2, and *r*C1−C2 distances;

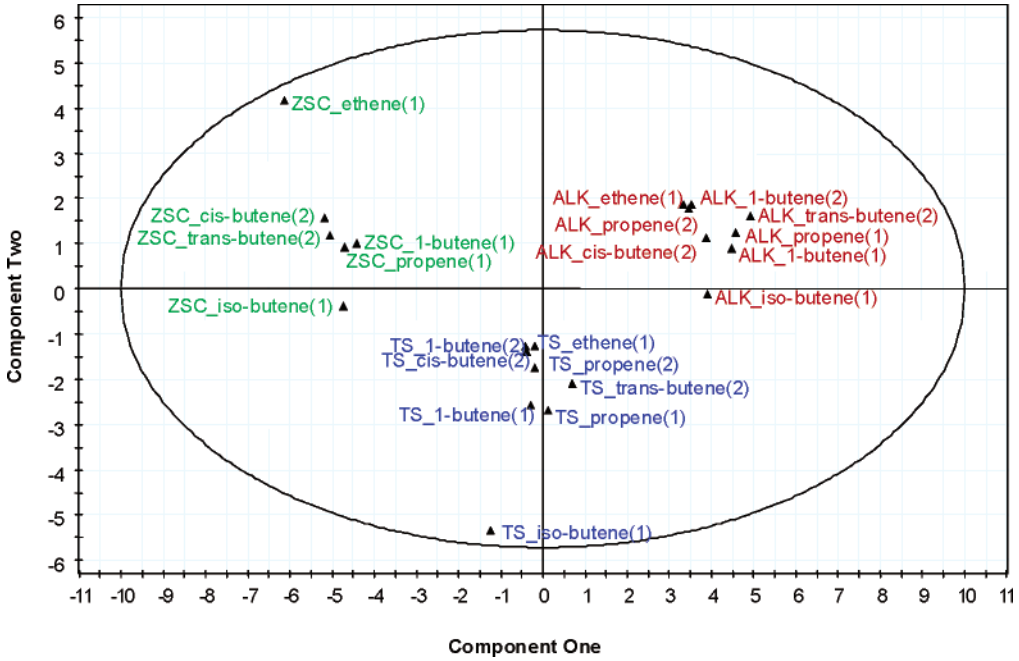

Figure 3. PCA scores plot: plot of component one against component two. The physisorbed zeolite−substrate complexes are denoted ZSC_*, the transition states are denoted TS_*, and the alkoxide intermediates are denoted ALK_*.

the *a*Al−O1−C2 angle; and the total Mulliken charge on the zeolite increase in magnitude. At the same time, the *r*Al−O1 distance, the *a*O2−C2−C1 and *a*Al−O1−C2 angles, and the overall charge on the substrate increase. With so many parameters changing in different ways, and to different extents, it becomes difficult to fully account for the multidimensional changes in structure and how they affect the energy. Fortunately, the correlated nature of the structural and electronic changes along the reaction coordinate makes this problem highly amenable to PCA or PLS regression.

PCA using the QM parameters derived from the zeolite−substrate complexes (ZSC), transition states (TS), and alkoxide intermediates (ALK) fits two components, describing approximately 70% of the total variance in the 26 descriptors. The model has a moderate cross-validated q^2; however, unlike QSARs, a large q^2 is not necessarily important because we are using the model for the purpose of interpretation rather than prediction. Projecting the 26 variables onto just two dimensions makes interpreting the relationship between the structural and electronic differences between the three stationary points of the different substrates much more facile. This information can be displayed on the scores and loadings plots.[62]

The zeolite−substrate complexes, transitions states, and alkoxide intermediates are clearly separated from each other in the scores plot (Figure 3) using the two PCA components, themselves derived from our DFT descriptors. Component one (*x* axis) describes the structural and electronic factors that allow us to separate the zeolite−substrate complexes from the alkoxide intermediates from each other, while the two minima are split from the transition states on the basis of component two (*y* axis). From an examination of the PCA loadings plot, we can determine the extent to which each of the 26 descriptors affects the two components that allow us to separate the three stationary points. The loading plot is a map of the structural and electronic features of our multidimensional system, in this case, projected into two easily interpretable dimensions as a result of the high degree of correlation between the descriptors. On the loading plot (Figure 4), descriptors found close to each other in both dimensions are highly correlated. Those that are similar on just one component are still correlated, but to a lesser extent, whereas those at opposite ends of the component(s) are inversely correlated. Furthermore, descriptors found at the extreme ends of the *x* or *y* axes have the most significant impact on the component that defines that axis. Those that are found close to the origin have a negligible effect. Using this information, and by a systematic examination of the loadings plot, we can extract a significant amount of information not easily obtained from a more traditional procedure. This allows us to easily chart the changes in the different QM parameter types, whether their variation is positively or negatively correlated, or whether the particular changes are negligible.

The first PCA component extracted describes 53% of the total variance in the descriptors, whereas the second describes just 17%. Many of the descriptors have large positive or negative loading (∼0.2) on the *x* axis, indicating that the descriptors in the dataset are highly intercorrelated. From the scores plot, we know the first component describes the structural differences between the zeolite−substrate complexes and alkoxide intermediates. Descriptors such as *r*C1−H, *r*Al−O1, and *r*C2−O2 and the zeolite and proton charges have significant negative loading on this component, indicating that as we go from zeolite−substrate complex to alkoxide intermediate, these values will increase in magnitude. We also find that descriptors such as *r*C1−C2, *r*O1−H, *r*Al−O2, and the Al charge are important to the first component, but in the positive sense. These descriptors are, therefore, inversely correlated to those with negative loadings and will decrease in magnitude on going from the zeolite−substrate complexes to alkoxide intermediates. Thus, component one describes the formation/breakage of the C1−H, C2−O2, and C1−C2 bonds and the associated changes in charge distribu-

