Inverse Molecular Design of Alkoxides and Phenoxides

for Aqueous Direct Air Capture of CO$_2$

Zisheng Zhang,[1] Amanda L. Kummeth,[3] Jenny Y. Yang,[3,*] Anastassia N. Alexandrova[1,2,*]

[1]Department of Chemistry and Biochemistry and [2]California NanoSystems Institute,

University of California, Los Angeles, Los Angeles, CA 90095

[3]Department of Chemistry, University of California, Irvine, Natural Sciences II, Irvine,

CA, 92697 USA

ABSTRACT: Aqueous direct air capture (DAC) is a key technology toward a carbon negative infrastructure. Developing sorbent molecules with water- and oxygen-tolerance and high CO$_2$ binding capacity is therefore highly desired. In this work, we analyze the CO$_2$ absorption chemistries on amines, alkoxides, and phenoxides with density functional theory (DFT) calculations and search for the optimal sorbent using an inverse molecular design strategy. The alkoxides and phenoxides are found to be more suitable for aqueous DAC than amines thanks to their water-tolerance and capture stoichiometry of 1:1 (2:1 for amines). All three molecular systems are found to obey the same linear scaling relationship (LSR) between pK_{CO_2} and pK_a, since both CO$_2$ and proton are bonded to the nucleophilic binding site through a majorly σ bonding orbital. Several high-performance alkoxides are proposed from the computational screening. In contrast, phenoxides have relatively poor correlation between pK_{CO_2} and pK_a, showing promise for optimization. We

apply genetic algorithm (GA) to search the chemical space of substituted phenoxides for the optimal sorbent. Several promising candidates that break the LSR are discovered. The most promising off-LSR candidate phenoxides feature bulky ortho substituents forcing the CO_2 adduct into a perpendicular configuration with respect to the aromatic ring. In this configuration, CO_2 utilizes a different molecular orbital for binding than does the proton, and the pK_{CO_2} and pK_a are thus decoupled. The $pK_{CO_2} - pK_a$ trend and off-LSR behaviors are then confirmed by experiments, validating the inverse molecular design framework. This work not only extensively studies the chemistry of the aqueous DAC, but also presents a transferrable computational workflow for understanding and optimization of other functional molecules.

INTRODUCTION

The climbing concentration of atmospheric CO_2 has presented tremendous challenges in aspects of environment, climate, and biodiversity.[1] Various technologies have been developed to convert CO_2 into value-added chemicals with energy input from clean and renewable resources, such as photo- and electro-catalytic CO_2 reduction[2] and water gas shift reactions.[3] However, the scalability of such approaches is heavily limited by the input side, as CO_2 input with high purity or concentration is need. Carbon dioxide Capture and Concentration (CCC), from industrial waste gas, vehicle emissions, or directly from air, is therefore a highly nontrivial rate-determining step for achieving net-zero, or ultimately negative, carbon emissions.[4]

The efficiency of CCC is dependent on a proper carrier molecule which can selectively bind CO_2 under one condition and release under another. The release method can include heat treatment (thermal cycle), electrochemical redox swing, electrochemical pH swings, photo-switches, and ligand exchange in metal-organic frameworks..[5–8] To date, the most industrially successful carrier family have been the amines which are able to bind CO_2 through the bicarbonate pathway and the carbamate pathway (*vide infra*). The captured CO_2 can then be released by heating up the solution in a thermal cycle system or by addition of metal cations in a metal-amine system.[9] However, the application of amines has been largely limited to treat industrial flue gas (8-12 % CO_2) instead of direct air capture (DAC) (0.4 % CO_2). The latter contains extremely dilute CO_2 but relatively high concentration of O_2 and H_2O. Uptaking CO_2 from a such dilute gas mixture is

thermodynamically challenging and requires a favorable and selective CO_2 binding energetics on the sorbent molecules. In addition, O_2 could oxidize functional groups such as the amino group, and H_2O could protonate the CO_2 binding site of the sorbent to render it inactive. Therefore, sorbent that is air- or water- stable with high CO_2 binding capabilities is highly desirable for aqueous DAC applications.

Since the sorbent are water soluble molecules, one could approach the optimal molecular candidate by two different strategies, bottom-up or top-down. For molecular systems where the relationship between the electronic structure and desired property is well understood, one could make modifications to the parent molecule accordingly to tune the molecular property based on the design principles, i.e., direct design or molecular engineering. If there is insufficient insight into the structure-activity relationship, the pool of possible candidates can be screened, or molecular design can be treated as an optimization task and evolutionary algorithms can be applied to efficiently optimize the desired property in a predefined chemical subspace, i.e., inverse design. Both strategies have proven successful in optimizing functional molecules for various applications, including solar cell,[10] redox flow cell,[11] solar heat battery,[12] molecular photocatalysis,[13] and electrocatalysis.[14–17]

Here, we focus on aqueous DAC and search for the optimal sorbent molecule from various molecular systems with a combination of inverse design strategies and modern computational methods. The bonding nature and energetics of CO_2 binding and protonation on substituted amines, alkoxides, and phenoxides are investigated using

density functional theory (DFT) calculations. We discovered that all three sorbent families obey a linear scaling relationship (LSR) between pK_{CO_2} and pK_a which originates in the same bonding nature of proton and CO_2 on the binding site. To exploit the off-LSR behavior of phenoxides, a genetic algorithm (GA) searcher combined with semiempirical quantum mechanical (SQM) calculations is performed to efficiently search for the air- and water- stable alkoxide species with optimal CO_2 binding capabilities. The top-scoring candidates are further refined with DFT calculations and analyzed computationally, to provide insights and design principles based on the search results. The data generated from the GA search can be further used to train predictive machine learning (ML) models for low-cost prediction of MO energy levels and binding free energies. For validation, several simple synthetically accessible substituted phenoxide and alkoxide molecules were selected to evaluate theirCO_2 absorption capacity. The overall pK_{CO_2}- pK_a trend and off-LSR behaviour of specific molecules are observed in the experiments, consistent with theoretical predictions. The inverse molecular design workflow not only provides trends and design principles, with minimal prior knowledge, but also is readily generalizable to explore and optimize other functional molecules for various applications.

RESULT AND DISCUSSION

Amines for aqueous DAC.

Amines are the current leading CO_2 sorbent family for post combustion capture.[5, 18] Amines are usually divided into groups based on the number of hydrogens on the nitrogen, namely primary, secondary, and tertiary. In aqueous solution, tertiary amines ($R_1R_2R_3N$) hydrolyze to produce hydroxide which acts as the capture agent:

$$R_1R_2R_3N + H_2O \rightarrow R_1R_2R_3NH^+ + OH^-$$

$$CO_2 + OH^- \rightarrow HCO_3^-$$

$$HCO_3^- + OH^- \rightarrow H_2O + CO_3^{2-}$$
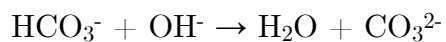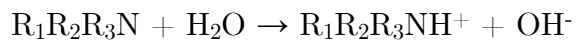
In these cases, the tertiary amino group function solely as a weak base instead of the direct binding site for CO_2. This makes tertiary amines a less interesting and promising system from the perspective of molecular design: substituents can only affect the basicity but not the CO_2 binding energetics.
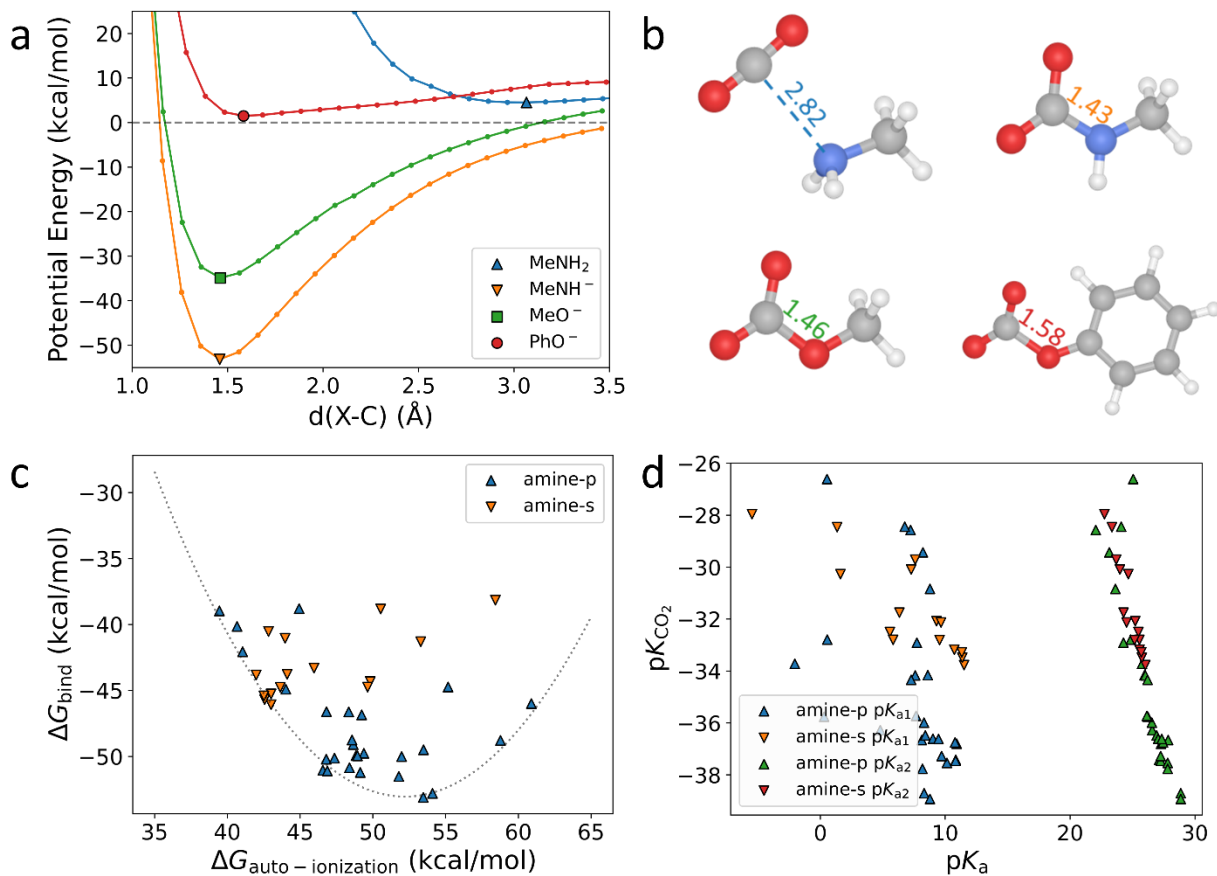
Figure 1. (a) Potential energy profile of CO_2 binding of the four species from a relaxed PES scan, referenced to the energy of free molecules (marked by dotted line). (b) Optimized geometries of CO_2 binding on $MeNH_2$, $MeNH^-$, MeO^-, and PhO^-, with the d(X-C) distances (X represent N for amine and O for alkoxide/phenoxide) labeled in Angstrom. (c) Scatter plot of DFT-calculated G_{bind} versus $G_{auto-ionization}$ for investigated primary and secondary amine molecules. (d) Scatter plot of DFT-calculated pK_{CO_2} values versus the pK_{a1} and pK_{a2} for investigated primary and secondary amine molecules.

