

Evaluating the Potential Environmental and Human Toxicity
of Solvents Proposed for use in Post-Combustion Carbon Capture

by

Fatima Ghiasi

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Master of Applied Science

in

Systems Design Engineering

Waterloo, Ontario, Canada, 2025

© Fatima Ghiasi 2025

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Carbon dioxide emitted by industrial activities is a growing concern due to the effects on global climate. For this reason, firms are being urged to lower their carbon footprint. Post combustion carbon capture is being explored as a method for the power and materials industries to decarbonize. The most mature technique of carbon capture is amine absorption.

Different amines are being explored to potentially be used within post-combustion carbon capture units. Many biological molecules are amines, and amines that resemble them can disrupt biological processes, harming organisms. In addition, if an amine is soluble within lipids, it can persist within the food chain and cause long term toxic effects that are not immediately visible. 151 solvents were compared based on four properties: volatility, lipophilicity, mutagenicity, and neuroactivity. Machine learning models were trained to predict these values. Due to their hydrophilicity, amino acids were determined to have the lowest potential of causing environmental toxicity.

Acknowledgements

The past few years have been difficult time for me, with both physical and psychological trauma, and metaphorical wounds from 13 years ago being opened. However, despite all of the turmoil, there were loyal friends that always had my back. I am very grateful for my professor, Dr. Ali Elkamel, for his unconditional support of me. I am also thankful to God, for helping me meet empathetic and kind people.

Table of Contents

List of Figures	vii
List of Tables.....	ix
Introduction	1
Post Combustion Carbon Capture.....	3
Carbon Dioxide Separation	4
Cryogenic Gas Separation	5
Carbonate Looping.....	6
Adsorption.....	7
Membranes.....	8
Amine-based Post Combustion Carbon Capture	9
Damage	11
Dissipation.....	12
Overview of Different Classes of Solvents Considered as Candidates for Carbon Capture.....	14
Non-Cyclic Amine Based Solvents.....	15
Alkanolamine Based Solvents	15
Amino Acid Based Solvents.....	15
Other Non-Cyclic Amines	16
Cyclic Amine Based Solvents.....	17
Piperazine and Derivatives.....	17
Piperidine and Derivatives	17
Pyrrolidine and Derivatives.....	18
Azoles	19
Morpholines and Derivatives	20
Pyridines and Derivatives.....	20
Criteria for Comparing Toxicity	21
Volatility	21
Lipophilicity.....	22
Mutagenicity	23
Neuroactivity.....	24
Review of Previous Models used to Predict Chemical Properties	26

Mechanistic Computational	26
Empirical Machine Learning.....	27
Convolutional Neural Networks	29
Methodologies	32
Training and Test Dataset	32
Solvents to be Analyzed	32
Encoding Chemical Structures	32
Chemical Fingerprints	33
2 Dimensional Image Representations	33
Candidate Machine Learning Models	34
Neural Networks	36
Optimization.....	37
Results of the Property Prediction Models.....	39
Volatility	39
Lipophilicity	43
Mutagenicity	47
Ames Classification Test	47
Nitramine and Nitrosamine Volatility	51
Neuroactivity.....	53
Acetylcholinesterase Affinity	53
Nicotinic Affinity.....	57
Cross-Comparison of Lipophilicity and Neuroactivity.....	62
Conclusion.....	66
References.....	68
Appendix	90

List of Figures

Figure 1 Post combustion carbon capture amine scrubbing process, diagram used with permission from Ghiasi et al. [107]	10
Figure 2 Overview of the Ames Mutagenicity Test	23
Figure 3 A binary decision tree with a depth of 3, diagram used with permission from Ghiasi et al. [107]	35
Figure 4 A single computational neuron, diagram used with permission from Ghiasi et al. [107]	36
Figure 5 A multilayer perceptron with a single hidden layer, diagram used with permission from Ghiasi et al. [107]	36
Figure 6 Convolution operation, diagram used with permission from Ghiasi et al. [107]	37
Figure 7 Distributing data into k train-test data groupings, for k-fold cross validation	38
Figure 8 Relationship between the temperature and the pressure in the train-test data	39
Figure 9 Structure of Modular Neural Networks created to predict vapor pressures	40
Figure 10 Pair plot of the predicted logarithm of vapor pressures at temperatures from 5 to 95 °C, for solvents and carbamates. Color corresponds with weight. 5 models with the least RMSE test error were included for space.	42
Figure 11 Histogram of the predicted lipophilicities of both cyclic and non-cyclic amines	45
Figure 12 Histogram of the predicted lipophilicities of different classes of non-cyclic amines	45
Figure 13 Mean mutagenicity predicted by all models for different numbers of amine groups	50
Figure 14 Pair plot of the predicted logarithm of vapor pressures at temperatures from 5 to 95 °C, for nitramines and nitrosamines of solvents and carbamates. The 5 models with the least RMSE test error were included for space.	52
Figure 15 Pairplot of predicted solvent and carbamate AChE affinity (log Ki nM) made using MACCS key trained models	55
Figure 16 Histogram of the difference between the logarithmic affinity to AChE between solvents and carbamates	56
Figure 17 Histogram of the AChE affinities (log Ki nM) predicted by different models for both cyclic and non-cyclic amines	56
Figure 18 Histogram of the predicted logarithmic affinities (log ki) for nAChR a4b2	59
Figure 19 Histogram of the predicted logarithmic affinities (log ki) for nAChR a7	59
Figure 20 Histogram of the difference in nicotinic affinity for solvents and their respective carbamates	60
Figure 21 All predicted affinities for nAChR a4b2 and nAChR a7	61
Figure 22 Comparison between lipophilicity and neuroactivity predictions for solvents and carbamates	62
Figure 23 Log Kaq of solvents and carbamates for Acetylcholinesterase	63
Figure 24 Correlations between different models for logarithm of vapor pressure	90
Figure 25 Correlations between different models for lipophilicity	91

Figure 26 Counts of solvents and carbamates in each bin for predicted mutagenicity probability	92
Figure 27 Correlations between different models for AChE affinity	93

List of Tables

Table 1 Major gaseous constituents of flue gas	4
Table 2 Optimized Parameters for neural network models used to predict the vapor pressures based on different chemical representations and temperature	40
Table 3 Solvents with highest and lowest vapor pressures at 40°C, based on mean of 5 best performing models	43
Table 4 Hyperparameters for Fingerprint Models trained to predict Lipophilicity	44
Table 5 Hyperparameters for CNNs trained to predict Lipophilicity	44
Table 6 Solvents predicted to be the most lipophilic and the least lipophilic	46
Table 7 Ames mutagenicity raw experimental classification, and pooled training data	47
Table 8 Test performance of different mutagenicity prediction models	47
Table 9 Mean solvent and carbamate mutagenicity probability predictions of different models	49
Table 10 Nitrosated solvents with highest and lowest vapor pressures at 40°C, based on mean of 5 best performing models	52
Table 11 Hyperparameters for Fingerprint Models trained to predict Acetylcholinesterase Affinity	53
Table 12 Hyperparameters for CNNs trained to predict Acetylcholinesterase Affinity	54
Table 13 Hyperparameters for Fingerprint Models trained to predict nAChR $\alpha 4\beta 2$ affinity	57
Table 14 Hyperparameters for Fingerprint Models trained to predict nAChR $\alpha 7$ affinity	57
Table 15 Hyperparameters for CNNs trained to predict Nicotinic Affinity	58
Table 16 Solvents predicted to be the most and the least neuroactive based on Kaq	63

Introduction

Energy is the most important commodity in the economy. It is a currency which can be exchanged for the ability to power industrial processes that extract raw materials, produce goods, and transport products. The industrial revolution was kickstarted by the development of technology that could convert chemical energy stored in coal into mechanical energy [1]. Ever since that era, fossil fuels have been a principal source of energy.

The combustion of fossil fuels produces carbon dioxide as a byproduct. This has caused the atmospheric carbon dioxide concentrations to increase from the pre-industrial level of 280 ppm [2] to 418 ppm as of 2023 [3]. It has been known for over a century that CO₂ absorbs and re-radiates radiant energy back to the earth's surface [4]. This phenomenon is the principal cause of the increase in global average temperature. The increase in average global temperature produces an array of consequences, including increases in the occurrence of extreme weather [5], heat waves [6], forest fires, habitat destruction [7], droughts, and extinction [8].

To minimize these effects, the rate of CO₂ emissions should be decreased. In the 2015 Paris agreement, the majority of countries pledged to limit their CO₂ emissions [9]. The release of greenhouse gases into the atmosphere can be mitigated by reducing the demand for power and by using less carbon-intensive methods of generating useable energy.

It is in the nature of private entities to reduce their costs. Since any form of useable energy has a monetary value, it is in the best interest of different industries to introduce updated technology and processes that lower the amount of energy that has to be used to provide the products or services. 35% of the energy consumed globally is used for transportation, of which 60% is for passenger cars [10]. Currently used ways of reducing the energy consumption and CO₂ emission of passenger cars include making the car lighter, designing it to be aerodynamic [11], and the use of automotive sensors and microcontrollers [12]. Electric vehicles have been marketed as carbon neutral, however, the electricity that they use likely originates from fossil fuel combustion plants: as of 2022, 63.3% of electricity was produced using fossil fuels, the majority being coal [13].

In 2019, the total global energy consumption was 418 EJ, of which 20% was in the form of electrical energy [10]. Historical trends show that both the amount and proportion of energy consumption in the form of electrical energy has been steadily increasing. It is reasonable to expect that this trend will continue. With the increasing market penetration for electric vehicles, electrical energy demand will quickly increase to include a portion of the 21% of total energy used for passenger vehicles [10]. A way to substantially cut down on total carbon emissions is to generate electricity using methods that produce less CO₂.

Thermonuclear electricity generation is the leading non-fossil fuel based electrical energy source. alternative energy source. It is a mature technology which has been widely adopted; accounting for 11% of worldwide [14] and 60% of Ontario's electricity generation [15]. However,

even though it does not directly contribute to the release of greenhouse gases, nuclear power is polluting. Uranium, the most commonly used fuel for nuclear power, is found as a component of minerals, such as pitchblende, below the earth's surface. The most common method of extracting uranium is in-situ leaching, where an acidic solution, such as sulfuric acid, is injected into rock formations to dissolve minerals containing valuable elements [16]. The solution is then pumped out through a recovery well, and the desired element is extracted from the solution. The process of in-situ leaching inherently transforms immobile heavy metal species into water soluble species. These mobile species contaminate groundwater [17].

Mine tailings are the by-products of the process of extracting valuable minerals from the raw ore. The process of uranium mining produces heavily contaminated mine tailings. Farmlands that are within the vicinity of uranium mine tailing ponds have been found to have extremely elevated concentrations of heavy metals including arsenic, cadmium, and mercury [18]. The byproduct of the thermonuclear reaction process, spent fuel, is both highly radioactive and retains this high radioactivity for over hundreds of thousands of years [19]. This makes waste disposal highly problematic. In addition to the controversy arising from constructing projects which are polluting and carry a risk of causing large environmental disasters [20], other barriers to the adoption of nuclear energy include both geopolitical restrictions in obtaining radioactive material, and the high initial cost of building nuclear reactors.

Hydroelectricity is another continuous renewable energy source. It accounts for 61% of the electricity produced in Canada [13] and 16% of the electricity produced globally [10]. Hydroelectric projects can be implemented in a wide array of sizes, from small setups that can power a house in a remote community, to large structures that can power multiple cities. As of 2023, the largest power station in the world, the Three Gorges Dam, is a hydroelectric dam with a capacity of 22.5 GW [21].

The construction of a dam changes the shape of the water distribution, increasing the height of the water upstream causing dry land to flood, decreasing the height of the water downstream, and dividing the aquatic habitat into two [22]. Habitat division due to dams has demonstrated impacts on the migration patterns of several species of fish, causing their populations to drop [23]. Structures such as fish ladders are designed to mitigate the disruptions to migration routes, however, not all fish can travel through them [24].

Hydroelectric plants also emit greenhouse gases. Methane is produced endogenously within waterways by the degradation of organic material [25]. In unaltered conditions, the majority of this methane becomes oxidized by aerobic bacteria. However, due to the rapid depressurization when water passes through the turbines of a dam, the dissolved methane is released as a gas [26]. These gaseous emissions can have a significant contribution to the greenhouse effect [27], and the amount of released methane depends on the habitat where the dams are constructed. In some habitats, such as tropical rainforests, the warming potential of methane emitted by dams is larger than the warming potential of CO₂ released by equivalently sized fossil fuel plants [28].

As of 2022, after hydroelectricity, wind power provides the second largest volume of renewable electricity production [29]. Wind turbines can be installed on land that is used for pastures, meadows and croplands, with minimal disruption. They can also be placed on floating offshore platforms in the middle of lakes and oceans [30] [31]. A problem with wind power is that wind speeds, and consequently wind power generation, is intermittent, and changes rapidly with local weather conditions [32] [33].

Another rising renewable energy technology is solar power. There are two types of solar energy generation systems, concentrated solar power and photovoltaics. Photovoltaic solar energy is a technology which has been rapidly improving, and the cost for each unit of solar PV capacity has been decreasing exponentially [34]. One of the major drawbacks of solar power is that the amount of electrical power that can be generated depends on the area of land that can be devoted for this purpose. This land use competes with land that can be used for agriculture or for carbon-capturing forests. In addition, solar power can only be generated intermittently, during daytime hours.

In 2019, the total amount of electric energy demand within Ontario was 138.8 TWh [35], or an average power demand of 15.8 GW. In order to be able to depend almost exclusively on intermittent renewable energy resources, there needs to be an extensive grid of energy storage units. If it's assumed that only 12 hours of mean energy usage needs to be stored by the network, that would amount to 190 GWh of energy storage needing to be installed. The construction of adequate renewable electricity production and energy storage infrastructure is a large investment. For this reason, dispatchable fossil fuel electricity generation plants are unlikely to be phased out in the near future. However, existing fossil fuel combustion plants can emit less carbon dioxide by being retrofitted with Carbon Capture units.

Post Combustion Carbon Capture

The essence of Carbon Capture Usage or Storage (CCUS) is to redirect carbon dioxide being emitted from an industrial activity into either being used for a purpose or being placed into long term storage. Carbon Capture is not limited to solely being used in the electricity generation sector; it can also be used in plants where fossil fuels are combusted for direct thermal [36] and mechanical energy. It is also a feasible way to mitigate the large greenhouse gas emissions associated with the production of steel [37] [38] and concrete [39] [40], two materials that form the backbone of modern infrastructure.

CO₂ can be used without being isolated from other gas constituents, such as when being used as a nutrient source for plants in greenhouses [41] and algae in bioreactors [40]. However, to minimize the amount of waste material that has to be handled, in most configurations for Post-Combustion Carbon Capture (PCCC), a flue gas stream is divided into two gas streams. One stream has a low concentration of CO₂ and will be released into the atmosphere. The other is predominantly composed of CO₂ and will be forwarded to be used or stored. Some applications where nearly pure CO₂ is required include producing dry ice for keeping products cold during

transportation, obtaining essential oils using supercritical fluid extraction [42] [43], synthesizing organic compounds such as salicylate [44], and building epoxide copolymers [45].

For the foreseeable future, the generation of CO₂ will be substantially larger than its usage by industrial processes. Therefore, a large portion of the captured carbon will have to be placed within a stable long-term storage sink. The natural weathering of silicate [46] and carbonate [47] rocks is a carbon sink. Carbonation of silicate minerals is a thermodynamically favorable process, and carbonate minerals are stable. Accelerated weathering is a technique which aims to increase the rate of carbon storage in minerals. There are two forms of mineral carbonation, ex-situ, carbonation outside of mineral reserves, and in-situ, carbonation inside mineral reserves.

Ex-situ carbon mineralization is an energy intensive operation, requiring the transportation of reactant rocks, grinding to increase the surface area, and disposal of the resulting material. It can be performed with both minerals mined from the subsurface, such as olivine, or industrial waste products, such as mine tailings, ash, and steel slag [48] [49]. One interesting mineral that can be used for carbon storage is chrysolite asbestos. Due to its large surface area, chrysolite asbestos reacts quickly with carbon dioxide, which also neutralizes the environmentally hazardous asbestos [50]. Carbonation also decreases the pH and harmfulness of alkaline industrial waste [49]. For these limited cases, the costs of ex-situ carbon mineralization can be offset by the benefits of decreasing environmentally hazardous materials.

In-situ carbon sequestration is a less energy intensive operation, since it requires that only the CO₂ is transported, then pumped into the mineral reservoirs. Initially, the carbon dioxide is stored in volumes inside the geologic formation, adsorbed by its surfaces, and dissolved in fluids inside the formation. The rate of mineralization depends on the characteristics of the minerals in the rock and the surface area exposed to the gas [51]. Shale gas reservoirs, which have stored methane, are particularly suited for carbon dioxide storage, since they have a cap rock that limited the escape of methane gas, and a microporous structure that can hold and adsorb carbon dioxide [52]. In addition, since the kerogen and clay surfaces preferentially adsorb CO₂, methane becomes desorbed and can be recovered [53] [54]. The enhanced recovery of natural gas further decreases the net cost of storage within low producing shale gas reservoirs.

Carbon Dioxide Separation

To minimize the amount of gas that has to be transported and stored, carbon dioxide should be separated from the other gases within the flue gas stream, which will be released into the environment. All separation techniques depend on a chemical or physical property of carbon dioxide that is different from those of other major constituents of the flue gas stream.

Table 1 Major gaseous constituents of flue gas

Flue Gas Source	Major Gaseous Constituents (> 1% volume)					References
	CO ₂	CO	H ₂ O	O ₂	N ₂	

Coal Combustion	11-14		6-23	3-5	72-77	[55] [56] [57] [58] [59] [60] [61]
Natural Gas Combustion	4-10		8-20	2-12	72-85	[56] [61] [62] [63] [64]
Wood Combustion	8	1-9	5	13-20	75	[65] [66]
Iron Production	22-38	20-51	4-27	3	46-50	[67] [68]
Steel Production	10-40	1	1-13	9-13	56	[37] [68]

Different separation techniques have unique benefits and drawbacks, and the choice to use a particular gas separation method should be based on both the process that is producing the flue gas and the demand for different carbon dioxide products.

Cryogenic Gas Separation

For instance, if an end product with unmet demand is solid CO₂, cryogenic gas separation can be used. Cryogenic gas separation is a method that has been commercially utilized for over a century to extract oxygen from air for use as an industrial feedstock [69]. It operates by the principle that at a given pressure, different species undergo phase transitions at different temperatures. In this method, a gas mixture is cooled to a temperature between the phase transition temperatures of two species to be separated. This causes the target molecules to exit the gaseous state, and thus be removed from the stream.