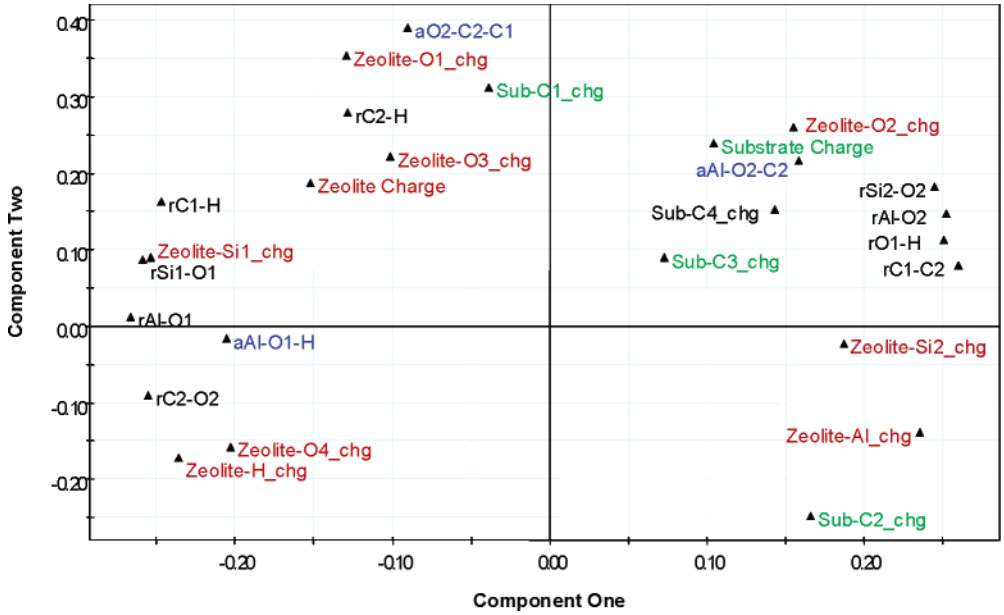

Figure 4. PCA loadings plot: plot of component one against component two.

tion. The component does not describe all the variations in these descriptors because they have nonzero values on the *y* axis. This means they have a small effect on the second component, which separates both the zeolite−substrate complexes and the alkoxide intermediates from the transition states.

There are only a small number of descriptors that have large loadings on the second component as given by the *y* axis, describing 17% of the total variation in the descriptors. No descriptors that directly describe bond formation or breakage load strongly onto the second component. Instead, we find charge-based descriptors and angles describing the interactions between the zeolite and substrates as being important. The key descriptors with large positive loading include the *a*O2−C2−C1, zeolite O2 and O3 atoms, the total zeolite and substrate charges, and the substrate C1 charge. Those with negative loadings are infrequent and considerably less influential because of their small loadings. These include the substrate C2 charge, the zeolite Al charge, and the acidic proton charge. Thus, component two essentially describes the changing hybridization state of the substrate that leads to a large increase in the net charge on the zeolite and substrate. This requires a quite significant reorientation of the substrate, as given by the *a*O2−C2−C1 descriptors. The appearance of the acidic proton charge with a negative coefficient, meaning its charge reduces on going from the minima to the transition states, can be rationalized on the grounds that it is close to forming an alkane-like CH bond to the substrate (∼1.25 Å). However, with the distinctly charged nature of the transition states compared to the minima, it is possible that the long-range electrostatic effects of the zeolite framework could alter what is a heavily concerted surface, potentially making the stepwise process more feasible.

The structural characteristics of all isobutene stationary points are distinctly different from all other alkenes and are, therefore, found as outliers in the scores plot. Similarly, the structural characteristics of the physisorbed ethene molecule at the 3T cluster are markedly different from those of other zeolite−substrate complexes. These differences are considerably larger than the random scatter within the clusters, suggesting that these are real structural differences and are not just attributable to noise in the PCA model. These stationary points are close to their related stationary points, but they are step-shifted along component two. The ethene zeolite−substrate complex is found step-shifted in the more positive direction, whereas all isobutene stationary points are step-shifted in the negative direction. On the basis of our analysis of component two, this would indicate that these stationary points must either be more charged or bind to the cluster in a somewhat distorted manner. An analysis of the data does, in fact, reveal that the ethene zeolite−substrate complex has the lowest net charge of all the bound substrates and has a relatively large *a*O2−C2−C1 angle, indicting a distinctly different conformation. Given that ethene is much less sterically hindered, it is understandable why this would be found. The converse argument can be used for isobutene, it being too sterically hindered to interact in the same way as the other less hindered alkenes. We find that the *a*O2−C2−C1 angle is significantly reduced for isobutene in all stationary points as is the overall charge found on the zeolite, the former having a significant influence on component two and the latter a more moderate effect.

Although PCA allows us to explore all the structural and electronic changes that occur in conjunction with the bond formation and breakage, the structural surface we have constructed does not take into account the relative energetics of the stationary points. We rectify this situation by studying how the energy changes with respect to our 26 QM parameters using both univariate and multivariate analyses.

3.2. Univariate Data Analysis. To determine how the energy of the different substrates changes with respect to structure, we consider the problem by separately evaluating what factors have a significant effect on the energy of the three different stationary points independently. This is because no single descriptor can account for the large differences in energy for all stationary points of all the substrates.