CO_2 capture using primary and secondary amines has been investigated extensively experimentally; however, a clear relation that could guide molecular design has not yet been proposed.(19, 20) Figure 1a shows the potential energy surface (PES) of CO_2 binding on $MeNH_2$ and its deprotonated form $MeNH^-$ obtained from relaxed scan at B3LYP-D3(BJ)/def2-TZVP. Despite some older reports that the primary and secondary amines could capture CO_2 via a zwitterionic (i.e. carbamic acid intermediate) pathway,(21, 22) our DFT calculation (Figure 1a) shows that the neutral amino group in $MeNH_2$ could only non-covalently physi-sorb a CO_2 molecule, as is evidenced by the N-C distance of 2.82 Å (Figure 1b). Only after deprotonation can a primary or secondary amino group bind CO_2 covalently into a carbamate intermediate, with a binding energy of -53.1 kcal/mol and a N-C bond length of 1.43 Å. Since the neutral amino groups are always more basic than water, the deprotonated amino group can only form from an autoionization process:

$$2\ R_1R_2NH \rightarrow R_1R_2NH_2^+ + R_1R_2N^-$$

The deprotonated amine group $R_1R_2N^-$ then readily capture a dissolved CO_2 molecule to form carbamate:

$$R_1R_2N^- + CO_2 \rightarrow R_1R_2NCOO^-$$
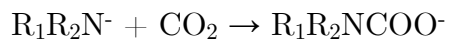
which in turn promotes the auto-ionization by consumption of the deprotonated species, until the system reaches equilibrium. In the meanwhile, the $R_1R_2NH_2^+$ produced in the auto-ionization step serves no absorbing function and remains in the protonated form

thereafter. Therefore, the primary and secondary amines bind CO_2 in an overall stoichiometry of 2:1, i.e., one mole of amine is "wasted" for each mole of CO_2 captured. Since the air is a dilute stream of CO_2, to achieve optimal absorption capacity, a sorbent needs favorable thermodynamics in both auto-ionization and CO_2 binding. To investigate the energetic trend, we perform DFT calculations on a library of mono-substituted primary and secondary amines, and Figure 1c shows the scatter plot of their CO_2 binding free energy G_{bind} versus auto-ionization free energy $G_{auto-ionization}$. Unfortunately, $G_{auto-ionization}$ lies between 39.5 and 60.9 kcal/mol, suggesting difficult autoionization and hence a low concentration of available species for CO_2 binding. In addition, no clear relation can be observed between the two reaction energies. As $G_{auto-ionization}$ increases, the G_{bind} decreases until $G_{auto-ionization}$ reaches c.a. 52 kcal/mol and increases thereafter, forming a convex hull. Since the unsubstituted methylamine (G_{bind} = 46.6 kcal/mol, $G_{auto-ionization}$ = −51.1 kcal/mol) is already located near the top of the "inversed volcano", the intrinsic limit leaves little room for further optimization of both auto-ionization and CO_2 binding.

To gain deeper insight into the energetic trends, we calculated the pK_{a1} ($R_1R_2NH_2^+ \rightarrow R_1R_2NH$), pK_{a2} ($R_1R_2NH \rightarrow R_1R_2N^-$), and pK_{CO_2} of the primary and secondary amines via the linear free energy relation (LFER)(23) and plotted them in Figure 1d. Note that pK_{a1} is usually reported as pK_a due to experimental difficulty in determining pK_{a2}. Previous experimental studies have reported a weak positive correlation between the pK_{a1} and CO_2 absorption capacity,(20) which is also observed in our calculations where pK_{a1} correlates

poorly with pK_{CO_2} with a R^2 of 0.22 (Figure S2a). However, pK_{a2} is found to correlate much better with pK_{CO_2} in a linear manner, with a R_2 of 0.86 (Figure S2b). The linear correlation is attributed to the similar bonding mode through which proton and CO_2 act as electrophile and bind to the nucleophilic anionic nitrogen site. The similar bonding characteristics results in coupling of the G_{bind} and G_{deprot}, i.e., linear scaling relationship (LSR).(24) The poor correlation between pK_{CO_2} and pK_{a1} originates in the poor correlation between pK_{a2} and pK_{a1} (Figure S2c). The nitrogen in neutral or deprotonated amino groups has different types of hybridization and hence is influenced to different extents by substituent effects. In addition, the first and the second proton to be attached to the anionic nitrogen is impacted by different steric effects from nearby groups. This is also the cause of the G_{bind}-$G_{auto-ionization}$ convex hull observed in Figure 1c.

Alkoxides for aqueous DAC

Despite the favorable CO_2 binding energetics of amine, the stoichiometry of 2:1 for the capturing agents and CO_2 caps the absorption capacity. The pK_{a2} of primary and secondary amines lies above 20, which makes them unsuitable for aqueous DAC due to susceptibility to protonation by water (pK_a=15.7) and blocking of the CO_2 binding site. In addition, primary and secondary amines could suffer from oxidation when in contact with air. In contrast, alkoxides are a promising family of CO_2 capturing agents that are more O_2-insensitive and can directly bind CO_2 to form a carbonate with 1:1 stoichiometry:
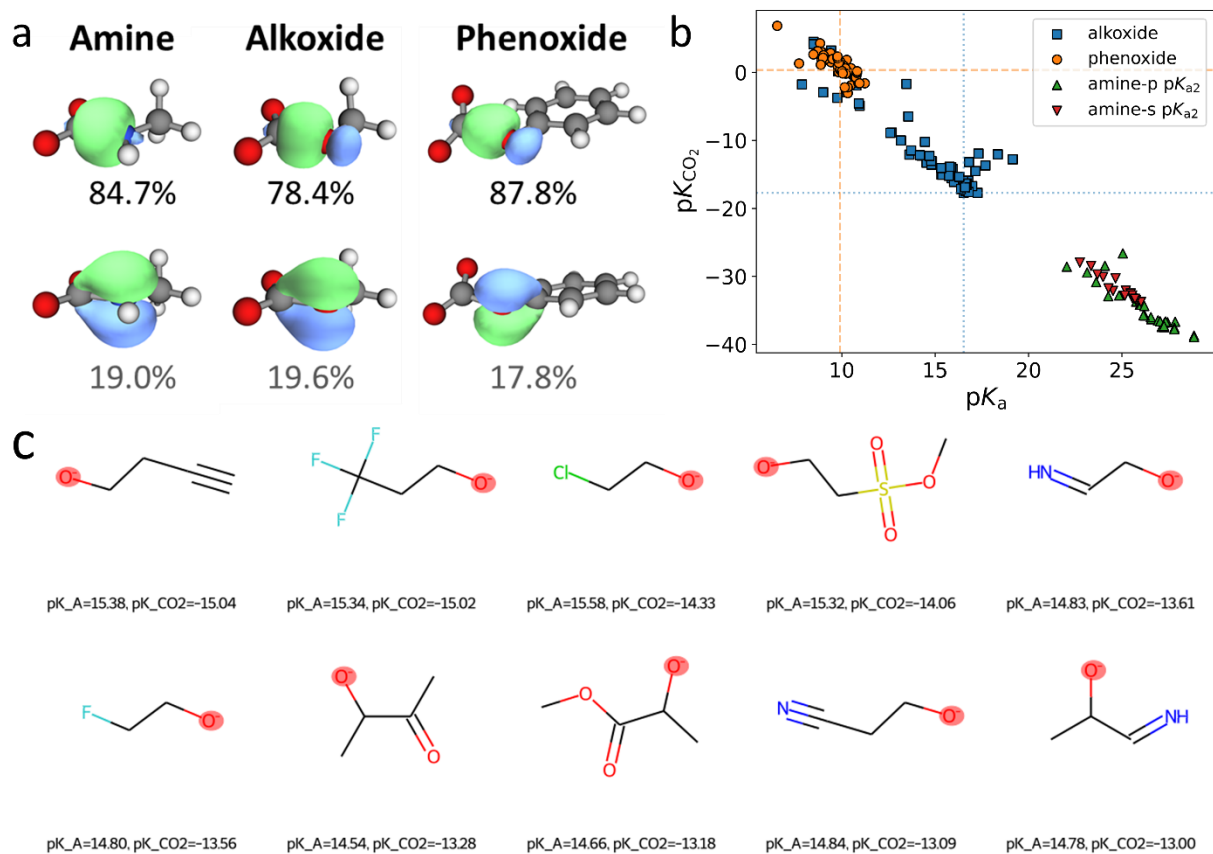
$RO^- + CO_2 \rightarrow ROCO_2^-$