Carbon dioxide can be separated as either a condensed liquid or a deposited solid. For conventional separation as a liquid, the pressure must be above the triple point of 5.1 atm [70]. This requires pressurizing the flue gas, an energy intensive process. By extracting CO₂ as a solid instead of a liquid, the energy requirement is reduced from 1472 kJ/kg to 810 kJ/kg [71]. In addition, since the freezing point of water is higher than the temperature at which CO₂ separation can occur, in liquid CO₂ capture, nearly all of the water must first be removed from the flue gas to prevent the buildup of ice on equipment. In contrast, in certain implementations of solid capture, water and CO₂ can be removed at the same time [72].

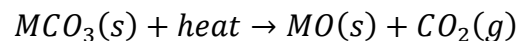
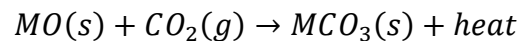
For removing CO₂ as a solid, there are three main methods. In one method, the carbon dioxide deposits on a cooling surface, which is then heated to release CO₂ gas [70]. In another method, moving packed bed CO₂ capture, there are multiple moving surfaces. The beds are first in a location where they are in contact with cooled flue gas. Carbon dioxide deposits on the surfaces. They are then moved to a location with a higher temperature, where carbon dioxide sublimates [73]. In a third method, CO₂ solidifies on the surface of a cooling liquid, forming a slurry [74]. The solid CO₂ is then separated from the liquid using a filter.

One advantage of cryogenic separation is that it yields a high CO₂ purity [71]. This means that the carbon dioxide does not have to be further purified for use in industries that require nearly pure feedstocks, which can increase the resale value of the product. Another advantage of cryogenic CO₂ separation is that liquid CO₂ has a density significantly larger than gaseous CO₂. This means that a larger quantity of carbon dioxide can be transported or stored for future use in a smaller volume. In addition, the captured CO₂ does not have to be cooled for use in applications that require liquid or solid carbon dioxide, such as electronics manufacturing [75] and food processing [76].

The main disadvantage of cryogenic separation is that it requires a large amount of energy to transfer heat from the flue gas to the environment. However, based on thermodynamics, for heat to flow from a warmer to a colder substance, no additional work needs to be performed. One potential heat sink, that can lead to energy savings, is the regassification of liquified natural gas [77]. However, the potential for cooling and work generated by the regassification process is limited by the quantity of transcontinental natural gas imports and is likely to be in demand for other cryogenic and cooling processes. Overall, it seems that cryogenic CO₂ capture is useful on a limited scale, enough to meet industrial demand of solid or liquid CO₂.

Carbonate Looping

In carbonate looping, a metal oxide is first exposed to and reacts with carbon dioxide in flue gas, to form a carbonate. The carbonate is then heated to release carbon dioxide and return to an oxide form.



One commonly used metal for these reactions is calcium. In the most commonly used setup, granular quicklime reacts with carbon dioxide within the flue gas to form calcium carbonate within the carbonator. The fluidized bed of calcium carbonate is then transferred to the calcinator, where heat is applied, and it is calcinated back into quicklime.

Heating carbonates to a temperature for significant thermal decomposition to occur requires a large amount of energy. For the system to be more energy efficient, thermal losses must be minimized. Calcium carbonate can be heated to above 900°C by placing it at the base of a small scale oxy-fuel combustion unit [78]. To minimize losses, heat from the oxy-fuel combustion can be used to generate steam and turn the turbines of an electric generator. A fraction of the heat generated by the endothermic carbonation process can be recovered, either by using a heat exchanger or mixing streams from both the carbonator and calciner, to heat up the carbonate [79].

Using calcium and magnesium looping, the energy required to capture a kilogram of carbon dioxide is 0.65 and 0.81 kWh [80]. This is significantly lower than the energy requirements of amine scrubbing. The absorbing material needs to be replenished frequently. As lime undergoes

cycles of repeated carbonation and calcification, it's reactivity with carbon dioxide decreases due to a reduction in the effective surface area [81]. Fortunately, lime is a relatively plentiful resource.

Adsorption

In order to function, adsorption-based systems require cyclical changes in the adsorption coefficient of a material. This is achieved by cyclically varying the environment of the adsorbent. For example, in vacuum swing adsorption, the feed gas is first applied to the adsorbent at a high pressure. At this high pressure, gas molecules become adsorbed to the surface. The pressure is then lowered to atmospheric pressure to desorb and remove species less strongly adsorbed to the substrate, such as nitrogen. A vacuum is then applied to desorb the CO₂ molecules [82]. In temperature swing adsorption, the adsorbent is first cooled. The low temperature adsorbent is exposed to the flue gas and adsorbs CO₂. It is heated to release the adsorbed CO₂ [83].

Different types of porous materials, including both organic and inorganic materials, can be used as adsorbents. However, each has its strengths and weaknesses that must be weighed. For example, if zeolites are used, the flue gas must be dehydrated first, since water has a strong tendency to adsorb on zeolite surfaces [84].

Adsorbents can be made of diverse types of non-toxic carbon containing waste material. Materials, such as local agricultural waste [85], wood, tires, PVC, and fabric [86], can be pyrolyzed to yield activated carbon. However, the properties of the carbon source can affect the properties of the adsorbent. Idrees et al. [87] used starch-based packaging peanuts to generate activated carbon. Their activated carbon derived from packaging had the added benefit of having high porosity, which means that there is more surface area for carbon to adsorb to.

Since sulfur dioxide is more strongly adsorbed to activated carbon than CO₂, it reduces the capacity of the adsorbent to hold CO₂ [88]. This means that for optimum performance, and to reduce the amount of sulfur in the desorbed CO₂ stream, the flue gas stream should be desulfurized prior to being in contact with the adsorbent. However, if the concentration of contaminants in the desorbed stream does not have strict limits, adsorbents can remove both SO₂ and aromatic compounds from flue gas [89]. This simplifies the design of the flue gas treatment since it only requires a single unit to capture multiple types of contaminants.

Adsorbents can be modified to increase their capacity, and the rate at which they adsorb carbon, by adding amine functional groups [90]. However, if amines are added, the disposed adsorbents risk leaching the additive into the environment.

Another factor to consider when selecting an appropriate adsorbent is selectivity. Zeolite has a higher CO₂/N₂ selectivity compared to activated carbon [91]. Activated carbon has low adsorption selectivity. When a gas separation technique has a low separation selectivity, the flue gas must go through multiple stages. However, it is resilient to exposure to many flue gas impurities, including water. Due to these properties, activated carbon adsorbents are promising as initial CO₂ and impurity removal agents in multi-stage processes.

Membranes

The main difference between membrane gas separation and other methods is that membrane gas separation does not require changes in phase. In membrane gas separation, there is a permeable membrane next to the flue gas stream. A proportion of the gas molecules cross the barrier. A selective membrane can separate different gaseous species since its permeability is different for each species. This means that gases for which the membrane is more permeable to have a higher likelihood of crossing the barrier and ending up on the other side.

Membranes can be divided into two general classes: dense or porous. Dense membranes consist of solid or suspended liquid material. There are no openings or spacings between the membrane's constituents. To cross the length of the membrane, a gas molecule must first dissolve into the material. Porous membranes have openings or spacings through which a gas molecule can cross without significant interaction with the material.

There are different mechanisms in which specific gas molecules can cross a selective barrier: solution-diffusion, molecular sieving, Knudsen diffusion, and selective surface diffusion [92]. In solution-diffusion, which occurs in dense membranes, the molecules are adsorbed at one surface, migrate through the material, and desorb at the other surface. This only allows gases that can be adsorbed to pass through. Molecular sieving occurs when a membrane has a pore size which only allows molecules with a smaller kinetic diameter to pass through.

Knudsen diffusion occurs in long narrow pores, where the molecules are more likely to collide with the boundary than with other gas molecules [93]. Heavier molecules pass through more slowly, which makes the diffusion flux inversely related to the mass of the molecule [94]. In selective surface diffusion, molecules which are adsorbed to the surface of the material pass through pores more quickly than other species [95].

A common class of material used to make membranes are polymers. Polymers can be flexible or rigid. Individual chains of flexible polymers have a high degree of freedom and can move and contort. For this reason, they are generally dense and do not have many pores. Instead, gases dissolve between the polymer chains, and transport is via the solution diffusion mechanism [96]. Rigid polymers, on the other hand, do not have many degrees of freedom. This can be achieved by having many double bonds in the polymer backbone. If the polymer includes monomers that bend the backbone, the material can be constructed to become unable to pack densely and have microscopic pores [97]. In this case, the transport mechanism is molecular sieving.

The selectivity of polymer membranes can be increased by adding amine carriers [98]. In facilitated transport by amine carriers, humidity is required. Unfortunately, humidity can cause membranes to swell and break, shortening their lifespan [99].

In some cases, a material that is selectively permeable may not have enough mechanical strength to be used as a membrane. To take advantage of the properties of different materials, membranes can be composed of multiple layers, each with a specific composition. The simplest

configuration is directly depositing polymers to a membrane surface for structural support [100]. However, in some cases, the support material may not bind directly to the selective material. To solve this problem, another configuration is having a porous base layer for structural support, a thin layer with selective porosity, and a middle gutter layer to aid the joining of the two outer layers [101].

Liquid solvents that were developed for CO₂ absorption processes can be held in membranes by capillary action. This allows solvents that have properties that make them unsuitable for conventional scrubbing processes, such as high viscosity or tendency to form solids, to be used [102].

The more permeability a membrane has, the more likely it is for a gas molecule to cross to the other side. The gaseous flux across a membrane is approximately proportional to the pressure difference across it and its permeability. This means that for more permeable membranes, lower pressures need to be applied, and less energy expended, for a given desired flux. Similarly, membranes that are more selective are more effective at lowering the amount of nontarget gases that pass into the permeate stream.

The ideal membrane would have both high permeability and high selectivity. Unfortunately, both goals are not physically realizable in existing materials. Robeson [103] noted the existence of an upper bound for a polymeric membrane's gas permeability and selectivity, which was downwards sloping. The more permeable the material was for a certain species, the less selective it could be for other species. For this reason, it is useful to use multiple stages of membranes.

The majority of membranes, including polymers, are not suitable for use in high temperature environments. For this reason, the flue gas must be cooled beforehand. In humid environments, certain types of membranes are prone to swelling [99], which can negatively impact their performance, or even cause them to break. For these types of membranes, it is also necessary to minimize moisture in the flue gas beforehand.

In addition, membranes are prone to corrosion by gases such as NO_x and SO_x [104]. For this reason, corrosive gases must be removed beforehand. One possibility is integrating an initial non-membrane removal stage with activated carbon adsorbents. The adsorbents can protect the downstream membrane from corrosion.

Amine-based Post Combustion Carbon Capture

CO₂ absorption using aqueous amine solvents, amine scrubbing, is the most mature method of removing carbon dioxide from gases. Amine scrubbing was originally developed in the 1930s to remove carbon dioxide and other impurities from natural gas [105]. In this method, shown in Figure 1, flue gas is fed in through the bottom of an absorber column. In the absorber column, a solution of water and an amine solvent is dripped from above. The solution has contact with the CO₂ rich flue gas, and carbon dioxide becomes dissolved within the liquid. However, the amine

solution is now saturated with CO₂, and must be regenerated. This regeneration process is power intensive. It involves heating the rich solution to above boiling point, then condensing it, which separates the CO₂ [106].

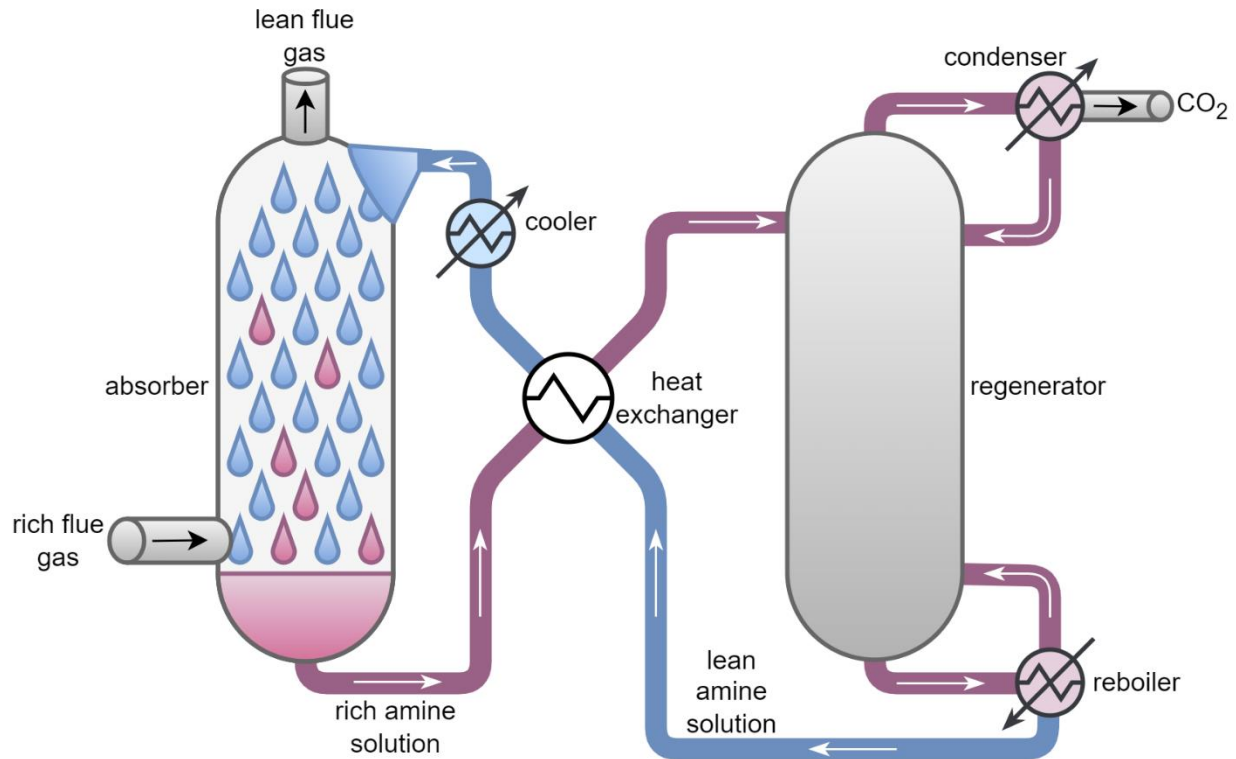


Figure 1 Post combustion carbon capture amine scrubbing process, diagram used with permission from Ghiasi et al. [107]

Using monoethanolamine (MEA), the most common solvent used for carbon capture, the energy demand is 1.05 kWh/kg CO₂ [108], the majority of which is used for amine regeneration. The energy penalty is the reduction in a currently existing power plant's efficiency due to the implementation of a carbon capture technology. Li et al. [109] estimated the energy penalty for installing MEA-based carbon capture units on currently existing coal fired power plants to range from 25% to 40%.

However, the energy cost can be reduced by using other solvents. There is active research in identifying and testing novel amines that can be applied in the scrubbing process. For this project, the identities of 151 different amine solvents that are considered for potential usage in PCCC were extracted from literature. These amines were organized into different classes based on their substructures and presence of functional groups.

An amine solvent as well as its breakdown products can escape into the atmosphere via the treated gas stream due to evaporation and mist formation [110]. The formation of suspended droplets of solution in the flue gas depend on the structure of the PCCC unit and the speed of the flue gas. A large proportion of these droplets can be removed by demisters which have a low energy requirement, including wire meshes [111] and electrostatic precipitators [112]. The

escape of solvents and their by-products as vapours depends on the vapour pressure at the temperature of the flue gas and scrubbing fluid interface. Solutions with lower vapour pressures release lower quantities of solvents into the atmosphere.

In addition, workers at plants that utilize amine based post-combustion capture will experience prolonged exposure to the vapors of the solvent. If daily exposure to these solvents causes harmful effects, firms which install post-combustion carbon capture units can expose themselves to litigation. In the event that litigation occurs, even for one specific amine solvent, carbon dioxide generating industries will become reluctant to adopt absorption-based carbon capture technologies.

With the higher temperature used in post combustion carbon capture, amine solvents react with each other and themselves at a higher rate. These products still have some carbon loading capacity; however, it can be significantly lower than the original solvent [113]. The thermal degradation product may be more corrosive [114]. For this reason, the solution must eventually be replaced. In addition, some of these products are more volatile and can possibly be more toxic than the parent compound [115]. For this reason, an in-depth toxicity analysis must not only analyze the specific chemical in question; it must also analyze the toxicity of both its decomposition products and its metabolites. This will provide insight into the appropriate measures to take to reduce the environmental impacts of the waste product.

Since it is infeasible to completely seal an industrial system, chemical production plants cannot avoid exposing workers and the environment to both precursors and by products. In a future analysis, the toxicity of chemical used within specific production methods can be analyzed for a holistic insight into the impact of the use of a specific solvent.

The toxicity of a chemical, once it has been released into the environment, or absorbed by a living organism, can be characterized by two properties, damage and dissipation.

Damage

Damage refers to the sum of the harms that are caused by exposure to a chemical species and its breakdown products. This text will focus on damage to living organisms. Damage can both be direct, due to the causticity of the substance, and indirect, due to the interaction between the chemical and molecules found in living organisms.

Many small biological signalling molecules, such as neurotransmitters and neuromodulators, have properties similar to amine solvents, including similar atomic compositions, atomic weights, and the presence of common substructures. This makes it highly likely that at least a subset of solvents under consideration interact with molecular signalling pathways. In fact, more than two solvents have been experimentally determined to interact with the cholinergic signalling system [116] [117]. These types of interactions pose a risk to a vast array of species that have exposure to the solvents.

Damage can also be caused by the degradation products of the amines, many of which are more toxic. An example of this are the nitramines and nitrosamines, which are the product of the oxidation reaction between nitrogen oxides and amine groups [118]. Both nitramines and nitrosamines are widely known to be carcinogenic, with nitrosamines generally being 15 times more potent [119]. Nitrosamines are metabolized into chemical intermediates which are strongly prone to forming covalent bonds with macromolecules including DNA [120]. This damages the DNA by forming adducts. If the damage is not detected, to either activate DNA repair mechanisms or cell apoptosis, mutations can pass on to daughter cells. If the mutations are in a sensitive part of the genome, they can lead to uncontrolled cellular replication, causing carcinogenesis.

