

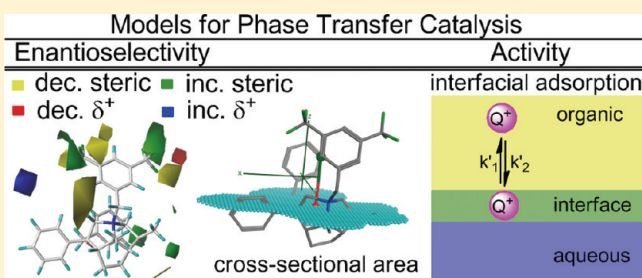
A Systematic Investigation of Quaternary Ammonium Ions as Asymmetric Phase-Transfer Catalysts. Application of Quantitative Structure Activity/Selectivity Relationships

Scott E. Denmark,* Nathan D. Gould, and Larry M. Wolf

Roger Adams Laboratory, Department of Chemistry, University of Illinois, Urbana, Illinois 61801, United States

Supporting Information

ABSTRACT: Although the synthetic utility of asymmetric phase-transfer catalysis continues to expand, the number of proven catalyst types and design criteria remains limited. At the origin of this scarcity is a lack in understanding of how catalyst structural features affect the rate and enantioselectivity of phase transfer catalyzed reactions. Described in this paper is the development of quantitative structure–activity relationships (QSAR) and –selectivity relationships (QSSR) for the alkylation of a protected glycine imine with libraries of quaternary ammonium ion catalysts. Catalyst descriptors including ammonium ion accessibility, interfacial adsorption affinity, and partition coefficient were found to correlate meaningfully with catalyst activity. The physical nature of the descriptors was rationalized through differing contributions of the interfacial and extraction mechanisms to the reaction under study. The variation in the observed enantioselectivity was rationalized employing a comparative molecular field analysis (CoMFA) using both the steric and electrostatic fields of the catalysts. A qualitative analysis of the developed model reveals preferred regions for catalyst binding to afford both configurations of the alkylated product.



under PTC. Third, QSAR methods are well suited for discovery-oriented, informatics-based research and hypothesis generation.^{9,10} Last, and most important, is that QSAR methods generate mathematical equations that facilitate the formulation of hypotheses, which logically leads to their application as a predictive tool. The ability to predict catalyst activity or selectivity *a priori* continues to serve as one of the “Holy Grails” of catalysis. This notion may be equally applied to APTC.

INTRODUCTION

A universal challenge in the chemical sciences is relating function to molecular structure. Linear free energy relationships (LFER) have served a fundamental role in physical organic chemistry by providing a quantitative correlation between reactivity and single group substitution.¹ Throughout the last century, the use of LFERs has been extended to include a multitude of parameters including steric² and electronic³ effects as well as lipophilicity⁴ and polarizability.⁵ Presently, extended forms of LFERs, namely, quantitative structure–activity relationships (QSARs), are a fundamental foundation upon which hypotheses of the biological function of small molecules are built.⁶ In contrast to the extensive application of QSAR methods to probe biological problems, these methods have only recently been applied to problems in chemical catalysis with respect to catalyst activity and selectivity.⁷ An area of catalysis for which QSAR methods exhibit high potential for applicability is phase-transfer catalysis (PTC).

A few interrelated aspects of QSAR methods are particularly attractive for application toward asymmetric phase-transfer catalysis (APTC) and warrant mention. First, QSAR methods have proven useful in understanding the relationship between the physicochemical properties of small molecules and the kinetics of their transfer across an interfacial barrier between two immiscible phases such as that present in all PTC systems.⁸ Second, QSAR methods have been extensively employed (and many descriptors developed) to investigate intermolecular, noncovalent interactions (such as drug-receptor binding) that are the hallmark of reactions

under PTC. Third, QSAR methods are well suited for discovery-oriented, informatics-based research and hypothesis generation.^{9,10} Last, and most important, is that QSAR methods generate mathematical equations that facilitate the formulation of hypotheses, which logically leads to their application as a predictive tool. The ability to predict catalyst activity or selectivity *a priori* continues to serve as one of the “Holy Grails” of catalysis. This notion may be equally applied to APTC.

The scarcity of design criteria for asymmetric phase-transfer catalysts in comparison to homogeneous catalysts has led to a rather unfortunate quandary for the methodological practitioner of organic chemistry hoping to develop new *asymmetric* phase-transfer catalysts. Currently, while one may consider what structural features should be included to impart enantioselectivity, the question of whether the envisioned catalyst will efficiently promote the desired reaction remains largely unanswered. For these reasons, we sought to investigate quantitative structure–activity/selectivity relationships to describe the rate enhancement and enantioselectivity exhibited by phase-transfer catalysts. The first part of this endeavor, the synthesis and evaluation of quaternary ammonium ion catalysts, has been extensively described in the preceding paper.¹¹ Herein, we report our efforts toward developing quantitative models for the enantioselectivity and activity of these catalysts.

Received: March 11, 2011

Published: March 29, 2011

BACKGROUND

1. Catalyst Activity. The primary objective of QSAR methods is to quantitatively model the variation in an activity observable as a function of variation in structure. Ideally, if physically meaningful descriptors are employed the origin of the relationship between structure and activity may be revealed. The most common experimental implementation of QSAR methods in the study of reactivity involves examining substrate reactivity as a function of systematic changes in a substituent. Apart from a few notable exceptions, the literature is deficient in reports on the application of quantitative methods to study catalyst activity. Among the most influential examples are studies that forged the concepts of general and specific acid and base catalysis.¹² Another recent example in the field of homogeneous catalysis is a systematic study of catalyst activity as a function of the hydrogen bond donating ability of a catalyst (pK_a).¹³ Finally, attempts to utilize QSAR methods to study catalysts for polymerization¹⁴ and homogeneous¹⁵ and heterogeneous catalysis are on record.¹⁶

For phase-transfer catalysts, structure–activity relationships have been established for simple, acyclic, achiral, quaternary ammonium ions that promote PTC reactions of small hydrophilic nucleophiles (e.g., cyanide, azide, thiolates).¹⁷ Such relationships for hydroxide-initiated PTC reactions are primitive by comparison.¹⁸ Typically, in these cases, the number of catalysts surveyed is less than 20 and the degree of structural variation is also limited.¹⁹ The most commonly used structural features are the number of carbons in the catalyst, and the ammonium ion accessibility.¹⁸ Accessibility is treated in a very limited, semi-quantitative manner that is applicable only to achiral, acyclic, unfunctionalized ammonium ions.²⁰ In a similar way, the hard–soft acid base principle (HSAB) is often employed in a qualitative sense to rationalize the difference in reactivity of small (hard or accessible) ammonium catalysts vs large (soft or inaccessible) quaternary ammonium phase transfer catalysts.²¹

In contrast to the small number of QSAR reports on catalyst activity, the literature is replete with QSAR studies on processes closely related to the fundamental steps of PTC including, inter alia, the rate of membrane permeation of small molecules,²² micelle formation,²³ and aqueous/organic phase-transfer rates.²⁴ The capacity for such descriptive models to be predictive is increasing rapidly. Furthermore, studies of these phenomena have elucidated many of the key structural features which are now included as descriptors in a number of computational suites for drug design.²⁵

2. Enantioselectivity. The field of computational drug design also offers methods capable of facilitating the elucidation of the structural features responsible for enantiotopic molecular recognition. Enantioselectivity has been modeled using geometrical descriptors such as steric size,²⁶ topological indices,²⁷ continuous chirality quantification methods²⁸ that indirectly account for 3-dimensionality, as well as various molecular interaction field (MIF) analyses. A MIF-based approach was chosen to initiate these studies as it typically provides a more direct and more information-rich representation of the 3-dimensional features necessary to reflect the enantiotopic differentiating capacity of chiral catalysts.

MIF-based approaches²⁹ are standard 3D-QSAR techniques employed in drug design. MIF algorithms encode variation in structure by positioning a probe atom and/or a point charge at fixed grid points around each molecule of interest, and the

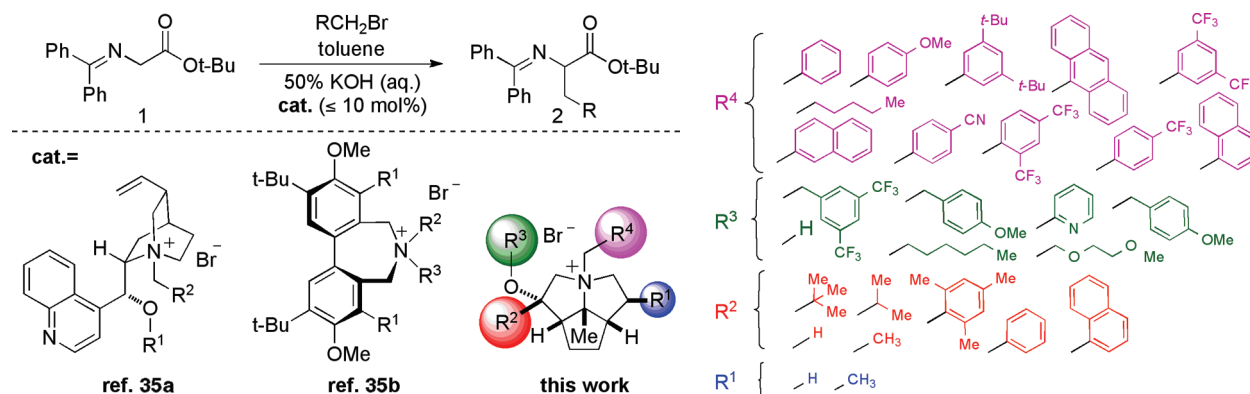
interaction energies with the probe are recorded at each grid point. The dependent variable (typically a free energy of binding) is then linearly correlated with the interaction energies for which the coefficients are extracted from a multivariate linear regression analysis, commonly by a partial least squares (PLS) method. Although MIF approaches have enjoyed a long-standing history of greater than 20 years in medicinal chemistry for the development of drug candidates, only relatively recently (<8 years) have they been applied to problems relating to asymmetric catalysis.

The most commonly employed MIF approach is comparative molecular field analysis (CoMFA). The method uses a molecular mechanics based force field to approximate van der Waals interactions and the standard Coulombic potential is used for electrostatic interactions for which the partial atomic charges are determined at the desired level of theory.³⁰ The appeal of this method stems from its predictive capacity and ability to allow one to visualize the developed model in terms of regions where field variation (either electrostatic or steric in origin) within the data set leads to a change in the dependent variable. In the field of asymmetric catalysis, this method was first applied in the analysis of the Diels–Alder reaction between cyclopentadiene and a 3-vinylloxazolidin-2-one using copper(II) Lewis acids with differing bisoxazoline and phosphinooxazoline ligands.³¹ Methods that incorporate semiempirical³² as well as ab initio quantum mechanical³³ interaction energies have also been developed and have found application in analyzing asymmetric diethylzinc additions using varying chiral amino alcohols³⁴ as well as asymmetric lithiation using varying chiral sparteine surrogates.³³

Two seminal reports employing CoMFA in the context of PTC involved the asymmetric alkylation of a protected glycine imine *tert*-butyl ester using different catalyst systems (Scheme 1).³⁵ In the first report,^{35a} a model was developed employing varying cinchona alkaloid derived catalysts. In the second report,^{35b} a model was generated for the same reaction while using different catalyst scaffolds developed by Lygo and co-workers.³⁶ The contributions of steric and electrostatic interactions to the variation in enantioselectivity were found to be roughly equivalent but the implications of electrostatic interactions were not discussed in detail in these reports, which are considered to be the predominate forces responsible for the ion-pair interaction strength.³⁷ Despite these important contributions, ambiguity is still present as to the dependence of electrostatic interactions on the observed enantioselectivity.

3. Objectives of This Study. The primary objectives of this study were to develop quantitative structure–activity and –selectivity relationships of quaternary ammonium ion asymmetric phase-transfer catalysts. These objectives were addressed in three stages of investigation: (1) the synthesis of a large number of diverse quaternary ammonium salts with variable physical properties; (2) accumulation of an internally consistent data set by evaluation of the catalysts ability to promote an enantioselective enolate alkylation; and (3) development of QSARs to correlate changes in catalyst structure to the observed rate and enantioselectivity of alkylation. The first two objectives were described in the accompanying paper.¹¹ The last objective was motivated by a number of interrelated queries. For example, would a QSAR approach generate testable hypotheses about the origin of rate and selectivity of the catalysts? If so, would these hypotheses be consistent or inconsistent with qualitative observations for other PTC enolate alkylations? Would a multivariate QSAR reveal any fundamentally important structural features inherent in desirable

Scheme 1



catalysts? Described herein are the development and analysis of QSAR models for the enantioselectivity and activity of the quaternary ammonium ion phase-transfer catalysts reported in the accompanying paper (DOI: 10.1021/jo2005445). The preceding questions are addressed throughout the Discussion as well as the Conclusion.

Before proceeding further, it must be stressed that we are well aware that the reaction chosen for this study can be performed highly enantioselectively with well-developed catalyst systems.³⁸ That success notwithstanding, the observation of high selectivity does not belie an *understanding* of the origin of such selectivity. For these reasons, the objective of this study, in the context of enantioselectivity, was not to develop catalysts for which the preparative utility is competitive with current systems; rather, the goal was to formulate meaningful structure–selectivity relationships employing a catalyst scaffold that is modifiable with respect to the steric and electrostatic environment circumscribing a quaternary, stereogenic nitrogen atom (Scheme 1).³⁹ If successful in generating meaningful relationships, these principles may serve as potential design criteria for the rational development of catalytic, enantioselective systems for which APTC variants do not exist.

COMPUTATIONAL METHODS

1. Conformations of Catalysts for QSAR. The choice of catalyst conformation is relevant for descriptors that are conformer-dependent including principle moment of inertia, dipole, surface area, and the spatial distribution of the interaction energies in the CoMFA analysis. Some ambiguity arises when considering the appropriate ring-flip geometry for each of the catalysts employed in this study, and thus, a conformational analysis proved necessary. Minimum conformers for each catalyst were generated using the MMFF force field and a Monte Carlo based conformation search through the use of an annealing algorithm as implemented in Spartan '08 v1.2.0.⁴⁰ To ensure that the MMFF conformers were converging to reasonably stable minima, full geometry optimizations for representative catalysts were carried out at the B3LYP/6-31G(d) level of theory for each ring-flip conformer obtained from the MMFF conformation distribution analysis which resulted in three stable local minima using the MMFF force field. Gratifyingly, none of the three optimizations changed the ring-flip conformations identified by the MMFF force field. Moreover, the local minima maintained the same relative energies, validating the use of the MMFF conformers (see the Supporting Information). Additionally, single point energies were determined including and excluding solvation

(SM8⁴¹ solvation model) in the absence of a counterion; again, the energetic ordering of the conformers was equivalent.

The lowest energy conformers of the isolated cations may not accurately reflect the major reactive conformation for the ammonium ion in the reaction medium. To address this deficiency, CoMFA analyses were performed on multiple conformer libraries with varying scaffold geometries obtained from the MMFF minimizations. The consequences of each developed model will be addressed in the Results.

2. Enantioselectivity Model Development. Models for enantioselectivity were generated using CoMFA to generate the interaction energies and the method of PLS was used for the regression analysis as implemented in SYBYL-X 1.1.⁴² All developed PLS models for enantioselectivity include five components (latent variables). The dependent variable was represented as a free energy term, $-\Delta G/RT = \ln(R/S)$, because it is expressed as a linear combination of energy terms in the PLS model. A full table of the enantioselectivity data is provided in the Supporting Information. The observed *er* range spans from 36:64 to 81:19 (*S*:*R*). Although the enantioselectivities observed in this investigation fall short of being synthetically useful, the range (1.2 kcal) has allowed for meaningful conclusions to be drawn from the models developed herein which, as will be revealed in the Results and Discussion sections, are consistent with qualitative observations.⁴³

The electrostatic fields were calculated from the partial atomic charges on all of the atoms. During preliminary studies, different partial charge methods were investigated including MMFF, Gasteiger–Marsili, and semiempirical methods (AM1, MNDO, and PM3). Models using MNDO charges generally provided better correlations. CoMFA models developed from semiempirical based charges have been demonstrated to provide consistently more predictive models than models developed from partial equalization of orbital electronegativity (PEOE, Gasteiger) and molecular mechanics based methods.⁴⁴ Thus, all of the remaining discussion of CoMFA model development will have incorporated MNDO ESP-based partial atomic charges as determined from single-point calculations on the MMFF conformers.

An oftentimes frustrating requirement in performing a MIF analysis is that the molecules should be aligned in a rational manner. Ideally, the molecules are aligned such that the amount of variation in the fields that can account for the variation in the dependent variable is maximized while the amount of variation in the fields that cannot account for the variation in the dependent variable is minimized. Frequently, the proper alignment scheme necessary to achieve this ideal is not apparent, and thus, multiple alignments are investigated. Alignment-independent methods have been developed to address this deficiency,^{34,45} although some of the information is inevitably lost in the process whether it belongs to signal or noise.