Figure 2. (a) The localized molecular orbitals with highest contribution to the Mayer bond order of the X-C bond (X represents N for amine and O for alkoxide/phenoxide), with the percentage contributions labeled below each orbital. (b) Scatter plot of DFT-calculated pK_{CO_2} values versus the pK_a for all investigated molecules. The orange dashed line and the blue dotted line represent the data points of phenoxide and methoxide, respectively. (c) The top-10 water-insensitive alkoxide candidates for CO_2 binding, with the pK_a and pK_{CO_2} labeled under each structural formula.

The CO_2 binding on alkoxide has the same bonding characteristics as on a deprotonated amino group. Figure 2a shows the two localized molecular orbitals (LMOs) with the

highest contribution to the Mayer bond order between CO_2 and the capturing agent. It can be seen that the major bonding orbital of amine- and phenoxide-CO_2 adduct both have s-pattern, accounting for 84.7% and 78.4% of the Mayer bond order, respectively. The LMO with the second highest bonding contribution has p-pattern for both amine and alkoxides, with a minor contribution of 19.0% and 19.6%, respectively. The analysis reveals the similar bonding pattern of CO_2 binding to an alkoxide as to an amine.

To explore the pK_{CO_2}-pK_a relation of alkoxide, the binding constants are calculated using DFT and plotted in Figure 2b for the library of mono-substituted methoxide and ethoxides, together with the datapoints of amines. The pK_{CO_2} of the majority of the alkoxides lie in the range of -10 to -20, suggesting less favorable CO_2 binding energetics compared to amines. However, most of the alkoxides have their pK_a below 15.7, which prevents deactivation from protonation by water. This could be attributed to the weaker nucleophilicity of the O in alkoxide (Hirshfeld charge: -0.84 e) than the N in amine (Hirshfeld charge: -1.12 e). As can be expected from the chemical bonding analysis, there also exists a LSR between pK_{CO_2} and pK_a of alkoxides that is similar to the case of amines. More intriguingly, the LSRs of alkoxides and amines lie in the same straight line, with a R^2 of 0.97 on the combined dataset. The shared LSR demonstrates that the weakening of CO_2 binding on alkoxide compared to on amine is just a compromise in exchange for water-tolerance. Alkoxides are not systematically inferior to amines, and since they follow the exact same pK_{CO_2}-pK_a curve we can access the upper left region that is inaccessible for amine via functional group substitution. Since the pK_a of ethanol (exp. pK_a=16) is

slightly higher than that of water, a straightforward approach is to move toward the upper left direction along the LSR by attaching weak electron withdrawing group (EWG) to the alkoxide. At the same time, the tradeoff must be kept small enough so that CO_2 binding capability is not excessively weakened. The top-10 water-tolerant alkoxide candidates ranked by pK_{CO_2} are listed in Figure 2c.

Optimizing Phenoxides with Genetic Algorithm Search

The LSR provides a clear correlation along which we could tune the binding energetics, however, it also imposes an intrinsic limitation that prevents optimization of both pK_a and pK_{CO_2}. This is similar to the activity volcano in catalysis.(25) Hence, we further extend the study to phenoxides whose distinct sterics and electronic structure characteristics (aromaticity) from alkoxides could span a different chemical space. The bonding nature of CO_2 on phenoxide is found to be similar to that on amines and alkoxide: major s-characteristics and minor p-characteristics (Figure 2a). A relatively low p-contribution could be attributed to the slightly out-of-plane CO_2 binding configuration (156.7° dihedral angle between the CO_2 plane and the benzene plane) and the lack of hyperconjugation effects by alkyl groups. Because the phenyl group is an inductive EWG, it also results in a less nucleophilic O in phenoxide (Hirshfeld charge: -0.58 e) compared to the alkoxide case, and that is also the reason why phenol is more acidic than alcohol. Therefore, phenoxides show a weaker CO_2 binding than alkoxides or amines, in terms of both energetics and bond length (Figure 1a, b). In Figure 2b, the datapoints of mono- and

di-substituted phenoxides lie in the upper left region in the pK_{CO_2}- pK_a plot, which means weak basicity (not prone to protonation by water) but small CO_2 adsorption capacity.
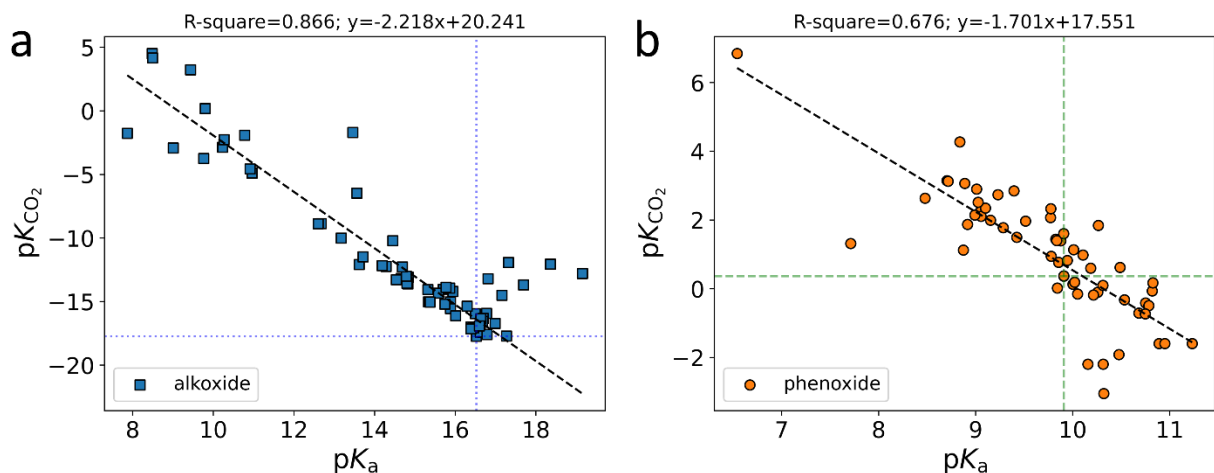

Figure 3. Scatter plot of DFT-calculated pK_{CO_2} values versus the pK_a of (a) substituted alkoxides and (b) substituted phenoxides, with the R^2 value and the formula of the LSR fitting labeled above each plot. The green dashed line and the blue dotted line represent the data points of phenoxide and methoxide, respectively.

A closer inspection of the phenoxide datapoints could reveal a pK_{CO_2}- pK_a distribution pattern that is distinct from that of the alkoxides. In Figure 3a, the datapoints of alkoxides closely obey a linear correlation with an R^2 of 0.86. However, the datapoints of phenoxides, although roughly showing a consistent trend with the LSR of alkoxides, are much more dispersed and have a lower R^2 of 0.67, suggesting a weak correlation between pK_{CO_2} and pK_a of phenoxides. This can be attributed to the rigidness of the phenyl group: the C-C in the phenyl ring cannot rotate as freely as C-C in the alkyl chains, which prevent ortho

substituents from adapting to configurations where their interaction with the CO_2 binding site is minimized. The aromaticity of the phenyl ring also allows for para and meta substituents to influence the electronic structure of the CO_2 binding site via the conjugated Π system, whereas the substituent effects in alkoxide tend to die out beyond the β carbon atom due to the saturated sp³ hybridization. In addition, the unsubstituted phenoxide is at about the middle of the distribution, unlike the case of alkoxides where the unsubstituted methoxide is at the lower-left end. To sum up, the phenoxides have greater room for further optimization and, more importantly, the potential to break the LSR and decouple the pK_a and pK_{CO_2}.