A significant proportion of nitrosamines are formed within the absorber [119]. There are two methods to avoid the release of nitramines and nitrosamines into the environment. The first is to prevent their formation, by removing nitrogen oxides from the flue stream using different methods including selective catalytic reduction, wet scrubbing, and adsorption [121]. The second method is to remove the evaporated nitramines and nitrosamines by water scrubbing [122]. However, both of these processes add plant complexity and additional energy costs to the flue gas treatment. For this reason, these treatments should only be used if the nitramine and nitrosamine analogs of a solvent are significantly volatile. If they have low volatility, before disposal of the scrubbing solution, the nitramines and nitrosamines that are dissolved within it can be destroyed by photolysis using ultraviolet light [123] [124].

Dissipation

Dissipation refers to the rate at which the specific chemical and its damaging decomposition products are removed from a region in the environment or a specific organism. There are two types of dissipation, dispersal and destruction. Both dispersal and destruction can be viewed from both the environmental scale and the scale of an individual organism.

In the environmental scale, dispersal is the process of a pollutant becoming spread out over a larger area, while becoming less concentrated near the source. Environmental dispersion can be either beneficial or detrimental, depending on both the properties of the pollutant and the pollution situation.

For example, in situations where a high concentration of pollutants, which are toxic at low levels, are limited to a small area, such as arsenic containing mine tailings [125], dispersion is detrimental for the surrounding environment. On the other hand, in situations where the pollutant is significantly more ecologically damaging when it is concentrated, such as in oceanic oil spills [126], dispersion is beneficial. In fact, agents which aid dispersion, such as surfactants for oil spills [127], are widely used to remediate pollution.

Within the environment, dissipation also includes the natural degradation of the amines. Degradation can occur by abiotic processes and by biological organisms such as bacteria [128].

Some degradation products are beneficial, and reduce the toxicity of the substance, while other degradation products are more toxic.

In the atmosphere, water molecules absorb ultraviolet photons and gain enough energy to break an H-O bond and form a hydroxyl radical. The dominant method of the destruction of amines within the atmosphere is oxidation by reactive oxygen species. In the atmosphere, during the daytime, the half life of amines interacting with hydroxyl radicals is slightly over 4 hours [123]. These hydroxyl radicals quickly take hydrogen atoms bonded to either carbon or nitrogen atoms. Additionally, sunlight can induce the breakage of carbon bonds, leading to formation of non-amines, such as formaldehyde.

Within an individual organism, the rate of dissipation is the sum of the rate of dispersion and the rate of destruction. Dispersion within organisms is commonly known as excretion. In excretion, the substance is either pumped out of or diffuses out of the body of the individual. There are many methods of excretion, including, but not limited to, renal, biliary, pulmonary excretion, and for species living underwater, diffusion into water. Destruction of xenobiotics within organisms is a subset of metabolism. The vast majority of metabolism occurs using catalysts called enzymes. Since the exact types of enzymes present depends on the species, the metabolism is different for different species. Metabolism can make some xenobiotics less toxic and others more toxic. One common activity in metabolism is to transform lipophilic compounds, which are difficult to excrete, into hydrophilic compounds, which are relatively easy to excrete [129] [130].

When the uptake of a substance by a living organism is higher than the rate of dissipation, bioaccumulation occurs. Chemicals that bioaccumulate tend to be lipophilic, since lipophilic molecules can settle within fat containing membranes and tissues of the organism. Bioaccumulation also makes organisms in higher trophic levels have exposure to a larger quantity of the substance, in a process called biomagnification [131]. To avoid biomagnification, the use of solvents which are either strongly lipophilic, or have breakdown products that are strongly lipophilic, should be avoided.

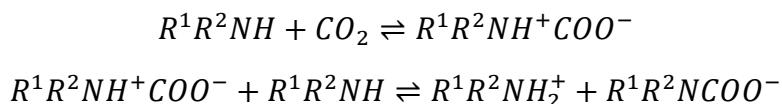
Overview of Different Classes of Solvents Considered as Candidates for Carbon Capture

Carbon dioxide is naturally found dissolved in water. However, its solubility in water is not high enough for water alone to be used as an effective scrubbing agent. For this reason, in aqueous scrubbing solutions, there needs to be a solute which can react with and convert the dissolved carbon dioxide to another form. A prototypical example of a reaction that allows an aqueous solution to hold on to an increased amount of CO₂ is the reaction of ammonia with carbon dioxide to produce ammonium carbamate.

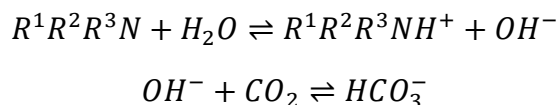


The main reason that ammonia is not widely used for carbon capture purposes is that it is highly volatile [132]. Organic derivatives of ammonia, amines, can engage in similar reactions with CO₂, while having lower volatility. Due to these properties, the majority of solutions in consideration for use within carbon capture units are aqueous solutions of amines.

Amines are classified as primary, secondary, and tertiary based on how many non-hydrogen constituents the nitrogen is attached to. The mechanisms of CO₂ absorption are similar for primary and secondary amines, which both have at least one hydrogen bonded to the nitrogen. In the first stage of the most prevalent process, a molecule of the amine reacts with CO₂ to form a zwitterion [133]. This zwitterion then reacts with another molecule of the amine to create a protonated ammonium cation, and a carbamate anion.



This process is not present for CO₂ absorption by tertiary amines. Instead, a molecule of the amine is protonated by water to form an ammonium cation and a hydroxide anion. The hydroxide anion then bonds with carbon dioxide to form bicarbonate.



In general, primary and secondary amines react more readily, however tertiary amines require a lower amount of heat for regeneration [134].

There are different classes of amines considered as candidates for carbon capture. These amines can be used as the sole solute in a solution, or they can be used in a blend [135]. In most cases, the selection of an amine solution is based on economic and performance properties, such as the heat needed to regenerate the solvent, the cost of the solvent, the degradation rate of the solvent, and the rate of CO₂ absorption. However, as post-combustion carbon capture becomes more widely adopted by power plants around the world, the process can become a

more significant source of pollution. In addition, as time goes by, effects that were not visible in a short time span, can become conspicuous. For this reason, it is important to predict and analyze the environmental toxicity profiles of the amines.

Non-Cyclic Amine Based Solvents

Non-cyclic amines are a diverse group of amines that contain no cycles that include nitrogen as a constituent. They can include other cyclical groups, such as phenyl and cycloalkyl groups.

Alkanolamine Based Solvents

Alkanolamines are a class of substances that contain attached amine groups and alcohol groups.

Monoethanolamine is the most commonly used solvent in post combustion carbon capture pilot tests, since it has a low-cost, high-water solubility, acceptable performance in removing carbon dioxide, and is resistant to thermal degradation [136]. One downside of using monoethanolamine is that it is corrosive to steel [137] [138]. Other ethanolamine derivatives used in carbon capture include diethanolamine and triethanolamine. Ethanolamine derivatives have displayed acute toxicity to marine species including diatoms and crustaceans, however this toxicity occurs at high concentrations, with LC50 values greater than 99 mg/L [139].

Alkanolamines other than ethanolamine derivatives have also been investigated. These include, but are not limited to, 2-amino-2-methyl-1-propanol [140] and diisopropanolamine.

Diethylamino-ethanol, a tertiary amine, is also being investigated, due to its large carbon dioxide capacity and its low regeneration energy [141]. However, since it has a slow reaction rate, it must be mixed with a faster amine [142]. Using alkanolamines with longer alkyl chain lengths have been proposed as a solution to reduce ammonia emissions [143].

Amino Acid Based Solvents

Amino acids tend to be resistant to oxidative degradation and to have a lower volatility than similar alkanolamines [144]. This means that amine losses can be lower when amino acid solutions are used, and the scrubbing fluid has to be replaced less frequently. In addition, due to the degradation resistance, using amino acid scrubbing solutions will release less ammonia into the atmosphere compared to using monoethanolamine.

Glycine is an amino acid analog of the commonly used monoethanolamine. In an experiment by Guo et al. [145] glycine solutions were twice as fast at absorbing CO₂ than alanine solutions of the same concentration and had a slightly larger rate constant than monoethanolamine.

However, the energy of regeneration was higher for glycine compared to monoethanolamine. A possible remedy to this problem is to add tertiary amine derivatives of glycine which have been used in gas scrubbing applications, such as dimethylglycine and diethylglycine [146], since, in general, tertiary amines require a lower amount of energy for regeneration.

Amino acids can also be used in solutions with other amines. Different amino acids were combined with an imidazolium derivative to create ionic liquids [147]. Amino acids can be mixed with other amino acids, to form salts where one species is an anion and the other is a cation

[148]. In addition, due to their faster reaction kinetics, amino acids can be used as promoters alongside slow but high absorption capacity solvents, to form scrubbing solutions that are both fast and have a high CO₂ capacity [149].

Certain amino acid salts, such as sodium taurate, form precipitates when the solution becomes more acidic due to absorbing carbon dioxide [150]. The solid salt is then collected by a mechanical separator, and then heated to release the absorbed carbon dioxide [151]. This saves energy, since only the precipitated solvent and its carbamate needed to be heated for regeneration.

Amino acids can also be used as anions for ionic liquids [152]. Ionic liquids are explained later.

In this study, for the purposes of comparing properties between different groups of amines, amino acids that contain cyclic amine groups, such as proline which has a pyrrolidine group, will be placed in one of the cyclic amine groups.

Other Non-Cyclic Amines

Alkyl polyamines, which contain multiple amine groups mounted on an alkane backbone, are being investigated due to their ability to hold onto a large amount of carbon dioxide. For a given concentration of a polyamine solution, the carbon dioxide capacity is roughly proportional to the number of nitrogen atoms within its molecules [153]. The absorption rate has a nonlinear relationship with the number of nitrogen atoms in a polyamine. It increases, until it reaches a molecule with 16 amine groups [153]. They have good performance as activators for tertiary ethanolamine solutions, with lower heat duty than the conventional activator piperazine [154]. However, one major problem with the polyamines is that they significantly increase the viscosity of a solution [141].

Etheramines are similar to both alkanolamines and polyamines, except that they contain at least one oxygen bridge connecting an alkanamine or alkanolamine section with an alkanamine, alkane, alkanol, or alkanolamine section. The etheramine, diglycolamine, has been used in gas sweetening applications [155]. Linear etheramines and alkanolamines can form complexes around an alkali metal, creating ionic liquids [156]. These complexes are more stable at higher temperatures than the raw alkanolamine or etheramine.

Another subclass of non-cyclic amines which is recently experiencing increased attention are the lipophilic thermomorphic biphasic amines. Lipophilic thermomorphic biphasic amines, such as cyclohexylamine derivatives, are dissolved in the aqueous phase at low temperatures such as 40°C but separate at higher temperatures such as 80°C [157] [158]. This allows the solvent to be mechanically separated from water using only a small amount of heat. After mechanical separation, only the solvent has to be heated for regeneration. When used at the same conditions as monoethanolamine, the oxidative degradation of cyclohexylamine derivatives are 1.2 to 3 times faster and the thermal degradation is twice as fast, however, since it can be used at lower temperatures, the thermal degradation is expected to be lower [159].

Structures with one central tertiary amine group bonded to three alkyl or alkylamine substituents, are being investigated to be used as cations in ionic liquids [160].

Cyclic Amine Based Solvents

A cyclic amine group is a cycle that contains at least one nitrogen with a lone pair. Molecules that contain cyclic amine groups can be either secondary or tertiary amines.

Piperazine and Derivatives

Both piperazine [161] and its derivatives [162] have been proposed as alternatives to monoethanolamine, since systems that use them have a better energy efficiency. One drawback of piperazine is its limited solubility in water [163]. However, it can be used in a blend, as an additive for other amine solutions, to speed up their rate of reaction [141] [164] [135]. This activator property is also present in piperazine derivatives such as 2-methylpiperazine [165]. However, blends of different amines, such as blends of piperazine and monoethanolamine, degrade at a much greater rate than solutions with one single type of amine [115].

One major drawback of piperazine is that its vapor concentration in the treated flue gas can exceed 1000 ppm [166]. This means that plants that use piperazine must include a treatment phase to remove it from the flue gas before it is released to the atmosphere.

Piperazine has been and is still used as a treatment for intestinal helminthiasis, and anthelmintic. It acts as a GABA agonist and gives nematodes muscle paralysis [167]. It produces stronger effects on nematodes that reside the large intestine than the ones that live in environments with higher O₂ levels. This is likely due to the fact that the non-carbamated form of piperazine is a weak GABA antagonist, while piperazine carbamate is a GABA agonist [168]. Even though specific groups of nematodes are beneficial for plant life [169], the interactions of piperazine and piperazine derivatives with GABA receptors are unlikely to be a significant source of their toxicity, since they are weaker than the native agonist [170].

Piperidine and Derivatives

One class of amine solvents that is proposed to help lower the energy needed for the absorption process are the piperidine derivatives. This is because they display demixing activity [171]. Demixing solvents react with carbon dioxide to form two new species. The unreacted solvent is hydrophobic, while the reacted species are hydrophilic. This property allows the unreacted and reacted phases to be mechanically separated. Since only the hydrophilic phase needs to pass through the regenerator, the energy costs of the regeneration process are reduced. Raynal et. Al. [172] used proprietary solvents in the DMX process and showed that the process requires only 0.58 kWh/kg CO₂. This is double the efficiency of standard MEA capture.

However, one large environmental and safety issue is that piperidine itself, and many of its derivatives are proven to be harmful to animal life. Piperidines are cholinergic substances; they bind as agonists to nicotinic acetylcholine receptors [173]. Nicotinic acetylcholine receptors (nAChRs) are excitatory receptors widespread within the animal kingdom, found in both the

central and peripheral nervous system [174]. This is why the most commonly used classes of pesticides, organophosphates, carbamates and neonicotinoids, are indirect agonists and direct agonists of nAChRs, respectively. Both naturally occurring [175] and synthetic [176] piperidine derivatives have displayed insecticidal properties. Numerous plants that contain naturally occurring derivatives of piperidine have shown teratogenicity towards livestock [177].

Piperidine displays both psychoactive and neuroactive properties. The effects depend on the amount of exposure, and can range from tremors at lower doses, to aggressive behavior, extrapyramidal signs and seizures at higher doses [178]. This can become very problematic for people that handle the substance, or inhale fumes from it. Piperidine derivatives that also act on nAChRs likely also have these properties. For this reason, it is important to screen all potential solvents, and to exclude solvents that have medium to high affinity to the nAChRs.

Pyrrrolidine and Derivatives

Pyrrrolidine itself is a candidate for carbon capture applications because it has a high reaction rate, it's solutions can hold onto a large concentration of carbon dioxide, and its regeneration energy is lower since it forms bicarbonate ions instead of carbamates [179]. In fact, pyrrolidine derivatives are the fastest solvents ever investigated for carbon capture. Based on simulations by Orllov et al. [164] the potential amine solvents with the highest reaction rates were both pyrrolidine derivatives, specifically 1-methyl and 1-ethyl-3-pyrrolidinol. The carbon capture pilot plant at the Irving Oil Whitegate Refinery in Ireland uses an aqueous solvent mixture which includes 1-(2-hydroxyethyl) pyrrolidine, a pyrrolidine derivative [180].

Pyrrrolidine is a direct nAChR agonist, however, it has half of the affinity of piperidine [173]. It also displays central nervous system activity similar to piperidine, however, it is less potent [117]. The use of pyrrolidines have the same risks of ecological harm as the use of piperidines. Therefore, they should also be thoroughly screened for cholinergic activity.

Pyrrrolizidines, a specific subset of pyrrolidine derivatives, have been proposed as solvents for carbon capture. An interesting feature is that a subset of the pyrrolizidines investigated by Hanusch et al. [181] became solid when they absorbed carbon dioxide, which they then released upon the application of heat. This means that the saturated solvent can be mechanically separated from both the unsaturated solvent and the rest of the aqueous solution. Since the solid is the only component that needs to be heated up, use of pyrrolizidines can potentially yield energy savings.

One major issue with using pyrrolizidines is that they covalently bind to DNA, becoming DNA adducts [182]. If the damage is detected, this can either activate DNA repair mechanisms, or trigger a cascade of molecular signalling that leads to preprogrammed cell death [183]. However, if the damage is not detected, and the cell undergoes mitosis, the chromosomes of the daughter cells may be mutated [184]. Since the mutations may lead to the development of neoplasms, these genotoxic substances are also carcinogenic. Therefore, to minimize the likelihood of workers that handle pyrrolizidines of developing carcinogenesis, only pyrrolizidines

that both have a low vapor pressure, and a low octanol-water partition coefficient, should be used.

Azoles

Azoles are a diverse class of cyclic amines, which have a 5 atom cycle that both contains at least one nitrogen atom and have a maximum of three carbon atoms. One advantage of azoles is that they can form ionic liquids.

Ionic liquids are pure salts that are in liquid form. In the context of carbon capture, the term ionic liquid specifically refers to salts which have a melting point below the normal temperatures of the flue gas and can remain liquid without an additional application of heat. One benefit of using ionic liquids is that they can absorb a large concentration of carbon dioxide [185]. Another benefit is that they have a low vapour pressure [186] [187]. This means that a lower amount is able to escape into the environment from the flue gas stream.

Salts of imidazole derivatives have been used to create ionic liquids [185]. However, one major problem with ionic liquids is that they are viscous. Wang et al. [147] overcame that barrier by binding 1-ethyl-3-methylimidazolium to the surfaces of porous poly (methyl methacrylate) spheres. Interestingly, in their experiments the capacity and rate of CO₂ uptake was strongly dependent on the amino acid anion that was used, with lysine providing the best performance. Another way of lowering the viscosity of imidazolium based ionic liquids is to include ether groups in the tail of the cation [188].

Alkyl-imidazoliums are readily and strongly adsorbed onto diverse types of soils, including marine sediment [189]. This property has both advantages and disadvantages. This means that if an area becomes polluted, it becomes impracticable to remediate the soil around that area. Since alkyl-imidazoliums exhibit bactericidal activity [190] and inhibit plant growth [191], this can negatively impact the agricultural productivity of the soil, sharply decreasing land values. This problem is exacerbated by the fact that imidazolium derivatives are relatively resistant to oxidative degradation in the environment. To minimize environmental threats, the substance should be broken down before disposal. The most effective way of doing this is by exposing the ionic liquids to both peroxide and ultraviolet radiation [192].

Imidazoliums also inhibit acetylcholinesterase. However, the degree of this inhibition strongly depends on the specific chemical species in question: while some species can be almost as powerful as commercial pesticides, such as Aldicarb, others require a concentration that is 100 times larger to provide the same degree of inhibition [193]. Overall, species that are more lipophilic pose a greater threat to invertebrates.