Our calculated adsorption energies follow the trend isobutene < *cis*-butene < *trans*-butene ≈ propene ≈

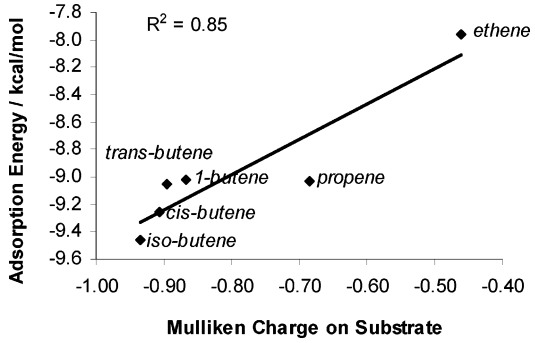

Figure 5. Plot of the substrate charge in the zeolite−substrate complex against the adsorption energy. $R^2 = 0.85$, $P = 0.009$.

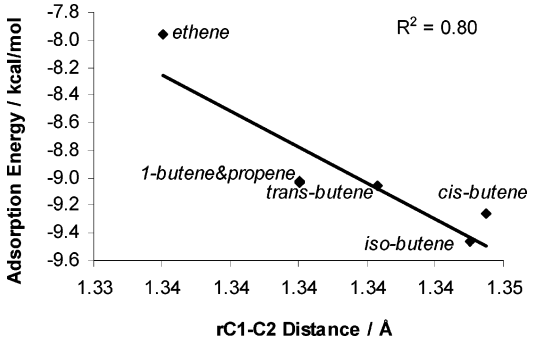

Figure 6. Plot of the C1−C2 distance in the zeolite−substrate complex against the adsorption energy. $R^2 = 0.80$, $P = 0.016$.

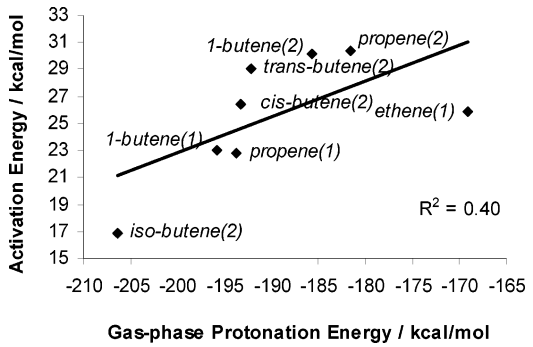

Figure 7. Plot of the gas-phase substrate protonation energy against the activation energy at the zeolite. $R^2 = 0.40$, $P = 0.092$. $R^2 = 0.85$ excluding ethene.

1-butene < ethene, suggesting the more sterically hindered the substrates, the greater their adsorption energy. Given that these 3T cluster adsorption results correlate relatively poorly with those derived from 10T-based ONIOM representations[44] or MM simulations[65] where the opposite is found, this suggests the model used here is unsuitable for ranking the adsorption energies of the alkene substrates. This most likely arises because the larger substrates do not experience an unfavorable energy penalty because the framework atoms adjacent to the 3T cluster are not present.

The trends we observe for adsorption energies may be unrealistic compared to those where the framework is included; however, understanding the reasons why this trend arises may help us to understand both the strength and deficiencies of the cluster approximation. From an analysis of the data, it is apparent that the adsorption energies of the substrates in the 3T models are governed by both steric and electronic factors. A plot of the absolute charge found on the substrate in the zeolite−substrate complex against the adsorption energies reveals a strong correlation ($R^2 = 0.85$, $P = 0.009$; Figure 5). An analysis of variance shows that the probability that this correlation could have occurred by chance is 0.01, above the commonly invoked 95% confidence level. Charge transfer onto the substrate is indicative of the strength of the interaction with the zeolite and allows us to differentiate between each of the substrates in terms of their predicted adsorption energies. Ethene followed by propene can accept the least amount of negative charge and, therefore, binds least effectively to ZSM-5. It is possible to explain the adsorption energies of the alkenes on the basis of the substrate charge. The strength of the interaction between the substrate and zeolite can also be followed by plotting the C1−C2 distance of the substrate against the adsorption energy (Figure 6). This parameter correlates more weakly with the energy compared to the latter parameter ($R^2 = 0.80$), although an analysis of variance still shows that this correlation is statistically significant, greater than the 95% confidence level.

The correlations found with both parameters show that the calculated adsorption energy is essentially governed by these key interactions between the substrate and the zeolite, giving rise to a weaker double bond between the C1 and C2 of the substrate and an increase in the negative charge delocalized onto the substrate. Including the effect of the framework will disfavor isobutene in particular and should bring these adsorption energies into line with others.[44,66] Even though there are clear inadequacies in our 3T model, the application of a statistical analysis on simpler systems can prove beneficial in understanding more complex models. We also might expect to observe similar correlations to parameters used here in more sophisticated models, albeit weaker because of other confounding effects due to the presence of the framework.

An analysis of the eight transition states shows that their energies do not correlate strongly with any geometrical or charge parameters in the same statistically significant way as found for the corresponding adsorption energies. This seems to be a result of the fact that the transition state energies are essentially independent of the zeolite itself, as modeled here. In general, we find that substrates that can protonate on a primary carbon are lower in energy than the corresponding secondary carbons; this is a result of the formation of a more stabilized secondary carbocation in the former. The relatively insignificant effect of the zeolite on the different substrates can be seen by plotting the gas-phase protonation energy of the substrates, which is the difference between the protonated and unprotonated substrates, against the corresponding transition state energy in the zeolite (Figure 7). This plot reveals a relatively weak correlation ($R^2 = 0.40$, $P = 0.09$), but it is not significant at the 95% confidence level. This is because ethene does not correlate as significantly as others, and this can be explained by the fact that this small alkene can avail of more stabilization from the zeolite, in relative terms, than the others because of its considerably smaller size. This manifests itself in a longer C1−H distance and a smaller absolute substrate charge compared to others. Removing this outlier improves the correlation ($R^2 = 0.85$), showing that, for the cluster results here, the barrier to reaction is governed by the basicity of the substrate and is essentially independent of the interaction with the zeolite. Boronat et al.[45] find a strong correlation ($R^2 = 0.85$) between the activation energy and the C2 charge

