Prediction of Thermochemical Properties of Long-Chain Alkanes Using Linear Regression Application to Hydroisomerization

Sharma, S.; Sleijfer, J.J.; Op de Beek, J.C.; van der Zeeuw, S.; Zorzos, D.; Lasala, Silvia; Rigutto, Marcello; Zuidema, Erik; Vlugt, T.J.H.; More Authors

DOI
10.1021/acs.jpcb.4c05355

Publication date
2024

Document Version
Final published version

Published in
The Journal of Physical Chemistry Part B (Biophysical Chemistry, Biomaterials, Liquids, and Soft Matter)

Citation (APA)
Sharma, S., Sleijfer, J. J., Op de Beek, J. C., van der Zeeuw, S., Zorzos, D., Lasala, S., Rigutto, M., Zuidema, E., Vlugt, T. J. H., & More Authors (2024). Prediction of Thermochemical Properties of Long-Chain Alkanes Using Linear Regression: Application to Hydroisomerization. *The Journal of Physical Chemistry Part B (Biophysical Chemistry, Biomaterials, Liquids, and Soft Matter)*, *128*(39), 9619-9629. https://doi.org/10.1021/acs.jpcb.4c05355

THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB
Article

Prediction of Thermochemical Properties of Long-Chain Alkanes Using Linear Regression: Application to Hydroisomerization

Published as part of The Journal of Physical Chemistry B special issue "Athanassios Z. Panagiotopoulos Festschrift".

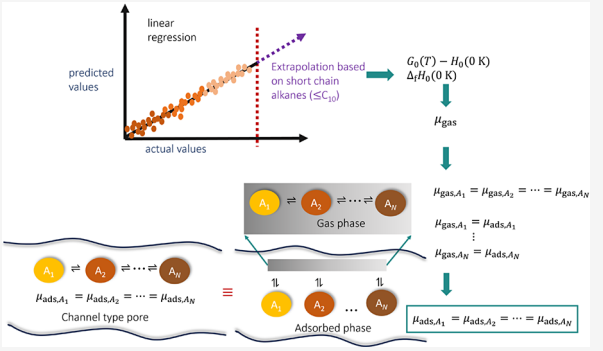
Shrinjay Sharma, Josh J. Sleijfer, Jeroen Op de Beek, Stach van der Zeeuw, Daniil Zorzos, Silvia Lasala, Marcello S. Rigutto, Erik Zuidema, Umang Agarwal, Richard Baur, Sofia Calero, David Dubbeldam, and Thijs J.H. Vlugt*

Cite This: J. Phys. Chem. B 2024, 128, 9619−9629

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Linear regression (LR) is used to predict thermochemical properties of alkanes at temperatures (0−1000) K to study chemical reaction equilibria inside zeolites. The thermochemical properties of C_1 until C_{10} isomers reported by Scott are used as training data sets in the LR model which is used to predict these properties for alkanes longer than C_{10} isomers. Second-order groups are used as independent variables which account for the interactions between the neighboring groups of atoms. This model accurately predicts Gibbs free energies, enthalpies, Gibbs free energies of formation, and enthalpies of formation for alkanes which exceeds the chemical accuracy of 1 kcal/mol and outperforms the group contribution methods developed by Benson et al., Joback and Reid, and Constantinou and Gani. Predictions from our model are used to compute the reaction equilibrium distribution of hydroisomerization of C_{10} and C_{14} isomers in MTW-type zeolite. Calculation of reaction equilibrium distribution inside zeolites also requires Henry coefficients of the isomers which can be computed using classical force field-based molecular simulations using the RASPA2 software for which we created an automated workflow. The reaction equilibrium distribution for C_{10} isomers obtained using the LR model and the training data set for this model are in very good agreement. The tools developed in this study will enable the computational study of hydroisomerization of long-chain alkanes ($>C_{10}$).

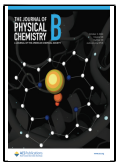


1. INTRODUCTION

In transitioning toward fuels and chemicals from renewable sources, platforms that provide clean hydrocarbon liquid energy carriers from carbon dioxide directly or via biocomponents can play an important role.[1] For sustainable aviation fuel and low carbon gas-oil or lubricants, iso-alkanes with a high degree of branching are the preferred constituents,[2] and hence shape-selective zeolite catalyzed hydroisomerization,[3] often called catalytic dewaxing, is likely to become a key step in the production of iso-alkanes, as it currently is for the classical analogue products.[4,5] To design processes, catalysts and equipment handling, reacting and separating (iso)alkanes and their mixtures, a detailed understanding of the thermochemical properties of these compounds, such as enthalpy, Gibbs free energy, entropy, heat capacity, and fugacity, is necessary.[6] To compute a reliable product distribution from hydroisomerization reactions, accurate prediction of thermochemical properties of alkanes is of utmost importance.[7] Another example of interest is alkane-based phase change materials for thermal energy storage systems.[8]

Thermochemical properties of all isomers until C_{10} are reported by Scott.[9] The work of Scott lists Gibbs free energies ($G_0 - H_0(0\ \mathrm{K})$), enthalpies ($H_0 - H_0(0\ \mathrm{K})$), absolute entropies S_0, Gibbs free energies of formation $\Delta_f G_0$, enthalpies of formation $\Delta_f H_0$, and constant pressure heat capacities c_{p_0} of these isomers at temperatures (0−1500) K. These properties are obtained from a correlation developed using statistical mechanics, which have been trained on experimental data.[10] The experimental data for thermochemical properties of different alkanes available in literature are reported in the

Received: August 8, 2024
Revised: September 13, 2024
Accepted: September 16, 2024
Published: September 23, 2024



https://doi.org/10.1021/acs.jpcb.4c05355
J. Phys. Chem. B 2024, 128, 9619−9629

NIST chemistry webbook.[11] Limited amount of experimental data can be found for alkanes longer than C_{10}.[12,13] Thermochemical properties of linear alkanes until C_{20} are listed in refs 14,15. Group contribution methods are commonly used to predict thermochemical properties of long chain alkanes.[16] These methods are additive in nature, where the structure of a molecule is fragmented into functional groups and the thermochemical properties are estimated by summing up the contributions of each functional group present in the molecule.[16] The contribution of an individual group remains the same in every molecule it appears.[16] A wide range of group contribution methods for thermochemical properties of organic molecules exists in literature[17] which includes methods developed by Benson et al.,[18] Joback and Reid,[19] Constantinou and Gani,[20] Marrero and Gani,[21] Hukkerikar et al.,[22] Albahri and Aljasmi,[23] and Domalski and Hearing.[24] Benson's group additivity method[18] is commonly used to predict $\Delta_f H_0$ and S_0.[25] Yaw's hand book[26] lists the entropies of formation $\Delta_f S_0$, $\Delta_f H_0$, $\Delta_f G_0$, and c_{p_0} in gas, liquid, and solid phases for a wide range of alkanes until C_{100}. The properties in this hand book are either collected from experimental data available in literature or predicted using Joback and Reid's method,[19] especially for long chain alkanes. The Design Institute for Physical Properties (DIPPR) database[27] also lists the thermochemical properties of a large number of alkanes obtained from quantum chemical calculations, experiments, and group contributions like Benson's method.[18] An alternative to predict thermochemical properties is using Machine Learning (ML) models. Yalamanchi et al. predicted $\Delta_f H_0$ of alkanes, alkenes, and alkynes at 298.15 K using a Support Vector Regression (SVR) model which provided better prediction than Benson's group additivity when compared with experimental data.[25] Trinh et al. also used an SVR model to predict $\Delta_f H_0$ of a wide variety of organic compounds. The training data set was obtained from the DIPPR database.[28] Aldosari et al. predicted $\Delta_f S_0$ $\Delta_f H_0$ and c_{p_0} of hydrocarbons using SVR, v-SVR, and Random Forest Regression (RFR) algorithms.[29] Alternatively, one could also consider a High-Dimensional Model Representation (HDMR) for the longer hydrocarbons.[30] As group contribution methods are more common, we opted for this approach.