Another subclass of azoles which have been computationally investigated to be able to absorb carbon and sulfur compounds and to form ionic liquids are the triazoles [194]. Outside of ionic liquids, solutions of tertiary imidazole derivatives have been found to require less energy for every quantity of carbon dioxide captured, compared to the benchmark monoethanolamine [195].

Morpholines and Derivatives

Oxazinanes are cyclic amines which have a 6 atom ring that contains both a nitrogen and an oxygen. 1,4 Oxazinane is commonly known as morpholine. In gas sweetening solutions that used diglycolamine, morpholine was found as a dehydration product that also aids in acid gas removal [155]. This amine is being investigated for use in carbon capture [196].

Morpholine has fast kinetics, with an absorption rate three times as fast as monoethanolamine [197]. This means that it can be used as an activator in solutions. It has a much higher resistance to thermal degradation compared to conventionally used amine solvents and can be used at higher temperature [137].

Tertiary derivatives of morpholine are also being investigated, due to their lower energy of regeneration needed [195]. Within the field of natural gas sweetening, a solution of morpholine derivatives, including N-formomorpholine, have been used in the Morphysorb solution, which removes both carbon dioxide and hydrogen sulfide [198].

Pyridines and Derivatives

Pyridine is a 6-atom aromatic ring, similar to benzene, except that one of the carbon atoms has been substituted with nitrogen. Pyridine derivatives have been used to form both anions and cations for ionic liquids.

Pyridinolates have been used as anions for ionic liquids. These ions can capture 1.6 moles of CO₂ per mole of pyridinolate, since they have two sites where an RCOO¹ group can attach [199]. Pyridinolates are also good at absorbing SO₂ [200], which can make them useful for locations where a significant amount of sulfur dioxide is present in the flue gas, such as coal combustion plants.

N-alkyl pyridiniums and their derivatives have been used to form cations for ionic liquids [187][201]. Similar to alkyl-imidazoliums, alkyl-pyridiniums are also strongly adsorbed by sediments. Another similarity with alkyl-imidazoliums is that pyridiniums are acutely toxic to bacteria, with their toxicity increasing with the length of the alkyl chain [191]. Interestingly, this trend also extends to toxicity to algae. This is likely due to the fact that longer chains make the solvent more lipophilic, and more capable of crossing cell membranes. Decreasing the lipophilicity by adding polar groups to an alkyl chain decreases the cytotoxicity [202]. Certain pyridinium derivatives, including as 1-butyl-2methylpyridinium, have displayed cholinergic activity by inhibiting the enzyme acetylcholinesterase [193].

Criteria for Comparing Toxicity

For this study, the toxicity of amine solvents will be compared using four properties, for which experimental data is available to train machine learning models. Dissipation is indicated by volatility and lipophilicity. Similarly, damage is indicated by mutagenicity and neuroactivity.

Volatility

Volatility is the tendency of a species within a liquid phase or solution to escape into the gas phase. A higher volatility increases the amine losses, which increases costs of operating the system, and it causes more of the chemical to leak into the environment. There are methods of reducing the amount of a specific solvent that escapes into the atmosphere. For example, water washing has been found to effectively reduce the monoethanolamine losses due to evaporation from 0.1-0.8 kg/tonne treated gas to 0.01-0.03 kg/tonne treated gas [203]. However, these methods are associated with their own energy costs. To avoid these costs, it is better to choose a solvent with lower volatility.

One way to quantify volatility is to use vapour pressure. However, vapour pressure depends on the temperature. In addition, a particular solution's vapour pressure at a specific temperature is a function of both the pure solute's vapour pressure, the mole fraction of the solute, and the CO₂ loading [110]. Since the vapor pressure is proportional to the solvent concentration, assuming that two solvents have the same toxicity profile, the solvent that has a lower vapour pressure in its pure form, can have a higher concentration before it begins to emit the same quantity of amines.

The vapor pressures of 28 different small molecules composed of solely hydrogen, carbon, oxygen, and nitrogen were obtained from experimental results by Lepori et al. [204]. Amines, alcohols, and ethers within the dataset had molecular weights from 43 to 116 daltons, and partial pressures ranging from 6.5 Pa to 36 kPa. 348 measurements were given for different temperatures from 0°C to 48°C. Additionally, 88 points of vapour pressure data for 10 ketones, aldehydes, esters and an alcohol, from temperatures of -40°C to 25°C were collected from the experimental measurements of Covarrubias-Carvantes, et al. [205], with molecular weights ranging from 58 to 144 Da. From the CRC handbook [206], 2872 vapor pressures of 1056 different small organic compounds, from 1 Pa to 100 kPa, were extracted.

In order to have amines better represented in the training dataset, experimentally measured equilibrium vapor pressures for 4 alkyl amines [207], 4 alkanolamines [208], 8 alkanolamines [209], 3 ethanolamine derivatives [210], 4 propanediamines [211], 4 phenylamines [212], were obtained from independent sources.

Vapor pressures at small intervals of different temperatures for 32 very small compounds were obtained from the NIST website. This will aid the algorithm in learning the relationship between temperature and pressure. The vapor pressure dataset was augmented by the boiling temperatures of 4552 organic compounds under 300 Da, 3387 of which were measured at the

standard 760 mmHg atmospheric pressure, and 1829 of which were measured at reduced pressures [206].

Lipophilicity

Lipophilicity is a measure of how much more readily a substance dissolves within lipids instead of within water. Chemicals which are lipophilic have a strong tendency to be sequestered within fatty tissues. This causes lipophilic toxins to accumulate within the lipids of an organism, which causes the concentration of a lipophilic substance to be higher inside of an organism than within the environment [213]. In addition, compounds that are stored within fatty tissues are continuously released into the rest of the body, the rate of release depends on the amount of metabolic stress [214].

The continuous release can expose the organism to chronic toxicity. The presence of chronic toxicity is especially dangerous since it takes a long time to show its effects. These effects are likely to go undetected in time-limited trials. This can make it appear like a higher amount of exposure is harmless to workers and the environment. In addition, after a lipophilic substance has been proven to have toxic effects, it takes a long time to exit both individual organisms and the ecosystem.

Their concentration also increases with each trophic level, putting lifeforms high on the food chain, such as carnivores, at greater risk than lifeforms lower on the food chain, such as plants and herbivores. Overall, lipophilic compounds have the potential to cause toxic effects which are not readily visible and can last for a long period of time.

A common method of quantifying lipophilicity is to use the partition coefficient (P) between a lipid phase and a water phase. Different substances that are primarily nonpolar in nature have been used to quantify the lipophilicity, however, the most commonly used is 1-octanol instead of nonpolar hydrocarbons, since it has a small polar OH group at the end and is more representative of biological membranes. The octanol-water partition coefficient is the ratio of the equilibrium concentration of a substance in two adjacent liquids, water and 1-octanol. It is commonly given in logarithmic form, logP.

The experimentally measured logP values for 80 organic compounds, with weights ranging from 160 to 300 daltons, and logP values ranging from 0.3 to 6.6, were collected from Martel et al. [215]. Since in their dataset, logP was correlated with the molecular weight, only molecules with weights below 300 Da were included. Sangster [216] collected and compiled experimental logP values for simple organic compounds from different sources. Similar to the Martel dataset, logP had a positive correlation with molecular weight. However, all of these compounds had weights below 300 Da. The average experimental logP value for 433 of these compounds were included in the train-test dataset.

For additional train-test data, the Physprop dataset[217], which had 18076 valid entries with Log P information, was also used. Excluding the molecules with elements other than C, H, O, N,

and S, or weights larger than 300 Da, the number of points in the Physprop dataset decreased to 9318. The three datasets combined has 9701 training points, with weights ranging from 16 to 299 Da, and a wide range of logP values from -5.08 to 9.95.

Mutagenicity

Carcinogenicity is not a straightforward property to quantify. There are many mechanisms that can induce carcinogenesis, including, but not limited to, the induction of mutations, the inhibition of DNA repair mechanisms, and the suppression of immune responses. Since the generation of the disease is usually multifactorial, and even individuals within a species can differ in their response to each factor, there is no universal measure of carcinogenicity.

One commonly used proxy for carcinogenicity is mutagenicity. A common experimental method to see whether a substance is mutagenic is the Ames test (Figure 2). In the Ames test, there are 5 mutant strains of salmonella, which are incapable of producing histidine because of different mutations. However, each strain can become capable of histidine production if there is a mutation. Each strain requires a different mutation, some require single nucleotide point mutations while others require frameshift mutations. The bacteria are first put in a histidine rich medium that contains the potential carcinogen to be tested. After a period of time, they are then moved into a histidine poor environment. If they thrive in the second environment, it means that a mutation occurred, and the substance is highly likely to be mutagenic.

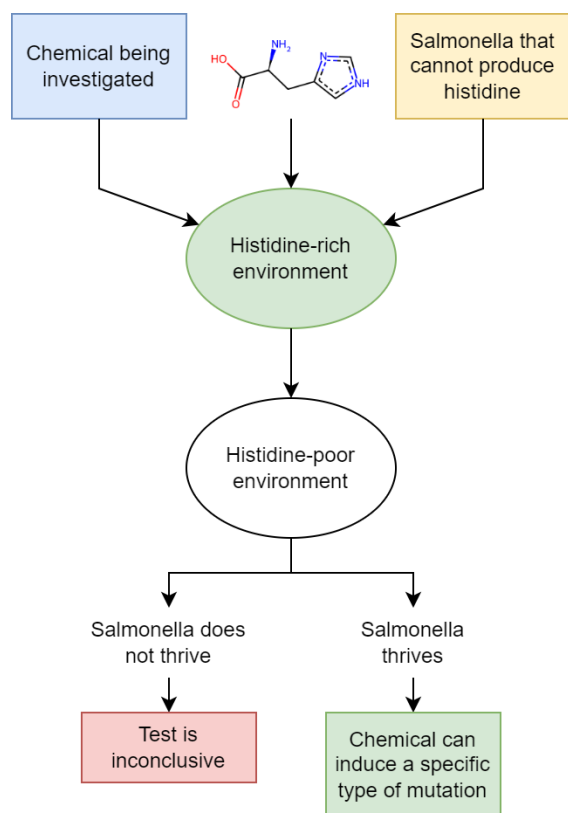


Figure 2 Overview of the Ames Mutagenicity Test

Martinez et al. [218] compiled a dataset of 5536 molecular compounds and their tested ability to induce mutations within 5 different salmonella strains used in the Ames test. A compound was assigned a label of “Mutagenic” if it was proven that it could induce mutations in at least one of the strains. It was assigned “Non-mutagenic” if there was no evidence that it could induce mutations in any of the five tested strains. Otherwise, the chemical’s mutagenicity was labelled as “Inconclusive”. The majority of the compounds within their dataset included metals and had to be eliminated. Of the dataset, only 380 compounds did not include metals or halogens. The majority of these substances had weights within the range of 150 to 300 Da. Within the limit of below 300 Da, there were 276 substances.

The results of the Ames test only provides categorical data, of whether or not a compound is able to induce a specific type of mutation that can allow a strain of salmonella to become capable of producing histidine. It doesn’t provide information about the relative strength of each carcinogen.

A weakness of the Ames test is that it is performed on bacteria. The metabolism of bacteria is vastly different from the metabolism of eukaryotes, such as plants and animals. Since certain types of carcinogens, such as nitrosamines, require metabolism before becoming active, the Ames test can fail to recognize substances that are carcinogenic for non-bacterial lifeforms.

The oxidation products, nitrosamines and nitramines, are widely agreed upon to be carcinogenic. Mutagenic nitrosamines and nitramines that are formed in the scrubbing solution can be destroyed using ultraviolet radiation [123]. However, some escape the solution as vapors, and cannot be destroyed. For the purposes of this comparison, each nitrosamine and each nitramine will be assumed to have the same mutagenic potency. The volatility, the ability for a specific solvent’s nitrosation product to escape into the environment, will be compared, as a proxy of how capable it is to cause nitrosamine and nitramine induced genotoxicity.

Neuroactivity

There are many biological molecules and proteins involved in neurological activity that the solvents can potentially interact with. However, it is not feasible to analyze every single neurological signaling pathway, even within a single species, both since there is a complex web of molecular signalling that has not been uncovered yet, and because many components have not been studied extensively. In addition, certain components of nervous system are present in some species, and absent or extensively modified in other species. For this reason, it is best to focus on a section of the neurological signalling web that is well-studied, applicable to as many species as possible, and that have literature suggesting interaction with certain subsets of solvents.

Nicotinic cholinergic receptors (nAChR) are an ancient class of ligand gated ion channel, conserved in distantly related animals including both insects and vertebrates. Due to its ubiquity, plants such as nightshades, have evolved to produce toxins that interfere with this ion channel. The cholinergic signalling system has been studied extensively for over a century by

both commercial and academic research teams [219]. Chemicals that interfere with nAChR are effective at disturbing ecosystems; the two most common classes of pesticides, acetylcholinesterase inhibitors and neonicotinoids are indirect and direct agonists of nAChR, respectively.

Information about chemical affinities to acetylcholinesterase and nAChR were obtained from BindingDB [220]. In all cases, the affinity is given in terms of inhibition constant, K_i . The smaller K_i is, the stronger the affinity. K_i does not give any indication of whether a ligand is an agonist, partial agonist, inverse agonist, or antagonist. However, since both strong agonists and strong antagonists are harmful, the classification of a ligand has little importance when discussing its potential for harm.

The experimental acetylcholinesterase affinity was found for 414 chemicals under 300 daltons and 1070 chemicals under 500 daltons. These chemicals ranged in weight from 86 to 499 daltons, however the majority clumped within the 200 to 400 Da range. The experimental animals were vertebrates, 57% of the data was based on human receptors, while the second common animal was the Electrophorus at 31%. Affinities ranged from 0.04 nM to 109 mM. There was insignificant correlation between the weight and affinity. Within the database, a large number of low affinity chemicals were assigned K_i values of 100 μ M [220]. Excluding those points, the log K_i formed a bell curve, ranging from -1 to 8, with a peak at 3. Medium affinity chemicals were common in the training data because researchers are incentivized to perform affinity measurements for species that they suspect are ligands.

Data for nAChR was found in different datasets, separated by the subtypes. Since they had the most entries, subtypes $\alpha 4\beta 7$ and $\alpha 7$ were used. Subtypes cannot be combined since they have different sensitivities to endogenous neurotransmitters.

The dataset for $\alpha 4\beta 7$ was based on human receptors, and included 1074 ligands under 300 Da and 1393 ligands under 500 Da. The weights mostly clumped in the 150 to 300 Da range. Affinities ranged from 0.015 nM to 1 mM. There was a small positive correlation between affinity and molecular weight.

The dataset for $\alpha 7$ was based on both rodent (70 %) and human (30%) receptors. It included 958 compounds below 300 Da, and 1741 chemicals below 500 Da. The weights were mostly concentrated in the 250 to 350 Da range, and the affinities ranged from 0.023 nM to 1 mM. There was a small negative correlation between weight and affinity.

In all cases, the larger values for K_i were less precise than the smaller values of K_i . In addition, they were more spaced out. Fortunately, the order of magnitude is more important than the precise values. As part of data pre-processing, before the values are put into a learning algorithm, the values of K_i will be transformed into log K_i .

Review of Previous Models used to Predict Chemical Properties

There are two types of models: mechanistic and empirical. Mechanistic models are built using theoretical relationships between different abstracted components of the systems. They require knowledge about the fundamentals of the system. An example is a model that uses quantum mechanics to predict thermodynamic properties of a fluid.

However, for many systems, prior knowledge about the fundamental components is not available, or mechanistic models contain too much complexity to be practical. For these systems, mechanistic models should be used. An empirical model is developed using experimental measurements. A large subset of empirical modelling is regression. Predicting properties of chemicals using machine learning is regression.

Supervised machine learning, the use of algorithms to generate an empirical model, can be used for both regression and classification. One of the major downsides to using machine learning is that it requires a large amount of training data in order to provide meaningful results [221]. However, once they have been optimized and trained, ML models require less computational resources to run compared to complex mechanistic models.

The rest of this section gives an overview of previous mechanistic and machine learning models that have been previously used to predict chemical properties.

Mechanistic Computational

Mechanistic computational models have been developed to simulate the interactions of molecules with each other. One of these models is the Conductor-like Screening Model for Real Solvents (COSMO-RS), which considers a molecule as a surface which can have interactions with neighbouring solvents. COSMO-RS is included as part of the Amsterdam Modelling Suite [222]. In the Amsterdam Modelling Suite, molecules are modelled by density functional theorem (DFT), with the choice to use different approximation methods and to model select properties of the electron orbitals. In DFT, electron position likelihoods are represented in 3 spatial dimensions for every atom in the molecule [223]. An optional step in the modelling process is to optimize the bond lengths and angles to lower the potential energy of a molecule.

Before using COSMO-RS, the DFT models are converted into equivalent surfaces of charges. COSMO-RS can simulate Van der Waals forces, electrostatic forces, and hydrogen bonding between a molecule and nearby molecules in the solvent [224]. It provides the energy difference between a molecule in the gas phase, in the liquid phase, and dissolved in a solution with adjustable mole fractions.

COSMO-RS has been used by Lin et al. [225] to predict vapour pressures and heats of evaporation for 371 organic substances. To find the vapor pressure at equilibrium, they equated the potential energy per mole of the gaseous and liquid phases, and rearranged to obtain a

function that depends on the solvation free energy and the molar volume. The COSMO model was used to determine the spacing between different molecules in the liquid phase, and the differences in energy between both phases due to electrostatic interactions. A limitation of their studies is that it is only applicable for pure substances in vacuum.

COSMO-RS has also been used to predict the vapor pressures of alcohols, carboxylic acids, aldehydes, and ketones at temperatures between 290 to 320 K [226]. The predictions were very close to experimental measurements for aldehydes and ketones. However, for alcohols and carboxylic acids, it tended to overestimate the vapor pressure. The vapor pressure prediction decreases with increased potential energy in gaseous phase and increases with increased potential energy in liquid phase. This suggests that the model's predictions for the potential energy of liquid alcohols and carboxylic acids may have had a positive error. This may occur if the model is underestimating the decrease in potential energy due to hydrogen bonding.

Warnau et al. [227] used COSMO-RS to determine the potential energies of a single molecule of a species being investigated within a solution of water, and within a solution of wet 1-octanol with a 0.27 mole fraction water. The potential energies were used in a thermodynamic equation to provide logP. The assumption of only a single molecule being dissolved in the medium provides a source of error, since in a real solution, there are different molecules of the solute interacting with each other, and the mol fraction of the solute affects the potential energy. Overall, for the majority of the 22 small drug-like molecules that they investigated, the predicted logP was very close to the experimental logP. The only exceptions were the isoxazoles.