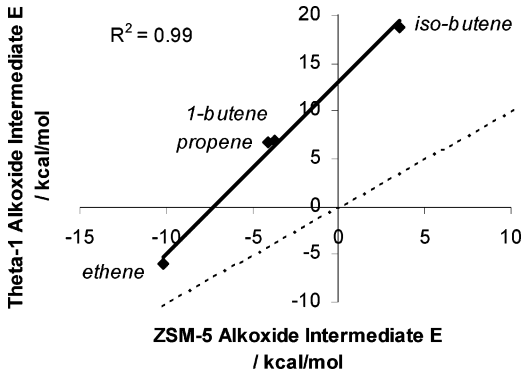

Figure 8. Plot of the 10T-based theta-1 alkoxide intermediate energies from Boronat et al.[44] against the 3T results from this study for the common substrates studied. $R^2 = 0.99$, $P = 0.003$. Line of unity shown as a dashed line.

for three different alkenes (ethene, propene, and isobutene) studied at a number of different acidic sites on the zeolite mordenite using the 10T-based ONIOM representation. We do not find as strong a correlation here with this particular variable ($R^2 = 0.66$, $P = 0.05$), which may be a manifestation of the greater range of substrates and the single acidic site investigated here, rather than a reflection of the different approaches.

These results show that the absolute charge found on the substrate is larger in the transition state compared to those of both minima, meaning the long-range electrostatic interactions with the zeolite framework would be expected to have a significant stabilizing effect on the barrier. This would suggest that using small zeolite clusters to describe this highly charged state may be a severe approximation, with our results differing from those where these effects are included.[44] Considering the common substrates in both studies, we find the activation energies in the latter study follow the trend 1-butene ~ propene < isobutene < ethene, compared to our 3T cluster results here where isobutene < propene ~ 1-butene < ethene. This suggests that the 3T cluster models can be used to differentiate between very different substrates, such that we can correctly rank isobutene and ethene but not substrates with more subtle differences such as isobutene and 1-butene.

In contrast to the adsorption energies or barriers reported here, we find a strong correlation ($R^2 = 0.99$, $P = 0.003$) between our 3T alkoxide intermediate energies and calculations that consider a more sophisticated 10T-based representation[44] (Figure 8). The 3T results correctly rank the relative energies of the different substrates; however, a significant overestimation of the calculated exothermicity is found (slope = 1.8, intercept = 13.1) for the four common substrates studied. Given the similar nature of the levels of theory and basis sets employed, this indicates that the predictions from the 3T cluster overestimate the exothermicity of the alkoxide intermediate, by ~10 kcal/mol on average. This is expected given that the unfavorable steric constraint imposed by the zeolite framework will not be present in the 3T system. However, the results do indicate that the key interactions that give rise to the differences in the alkoxide intermediate energies of these four substrates, ethene, propene, 1-butene, and isobutene, are primarily governed by the short-range effects associated with the 3T core.

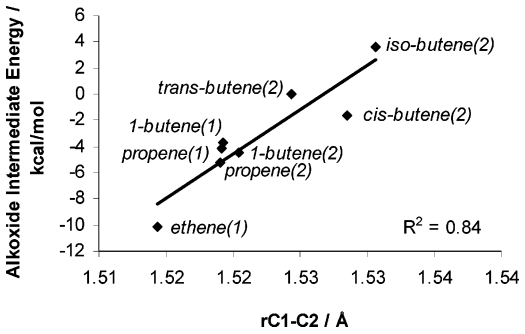

Figure 9. Plot of the C1−C2 distance in the alkoxide intermediate against the alkoxide intermediate energy. $R^2 = 0.84$, $P = 0.001$.

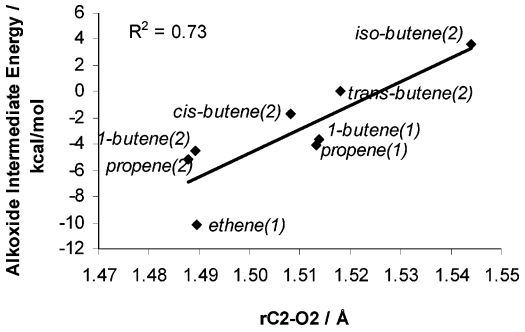

Figure 10. Plot of the C2−O2 distance in the alkoxide intermediate against the alkoxide intermediate energy. The larger sterically hindered substrates form the less exothermic alkoxide intermediates. $R^2 = 0.73$, $P = 0.006$.

Considering all the substrates studied, the energies of the alkoxide intermediates are generally lower for the smaller, less sterically hindered substrates. This is in contrast to the DFT adsorption energies where, isobutene > *cis*-butene > *trans*-butene ≈ 1-butene ≈ propene > ethene. This can simply be understood in terms of the less significant interaction between the smaller substrates with the atoms of the zeolite. Furthermore, we find that the relative energies of the substrates, which can protonate on either a primary or secondary carbon, have reversed compared to that observed in the transition state. A substrate that protonates on its primary carbon will require the zeolite O2 atom to attack the more saturated of the two possible carbon atoms, leading to a less stable intermediate. The difference between the two protonation states in the alkoxide intermediates is approximately −1 kcal/mol, considerably less than the differences in the barriers at approximately 7.5 kcal/mol.