Due to a lack of insights into the origin of such off-LSR behavior, we turn to the inverse design strategy and expend the chemical subspace from mono- and di-substitution to all possible substitutions. This would include a total of c.a. 8 million unique molecules (24 substituents, 5 sites), which is beyond the capability of brute force exhaustion. To efficiently explore the vast chemical space, we employ genetic algorithm (GA), an evolutionary algorithm that has been successfully applied to structural prediction and property optimization of molecular systems,[15, 26] to search for the substituted phenoxide with minimal ΔG_{bind} on condition that its $pK_a$is lower than 15.74. To lower the computational cost and speed up the GA search, semi-empirical quantum mechanical (SQM) method GFN1-xtb, with GBSA implicit solvation, is adopted. The GFNn-xtb method has been reported to predict pK_a at semi-quantitative level and is suitable for trend recognition on large datasets where QM methods are unaffordable.[27] In our tests,

the adopted SQM method could predict experimental pK_a via LFER with a R^2 of 0.9486

(Figure S3), which is comparable to the DFT calculations.
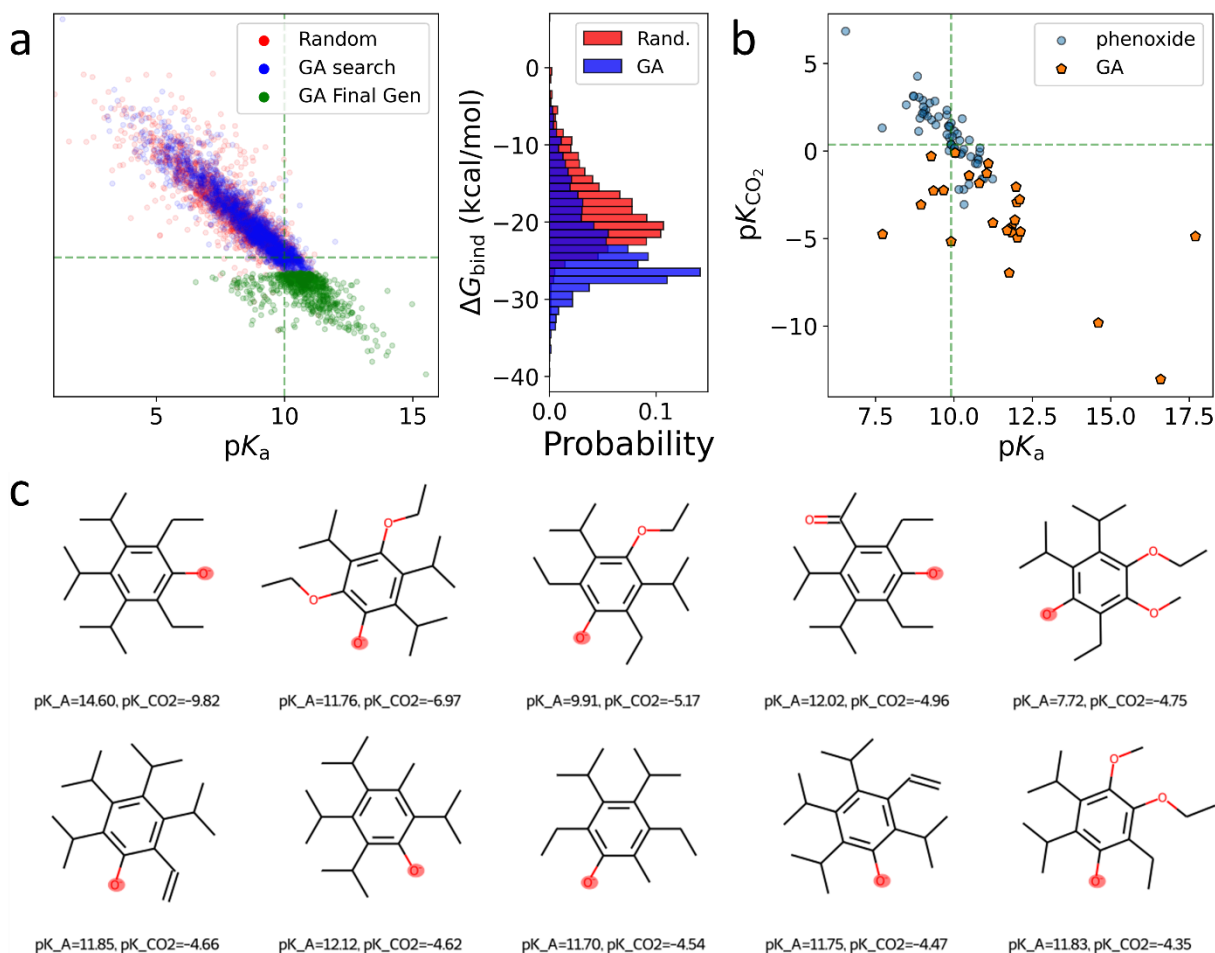
Figure 4. (a) Scatter plot of SQM-calculated ΔG_{bind} versus the pK_a of substituted phenoxides from random sampling, GA search, and the final generation of the GA search. The histogram showing distribution of ΔG_{bind} for random sampling and GA search is on the right panel. (b) Scatter plot of DFT-calculated pK_{CO_2} values versus the pK_a of top-scoring candidates from the GA search, together with the datapoints of simple substituted phenoxides (blue dots). (c) The top-10 water-insensitive phenoxide candidates for CO_2

binding from the GA search, with the pK_a and pK_{CO_2} labeled under each structural formula.

Ten independent GA searches are performed to avoid premature convergence in local optima, and the evolution of lowest ΔG_{bind} is shown in Figure S4. The optimal candidate from each individual GA search outperforms the reference species by 9~15 kcal/mol in terms of SQM-calculated ΔG_{bind}. It can be seen from Figure 4a that the GA not only samples sufficiently into the lower right region in the plot along the LSR but also explores the lower left region where the binding energies of proton and CO_2 are decoupled. The histogram in Figure 4a could more clearly demonstrate the sampling efficiency of the GA compared to random sampling, with 51% of the sampled candidates outperforming the reference molecule, which is a significantly higher percentage than 6% in random sampling. To verify the GA search results, DFT calculations are performed on the top-scoring candidates from the GA searches at the same level of theory as in the previous sections. It is shown in Figure 4b that all the candidates found in GA searches outperform the reference molecule, with a large portion of the datapoints distributed into the lower left off-LSR region. The datapoint with the most negative pK_{CO_2} is located far from the reference species and almost enters the regime of alkoxides. After filtering out the water-sensitive compounds, we rank the candidates from GA searches and show their molecular structure and binding constants in Figure 4c.
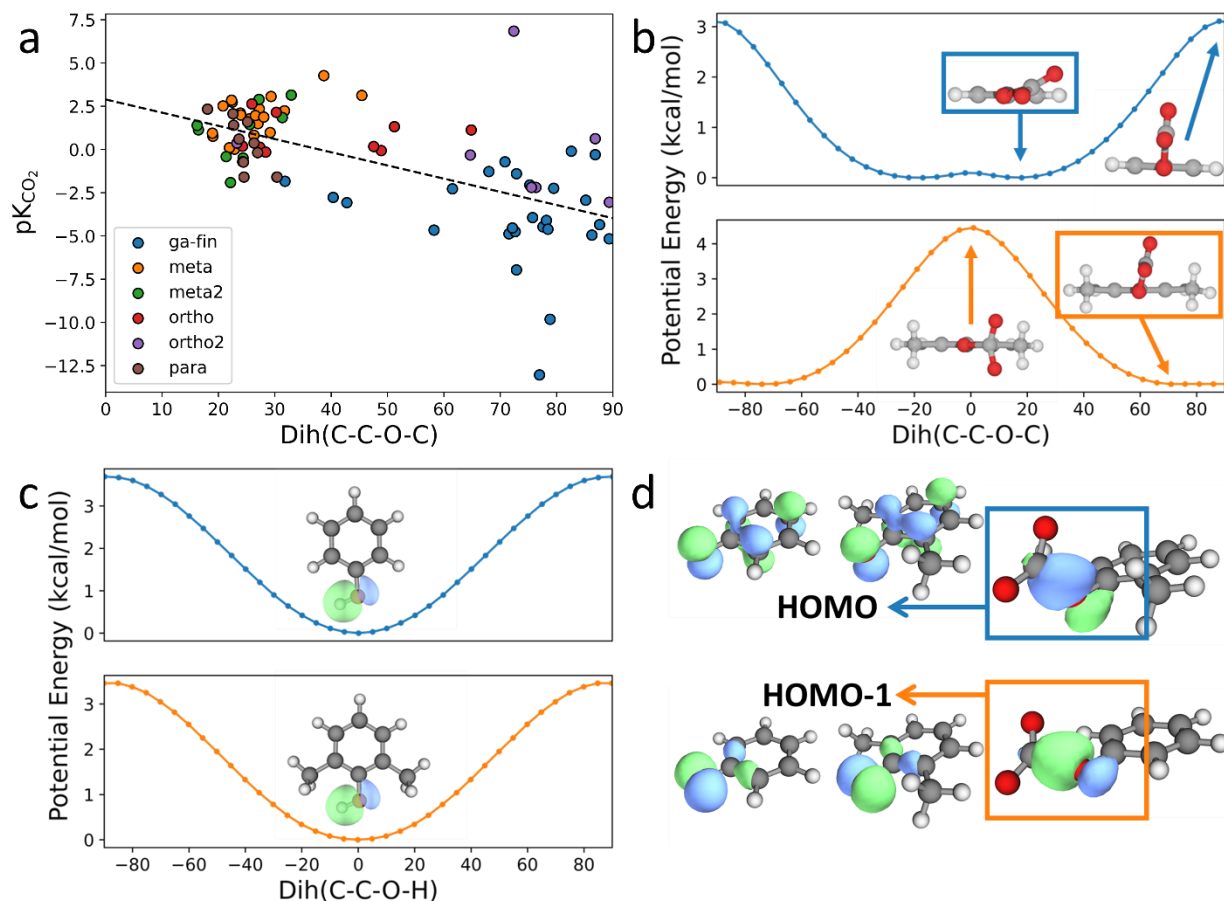
Figure 5. (a) Scatter plot of pK_{CO_2} versus phenoxide-CO_2 C-C-O-C dihedral angle of substituted phenoxides. The fitted line and the fitting formula are label in the plot. (b) Potential energy profile of phenoxide-CO_2 C-C-O-C dihedral angle rotation of the phenoxide and 2,6-xylenoxide from relaxed PES scans. Geometry of key extrema are shown along the profile. (c) Potential energy profile of phenyl-OH C-C-O-H dihedral angle rotation of the phenoxide and 2,6-xylenoxide from relaxed PES scans. The LMO that contributes the most to the proton binding is shown in the profile. The energies are referenced relative to global minima. (d) The HOMO and HOMO-1 isosurface of

phenoxide and 2,6-xylenoxide, and the bonding LMO isosurface that contributes the most to the CO_2 binding on them.