Most of the group contribution methods available in literature predict thermochemical properties for a wide range of hydrocarbons. These methods are not always accurate due to either considering only first order groups ($-CH_3$, $-CH_2$, $-CH$, -C) or combining first order group contributions with a very few second order groups.[25] Figure 1 shows typical examples of second order groups (a) $CH_2(CH3)(CH)$ and (b) $CH_2(CH)(CH)$ which consider the interaction between the central atom (here CH_2) and the neighboring united atoms. The united atoms inside the brackets are the first order neighboring groups. Considering only the first order groups leads to less reliable prediction for highly branched isomers. Increasing the number of second order groups in a group contribution method provides a better prediction of the thermochemical properties. Thermochemical property predictions using Scott's correlation[10] based on statistical mechanics are in very good agreement with experimental data for chains up to C_{10} isomers. This correlation is very complex and involves many different types of functions and fitting parameters, which makes it hard to apply for long chains ($>C_{10}$) that were not in the training set. ML based predictions

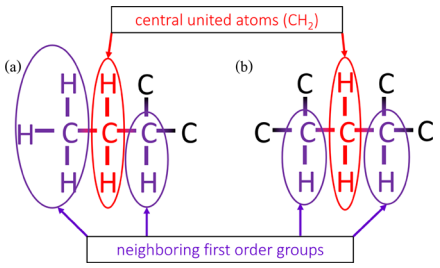


Figure 1. Typical examples of second order groups (a) $CH_2(CH_3)$-(CH) and (b) $CH_2(CH)(CH)$ with CH_2 as the central united atom. The united atoms inside the brackets are the neighboring first order groups. In first order group contribution methods, only the central united atom is considered which is CH_2 in cases (a,b). In both cases, first order group contributions will be identical and lead to inaccurate prediction of thermochemical properties. Unlike first order group contribution methods, second order group contribution methods consider the interactions between the central atom (here CH_2) and the neighboring groups which leads to more accurate prediction of thermochemical properties.

have the potential to provide better predictions compared to group contribution methods. However, ML could perform poorly in case of extrapolation[29,31] which is required for long chain alkanes due to the absence of experimental data. If sufficient number of independent variables are considered and the output is linearly dependent on these variables, Linear Regression (LR) will perform better than the ML models in the extrapolated region.[31] In this study, an user-friendly LR model is developed, where the occurrences of all possible second order groups are considered as independent variables to accurately predict the thermochemical properties of alkanes longer than C_{10}.

Here, we aim to predict the thermochemical properties ($G_0 - H_0(0\ K)$), ($H_0 - H_0(0\ K)$), $\Delta_f G_0$, and $\Delta_f H_0$ of alkanes longer than C_{10} isomers using LR with second order groups as molecule descriptors. The training data set includes the ideal gas thermochemical properties of C_1 to C_{10} isomers at temperatures ranging from (0−1000) K, listed by Scott.[9] LR is performed at each temperature separately. To account for the effect of temperature, a quadratic polynomial as a function of temperature is fitted to the coefficients of the second order groups for each thermochemical property. The predictions using the second order group contribution method outperform the first order group contribution methods and exceed the chemical accuracy of 1 kcal/mol. This is because the second order groups include the interactions between the neighboring groups of atoms. This study also aims toward using the predicted thermochemical properties $\Delta_f H_0$ at 0 K and ($G_0(T) - H_0(0\ K)$) to compute the reaction equilibrium distribution of hydroisomerization of alkanes longer than C_{10} isomers at that temperature.

To optimize the yield of branched isomers in hydroisomerization, it is important to understand the reaction product distribution at chemical equilibrium.[7,32] Hydroisomerization reactions involves adsorption of linear alkanes and dehydrogenation of these alkanes in the metal sites of the zeolites forming alkenes.[33] Protonation of alkenes take place at the acid sites of the zeolites to form alkylcarbenium ions.[34] These ions are transferred to the metal sites where alkanes are produced via hydrogenation. This indicates that alkenes as intermediates can play an important role at reaction equilibrium. However, due to the absence of alkenes in the

https://doi.org/10.1021/acs.jpcb.4c05355
J. Phys. Chem. B 2024, 128, 9619−9629

final product distribution and lack of experimental data of thermochemical properties of alkenes, the reaction equilibrium of only alkanes is considered in this study. In our previous study, Sharma et al.[7] studied the shape selectivity effects of zeolites on the reaction equilibrium distribution of hydroisomerization of C_7 and C_8 isomers. It was shown that the reaction equilibrium distribution of alkanes is useful for understanding the shape selectivity effects of zeolites on hydroisomerization of alkanes. The reaction equilibrium distribution of this reaction is determined by establishing reaction equilibrium in the gas phase and phase equilibrium between the gas and the adsorbed phase.[7] For applications such as production of sustainable aviation fuels, long chain alkanes (e.g., C_{16}) with high degree of branching[2] are desirable because of high energy density, low freezing point, and good thermal stability.[35] The LR model is used to predict $\Delta_f H_0$ at 0 K and $(G_0 - H_0(0\ \mathrm{K}))$ at a specified temperature which are used to compute the gas phase distribution, and classical force field-based simulations are used to quantify the interactions between isomers and the zeolite. To automate the workflow and handling of a large number of isomers, alkanes are represented using SMILES strings.[36,37] A Python function is developed to generate SMILES strings from the IUPAC names of alkanes which is required as an input in the source code for the LR model (Supporting Information SI2.py). The reaction equilibrium distribution inside constraining pore zeolites differs significantly from the gas phase distribution.[7] At infinite dilution, the reaction equilibrium distribution is strongly influenced by Henry coefficients of alkanes in constraining pore zeolites such as MTW-type zeolite with pore diameters 5.6 × 6.0 Å[38] (Figure S1 in the Supporting Information SI5.pdf).[7] The computation of Henry coefficients using classical force field-based Monte Carlo simulations requires interaction terms as input to account for both bonded and nonbonded interactions of alkanes and also nonbonded interactions between the zeolite atoms and the alkanes. A source code (Supporting Information SI3.py) for automated force field file generation for use with the RASPA2 software[39,40] is developed to avoid the manual entry of a large number of interaction terms, which is especially useful for long chain alkanes. This code also uses the Python function to generate SMILES strings for alkanes. Using the values for $\Delta_f H_0$ at 0 K and $(G_0 - H_0(0\ \mathrm{K}))$ at 500 K obtained from the LR model and Henry coefficients of alkanes calculated using classical Monte Carlo simulations in the RASPA2 software,[39,40] the reaction equilibrium distributions of hydroisomerization of C_{10} and C_{14} isomers in MTW-type zeolite at infinite dilution are computed. The reaction equilibrium distribution of C_{10} isomers in MTW-type zeolite computed using the thermochemical properties obtained from our LR model and the training data set are in very good agreement. This suggests that the thermochemical properties predicted using the LR model can be reliably used to compute reaction equilibrium distribution of hydroisomerization of long chain alkanes. In this study, hydroisomerization of C_{14} is shown as an example for long chain alkanes. In future studies, the reaction equilibrium distribution of hydroisomerization of alkanes longer than C_{14} will be analyzed in constraining pore zeolites such as MTW- and MRE-type zeolites.

This article is organized as follows. The important concepts and the theory behind linear regression, and simulation details are provided in Section 2. Our main results are discussed in Section 3. It is observed that LR with second order group contributions outperforms methods based on first order group contributions. The variations in the thermochemical properties due to the differences in branching patterns of isomers are well captured by this method, in sharp contrast to the predictions using group contribution methods available in literature. The coefficients of the occurrences of the second order groups of each thermochemical property are fitted using temperature dependent quadratic polynomials. In Section 4, conclusions on the performance of the LR model in predicting different thermochemical properties of long chain alkanes and the use of these properties for computing reaction equilibrium distribution of hydroisomerization reactions are discussed. The predicted thermochemical properties using this LR model can be reliably used to compute reaction equilibrium distribution of hydroisomerization of alkanes longer than C_{10}. This article also contains Supporting Information SI1.xlsx, SI2.py, SI3.py, SI4.xlsx, and SI5.pdf. The training data set for the thermochemical properties of all isomers ranging from C_1 to C_{10} are listed in SI1.xlsx. In the Supporting Information SI2.py, the source code for the automatic generation of force field files to compute Henry coefficients in the RASPA2 software[39,40] is included. The source code for the LR model is provided in SI3.py. SI4.xlsx contains the predicted thermochemical properties of all isomers until C_{14} molecules with temperature ranging from (0−1000) K. The coefficients of the occurrences of the second order groups obtained using LR and the corresponding temperature dependent quadratic polynomial fits are also listed in SI4.xlsx. The reaction equilibrium distribution data of hydroisomerization of C_{10} and C_{14} molecules in the gas phase and MTW-type zeolite at 500 K are tabulated in SI4.xlsx. This includes the ideal gas chemical potentials, Henry coefficients, and selectivities of these isomers in both the gas and the adsorbed phase. SI5.pdf contains figures showing the pore structure in MTW-type zeolite, $\Delta_f H_0$ values from Scott's tables, our LR model, and the DIPPR database. SI5.pdf also shows the variations of the coefficients of second-order groups with C as the central atom for $(G_0 - H_0(0\ \mathrm{K}))$ at different temperatures and the reaction equilibrium distribution imposed on the gas phase for C_{10} isomers.