An analysis of the relationship between alkoxide intermediate energies and our QM parameters shows that the greater the single bond character of the C1−C2 bond of the substrate, the lower the energy of the alkoxide intermediate ($R^2 = 0.84$, $P = 0.001$; Figure 9). Not surprisingly, the relative stability of the alkoxide intermediates also correlates with the strength of the O2−C2 bond ($R^2 = 0.73$, $P = 0.006$) formed between the zeolite and the substrate (Figure 10), as well as with the extent of charge delocalization onto the substrate (Figure 11; $R^2 = 0.70$, $P = 0.009$), the latter in good agreement with that found by others.[45] Furthermore, plotting the protonation energy of the substrate in the gas phase against the alkoxide intermediate formed in the zeolite reveals an inversely correlated relationship ($R^2 = 0.83$, $P = 0.002$; Figure 12). The smaller the substrate in the gas phase, the less stable the protonated form because of the lack of

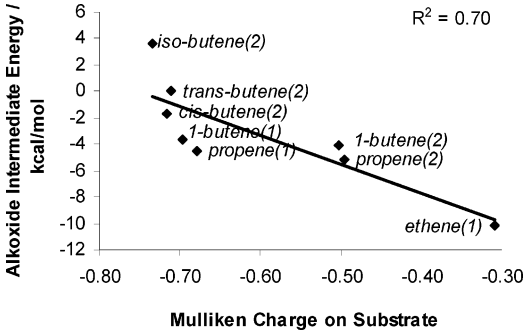

Figure 11. Plot of the charge delocalized onto the substrate in the alkoxide intermediate against the alkoxide intermediate energy. $R^2 = 0.70$, $P = 0.009$.

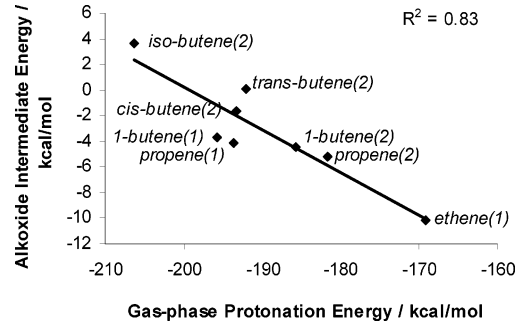

Figure 12. Plot of the substrate protonation energy in the gas phase against the energy of the alkoxide intermediate formed within the zeolite. $R^2 = 0.83$, $P = 0.002$.

Table 3. PLS and PCA Model Statistics[a]

model	r^2X	r^2Y	q^2	comp.
PCA	0.706		0.485	2
PLS	0.681	0.902	0.848	2

[a] PLS and PCA model statistics. r^2X refers to the variance in the descriptor matrix described by the PCA and PLS models. r^2Y refers to the variance in the energy explained by the PLS model. q^2 refers to the cross validated r^2X for the PCA model and r^2Y for the PLS model.

inductive stabilization by sidegroups. Yet, in the zeolite, the protonation and nucleophilic attack of a small substrate gives rise to the most stable alkoxide intermediate. Thus, the smaller the substrate, the less geometrically constricted the resultant alkoxide intermediate and the greater the charge transfer.

3.3. Partial Least-Squares Regression. A univariate analysis of data offers a rather limited insight into how the energy of stationary points on the reaction surface change with respect to the 3D structure. This is particularly true where we want to evaluate all stationary points of all the substrates at the same time because no single descriptor can be used to describe the considerable variability in the structure and the electronic distribution. In contrast, PLS provides us with a systematic way to assess how the energy of the stationary points varies with respect to the structure and the charge distribution.

A PLS model was generated using all 26 QM parameters and the energies of the stationary points expressed relative to the physisorbed zeolite−substrate complexes. The model consists of two components, describing 68% of the variance in the descriptors (r^2X) and 90% of the variance in the energy (r^2Y) (Table 3). The performance in cross validation is strong, as indicated by the model q^2 values approaching the r^2 values in magnitude, giving us confidence in the results. Additionally, *Y*-randomization trials show that no model built at random could approach the performance of the true model in fit or cross validation, giving us further confidence in the results.

The scores plot derived from the 26-descriptor PLS models differs quite significantly from the PCA model built using the same dataset (Figure 13). This is not unexpected given that PLS extracts components in order to maximize the correlation between the response and the extract components. An analysis of Figure 13 reveals that component one no longer splits the zeolite−substrate complexes from the alkoxide intermediates; instead, it separates the latter from the transition states. The second component separates the transition state and alkoxide intermediates from the minima. The difference between the PLS and PCA components can be understood by an analysis of the PLS weights plot, the equivalent of the loading plot in PCA (Figure 14). Component one in the PLS model is very similar to component two from the PCA model and vice versa. This switch occurs because PCA extracts components in order of how much variance they describe in the descriptor matrix, whereas in PLS, the component that describes the greatest amount of variation in the response (energy) is extracted from the matrix first. The zeolite−substrate complex and alkoxide intermediate energies are close together compared to that of the transition states, so PLS first takes into account the factors that lead to the transition states having significantly elevated energies compared to the minima. It is, therefore, not surprising that component one describes 82% of the total variance in energy compared to just 11% for the second component. By determining what descriptors are involved in component one, we can ascertain what the key factors are that give rise to the large activation energies. This we do from an analysis of the weights plot.

The weights plot in PLS differs from the loadings plot in PCA in that the response is also displayed (Figure 14). One can determine the effect of a particular descriptor on the response on the basis of where it is found relative to the latter. In Figure 14, the response is found close to zero on the *y* axis, meaning it is only marginally affected by the second component, as already mentioned above. This means descriptors found at the extreme ends of the *x* axis affect the energy to the greatest extent. The five most important descriptors for the first component are the zeolite O1, O2, and O4; proton charge; and the *r*C2−H distance. This means the greater the negative charge on the zeolite oxygen atoms that can directly interact with the substrate, the larger the barrier to the reaction.