Empirical Machine Learning

Machine learning regression and classification models are fundamentally mathematical equations, where the coefficients are learned from an algorithm. Machine learning models can be as simple as a linear or polynomial equation, and still be useful.

Olsen and Nielsen developed a model to predict vapor pressures at 20°C for organic compounds based on their modified UNIFAC structures. UNIFAC structures are a string of count numbers for preselected molecular substructures [228], and for their model, nine substructures were used. Their data consisted of the vapour pressures of 294 organic compounds at ambient temperatures and was split into a train and test set. Six different polynomial equations were built to account for both general and special cases, such as when the chemical has a polar group. Coefficients of these equations were obtained from the training data using partial least squares. For relevant compounds, the corresponding special equations had lower prediction errors than the general-use equation.

Gharagheizi et al. [229] trained a neural network to predict the vapor pressures of different chemicals using the temperature and the absence or presence of 202 groups. They found that the ideal structure for their task had a single hidden layer with 10 neurons. The training data consisted of 42 000 vapour pressure and temperature pairs, belonging to 1405 different chemicals. This is almost 30 pressure-temperature pairs for each compound. The data points

have been randomly divided into train, test, and validation sets. This part of their methodology is problematic, since there is a high chance that every single species is represented within the training data. It is no surprise that the predicted and database vapour pressure were strongly correlated with each other and appeared like a line when plotted. This does not mean that the model will have good performance for chemicals outside of the 1405 species. For making the model more capable of generalizing, the researchers should have split the train, test, and validation sets in a way so that all points corresponding to a single chemical appear on one side of the split.

Prasad and Brooks [230] used neural networks to predict logP values. 12 000 compounds were used for training, and 2000 were used for testing. The compounds were translated into 1024 bit ECFP fingerprints with a radius of 4. Architectures with the highest performance, one with 5 hidden layers and 2048 hidden neurons and the other with 3 hidden layers and 896 hidden neurons, were found using 5-fold cross validation. They had root mean square errors of 0.62 and 0.82 for logP, respectively, which translates to differences in partition coefficients of 4.1 times and 6.6 times.

Isert et al. [231] produced machine learning models that are more precise, with logarithmic root mean square errors of 0.44 and 0.34, for models trained on publicly available compounds and compounds available for Novartis researchers. This discrepancy can be attributed to the fact that the public dataset contained 2114 species while the Novartis dataset contained 25642 species. However, the Novartis dataset contained both logP values which were experimentally determined, and logP values which were obtained using the mechanistic model COSMO-RS. The authors acknowledged that even though the quality of the ML model is limited by the quality of the mechanistic model, it provides values using less computational time. In both cases, the ML model with the best performance was Chemprop.

Machine learning has also been used for chemical classification. Martinez et al [232] trained deep neural networks (DNN) to predict the mutagenicity, as indicated by the Ames test results, of organic compounds. The neural networks were trained and tested on 5536 chemicals. Inputs included chemical descriptors calculated using Mordred [233], and the outputs were whether a chemical could induce one of the five strains to begin synthesizing histidine and the overall mutagenicity. Three different architectures were used, with the ideal values selected using grid search.

The simplest architecture was a single neural network to predict the overall mutagenicity of a compound. However, the drawbacks of this architecture was that it was biased to give false positives due to 93% of the compounds being mutagenic and it provided no insight into the mechanisms of mutagenicity. Another architecture was to build five separate neural networks to predict the results of each strain. The third architecture was a multitask learning DNN, composed of a common layer whose outputs were fed as inputs into task specific cores. The performance of the multitask learning DNN was better than the performance of the single task

DNNs [232]. This suggests that there are common factors that cause different types of mutagenicity.

Another technique to predict classes is to perform clustering, which was used by Benigni to predict the Ames mutagenicity of compounds [234]. The QSAR toolbox [235] was used to cluster the molecules into groups based on the following criterion: the similarity of the computed lipophilicity ($\Delta \log P < 1$), molecular weight within 20% of each other, and a structural similarity index greater than 70%. The majority of clusters contained training data with all negative or all positive Ames mutagenicity. For test predictions, the accuracy was 85%. One drawback of this method was that, for molecules that cannot be placed in a cluster, no predictions could be made.

Sharma et al. [236] trained deep neural networks, both single task and multi-task, to both in vitro and in vivo toxicity results. The in vitro part included predicting the ability of molecules to interact with 7 different nuclear receptors and to activate 5 different stress responses. The in vivo part of the study was based on the acute lethal dose for 50% of mice. Their training and test set included 10 000 molecules, obtained from 3 sources, the Tox21 database [237], ClinTox [238], and the Registry of Toxic Effects of Chemical Substances [239]. The neural networks were trained on 4096 bit Morgan fingerprints with radii of 2.

Their neural network consisted of two parts, which were trained together [236]. The first part was a bidirectional gated recurrent unit (biGRU). The output of the biGRU was a molecular representation that was 128 units long. This representation, along with properties predicted by other neural networks, such as logP and polar surface area, were fed into single task and multi-task deep neural networks. Genetic algorithms were used to choose the ideal hyperparameters for each neural network. Interestingly, their multi-task neural networks had better performance than the single-task neural networks that they trained. This suggests that different toxicity measures had common mechanisms.

Convolutional Neural Networks

2 dimensional chemical graphs have been used as inputs to CNNs to predict molecular properties. Goh et al. [240] developed Chemception, which requires no former knowledge of chemistry to predict chemical properties, such as toxicity classification and solvation energy. It's inputs are 40 angstrom squares with 0.5 angstrom grids that have a single channel. The value of the grayscale input pixel is the mass of the particular atom that exists in the position. Pixels that did not contain any atoms were given a value of 2 if they contained a bond or a value of 0 if they were empty. Using 5 fold cross-validation, they optimized the architecture of their models. They found that changing the width of the kernel did not change the performance but decreasing the number of filters in each layer to 16 provided significant benefit. The performance improved with increasing number of layers, up to a certain point, 21 layers. They compared their results with multilayer perceptrons trained on ECFP fingerprints and noted that the AUC was slightly lower for the CNN.

Improvements were made to the Chemception model, to create AugChemception [241]. In AugChemception, three additional channels were added to the image, creating a four channel image. Different possible properties to be represented by the additional channels were tested, including bond order, partial charge, orbital hybridization, and valency. For the toxicity prediction, the use of three extra channels slightly improved the AUC. However, the four different models using different information had indistinguishable performance. For the prediction of solvation energy, all four multi channel models had better performance than the single channel Chemception model. The two specific models that had better performance than the rest included information about both partial charge and hybridization. This means that the distribution of partial charges within a molecule and orbital hybridization have an impact on the solvation energy.

Shi et al. [242] used four separate CNNs for classifying organic chemicals as cytochrome P450 1A2 enzyme inhibitors, P glycoprotein pump blockers, blood brain permeable substances, and mutagens. The inputs were 300 px by 300 px molecular graphs with 3 color layers as the input to a 2 dimensional image processing CNN. They tested multiple neural networks and found that increasing the number of convolutional layers above 1 and increasing number of fully connected layers above 1, did not improve the model. The optimal structure had 14 convolutional filters, with a 21x21 kernel shape, in it's one layer, and a 14x14 max pooling layer. For predicting the enzyme inhibition, pump blocking, and blood brain barrier permeability, the test area under the curves (AUC) were 0.81, 0.83, and 0.75. However, for the Ames mutagenicity, the train and test AUC were 0.997 and 0.95, respectively. This means that the model is good at identifying patterns that are linked with mutagenicity.

Fernandez et al. [243] trained convolutional neural networks to perform classification of a compound as having or not having each of the 12 biological properties specified in the Tox21 challenge [244], including receptor agonism, antagonism, and enzyme inhibition. 300x300 images of the chemicals were used, with three methods of generation, indicating non-carbon atoms by colored letters, indicating non-carbon atoms by colored dots, and indicating non-carbon atoms by black letters and using color to indicate charge. None of the three different techniques of image generation had overall superior performance for all tasks, for example, colored letter graphs had better performance for predicting androgen receptor antagonism, while charge indicated graphs had better performance for predicting androgen receptor agonism. Grid search was used to determine the optimal input graph type, number of filters and kernel size for the convolutional layer. Interestingly, rotation of molecule images within the training and test set did not yield any changes in performance.

3 dimensional images can also be used as inputs for CNNs. Wallach et al. developed AtomNet [245], a deep CNN, which uses the previously generated 3 dimensional structures of a molecule, down sampled to a grid of 1 angstrom width, as the input. The goal of the network was to classify molecules as ligands with affinities greater than or less than K_i of 50 μ M. Chemicals that could be analyzed by the model were limited, since they had to fit within a 20 angstrom width

cube. The structure of the neural network consisted of four three dimensional convolutional layers and two fully connected layers. It provided test AUCs greater than 0.9 for over half of the 59 targeted receptors. One weakness of the model was that it had solely rgb channels as the input channels. The authors acknowledged that spurious connections were made between atoms that share little similarity, based on the similarity of the colors used to indicate them. This problem could have been resolved if, before down sampling, the color channels were expanded to have one channel for each element present in the set of molecules investigated.

In addition to being able to be represented as 2 and 3 dimensional images, molecules can be represented as graphs, where each atom is a vertex, and each bond is an edge. Convolutional neural networks can be used to generate custom molecular fingerprints, which can then be used as inputs to diverse machine learning algorithms similar to previously fingerprints. The benefit is that the method of generating these custom fingerprints can be trained, to provide the best performance for a particular task. Duvenaud et al. [246] determined that when linear regression was used to predict water solubility, CNN generated fingerprints had a lower root mean square error than extended-connectivity circular fingerprints, regardless of the fingerprint radii.

Ulrich et. al [247] used custom fingerprints from Duvenaud convolution as inputs to a neural network used to predict the octanol water partition coefficient. The neural network had two hidden layers, with 64 and 128 neurons. They obtained a root mean squared error of 0.80 logarithmic units. For comparison, they also computed the lipophilicity of the compounds in their test dataset using tools including ALOGPS[248], KOWWIN[249], and COSMO-RS[224]. The computational tools had RMSE errors ranging from 0.34 to 0.97.

Chen et al. used Graph Convolutional Neural Networks (GCNN) [250] to predict the values of 12 different toxicity indicators, including interactions with nuclear receptors and activating stress responses. Their training and test data consisted of 7831 compounds from the Tox21 dataset in MoleculeNet [238]. In their model, based on a method developed by Kipf and Welling [251], each atom represented a node, and each covalent bond was an edge. The information about nodes and edges was placed into a matrix, which was used as an input to the first layer of the convolutional network. Their model consisted of two parts. The first part consisted of convolutional layers and output a tensor representation of the molecule. Between each convolutional layer, noise was added. The second part was a two-layer perceptron that output the predicted value. They also trained additional machine learning models, such as kNN, Random Forest, and XGBoost with ECFP4 fingerprints, to set a baseline performance. Interestingly, the GCNN model performed only slightly better for 5 predicted values, while performing significantly worse for 5 other values [250]. Their conclusion was that using GCNNs provided no benefit over using conventional machine learning techniques.

Methodologies

Training and Test Dataset

The first step to forming machine learning (ML) models is to decide on the domain of the training and test data points that will be used. Models that are trained on a particular domain will only be effective in performing regression or classification on inputs within that domain.

In this particular case, since the solvents found to be used in amine-based post combustion capture and their degradation products contain solely carbon, oxygen, hydrogen, sulfur and nitrogen, the only molecules that will be used for model development and use will contain only those elements. In addition, the model will only cover molecules that contain carbon.

Since amines used as solvents in aqueous solutions must be significantly soluble in water, they are generally smaller molecules. A molecular size cut-off is also necessary for the training and test dataset. A reasonable size for this cutoff is 300 daltons, twice the weight of triethanolamine. This value will exclude molecules too heavy to become potential solvents, while not unnecessarily limiting the training dataset, allowing us to learn trends from smaller biological molecules and pharmaceuticals. If the property to be predicted does not have a significant dependence on weight, and the training dataset is significantly increased with a larger cut-off weight, a second version of the model will be trained with a cut-off of 500 daltons.

Solvents to be Analyzed

151 solvents found referenced within literature will be analyzed. 91 of them are non-cyclic amines, while 60 are cyclic amines. The most common class to be analyzed are the alkanolamines, with 29 members. For molecules with multiple amine groups, the degree was assigned the lowest degree of any of its amine groups. The numbers of primary, secondary, and tertiary amines are 63, 42, and 43, respectively. 63 of the solvents have more than one amine group.

Since it is not economically efficient to synthesize stereopure solutions from inorganic substances, the solvents that will be analyzed will be assumed to be composed of equal amounts of the different stereoisomers, with properties midway between the properties of each stereoisomer.

Primary and secondary amines form carbamates when exposed to carbon dioxide. For this reason, 139 carbamates of the solvents will also be analyzed. As a proxy for mutagenicity, the volatility of the nitramine and nitrosamines of primary and secondary amines will be determined.

Encoding Chemical Structures

Many organic chemicals found in databases are specified by descriptors such as SMILES strings. These formats have the benefit that they are compactly stored, can be displayed by general purpose software such as browsers, are somewhat human readable, and can be reconstructed

and analyzed by appropriate software. However, these strings are not suitable for use within general purpose machine learning models.

Since the working representations of the molecules do not need to be stored alongside others in large databases or transmitted through the internet, each one has a higher acceptable data size. Many ML models require highly structured data as inputs. Molecular fingerprints are expanded representations that follow structured formats, such as vectors, where every position has a unique meaning.

Chemical Fingerprints

Unlike descriptors like SMILES, for every specific molecular fingerprint format, each specific molecule has a single fingerprint representation. Similar molecules also have similar fingerprints. Using functions included in cheminformatics packages, such as RDKit, this degree of similarity can be computed [252]. This is particularly useful for ML models that use the degree of similarity between inputs and the training set, such as k-Nearest Neighbours. One problem with degree of similarity functions is that different fingerprints for a pair of molecules can have different degrees of similarity. Also, degree of similarity varies depending on which specific function is used to calculate similarity.

One method of generating chemical fingerprints is to consider whether certain substructures exist. The chemical structure can be converted to an array of Booleans, with each index specifying the existence of a certain substructure. This method is appropriate for tree-based regression and classification models, since they use decision branches based on the Boolean result of a comparison.

An example of a chemical fingerprint scheme that focuses on chemical substructures is the Morgan Algorithm, and derivatives such as Extended Connectivity Fingerprints (ECFP) [253]. In the ECFP algorithm, each atom is looked at individually. The individual atom and its first order neighbours of the atom are then analyzed to determine the specific substructure they form; the existence of this substructure is then noted. This process is then repeated with its first and second order neighbours, and then its 1st, 2nd, 3rd order neighbours, depending on the fingerprint radius specified by the user.

A more condensed chemical fingerprint generation algorithm is the MACCS key fingerprint. It consists of a set of 166 or 960 bits, each of which represents the presence or absence of a specific molecular substructure. The specific substructures represented by the bits of the key were manually curated and were chosen for their relevance to biological activity [254]. For this project, both ECFP and MACCS keys will be used and compared.

2 Dimensional Image Representations

Chemical structures can also be represented by 2 dimensional molecular graphs. Convolutional neural networks (CNNs) are deep neural networks that consist of repeated units, called kernels. CNNs are adept at learning to extract patterns that are relevant to the task, making them useful

for tasks with multidimensional data that have an autocorrelation between nearby data points, such as image classification [255].

For this project, 2 dimensional representations of chemicals will be used as inputs to the CNNs. The molecular graphs are in a 128 by 128 pixel rgb image, where carbon is indicated by black, oxygen by red, nitrogen by blue, and sulfur by yellow. The representation will not be kekulized, and aromatic bonds will be represented by dashed lines. To augment the training datasets, molecular graphs will be scaled and rotated.

Candidate Machine Learning Models

All of the machine learning models that will be used are supervised models. Both classification models and regression models will be used, depending on the type of data available.

The simplest type of model that will be used is k-Nearest Neighbours (kNN) regression and classification. In kNN, the training data is arranged in an n-dimensional plane, based off of the values of their input vector. To find the output for a set of inputs, the distance of the inputs and the training data points are looked at. The k closest training values are selected, where k is a parameter that can be optimized. For classification, the output is assigned the most common class within the k training points. For regression, the output becomes the average or weighted average y-value of the k training points.

Distance can be interpreted in many different ways. The most common way of defining distance is the Euclidean distance, which is defined as:

$$Euclidean\ distance = \sqrt{\sum(x_i - y_i)^2}$$

Where x and y are different points, and x_i represents the coordinate along the i^{th} axis. Another method of defining distance is the Manhattan distance:

$$Manhattan\ distance = \sum(x_i - y_i)$$

In some cases, the Manhattan distance can be more useful. In cases where the value of features can only be either 1 or 0, such as fingerprints, the k neighbours are the same regardless of whether Euclidean or Manhattan distance are used. However, if the neighbours are weighted, since the Euclidean distances are lower, more distant k-neighbours will be assigned a higher weight. This can affect performance.

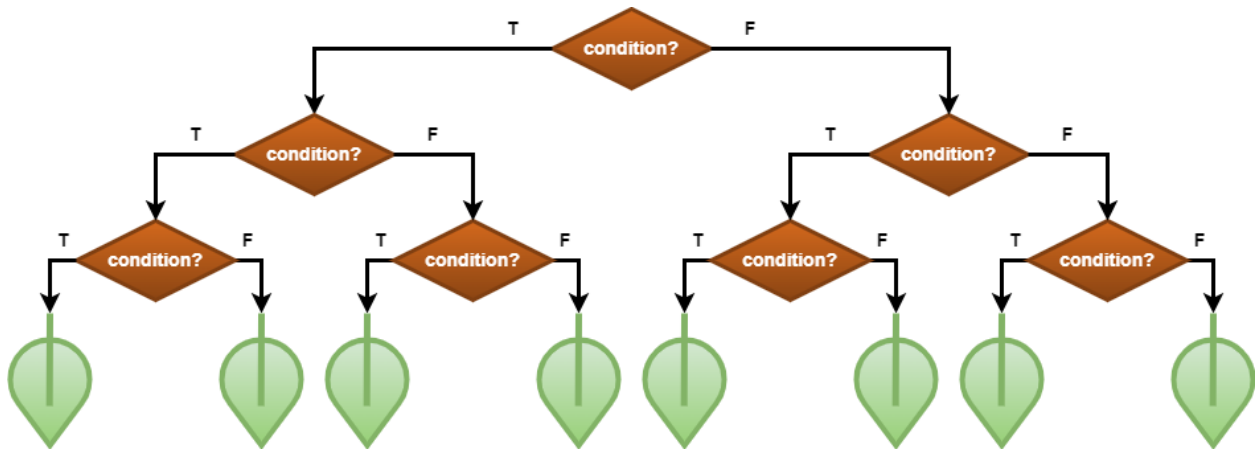


Figure 3 A binary decision tree with a depth of 3, diagram used with permission from Ghiasi et al. [107]

Another commonly used class of machine learning models are the tree-based methods. All tree-based methods are based on a simple learning unit called a decision tree. A decision tree consists of decision nodes, branches, and leaves. Starting at the topmost decision node, based on the results of a comparison, the processing travels down a particular branch to the next decision node or a leaf. The leaf it ends up at provides the predictions, whether is a value for a regression model or a class label for a classification model. Decision trees are trained using the greedy algorithm, growing the tree one layer at a time. However, decision trees are prone to overfitting.