2. THEORY

An LR model is used to predict thermochemical properties of alkanes longer than C_{10} with the occurrences of the first or the second order groups as independent variables as shown below. This linear regression is performed using the SciPy library in Python[41]:

$$y = \alpha_0 + \sum_{k=1}^{N_{\mathrm{descriptor}}} \alpha_k x_k \tag{1}$$

In eq 1, y is the thermochemical property predicted using LR, x_k is the independent variable which refers to the occurrence of a first or a second order group k in a molecule, and $N_{\mathrm{descriptor}}$ is the total number of such independent variables or descriptors. α_0 is the intercept of the equation for LR. α_k refers to the coefficients of the independent variables. In first order group contribution methods, the occurrences of the united atoms CH_4, CH_3, CH_2, CH, and C are considered as independent variables. In second order group contribution methods, the combination of the central united atom and its nearest neighbors is considered as an independent variable. There are 69 second order groups for alkanes which are listed

https://doi.org/10.1021/acs.jpcb.4c05355
J. Phys. Chem. B 2024, 128, 9619−9629

in the Excel worksheet Second_order_grps of the Supporting Information SI1.xlsx. In addition to these groups, CH_4 is also included in the second order group contribution to predict the thermochemical properties of methane.

The training data set contains isomers in the range C_1−C_{10}. This data set includes the occurrences of only 46 second order groups. The remaining second order groups are present in alkanes of chain length longer than C_{10}. The list of the second order groups which are not present in the training data set along with the smallest isomers which contain these groups are shown in the worksheet Second_order_grps of the Supporting Information SI1.xlsx. To account for the contribution of these groups in the longer chains, these groups are approximated by a similar second order group present in the training data set. For example, the second order group $C(C)(CH_2)(CH_2)$-(CH_2) is approximated by $C(C)(CH_2)(CH_2)(CH_3)$. The list of all such approximations are also included in the Excel worksheet Second_order_grps of the Supporting Information SI1.xlsx. These second order groups correspond to highly branched isomers which are unlikely to form inside constraining pore zeolites like MTW-type zeolite during hydroisomerization of alkanes. This is due to constraining pores present in MTW-type zeolite and steric hindrance caused by the proximity of the branches present in these isomers.[7] Therefore, approximate predictions of the thermochemical properties of such alkanes are sufficient to compute the reaction equilibrium distribution of all hydroisomerization reactions inside zeolites.

Linear regressions for the thermochemical properties are performed at a specific temperature and the coefficients of the second order groups (α_k) obtained from this model are specific to that temperature. To include the effect of temperature, these coefficients are refitted to a temperature dependent quadratic polynomial.

$$\alpha = A \times T^2 + B \times T + C \tag{2}$$

In eq 2, A, B, and C are constants and T is the temperature in K. The values of the coefficients for the thermochemical properties $\Delta_f H_0$, $\Delta_f G_0$, $(G_0 - H_0(0\ \mathrm{K}))$, and $(H_0 - H_0(0\ \mathrm{K}))$ are listed in the Excel worksheets DHf0_coeff, DGf0_coeff, G0-H0(0K)_coeff, and H0−H0(0K)_coeff of the Supporting Information SI2.xlsx, respectively. The corresponding temperature dependent polynomials are listed in the Excel worksheets DHf0_coeff_poly, DGf0_coeff_poly, G0-H0(0K)_coeff_poly, and H0−H0(0K)_coeff_poly of SI4.xlsx. The source code to predict thermochemical properties of alkanes using LR is provided in the Supporting Information SI2.py. To handle a large number of isomers during linear regression, alkanes are represented as SMILES strings.[36,37] This code includes a function to generate SMILES strings for alkanes with maximum of 5-carbon alkyl branches. The thermochemical properties, $\Delta_f H_0$ at 0 K and $(G_0 - H_0(T_{ref}))$ at a specified temperature are used to compute ideal gas chemical potentials[42−44] which are further used to compute the reaction equilibrium distribution of hydroisomerization of alkanes. The ideal gas chemical potential of component i equals[42,43]

$$\begin{aligned}\mu_{id,i} &= \mu_{ref,i} + RT\ln\left(\frac{\rho_i}{\rho_0}\right)\\ &= (G_{0,i}(T) - H_{0,i}(T_{ref})) - D_{0,i} + RT\ln\left(\frac{\rho_i}{\rho_0}\right)\end{aligned} \tag{3}$$

In eq 3, $\mu_{ref,i}$ is the reference chemical potential of component i, ρ_i is the number density of component i, and ρ_0 is the reference density which is chosen to be 1 molecule/Å^3. T_{ref} is the reference temperature which is 0 K in this study. $D_{0,i}$ is the atomization energy[44] which is

$$D_{0,i} = a_C \Delta_f H_{0,C}\ (0\ \mathrm{K}) + a_H \Delta_f H_{0,H}\ (0\ \mathrm{K}) - \Delta_f H_{0,i}\ (0\ \mathrm{K}) \tag{4}$$

In eq 4, a_C is the number of C atoms and a_H is the number of H atoms present in the alkane isomer i. $\Delta_f H_{0,C}$, $\Delta_f H_{0,H}$, and $\Delta_f H_{0,i}$ are the enthalpies of formation of the C atom, the H atom, and alkane isomer i, respectively, at 0 K. The value of $\Delta_f H_{0,C}$ is 711.185 kJ/mol and $\Delta_f H_{0,H}$ is 216.035 kJ/mol at 0 K which are obtained from the JANAF tables.[45] $\Delta_f H_{0,i}$ values of alkanes are provided in the Supporting Information SI1.xlsx.

The reaction equilibrium distribution of hydroisomerization reactions in zeolites can be computed by imposing reaction equilibrium in the gas phase and phase equilibrium between the gas and the adsorbed phase.[7,32,46] At reaction equilibrium, the chemical potentials of the reactants and the reaction products are equal in the gas phase.[7] Equating the ideal gas chemical potentials of the reactants and the reaction products leads to the gas phase reaction equilibrium distribution at infinite dilution. The equilibrium loadings in the adsorbed phase are computed using Henry's law at infinite dilution and mixture adsorption isotherm models such as Ideal Adsorbed Solution Theory (IAST)[47,48] at finite loadings. At high temperatures (≥500 K), the effect of pressure on the reaction equilibrium distribution of hydroisomerization of alkanes is negligible.[7] High temperature leads to negligible variations in the gas phase distribution of alkane isomers with pressure and a decrease in the amount of molecules adsorbing in the zeolites.[7] Therefore, only infinite dilution is considered in this study. For details on computing reaction equilibrium distribution of hydroisomerization, the reader is referred to our previous study.[7]

The required Henry coefficients are computed using the Widom test particle insertion method[49] combined with the Configurational-Bias Monte Carlo (CBMC) method[50−53] in the RASPA2 software.[39,40] Alkanes are modeled using a united-atom model[54] and the Coulomb interactions are neglected because alkanes are nonpolar.[39] The intramolecular nonbonded interactions of alkanes and the intermolecular nonbonded interactions between the alkanes and the zeolite atoms are modeled using the Lennard-Jones interactions.[55] The Lennard-Jones parameters for alkanes are obtained from Dubbeldam et al.[56] The Lennard-Jones parameters for the zeolite atoms are taken from the TraPPE-zeo force field.[57] The intramolecular bonded interactions which include bond-stretching, bond-bending, and torsion interactions are obtained from refs 58,59. Both bonded and nonbonded interaction parameters are listed in the Excel worksheet force field_param of the Supporting Information SI1.xlsx. Files containing force field parameters and the list of bonded and nonbonded

https://doi.org/10.1021/acs.jpcb.4c05355
J. Phys. Chem. B 2024, 128, 9619−9629

interactions are required as input in the RASPA2 software.[39,40] The number of intramolecular interactions increases tremendously with increasing chain lengths of alkanes (e.g., n-C_{14} contains 91 intramolecular interactions). Therefore, a Python code for automatic generation of force field files for alkanes is provided in the Supporting Information SI3.py. This code requires alkanes to be represented as SMILES strings.[36,37] Therefore, this code also includes the Python function to generate SMILES strings for alkanes. All Lennard-Jones interactions are truncated and shifted at 12 Å without applying tail corrections. The number of unit cells in the simulation box of MTW-type zeolite are 2 × 15 × 2 for C_{10} isomers and 2 × 25 × 2 for C_{14} isomers. The zeolite is considered to be a rigid structure as the effect of zeolite flexibility is negligible on adsorption processes, especially at infinite dilution.[60]