The presence of the O4 charge with a significant positive weight is surprising, as it is not directly involved in the reaction. The O4 charge is only found to have an affect on the alkoxide intermediate energy, as this atomic charge remains essentially constant in the zeolite−substrate complex and transitions states. Although it might be expected that the O3 and O4 atoms of the zeolite cluster are equivalent, this is not the case as the O3 atom can interact somewhat with the substrate, particularly when the alkoxide intermediate forms. Thus, as the substrate interacts to a greater extent with the O3 atom in the product, it seems a greater proportion of the negative charge is delocalized onto the O4 atom. The model predicts that as the charge delocalized onto the O4

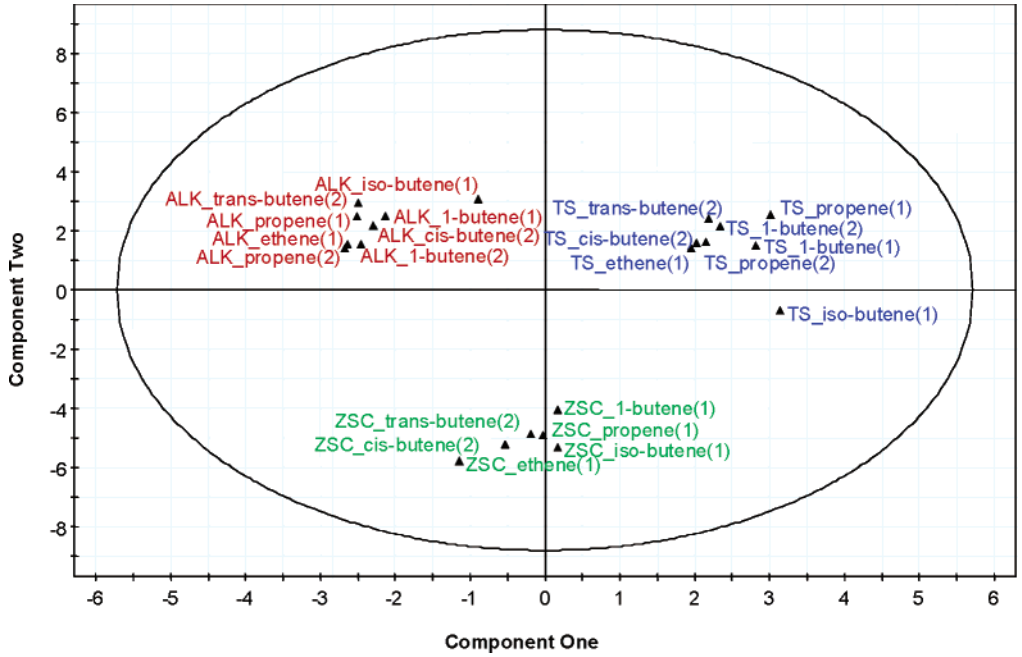

Figure 13. PLS scores plot: plot of component one against component two. Zeolite−substrate complexes denoted react_*, transition states denoted TS_*, and alkoxide intermediates denoted prod_*. A 1 or 2 indicates the carbocation type where possibilities exist.

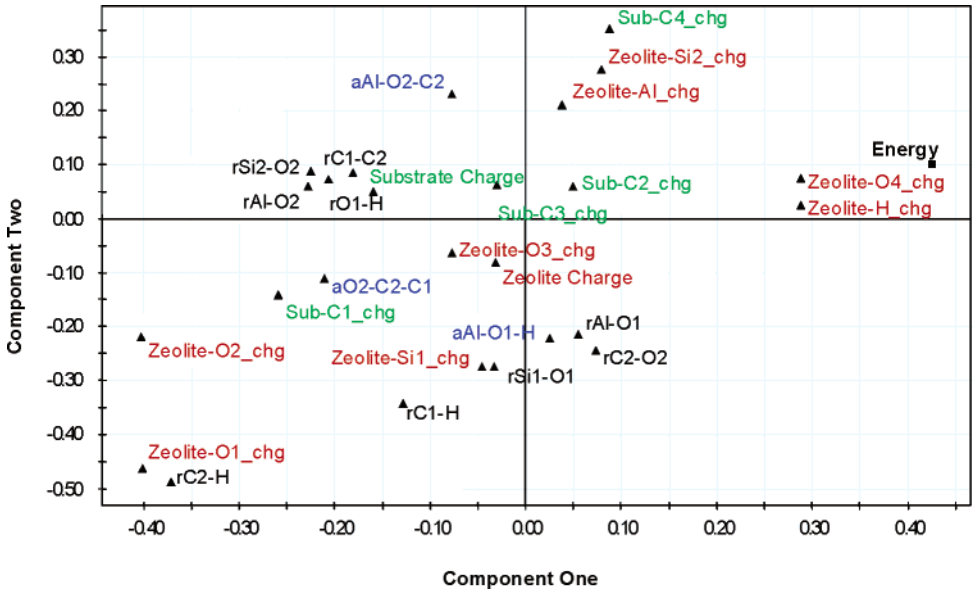

Figure 14. PLS weights plot: plot of component one against component two. The order in which the components are extracted differs compared to PCA. That is, component one from PLS resembles component two from PCA and vice versa.

atom increases, the energy of the alkoxide intermediate decreases. It is possible that this correlation occurred by chance or that the variable is acting as a surrogate for another factor.

4. CONCLUSIONS

Our results show that the relative energies of the alkoxide intermediates of ethene, propene, 1-butene, and isobutene are similar in both the 3T- and 10T-based calculations, suggesting that long-range effects are not as important to differentiate between the relative energies of the alkoxide intermediates formed within zeolite pores. However, the calculated adsorption and activation energies are poorly correlated, suggesting that the rather small 3T model used here is not suitable to compare different substrates when the treatment of medium-/long-range interactions is required to deal with stationary points where nonbonded electrostatic interactions are important.