After obtaining and validating the optimal candidates from the inverse design strategy, we proceeded to study their molecular structures, aiming to understand the origin of their favorable and off-LSR energetics. A quick glance at Figure 4c reveals the common characteristics of the top-scoring and off-LSR candidates: bulky alkyl groups on the ortho positions. This is quite unexpected since alkyl groups are neither strong EWGs nor EDGs, and they are usually considered to weaken binding due to steric hinderance. During inspection of the DFT-optimized geometries, we notice that the CO_2 binding configuration on the top-scoring candidates is quite different from on the simple substituted phenoxides. Specifically, the dihedral angle between CO_2 plane and phenyl plane is c.a. 25° (noted as in-plane) for mono- and di-substituted phenoxides but 60~90° (noted as perpendicular) for the top-scoring candidates with bulky ortho groups from GA search. Moreover, as is shown in Figure 5a, the pK_{CO_2} is negatively correlated with the dihedral angle, i.e., perpendicular binding configurations are more energetically favorable compared to in-plane binding configurations.

To investigate how the bulky groups at ortho positions could alter the binding energetics, we use phenoxide and 2,6-xylenoxide as model systems. Note that the optimal candidate from the GA search is not used here for clarity and controlling variables. Relaxed PES scans are performed to explore the energy profile as the Dih(C-C-O-C), the dihedral angle

between phenyl plane and CO_2 planes, rotates. The PES of phenoxide-CO_2 adduct (Figure 5b) is quite flat in the 0~40° region, with the global minimum configuration at c.a. 20°. The exact in-plane 0° configuration is a local maximum due to the steric repulsion between CO_2 and the ortho C-H. As the Dih(C-C-O-C) increases, the phenoxide-CO_2 adduct gets increasingly unstable until it passes through the global maximum with respect to Dih(C-C-O-C), at 90°. However, this unstable perpendicular configuration is the global minimum configuration on the PES of 2,6-xylenoxide-CO_2 adduct (Figure 5b). The steric effect of ortho methyl groups destabilizes the in-plane configuration into a local maximum. On the contrary, the perpendicular configuration is not affected as much and becomes the global minimum.

In comparison, the PES of proton binding on phenoxides is not reshaped by bulky ortho groups. As is shown in Figure 5c, the in-plane configuration is the global minimum for proton binding on both phenoxide and 2,6-xylenoxide. The in-plane proton bindings on both molecules are contributed majorly by a σ bonding orbital formed from H 1s and the HOMO-1 of phenoxides (Figure 5d). In contrast, the perpendicular binding is contributed majorly by the σ bonding orbital that is formed from the HOMO of phenoxides. To sum up, protons bind to all substituted phenoxides through their HOMO in the same way, but CO_2 binds to their HOMO or HOMO-1 depending on the sterics of the ortho substituents. For the substituted phenoxides without bulky ortho substituents, the proton and CO_2 bindings are both associated to the HOMO, which is the origin of the previously observed LSR between pK_{CO_2} and pK_a. For the substituted phenoxides with bulky ortho

substituents, CO_2 and proton binds through different MOs whose energy level is influenced differently by substituents with different σ or π characteristics (inductive or resonance). In addition, due to the negative electrostatic potential (ESP) around the oxygens in CO_2 (Figure S5a), the phenoxide-CO_2 adduct would be less stabilized by the ortho C-H in phenoxide with ESP of -80 kcal/mol (Figure S5b) than by the ortho methyl in 2,6-xylenoxide with ESP of -73 kcal/mol (Figure 5c). The stabilization effect of ortho alkyl groups through non-covalent interaction (NCI) could be explicitly visualized by the NCI map. As the ortho substituent gets bulkier from -H (Figure S6a) to methyl (Figure S6b) and then to tert-butyl (Figure S6c), the green isosurface representing attractive NCI between the bound CO_2 and the ortho substituent becomes larger. As a result of the discussed effects, increasing the bulkiness of ortho substituents on phenoxide can decouple the CO_2 and proton binding energies. The system will then be allowed to move beyond the LSR and access the regions with more favorable energetics, namely, higher water-tolerance and stronger CO_2 binding.

Another advantage of the GA search is the large and diverse dataset it generates that could be utilized to train predictive machine learning (ML) models. As is shown in Figure S7, the neural network (NN) model trained on the GA dataset does an excellent job in predicting HOMO energy levels with R^2 of 0.926 and mean absolute error (MAE) of 86 meV. This could be especially helpful for scenarios where the MO energy level is a major descriptor of the target property, such as redox potential, excitation energies, and energetics of molecular binding that is associated with a specific MO. However, the NN

model turns out to predict the binding free energies rather poorly, with R² of 0.555 and MAE of 2.64 kcal/mol for ΔG_{bind}. To address the free energy, descriptors that contain information on the electronic structure and dynamics of the molecular fragments would be required. Still, the predictive ML model is cost-wise suited for initial screening for favorable energetics at the semi-quantitative level.

Experimental validation of theoretical trend and design principle
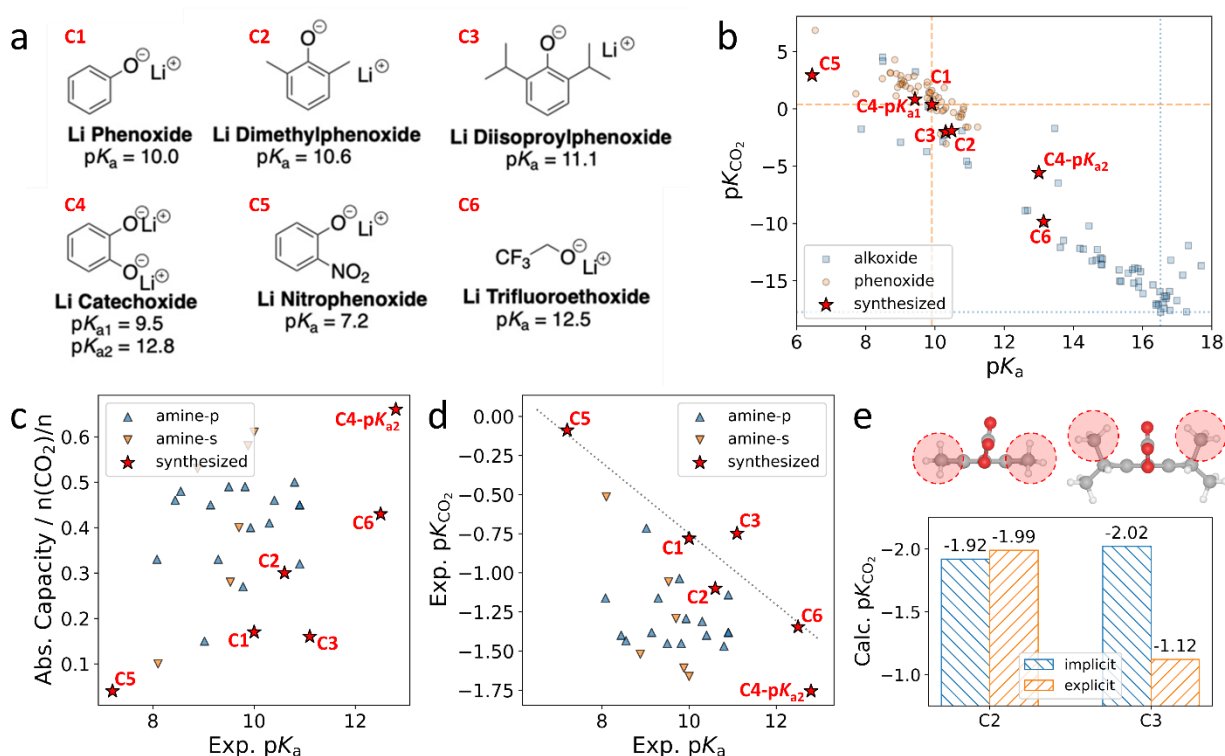

Figure 6. (a) Structures and experimental pK_a's of the phenoxide and alkoxide carriers synthesized in this work, labeled as C1-6. (b) Scatter plot of DFT-calculated pK_{CO_2} values versus the pK_a for synthesized molecules, with datapoints in Figure 2B also plotted as reference. The orange dashed line and the blue dotted line represent the data points of

phenoxide and methoxide, respectively. The experimental (c) CO_2 adsorption capacity and (d) pK_{CO_2} of the synthesized molecules, with datapoints of noncyclic primary and secondary amines from ref (19) as reference. (e) Geometry of global minimum configuration of the C2 and C3 molecules, and their corresponding pK_{CO_2} calculated by implicit or explicit solvation methods. Solvent inaccessible regions due to alkyl sterics are marked by red circles.