The rigidity of the scaffold led to a relatively straightforward decision on how to align the molecules in an effort to minimize the variation in

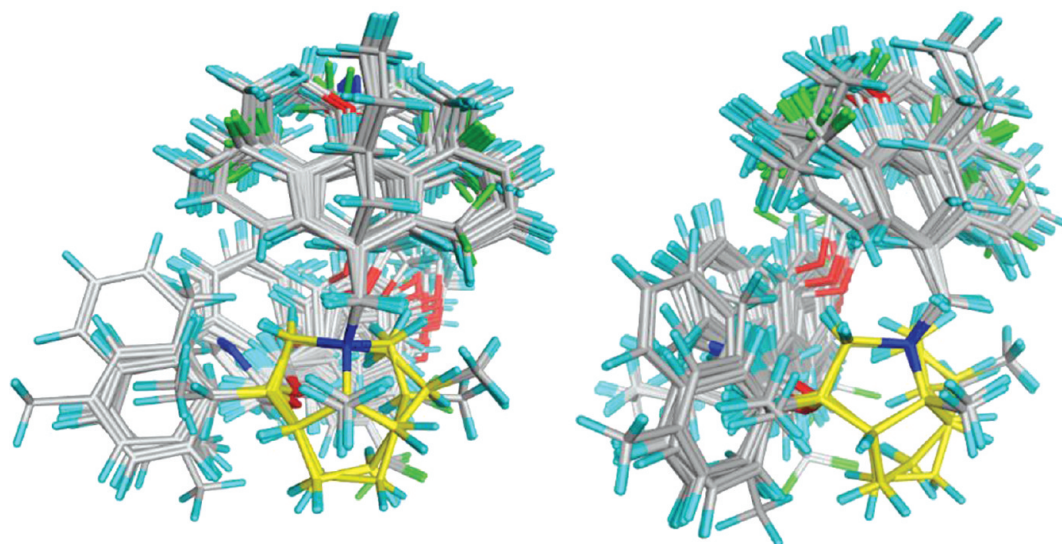


Figure 1. Rigid rms alignment for a representative conformation library (101 catalysts) from two perspectives. The core 5–5–5 scaffold is highlighted in yellow.

their fields (maximize the ratio of variation accountable versus variation that is unaccountable). The structures were aligned employing a simple rms rigid-body alignment. The common substructure used for the alignment is represented by the nine atoms that make up the core 5–5–5 ring scaffold. An example of an alignment for one of the conformation libraries (see the Results) is illustrated (Figure 1).

The electrostatic and van der Waals energies were calculated using two separate methods: Tripos Standard and Indicator field⁴⁶ classes. Under the Tripos Standard field method, the energies at each lattice point are evaluated from the Lennard-Jones (LJ) (6–12) (van der Waals) and the Coulombic potentials (electrostatic). Cutoff energies are applied so that energies are not evaluated within the van der Waals surface of the atoms where the energy would approach unreasonably high numbers that are not fit for comparison. Under the indicator field method, grid points are assigned as having either the preassigned cutoff value or zero. If the energy is calculated (by the LJ or Coulomb potentials) to be above the cutoff value, the energy at that grid point is set to the cutoff value. If the energy is calculated to be below the cutoff value, the energy at that grid point is set to zero. An advantage to the indicator field method is that only grid points of intermediate energy are selected which has a tendency to reduce model noise.^{35b}

The resultant CoMFA models were further refined by placing weights on grid points that were more pertinent to the model through the region focusing technique as implemented in Sybyl-X 1.1. The grid points were weighted by the discriminate power option which weights each grid point by its contribution to the variation in the components of the model. This effectively enhances grid points with larger contributions while attenuating grid points that are less pertinent to the model. The exponent that gauges the steepness of the applied weights was varied between 0.2 and 0.8. Application of this region focusing procedure was carried out iteratively until no improvement in the q^2_{LOO} was observed.

3. Catalyst Activity Model Development. *3.1. Descriptors.* A molecule can be characterized in an infinite number of ways. Because many descriptions of a molecule are neither a physical nor a chemical molecular property, the term “descriptor” is preferred over “property” to relate a calculated numerical characterization of a molecule for the purposes of a QSAR study. For the remainder of this report, the term “descriptor” will be used exclusively for that purpose. The computational package Molecular Operating Environment (MOE) was chosen for this study.⁴⁷ The MOE computational package contains 319 descriptors ranging from the most simple (1-D) atom counts (e.g., number of

carbons) to complex (3-D) surface area and volume descriptors (e.g., amphiphilic moment), and all were included in this analysis.⁴⁸

Many studies suggest that the solubility of the quaternary ammonium ion in the organic phase is an important catalyst structural feature (vide supra).¹⁷ To address thermodynamic solubilities, a variety of solvation parameters were included in the analysis, connectivity dependent but conformation independent methods (2D) and conformation dependent DFT methods (3D).⁴⁹ Solvation energies were determined by the SM8 solvation model (B3LYP/6-31+G(d)) for each catalyst in water and benzene.^{41,50}

Particular emphasis was placed on addressing ammonium accessibility and polarizability in a quantitative fashion. Therefore, a number of customized quaternary ammonium ion descriptors were developed on the basis of the accessibility of the α carbon(s) of the ammonium ion (represented in terms of solvent accessible surface area (A^2)) with and without various charge weights. These will be discussed in detail in a following section. In addition, 543 surface area and charge density descriptors were included.⁵¹ Similarly, the overall polarizabilities (and hyperpolarizabilities) of each quaternary ammonium ion were calculated quantum mechanically,⁵² and 50 HSAB and inductive descriptors were included.⁵³ In total, 1102 descriptors were compiled.

3.2. Data Manipulation and Statistical Methods. The kinetic data reported in the previous paper (DOI: 10.1021/jo2005445) were transformed into data suitable for QSAR by taking the logarithm of the ratio of the observed half-life relative to the half-life of the background reaction.⁹ Initial descriptor evaluation was conducted utilizing a genetic algorithm (GA)⁵⁴ in combination with multiple linear regression (MLR) as a preliminary search for descriptors and pairwise combinations thereof (e.g., solubility + polarizability) that account for the greatest amount of variation in the data. Each evolution was allowed to run for 50000 generations or until no change was observed over 1000 generations. During evolution, the “quality” of models was evaluated by comparison of lack of fit.⁵⁵ Final models were analyzed by internal and external validation in the same manner as in the CoMFA procedure (vide infra). The linear models were evaluated for coefficient of determination (R^2), root mean squared error (RMSE), and fit (F) and will be discussed on a per-model basis. The descriptors utilized for developing a model for catalyst activity are intrinsically less dependent on conformation (lower dimension) than MIF approaches; nonetheless, the same conformations, charges, and alignment scheme were utilized as in the CoMFA model. Molecules included in the rate data that do not

share the nine-carbon scaffold were rigidly aligned by superposition of the four ammonium α carbons.⁵⁶

4. Model Validation. The predictive capacity of the models was assessed through internal and external cross-validation.⁵⁷ The internal cross-validation was performed employing the leave-one-out (LOO) and leave-multiple-out (LMO) cross-validation methods. The q^2_{LMO} is subject to variation, especially for smaller training and prediction set splits. Thus, 100 LMO cross-validation runs were performed, and the corresponding q^2_{LMO} is reported as the average over the 100 runs. External validation was performed upon the judicious division of the entire data set into training and test sets.

The model robustness was assessed through γ -scrambling analysis.⁵⁸ The dependent variable data were completely scrambled such that each half-life or enantioselectivity value is paired up with the incorrect set of descriptors that are calculated for a particular catalyst. This process was performed 100 times, and the average R^2 and q^2_{LOO} are reported.

RESULTS

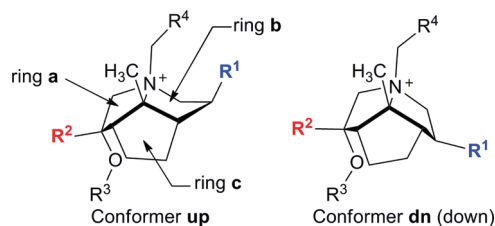
The model development for enantioselectivity and rate involved fundamentally different approaches and will be presented and discussed separately with the enantioselectivity model development being presented first. The CoMFA modeling required multiple stages of development. The first stage involved the choice of conformers to use to represent each catalyst. Preliminary CoMFA modeling was then performed on each conformer library to obtain knowledge on the optimal representations of the catalysts for which the greatest variation in selectivity may be explained by the variation in their fields. Since both the rate and enantioselectivity models include conformation-dependent descriptors, the conformations identified from the preliminary CoMFA modeling may be used for the model development for catalyst activity. After the ideal conformational representations for each catalyst was established from preliminary CoMFA modeling, the different CoMFA parameters (cutoff energies, field types, dielectric, etc.) were explored on the optimal conformer library to generate the best models as determined from internal cross-validation methods.

As may be gathered from the information that will follow, the results of the enantioselectivity model development revealed that (1) a statistically significant CoMFA model could be developed with this data set given the appropriate conformational representations, (2) these representations may provide clues as to the stereochemical course of the reaction, and (3) the preference of either configuration of the product may have both steric and electrostatic components.

1. Enantioselectivity Model Development. *1.1. Establishing Catalyst Conformations.* This stage often represents a significant challenge for 3-D QSAR development, and various approaches have been adopted.⁵⁹ The catalysts studied herein have relatively few degrees of freedom at ambient temperature (within the scaffold) allowing for a relatively unambiguous conformational investigation. The most significant conformational differences are those represented as the “up” and “dn” conformers (Table 1). The preference for ring a to be in the “down” conformation is large enough such that the “up” conformer for this ring need not be considered.⁶⁰ The preference for either the up or dn conformation as shown is primarily a function of the identities of the R^1 and R^2 substituents and the libraries in Table 1 (libraries A–E) are organized as such. A first approximation of reasonable conformations was the global minimum for each catalyst (library A, Table 1).⁶¹ The geometry optimization for each catalyst was carried out in the absence of a

counterion, and thus, the probability with which they represent active conformations is uncertain. Hence, alternative conformational representations and different combinations thereof were investigated. Taking into consideration the minimal energy difference between the “up” and “dn” conformers,⁶¹ two additional conformer libraries were generated, one with all of the catalysts in the “dn” conformation (library B) and another with all of the catalysts in the “up” conformation (library C). The investigation of libraries D and E only became apparent after the initial CoMFA modeling was carried out and will be described in that order.

Table 1. Two Primary Conformational Representations of Catalysts and Table of Libraries Representing Differing Combinations of Conformers Up and Dn Dependent on R^1 and R^2



library	R^2/R^1				
	H/H	H/Me	Me/Me	(<i>i</i> -Pr or <i>t</i> -Bu)/Me	aryl ^a /Me
A ^b	up ^c	up	dn	dn	dn
B	dn	dn	dn	dn	dn
C	up	up	up	up	up
D	up	up	up	dn	up
E	dn	dn	dn	up	up

^a aryl = Ph, 1-naphthyl, mesityl. ^b Library containing differing conformer combinations. ^c Conformation of scaffold.

CoMFA modeling of the enantioselectivity was carried out using a limited number of cutoff energy combinations (15 or 30 kcal/mol) for both electrostatic and steric fields of standard or indicator field types. An example of a rigid body alignment is illustrated in Figure 1 (conformer library D). A condensed summary of these results in terms of their coefficients of determination (R^2/q^2_{LOO}) is compiled in Table 2. Generally, MNDO semiempirical-based ESP partial charges provided models with the highest correlations and are considered here in calculating the electrostatic energies for the CoMFA analysis (Table 2).⁶² The correlations for libraries A–C fall short of the minimum for statistically significant predictions ($q^2_{LOO} \geq 0.6$). Inspection of the residuals for the cross-validated runs (q^2_{LOO}) revealed that catalysts with $R^2 = i$ -Pr and the most selective catalysts ($R^2 = \text{aryl}$, $R^4 = 3,5$ -bis(trifluoromethyl)benzyl) exhibited the largest error in the predictions. Although library C does contain one model with a $q^2_{LOO} > 0.6$, the result is quite sensitive to the cutoff energies. Therefore, additional conformer libraries were envisioned to address the low correlations.

Conformer library D (Table 1), in which $R^2 \neq i$ -Pr or *t*-Bu, possesses the up conformation, while catalysts with $R^2 = i$ -Pr or *t*-Bu possess the dn conformation, was envisioned to address the error in predictions for these catalysts. Additionally, library E, which possesses the opposite ring-flip geometries with respect to

Table 2. CoMFA Modeling (R^2/q^2_{LOO}) of Different Conformer Libraries from Table 1 with Varying Cutoff Energies and Field Types (Standard and Indicator)

cutoff energy (field type)	conformation library ^a				
	A	B	C	D	E
30/30 ^b Std ^c	0.728 ^d /0.547 ^e	0.697/0.465	0.711/0.493	0.815/0.612	0.794/0.574
30/15 Std	0.729/0.551	0.700/0.481	0.711/0.485	0.814/0.598	0.794/0.586
15/30 Std	0.734/0.545	0.705/0.498	0.709/0.460	0.803/0.641	0.812/0.636
30/30 Ind ^f	0.724/0.474	0.697/0.423	0.712/0.477	0.835/0.648	0.768/0.527
30/15 Ind	0.797/0.484	0.754/0.442	0.835/0.619	0.924/0.778	0.782/0.474
15/30 Ind	0.711/0.462	0.693/0.416	0.724/0.466	0.810/0.604	0.795/0.557

^a See Table 1. ^b Steric cutoff energy (kcal/mol)/electrostatic cutoff energy (kcal/mol). ^c Standard Tripos field. ^d R^2 . ^e q^2_{LOO} . ^f Indicator field.

Table 3. CoMFA Models with Optimal Cutoff Energies, Standard and Indicator Field Types, Internal Cross-Validation, Y-Scrambling, with and without the Application of Region Focusing

field(s)_field type	cutoff energy ^a (steric/elec)	R^2	q^2_{LOO}	q^2_{LMO} ^b	y-scramb ^c R^2_{scramb}	y-scramb $q^2_{\text{LOO,scram}}$
Both ^d _STD ^e	35/5	0.846 ^f /0.875 ^g	0.669/0.738	0.615/0.705	0.383/0.338	-0.090/-0.050
elec ^h _STD ⁱ	15/30	0.865 ^j	0.543	0.506	0.413	-0.073
Steric ^k _STD	40	0.794/0.829	0.605/0.710	0.583/0.697	0.350/0.301	-0.103/-0.084
Both_IND ^l	20/5	0.940/0.944	0.794/0.890	0.760/0.878	0.579/0.429	-0.137/-0.096
elec_IND	10/5	0.923/0.887	0.750/0.799	0.705/0.766	0.594/0.388	-0.069/-0.030
steric_IND	25	0.840/0.867	0.648/0.749	0.627/0.732	0.431/0.378	-0.122/-0.102

^a kcal/mol. ^b Leave 20% out cross-validation average over 100 runs. ^c Average correlation coefficient over 100 completely scrambled iterations. ^d Both = electrostatic and steric fields. ^e R^2 for model constructed from an unfocused region. ^f R^2 for model constructed from a focused region. ^h Electrostatic field only. ⁱ STD = standard Tripos field. ^j Unfocused region only. Region focusing did not improve the correlations. ^k Steric field only. ^l IND = indicator field.

library **D**, was investigated for purposes of comparison. Clearly, conformer library **D** consistently provided the highest correlations, with one model (30/15; indicator field) exhibiting a correlation that may be statistically relevant ($q^2_{\text{LOO}} > 0.778$). Interestingly, conformer library **E** (possessing conformers opposite to that of **D**) provided the second highest correlations, alluding to the importance of the conformational difference between ring **a** and ring **b** of the catalysts at the extremes of the er spectrum. The physical justification for the nature of the conformations in library **D** will be addressed in the Discussion.

To attest to the statistical prevalence of the conformations present in library **D**, 500 libraries of random (but unique) distributions of **up** and **dn** conformers were generated. The average q^2_{LOO} over the 500 runs was 0.409 with a maximum q^2_{LOO} of 0.679. Reassuringly, the library with $q^2_{\text{LOO}} = 0.679$ contained conformers that closely mimicked that present in library **D** in that the most selective catalysts (>75:25) possessed the **up** conformation and the most selective in the opposite direction (<38:62) possessed the **dn** conformation.

1.2. Internal Cross-validation. With conformations that lead to models possessing statistically significant correlations (represented by library **D**, Table 1), optimization of the various CoMFA parameters (cutoff energies, dielectric, exponent of the repulsive term in the Lennard-Jones potential, etc.) can be carried out. Models incorporating both indicator and standard field types were generated. Additionally, models incorporating only the electrostatic or steric field, individually, were generated. The assumption in the development of models with individual fields is that all of the interaction responsible for the variation in enantioselectivity is either electrostatic or steric in origin. Because overlap in these fields is possible (an interaction energy at a

grid point may be interpreted as steric but also be interpreted as electrostatic and vice versa), analysis of models constructed from individual fields may be informative.