The selectivity of a component in the gas phase or the adsorbed phase is computed as[61]

$$s_i^{\text{gas}} = \frac{y_i}{\left(\sum_{n=1}^{N} y_n\right) - y_i} \tag{5}$$

The selectivity of a component (eq 5) in both gas phase and adsorbed phase is defined as the ratio of the mole fraction y_i of the component to the sum of the mole fractions of all other components present in the same phase. To compare the selectivity of a component relative to another, the term relative selectivity ($s_{\text{rel},i}$) is defined as the ratio of the absolute selectivity of that component to a reference component as shown in eq 6 for both gas and adsorbed phase.

$$s_{\text{rel},i} = \frac{s_i}{s_{\text{ref}}} = \frac{\frac{y_i}{\left(\sum_{n=1}^{N} y_n\right) - y_i}}{\frac{y_{\text{ref}}}{\left(\sum_{n=1}^{N} y_n\right) - y_{\text{ref}}}} \tag{6}$$

In this study, n-C_{10} and n-C_{14} molecules are chosen as reference components for computing relative selectivities.

3. RESULTS AND DISCUSSION

Figure 2 shows the comparison between $(G_0 - H_0(0\text{ K}))$ of C_7 isomers at 400 K predicted by the LR model using first and second order group contribution methods. The predicted values are also compared with the data obtained from the tables by Scott.[9] The thermochemical properties obtained using the second order groups as descriptors are in very good agreement with the training data set (Figure 2) These predictions are much better than those obtained using the first order group contributions. This is because the influence of the neighboring groups of atoms are neglected in the first order group contribution which is clearly shown in Figure 2.

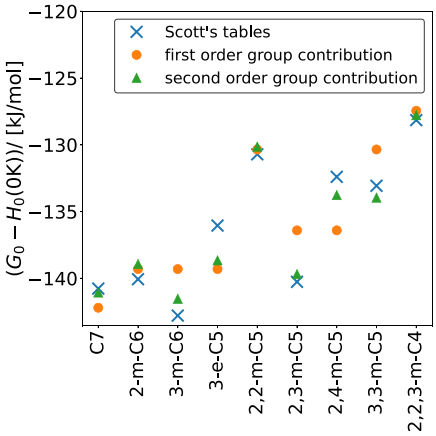


Figure 2. Comparison between the thermochemical property $(G_0 - H_0(0\text{ K}))$ predicted using the LR model with first and second order group contribution methods for C_7 isomers at 298.15 K. The predictions using the second order group contributions are in very good agreement with the data from the Scott's tables.[9]

The Mean Absolute Errors (MAEs) for the prediction of the thermochemical properties of alkanes using the first and the second order group contributions are listed in Table 1. The use of the second order groups as descriptors leads to smaller MAEs compared to the first order group contributions. Therefore, only second order group contributions are considered in this study. For comparison, the values of $\Delta_f H_0$ at 400 K for a few selected C_{10} isomers obtained from the training data set[9] and computed using the LR model are listed in Table 2. The values computed using the LR model and those obtained from the training data set are in excellent agreement.

Table 1. Mean Absolute Errors (MAEs) of the Thermochemical Properties $\Delta_f G_0$, $\Delta_f H_0$, $(G_0 - H_0(0\text{ K}))$, $(H_0 - H_0(0\text{ K}))$ Predicted Using Linear Regression (LR) with the First and the Second Order Group Contributions

thermochemical property	MAE (first order)/ (kJ/mol)	MAE (second order)/ (kJ/mol)
$\Delta_f G_0$	7.529	1.029
$\Delta_f H_0$	5.834	0.152
$(G_0 - H_0(0\text{ K}))$	3.087	1.012
$(H_0 - H_0(0\text{ K}))$	1.041	0.181

Table 2. Comparison between the Values of Enthalpy of Formation $\Delta_f H_0$ for C_{10} Isomers at 400 K Obtained from the Training Dataset[9] and the Linear Regression Model

	$\Delta_f H_0$/(kJ/mol)	
isomer	training data	linear regression data
n-C_{10}	−265.47	−265.47
4-m-C_9	−270.04	−270.07
4-e-C_8	−267.48	−267.49
2,2-m-C_8	−282.55	−282.56
3-e-4-m-C_7	−266.60	−266.53
2,2,5-m-C_7	−287.19	−287.09
3,3-e-C_6	−266.10	−266.08
4-e-3,3-m-C_6	−261.75	−261.76
2,2,3,5-m-C_6	−284.68	−284.57
3-e-2,2,4-m-C_5	−261.37	−261.37
2,2,3,4,4-m-C_5	−261.54	−261.54

Several group contribution methods are available in literature such as Benson's,[18] Constantinou and Gani's,[20] Joback's,[19] and Domalski and Hearing's[24] methods which either consider first order group contributions or a combination of first order groups and a few second order groups. Figure 3 shows the variations in $\Delta_f H_0$ of a few C_9 isomers at 298.15 K predicted using the LR model, Benson's,[18] and Constantinou and Gani's methods.[20] The properties predicted using Benson's and Constantinou and Gani's methods are computed using the SPLIT software by AmsterChem.[62] Figure 3 also includes data obtained from

https://doi.org/10.1021/acs.jpcb.4c05355
J. Phys. Chem. B 2024, 128, 9619−9629

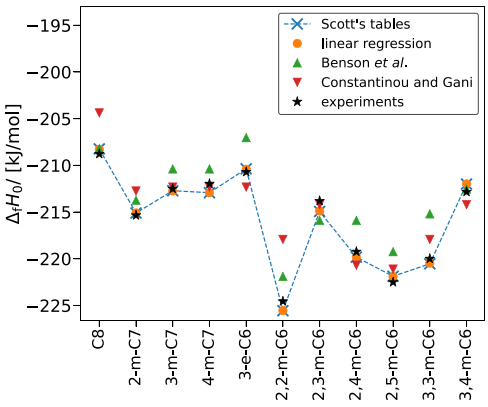


Figure 3. Prediction of $\Delta_f H_0$ for C_8 isomers at 298.15 K using the LR model, Benson et al.'s group additivity method,[18] Constantinou and Gani's group contribution,[20] the Scott's tables,[9] and the experimental data listed by Scott.[10] The predictions using the LR model are in excellent agreement with the Scott's tables and the data from the experiments[10] For the $\Delta_f H_0$ values of C_8 isomers at 298.15 K, using the experimental data as the reference, the Mean Absolute Errors (MAEs) are the smallest for our LR model (0.62 kJ/mol) and the Scott's Tables (0.64 kJ/mol). This is followed by Constantinou and Gani's group contribution method (2.04 kJ/mol), and Benson et al.'s group additivity method (2.37 kJ/mol) respectively. The dashed blue line through the data points obtained from the Scott's tables is a guide to the eye.

the Scott's tables,[9] and Yaws' handbook.[26] Yaws' handbook[26] uses Joback's group contribution method[19] for long chain isomers. Benson's[18] and Constantinou and Gani's[20] group contribution methods are not always able to distinguish between isomers based on the positions and the types of branches these isomers possess. For example, the Constantinou and Gani's method[20] provided nearly identical values of $\Delta_f H_0$ for 3-m-C_8, 4-m-C_8, 3-e-C_7, and 4-e-C_7 whereas experimental data from Scott's tables clearly shows variations due to the presence of the methyl and the ethyl groups as branches in these isomers. Similarly, Benson's method[18] does not distinguish between 2,3-m-C_7, 2,4-m-C_7, and 2,5-m-C_7. Such variations are well captured by our LR model and are necessary to determine the selectivities of reaction products in hydroisomerization of alkanes.[7] $\Delta_f H_0$ values of a few C_9 and C_{10} isomers at 298.15 K obtained from the Scott's tables and predicted by LR are also compared with those listed in the DIPPR database[27] as shown in Figure S2 in the Supporting Information SI5.pdf. $\Delta_f H_0$ of C_{10} isomers listed in the DIPPR database[27] are computed using the Domalski and Hearing's method.[24] The values predicted using our LR model are in excellent agreement with the experimental data from the Scott's tables. The values obtained from the DIPPR database[27] are also in good agreement with the Scott's tables[9] with small deviations for 4-m-C_8, 2-m-C_9, and 3-m-C_9.