We have employed PCA and PLS to help us understand the structural and electronic factors that affect the adsorption and reactivity of alkenes in acidic zeolites. PCA/PLS also allows us to chart the correlated nature of the structural and electronic changes along the reaction coordinate in a facile, systematic manner.

We find that the adsorption and the alkoxide intermediate energies correlate strongly with the charge delocalized from the zeolite to the substrate and the subsequent weakening of its double bond. There is significant correlation between the structural and electronic parameters for the transition state. The results also clearly show the concerted nature of the

transition state, involving elements of both proton transfer and nucleophilic attack.

These results show that it is crucial to account for all of the factors involved in zeolite catalysis to obtain realistic results. Even so, relatively small clusters can be used to elucidate important physical phenomenon in zeolites because of their greater applicability to a larger number of substrates with reduced computational expense.

ACKNOWLEDGMENT

The present research has been supported in part by a TRF Senior Research Scholar Grant and a TRF New Research Scholar Grant from the Thailand Research Fund, The Ministry of University Affairs, under the Science and Technology Higher Education Development Project (MUA-ADB funds), and Kasetsart University Research and Development Institute (KURDI).

Note Added after ASAP Publication. This article was released ASAP on June 29, 2005 with several incorrect figures and errors in the footnote of Table 3 and in the paragraph following the caption of Figure 10. The correct version was posted on July 25, 2005.