In a typical CoMFA analysis, the descriptors significantly outnumber the dependent variables, thus rigorous cross-validation methods are necessary to substantiate the predictive capacity of the models. Accordingly, different internal cross-validation methods were carried out (q^2_{LOO} and q^2_{LMO}) in conjunction with y-scrambling (which assesses model robustness) (Table 3). The coefficients in the absence and in the presence of region focusing are presented. The results clearly show that indicator fields provide models that exhibit better overall correlations. The y-scrambling results (low average R^2 and q^2 over 100 runs) suggest that the models are statistically significant and are not subject to chance correlation between the randomized enantioselectivities and the descriptors as the $q^2_{\text{LOO,scramb}}$ for each model is never a positive value and the corresponding R^2_{scramb} values are minimal ($R^2_{\text{scramb}} < R^2/2$). Additionally, external cross-validation has been performed through analysis of successive training and test set splits and further supports the sufficient predictive capacity of the model ($R^2_{\text{test,avg}} = 0.880$).⁶³

2. Catalyst Activity Model Development. Before undertaking the development of a full QSAR for catalyst activity, two important questions had to be addressed, namely, (1) what descriptors are capable of reflecting ammonium ion accessibility, and (2) is a multidimensional QSAR even necessary or is a one-dimensional QSAR (LFER) possible? That is, would an accessibility descriptor for an ammonium ion sufficiently account for all of the variation in catalyst activity expressed in this data set?

2.1. Investigation of Ammonium Ion Accessibility. Previous studies employing unfunctionalized ammonium ion catalysts

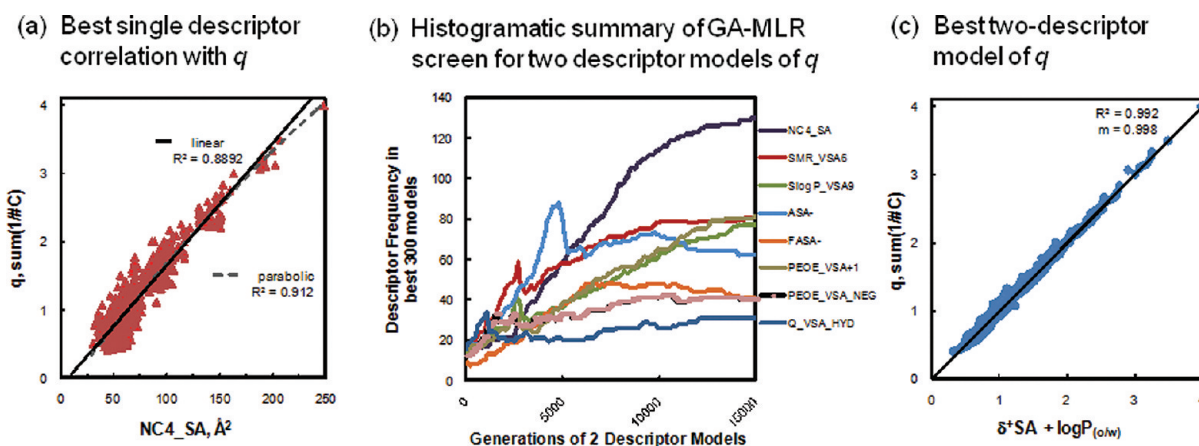


Figure 2. Summary of the descriptor and model screening strategy. (a) A comparison of a linear and nonlinear single descriptor model of q ; NC4_SA = the water accessible surface area of the ammonium α -carbons. (b) GA-MLR run terminated at 15000 generations. The frequency of inclusion of the 300 “best” descriptors in the 300 best models is shown (y -axis) versus the number of generations (x -axis). (c) Best two-descriptor model for $q = -7.502 - 6.03 \times \log P(o/w) + 5.27 \times \text{PEOE_VSA_POS}$.

showed that the PTC alkylation of deoxybenzoin and phenylacetonitrile have similar catalyst structure–activity relationships.⁶⁴ In these cases, the catalyst activity is best rationalized in terms of ammonium ion accessibility as defined by the parameter (q).²⁰ The following preliminary descriptor and model survey for ammonium accessibility serves as an example description of the general model development strategy utilized throughout this study. Because q is only defined for linear quaternary ammonium ions, it was essential to identify descriptors that reflect ammonium ion accessibility that could be applied to *all* quaternary ammonium ions. The initial approach sought to identify which *single* descriptor used in this study is most highly correlated with q . If no single descriptor could be identified that gave an exceptionally high correlation (i.e., an LFER) then multidescrptor models would be developed until a near-perfect fit could be realized (i.e., a QSAR).⁶⁵ The possibility of an LFER (i.e., single descriptor) was addressed by generating a database of all straight-chain quaternary ammonium ions with 4–40 carbons ($n = 715$) and then calculating the same descriptors for them as for the ammonium ions investigated in this study. The solvent accessible surface area of the ammonium ion center (NC4_SA) was the descriptor most highly correlated with q (Figure 2a, $R^2 = 0.889$).⁶⁶ To improve the correlation, two component models that accounted for all of the variance in q were generated using a genetic algorithm in combination with multilinear regression (GA-MLR) (Figure 2b). As shown in Figure 2c, the best resultant model consisted of $\log P(o/w)$ and the van der Waals surface area bearing a partial positive charge ($\delta + \text{SA}$). Surprisingly, this model does not contain NC4_SA despite the fact that it appears with the highest frequency in the two-component models (dark blue bar in Figure 2b). At this point in the model development methodology schema nonlinear correlations with q (the example “experimental data”) would be sought out. It is observed that a parabolic fit of q as a function of NC4_SA has a higher correlation coefficient than the linear fit (see the gray bars in Figure 2a).

The QSAR model development strategy employed for the catalyst activity data set is summarized below. The method introduces no bias for any particular descriptor a priori and no intervention by the practitioner is required during the model screening process. To facilitate the survey and comparison of

large numbers of descriptors and models in short order, a genetic algorithm was utilized. For example, in the investigation of q described above, a GA-MLR allowed for the rapid comparison of 3×10^6 linear models from a population of 200 models over 15000 generations. The process consists of iterative application of four basic steps: (1) inspection of linear single descriptor correlations (Figure 2a), (2) screening of linear QSAR models (combinations of 2 or more descriptors) by application of a GA-MLR algorithm (follow Figure 2a to Figure 2b), (3) inspection of both the frequency of descriptor inclusion in “good” models (Figure 2b) as well as the resultant models (Figure 2c), and last, (4) comparison of the higher dimensional models (more descriptors) to the lower dimensional ones (follow Figure 2c to Figure 2a). Typically, for the purposes of this study, the descriptor/model surveys consisted of two to three iterations of the process.

To summarize, the high correlation of q and the solvent accessible surface area of the ammonium ion center (NC4_SA) confirms that q is a good reflection of accessibility of the ammonium ion. Ammonium ion accessibility is also highly correlated to a combination of solubility and partial charge exposure, which has important mechanistic implications (vide infra). Lastly, and perhaps most importantly, nonlinear relationships often remain hidden until manual inspection of the model and residuals. Nonlinear correlations will be presented and discussed on a case by case basis below.

2.1. Investigation of LFERs. The next questions to be addressed were as follows: (1) Is a multidimensional QSAR necessary or is a one-dimensional QSAR (LFER) possible? (2) Would an ammonium ion accessibility descriptor sufficiently account for all of the catalyst activity in *this* data set? As before (vide supra), inspection of a correlation matrix of catalyst activity and each descriptor initiated the analysis. In no case was a single descriptor found that was highly correlated ($R^2 > 0.8$) to catalyst activity. However, it is useful to identify descriptors that exhibit the strongest linear correlations to catalyst activity of the total pool of descriptors investigated. A representative selection of single descriptor correlations is summarized in Table 4 along with their associated regression statistics ($n = 102$, R^2 , q^2_{LOO} , RMSE, and F).

The most highly, linearly correlated descriptors can be categorized into three groups, namely (1) those pertaining to

Table 4. Representative Summary of Descriptors Linearly Correlated to Catalyst Activity

descriptor	description	correlation	R^{2a}	$q^2_{\text{LOO}}^b$	RMSE ^c	F^d
vsurf_D1	hydrophobic volume	+	0.413	0.400	1.17	74.41
vsurf_W1	hydrophilic volume	+	0.381	0.370	1.20	66.95
vsurf_WP1	polar volume	+	0.466	0.460	1.11	87.97
ASA	accessible surface area	+	0.373	0.368	1.20	65.81
RA_2D_PIP2	Politzer ionization potential	-	0.400	0.371	1.20	67.88
elstat_min	minimum electrostatic potential	-	0.366	0.337	1.21	57.73
Aq_Solv_E	Aqueous solvation energy	+	0.419	0.402	1.16	74.22
clogP(o/w)	Partition coefficient	+	0.221	0.185	1.34	29.00
TPSA	Total polar surface area	+	0.070	0.045	1.47	9.50

^aSquare of the correlation coefficient. ^bSquare of the leave-one-out cross-validated correlation coefficient. ^cRoot mean square error. ^dFischer number.

nonpolar surface area/volume, (2) polar surface area/volume, and (3) electrostatic interactions. The three descriptors with the highest positive linear correlations are polar volume, hydrophobic volume, and hydrophilic volume (vsurf_WP1, vsurf_D1, and vsurf_W1 respectively). Similarly, the water-accessible surface area (ASA) is correlated positively with catalyst activity. Many descriptors related to catalyst polarity (electrostatic interactions) were found to be negatively correlated to catalysts activity. Two representative descriptors include the Politzer ionization potential (RA_2D_PIP2) and the minimum electrostatic potential (elstat_min, most negative or electron repelling energy). Two descriptors that are not correlated with catalyst activity are the calculated octanol/water partition coefficient (clogP(o/w))⁶⁷ and total polar surface area (TPSA).^{68,69} The high frequency of correlation observed between multiple subdivided surface area/volume descriptors was expected as these are useful in prediction of pharmacokinetic properties.⁷⁰ Similarly, interactions between an ammonium catalyst and an anion are necessarily noncovalent; thus electrostatic terms are chemically and physically justified. Thus, prior to investigation of a multicomponent QSAR it was established that no single descriptor could account for all of the catalyst activities. The question therefore remained, how many descriptors are necessary to account for the variation in catalyst activity?

2.3. Multidimensional QSAR Analysis

2.3.1. Establishing an Upper Limit for the Number of Descriptors per Model. It has been proposed that six descriptors are required to model neutral solute behavior and seven are required to model ionic solute behavior in a biphasic system.⁷¹ Multicomponent QSAR model development was initiated by dividing the database into two sets of descriptors (3D and 2D) and developing models with a variable number of components (1–10). The 3D descriptor set performed slightly better than did 2D descriptor models with fewer than five components.⁵⁵ When more than five components were included, 3D and 2D models exhibited similar performance. Inclusion of more than seven descriptors in a model did not lead to a significantly better fit, completely consistent with observations for many phase transfer related processes.⁷¹ Therefore, it was decided to limit further model development to combinations of seven descriptors or less.

The 2D descriptors that were most frequently included (i.e., “survived” the evolution) were (1) number of rotatable bonds, (2) molar refractivity, (3) clogP(o/w), (4) molecular volume, and (5) molecular weight as well as various descriptors for partial charge distribution and electrostatic potential interaction energy.

The 3D descriptors that were included most frequently were (1) molecular dipole, (2) cross-sectional area (XSA),⁷² (3) ionization potential, and various descriptors encoding electrostatic potential interaction energies.⁵¹

2.3.2. Models with Descriptor Subsets. In addition to comparing 3D and 2D descriptor set models, comparison of models derived from different descriptor classes is informative. Table 5 is a summary of models with variable numbers of components based on VolSurf,⁷³ electrostatic surface area,⁵¹ charged surface area,⁷⁴ inductive,⁵³ and SMR/SlogP_VSA descriptor classes. For

Table 5. Comparison of Models Based on Various Descriptor Classes

descriptor class	no. of descriptors	R^2	$q^2_{\text{LOO}}^a$	RMSE ^b	F^c	
inductive	2	0.461	0.412	1.15	43.23	
	4	0.583	0.543	1.01	34.65	
	5	0.613	0.537	0.98	31.07	
	6	0.635	0.552	0.95	28.07	
	7	0.648	0.564	0.93	25.23	
	δ^\pm SA	2	0.483	0.458	1.09	46.16
		4	0.561	0.524	1.00	30.93
5		0.591	0.546	0.97	27.70	
6		0.618	0.562	0.94	25.60	
7		0.642	0.569	0.91	24.07	
volsurf	2	0.589	0.567	1.01	72.33	
	4	0.655	0.610	0.92	47.02	
	5	0.686	0.641	0.88	42.76	
	6	0.707	0.664	0.85	38.95	
	7	0.721	0.674	0.83	35.48	
	SMR & SlogP	2	0.541	0.518	1.06	59.55
		4	0.671	0.651	0.90	50.57
5		0.703	0.666	0.86	46.35	
6		0.726	0.685	0.82	42.84	
7		0.746	0.692	0.79	40.32	
electrostatic surface area		2	0.550	0.538	1.05	61.67
		4	0.710	0.604	0.85	60.54
	5	0.746	0.712	0.79	57.48	
	6	0.786	0.635	0.73	59.41	
	7	0.808	0.782	0.69	57.75	

^aAverage square of the correlation coefficient after leaving one data point out. ^bSquare of the leave-one-out cross-validated correlation coefficient. ^cFischer number.

this data set, the lowest correlations were found with models based solely on partial charge distribution (inductive or surface area). None of the models based on inductive effects or partially charged surface areas had R^2 greater than 0.7. Also, in both of these cases, as the number of descriptors in the model was increased the variance in the data that was accounted for decreased at a greater rate than for other descriptor sets. The VolSurf and SMR/SlogP descriptor sets performed well, generating models with $R^2 > 0.7$ and similar errors (RMSE and F). The electrostatic descriptor set generated the best models with the smallest deviations.

Many of the models in Table 5 are quite good by QSAR standards⁹ in that they exhibit a high fit of the experimental data, small error, and a high data to descriptor ratio.⁷⁵ Each of the most successful descriptor sets share a common feature; the independent variable inputs consist of descriptor values that are “binned”. Binning⁷⁶ descriptor values by surface area or volume is an extremely useful QSAR method capable of separating out the molecular surface area and volume properties responsible for an observed effect from noise.^{9f,g} The process is analogous to compressing an image where the important contrasts are retained and the unimportant discarded.

The disadvantage of such an analysis is that deconvoluting the resultant property values and correlating them with a particular structural change can be challenging. That is, the interpretability of the resultant QSAR model is often difficult.

For this reason, a final series of models was sought that would provide a higher degree of interpretability. Interestingly, $\text{clog}P_{(o/w)}$ was not identified as a useful descriptor by randomly screening and comparing linear models, which certainly warrants further investigation.

2.4. Systematic Investigation of $\text{clog}P$ and XSA

2.4.1. Comparison of $\text{clog}P_{(o/w)}$ and $\text{clog}P_{(b/w)}$. The relevance of $\text{clog}P_{(o/w)}$ for the catalysts in this study was evaluated by comparing the calculated octanol/water partition values⁷⁷ to calculated benzene/water partition values that are known to be very accurate for charged organic solutes (Figure 3a).⁴¹ A number of systematic deviations are immediately apparent. For most of the data set, a good linear correlation is seen (blue circle, 81 of 102). However, even for these cases, there is significant systematic deviation indicated by the slope of ~ 0.6 (not 1.0), which reflects the greater solvating ability of octanol than benzene. Also, catalysts bearing strongly electron-withdrawing R^4 substituents (Figure 3a, red triangle, 18 of 102) make up another group of catalysts with systematic deviation and are significantly less lipophilic in the benzene/water system than an octanol/water partition would predict. For catalysts bearing two strongly electron-withdrawing substituents (blue diamond, Figure 3a) the effect is multiplied. Ammonium ions with *N*-methyl substituents showed similar deviation (orange square, Figure 3a), but to a lesser extent. Taken together, these results indicate that the dominant physical origin of the difference between the two calculated values is an increased electrostatic contribution to the partitioning of ammonium ions in benzene/water in comparison to octanol/water.

Plotting catalyst activity versus $\text{clog}P_{(b/w)}$ partition reveals a potential parabolic relationship (blue circle, Figure 3b). Roughly 75% of the data set conforms to this fit, but the remaining 25% does not (red \times). The ammonium ions whose activity deviated the most from the parabolic fit were those with *N*-methyl groups (e.g., $\text{cetylMe}_3\text{N}^+$), the parent cyclopenta[*gh*]pyrrolizidinium ion ($R^1, R^2 = \text{H}$ and no β -oxygen), and those containing a β -hydroxy group ($R^1, R^2, R^3 = \text{H}$). The fit indicates that a

maximum catalyst activity is observed when $\text{clog}P_{(b/w)}$ is between -3 and -1 . Therefore, this partially parametrized parabolic fit was included in another round of model screening. The effect of model screening in this way can be visualized as scalar multiples of the partially parametrized parabolic fit (gray lines, Figure 3b). To this end, the most successful descriptor sets and singly, highly correlated descriptors were compiled (see Results sections 2.3.1 and 2.3.2) and allowed to compete in another round of evolution, inclusive of nonlinear descriptor operations (e.g., square, inverse, log, etc.).