To further validate the trends and design principles obtained from computational screening and genetic algorithm search, we synthesized six compounds (Figure 6a) including lithium phenoxide (C1),(28) 2,6-dimethylphenoxide (C2), 2,6-diisopropylphenoxide (C3), catechoxide (C4),(29) 2-nitrophenoxide (C5), and lithium trifluoroethoxide (C6).(30) Their corresponding DFT-calculated pK_{CO_2} and pK_a are plotted in Figure 6b. C2 and C3 are located in the lower right region of the phenoxide cluster, with slight off-scaling relationship behaviour due to the sterics-induced change in bonding orbital as discussed in the previous section. C4 in di-anionic form (pK_{a2}) binds both proton and CO_2 more strongly than regular phenoxides. C6 is moved to the upper left along the $pK_{CO_2} - pK_a$ scaling relation of alkoxide by introducing trifluoromethyl (EWG), thereby trading part of the CO_2 binding strength for higher water-tolerance.

The CO_2 capturing ability of the compounds are characterized by absorption capacity, defined as the molar ratio of absorbed CO_2 to the capturing agent, determined using the experimental set-up (Figure S8) proposed in ref (19). For synthesized compounds in this

work, there exists a positive relationship between pK_a and absorption capacity. C2 and C6 have absorption capacities comparable to commercial amines (noncyclic, primary and secondary), and C4 outperforms all commercial amines. Notably, the time needed to reach the absorption equilibrium is c.a. 1 h for alkoxides and phenoxides, which is significantly faster than the case of amines (c.a. 3h) in previous reports.(19, 31) This can be attributed to the direct one-step adsorption pathway on phenoxide, unlike the amines which need to undergo protonation first. The facile CO_2 adsorption kinetics adds to the merit of alkoxide and phenoxides for DAC applications.

In Figure 6d, the absorption capacity is converted to the pK_{CO_2}, assuming the contribution from carbonate and bicarbonate pathways are minor. A similar $pK_{CO_2} - pK_a$ relation as in Figure 4b can be observed of the synthesized compounds. The trend for absorbance capacity of amines versus pK_a is less defined due to the use of pK_{a1} instead of pK_{a2} (no experimental values available). C4 shifts towards the lower left, away from the scaling relationship, probably due to the strong CO_2 binding in its di-anionic form, and more binding sites available (2 sites per molecule). The weaker correlation for C3 compared to C4 can be attributed to the weakness of the implicit solvation method employed throughout the study. Since propyl group is bulkier, the H-bond interaction between the bound CO_2 and its solvent environment is weaker for C3 than for C2 (Figure S9). In addition, the water-inaccessible regions caused by the sterics of the ortho groups are closer to the bound CO_2 in C3 than in C2, thus lowering the solvent-accessibility of the bound CO_2 more for C3. As the result, pK_{CO_2} calculated considering explicit water molecule is

lower than the implicit solvation results by 0.9 units for C3. In contrast, little change is observed for C2. These results suggests the overestimation of CO_2 binding for phenoxide with highly bulky ortho groups (> methyl), due to implicit solvation model giving the incorrect solvation free energy. After correcting the pK_a in Figure 6d using the explicit solvation results, C3 would shift closer to the scaling relation, while C2 stays about the same. The explicit solvation model, however, requires sampling hundreds of configurations per compound, which is not suitable for screening or global optimization search purposes. Hence, we use it only in the final refinement step after narrowing down the candidate pool using lower level methods. Possible approaches to incorporate evaluation of more realistic solvation free energy into the screening process includes QM/MM Monte Carlo (or other advanced sampling techniques)(32) and machine learning models trained on all-QM datasets.(33)

Although C2 is still not as good as C4, we note that it has been improved by 76.5% in absorption capacity (from 0.17 to 0.30) compared to the unsubstituted phenoxide C1, and its meta/para sites are still available for further substitution. Meanwhile, the catechoxide (C4) may also be further improved by introducing ortho bulky groups to exploit the steric effect which we are currently pursuing. In this work, we focus on presenting the computational workflow for inverse molecular design, hence compounds with relatively simple synthetic route are chosen mainly for the purpose of validation.

CONCLUSIONS

In conclusion, we explored amines, alkoxides, and phenoxides with a series of theoretical and computational methods in search of the optimal sorbent for aqueous DAC of CO_2. DFT calculations are first performed to study the bonding nature and energetics of auto-ionization, CO_2 binding, and deprotonation on substituted amines. The anionic deprotonated amino group is found to be the species that binds CO_2. We discovered a convex hull relationship that prohibits optimization of both ΔG_{bind} (binding strength) and $\Delta G_{auto-ionization}$ (concentration of the binding species), and a LSR between pK_{CO_2} and pK_{a2} of amines. Alkoxides and phenoxides are then proposed as better sorbents for their improved water-tolerance and more favorable capture stoichiometry. All three molecular systems are found to bind CO_2 in a primarily s-fashion and follow the same pK_{CO_2}- pK_a LSR. Several high-performance alkoxides are proposed from the computational screening. Since there is an opportunity of pK_{CO_2}- pK_a decoupling on phenoxides, we apply GA global optimization combined with SQM calculations to explore the large chemical space of substituted phenoxides. Several promising molecules with off-LSR energetics are discovered from the GA searches and validated by DFT calculations. The top-scoring molecules are then studied by bonding analysis and PES scan to understand the origin of the off-LSR behavior: bulky ortho substituents force the CO_2 adduct into a perpendicular configuration that binds through HOMO, while the proton still binds through HOMO-1 in the in-plane configuration, thus decoupling the pK_{CO_2} and pK_a. Several substituted phenoxide and alkoxide molecules are synthesized for CO_2 absorption capacity measurement to validate the computational results. The overall pK_{CO_2}- pK_a trend and off-

LSR behaviour of specific molecules (bulky groups or hydroxyl on ortho positions) is observed in the experiments, consistent with theory. The inverse molecular design workflow presented in this work is highly generalizable and can be readily adapted for optimizing other functional molecules for various applications.

METHODS

Computational Methods

Geometry optimization, vibrational analysis, and implicit solvation model calculations are performed using the Gaussian 16 program(34) (Revision C.01). The geometry optimizations and potential energy surface (PES) scans are performed using B3LYP functional(35, 36) with def2-TZVP basis sets(37) and D3 correction (Becke-Johnson damping)(38) to better account for the dispersion interactions. Harmonic vibrational frequencies are computed on each optimized geometry to make sure that all reaction intermediates have no imaginary frequency. The entropic and thermo-statistical contributions (δG_{RRHO}) are calculated by the rigid rotor harmonic oscillator (RRHO) approximation on optimized geometries at 298.15 K and 1 atm. To account for the aqueous environment, solvation free energies (δG_{solv}) are calculated by the implicit SMD model(39) with M05-2X functional(40) and 6-31G* basis set, which is the level of theory where the SMD is parameterized against experimental datasets.(41) The accurate electronic energies (E_{el}) are computed at PWPB95 double hybrid functional(42) with def2-QZVPP basis

sets(37) and D3 correction using ORCA program(43) (Version 4.2.1). The Gibbs free energy of solvated species at 298.15 K and 1 M can then be calculated by:

$$G = E_{el} + \delta G_{RRHO} + \delta G_{solv} + \delta G^{1\,atm\to1\,M}$$

where the $\delta G^{1\,atm\to1\,M}$ term is the needed free energy to increase concentration of the species from gas phase (4.088×10^{-2} mol/L) to 1 M in solution, which is 1.89 kcal/mol at 298.15 K and 1 atm pressure. The Gibbs free energy of proton, although cannot be calculated quantum mechanically, can be derived using thermodynamics and the Sackur-Tetrode equation, giving $H_{gas}^o(\text{H}^+)$=5/2 RT=1.48 kcal/mol and S_{gas}^o=26.5 cal/(mol·K) at 298.15 K and 1 atm. The value of experimentally determined δG_{solv} is taken from reference (44).

The free energy change of the proton dissociation process of HA can then be calculated by:

$$\Delta G_{deprot} = G(\text{H}^+) + G(\text{A}^-) - G(\text{HA})$$

The acidity constant $\text{p}K_\text{a}$ can be calculated from ΔG_{deprot} via the linear free energy relationship (LFER):(23)

$$\text{p}K_\text{a} = c_0 \cdot \frac{\Delta G_{deprot}}{\ln(10)\,\text{RT}} + c_1$$

where RT is the ideal gas constant times the temperature (298.15 K). The c_0 and c_1 parameters are fitted to experimental data obtained from the iBonD database (http://ibond.chem.tsinghua.edu.cn). For amines, the $\text{p}K_{\text{a1}}$ and $\text{p}K_{\text{a2}}$ are related to the

first and second deprotonation of the corresponding protonated primary or secondary amino group, respectively.