LR is performed at a specific temperature for each thermochemical property. To account for the effect of temperature on the thermochemical properties and compute these properties at any temperature, the coefficients which correspond to the occurrence of the second order groups are fitted to a temperature dependent quadratic polynomial (eq 2). Figure S3 in the Supporting Information SI5.pdf shows the variations in the magnitudes of the coefficients of the occurrences of the second order groups for $(G_0 - H_0(0\ \mathrm{K}))$ with C as the center atom present in the training data set. The fitted coefficients using the quadratic polynomial are in excellent agreement with those predicted using the LR model. The values of the coefficients are different at a specific temperature (Figure S3 in the Supporting Information SI5.pdf). In case of first order group contributions, the variations will be identical as each group has the same central united atom C. Therefore, combining these coefficients into a single coefficient to reduce the number of independent variables or simply using first order group contributions will lead to erroneous predictions of the thermochemical properties. This clearly indicates the need for a second order group contribution method.

The predicted thermochemical properties $(G_0 - H_0(0\ \mathrm{K}))$ at a specified temperature and $\Delta_f H_0$ at 0 K for alkanes longer than C_{10} are used to compute the ideal gas chemical potentials which are further used in calculating the reaction equilibrium distribution of hydroisomerization of long chain alkanes. Figure S4 in the Supporting Information SI5.pdf shows the reaction product distribution of C_{10} isomers in the gas phase at infinite dilution and 500 K. The reaction product distribution obtained using LR is in very good agreement with the training data set (Figure S4 in the Supporting Information SI5.pdf). In both cases, the variations in the selectivities of C_{10} isomers relative to n-C_{10} are similar. The gas phase distribution and Henry coefficients are used to compute the reaction equilibrium distribution in MTW-type zeolite at infinite dilution and 500 K (Figure 4). For monobranched isomers, 4-p-C_7 and 4-ip-C_7 have the smallest preferences in MTW-type zeolite as shown in Figure 4b. This indicates that isomers with branches longer than propyl or isopropyl groups such as butyl and pentyl groups will have even lower selectivities in MTW-type zeolite. Such isomers can be eliminated form the calculation of the reaction equilibrium distribution of hydroisomerization of long chain alkanes. The ethyl-trimethyl isomers (3-e-2,2,3-m-C_5, 3-e-2,2,4-m-C_5, and 3-e-2,3,4-m-C_5) and pentamethyl isomers (2,2,3,3,4-m-C_5 and 2,2,3,4,4-m-C_5) have the least preference compared to all other isomers. This indicates that isomers with multiple branches present in the vicinity of each other are not preferably formed inside constraining pore zeolites such as MTW-type zeolite. Reaction product distributions obtained using both LR and Scott's tables provide very similar variations in selectivities. This suggests that the thermochemical properties obtained using LR are reliable and can be used in computing the reaction equilibrium distribution for hydroisomerization of long chain alkanes.

Figure 5 shows the reaction product distribution of C_{14} isomers in MTW-type zeolite at infinite dilution and 500 K. Monomethyl branched isomers have higher preference compared to the monoethyl and dimethyl isomers. This is because of the smaller pore diameters present in MTW-type zeolites (5.6 × 6.0) Å.[38] The order of magnitude of relative selectivities varies for dibranched isomers which depends on the proximity of the methyl branches. The geminal alkanes (2,2-m-C_{12}, 3,3-m-C_{12}, 4,4-m-C_{12}, 5,5-m-C_{12}, and 6,6-m-C_{12}) have the least selectivity due to steric hindrance posed by the presence of two methyl groups at the same position in the alkane chain. Isomers with methyl groups far apart (2,10-m-C_{12} and 2,11-m-C_{12}) have the highest selectivities compared to other dimethyl isomers. In future studies, such criteria will be used to identify relevant isomers for computing reaction equilibrium distribution of alkanes longer than C_{14}. Such filtering of isomers is necessary because the number of isomers

https://doi.org/10.1021/acs.jpcb.4c05355
J. Phys. Chem. B 2024, 128, 9619−9629

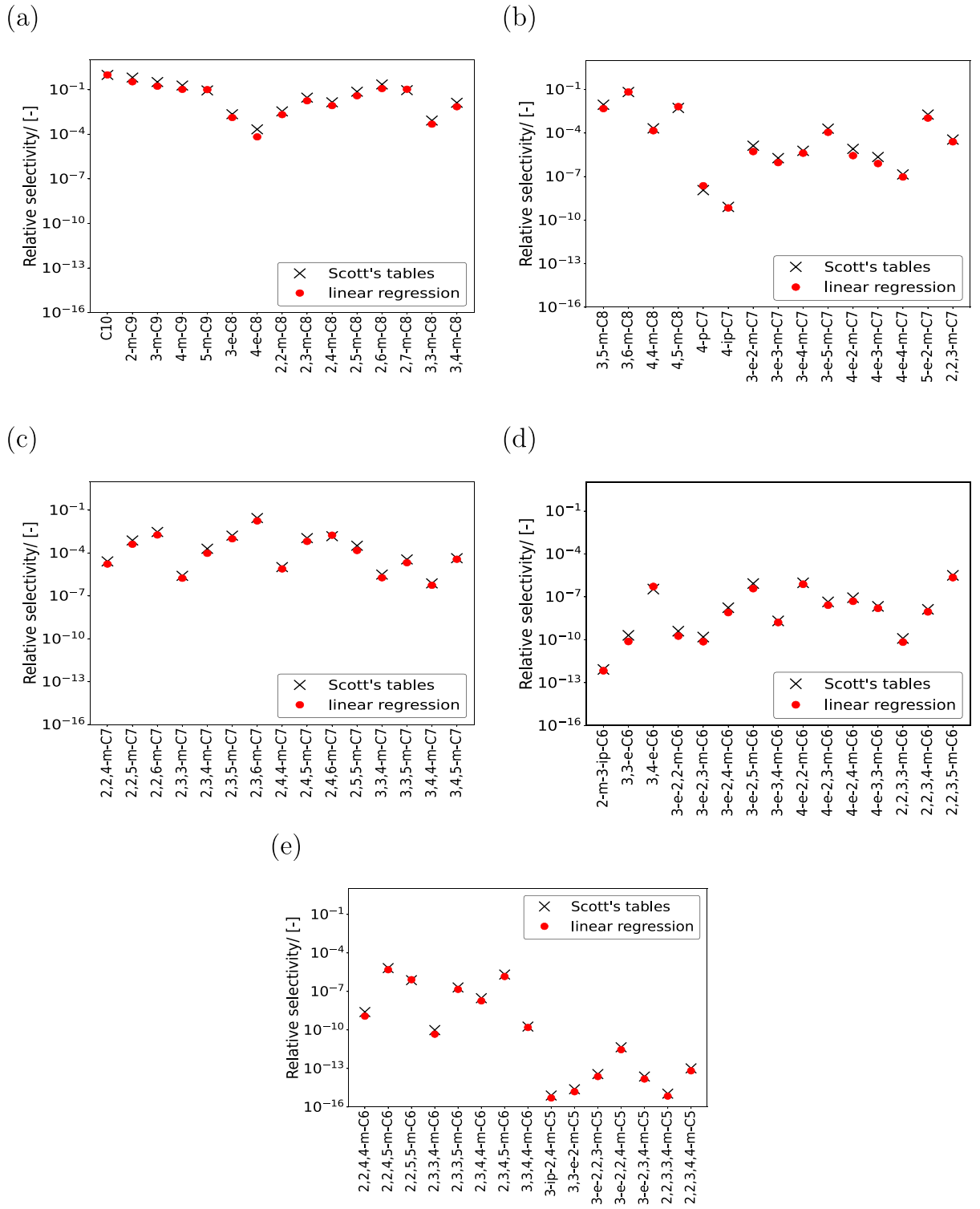


Figure 4. Selectivities of C_{10} isomers relative to n-C_{10} at reaction equilibrium in MTW-type zeolite at infinite dilution and 500 K. The gas phase reaction equilibrium distribution required in computing the adsorbed phase reaction equilibrium distribution are obtained using the Scott's tables (black crosses) and the LR model (red filled circles). The raw data is listed in the Excel worksheet xi_iC10_500 K of the Supporting Information SI4.xlsx.

increases enormously for alkanes longer than C_{14} and it is difficult to consider each isomer experimentally.