The best models resulted when either $\text{clog}P_{(b/w)}$ or XSA were forcibly included and are summarized in Table 6. Overall, a parabolic fitting of $\text{clog}P_{(b/w)}$ generated models with similar performance as observed for $\log(\text{XSA})$. Within the $\text{clog}P_{(b/w)}$ -based models, the most frequently encountered descriptors were those dealing with motion, specifically the standard dimensions (std_dim_n)⁷⁸ and principle moments of inertia (pmi). A variety of partial charge descriptors, such as the surface area of the ammonium ion and other subdivided surfaces were also included. Most interestingly, the best model with three descriptors was $-\text{XSA}^2 + \text{XSA}$. Overall, the $\text{clog}P_{(b/w)}$ models had a significantly decreased RMSE, about half that of the XSA models, but the overall fit decreased only slightly in all cases relative to the XSA models.

Table 6. Comparison of XSA, $\log(\text{XSA})$, and $\text{clog}P_{(b/w)}$ Models

dominant ^a descriptor	descriptors	R^2	q^2_{LOO} ^b	RMSE ^c	F^d
$-\text{aclog}P_{(b/w)}^2 + \text{bclog}P_{(b/w)}$	2	0.564	0.539	0.44	63.90
	3	0.635	0.626	0.40	56.76
	4	0.702	0.690	0.36	57.03
	5	0.734	0.718	0.34	53.10
	6	0.779	0.754	0.31	55.86
	7	0.791	0.765	0.302	50.79
	XSA	2	0.646	0.627	0.90
3		0.688	0.667	0.85	72.19
4		0.723	0.690	0.80	63.29
5		0.759	0.721	0.75	60.36
6		0.780	0.750	0.71	56.03
7		0.803	0.772	0.68	54.77
$\log(\text{XSA})$		2	0.629	0.614	0.93
	3	0.751	0.732	0.76	98.68
	4	0.781	0.753	0.71	86.71
	5	0.799	0.764	0.68	76.36
	6	0.824	0.799	0.64	73.97
	7	0.845	0.759	0.60	73.39

^a Here, $\text{aclog}P^2 + \text{bclog}P$ from the partially parametrized parabolic fit above is treated as a single descriptor. ^b Square of the leave-one-out cross-validated correlation coefficient. ^c Root mean square error. ^d Fischer number.

In general, it was discovered that taking the log of the cross-sectional area (XSA) resulted in better models than using the native cross-sectional area. In these cases, the remaining descriptors were almost exclusively composed of descriptors encoding electrostatic interactions.⁵¹ The high correlation of models containing cross-sectional area or the log of the cross sectional area provided the impetus for a more detailed investigation of this descriptor.⁷⁹

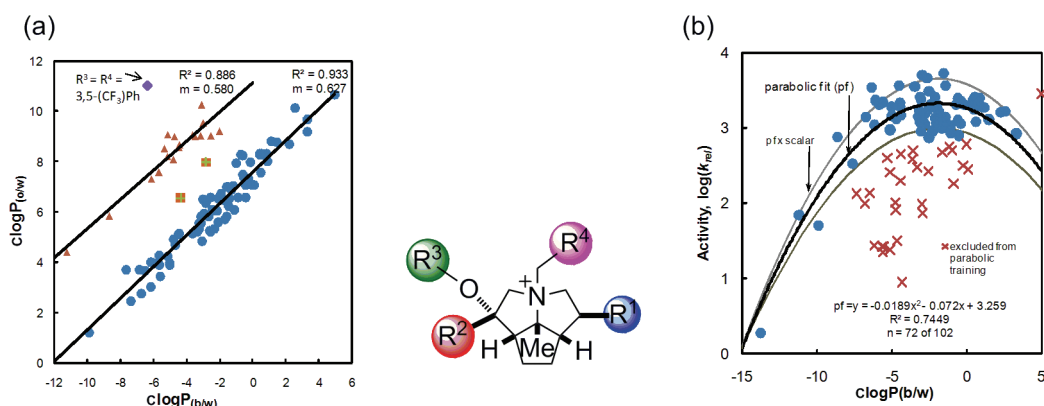


Figure 3. (a) Comparison of thermodynamic ammonium ion partitioning values; $\log P_{(o/w)}$ and $\log P_{(b/w)}$. (b) Possible parabolic relationship between catalyst activity and $\log P_{(b/w)}$.

The cross-sectional area descriptor has been developed specifically to relate the surface activity of amphiphilic molecules.⁷² After correcting for conformational differences to resemble a surface bound ammonium ion, it was found that cross-sectional area alone was sufficient to account for most of the differences in catalyst activity (Figure 4a). A nonlinear correlation with XSA was observed, similar to the observed maximum for $\log P_{(b/w)}$ partition coefficient (Figure 3b). In the case of XSA, the optimal value for catalyst activity is between 80 and 100 Å². Thus, the model chosen for further analysis is (Figure 4b):

$$\begin{aligned} \log(k_{rel}) = & +0.079(-0.0189 \times \log P_{(b/w)}^2 - 0.072 \\ & \times \log P_{(b/w)} + 3.259) + 0.833(-0.0004 \\ & \times XSA^2 + 0.0824 \times XSA - 1.154) \\ & - 0.46778 \end{aligned} \quad (1)$$

Although higher regression coefficients and overall fits could be obtained with more complicated models, the high correlation (Figure 4b) with only two descriptors and high degree of physical interpretability favors the use of this model.

DISCUSSION

1. Catalyst Enantioselectivity. *1.1. Conformations from CoMFA Modeling.* As noted previously, the catalysts in this study exhibit two primary conformational preferences manifested as the **up** and **dn** conformers (Table 1). The distribution of the conformers in a given library significantly impacts the integrity of the CoMFA models as illustrated in Table 2. These conformational preferences in the CoMFA models necessitate further analysis and physical justification.

The underlying hypothesis is that the preference for the catalyst to exist in the **up** or **dn** conformation depends largely on the face of the catalyst to which the enolate associates. As illustrated in the accompanying paper (DOI: 10.1021/jo2005445), the oxygen substituent at C(1) polarizes a larger amount of positive potential to the face over ring **a** relative to the face over ring **b** of the catalyst scaffold. Thus, association of the enolate with the face over ring **a** would be expected with small groups as the R^2 substituent (Table 1). If R^2 is isopropyl or *tert*-butyl, it may be expected that the enolate would associate preferentially with the face over ring **b** of the catalyst as less positive potential is “screened” by a methyl group (at R^1) relative

to the isopropyl or *tert*-butyl groups (at R^2). Association with the face over ring **b** would require the catalyst to exist in the **dn** conformation to maximize exposed positive potential and hence the electrostatic interaction. The conformational preference (**dn** or **up**) for catalysts with R^2 not equal to isopropyl or *tert*-butyl does not have a strong electrostatic component and may primarily be dictated by the energy intrinsic to the scaffold geometry. These energy differences vary depending on R^1 , but it is not expected to greatly influence the enantioselectivity.

When taking the reaction medium into consideration, it is unlikely that the strength of the interaction with the enolate (when R^2 is not equal to isopropyl or *tert*-butyl) would be affected by the conformation of ring **b** significantly if the catalyst is in the toluene layer. On the other hand, if the catalyst is present in the aqueous or interfacial layer, it is probable that the catalyst would prefer to adopt the **dn** conformation to enable partial solvation with water molecules. However, of primary concern is the thermodynamic preference of the ammonium enolate in the toluene layer as that is primarily the medium where the intrinsic alkylation step (and also the stereodetermining step) is believed to occur.

The complication of unknown reactive conformations may be alleviated by including a distribution of both conformers weighted by their corresponding energies. However, this approach would require modeling in the presence of an anion as the energies would not reflect the desired Boltzmann distributions in the absence of an anion (since the **dn** conformation would be largely preferred if the enolate is associating with the face over ring **b** which is presumed to be occurring for catalysts with $R^2 = i\text{-Pr, } t\text{-Bu}$), which would be outside the computational rigor intended for this study.

1.2. Contour Maps. The CoMFA models may be physically interpreted in the form of contours encompassing the catalysts. The most common method for illustrating the contours are as products of the standard deviation of the interaction action energies and the coefficients in the PLS model at each grid point (StDev*Coeff). This product represents regions where and how the variation in the interaction energies can be explained by the variation in the enantioselectivity. The contours produced can be physically interpreted as spatial regions encompassing the catalyst where an increase or decrease in steric bulk (green for increase, yellow for decrease) and positive charge (blue for increase, red for decrease) leads to an increase and decrease in enantioselectivity respectively. Both the steric and electrostatic

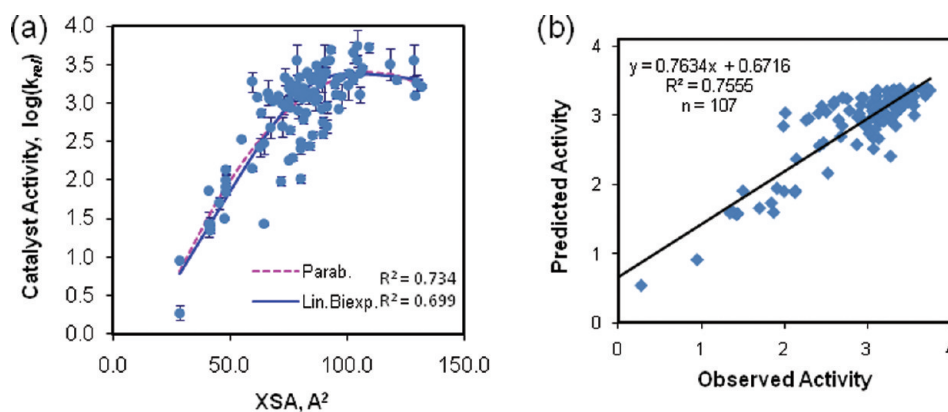


Figure 4. (a) Comparison of a parabolic and bilinear correlation between catalyst activity and catalyst cross-sectional area. (b) Plot of the predicted catalyst activity versus the observed catalyst activity for a double parabolic QSAR model with $\log P_{(b/w)}$ and XSA.

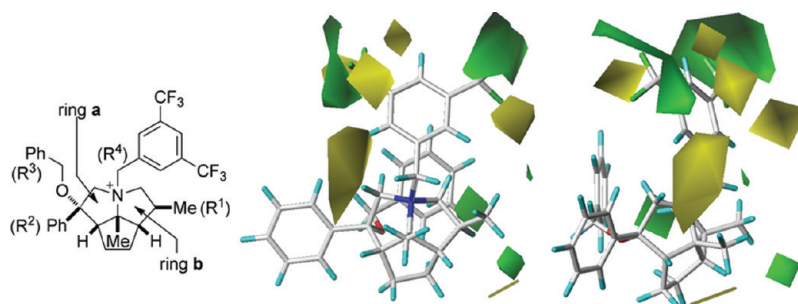


Figure 5. Steric contour maps from two different perspectives. Green contours indicate regions where increased steric bulk leads to increased enantioselectivity, while yellow contours indicate regions where decreased steric bulk leads to increased enantioselectivity.

contour maps derived from indicator fields (boxed model in Table 3) are illustrated below (Figures 5 and 6). It is noteworthy that contour maps from models derived from Tripos standard fields display information that is similar to the contour maps derived from indicator-based fields below.

1.2.1. Steric Contour Maps. Upon inspection of the steric contour maps (Figure 5), the relationship between the substitution pattern and the contours becomes immediately apparent. A catalyst among the most selective is illustrated to serve as a template for interpreting how the contours are manifested. Two 3-D representations of different perspectives are illustrated for gauging relative depth. The location of the contours will be discussed with respect to the perspective of the left illustration in which the 5–5–5 scaffold is presented. The green contour on the front of ring **b** overlaps with the methyl group which is necessary for higher selectivity. The data is consistent with the hypothesis that this methyl group is necessary to shield the right front of ring **b** from anion association. The green contours overlapping with the substituents at the 3 and 5 positions of the phenyl ring of the benzyl group attached to the nitrogen atom are consistent with increased enantioselectivity with substitution at those positions. Only *tert*-butyl and trifluoromethyl groups were explored with latter groups bestowing greater enantioselectivity. The green contour in the back, overlapping with the phenyl ring of the benzyl group attached to the oxygen atom, is an indirect indicator of the presence of a non-hydrogen group at the R^2 substituent. If the R^2 substituent is hydrogen, the benzyl group attached to the oxygen atom occupies the region of space

after rotation of $\sim 90^\circ$ counterclockwise about the O–CH₂Ph bond with respect to the conformation adopted by the corresponding catalyst with $R^2 \neq$ hydrogen.⁵⁶ All catalysts possessing a hydrogen atom at R^2 exhibit poor enantioselectivity in which case the phenyl ring would not overlap with the green contour of interest. The yellow contours surrounding the group attached to the nitrogen atom (R^4) reflect substitution patterns (1-naphthyl, hexyl, and 9-anthracenyl groups) unfavorable for enantioselectivity. The large yellow contour over ring **A** may reflect both unfavorable interactions with the 9-anthracenyl substituent as well as bulky aliphatic groups at the R^2 position (*i*-Pr and *t*-Bu). The effect of this steric contour will be further rationalized in conjunction with the electrostatic contours as discussed below.

1.2.2. Electrostatic Contour Maps. Although the steric contours enabled the rationalization of the direct influence of specific substitution patterns, the electrostatic contours tend to reveal more indirect information relating to stereocontrolling features including differential binding affinities. After eliminating the grid points with minimal field variation (1.3 kcal/mol column filtering), only two primary contours remain. The presence of a large blue contour over ring **a** suggests the potential for this region to serve as a reasonable binding site for the reactive enolate. This conclusion is consistent with the fact that the largest degree of positive potential is localized over this ring in accordance with electrostatic potential (ESP) maps determined using *ab initio* theory.⁸⁰ The blue contour may additionally reflect the enantioselectivity enhancing effects of an aryl group at the R^1 position as the quadrupole moment of arenes results in a region

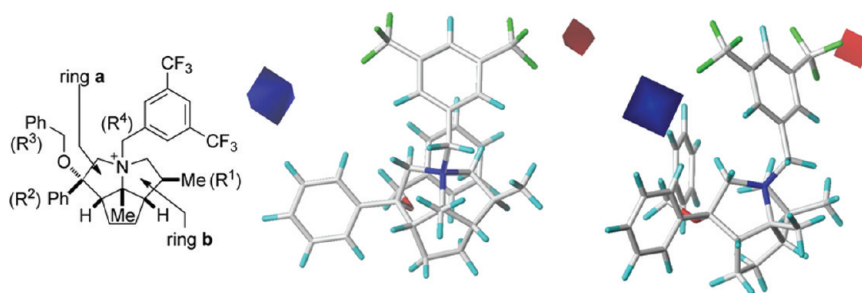


Figure 6. Electrostatic contour maps from two different perspectives. Blue contours indicate regions where increased positive charge leads to increased enantioselectivity, while red contours indicate regions where decreased positive charge (increased negative charge) leads to increased enantioselectivity.

of positive potential along the circumference of the ring and may serve as an indicator for the regression analysis.⁸¹ The ESP partial charges should reflect this region of increased positive potential (relative to catalysts with $R^2 \neq$ arene) which is thus detected by the indicator fields. The red contour is coincident with a trifluoromethyl group on the right side and is consonant with the enantioselectivity enhancing effects of this group in conjunction with previously mentioned substitution requirements necessary for enantioselectivity. Interestingly, a red contour does not overlap with the trifluoromethyl group on the left-hand side, which may be a result of the nearby blue contour and it having a larger contribution to the model. The absence of a contour in this position may suggest the undesirable negative electrostatic potential and that the enantioselectivity might be enhanced in its absence. These two electrostatic contours may also reveal an intrinsic dipole preference of the catalyst for enantioselectivity. The yellow steric contour on the front of ring a in Figure 5 is complementary to the electrostatic contours in that increased positive potential and decreased steric potential should lead to a more favorable binding interaction with an associating anion. The combined effects of the steric and electrostatic contours are supportive of the conclusion that the most favorable binding site for increased enantioselectivity is that on the convex side of ring a for an associating enolate.

1.3. Stereochemical Analysis of QSSR Model. The information revealed by considering both the steric and electrostatic fields allows some generalizations to be drawn regarding the mode with which the enolate should associate with the catalyst that leads to increased enantioselectivity. The blue contour coincident with the R^2 group (Figure 6), the green contour coincident with the R^1 substituent (Figure 5), and the large yellow contour on the front of ring a of the catalyst (Figure 5) are all consistent with the preferred region of the catalyst for association that leads to increased enantioselectivity to be the front of ring a of the catalyst scaffold. The red contour (Figure 6) may enforce a preferred dipole orientation of the catalyst for aligning the enolate.