The CO_2 binding constant pK_{CO_2} of a capturing agent A^- is calculated similarly from the free energy change of the CO_2 binding process (LFER parameters for pK_{CO_2} are set to $c_0 = 1$ and $c_1 = 0$ due to lack of available experimental data):

$$\Delta G_{bind} = G(A^- * CO_2) - G(A^-) - G(CO_2)$$

$$pK_{CO_2} = c_0 \cdot \frac{\Delta G_{bind}}{\ln(10)\,RT} + c_1$$

According to our test, the DFT protocol of this work outperforms DLPNO-CCSD(T)/def2-QZVPP(45) and quantum chemistry composite method CBS-QB3(46) in reproducing experimental pK_a of alkoxides (Figure S1).

Molecular orbital analysis, Hirshfeld population analysis, non-covalent interaction analysis, and electrostatic potential mapping are performed using the Multiwfn program on the converged wavefunctions from DFT calculation.(47) The configurational sampling under explicit solvation is performed using genmer module in the Molclus program.(48)

In the computational screening section, the substituent pool includes (24 in total): -H, -OCH$_3$, -CN, -CH$_3$, -C$_2$H$_5$, -CH(CH$_3$)$_2$, -CHCH$_2$, -CCH, -CH$_2$Cl, -CF$_3$, -CHO, -COCH$_3$, -COOCH$_3$, -OCOCH$_3$, F, Cl, Br, -OC$_2$H$_5$, -OCHO, -SO$_2$CH$_3$, -SO$_2$OCH$_3$, -SOCH$_3$, -NO$_2$, -CHNH. Note that the hydroxyl and amino groups are not included in the pool to avoid difficulty in pK_a determination of polyamines and alkanolamines.

Each molecule is represented by a 1D vector with 5 (number of substitution sites) elements, each representing a substituent. The representation (noted as gene representation of molecules) can reversibly interconvert into or from a SMILES representation.(49) The SMILES string is converted to XYZ coordinate using Open Babel package,(50) and sufficient stochastic conformational search is performed at MMFF94 level to obtain the most stable conformation.(51)

Connectivity is checked after geometry optimization to make sure there is no unexpected bond dissociation which suggests instability of the molecule under certain charge states. A data point is discarded if any of the involved species (neutral, deprotonated, CO_2 adduct) is found unstable.

Semiempirical quantum mechanical (SQM) calculations are performed using the xTB package(52) for the geometry optimization and energy evaluation throughout the genetic algorithm search. GFN1-xTB tight binding method with GBSA model for describing implicit solvation by water is employed in the high-throughput computational screening section for its low computational cost and comparable accuracy to DFT methods in terms of geometry and thermochemistry.(53) The pK_a and pK_{CO_2} values are calculated following the fitting and LFER procedure as described in the DFT section.

The genetic algorithm search is performed using an adapted version of the molGA code.(15) The population size, mutation rate, and convergence criterion are set to 100 candidates, 33%, and 100 generations, respectively. The search goal is set to minimize the ΔG_{bind} while keeping the ΔG_{deprot} higher than that of OH⁻. Ten independent GA searches are

performed, and the top-scoring candidates are collected from the final populations. DFT calculations are performed on those candidates thereafter to obtain more accurate energetics.

The machine learning (ML) model for fast prediction of molecular properties is a multi-layer perceptron neural network (NN) with four rectified linear units (ReLU) and two hidden linear layers, implemented using pyTorch library.(54) Each phenoxide molecule is converted into a 1D vector with 24 (number of substituents) ×5 (number of substitution sites)=120 binary elements via positional one-hot encoding (Scheme S1). The NN is trained on the dataset of sampled substituted phenoxides labeled with SQM-calculated properties, with 80% of the data as training set and 20% of the data as test set. Data augmentation is achieved by "flipping" the molecule to generate two equivalent representations for each data point.

Experimental Methods

Synthesis and manipulation of compounds were carried out in open air unless otherwise mentioned. For air- and moisture-sensitive procedures, manipulations were carried out in a glovebox or using standard Schlenk techniques under inert atmosphere of nitrogen. Solvents used during inert atmosphere synthesis and/or manipulations were degassed by sparging with argon and dried by passing through columns of neutral alumina or molecular sieves. All deuterated solvents were purchased from Cambridge Isotope Laboratories, Inc. Deuterated methanol was degassed and stored over activated 3 Å molecular sieves prior to use. All solvents and reagents were purchased from commercial vendors and used

without further purification unless otherwise noted. The compounds in Figure 6a are synthesized using two routes:

Synthesis using LiOH: This synthetic route was used for deprotonation of trifluoroethanol, hexafluoropropanol, 2-nitrophenol, and 2,6-dimethylphenol. Under an N_2 atmosphere, 25 mmol of alcohol was combined with 25 mmol of lithium hydroxide in 20 mL dry methanol. The reaction was refluxed overnight then dried under vacuum to give solid.

Synthesis using n-butyl lithium: This synthetic route was used to deprotonate phenol, 2,6-diisopropylphenol, and catechol. In a rigorously dried N_2 atmosphere with 50 mL dried pentane from the solvent system and 25 mmol of dried alcohol, 25 mmol (or 50 mmol for catechol) 1.6 M n-butyl lithium in hexane was added dropwise at -78°C. The mixture was stirred at -78°C for one hour, and then stirred for an additional 24 hours at room temperature. Solvent was removed under reduced pressure to leave a colorless solid.

NMR spectroscopy was used to confirm the identity and purity of the synthesized compounds. ^{1}H NMR spectroscopy was performed on a 500MHz Bruker Avance GN500 with a BBO probe or on a 500 MHz Bruker DRX 500 spectrometer with a TCI cryoprobe. ^{13}C{^{1}H} NMR spectra were recorded on a 500MHz Bruker DRX 500 fitted with a TCI cryoprobe. All NMR spectra were acquired at room temperature and referenced to residual ^{1}H or ^{13}C resonances of the deuterated solvent (^{1}H: CD$_3$OD, δ 3.31; D$_2$O, δ 4.79; DMSO-D$_6$, δ 2.50) (^{13}C: CD$_3$OD, δ 49.00; D$_2$O, δ -; DMSO-D$_6$, δ 39.52). Electrospray ionization mass spectrometry was preformed using an ESI LC-TOF Micromass LCT. Infrared spectroscopy was performed using a Thermo Scientific Nicolet iS5 spectrophotometer with

an iD5 ATR attachment. UV-Visible absorption spectra were collected in methanol using an Agilent Technologies Cary 60 UV-Vis.

The CO_2 absorption capacity measurements were made following the procedure previously described in ref 19 by Puxty *et al.* 10 mL 0.5 M alkoxide solution was placed in weighed 20 mL vial with septum screw top and stir bar (Figure S8a). First, the mass change due to evaporation was recorded by placing the vial in 40 °C bath and sparging with N_2. The inlet needle was never placed directly into solution, only the headspace. The change in mass was measured 8 times over a period of 20 minutes. Next, the gas inlet was changed to 10% CO_2. The change in mass was measured every minute for the first 10 minutes, and then every 5 minutes for an hour. To determine the overall CO_2 absorption, the mass change due to evaporation alone was subtracted from mass change when 10% CO_2 was used to give the total mass gained due to CO_2 absorption (Figure S8b). This method was validated against the original data from Puxty *et al.* using ethylenediamine as the standard.(19)

ASSOCIATED CONTENT

Supporting Information.

The following files are available free of charge.

Schematics of the conversion between data representations; LFER fitting plots of DFT and SQM calculations; Scatter plots showing correlation among pK_{CO_2}, pK_{a1}, and pK_{a2} of amines; the evolution diagram of the ΔG_{bind} during the GA searches; the ESP and NCI

map of the investigated substituted phenoxides and CO_2 adducts; validation of the trained NN model; experimental set-up for CO_2 absorption capacity measurement; the H-bond geometry and bond order between water and CO_2-bound phenoxides.

AUTHOR INFORMATION

Corresponding Author

*Corresponding Author's email:

Anastassia N. Alexandrova: ana@chem.ucla.edu ,

Jenny Y. Yang: j.yang@uci.edu

Notes

The authors declare no competing financial interest.

References

1. T. L. Froelicher, D. J. Paynter, Extending the relationship between global warming and cumulative carbon emissions to multi-millennial timescales. *Environ. Res. Lett.* 10, 75002 (2015).

2. W. Zhang, *et al.*, Progress and perspective of electrocatalytic CO2 reduction for renewable carbonaceous fuels and chemicals. *Adv. Sci.* 5, 1700275 (2018).

3. P. Kaiser, R. B. Unde, C. Kern, A. Jess, Production of liquid hydrocarbons with CO2 as carbon source based on reverse water-gas shift and Fischer-Tropsch synthesis. *Chemie Ing. Tech.* 85, 489–499 (2013).

4. R. J. Millar, M. R. Allen, Understanding the Role of CCS Deployment in Meeting Ambitious Climate Goals. *Carbon Capture and Storage* 26, 8 (2019).

5. B. Dutcher, M. Fan, A. G. Russell, Amine-based CO2 capture technology development from the beginning of 2013⊠ A Review. *ACS Appl. Mater. Interfaces* 7, 2137–2148 (2015).