4. CONCLUSIONS

A linear regression model with second order group contributions is developed to predict the thermochemical properties $\Delta_f H_0$, $\Delta_f G_0$, $(G_0 - H_0(0\ \mathrm{K}))$, and $(H_0 - H_0(0\ \mathrm{K}))$ of alkanes. The predicted properties are in excellent agreement with the Scott's tables[9] and exceed the chemical accuracy of 1 kcal/mol. The maximum mean absolute error is observed for $\Delta_f G_0$ which is 1.03 kJ/mol. The second order group contribution method outperforms first order group contribu-

https://doi.org/10.1021/acs.jpcb.4c05355
J. Phys. Chem. B 2024, 128, 9619−9629

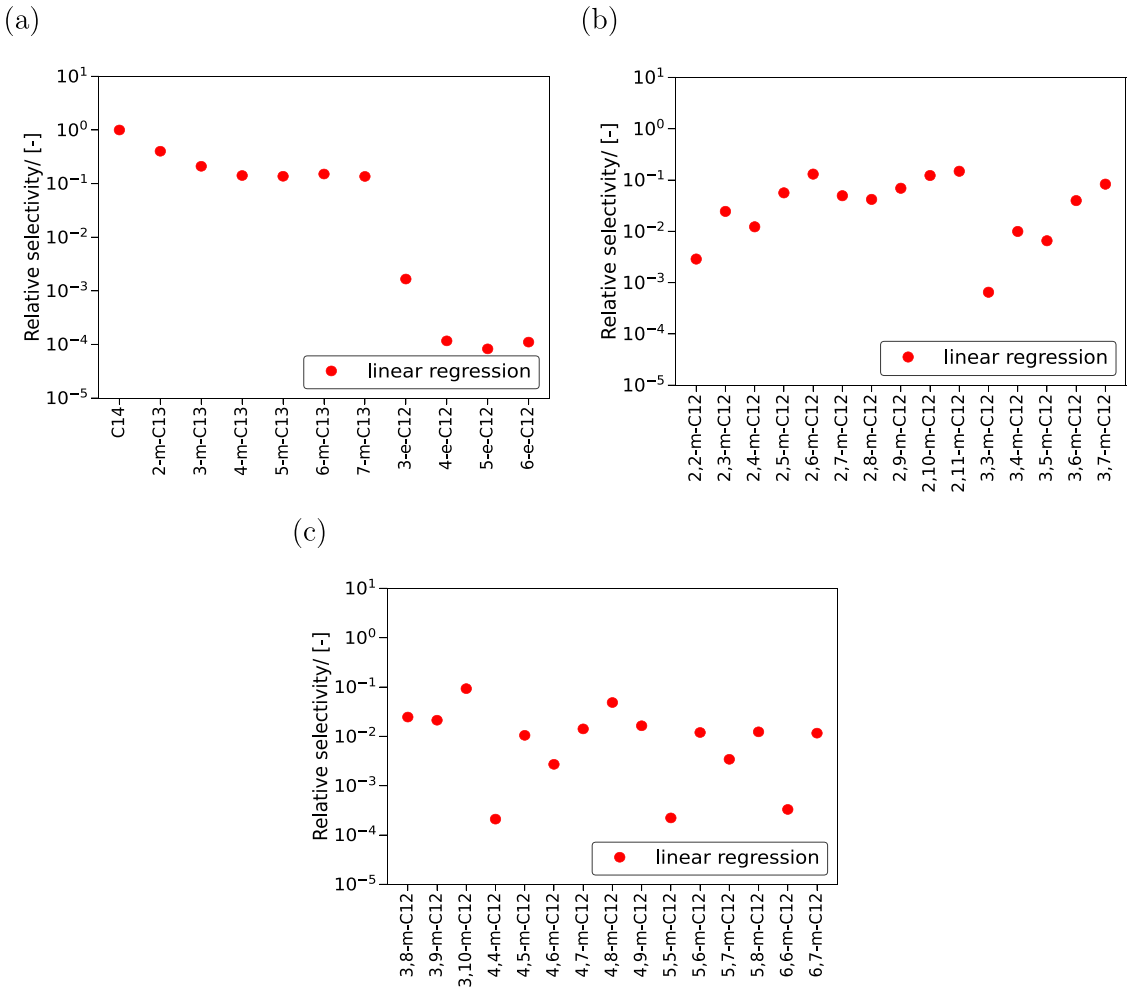


Figure 5. Selectivities of different C_{14} isomers relative to n-C_{14} at reaction equilibrium in MTW-type zeolite at infinite dilution and 500 K. The LR model is used to predict $(G_0 - H_0\ (0\ \mathrm{K}))$ at 500 K and $\Delta_f H_0$ at 0 K which are used to compute the ideal gas chemical potentials of C_{14} isomers. These chemical potentials are used to compute the gas phase reaction equilibrium distribution of hydroisomerization of C_{14}. The raw data is listed in the Excel worksheet xi_iC14_500 K of the Supporting Information SI4.xlsx.

tion methods for predicting these properties. Our LR model performs better than many existing group contribution methods in literature such as the Benson's,[18] Constantinou and Gani's,[20] and Joback's methods.[19] The influence of temperature is considered by fitting the coefficients of the occurrences of the second order groups with a temperature dependent quadratic polynomial. The reaction equilibrium distributions computed using the LR model and the Scott's tables[9] in combination with automatic computation of Henry coefficients from RASPA2[39,40] are in very good agreement for the hydroisomerization of C_{10} isomers in MTW-type zeolite. This indicates that the LR model can be used to compute reaction equilibrium distribution of long chain alkanes in zeolites. The isomers with propyl and the isopropyl groups as branches have low selectivity in MTW-type zeolites. This indicates that isomers with branches longer than propyl group such as butyl or pentyl groups will have very low selectivity and can be neglected from reaction product distribution. The reaction equilibrium distribution for the hydroisomerization of C_{14} isomers in MTW-type zeolite is also computed using the thermochemical properties predicted by the LR model. Isomers with branches far apart (e.g., 2,10-m-C_{12} and 2,11-m-C_{12}) have larger selectivities compared to isomers with branches present close to each other (2,2-m-C_{12} and 3,3-m-C_{12}). Such criteria are necessary to exclude isomers with very low selectivities from the analysis of reaction equilibrium distributions of alkanes longer than C_{14}. In future studies, hydroisomerization of alkanes longer than C_{14} will be considered for which the thermochemical properties will be computed using our LR model. As this involves a large number of isomers, automatic generation of force field files and other input files for RASPA2 will be essential (SI3.py).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.4c05355.

Contains Supporting Information (SI1.xlsx, SI2.py, SI3.py, SI4.xlsx, and SI5.pdf); training data sets of the thermochemical properties of all isomers ranging from C_1 to C_{10} listed in SI1.xlsx; list of the second-order groups present in the training data set listed in the Supporting Information (SI1.xlsx); force field parameters to compute Henry coefficients of alkanes in zeolite listed in SI1.xlsx; source code for the LR model listed in SI2.py; code provides thermochemical properties of alkanes directly from the SMILES string; source code for

https://doi.org/10.1021/acs.jpcb.4c05355
J. Phys. Chem. B 2024, 128, 9619−9629

automated force field file generation, using the SMILES string as input, of branched alkanes for use with the RASPA2 software provided in SI3.py; predicted thermochemical properties of all alkane isomers until C_{14} for temperatures ranging from (0−1000) K listed in SI4.xlsx; linear regression coefficients of the second-order groups and the corresponding temperature-dependent quadratic polynomial fits also listed in SI4.xlsx; and reaction equilibrium distribution data of hydroisomerization of C_{10} and C_{14} in the gas phase and MTW-type zeolite at 500 K tabulated in SI4.xlsx (ZIP)
SI5.pdf contains figures showing the pore structure in MTW-type zeolite, $\Delta_f H_0$ values obtained from the Scott's tables, our LR model, and DIPPR database, and the coefficients of second-order groups with C as the central atom at different temperatures for $(G_0 - H_0(0\ K))$; reaction equilibrium distribution imposed on the gas phase for hydroisomerization of C_{10} isomers also shown in SI5.pdf (PDF)