To minimize the likelihood of overfitting occurring, an array of trees, called a random forest, can be created. Random forests avoid creating identical trees by independently training each tree on its own set of data points generated by bootstrap aggregation. In bootstrap aggregation, data points are chosen randomly, with replacement, from the original training dataset. Since the size of the datasets generated by bootstrap aggregation is the same as the original dataset, some data points are found as duplicates. Bootstrap aggregation prevents overfitting, since not every point is found in every tree's training dataset. The number of trees in a random forest is a parameter that can be optimized. Random forest models can perform both classification and regression. In random forest classification, the modal predicted class becomes the output. In random forest regression, the output is the average of the values predicted by the individual trees.

Boosting is a method where weak learners, such as small decision trees, are iteratively created, one at a time [256]. After the first weak learner is created, the training dataset is weighted, to give more weight to data points that were misclassified, or for regression, had predicted values far from their actual value. One popular tree-based boosted ML model is XGBoost. To reduce the impact of a single tree on training, and to prevent overfitting, XGBoost includes shrinkage, where the weights are multiplied by a parameter, η , and column subsampling, where only a fraction of the inputs are used for each particular tree [257].

Neural Networks

Neural networks are one of the most well-known machine learning techniques. A neural network is a structure that has subunits, called neurons (Figure 4). Each neuron accepts a series of inputs, which are then weighted. The sum of the weighted inputs is fed into the activation function, which provides the output of the neuron. Different functions can be used as the activation function; the most popular are the Rectified Linear Unit (ReLU), sigmoid, and tanh.

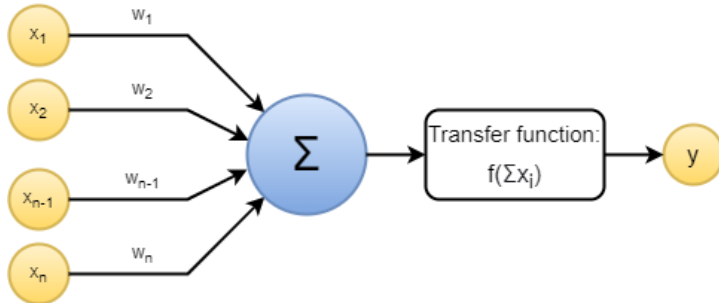


Figure 4 A single computational neuron, diagram used with permission from Ghiasi et al. [107]

The simplest neural network is the perceptron. It consists of a set of input nodes and single neuron to process the inputs and generate a single output. However, this low level of complexity is usually insufficient to model most phenomena. To model more complex phenomena, neural networks can have hidden neurons between the input nodes and output neuron. In this study, only feed forward neural networks will be used.

A simple form of the feed forward network is the multilayer perceptron (MLP). In a multi-layer perceptron, neurons are arranged into multiple layers: the inputs, hidden layers, and the outputs. MLPs are fully connected, the inputs to all of the neurons in a single layer are the same. The outputs of a layer become inputs of the next layer. MLPs are useful since they can take advantage of matrix multiplication capabilities of GPUs. A diagram of an MLP with a single hidden layer is shown in Figure 5.

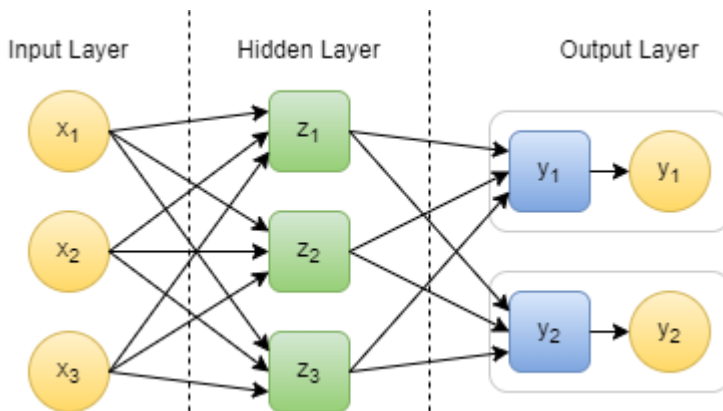


Figure 5 A multilayer perceptron with a single hidden layer, diagram used with permission from Ghiasi et al. [107]

One weakness of MLPs is that the number of trainable weights of the first hidden layer is proportional to the number of inputs. For input data forms with many inputs, such as images, it is infeasible to use MLPs. Instead, if the data is arranged into one or two dimensions, a Convolutional Neural Network (CNN) can be used. The basis for CNNs is the convolutional layer, also known as a filter. A convolutional layer has a data window, which slides over the data. For each position of the window, the inputs within it are multiplied by trained weights within a kernel, then added together, to form an output point, as can be seen in Figure 6.

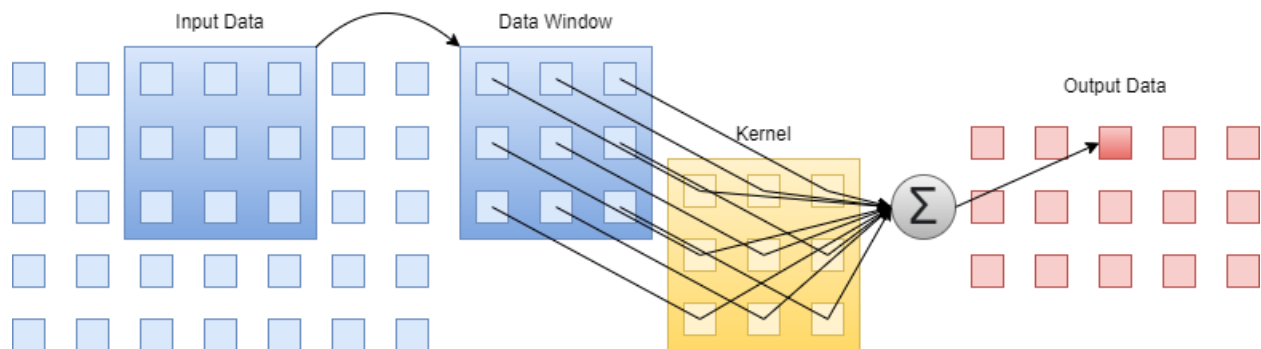


Figure 6 Convolution operation, diagram used with permission from Ghiasi et al. [107]

CNNs can have filters in parallel and in series. They also commonly include elementwise non-linear operation layers in between filters. One strength of neural networks is that they can be modularized and customized using libraries such as TensorFlow [258]. Different types of neural structures can be included within a single network. A common implementation is to have convolutional layers at the beginning and fully connected layers at the end.

Optimization

To find the best hyperparameters for an ML model, optimization is used. For optimization, versions of the same model with different hyperparameters are created. These models are tested using the validation data. The hyperparameters that yield the lowest validation error are then used for training, testing and making predictions.

To minimize biases due to the particular validation dataset used, training and test data are commonly split in different ways. For k-cross validation, as can be seen in Figure 7. A copy of the model is trained on each of the training sets and is tested on its respective validation set. This yields a series of k errors. The mean, maximum, and standard deviation of these errors can be taken into account when picking the ideal parameters to use in the rest of the study.

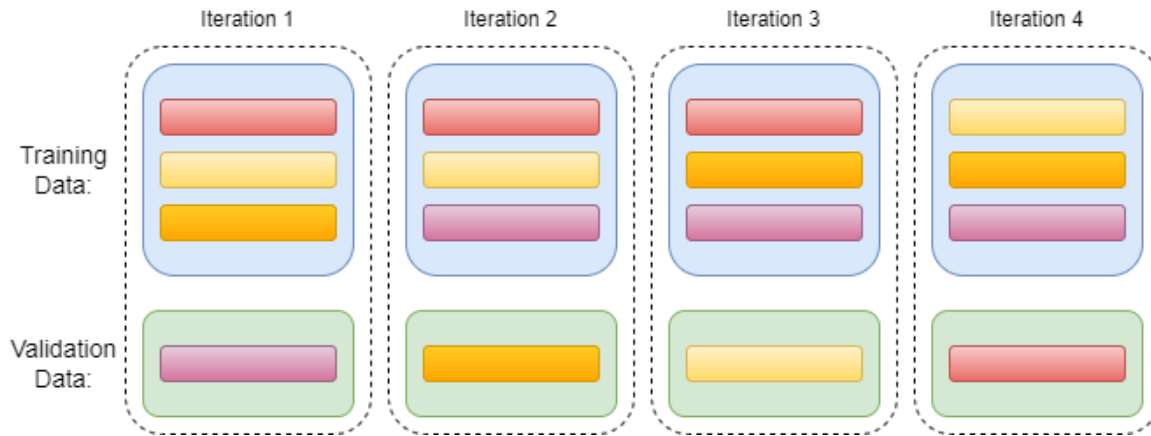


Figure 7 Distributing data into k train-test data groupings, for k -fold cross validation

Grid search is the most straightforward hyperparameter optimization method. In grid search, possible values for each hyperparameter are specified in a vector. Every possible set is then tested with k -fold cross validation, and their performance is compared with each other. The set of hyperparameters with the best performance is used to build a model. However, for certain models that take a long time to train, such as XGBoost, or for a large set of possible hyperparameter values, it is not practical to use grid search. Instead, random search can be used to look at a random selection of the possible sets of variables.

Results of the Property Prediction Models

For the primary and secondary amines, the carbamates will be analyzed concurrently.

Volatility

To detect overfitting, where the machine learning model learns to accurately predict the vapor pressures of the chemicals represented in the training data set, but cannot generalize to chemicals outside of the dataset, the data was split in a manner, where each chemical and all of its data points were placed into a single bin, the test bin or one of the five train-validation bins.

As can be seen in Figure 8, the logarithm of the pressure for each single chemical has an almost linear relationship to the inverse of the temperature. The Clausius-Clapeyron equation for ideal gases is [259] :

$$\frac{1}{P} dP = \frac{\Delta H_{vap}}{R} \frac{1}{T^2} dT$$

Equation 1 Clausius Clapeyron equation for ideal gases

Where ΔH_{vap} is the enthalpy of vaporization and R is the ideal gas constant. Assuming that the enthalpy of vaporization is constant over a small temperature range, and integrating, Equation 2 represents the relationships seen for a single molecule in the training dataset.

$$\log_{10} P - \log_{10} P_0 = -\frac{\Delta H_{vap}}{R \ln 10} \left(\frac{1}{T} - \frac{1}{T_0} \right)$$

Equation 2 Clausius Clapeyron equation assuming enthalpy of vaporization is constant

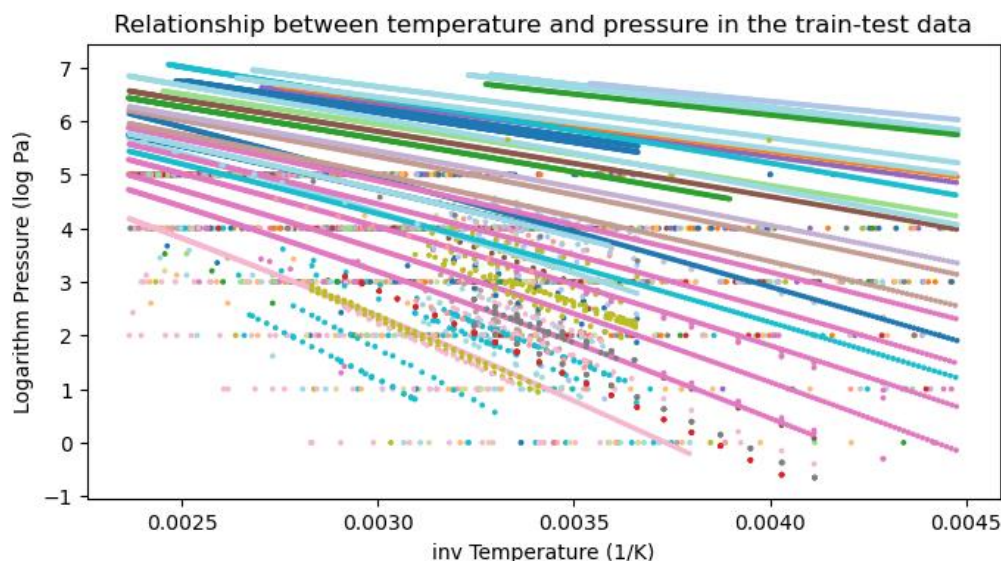


Figure 8 Relationship between the temperature and the pressure in the train-test data

Modular neural networks (Figure 9) were trained to predict the logarithm of the vapor pressure based on the inverse temperature and a chemical representation, graph or fingerprint. Random

search and 5 fold cross validation were used to find the optimal parameters. When training, early stopping was used to minimize the degree of overfitting to the molecules in the training dataset, and not being able to generalize.

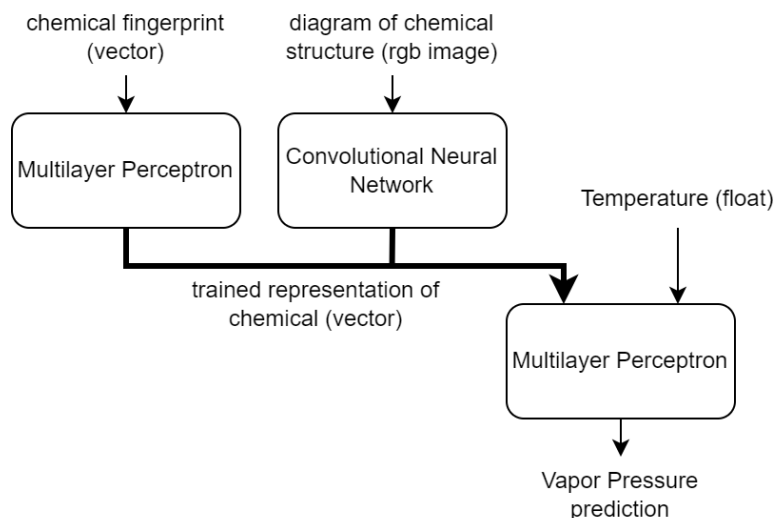


Figure 9 Structure of Modular Neural Networks created to predict vapor pressures

Models were also trained using solely the molecular mass and the number of atoms of each element in a molecule. Table 2 displays the train, validation, and test root mean square error for logarithmic pressures. As can be seen, the fingerprint models have low prediction errors for the training molecules. However, they have high test errors. This suggests that the models were overfitting, and memorizing the parameters for the specific molecules in the training set instead of general trends. Moreover, the models trained on weight and the models trained on numbers of atoms of each element had strong correlations with each other, which was not seen between any other two models (Figure 24).

Table 2 Optimized Parameters for neural network models used to predict the vapor pressures based on different chemical representations and temperature

Chemical input format	Optimized chemical vector length	Number of Neurons/Filters in Layers of Part 1	Number of Neurons in Layers of Part 2	Train RMSE	Validation RMSE	Test RMSE
MACCS keys	7	[8]	[32, 16, 4]	0.135	1.852	1.349
MACCS keys	5	[32, 16, 4]	[32, 16]	0.534	1.645	2.215
MACCS keys	3	[4]	[32, 8]	0.238	0.820	1.267
MACCS keys	5	[4]	[32, 16]	0.190	2.055	3.490
ECFP4 fingerprint	3	[16, 8, 4]	[32, 16]	0.216	0.738	1.947

ECFP4 fingerprint	3	[8, 4]	[32, 16]	0.793	1.914	1.692
ECFP4 fingerprint	7	[8, 4]	[32, 16, 4]	0.763	1.807	1.602
ECFP4 fingerprint	5	[32, 8]	[16, 8]	0.455	1.384	1.857
128x128 bitmap	7	[16, 32, 8, 8, 32, 8]	[16, 8]	0.848	1.288	2.131
128x128 bitmap	3	[16, 32, 32, 8, 16, 16]	[16]	1.194	2.314	1.952
128x128 bitmap	5	[8, 8, 32, 16, 8, 8]	[32, 4]	0.814	2.210	2.392
128x128 bitmap	3	[16, 4, 16, 8, 8, 32]	[4]	1.042	2.658	2.761
Molecular Mass	1	[]	[4, 32]	0.611	0.823	0.928
Molecular Mass	1	[]	[32]	0.588	0.903	0.952
Molecular Mass	1	[]	[8, 4]	0.722	0.968	1.037
Molecular Mass	1	[]	[1]	1.196	1.645	1.538
Number of Atoms of each Element	5	[]	[8, 32, 4]	1.098	2.564	2.230
Number of Atoms of each Element	5	[]	[32, 32, 8]	0.804	1.488	1.509
Number of Atoms of each Element	5	[]	[32]	1.038	1.203	1.242
Number of Atoms of each Element	5	[]	[1]	1.241	1.476	1.536

As expected, vapor pressure predictions were continuous with temperature. For all of the models, a lower molecular mass was associated with higher vapor pressures (Figure 10). This is reasonable since the heavier solvents have a larger molecular surface area which can interact with other molecules in the liquid phase, increasing the Van der Waals forces. All models, save for the two lowest performing CNN models, predicted that the vapor pressure at any specific temperature decreases with carbamation. The best performing models predicted a geometric mean of approximately 1 logarithmic unit, or a tenfold decrease in vapor pressure. This is due to carbamation both increasing the surface area of the molecule, and due to hydrogen bonding.