The observed enantioselectivity may be explained by considering one of two limiting scenarios:⁸² (1) the binding of the enolate to the catalyst is completely selective for one of the four faces of the imaginary tetrahedron inscribing the ammonium nitrogen and the enantioselectivity is directly related to the capacity of the face of the catalyst in question to differentiate the *Re* or *Si* faces of the enolate or (2) the capacity with which the catalyst differentiates the *Re* or *Si* faces of the enolate with respect to the faces of the ammonium ion is maximal (each face selects either *Re* or *Si* completely), and the enantioselectivity is directly related to the binding selectivity of the enolate to one of the four faces of the catalyst. A combination of (1) and (2) is likely

operative; however, the observation that bulky aliphatic groups at R^2 result in a reversal in enantioselectivity suggests that a large portion of the observed enantioselectivity is governed by limiting scenario (2).

Assuming that limiting scenario (2) is primarily operative, some generalizations about relative binding affinities and enantioselectivity can be drawn. Both the electrostatic and steric contours are then consistent with the relative association preferences outlined in Figure 7. Because of the polarization effects of the oxygen substituent, association mode (A) is preferred in structures of type 5 and to a lesser extent for 1, 2, and 3 (preferred pathway is indicated by a larger arrow). When the face over ring a is significantly sterically hindered, as in structures of type 4, association mode (B) is preferred even though the electrostatic preference favors association mode (A). The ordering outlined in Figure 7 is most consistent with the enantiomeric ratios when R^4 is significantly electron withdrawing such as a bis-(trifluoromethyl)phenyl group. Presumably, this group serves to create a stronger electrostatic interaction (decreased separation distance) with the enolate which enhances the steric influence of the R^1 and R^2 substituents with respect to catalysts with less electron-withdrawing groups at R^4 . The stronger interaction for catalysts with $R^2 =$ aryl is not likely a consequence of increased steric or electrostatic (with respect to the electrostatic potential associated with the ammonium nitrogen atom) interactions but may be attributed to the intervention of π - π interactions. An alternative to this analysis (limiting scenario (1)), would be that the erosion of enantioselectivity may be attributed to decreased enantiotopic differentiation of the faces of ring a or ring b of the

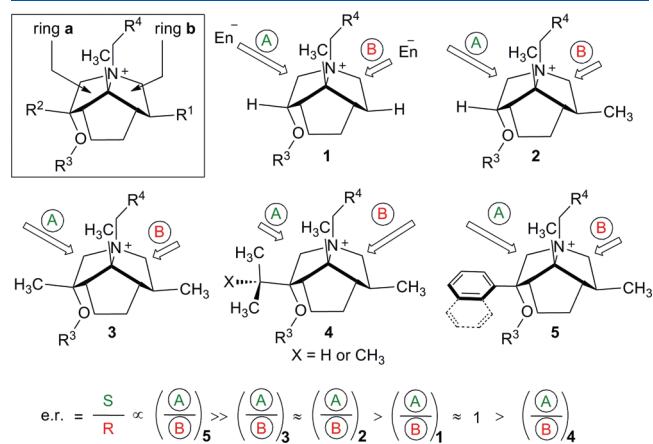


Figure 7. Association preferences for the enolate (En^-) as a function of the R^1 and R^2 groups.

catalyst for the *Re* and *Si* faces of the enolate which leads to a more challenging stereochemical rationalization.

Although the highest enantioselectivity observed in this report (*er* = 81:19) does not currently allow for the development of a well-defined stereochemical model for this catalyst system, the contour maps do allow the generation of *qualitative* assessments of how to increase the enantioselectivity further by introducing substituents that render the interactions noted above even more favorable. The model may also be used to make *semiquantitative* predictions regarding the extent of enantioselectivity to which certain substituents may provide. While such extrapolation can be susceptible to erroneous predictions, some confidence may be placed in the predictions if the catalysts identified to be more selective are proximal to the structural and dependent variable domain of the training set. For example, this model would be incapable of predicting enantioselectivities imparted by cinchona derived quaternary ammonium ions and any catalysts that furnish enantioselectivities significantly greater than 81:19.

1.4. Summary of QSSR Analysis of Enantioselectivity. The observed variation in enantioselectivity could be explained by analysis of the electrostatic and steric environments of the catalysts through CoMFA modeling. The contour maps (graphical representations of the model) may be interpreted as relating to the association preferences of the enolate for the catalyst. The relationship is more evident when the electrostatic attraction is strongest which is the case when R⁴ is a bis-(trifluoromethyl)phenyl group (Figure 7). This attraction enhances the influence (whether it be steric or electrostatic) of the R¹ and R² substituents. The determination of catalyst·enolate interaction energies through *ab initio* molecular modeling might reveal a more convincing quantitative relationship between *er* and the relative association strengths of the two most exposed catalyst faces but is beyond the scope of this report.

The results from the study on enantioselectivity suggest some design considerations for designing future catalysts for this system and may be extrapolated to underdeveloped APTC systems. The design criteria can be broken into steric and electrostatic (which is further decomposed in a truncated multipole expansion).

Electrostatic:

- monopole: not variable
- dipole: enforced by strongly polarizing groups proximal to the quaternary nitrogen (e.g., R⁴ = bis(trifluoromethyl)phenyl group; β -alkoxy group)
- quadrupole (locally):⁸³ may comprise of π - π interactions in the context of the system studied in this investigation (e.g., R² = aryl)

Steric:

- addition of steric bulk for three of the four faces of (R¹ = Me)
- removal of steric bulk for one of the four faces (R² \neq *i*-Pr, *t*-Bu)

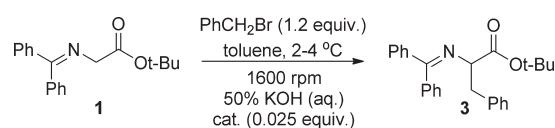
Either electrostatic or steric interactions may be predominant depending on the nature of the structural variation in the training set. For the data set investigated in this study, the contributions of electrostatic and steric interactions to the variation in the observed enantioselectivity are 45% and 55%, respectively. All of the aforementioned principles are necessary to account for the relatively elevated enantioselectivities imparted by catalysts of type 5 in Figure 7. Although these principles might seem apparent from chemical intuition and a moderate understanding

of intermolecular interactions, the application of this analysis is rarely encountered in APTC catalyst design, particularly for all the aforementioned interactions.

2. Catalyst Activity. At the final stage of any QSAR endeavor, the same fundamental questions must be answered, namely, do the statistically derived relationships have any meaning, and, if so, what? Given the statistical labyrinth that defines the modus operandi of QSAR analyses, the pitfall of associating correlation with causation is often encountered.^{84,85} One of the most common sources of this problem is to search for trends in large data sets while ignoring trends in the chemotype subpopulations.⁸⁶ For this reason, the following discussion is divided into two sections, one that involves the entire data set and the other that compares library subpopulations with common chemotypes. The two descriptor types common to most good models for this data set were molecular cross-sectional areas and those derived from electrostatic potential maps. The differences in electrostatic potential maps of the catalysts were discussed in detail in the preceding paper.⁸⁰ The most intuitive model developed was based on two descriptors, namely XSA and $\log P_{(b/w)}$. Therefore, the following discussion will focus on the differences and physical interpretation of the cross-sectional areas and $\log P_{(b/w)}$ of the catalysts.

Prior to a detailed discussion and interpretation of the QSAR results, a brief review of the data collected would be helpful. The previous paper reported, in detail, the manner in which the data for this study was collected (Scheme 2).¹¹ Most importantly, the catalyst activity data set was collected at a single stir-rate (1600, rpm) under a standard set of conditions. Care was taken to collect data under conditions such that the reaction is stir rate dependent and with a highly active nucleophile (an enolate alkylation). *Therefore, under no circumstances should these data be interpreted to represent a rate-limiting alkylation step.*⁸⁷

Scheme 2



2.1. $\log P$. Because the basis of phase-transfer catalysis is extraction of a hydrophilic substrate anion into a lipophilic organic phase, a correlation of catalyst activity with organic phase solubility is perhaps not surprising. Indeed, it has long been appreciated that catalyst solubility is an important factor to consider. Early reports on PTC indicated that the catalyst activity was correlated to the partition coefficient *P*, most commonly represented as the ratio of the organic and aqueous phase concentrations (Figure 8).^{17,88} Also, the number of carbons in a catalyst has been proposed as a useful predictor of catalyst activity.^{18,89} Studies in medicinal chemistry have shown time and time again that the contribution of a methylene unit to $\log P_{(o/w)}$ is additive for homologous series like quaternary ammonium ions.^{4c} In other words, phase-transfer catalyst activity is correlated to the number of carbons because $\log P$ is linearly related to the number of carbons in an unfunctionalized ammonium ion.⁹⁰

Partition coefficients have been thoroughly studied from both a thermodynamic and kinetic perspective because they facilitate the description (and prediction) of the pharmacokinetic

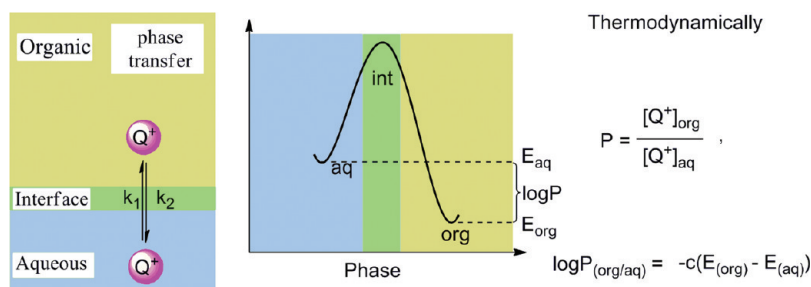


Figure 8. Thermodynamic representation of the partition coefficient ($\log P$).

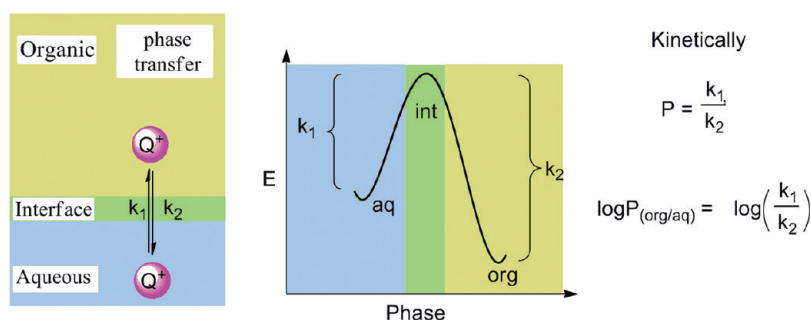


Figure 9. Kinetic representation of the partition coefficient ($\log P$).

behavior of drugs. Thermodynamically, the distribution of a solute between two immiscible phases can be represented as the difference in the solvation energy of the solute in the two phases (Figure 8).⁹¹ Because $\Delta G = -RT \ln(K)$, this energy difference is proportional to $\log P$ and the coefficient c contains the constant R as well as the absolute temperature, T . Interpretation of PTC in terms of a static, thermodynamic partition coefficient has been done and leads to conclusions such as “the effectiveness of a phase-transfer catalyst depends mainly on its organophilicity.”⁹² Although true, in part, a number of drawbacks to this representation and subsequent interpretation are readily apparent. For example, a thermodynamic explanation does not necessarily indicate *why* the correlation is observed. The typical interpretation is that as the lipophilicity of the catalyst increases, the concentration of the active substrate in the organic phase increases with it. Considered in the context of catalysis, where the observed activity is necessarily a balance of multiple rates, this interpretation is unsettling because it leads to the incorrect conclusion that the best catalysts should exhibit both infinite organic solubility and water insolubility.

Although a kinetic interpretation of the partition coefficient is much less common in the PTC literature, it is far more valuable (Figure 9) for understanding the catalytic cycle of PTC.⁹³ Thus, the thermodynamic partition coefficient is expressed as the ratio of the forward (k_1) and reverse (k_2) rates of the catalyst traversing the interfacial region.⁹⁴ This type of analysis allows the application of transition-state theory and is known to be an accurate representation, even for ionic solutes (e.g., ammonium ions).⁹⁵ In this interpretation, the two solvated states (organic or aqueous) are energetic wells and should be observable intermediates on the catalytic cycle. Also, a catalyst traversing the interphase is in a third state, namely a transition state, which has an infinitesimal population during the catalytic cycle. Or more succinctly, this interpretation is consistent with the extraction mechanism of PTC.⁹⁶

A number of questions can be quantitatively addressed following this line of reasoning. For example, theoretically, at what $\log P$ should the greatest catalytic rate of phase transfer (sum of k_1 and k_2) be observed? Application of the Hammond postulate reveals that, theoretically, the greatest rate of phase transfer should be observed when $k_1 = k_2$, corresponding to a partition coefficient of 1 ($\log P = 0$). That is, a maximum PTC rate should be observed when the transfer of the catalyst is energetically neutral, corresponding to a perfectly symmetric transition state (neither early nor late). What is observed in this data set is a maximum catalyst activity when the calculated $\log P_{(b/w)} = -1.90$ (Figure 3) out of a range from -15 to 5 . Given the known high precision of the SM8 solvation model, the deviation from the theoretical is most likely the result of calculating $\log P$ with water as the aqueous phase, rather than the strong electrolyte solution (50% KOH in water) present under the reaction conditions. There are a number of other less likely possibilities and they will not be elaborated here.⁹⁷

Thus, we conclude that that $\log P_{(org/aq)}$ serves as a useful predictor of phase transfer catalyst activity because it can be interpreted kinetically. It can then readily be inferred that a maximum in catalyst activity as a function of $\log P$ represents an “optimal” balance between catalyst lipophilicity and hydrophilicity (Figure 3b) or optimal rate of phase transfer. Catalysts with a negative $\log P$ (hydrophilic) will be rate limiting in an aqueous-to-organic phase transfer step (k_1), and catalysts with large $\log P$ (lipophilic) will be rate limited by an organic-to-aqueous phase transfer step (k_2). It seems likely that any phase-transfer catalysis reaction employing quaternary ammonium ions would be amenable to such an analysis as long as the dominant mechanism involved a direct aqueous-to-organic phase transfer (and vice versa). However, in the QSAR derived above, $\log P$ was not the dominant term and many catalysts were not well correlated to $\log P$ (Results, Section 2.4.1). We therefore propose that the

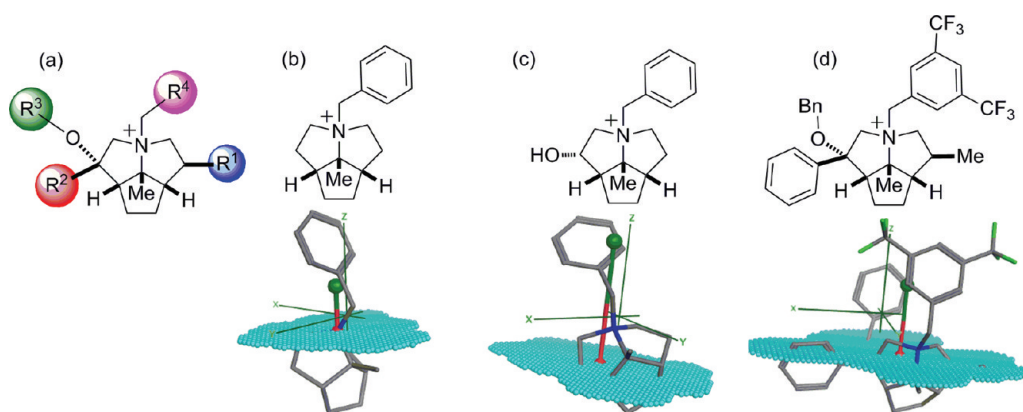


Figure 10. (a) Catalyst scaffold and substitution pattern included in this study. (b) Catalyst derived from the parent scaffold. (c) C(1) hydroxyl catalyst. (d) More highly adorned catalyst.

dominant mechanism does not involve a direct aqueous to organic phase transfer, but rather, interfacial adsorption and desorption type mechanism. The following section will consider this aspect of proposed PTC mechanisms on a more quantitative level.

2.2. XSA as a Function of Catalyst Substitution. The descriptor with the highest correlation to catalyst activity was found to be the cross-sectional area (XSA). This descriptor is defined as the cross-sectional area perpendicular to the amphiphilic axis and through the centroid of the polar atoms.⁷² Correlation of catalyst activity with the logarithm of the cross-sectional area was arrived by empirical screening. Ultimately, it was found that a parabolic relationship of catalyst activity with XSA gave the best description of catalyst activity (Results section 2.4 and Figure 4). Nonlinear correlations (parabolic,^{5,98} sigmoidal,⁹⁹ hyperbolic,¹⁰⁰ and bilinear¹⁰¹) of the phase-transfer rate of small molecules with polar surface area and lipophilicity are often observed. In this case, the catalyst activity levels off between XSA values of ~ 80 and 100 \AA^2 (Figure 4). Prior to analysis of the physical interpretation of catalyst XSA some appreciation for the effect of catalyst substitution on XSA is needed.