AUTHOR INFORMATION

Corresponding Author

Thijs J.H. Vlugt − *Engineering Thermodynamics, Process & Energy Department, Faculty of Mechanical Engineering, Delft University of Technology, 2628CB Delft, The Netherlands;* orcid.org/0000-0003-3059-8712; Email: t.j.h.vlugt@tudelft.nl

Authors

Shrinjay Sharma − *Engineering Thermodynamics, Process & Energy Department, Faculty of Mechanical Engineering, Delft University of Technology, 2628CB Delft, The Netherlands;* orcid.org/0000-0001-8345-7433

Josh J. Sleijfer − *Delft Institute of Applied Mathematics, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, 2628CD Delft, The Netherlands; Faculty of Applied Sciences, Delft University of Technology, 2628 CJ Delft, The Netherlands*

Jeroen Op de Beek − *Delft Institute of Applied Mathematics, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, 2628CD Delft, The Netherlands*

Stach van der Zeeuw − *Faculty of Applied Sciences, Delft University of Technology, 2628 CJ Delft, The Netherlands*

Daniil Zorzos − *Faculty of Aerospace Engineering, Delft University of Technology, 2629 HS Delft, The Netherlands*

Silvia Lasala − *Université de Lorraine, CNRS, LRGP, F-54000 Nancy, France*

Marcello S. Rigutto − *Shell Global Solutions International B.V., 1031HW Amsterdam, The Netherlands;* orcid.org/0000-0002-3671-3446

Erik Zuidema − *Shell Global Solutions International B.V., 1031HW Amsterdam, The Netherlands*

Umang Agarwal − *Shell Chemical LP, Monaca, Pennsylvania 15061, United States;* orcid.org/0000-0002-5182-3141

Richard Baur − *Shell Global Solutions International B.V., 1031HW Amsterdam, The Netherlands*

Sofia Calero − *Department of Applied Physics, Eindhoven University of Technology, 5600MB Eindhoven, The Netherlands;* orcid.org/0000-0001-9535-057X

David Dubbeldam − *Van't Hoff Institute of Molecular Sciences, University of Amsterdam, 1098XH Amsterdam, The Netherlands;* orcid.org/0000-0002-4382-1509

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.4c05355

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was sponsored by NWO Domain Science for the use of supercomputer facilities. This work is part of the Advanced Research Center for Chemical Building Blocks, ARC-CBBC, which is cofunded and cofinanced by The Netherlands Organization for Scientific Research (NWO) and The Netherlands Ministry of Economic Affairs and Climate Policy. The authors acknowledges Dr.ir. Jasper van Baten for providing data on thermochemical properties of alkanes computed using SPLIT software by AmsterCHEM (https://www.amsterchem.com/). The authors also acknowledge the use of computational resources of the DelftBlue supercomputer, provided by Delft High Performance Computing Center (https://www.tudelft.nl/dhpc).[63]

REFERENCES

(1) van Bavel, S.; Verma, S.; Negro, E.; Bracht, M. Integrating CO_2 electrolysis into the gas-to-liquids−power-to-liquids process. *ACS Energy Letters* **2020**, *5*, 2597−2601.

(2) Calis, H.; Lüke, W.; Drescher, I.; Schütze, A. Synthetic Diesel Fuels. In *Energy Sources for Transportation: Handbook of Fuels*; 2021; pp 161−200.

(3) Smit, B.; Maesen, T. L. M. Towards a molecular understanding of shape selectivity. *Nature* **2008**, *451*, 671−678.

(4) Rigutto, M. S.; van Veen, R.; Huve, L. Zeolites in hydrocarbon processing. In *Studies in surface science and catalysis*; Elsevier: 2007; pp 855−913.

(5) Mäki-Arvela, P.; Kaka khel, T. A.; Azkaar, M.; Engblom, S.; Murzin, D. Y. Catalytic hydroisomerization of long-chain hydrocarbons for the production of fuels. *Catalysts* **2018**, *8*, 534.

(6) Letcher, T. M. *Chemical thermodynamics for industry*, 1st ed.; Royal Society of Chemistry: Cambridge, 2004.

(7) Sharma, S.; Rigutto, M. S.; Zuidema, E.; Agarwal, U.; Baur, R.; Dubbeldam, D.; Vlugt, T. J. H. Understanding shape selectivity effects of hydroisomerization using a reaction equilibrium model. *J. Chem. Phys.* **2024**, *160*, 214708.

(8) Peng, H.; Zhang, D.; Ling, X.; Li, Y.; Wang, Y.; Yu, Q.; She, X.; Li, Y.; Ding, Y. n-Alkanes phase change materials and their microencapsulation for thermal energy storage: a critical review. *Energy Fuels* **2018**, *32*, 7262−7293.

(9) Scott, D. W. *Chemical thermodynamic properties of hydrocarbons and related substances: properties of the alkane hydrocarbons, C1 through C10, in the ideal gas state from 0 to 1500 K*, 1st ed.; US Department of the Interior, Bureau of Mines: Washington DC, 1974.

(10) Scott, D. W. Correlation of the chemical thermodynamic properties of alkane hydrocarbons. *J. Chem. Phys.* **1974**, *60*, 3144−3165.

(11) Linstrom, P. J.; Mallard, W. G. The NIST Chemistry WebBook: A chemical data resource on the internet. *Journal of Chemical & Engineering Data* **2001**, *46*, 1059−1063.

(12) Fraser, F. M.; Prosen, E. J. Heats of combustion of liquid n-hexadecane, 1-hexadecene, normal-decylbenzene, normal-decylcyclohexane, normal-decylcyclopentane, and the variation of heat of combustion with chain length. *Journal of Research of the National Bureau of Standards* **1955**, *55*, 329−333.

(13) Prosen, E.; Pitzer, K.; Rossini, F. Heats and free energies of formation of the paraffin hydrocarbons, in the gaseous state, to 1500 degree. *Natl. Bur. Stand* **1945**, *34*, 403.

(14) Stull, D. R.; Westrum, E. F.; Sinke, G. C. *The chemical thermodynamics of organic compounds*; Elsevier: New York, 1969.