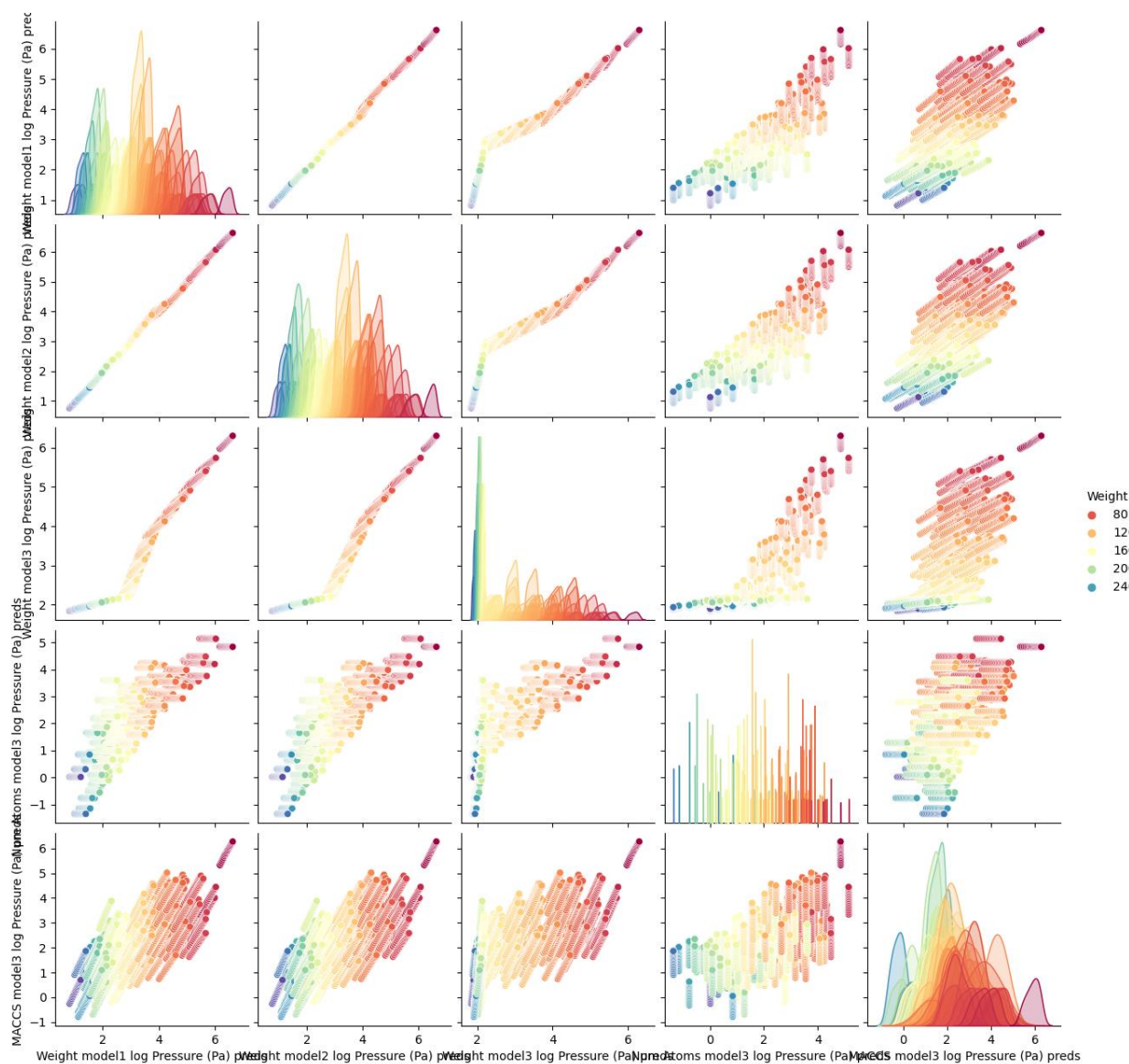


Figure 10 Pair plot of the predicted logarithm of vapor pressures at temperatures from 5 to 95 °C, for solvents and carbamates. Color corresponds with weight. 5 models with the least RMSE test error were included for space.

In every model, the non-cyclic and cyclic amine solvents had significantly overlapped distributions. Within the non-cyclic amines, amino acids and alkyl polyamines tended to have the lowest vapor pressures. This can be partially explained by size, with a mean mass of 153 Da and 162 Da, for the amino acids and alkyl polyamines, compared to a mass of 130 Da for other noncyclic solvents. Additionally, amino acids have dipole-dipole intermolecular forces and hydrogen bonding. In the cyclic amines, the pyrrolizidines had the lowest vapour pressures. However, this can be attributed to size, since the mean weight of the potential solvent pyrrolizidines were 198 Da, while the mean weight of the other cyclic solvents were 135 Da.

Based on an absorber temperature of 40°C [260], the equilibrium vapor pressures of pure solvents were sorted and compared to each other. The 5 substances with highest and lowest

vapor pressures in pure form are listed in Table 3. Vapor pressures in the absorber are lower than the listed vapor pressures, since in the absorber, the amine is in the aqueous phase. However, at a single concentration, the equilibrium vapour pressures will be approximately proportional. In addition, amines will not reach equilibrium pressure in the absorber due to continuous gas flow. The highest vapor pressures were populated by small molecules, and the lower vapor pressures were populated by relatively larger molecules. Overall, an important predictor of whether a solvent will have a high or low vapor pressure is its size.

Table 3 Solvents with highest and lowest vapor pressures at 40°C, based on mean of 5 best performing models

Highest Vapor Pressure at 40 °C			Lowest Vapor Pressure at 40 °C		
Name	SMILES	Pressure	Name	SMILES	Pressure
Dimethylamine	CNC	826 kPa	N-[2-(tetrahydro-1H-pyrrolizin-7a(5H)-yl)ethyl]aniline	c1ccc(NCCC23CCCN2CCC3)cc1	19 Pa
Ethylenediamine	NCCN	160 kPa	1-propylamido-3-butyl imidazolium	CCCCn1cc[n+](CCC(N)=O)c1	21 Pa
Monoethanolamine	NCCO	80 kPa	Tryptophan	NC(Cc1c[nH]c2ccccc12)C(O)=O	22 Pa
Pyrrolidine	C1CCNC1	66 kPa	Bis-tris methane	OCCN(CCO)C(CO)(CO)CO	23 Pa
N,N-Dimethylethanamine	CCN(C)C	63 kPa	N-(Tetrahydro-1H-pyrrolizin-7a(5H)-ylmethyl)aniline	c1ccc(NCC23CCCN2CCC3)cc1	27 Pa

Lipophilicity

To predict lipophilicity, as indicated by the octanol-water partition, machine learning models were used and compared with a computational model. 8 different fingerprint models were trained. For each model, the found experimental data were split into three datasets for training, validation, and testing. For the kNN, RF, and MLP models the ideal hyperparameters were determined using grid search with 5-fold cross validation. However, for the two XGBoost models, the ideal hyperparameters were found using random search. The neural networks used a ReLU activation function and were trained using the Adam algorithm. They are shown below in Table 4.

Table 4 Hyperparameters for Fingerprint Models trained to predict Lipophilicity

Model	Fingerprint Scheme	neighbours/estimators/hidden neurons	weights and distance metric /maximum depth/ layers	column subsample	row subsample	test RMSE
kNN	MACCS	8	1/d, Manhattan	-	-	0.9250
kNN	ECFP4	8	1/d, Manhattan	-	-	1.3869
RF	MACCS	256	19	Sqrt	-	0.7642
RF	ECFP4	256	19	Sqrt	-	1.0311
XGBoost	MACCS	256	8	0.5	1	0.7397
XGBoost	ECFP4	256	9	0.5	1	0.8618
MLP	MACCS	(32)	1	-	-	0.7914
MLP	ECFP4	(128)	1	-	-	1.0097

For the convolutional neural networks, 7 layer CNNs were used, in which after each convolutional layer max pooling was used. Since the train-test dataset was large, with 7761 training points and 1940 test points, the data had to be fed in batches. A batch size of 32 points was found to optimize the training speed of each epoch. Optimal parameters were found using 5 fold cross validation. Three models were selected and used for further training, listed below in Table 5. For the training, early stopping was used.

Table 5 Hyperparameters for CNNs trained to predict Lipophilicity

CNN label	Convolutional Layers	Fully Connected Layers	Training Epochs	Test RMSE
1	[32, 32, 16, 32, 8, 16, 16]	[8]	44	0.9768
2	[32, 32, 32, 16, 16, 16, 16]	[8]	30	0.9165
3	[32, 32, 32, 16, 16, 32, 16]	[8]	43	0.9971

The atomic block contribution model developed by Wildman and Crippen in 1999 [261] was used to form baseline logP calculation to compare with the newly generated ML models. In the atomic block contribution model, each atom in the molecule is categorized within one of 68 atom types based on their element and their neighbours. The presence of each atom type has an additive effect on the net prediction, and the 68 additive factors were determined using linear least squares on a set of 9920 molecules and their experimental logP values. When comparing to experimental measurements of logP, the predictions from the model had an RMSE of 0.868.

Lipophilicity predictions from different models are positively and moderately to strongly correlated. As can be seen in Figure 25 in the appendix, the strongest correlations are between models trained with the same input data format.

Carbamation of a solvent did not cause consistent changes in its lipophilicity, based on most models. For some molecules, the change was decreased lipophilicity, while for others it was increased lipophilicity. The hyperclass, non-cyclic or cyclic, did not change this outcome. Altogether, the predicted change of log P due to carbamation had a mean of 0.06 and standard deviation of 1.02. With the exception of the atomic block contribution model, for individual models, the standard deviation of the change in log P ranged from 0.65 to 1.120. In the atomic block contribution model, the standard deviation was 0.099, and there were a finite number of possible differences due to the nature of the model itself; the values that were present were around the median of 1.0258.

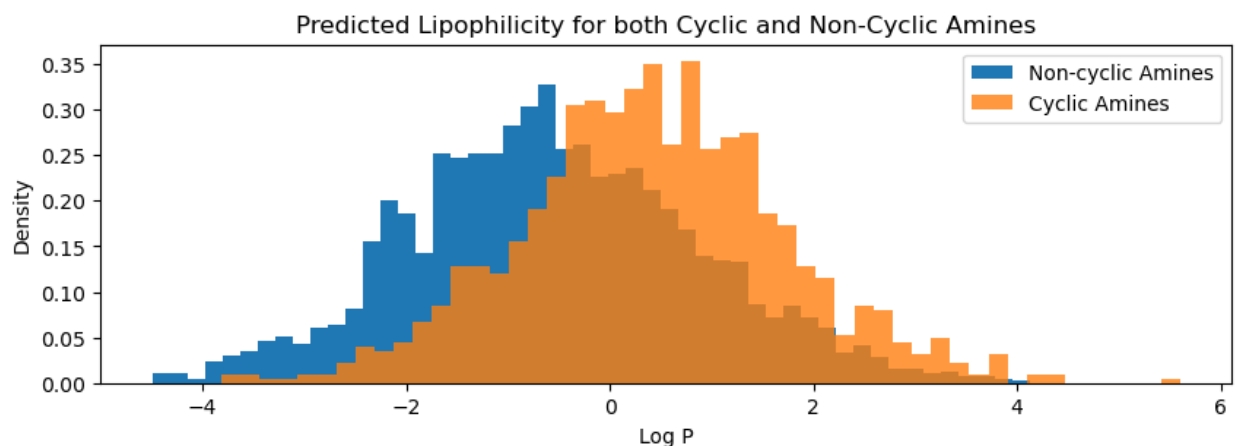


Figure 11 Histogram of the predicted lipophilicities of both cyclic and non-cyclic amines

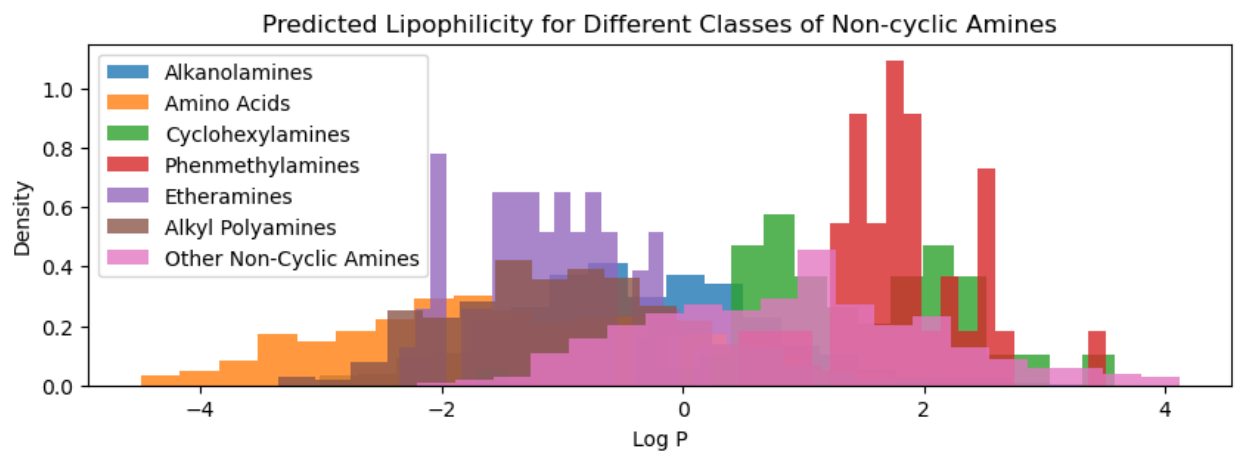


Figure 12 Histogram of the predicted lipophilicities of different classes of non-cyclic amines

Non-cyclic amines tended to have lower lipophilicities, as can be seen in Figure 11. Based on the recently trained machine learning models, within both non-cyclic amines (Figure 12) and all amines, amino acid solvents and carbamates had the lowest lipophilicities. This is reasonable, since the carboxylic acid group is strongly polar, which increases water solubility. In contrast, for predictions made using the atomic block contribution model, alkyl polyamine carbamates had

lower lipophilicities than amino acids. This is likely due to model linearly adding negative log P contributions from similar amine groups, causing an underestimation of the lipophilicity.

Based on both the newly trained ML models and the atomic block contribution model, the most lipophilic molecules were solvents that had one or more terminal alkyl groups with the same or larger size than butyl, phenyl, or terminal cyclohexyl groups. These three types of substituents are nonpolar, which decreases their solubility in water, and increases log P. Terminal positions have a stronger affect on lipophilicity, since a larger portion of the group is available to interact with surrounding molecules.

The most lipophilic compound was tributylamine, with models predicting a log P from 2.75 to 4.12, with a mean of 3.40. The next molecules with strongest lipophilicity were 1-butyl-3-methylpyridinium, 1-octyl-3-methylimidazolium, and N-(Tetrahydro-1H-pyrrolizin-7a(5H)-ylmethyl)aniline, with mean predicted log P values of 2.73, 2.61, and 2.58, respectively. The least lipophilic compounds were amino acids (Table 6).

Table 6 Solvents predicted to be the most lipophilic and the least lipophilic

Most Lipophilic			Least Lipophilic		
Name	SMILES	logP prediction	Name	SMILES	logP prediction
tributylamine	<chem>CCCCN(CCCC)CCCC</chem>	3.37±0.42	Arginine	<chem>N=C(N)NCCCC(N)C(=O)O</chem>	-3.13±1.18
1-butyl-3-methylpyridinium	<chem>CCCC[n+]1cccc(C)c1</chem>	2.79±1.29	Aspartic Acid	<chem>NC(CC(=O)O)C(=O)O</chem>	-3.06±0.98
1-octyl-3-methylimidazolium	<chem>CCCCCCCn1cc[n+](C)c1</chem>	2.61±0.96	Glutamine	<chem>NC(=O)CCC(N)C(=O)O</chem>	-2.98±0.87
N-(Tetrahydro-1H-pyrrolizin-7a(5H)-ylmethyl)aniline	<chem>c1ccc(NCC23CCCN2CCC3)c1</chem>	2.57±0.61	Asparagine	<chem>NC(=O)CC(N)C(=O)O</chem> <chem>NC(CO)C(=O)O</chem>	-2.93±0.98
1-methyl-3-octylimidazolium	<chem>CCCCCCC[n+]1ccn(C)c1</chem>	2.57±0.89	Serine	<chem>NC(CO)C(=O)O</chem>	-2.91±0.45

Based on the result, to lower lipophilicity, and the potential for solvents to bioaccumulate and cause long term toxicity, large nonpolar groups such as alkyl groups with 4 or more carbons, phenyl, and cyclohexyl groups should be avoided. The presence of polar groups, such as carboxylic acid, can further lower the lipophilicity.

One of the goals of choosing a solvent to use in a carbon capture system is to maximize the amount of carbon dioxide that a given volume of solution can absorb while falling down a given height of the absorber. Ignoring effects of drop size, for two solvents with the same carbon absorption capacity and kinetics, the solvent with higher water solubility can have a higher

concentration, and thus absorb more carbon dioxide at a faster rate. Since water solubility and lipophilicity are related, the environmental goal of minimizing bioaccumulation by minimizing lipophilicity is aligned with the functional goal of maximizing the concentration of carbon a solution can absorb in a given amount of time.

Mutagenicity

Mutagenicity will be compared using predictions for both the Ames Classification test and the volatility of nitramines and nitrosamines.

Ames Classification Test

Strains TA98 and TA1537 detect frameshift mutations [262], while strains TA100, TA102, and TA1535 detect base pair substitutions [263] [264]. To increase the number of training points, the results from each class of mutation type are combined together (Table 7). When the results of each test type in a mutagenicity group are the same, the class of the training data point is assigned to that result. However, when they have conflicting results, the training data point is assigned to a class probabilistically before the training of a particular model. Since the training dataset is biased, most of the chemicals it contains are mutagenic, it is expected that the models will mislabel many non-mutagenic chemicals as being mutagenic and have a low specificity.

Table 7 Ames mutagenicity raw experimental classification, and pooled training data

Strain	Number of Valid Samples	Detected Mutagenicity	No Detected Mutagenicity	% Detected Mutagenicity
TA98	253	192	61	76 %
TA1537	88	47	41	53 %
TA100	218	152	66	70 %
TA102	27	10	17	37 %
TA1535	105	21	84	20 %
Common Mechanism				
Frameshift	255	189.5*	65.5*	74%
Base pair substitution	223	139*	84*	62%

*signifies expected number of train-test data points in the class

For the fingerprint trained models, ten instances of each model will be trained in parallel, each with a different version, with classes picked at random, of the training and test datasets. However, since the CNNs can be trained and output probabilities, only a single instance of the optimized network will be used. Parameters of these sets will be found using random search. The mean accuracy, sensitivity, and specificity of the parallel models are listed below in Table 8. The parameters were found using 5-fold cross validation.