6. Y. W. Abraha, C.-W. Tsai, J. W. H. Niemantsverdriet, E. H. G. Langner, Optimized CO2 Capture of the Zeolitic Imidazolate Framework ZIF-8 Modified by Solvent-Assisted Ligand Exchange. *ACS omega* 6, 21850–21860 (2021).

7. R. Lyndon, *et al.*, Visible light-triggered capture and release of CO2 from stable metal organic frameworks. *Chem. Mater.* 27, 7882–7888 (2015).

8. S. E. Renfrew, D. E. Starr, P. Strasser, Electrochemical Approaches toward CO2

Capture and Concentration. *ACS Catal.* 10, 13058–13074 (2020).

9.   C. Cheng, *et al.*, Amine-based post-combustion CO2 capture mediated by metal ions: Advancement of CO2 desorption using copper ions. *Appl. Energy* 211, 1030–1038 (2018).

10. S. Mathew, *et al.*, Dye-sensitized solar cells with 13% efficiency achieved through the molecular engineering of porphyrin sensitizers. *Nat. Chem.* 6, 242 (2014).

11. Y. Ding, C. Zhang, L. Zhang, Y. Zhou, G. Yu, Molecular engineering of organic electroactive materials for redox flow batteries. *Chem. Soc. Rev.* 47, 69–103 (2018).

12. M. Koerstz, A. S. Christensen, K. V Mikkelsen, M. B. Nielsen, J. H. Jensen, High throughput virtual screening of 230 billion molecular solar heat battery candidates. *PeerJ Phys. Chem.* 3, e16 (2021).

13. C. Yang, *et al.*, Molecular engineering of conjugated polybenzothiadiazoles for enhanced hydrogen production by photosynthesis. *Angew. Chemie Int. Ed.* 55, 9202–9206 (2016).

14. X. Zhang, *et al.*, Molecular engineering of dispersed nickel phthalocyanines on carbon nanotubes for selective CO 2 reduction. *Nat. Energy* 5, 684–692 (2020).

15. Z. Zhang, Y.-G. Wang, Molecular Design of Dispersed Nickel Phthalocyanine@Nanocarbon Hybrid Catalyst for Active and Stable Electroreduction of CO2. *J. Phys. Chem. C* 125, 13836–13849 (2021).

16. Z. Zhang, T. Yang, P. Qin, L. Dang, Nickel bis (dithiolene) complexes for electrocatalytic hydrogen evolution: A computational study. *J. Organomet. Chem.* 864, 143–147 (2018).

17. Y. Wang, *et al.*, Theory-driven design of electrocatalysts for the two-electron oxygen reduction reaction based on dispersed metal phthalocyanines. *CCS Chem.*, 1–21 (2021).

18. G. T. Rochelle, Amine scrubbing for CO2 capture. *Science (80-.).* 325, 1652–1654 (2009).

19. G. Puxty, *et al.*, Carbon dioxide postcombustion capture: a novel screening study of the carbon dioxide absorption performance of 76 amines. *Environ. Sci. Technol.* 43, 6427–6433 (2009).

20. I. M. Bernhardsen, H. K. Knuutila, A review of potential amine solvents for CO2 absorption process: Absorption capacity, cyclic capacity and pKa. *Int. J. Greenh. Gas Control* 61, 27–48 (2017).

21. G. F. Versteeg, L. A. J. Van Dijck, W. P. M. van Swaaij, On the kinetics between CO2 and alkanolamines both in aqueous and non-aqueous solutions. An overview. *Chem. Eng. Commun.* 144, 113–158 (1996).

22. M. Caplow, Kinetics of carbamate formation and breakdown. *J. Am. Chem. Soc.* 90, 6795–6803 (1968).

23. A. Klamt, F. Eckert, M. Diedenhofen, M. E. Beck, First principles calculations of

aqueous p K a values for organic and inorganic acids using COSMO– RS reveal an inconsistency in the slope of the p K a scale. *J. Phys. Chem. A* 107, 9380–9386 (2003).

24. M. Anand, B. Rohr, M. J. Statt, J. K. Nørskov, Scaling Relationships and Volcano Plots in Homogeneous Catalysis. *J. Phys. Chem. Lett.* 11, 8518–8526 (2020).

25. T. Bligaard, *et al.*, The Brønsted–Evans–Polanyi relation and the volcano curve in heterogeneous catalysis. *J. Catal.* 224, 206–217 (2004).

26. A. N. Alexandrova, H·(H2O) n clusters: Microsolvation of the hydrogen atom via molecular ab initio gradient embedded genetic algorithm (GEGA). *J. Phys. Chem. A* 114, 12591–12599 (2010).

27. P. Pracht, S. Grimme, Efficient Quantum-Chemical Calculations of Acid Dissociation Constants from Free-Energy Relationships. *J. Phys. Chem. A* 125, 5681–5692 (2021).

28. R. Den Besten, S. Harder, L. Brandsma, "A method for the determination of the degree of association of organolithium compounds in liquid ammonia" (Elsevier Sequoia S.A, 1990).

29. A. E. Bayliff, M. R. Bryce, R. D. Chambers, "Polyhalogenoheterocyclic Compounds. Part 38.' Reactions of Fluorinated-Alkenes and-Cycloalkenes with Difunctional Nucleophiles" (1987).

30. Z. Z. Yang, L. N. He, Efficient CO2 capture by tertiary amine-functionalized ionic liquids through Li,+-stabilized zwitterionic adduct formation. *Beilstein J. Org. Chem.* 10, 1959–1966 (2014).

31. X. Yang, *et al.*, Computational Modeling and Simulation of CO2 Capture by Aqueous Amines. *Chem. Rev.* 117, 9524–9593 (2017).

32. A. N. Alexandrova, W. L. Jorgensen, On the Mechanism and Rate of Spontaneous Decomposition of Amino Acids. *J. Phys. Chem. B* 115, 13624–13632 (2011).

33. Y. Basdogan, *et al.*, Machine Learning-Guided Approach for Studying Solvation Environments. *J. Chem. Theory Comput.* 16, 633–642 (2020).

34. M. J. Frisch, *et al.*, Gaussian 16 (2016).

35. A. D. Becke, Becke's three parameter hybrid method using the LYP correlation functional. *J. Chem. Phys* 98, 5648–5652 (1993).

36. J. Tirado-Rives, W. L. Jorgensen, Performance of B3LYP density functional methods for a large set of organic molecules. *J. Chem. Theory Comput.* 4, 297–306 (2008).

37. F. Weigend, R. Ahlrichs, Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* 7, 3297–3305 (2005).

38. S. Grimme, S. Ehrlich, L. Goerigk, Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* 32, 1456–1465 (2011).

39. A. V Marenich, C. J. Cramer, D. G. Truhlar, Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* 113, 6378–6396 (2009).

40. Y. Zhao, N. E. Schultz, D. G. Truhlar, Design of density functionals by combining the method of constraint satisfaction with parametrization for thermochemistry, thermochemical kinetics, and noncovalent interactions. *J. Chem. Theory Comput.* 2, 364–382 (2006).

41. J. Ho, A. Klamt, M. L. Coote, Comment on the correct use of continuum solvent models. *J. Phys. Chem. A* 114, 13442–13444 (2010).

42. L. Goerigk, S. Grimme, Efficient and Accurate Double-Hybrid-Meta-GGA Density Functionals—Evaluation with the Extended GMTKN30 Database for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions. *J. Chem. Theory Comput.* 7, 291–309 (2011).

43. F. Neese, F. Wennmohs, U. Becker, C. Riplinger, The ORCA quantum chemistry program package. *J. Chem. Phys.* 152, 224108 (2020).

44. K. S. Alongi, G. C. Shields, "Theoretical calculations of acid dissociation constants: a review article" in *Annual Reports in Computational Chemistry*, (Elsevier, 2010), pp. 113–138.

45. E. Paulechka, A. Kazakov, Efficient DLPNO–CCSD(T)-Based Estimation of

Formation Enthalpies for C-, H-, O-, and N-Containing Closed-Shell Compounds Validated Against Critically Evaluated Experimental Data. *J. Phys. Chem. A* 121, 4379–4387 (2017).

46.   G. P. F. Wood, *et al.*, A restricted-open-shell complete-basis-set model chemistry. *J. Chem. Phys.* 125, 94106 (2006).

47.   T. Lu, F. Chen, Multiwfn: a multifunctional wavefunction analyzer. *J. Comput. Chem.* 33, 580–592 (2012).

48.   T. Lu, Molclus program. *Beijing Kein Res. Cent. Nat. Sci. China* (2016).

49.   D. Weininger, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* 28, 31–36 (1988).

50.   N. M. O'Boyle, *et al.*, Open Babel: An open chemical toolbox. *J. Cheminform.* 3, 33 (2011).

51.   N. Yoshikawa, G. R. Hutchison, Fast, efficient fragment-based coordinate generation for Open Babel. *J. Cheminform.* 11, 49 (2019).

52.   S. Grimme, XTB, a tight-binding quantum chemistry code for large molecules. *Univ. Bonn* (2019).

53.   C. Bannwarth, S. Ehlert, S. Grimme, GFN2-xTB—An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *J. Chem.*

Theory Comput. 15, 1652–1671 (2019).

54.   A. Paszke, *et al.*, Pytorch: An imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* 32, 8026–8037 (2019).