(15) Rossini, F. D. *Selected Values of Properties of Hydrocarbons: Prepared as Part of the Work of the American Petroleum Institute Research Project 44*; US Department of Commerce, National Bureau of Standards: Washington, D.C., 1947; pp 175−356.
(16) Gani, R. Group contribution-based property estimation methods: advances and perspectives. *Current Opinion in Chemical Engineering* **2019**, *23*, 184−196.
(17) Morales-Rodriguez, R. *Thermodynamics: Fundamentals and Its Application in Science*; InTech: Rijeka, 2012.
(18) Benson, S. W.; Cruickshank, F.; Golden, D.; Haugen, G. R.; O'neal, H.; Rodgers, A.; Shaw, R.; Walsh, R. Additivity rules for the estimation of thermochemical properties. *Chem. Rev.* **1969**, *69*, 279−324.
(19) Joback, K. G.; Reid, R. C. Estimation of pure-component properties from group-contributions. *Chem. Eng. Commun.* **1987**, *57*, 233−243.
(20) Constantinou, L.; Gani, R. New group contribution method for estimating properties of pure compounds. *AIChE J.* **1994**, *40*, 1697−1710.
(21) Marrero, J.; Gani, R. Group-contribution based estimation of pure component properties. *Fluid Phase Equilib.* **2001**, *183*, 183−208.
(22) Hukkerikar, A. S.; Meier, R. J.; Sin, G.; Gani, R. A method to estimate the enthalpy of formation of organic compounds with chemical accuracy. *Fluid Phase Equilib.* **2013**, *348*, 23−32.
(23) Albahri, T. A.; Aljasmi, A. F. SGC method for predicting the standard enthalpy of formation of pure compounds from their molecular structures. *Thermochim. Acta* **2013**, *568*, 46−60.
(24) Domalski, E. S.; Hearing, E. D. Estimation of the Thermodynamic Properties of Hydrocarbons at 298.15 K. *J. Phys. Chem. Ref. Data* **1988**, *17*, 1637−1678.
(25) Yalamanchi, K. K.; Van Oudenhoven, V. C.; Tutino, F.; Monge-Palacios, M.; Alshehri, A.; Gao, X.; Sarathy, S. M. Machine learning to predict standard enthalpy of formation of hydrocarbons. *J. Phys. Chem. A* **2019**, *123*, 8305−8313.
(26) Yaws, C. L. *Yaws' Handbook of Thermodynamic Properties for Hydrocarbons and Chemicals*, 1st ed.; Knovel: New York, 2009.
(27) Bloxham, J. C.; Redd, M. E.; Giles, N. F.; Knotts, T. A., IV; Wilding, W. V. Proper use of the DIPPR 801 database for creation of models, methods, and processes. *Journal of Chemical & Engineering Data* **2021**, *66*, 3−10.
(28) Trinh, C.; Meimaroglou, D.; Lasala, S.; Herbinet, O. Machine Learning for the prediction of the thermochemical properties (enthalpy and entropy of formation) of a molecule from its molecular descriptors. *Comput.-Aided Chem. Eng.* **2022**, *51*, 1471−1476.
(29) Aldosari, M. N.; Yalamanchi, K. K.; Gao, X.; Sarathy, S. M. Predicting entropy and heat capacity of hydrocarbons using machine learning. *Energy and AI* **2021**, *4*, No. 100054.
(30) Hayes, M. Y.; Li, B.; Rabitz, H. Estimation of molecular properties by high-dimensional model representation. *J. Phys. Chem. A* **2006**, *110*, 264−272.
(31) Muckley, E. S.; Saal, J. E.; Meredig, B.; Roper, C. S.; Martin, J. H. Interpretable models for extrapolation in scientific machine learning. *Digital Discovery* **2023**, *2*, 1425−1435.
(32) Estrada-Villagrana, A. D.; De La Paz-Zavala, C. Application of chemical equilibrium for hydrocarbon isomerization analysis. *Fuel* **2007**, *86*, 1325−1330.
(33) Gunawan, M. L.; Novita, T. H.; Aprialdi, F.; Aulia, D.; Nanda, A. S. F.; Rasrendra, C. B.; Addarojah, Z.; Mujahidin, D.; Kadja, G. T. M. Palm-oil transformation into green and clean biofuels: Recent advances in the zeolite-based catalytic technologies. *Bioresource Technology Reports* **2023**, *23*, No. 101546.
(34) Poursaeidesfahani, A.; de Lange, M. F.; Khodadadian, F.; Dubbeldam, D.; Rigutto, M.; Nair, N.; Vlugt, T. J. H. Product shape selectivity of MFI-type, MEL-type, and BEA-type zeolites in the catalytic hydroconversion of heptane. *J. Catal.* **2017**, *353*, 54−62.
(35) Holladay, J.; Abdullah, Z.; Heyne, J. *Sustainable aviation fuel: Review of technical pathways*; DOE EERE; Pacific Northwest National Lab.(PNNL): Richland, WA, United States, 2020.
(36) Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31−36.
(37) Weininger, D.; Weininger, A.; Weininger, J. L. SMILES. 2. Algorithm for generation of unique SMILES notation. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 97−101.
(38) Baerlocher, C.; McCusker, L. B.; Olson, D. H. *Atlas of zeolite framework types*, 6th ed.; Elsevier: Amsterdam, 2007.
(39) Dubbeldam, D.; Torres-Knoop, A.; Walton, K. S. On the inner workings of Monte Carlo codes. *Mol. Simul.* **2013**, *39*, 1253−1292.
(40) Dubbeldam, D.; Calero, S.; Ellis, D. E.; Snurr, R. Q. RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials. *Mol. Simul.* **2016**, *42*, 81−101.
(41) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P. SciPy 1.0 Contributors, SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261−272.
(42) Hens, R.; Rahbari, A.; Caro-Ortiz, S.; Dawass, N.; Erdös, M.; Poursaeidesfahani, A.; Salehi, H. S.; Celebi, A. T.; Ramdin, M.; Moultos, O. A.; Dubbeldam, D.; Vlugt, T. J. H. Brick-CFCMC: Open source software for Monte Carlo simulations of phase and reaction equilibria using the Continuous Fractional Component method. *J. Chem. Inf. Model.* **2020**, *60*, 2678−2682.
(43) Polat, H. M.; Salehi, H. S.; Hens, R.; Wasik, D. O.; Rahbari, A.; De Meyer, F.; Houriez, C.; Coquelet, C.; Calero, S.; Dubbeldam, D.; Moultos, O. A.; Vlugt, T. J. H. New features of the open source Monte Carlo software Brick-CFCMC: thermodynamic integration and hybrid trial moves. *J. Chem. Inf. Model.* **2021**, *61*, 3752−3757.
(44) Polat, H. M.; de Meyer, F.; Houriez, C.; Moultos, O. A.; Vlugt, T. J. H. Solving chemical absorption equilibria using free energy and quantum chemistry calculations: methodology, limitations, and new open-source software. *J. Chem. Theory Comput.* **2023**, *19*, 2616−2629.
(45) Chase, M. W. *NIST-JANAF thermochemical tables*, 4th ed.; American Chemical Society: Washington DC, 1998; Vol. *9*; pp 550−1261.
(46) Matito-Martos, I.; Rahbari, A.; Martin-Calvo, A.; Dubbeldam, D.; Vlugt, T. J. H.; Calero, S. Adsorption equilibrium of nitrogen dioxide in porous materials. *Phys. Chem. Chem. Phys.* **2018**, *20*, 4189−4199.
(47) Myers, A. L.; Prausnitz, J. M. Thermodynamics of mixed-gas adsorption. *AIChE J.* **1965**, *11*, 121−127.
(48) Simon, C.; Smit, B.; Haranczyk, M. pyIAST: Ideal Adsorbed Solution Theory (IAST) Python Package. *Comput. Phys. Commun.* **2016**, *200*, 364−380.
(49) Widom, B. Some topics in the theory of fluids. *J. Chem. Phys.* **1963**, *39*, 2808−2812.
(50) Siepmann, J. I.; Frenkel, D. Configurational Bias Monte Carlo: a new sampling scheme for flexible chains. *Mol. Phys.* **1992**, *75*, 59−70.
(51) De Pablo, J. J.; Laso, M.; Siepmann, J. I.; Suter, U. W. Continuum-configurational-bias Monte Carlo simulations of long-chain alkanes. *Mol. Phys.* **1993**, *80*, 55−63.
(52) Vlugt, T. J. H.; Krishna, R.; Smit, B. Molecular simulations of adsorption isotherms for linear and branched alkanes and their mixtures in silicalite. *J. Phys. Chem. B* **1999**, *103*, 1102−1118.
(53) Vlugt, T. J. H. Efficiency of parallel CBMC simulations. *Mol. Simul.* **1999**, *23*, 63−78.
(54) Ryckaert, J.-P.; Bellemans, A. Molecular dynamics of liquid alkanes. *Faraday Discuss. Chem. Soc.* **1978**, *66*, 95−106.
(55) Lennard, J. E.; Jones, I. On the determination of molecular fields.—I. From the variation of the viscosity of a gas with temperature. *Proc. R. Soc. A* **1924**, *106*, 441−462.

https://doi.org/10.1021/acs.jpcb.4c05355
J. Phys. Chem. B 2024, 128, 9619−9629

(56) Dubbeldam, D.; Calero, S.; Vlugt, T. J. H.; Krishna, R.; Maesen, T. L. M.; Smit, B. United atom force field for alkanes in nanoporous materials. *J. Phys. Chem. B* **2004**, *108*, 12301−12313.
(57) Bai, P.; Tsapatsis, M.; Siepmann, J. I. TraPPE-zeo: transferable potentials for phase equilibria force field for all-silica zeolites. *J. Phys. Chem. C* **2013**, *117*, 24375−24387.
(58) Martin, M. G.; Siepmann, J. I. Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes. *J. Phys. Chem. B* **1998**, *102*, 2569−2577.
(59) Nath, S. K.; Khare, R. New forcefield parameters for branched hydrocarbons. *J. Chem. Phys.* **2001**, *115*, 10837−10844.
(60) Vlugt, T. J. H.; Schenk, M. Influence of framework flexibility on the adsorption properties of hydrocarbons in the zeolite silicalite. *J. Phys. Chem. B* **2002**, *106*, 12757−12763.
(61) Levenspiel, O. *Chemical reaction engineering*, 3rd ed.; John Wiley & Sons: New York, 1998.
(62) van Baten, J. *AmsterChem: tailor made engineering software solutions*. https://www.amsterchem.com/ (accessed on Jul 24, 2024).
(63) *Delft High Performance Computing Centre (DHPC), DelftBlue Supercomputer (Phase 2)*. https://www.tudelft.nl/dhpc/ark:/44463/DelftBluePhase2 (accessed on Jul 26, 2024).

https://doi.org/10.1021/acs.jpcb.4c05355
J. Phys. Chem. B 2024, 128, 9619−9629