A representative selection of catalysts of increasing complexity and substitution is shown in Figure 10. Projected onto each of the catalysts are the Cartesian coordinate axes (green), the amphiphilic axis (red-green), and the cross-sectional area (teal). In each projection, the catalysts are oriented such that the amphiphilic axis is parallel with the z-coordinate axis and is depicted with red (polar centroid) and green (lipophilic centroid) spheres. The dependence of catalyst amphiphilicity on R^4 substitution is reflected by the direction and magnitude of the amphiphilic axis (compare Figure 10b, c, and d). In ammonium ions with no other heteroatom-derived functional groups, the XSA passes through the ammonium nitrogen (Figure 10b), which is the polar centroid. For catalysts derived from the cyclopenta[gh]pyrrolizidinium scaffold ($R^1, R^2, R^3 = \text{H}$) the resulting XSA is small ($\sim 40 \text{ \AA}^2$); comparable to the minimum possible area for an ammonium ion ($\text{XSA}(\text{Me}_4\text{N}^+) = 28.7 \text{ \AA}^2$). Functionalization of the ammonium scaffold with a hydroxyl group at C(1) (Figure 10c) causes the polar centroid to shift toward this group. The now-shifted XSA plane traverses a region of the scaffold with a greater circumference resulting in a small increase in XSA ($\sim 55 \text{ \AA}^2$). Addition of a non-hydrogen substituent at C(1) causes a further increase in XSA ($\sim 85 \text{ \AA}^2$, Figure 10d). The increase in XSA from no substituents at C(1) to hydroxy and alkoxy groups (up to $\sim 80 \text{ \AA}^2$) correlates with a proportional increase in catalyst activity. Further

substitution at C(1) increases the XSA of the catalyst but a proportional increase in catalyst activity is not observed (Figure 4a). A slight decrease in catalyst activity is observed for catalysts with a cross-sectional area significantly greater than 100 \AA^2 . Thus, the XSA of the catalysts reflects the dependence of catalyst activity on substituents at C(1), albeit in a nonlinear fashion. This correlation warrants a more quantitative investigation by comparing catalyst subsets, which is the subject of the next section.

2.3. Physical Interpretation of XSA. The cross-sectional area descriptor was developed specifically to correlate the thermodynamic tendency of a molecule to be in an adsorbed state at the air/water interface.^{102–104} Similarly, the tendency of a catalyst to be concentrated at the interfacial space is proposed to be a key feature for catalysts of hydroxide-initiated PTC reactions and is well represented as an adsorbed state.¹⁰⁵ The two processes, adsorption at an air/water interface and adsorption at an aqueous/organic interface, are fundamentally very similar in that they describe the thermodynamic distribution of an amphiphilic molecule in a highly anisotropic medium.¹⁰⁶ Concentration of an amphiphile at the air/water interface reflects the thermodynamic tendency of the amphiphile to be repelled from a highly polar medium (for water; $\epsilon = 80$) to a less polar one (for air; $\epsilon \sim 1$). Expression of the same fundamental behavior in the context of phase transfer catalysis, i.e., repulsion from a highly polar medium (50% KOH, $\epsilon > 80$) to a less polar one (for toluene $\epsilon \sim 2.4$), is shown graphically in Figure 11. In essence, this is the defining feature of the interfacial mechanism of PTC.¹⁰⁵

The only difference between this representation and one involving a direct aqueous to organic phase transfer (Figure 9) is the presence of an intermediate adsorbed state. That is, the adsorbed state is an energetic well rather than a transition state. However, it is not clear which of the elementary steps of the catalytic cycle are most influenced by the XSA of the catalysts. Interpreting the physical meaning of XSA in the context of a phase-transfer catalytic cycle requires a few basic assumptions as well as both kinetic and thermodynamic interpretations of $\text{dlog}P$. The following rationale suffices to do so with a minimal number of well-justified assumptions. This analysis begins with the experimental observation that the largest deviations in catalyst activity from what would be predicted by $\text{dlog}P_{(b/w)}$ alone occurred when $\text{dlog}P_{(b/w)}$ was at the optimum value (compare blue circle to red \times in Figure 3b). As previously explained, at the experimentally observed optimum $\text{dlog}P$ value, the catalytically

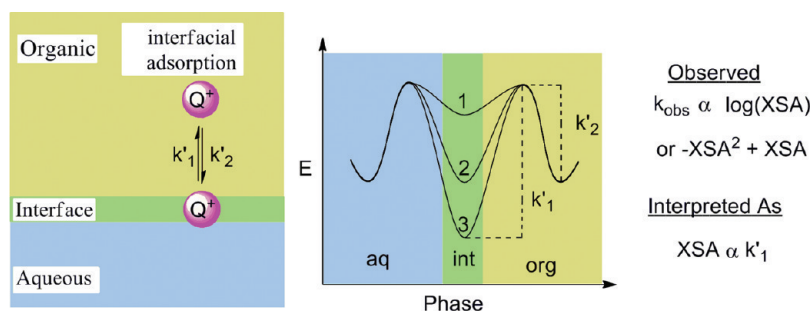


Figure 11. Analysis of XSA in terms of the relative rates of interfacial adsorption (k'_2) and desorption (k'_1).

relevant direct aqueous/organic phase transfer rates *must* be nearly equal which means that the aqueous and organic solvation energies of the catalyst must be equal as well (see Discussion section 2.1). Also, because the only descriptor other than $\log P$ in the derived QSAR is XSA, this descriptor *must* account for the deviation of catalyst activity from $\log P$ (and vice versa). Moreover, this descriptor must be reflecting a *different* rate-determining step than that which is reflected in cases where $\log P$ alone can describe activity.

The largest deviation of the observed catalyst activity was (conveniently) observed at the optimal $\log P$ value (Figure 3b). At this value, the thermodynamic solvation energies of the catalysts are the same, so this discussion can be reduced to only those scenarios where the aqueous and organic solvation energies are equal. Three limiting cases representing three different energies of an interfacially adsorbed state relative to the two solvated states in the bulk media are considered (Figure 11): case 1, wherein the adsorbed state is higher in energy; case 2, wherein the adsorbed state is equal in energy; and case 3, wherein the adsorbed state is lower in energy than the solvated states. In any scenario wherein the adsorbed state is equal to (case 2) or greater than (case 1) the solvated states, a good correlation with $\log P$ should hold because they represent the same rate determining step.¹⁰⁷ However, in any scenario wherein the adsorbed state is lower in energy than the solvated states (case 3), two new, potentially rate-limiting steps arise, namely, desorption from the interface into either the organic or aqueous phases. Because XSA describes the movement of an amphiphilic molecule from a state of high polarity (water) to a state of low polarity (air), the interpretation most consistent with the observed correlation of catalyst activity is rate limiting desorption from the interface and into the organic phase (k'_1 , case 3, Figure 11). Recall that with LFERs and QSARs, a correlation between activity and a descriptor is a reflection of that descriptors relatedness to the $\Delta\Delta G^\ddagger$ of the rate-determining step. Just as with the rationalization of $\log P$ (Discussion section 2.1), both kinetic and a thermodynamic interpretations are possible. Given that this is a study of catalysis, a kinetic interpretation is more meaningful and is described below.¹⁰⁸

From a kinetic perspective, rate-determining interfacial desorption of the catalyst-reactant complex is consistent with the interfacial mechanism forwarded by Makosza.¹⁰⁵ Because the PTC reaction must occur in the organic phase, or at least on the organic side of the interface where the electrophile is located, interpretation of XSA in terms of the rate of interfacial desorption (k'_1 , Figure 11) makes sense.^{109,110} Therefore, it may be concluded that the greater the cross-sectional area of the catalyst, the faster it is repelled from the interfacial region into the organic

phase. The observed leveling of catalyst activity with increasing XSA may be interpreted in two nearly equivalent ways. In the first interpretation, the observed leveling off of catalytic activity with increasing XSA is the result of approaching the energetic scenario where the rate of interfacial desorption and adsorption are equal (case 2, Figure 11). In a subtly different case, one could interpret the observed leveling of catalytic activity with increasing XSA as approaching an upper limit in the rate of interfacial desorption, namely, the rate of diffusion (see Figure 4a). Both of these interpretations are consistent with the proposed interfacial mechanism.^{111–113} In this way, the catalyst XSA may be treated as an analogue of the aqueous to organic phase-transfer rate (compare k'_1 in Figure 11 to k_1 Figure 9). However, what is not clear from this conclusion is whether XSA also incorporates the competing microscopic reverse step, k'_2 . That is, if XSA is a reflection of the interfacial desorption rate (k'_1) and *only* the interfacial desorption rate, then another descriptor, one to reflect the interfacial adsorption rate (k'_2) would be expected. In other words, why did the derived QSAR have two descriptors and not three?

To investigate this question, the entire data set was distilled down to a handful of data points by generating a large conformational library (6372 total, $\sim 62/\text{catalyst}$)¹¹⁴ and then taking the average of the XSA and catalytic activity for multiple catalyst subtypes. These averaged catalytic activities and cross-sectional areas are plotted against each other so that they can be compared (Figure 12). In general, the catalyst activity ($\log(k_{\text{rel}})$) increases linearly as the XSA of the catalyst increases and then levels off. There is a linear dependence of activity on catalyst XSA when XSA is less than or equal to $\sim 75 \text{ \AA}^2$ (also compare Figure 10b to Figure 10c). Once the XSA of the catalysts reaches $\sim 70\text{--}80 \text{ \AA}^2$ the observed activity of the catalysts changes only slightly. For the aliphatic ammonium ions, this value corresponds to catalysts with three or more butyl groups. For cyclopenta[gh]pyrrolizidinium ions, this value corresponds to catalysts that have a non-hydrogen substituent on oxygen (R^3). Catalysts with $R^2 = \text{H}$ or aryl (red circle) exhibited nearly identical activity as the aliphatic catalysts (blue diamond), however, catalysts with $R^2 = \text{aliphatic groups}$ (green square) exhibited slightly lower activities. Thus, as pointed out in a qualitative way in the accompanying paper,¹¹ catalysts bearing little or no substitution beyond the cyclopenta[gh]pyrrolizidinium scaffold (green triangle) behave similarly to tetraethyl- and tetrapropylammonium ions.

The most important part of analyzing the data in this way is inspection of which types of fits best represent the summarized data. The scatter plot of the average catalyst activity versus the average catalyst XSA generates a distinct type of nonlinear correlation, which is commonly encountered in enzyme

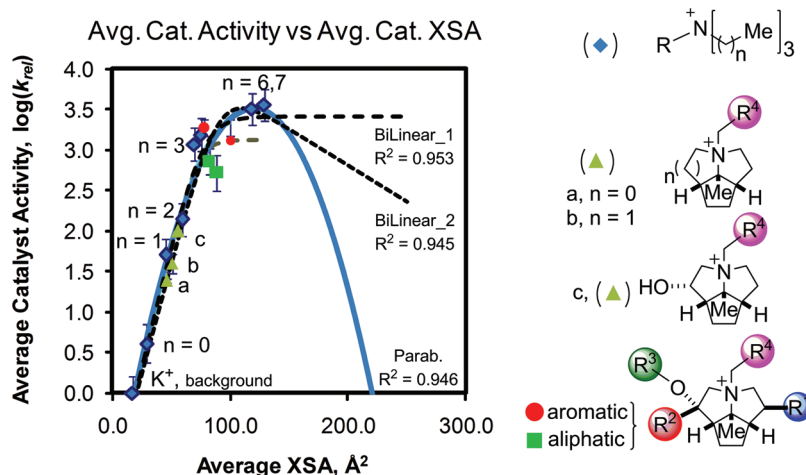


Figure 12. Plot of average catalyst activity versus average XSA for multiple catalyst chemotypes and corresponding regressions.

kinetics,^{115,116} surface adsorption,¹¹⁷ as well as the kinetics of transfer across an interfacial barrier.^{101,118} Specifically, the resulting scatter plot was fit well by rectangular bilinear functions (Bilinear_1 (dark blue line) and Bilinear_2 (dark blue dashed line) in Figure 12). Also, the fact that the parabolic fit (light blue line, Figure 12) arrived at empirically (Figure 4) was retained and the data collected to date fills only one side of the parabola means that more data would be needed to differentiate these interpretations. However, unlike the empirical screening that necessarily motivated the data collection process for this study, further data collection can be done in a quantitative manner with catalyst structures designed to probe specific hypotheses. For a mechanistic interpretation of XSA, it must be recognized that each of the phenomena described by more complex functions in Figure 12 involve processes composed of two rates that are the microscopic reverse of one another.¹¹⁹ Specifically, given the data collected thus far, three scenarios can be envisioned for catalysts with significantly larger cross-sectional areas ($XSA > 150 \text{ \AA}^2$) than those employed in the initial study. The simplest is that the parabolic fit of XSA found empirically (light blue line, Figure 12 and Figure 4) would hold and no other descriptor would be required. Mechanistically, this would mean that XSA is a reflection of k'_1/k'_2 . The second possibility is that the parabolic model would not be predictive, for which two “subclasses” can be proposed. In the simpler of these two subclasses (BiLinear_1) further increasing the cross-sectional area of the catalyst would cause no change in catalyst activity. Mechanistically this would mean that XSA is a reflection of k'_1 , and k'_1 is the only rate that matters. The reaction can never be rate determining in the adsorption step (k'_2). The second more complicated subclass is that the catalyst activity would decrease and be inversely proportional to the ammonium ion accessibility (BiLinear_2). Taken together with the known correlation of ammonium ion accessibility to catalyst activity (represented by q) for other PTC enolate alkylations, it is the latter possibility that seems most likely. That is, for larger catalysts, the rate determining step will likely change to adsorption to the interface (k'_2) and would be better described by structural features such as polar surface area. Fortunately, the number of possibilities is limited to three, so with only a few carefully designed catalysts (~ 3) these different interpretations can be distinguished.

2.4. Summary of QSAR Analysis of Catalyst Activity. In summary, a two-descriptor QSAR model of phase-transfer catalyst activity was derived. The two descriptors were cross-sectional area and $\log P$ and both were correlated to catalyst activity in a parabolic manner. Because the data was collected in a stir-rate dependent regime, the model was subsequently interpreted in the context of possible rate-determining steps involving a physical distribution of the catalyst. The QSAR modeling studies indicate the catalytic activity of ammonium ions can be expressed as a sum of two pairs of relative rates which define the two mechanistic extremes of PTC. The first pair of relative rates, those of aqueous/organic phase transfer, defines the extraction mechanism and are readily expressed as thermodynamic partition coefficients. The other pair of relative rates, those of adsorption/desorption from the interface, describe the interfacial mechanism. The descriptor most highly correlated to catalyst activity was XSA and was interpreted in terms of catalyst desorption from the interface and into the organic layer. The relative magnitudes of the coefficients in the QSAR model indicate that for the hydroxide-initiated PTC alkylation of glycine imine **1** the interfacial mechanism is ~ 10.5 times (the ratio of the coefficients in eq 1) more active pathway for non- T_d symmetrical quaternary ammonium ion catalysts. The fact that similar correlations are on record for PTC reactions conducted well above a stir-rate dependent regime indicates that the general form of the QSAR model derived herein may be applicable to other reaction types and catalysts.¹⁸

CONCLUSIONS AND OUTLOOK

Multiple quantitative models were developed for the selectivity and activity of the asymmetric phase transfer catalysts reported in the accompanying paper.¹¹ Two fundamentally different approaches were taken to develop the activity and selectivity models. The catalyst selectivities were modeled by CoMFA and the catalyst activities were modeled with multilinear regression of descriptors.

Correlations of the enantioselectivity with the steric and electrostatic fields (55% steric; 45% electrostatic) of the catalysts were statistically significant as determined by rigorous cross-validation. The model was readily interpreted qualitatively as relating to the differential binding preferences of the reactive enolate for the catalyst in accordance with contour maps represented as multiplicative products of coefficients of the

model and the variation in the molecular interaction fields of the catalysts. The steric field revealed differential van der Waals repulsive interactions between the two active faces of the catalysts available for binding as being relevant in explaining the observed reversal in enantioselectivity with the inclusion of bulky aliphatic groups shielding one of the two faces. The electrostatic field revealed that higher enantioselectivity for the *S*-configured product is obtained by further differentiating one face over the other through charge polarization. The inclusion of aryl groups at one of the faces selectively introduces positive potential in this region as a result of the quadrupole moment imparted by arenes, which enforces the preference for the *S*-configured product. This observation may indicate an important role for π - π interactions at this position for this reaction system. Additionally, the inclusion of a region of preferred negative potential for the *S* configured product, coupled with the region of preferred positive potential, may reveal a preferred dipole orientation that allows for facial discrimination of the enolate. Future development of more enantioselective catalysts based on the current scaffold would involve enhancing substrate interaction with the desired face for binding through the qualitative and quantitative application of the QSSR model developed here. The principles borne from this QSSR study may serve as criteria for the rational design of future APTC catalysts.