Supporting Information Available: The structural and electronic parameters of the species on the reaction coordinate for PCA and PLS models (Tables S1−S9). Key bond distances and angles as well as the Mulliken charges found on the heavy atoms of the zeolite and substrate and the acidic proton being transferred between them (Table S10). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Lukyanov, D. B.; Shtral, V. I.; Khadzhiev, S. N. *J. Catal.* **1994**, *146*, 87.
(2) Zygmunt, S. A.; Curtiss, L. A.; Zapol, P. *J. Phys. Chem. B* **1999**, *103*, 10417.
(3) Antia, J. E.; Israni, K.; Govind, R. *Appl. Catal., A* **1997**, *159*, 89.
(4) Rozanska, X.; Saintigny, X.; van Santen, R. A.; Clemendot, S.; Hutschka, F. *J. Catal.* **2002**, *208*, 89.
(5) Sassi, A.; Wildman, M. A.; Haw, J. F. *J. Phys. Chem. B* **2002**, *106*, 8768.
(6) Tielens, F.; Langenaeker, W.; Geerlings, P. *THEOCHEM* **2000**, *496*, 153.
(7) Andzelm, J. W.; Alvarado-Swaisgood, A. E.; Axe, F. U.; Doyle, M. W. *Catal. Today* **1999**, *50*, 451.
(8) Corma, A. *Chem. Rev.* **1995**, *95*, 559.
(9) Stepanov, V. G.; Ione, K. G.; Snytnikova, G. P. *Stud. Surf. Sci. Catal.* **1996**, *100*, 477.
(10) Rigby, A. M.; Kramer, G. J.; van Santen, R. A. *J. Catal.* **1997**, *170*, 1.
(11) Vollmer, J. M.; Troung, T. N. *J. Phys. Chem. B* **2000**, *104*, 6308.
(12) Rigby, A. M.; Frash, M. V. *J. Mol. Catal. A* **1997**, *126*, 61.
(13) Parker, W. O. *Inorg. Chem.* **2000**, *22*, 31.
(14) Lutz, W.; Ruescher, C. H.; Heidemann, D. *Microporous Mesoporous Mater.* **2002**, *55*, 193.
(15) van Bokhoven, J. A.; Koningsberger, D. C.; Ramaker, D. E. *J. Phys.: Condens. Mater.* **2001**, *13*, 10383.
(16) Simperler, A.; Bell, R. G.; Anderson, M. W. *J. Phys. Chem. B* **2004**, *108*, 7142.
(17) Simperler, A.; Bell, R. G.; Foster, M. D.; Gray, A. E.; Lewis, D. W.; Anderson, M. W. *J. Phys. Chem. B* **2004**, *108*, 7152.
(18) Yoo, K.; Burckle, E. C.; Smirniotis, P. G. *J. Catal.* **2002**, *211*, 6.
(19) Vos, A. M.; Nulens, K. H. L.; De Proft, F.; Schoonheydt, R. A.; Geerlings, P. *J. Phys. Chem. B* **2002**, *106*, 2026.
(20) van Donk, S.; Bitter, J. H.; de Jong, K. P. *Appl. Catal., A* **2001**, *212*, 97.
(21) Seo, Q.; Jeong, H. S.; Hong, S. B.; Uh, Y. S. *Catal. Lett.* **1996**, *36*, 249.
(22) Hancsok, J.; Hollo, A.; Debreczeni, E. *Stud. Surf. Sci. Catal.* **1999**, *125*, 417.
(23) Rozanska, X.; Saintigny, X.; van Santen, R. A.; Hutschka, F. *J. Catal.* **2001**, *202*, 141.
(24) Perego, C.; Ingallina, P. *Catal. Today* **2002**, *73*, 3.
(25) Gayubo, A. G.; Aguayo, A. T.; del Campo, A. E. S. *Ind. Eng. Chem. Res.* **2000**, *39*, 292.
(26) Sommer, J.; Jost, R.; Hachoumy, M. *Catal. Today* **1997**, *38*, 309.
(27) Frash, M. V.; van Santen, R. A. *Top. Catal.* **1999**, *9*, 191.
(28) Cejka, J.; Vondrova, A.; Wichterlova, B. *Zeolites* **1994**, *14*, 147.
(29) Cavani, F.; Trifiro, F.; Giordano, G. *Appl. Catal., A.* **1993**, *94*, 131.
(30) El Morsi, A. K.; Shokry, S. A. *Petrol. Sci. Technol.* **2000**, *18*, 1195.
(31) Arnstad, B.; Kolboe, S.; Swang, O. *J. Phys. Chem. B* **2002**, *106*, 12722.
(32) Svelle, S.; Kolboe, S.; Olsbye, U.; Swang, O. *J. Phys. Chem. B* **2003**, *107*, 5251.
(33) Svelle, S.; Arnstad, B.; Kolboe, S.; Swang, O. *J. Phys. Chem. B* **2003**, *107*, 9281.
(34) Arnstad, B.; Kolboe, S.; Swang, O. *J. Phys. Chem. B* **2004**, *108*, 2300.
(35) Arnstad, B.; Kolboe, S.; Swang, O. *J. Phys. Chem. B* **2004**, *108*, 2953.
(36) Zygmunt, S. A.; Curtiss, L. A.; Zapol, P.; Iton, L. E. *J. Phys. Chem. B* **2000**, *104*, 1944.
(37) Viruela, P.; Zicovich-Wilson, C. M.; Corma, A. *J. Phys. Chem. B* **1993**, *97*, 13713.
(38) Boronat, M.; Zicovich-Wilson, C. M.; Corma, A. *J. Phys. Chem. A* **1998**, *102*, 982.
(39) Sinclair, P. E.; de Vries, A.; Sherwood, P.; Catlow, C. R.; van Santen, R. A. *J. Chem. Soc., Faraday Trans.* **1998**, *94*, 3401.
(40) Correa, R. J.; Mota, C. J. A. *Phys. Chem. Chem. Phys.* **2002**, *4*, 375.
(41) Soscun, H.; Hernandez, J.; Castellano, O.; Arrieta, F.; Ruette, F.; Sierralta, A.; Machado, F.; Rosa-Brusin, M. *J. Mol. Catal. A.* **2003**, *192*, 63.
(42) Boronat, M.; Zicovich-Wilson, C. M.; Viruela, P. *Eur. J. Chem.* **2001**, *7*, 1295.
(43) Boronat, M.; Viruela, P.; Corma, A. *Phys. Chem. Chem. Phys.* **2000**, *2*, 3327.
(44) Boronat, M.; Zicovich-Wilson, C. M.; Viruela, P.; Corma, A. *J. Phys. Chem. B* **2001**, *105*, 11169.
(45) Boronat, M.; Viruela, P. M.; Corma, A. *J. Am. Chem. Soc.* **2004**, *126*, 3301.
(46) Sillar, K.; Burk, P. *J. Phys. Chem. B* **2004**, *108*, 9893.
(47) Rozanska, X.; van Santen, R. A.; Demuth, T.; Hutschka, F.; Haftner, J. *J. Phys. Chem. B* **2003**, *107*, 1309.
(48) Benco, L.; Haftner, J.; Hutschka, F.; Toulhoat, H. *J. Phys. Chem. B* **2003**, *107*, 9756.
(49) Höskuldsson, A. *Prediction Methods in Science and Technology*; Thor Publishing: Penhagen, Denmark, 1996.
(50) Wold, S.; Geladi, P.; Esbensen, K.; Öhman, J. *J. Chemom.* **1987**, *1*, 41.
(51) Wold, H. Path models with latent variables: The NIPALS approach. In *Quantitative Sociology: International perspectives on mathematical and statistical model building*; Academic Press: New York, 1975; pp 307−357.
(52) Wold, S.; Albano, C.; Dunn, W. J.; Edlund, U.; Esbensen, K.; Geladi, P.; Hellberg, S.; Johansson, E.; Lindberg, W.; Sjöström, M. Multivariate Data Analysis in Chemistry. In *Chemometrics: Mathematics and statistics in Chemistry*; Kowalski, B. R., Ed.; D. Reidel Publishing Company: Dordrecht, The Netherlands, 1984.
(53) Wold, S.; Eriksson, L.; Sjöström, M. *PLS in Chemistry, Encyclopedia of Computational Chemistry*; Wiley: New York, 2000.
(54) Cruciani, G.; Pastor, M.; Mannhold, R. *J. Med. Chem.* **2002**, *45*, 2685.
(55) Guha, R. I.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2179.
(56) de Paula da Silva, C. H. T.; Sanches, S. M.; Taft, C. A. *J. Mol. Graph. Model.* **2004**, *23*, 89.
(57) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
(58) Sutherland, J. J.; O'Brien, L. A.; Weaver, D. F. *J. Med. Chem.* **2004**, *47*, 5541.
(59) Buolamwini, J. K.; Assefa, H. *J. Med. Chem.* **2002**, *45*, 841.
(60) Wold, S.; Sjostrom, M.; Eriksson, L. *Chemom. Intell. Lab. Syst.* **2001**, 109.
(61) Seigbahn, P. E. M.; Pelmenschikov, V. *Inorg. Chem.* **2002**, *41*, 5659.
(62) Eriksson, L.; Johansson, E.; Kettaneh-Wold, N.; Wold, S. *Multi- and Megavariate Data Analysis: Principles and Applications*; Umetrics: Umeå, Sweden; ISBN 91-973730-1-X.
(63) *SIMCA-P 10*; Umetrics: Umeå, Sweden. www.umetrics.com.
(64) Stanton, R. V.; Perakyla, M.; Bakowies, D.; Kollman P. *J. Am. Chem. Soc.* **1998**, *120*, 3448.
(65) Jousse, F.; Leherte, L.; Vercauteren, D. P. *J. Mol. Catal. A* **1997**, *119*, 165.
(66) Asensi, M. A.; Corma, A.; Martinez, A. *J. Catal.* **1996**, *158*, 561.

CI0500583