For this initial report, a discovery-oriented, statistical approach was taken to identify molecular descriptors and combinations thereof that best accounted for the variation in catalyst activities. Throughout the process, thousands of models with variable types and numbers of descriptors were found that could describe the variation in catalyst activity. It was found that models with more than seven descriptors were not able to fit the data much better than models with fewer descriptors. Attempts to assess the "goodness" of models were made throughout the process by various validation methods. The ultimate question of what models have predictive capacity is currently being probed experimentally with other catalyst scaffolds and reactions.

One of the descriptors identified by the "statistical screen" to be correlated with catalyst activity was the cross-sectional area of the catalyst. A nonlinear correlation of catalyst activity and XSA was noted and interpreted in terms of interfacial adsorption/desorption. Throughout the course of investigation of QSAR models for catalyst activity, many challenges were encountered. One of the key challenges was the identification of nonlinear relationships, and it seems likely that similar challenges will be intrinsically tied to any QSAR study of catalysis. The catalyst cross-sectional area descriptor was found to be uniquely capable of reflecting the size of the ammonium ion relevant to catalytic activity. Future developments toward understanding and modeling PTC activity will focus on and test different descriptors for the steric and electronic environment around the central ammonium ion. Specifically, custom molecular shadow indices¹²⁰ or sterimol algorithms¹²¹ capable of providing information about the steric and electrostatic environment of the central ammonium would be worth investigating.

Drawing specific mechanistic conclusions from a QSAR study is necessarily speculative; nonetheless, a significant effort was put forth to do so. Accordingly, it is proposed that the greatest contributors to the activity of ammonium ion catalysts in hydroxide-initiated PTC reactions is their relative interfacial adsorption/desorption abilities and to a lesser extent the complete aqueous/organic phase transfer rates, two easily testable hypotheses. The design, synthesis, evaluation, and modeling of

catalysts to test these hypotheses are currently underway and the results will be disclosed in due course. Given the paucity of QSAR approaches to optimizing and understanding catalytic reactions, having a quantitative model for hypotheses testing is apparent.

■ ASSOCIATED CONTENT

Supporting Information. Cartesian coordinates for all calculated structures, model development, and cross validation. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: sdenmark@illinois.edu.

■ ACKNOWLEDGMENT

We are grateful to Professor Marissa Kozlowski for insightful discussion relating to QSSR modeling. We are grateful to the National Institutes of Health (R01 GM30938) and the American Chemical Society Petroleum Research Fund (ACS PRF 49668-ND1) for generous financial support. N.D.G. thanks Amgen for a Graduate Fellowship in Synthetic Organic Chemistry.

■ REFERENCES

- (1) (a) Wells, P. R. *Linear Free Energy Relationships*; Academic Press: New York, 1968. (b) Anslyn, E. V.; Dougherty, D. A. *Modern Physical Organic Chemistry*; University Science: Sausalito, CA, 2006; pp 442–471. (c) March, J. *Advanced Organic Chemistry*, 4th Ed.; Wiley: New York, 2001; pp 368–375. (d) Sundberg, C. A. *Advanced Organic Chemistry*, 4th ed.; Plenum: New York, 2000; Part A, pp 204–215. (e) Lowry, T. H. Richardson, K. S. *Mechanism and Theory in Organic Chemistry*, 3rd ed.; Harper and Row: New York, 1987; pp 143–158. (f) Page, M. I. *The Chemistry of Enzyme Action*; Elsevier: New York, 1984; pp 143–166.
- (2) (a) Taft, R. W., Jr. *J. Am. Chem. Soc.* **1952**, *74*, 2729–2732. (b) Taft, R. W., Jr. *J. Am. Chem. Soc.* **1953**, *75*, 4538–4539. (c) Taft, R. W., Jr. In *Steric Effects in Organic Chemistry*; Newman, M. S., Ed.; John Wiley and Sons: New York, 1956; pp 556–675. (d) Charton, M. *J. Am. Chem. Soc.* **1969**, *91*, 615–618. (e) Charton, M. *J. Am. Chem. Soc.* **1975**, *97*, 3691–3693. (f) Charton, M. *J. Am. Chem. Soc.* **1975**, *97*, 3694–3697. (g) Charton, M. *J. Am. Chem. Soc.* **1975**, *97*, 1552–1556.
- (3) (a) Hammett, L. P. *J. Am. Chem. Soc.* **1937**, *59*, 96–103. (b) Jaffe, H. H. *Chem. Rev.* **1953**, *53*, 191–261. (c) Hammett, L. P. *J. Chem. Educ.* **1966**, *43*, 464–469.
- (4) (a) Hansch, C.; Maloney, P. P.; Fujita, T.; Muir, R. M. *Nature* **1962**, *194*, 178–180. (b) Hansch, C. *Acc. Chem. Res.* **1969**, *2*, 232–239. (c) Leo, A.; Hansch, C.; Elkins, D. *Chem. Rev.* **1971**, *71*, 525–616.
- (5) (a) Hansch, C.; Steinmetz, W. E.; Leo, A. J.; Mekapati, S. B.; Kurup, A.; Hoekman, D. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 120–125. (b) Verma, R. P.; Kurup, A.; Hansch, C. *Bioorg. Med. Chem.* **2004**, *13*, 237–255. (c) Verma, R. P.; Hansch, C. *Bioorg. Med. Chem.* **2005**, *13*, 2355–2372.
- (6) Kubinyi, H. *Drug Discovery Today* **1997**, *2*, 457–467.
- (7) Sigman, M. S.; Miller, J. J. *J. Org. Chem.* **2009**, *74*, 7633–7643.
- (8) (a) Brändström, A. *J. Chem. Soc., Perkin Trans. 2* **1999**, 2419–2422. (b) Camenisch, G.; Folkers, G.; van de Waterbeemd, H. *Pharm. Acta Helv.* **1996**, *71*, 309–327.
- (9) (a) Hansch, C.; Kurup, A.; Garg, R.; Gao, H. *Chem. Rev.* **2001**, *101*, 619–672. (b) Gasteiger, J. *Handbook of Chemoinformatics: From Data to Knowledge*; Wiley-VCH: Weinheim, 2003. (c) Gasteiger, J.; Engel, T. *Chemoinformatics: A Textbook*; Wiley-VCH: Weinheim, 2003. (d) Engel, T. *J. Chem. Inf. Model.* **2006**, *46*, 2267–2277. (e) Upmanyu,

N.; Gupta, S.; Prakash, B. N.; Garg, G.; Mishra, P. *Res. J. Pharm. Technol.* **2008**, *1*, 2–5.(f) *Molecular Descriptors for Chemoinformatics*, 2nd ed.; Todeschini, R., Consonni, V., Eds.; Wiley-VCH: Weinheim, 2009; Vol. I, Alphabetical Listing. (g) *Molecular Descriptors for Chemoinformatics*, 2nd ed.; Todeschini, R., Consonni, V., Eds.; Wiley-VCH: Weinheim, 2009; Vol. II: Appendices, References.

(10) As such, no testable hypotheses have been proposed to correlate catalyst structure to activity for hydroxide-initiated PTC reactions with chiral quaternary ammonium ion catalysts.

(11) The preceding paper in this issue (DOI: 10.1021/jo2005445).

(12) Jencks, W. P. *Acc. Chem. Res.* **1976**, *9*, 425–432. (b) Jencks, W. P. *Acc. Chem. Res.* **1980**, *13*, 161–169.(c) Jencks, W. P. *Catalysis in Chemistry and Enzymology*; McGraw-Hill: New York, 1969.

(13) Jensen, K. H.; Sigman, M. S. *Angew. Chem., Int. Ed.* **2007**, *46*, 4784–4750.

(14) (a) Cruz, V.; Ramos, J.; Muñoz-Escalona, A.; Lafuenta, A.; Peña, B.; Martínez-Salazar, J. *Polymer* **2004**, *45*, 2061–2072. (b) Cruz, V. L.; Ramos, J.; Martínez, S.; Muñoz-Escalona, A.; Martínez-Salazar, J. *Organometallics* **2005**, *24*, 5095–5102. (c) Cruz, V. L.; Martínez, S.; Martínez-Salazar, J.; Polo-Ceron, D.; Gomez-Ruiz, S.; Fajardo, M.; Prashar, S. *Polymer* **2007**, *48*, 4663–4674.

(15) (a) Rothenberg, G.; Burello, E. *Adv. Synth. Catal.* **2003**, *345*, 1334–1340. (b) Rothenberg, G.; Burello, E.; Farrusseng, D. *Adv. Synth. Catal.* **2004**, *346*, 1844–1853. (c) Rothenberg, G. *Catal. Today* **2008**, *137*, 2–10. (d) Rothenberg, G.; Maldonado, A. G. *Chem. Soc. Rev.* **2010**, *39*, 1891–1902. (e) Tibiletti, D.; Graaf, E. A. B. de; Pheng Teh, S.; Rothenberg, G.; Farrusseng, D.; Mirodatos, C. *J. Catal.* **2004**, *225*, 489–497.

(16) For a recent example concerning the water gas shift reaction, see: Baumes, I.; Farrusseng, D.; Lengliz, M.; Mirodatos, C. *QSAR Comb. Sci.* **2004**, *23*, 767–778.

(17) (a) Herriot, A. W.; Picker, D. *J. Am. Chem. Soc.* **1975**, *97*, 2345–2349. (b) Landini, D.; Maia, A.; Montanari, F. *J. Chem. Soc., Chem. Commun.* **1977**, 112–113.

(18) Starks, C. M.; Liotta, C. L.; Halpern, M. *Phase-Transfer Catalysis: Fundamentals, Applications and Industrial Perspectives*; Chapman and Hall: New York, 1994.

(19) Halpern, M.; Sasson, Y.; Rabinovitz, M. *Tetrahedron* **1982**, *38*, 3183–3187.

(20) (a) q is defined as the reciprocal sum of the number of carbons in each chain of the quaternary ammonium ion. (b) Halpern, M. In *Phase-Transfer Catalysis, Mechanism and Synthesis*; Halpern, M. E., Ed.; ACS Symposium Series 659; American Chemical Society: Washington, DC, 1996; Chapter 8. (c) For a discussion on the use of this relationship see: Starks, C. M.; Liotta, C. L.; Halpern, M. *Phase-Transfer Catalysis: Mechanisms and Syntheses*; Chapman and Hall: New York, 1997; pp 270–287.

(21) Dehmlow, E. V. In *Phase-Transfer Catalysis, Mechanism and Synthesis*; Halpern, M. E., Ed.; ACS Symposium Series 659; American Chemical Society: Washington, DC, 1996; 108–122.

(22) Camenisch, G.; Folkers, G.; van de Waterbeemd, H. *Pharm. Acta. Helv.* **1996**, *71*, 309–327.

(23) Sprunger, L. M.; Gibbs, J.; Acree, W. E., Jr.; Abraham, M. H. *QSAR Comb. Sci.* **2009**, *28*, 72–88.

(24) Abraham, M. H.; Acree, W. E., Jr. *J. Org. Chem.* **2010**, *75*, 1006–1015.

(25) Bunin, B. A.; Siesel, B.; Morales, G. A.; Bajorath, J. *Chemoinformatics: Theory, Practice, & Products*; Springer: Dordrecht, 2007.

(26) (a) Miller, J. J.; Sigman, M. S. *Angew. Chem., Int. Ed.* **2008**, *47*, 771–774. (b) Sigman, M. S.; Miller, J. J. *J. Org. Chem.* **2009**, *74*, 7633–7643.

(27) Jiang, C.; Li, Y.; Tian, Q.; You, T. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1876–1881.

(28) (a) Lipkowitz, K. B.; Schefzick, S.; Avnir, A. *J. Am. Chem. Soc.* **2001**, *123*, 6710–6711. (b) Alvarez, S.; Schefzick, S.; Lipkowitz, K.; Avnir, D. *Chem.—Eur. J.* **2003**, *9*, 5832–5837.

(29) *Molecular Interaction Fields: Applications in Drug Discovery and ADME Prediction*; Cruciani, G., Ed.; Series: Methods and Principles in Medicinal Chemistry; Wiley-VCH: Weinheim, 2006; Vol 27.

(30) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.

(31) Lipkowitz, K. B.; Pradhan, M. *J. Org. Chem.* **2003**, *68*, 4648–4656.

(32) Dixon, S.; Merz, K. M., Jr.; Lauri, G.; Ianni, J. C. *J. Comput. Chem.* **2004**, *26*, 23–24.

(33) Phaun, P. –W.; Ianni, J. C.; Kozlowski, M. C. *J. Am. Chem. Soc.* **2004**, *126*, 15473–15479.

(34) (a) Kozlowski, M. C.; Dixon, S. L.; Panda, M.; Lauri, G. *J. Am. Chem. Soc.* **2003**, *125*, 6614–6615. (b) Ianni, J. C.; Annamalai, V.; Phuan, P.-W.; Panda, M.; Kozlowski, M. C. *Angew. Chem., Int. Ed.* **2006**, *45*, 5502–5505. (c) Huang, J.; Ianni, J. C.; Antoline, J. E.; Hsung, R. P.; Kozlowski, M. C. *Org. Lett.* **2006**, *8*, 1565–1568. (d) Urbano-Cuadrado, M.; Carbo, J. J.; Maldonado, A. G.; Bo, C. *J. Chem. Inf. Model.* **2007**, *47*, 2228–2234.

(35) (a) Melville, J. L.; Andrews, B. I.; Lygo, B.; Hirst, J. D. *Chem. Commun.* **2004**, 1410–1411. (b) Melville, J. L.; Lovelock, K. R. J.; Wilson, J.; Allbutt, B.; Burke, E. K.; Lygo, B.; Hirst, J. D. *J. Chem. Inf. Model.* **2005**, *45*, 971–981.

(36) Lygo, B.; Allbutt, B.; James, R. S. *Tetrahedron Lett.* **2003**, *44*, 5629–5632.

(37) Lipkowitz, K. B.; Cavanaugh, M. W.; Baker, B.; O'Donnell, M. J. *J. Org. Chem.* **1991**, *56*, 5181–5192.

(38) *Asymmetric Phase Transfer Catalysis*; Maruoka, K., Ed.; Wiley-VCH: New York, 2008.

(39) The advantages of investigating catalyst systems that are stereogenic at the ammonium nitrogen atom are addressed in the accompanying paper.

(40) Spartan '08 is a product of Wavefunction, Inc.: 18401 Von Karman Ave, Suite 370. Irvine, CA 92612, <http://www.wavefun.com/index.html>.

(41) Cramer, C.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 760–768.

(42) SYBYL X 1.1, Tripos International, 1699 South Hanley Rd, St. Louis, MO, 63144.

(43) See the previous paper in this issue (DOI: 10.1021/jo2005445) for more qualitative rationalizations in the absence of QSAR models.

(44) Mittal, R. R.; Harris, L.; McKinnon, R. A.; Soric, M. J. *J. Chem. Inf. Model.* **2009**, *49*, 704–709.

(45) (a) Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S. *J. Med. Chem.* **2000**, *43*, 3233–3243. (b) Sciabola, S.; Alex, A.; Higginson, P. D.; Mitchell, J. C.; Snowden, M. J.; Morao, I. *J. Org. Chem.* **2005**, *70*, 9025–9027.

(46) Kroemer, R. T.; Hecht, P. J. *Comput.-Aided Mol. Des.* **1995**, *9*, 205–212.

(47) MOE (The Molecular Operating Environment), Version 2006.10, software available from Chemical Computing Group Inc., 1010 Sherbrooke Street West, Suite 910, Montreal, Canada H3A 2R7, <http://www.chemcomp.com>.

(48) For the purposes of this analysis descriptors that differentiate molecules by graph theory and/or connectivity indices were excluded.

(49) Initial attempts to develop a high-throughput HPLC method for the determination of log(toluene/water) values for quaternary ammonium ions has not been successful.

(50) The logP(benzene/water) values were calculated by the SM8 method⁴¹ as implemented in Jaguar (6-31+G(d)//B3LYP) and are scaled appropriately to a temperature of 3 °C.

(51) (a) For lead references on the TAE/RECON descriptor set, see: Breneman, C. M.; Rkhem, M. *J. Comput. Chem.* **1997**, *18*, 182–197. (b) Oloff, S.; Zhang, S.; Sukumar, C. M.; Breneman, C. M. *J. Chem. Inf. Model.* **2006**, *46*, 844–851.

(52) First (α) and second (β) polarizabilities were calculated in Spartan 08 by the default method.

(53) Cherkasov, A.; Shi, Z.; Fallahi, M.; Hammond, G. L. *J. Med. Chem.* **2005**, *48*, 3203–3213.

(54) All descriptor calculation software, algorithms, and statistical fitting software utilized in this study are from the commercial MOE package or available via the svl exchange at <http://svl.chemcomp.com>.

(55) See the Supporting Information for a summary of the specific settings for GA runs (e.g., population sizes, eugenic factors, etc).

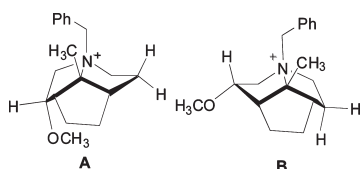
(56) The 3D coordinates for all molecules included in the enantioselectivity and activity data set are included in the Supporting Information.

(57) (a) Golbraikh, A.; Tropsha, A. *J. Mol. Graphics Modell.* **2002**, *20*, 269–276. (b) Golbraikh, A.; Tropsha, A. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 357–369. (c) Tropsha, A.; Gramatica, P.; Gombar, V. K. *QSAR Comb. Sci.* **2003**, *22*, 69–77.

(58) Wold, S.; Eriksson, L. In *Chemometric Methods in Molecular Design*; van de Waterbeemd, H., Ed.; Wiley-VCH: Weinheim, 1995; pp 309–318.

(59) (a) Hopfinger, A. J.; Wang, S.; Tokarski, J. S.; Jin, B.; Albuquerque, M.; Madhav, P. J.; Duraiswami, C. *J. Am. Chem. Soc.* **1997**, *119*, 10509–10524. (b) See ref 35b for an implementation of the method of ref 59a and variants thereof applied to APTC.

(60) The energy difference between conformations **A** and **B** is calculated to be 2.4 kcal in favor of **A** (B3LYP/6-31G(d)//M06-2X/6-31G(d) in toluene solvent (SM8)). This difference is presumed to be larger when $R^2 \neq H$. Preliminary modeling studies in the presence of an anion support a significantly larger energy difference between **A** and **B**.



(61) The energy difference ranges from 2.6 kcal ($R^1 = R^2 = H$) favoring the **up** conformation to 0.8 kcal favoring the **dn** conformation depending on the identities of R^1 and R^2 (B3LYP/6-31G(d)//M06-2X/6-31G(d) in toluene solvent (SM8), without a counterion, and with $R^3 = Me$; $R^4 = Ph$ for reduction in the number of basis functions since R^3 is unlikely to influence the ring conformation significantly).

(62) A full table including the investigation of different semiempirical partial charges (including PM3, AM1) is available in the Supporting Information.

(63) Additional data relating to the external cross-validation and the methods by which it was performed are available in the Supporting Information.

(64) (a) Halpern, M.; Sasson, Y.; Rabinovitz, M. *Tetrahedron* **1982**, *38*, 3183–3187. (b) Makosza, M.; Serafinowa, B. *Rocz. Chem.* **1965**, *39*, 1223–1230.

(65) It is important to note that a QSAR model that exhibits a “near perfect fit” is unlikely to be generally applicable and is, rather, likely overfit because a QSAR model is based on experimental data with error. In this example, the “data” being modeled is not experimental though.

(66) The NC4_SA descriptors were coded in scientific vector language (svl) utilizing a smiles string search ([N+](C)(C)(C)(C)) and the AtomSurfaceArea function with variable probe radii.

(67) All of the methods surveyed gave similar correlations. (a) Hou, T. J.; Xia, K.; Zhang, W.; Xu, X. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 266–275. (b) Wildman, S. A.; Crippen, G. M. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 868–873.

(68) In most medicinal chemistry applications a quaternary ammonium ion is assigned a polar surface area of zero. In this study, the quaternary ammonium ion was assigned a value of ten. No changes to the other standard TPSA values were made.⁶⁹

(69) Ertl, P.; Rohde, B.; Selzer, P. *J. Med. Chem.* **2000**, *43*, 3714–3717.

(70) Crivori, P.; Cruciani, G.; Carrupt, P.-A.; Testa, B. *J. Med. Chem.* **2000**, *43*, 2204–2216.

(71) For an excellent summary of QSAR models of solute/solvent interactions, see: Abraham, M. H.; Ibrahim, A.; Zissimos, A. M. *J. Chromatogr., A* **2004**, *1037*, 29–47.

(72) Gerebtzoff, G.; Seelig, A. *J. Chem. Inf. Model* **2006**, *46*, 2638–2650.

(73) Cruciani, G.; Crivori, P.; Carrupt, P.-A.; Testa, B. *THEOCHEM* **2000**, *503*, 17–30.

(74) Based on the MOE partial charge descriptor set including those from: Stanton, D.; Jurs, P. *Anal. Chem.* **1990**, *62*, 2323–2329.

(75) (a) Topliss, J. G.; Costello, R. J. *J. Med. Chem.* **1972**, *15*, 1066–1068. (b) Unger, S. H.; Hansch, C. *J. Med. Chem.* **1973**, *16*, 745–749.

(76) In a typical descriptor binning process, the full range of a surface or volume value, e.g., electron interaction energy, is divided incrementally. The resultant sum of the surface areas or volume that occupy the incremental ranges (e.g., 1–2, 2–3, and so on) are used as individual descriptors.

(77) For this comparison, we chose the $\log P(o/w)$ values as determined by the MOE partial atom contribution method since it was the value most frequently included by the GA-MLR model screens.

(78) A standard dimension is equivalent to the standard deviation along a principal component axis.

(79) Various, derivatives of the cross-sectional area descriptor were investigated as well.

(80) See the discussion in the previous paper in this issue (DOI: 10.1021/jo2005445) for a comparison of electrostatic potential maps.

(81) (a) Anslyn, E. V.; Dougherty, D. A. *Modern Physical Organic Chemistry*; University Science: Sausalito, 2006; p 181. (b) An alternative explanation for the enantioselectivity enhancing effects for arenes at the R^2 position is the provision of a hydrophobic pocket for enolate association. However, the absence of any steric contours in this region in the contour maps suggested an electrostatic interpretation.

(82) The explanation provided is applicable under the assumption that the outcome is dependent on the difference in ground state energies of association complexes rather than the difference in transition state energies of the intrinsic alkylation reaction.

(83) See ref 81b, in which case this interaction would more appropriately belong in the steric subsection.

(84) Aldrich, J. *Stat. Sci.* **1995**, *10*, 364–376.

(85) Aldridge, J. W. *Science* **2005**, *308*, 954.

(86) Dowejko, A. M. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 81–89.

(87) Starks, C. M.; Owens, R. M. *J. Am. Chem. Soc.* **1973**, *95*, 3613–3617. (b) Starks, C. M. *J. Am. Chem. Soc.* **1971**, *93*, 195–199.

(88) Herriott, A. M.; Picker, D. *Tetrahedron Lett.* **1974**, *16*, 1511–1514.

(89) Starks, C. In *Phase-Transfer Catalysis, Mechanism and Synthesis*; Halpern, M. E., Ed.; ACS Symposium Series 659; American Chemical Society: Washington, DC, 1996; Chapter 2.

(90) (a) Leo, A.; Hansch, C.; Elkins, D. *Chem. Rev.* **1971**, *71*, 525–616. (b) Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR. Hydrophobic, electronic and steric constants*; American Chemical Society: Washington, DC, 1995; p 557.

(91) Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*; John Wiley & Sons: Hoboken, 2004; pp 385–485.

(92) Landini, D.; Mai, A.; Montanari, F. *J. Am. Chem. Soc.* **1978**, *100*, 2796–2801.

(93) Significant effort has been devoted the two understanding catalyst partitioning by kinetic modeling of the catalytically relevant mass transfer events of the catalysts. For lead references, see: (a) Melville, J. B.; Goddard, J. D. *Ind. Eng. Chem. Res.* **1988**, *27*, 551–555. (b) Wang, M. L.; Yang, H. M. *Chem. Eng. Sci.* **1991**, *46*, 509–517. (c) Wang, M. L.; Hsieh, Y. M. *Chem. Eng. Jpn.* **1993**, *26*, 374–381. (d) Naik, S. D.; Doraiswamy, L. K. *Amer. Inst. Chem. Eng. J.* **1998**, *44*, 612–646.

(94) (a) Lippold, B. C.; Schneider, G. F. *Arzneim.-Forsch.* **1975**, *25*, 843–852. (b) Lippold, B. C.; Schneider, G. F. *Arzneim.-Forsch.* **1975**, *25*, 1683–1686. (c) Lippold, B. C.; Schneider, G. F. *Pharmazie* **1976**, *31*, 237–239.

(95) (a) van de Waterbeemd, J.; Th., M.; Boeckel, S.; Jansen, A.; Gerritsma, K. *Eur. J. Med. Chem.* **1980**, *15*, 279–282.

(96) (a) Starks, C. M.; Liotta, C. L.; Halpern, M. *Phase-Transfer Catalysis: Fundamentals, Applications and Industrial Perspectives*; Chapman and Hall: New York, 1994. (b) *Handbook of Phase Transfer Catalysis*; Sasson, Y., Neumann, R., Eds.; Chapman & Hall: London, 1997.

(97) An alternative interpretation of the deviation of the calculated $\log P_{(b/w)}$ from the optimum theoretical $\log P$ invokes a perturbation of reaction preequilibria. In this analysis only the quaternary ammonium ion was considered without inclusion of a counterion. Under conditions with substoichiometric amounts of phase-transfer catalyst (Q^+), at least three quaternary ammonium ion pairs are present, ammonium·substrate, ammonium·hydroxide, and ammonium·halide, and each of them has associated competing k_1 's and k_2 's. Under PTC reaction conditions all three can interconvert, either inter- or intrafacially. Thus, an alternative interpretation of the difference between the theoretical and observed optimum $\log P$'s is a perturbation of these preequilibria.

(98) (b) Hansch, C. In *Drug Design*; Ariens, E. J., Ed.; Academic Press: London, 1971; Vol. 1; pp 271–342.

(99) (a) Ho, N. F. H.; Park, J. Y.; Morozowich, W.; Higuchi, W. I. In *Design of Biopharmaceutical Properties through Prodrugs and Analogs*; Roche, E. B., Ed.; APhA/APS: Washington, DC, 1977; pp 136–227. (b) Ho, N. F. H.; Park, J. Y.; Ni, P. F.; Higuchi, W. I. In *Animal Models for Oral Drug Delivery in Man: In Situ and In Vivo Approaches*; Crouthamel, W. G., Sarapu, A., Eds.; APBA/APS: Washington, DC, 1983; pp 27–106.

(100) (a) Leahy, D. E.; Lynch, J.; Taylor, D. C. In *Novel Drug Delivery and its Therapeutic Application*; Prescott, L. F.; Nimmo, W. S., Eds.; Wiley: New York, 1989; pp 33–44. (b) Wagner, J. G.; Sedman, A. J. *J. Pharmacokin. Biopharm.* **1973**, *1*, 23–50.

(101) (a) Kubinyi, H. *Arzneim. Forsch. (Drug Res.)* **1976**, *26*, 1991–1997. (b) Kubinyi, H. *J. Med. Chem.* **1977**, *20*, 625–629. (c) Buckwald, P. *J. Pharm. Sci.* **2005**, *94*, 2355–2379.

(102) (a) Fischer, H.; Gottschlich, R.; Seelig, A. *J. Membr. Biol.* **1998**, *165*, 201–211. (b) Seelig, A. *J. Mol. Neurosci.* **2007**, *33*, 32–41.

(103) Seelig, A.; Gottschlich, R.; Devant, R. M. *Proc. Natl. Acad. Sci.* **1994**, *91*, 68–72.

(104) Studies of this sort utilize the Gibbs adsorption isotherm, Langmuir isotherm, and ultimately the Szyszkowski equation to determine the cross-sectional area of the interface occupied by an amphiphile and the surface adsorption equilibrium constant. For a lead reference, see: Rosen, M. J. *Surfactants and Interfacial Phenomena*; John Wiley & Sons: New York, 1989.

(105) Makosza, M.; Lasek, W. *J. Phys. Org. Chem.* **1993**, *6*, 412–420.

(106) For lead references on the fundamentals of and mathematics used to describe adsorption/desorption processes, see: (a) *Fundamentals of Adsorption*; Mersmann, A. B., Scholl, S. E., Eds.; United Engineering Trustees, Inc.: Sonthofen, 1991. (b) Fawcett, R. W. *Liquids, Solutions, and Interfaces*; Oxford University Press: New York, 2004.

(107) Scenario 2 wherein the energy of the adsorbed state is equal to the aqueous and organic solvated states is the special case where the catalytic activity predicted by $\log P$ would be observed even though the mechanistic pathway involves an intermediate adsorbed state. As such, scenario 2 would not result in a significant deviation from the activity predicted by $\log P$ and can be discounted.

(108) Although less relevant, a thermodynamic interpretation is more intuitive. On the basis of thermodynamic stability considerations, the ammonium ions whose activities deviate the most from that predicted by $\log P_{(b/w)}$ are expected to prefer the interfacially adsorbed state. For example, the activity of cetyltrimethylammonium ion deviates the most from what would be predicted by $\log P$ alone (see lowest X in Figure 3b), which implies that it is strongly adsorbed at the interface. This characteristic is consistent with its known behavior as a surfactant. For a recent reference characterizing both the thermodynamics and kinetics of cetyltrimethylammonium ion surface adsorption, see: Howard, S. C.; Craig, V. S. *J. Langmuir* **2009**, *25*, 13015–13024.

(109) Norton, P. R.; Zhdanov, V. P. *Langmuir* **1994**, *10*, 1292–1296.

(110) (a) Dopierala, K.; Prochaska, K. *J. Colloid Interface Sci.* **2008**, *321*, 220–226. (b) Tamaki, K. *Bull. Chem. Soc. Jpn.* **1967**, *40*, 38–41. (c) Tamaki, K. *Bull. Chem. Soc. Jpn.* **1974**, *47*, 2764–2767.

(111) Along this line of reasoning, it is interesting to note the observation that interfacial surface tension is correlated with catalyst activity in a nonlinear manner. (a) Masson, D.; Magdassi, S.; Sasson, Y. *J. Org. Chem.* **1990**, *55*, 2714–2717. (b) Moberg, R.; Bokman, F.; Bohman, O.; Siegbahn, H. O. *G. J. Am. Chem. Soc.* **1991**, *113*, 3663–3667.

(112) Starks, C. M.; Liotta, C. L.; Halpern, M. *Phase-Transfer Catalysis: Fundamentals, Applications and Industrial Perspectives*; Chapman and Hall: New York, 1994; pp 383–438.

(113) (a) For a lead reference, see: Sirovski, F. S. In *Phase-Transfer Catalysis: Mechanisms and Synthesis*; Halpern, M. E., Ed.; ACS Symposium Series 659; American Chemical Society: Washington, DC, 1997; pp 68–88. (b) Rate enhancement by micellar catalysis, employing cationic surfactants, such as trimethylcetylammmonium (TMCA) is typically ~10–100 fold and attributable to compression from lipophilic forces and effective local dielectric of the media. The rate enhancement of TMCA in this reaction is 9.2 times the background and only 4.7 times greater than Me_4N^+ , indicating that a rate increase of 4.5-fold can be explained solely by the formation of micelles. For a lead reference, see: Rathman, J. F. *Curr. Opin. Colloid Interface Sci.* **1996**, *1*, 514–518.

(114) The conformational library was generated with MOE in low mode molecular dynamics with a limit of 200 conformers per catalyst, and only conformers within 7 kcal/mol of the lowest energy conformer were kept. Stereogenic centers were not allowed to invert. For a tabular summary of the averaged data, see the Supporting Information.

(115) Hill, A. V. *J. Physiol. (Oxford, U.K.)* **1910**, *40*, 4–7.

(116) Michaelis, L.; Menten, M. L. *Biochem. Z.* **1913**, *49*, 333–369.

(117) Langmuir, I. *J. Am. Chem. Soc.* **1919**, *30*, 2221–2295.

(118) (a) Walsh, C. *Enzymatic Reaction Mechanisms*; W. H. Freeman: San Francisco, 1979. (b) Cornish-Bowden, A. *Fundamentals of enzyme kinetics*, 3rd ed.; Portland Press: London, 2004. (c) Cleland, W. W.; Cook, P. *Enzyme Kinetics and Mechanism*; Garland Science: New York, 2007.

(119) (a) For a lead reference, see: Ben-Naim, A. *Cooperativity and Regulation in Biochemical Processes*; Kluwer Academic: New York, 2001. (b) For a historical summary of the interrelatedness and derivations of surface saturation (Langmuir isotherm) and enzyme saturation kinetics (Michaelis–Menton and Hill equations), see: Colquhoun, D. *Trends Pharmacol. Sci.* **2006**, *27*, 149–157.

(120) Verloop, A.; Hoogenstraaten, W.; Tipker, J. In *Drug Design*; Ariens, E. J., Ed.; Academic Press: New York, 1976; Vol. 6; pp 165–207.

(121) Jurs, P.; Rohrbaugh, R. H. *Anal. Chim. Acta* **1987**, *199*, 99–109.