Table 8 Test performance of different mutagenicity prediction models

Mutation Type	Input data	Model Type	Parameters	Accuracy	Sensitivity	Specificity
---------------	------------	------------	------------	----------	-------------	-------------

Frameshift	MACCS keys	kNN	n_neighbors=8, weights='distance', metric='manhattan'	0.833	0.976	0.211
Frameshift	MACCS keys	RF	n_estimators=50, max_depth=19	0.861	0.947	0.467
Frameshift	MACCS keys	XGB	n_estimators=30, max_depth=19, subsample=0.75, colsample_bytree=0.25	0.847	0.986	0.227
Frameshift	MACCS keys	MLP	hidden_layer_sizes=(150, 50), activation='relu'	0.867	0.983	0.374
Frameshift	ECFP4	kNN	n_neighbors=7, weights='uniform', metric='euclidean'	0.829	0.889	0.563
Frameshift	ECFP4	RF	n_estimators=125, max_depth=9	0.865	0.957	0.469
Frameshift	ECFP4	XGB	n_estimators=90, max_depth=3, subsample=0.75, colsample_bytree=0.25	0.888	0.964	0.562
Frameshift	ECFP4	MLP	hidden_layer_sizes=(50), activation='logistic'	0.875	0.966	0.463
Base pair substitution	MACCS keys	kNN	n_neighbors=2, weights='uniform', metric='manhattan'	0.744	0.709	0.796
Base pair substitution	MACCS keys	RF	n_estimators=150, max_depth=7	0.742	0.9924	0.409
Base pair substitution	MACCS keys	XGB	n_estimators=110, max_depth=2, subsample=0.5, colsample_bytree=1	0.731	0.939	0.451
Base pair substitution	MACCS keys	MLP	hidden_layer_sizes=(50, 25), activation='tanh'	0.782	0.943	0.562
Base pair substitution	ECFP4	kNN	n_neighbors=9, weights='distance', metric='euclidean'	0.656	0.791	0.481
Base pair substitution	ECFP4	RF	n_estimators=125, max_depth=12	0.713	0.931	0.451
Base pair substitution	ECFP4	XGB	n_estimators=100, max_depth=6, subsample=0.75, colsample_bytree=1	0.656	0.768	0.516
Base pair substitution	ECFP4	MLP	hidden_layer_sizes=(150), activation='logistic'	0.727	0.861	0.548
Frameshift	64x64 bitmap	CNN	n_filters=[4,16,16,16,16,8,16], n_neurons=[16]	0.745	0.850	0.364
Frameshift	64x64 bitmap	CNN	n_filters=[8,4,16,4,16,16,8], n_neurons=[16]	0.804	0.949	0.333

Frameshift	64x64 bitmap	CNN	n_filters= [4,16,4,16,16,8,8], n_neurons=[8]	0.863	0.974	0.500
Base pair substitution	64x64 bitmap	CNN	n_filters=[16,8,8,16,16,4,4], n_neurons=[16]	0.644	0.731	0.526
Base pair substitution	64x64 bitmap	CNN	n_filters=[8,8,8,16,16,4,16], n_neurons=[16]	0.578	0.643	0.471
Base pair substitution	64x64 bitmap	CNN	n_filters=[16,8,16, 8,16,16,16], n_neurons=[16,8]	0.600	0.704	0.444

Different models had different predictions for the mutagenicity of solvents and carbamates. As can be seen in Table 9, for most models, the mean predicted probability of a specific type of mutagenicity for the set of chemicals was below 0.5. The exceptions were base pair substitution models trained on ECFP4 fingerprints. For most models, a large proportion of the chemicals had mutagenicity probabilities less than 0.05, with a smaller cluster of having high mutagenicity probabilities. However, this U-shaped pattern is not present in base pair substitution models trained on ECFP4 fingerprints (Figure 26). This suggests that the ECFP4 fingerprints do not contain an element which is a consistent predictor of mutagenicity.

Table 9 Mean solvent and carbamate mutagenicity probability predictions of different models

Frameshift			Base pair substitution		
Input Data	Model	Mean Mutagenicity Probability	Input Data	Model	Mean Mutagenicity Probability
MACCS keys	kNN	0.226	MACCS keys	kNN	0.069
MACCS keys	RF	0.047	MACCS keys	RF	0.387
MACCS keys	XGB	0.133	MACCS keys	XGB	0.291
MACCS keys	MLP	0.278	MACCS keys	MLP	0.412
ECFP4	kNN	0.016	ECFP4	kNN	0.553
ECFP4	RF	0.044	ECFP4	RF	0.589
ECFP4	XGB	0.035	ECFP4	XGB	0.485
ECFP4	MLP	0.426	ECFP4	MLP	0.697
bitmap	CNN	0.253	bitmap	CNN	0.172
bitmap	CNN	0.157	bitmap	CNN	0.499
bitmap	CNN	0.187	bitmap	CNN	0.232

An interesting pattern was that, for both frameshift and base-pair substitution mutagenicity, molecules with multiple amine groups were predicted to be mutagenic by a larger proportion of models (Figure 13). In fact, the 19 and 17 molecules predicted to have frameshift and base pair substitution mutagenicity, respectively, by the most models were all multi-amines. The molecules predicted by the most models to have frameshift and base pair substitution

mutagenicity were the carbamate 3-(2-amino-2-carboxyethyl)-1H-indole-1-carboxylate and the solvent 3-(3-amino-3-oxopropyl)-1-butyl-1H-imidazol-3-ium, respectively.

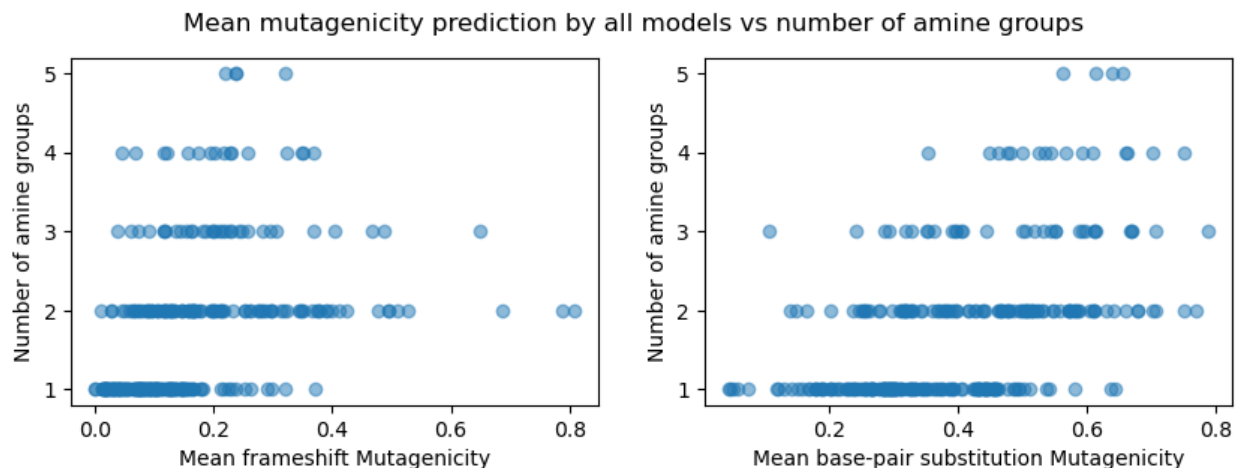


Figure 13 Mean mutagenicity predicted by all models for different numbers of amine groups

Distributions of models predictions had the same shapes for solvents and carbamates, and noncyclic and cyclic amines. This means that those classes do not have a consistent impact on the Ames mutagenicity of a molecule. Overall, to avoid the use of amines that are likely to be mutagenic, monoamines should be used.

Most base pair substitution mutagenic			Least base pair substitution mutagenic		
Name	SMILES	Mean mutagenicity probability	Name	SMILES	Mean mutagenicity probability
1-propylamido-3-butyl imidazolium	<chem>CCCCn1cc[n+](CCC(N)=O)c1</chem>	0.79	2-Pyridinolate	<chem>[O-]c1ccccn1</chem>	0.045
Tryptophan carbamate	<chem>N(C([O-])=O)C(Cc1c[nH]c2cccc12)C(O)=O</chem>	0.77	Benzyl methyl amine carbamate	<chem>CN(C([O-])=O)Cc1cccc1</chem>	0.046
2-(Tetrahydro-1H-pyrrolizin-7a(5H)-yl) ethanamine	<chem>NCCC12CCCN1CCC2</chem>	0.75	Aniline	<chem>Nc1cccc1</chem>	0.052
1,2-Bis(3-aminopropylamino) ethane	<chem>NCCCNCNCCCN</chem>	0.75	Butyl pyridinium	<chem>CCCC[n+]1cccc1</chem>	0.058
3-(Aminopropyl) ethylenediamine	<chem>NCCCNCN</chem>	0.71	3-Pyridinolate	<chem>[O-]c1ccncc1</chem>	0.076
Most frameshift mutagenic			Least frameshift mutagenic		
Name	SMILES	Mean mutagenicity probability	Name	SMILES	Mean mutagenicity probability

Tryptophan carbamate	<chem>NC(Cc1cn(C([O-])=O)c2ccccc12)C(O)=O</chem>	0.81	N-tert-Butyl ethanolamine carbamate	<chem>CC(C)(C)N(C([O-])=O)CCO</chem>	0.00045
Tryptophan carbamate	<chem>N(C([O-])=O)C(Cc1c[nH]c2ccccc12)C(O)=O</chem>	0.79	1-methyl piperidine	<chem>CN1CCCCC1</chem>	0.00049
Tryptophan	<chem>NC(Cc1c[nH]c2ccccc12)C(O)=O</chem>	0.69	1-methyl piperazine carbamate	<chem>CN1CCN(C([O-])=O)CC1</chem>	0.011
1H-Imidazo[4,5-b]pyridine	<chem>c1cnc2[nH]cnc2c1</chem>	0.65	Threonine	<chem>CC(O)C(N)C(=O)O</chem>	0.012
N-Methyl-1-(tetrahydro-1H-pyrrolizin-7a(5H)-yl)methanamine	<chem>CNCC12CCCN1CCC2</chem>	0.53	Methyl diethanolamine	<chem>CN(CCO)CCO</chem>	0.012

Nitramine and Nitrosamine Volatility

Vapor pressures of nitramines and nitrosamines were predicted using the same models used to predict vapor pressures of solvents and carbamates. As expected, the same patterns of heavier molecules having lower vapor pressures, and carbamates having reduced vapor pressures, was present.

Nitrosamines had larger vapor pressures than nitramines (Figure 14). This can be explained by both the larger surface area of nitramines which can interact with other molecules, and the fact that nitramines contain a resonant electron which can become polarized and contribute to dipole-dipole forces. When the vapor pressures at 40 °C were listed, the lowest vapor pressures belonged to carbamate nitramines and the highest vapor pressures belonged to solvent nitrosamines. In addition, nitrosamines are more mutagenic than nitramines [265]. For this reason, nitrosated solvents will be compared.

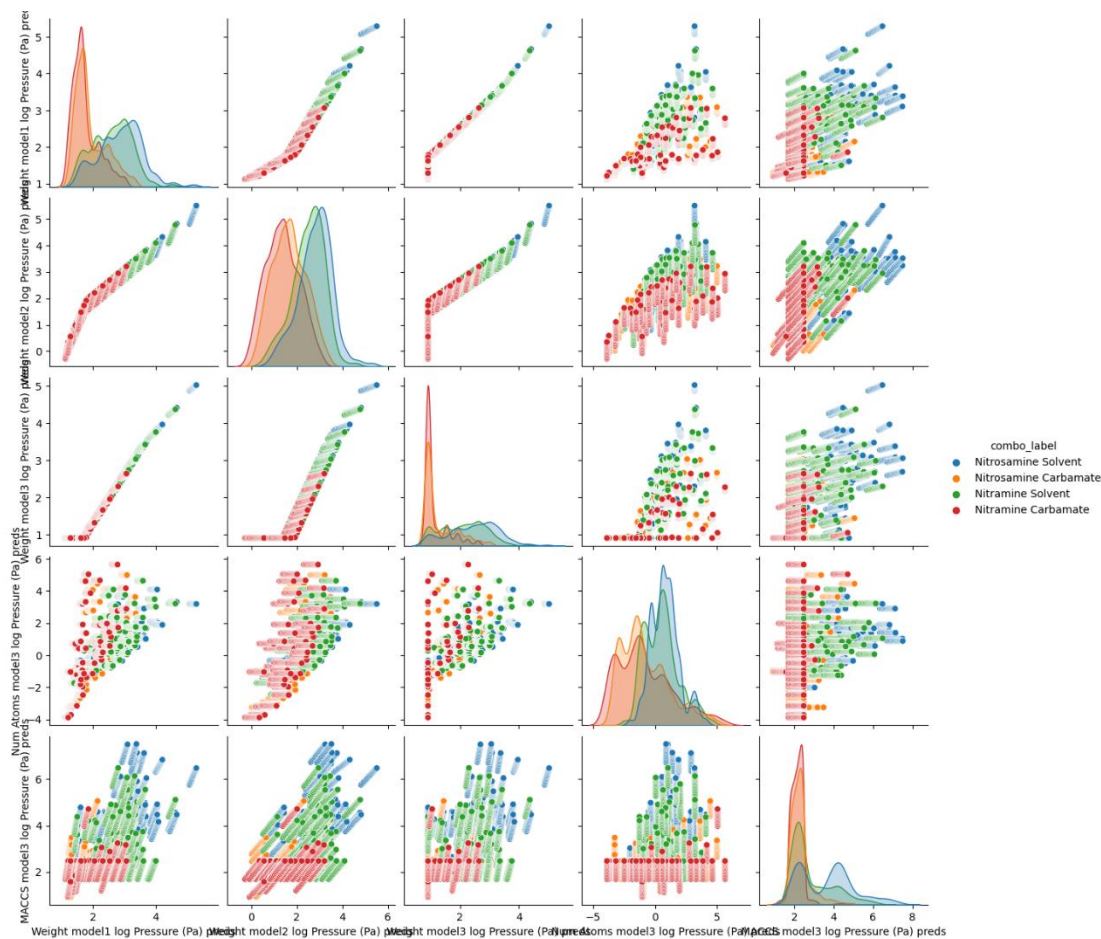


Figure 14 Pair plot of the predicted logarithm of vapor pressures at temperatures from 5 to 95 °C, for nitramines and nitrosamines of solvents and carbamates. The 5 models with the least RMSE test error were included for space.

The most and least volatile nitrosated solvents, based on the equilibrium vapor pressure at 40 °C, are listed in Table 10. As can be seen, solvents with high volatility were predicted to have nitrosamines with high volatility, and solvents with low volatility were predicted to have nitrosamines with low volatility. Overall, larger solvents had lower equilibrium vapor pressures for nitramines and nitrosamines.

Table 10 Nitrosated solvents with highest and lowest vapor pressures at 40°C, based on mean of 5 best performing models

Highest Vapor Pressure at 40 °C			Lowest Vapor Pressure at 40 °C		
Name	SMILES	Pressure	Name	SMILES	Pressure
Dimethylamine nitrosate	<chem>CN(N=O)C</chem>	81 kPa	1-propylamido-3-butyl imidazolium nitrosate	<chem>CCCCn1cc[n+](CCC(N(N=O))=O)c1</chem>	8.7 Pa
Ethylenediamine nitrosate	<chem>N(N=O)CCN</chem>	14 kPa	Tryptophan nitrosate	<chem>N(N=O)C(Cc1c[nH]c2ccccc12)C(O)=O</chem>	11.3 Pa

Monoethanolamine nitrosate	<chem>N(N=O)CCO</chem>	13 kPa	N-(Tetrahydro-1H-pyrrolizin-7a(5H)-ylmethyl)aniline nitrosate	<chem>c1ccc(N(N=O)CC23CCCN2CCC3)cc1</chem>	12.6 Pa
Pyrrolidine nitrosate	<chem>C1CCN(N=O)C1</chem>	12 kPa	N-(3-Aminopropyl)-N'-[3-(nitrosoamino)propyl]-1,4-butanediamine	<chem>N(N=O)CCCNCCCCNCCCN</chem>	12.7 Pa
Alanine nitrosate	<chem>CC(N(N=O))C(=O)O</chem>	9.5 kPa	N-[2-(tetrahydro-1H-pyrrolizin-7a(5H)-yl)ethyl]aniline nitrosate	<chem>c1ccc(N(N=O)CCC23CCCN2CCC3)cc1</chem>	13.0 Pa

Neuroactivity

For this study, neuroactivities on the cholinergic signalling pathway will be used due to acetylcholine being used as a neurotransmitter in the vast majority of animal life. Both nicotinic receptors[266] and acetylcholinesterase enzymes[267][268] are found in diverse taxa, including chordates, insects, molluscs and nematodes.

Acetylcholinesterase Affinity

To predict the acetylcholinesterase affinities, 16 different models were trained on molecular fingerprints, and 3 convolutional models were trained on molecular images. For each model, data was split into train, validation, and test sets. Ideal hyperparameters were found using grid search and 5-fold cross validation. The ideal hyperparameters for the models trained using molecular fingerprints are listed below in Table 11. Solvent affinities predicted by each model were compared with each other to determine general trends.

Table 11 Hyperparameters for Fingerprint Models trained to predict Acetylcholinesterase Affinity

Model	Fingerprint Scheme	Maximum Weight (Da)	neighbours/estimators/hidden neurons	weights and distance metric /maximum depth/ layers	column subsample	row subsample	test RMSE
kNN	MACCS	300	7	1/d, manhattan	-	-	0.9308
kNN	MACCS	500	4	1/d, manhattan	-	-	1.1307
kNN	ECFP4	300	3	uniform, euclidean	-	-	1.1697

kNN	ECFP4	500	2	uniform, euclidean	-	-	1.1075
RF	MACCS	300	125	9	sqrt	-	0.9200
RF	MACCS	500	175	18	sqrt	-	1.1144
RF	ECFP4	300	200	18	sqrt	-	1.2330
RF	ECFP4	500	150	19	sqrt	-	0.9696
XGBoost	MACCS	300	100	12	0.75	0.25	0.9872
XGBoost	MACCS	500	50	12	0.5	1	1.1429
XGBoost	ECFP4	300	50	7	0.75	1	1.2562
XGBoost	ECFP4	500	50	12	1	0.75	0.9460
MLP	MACCS	300	(80,60,40)	3	-	-	0.9345
MLP	MACCS	500	(100)	1	-	-	1.1531
MLP	ECFP4	300	(80,60,20)	3	-	-	1.3139
MLP	ECFP4	500	(100,60,20)	3	-	-	1.0250

All neural networks used ReLU as the activation function and were trained using Adam optimization. The CNNs had 7 layers, with pooling layers in between convolutional layers. 5 fold cross validation was also used to find the optimal number of filters in each layer. To increase the amount of training data available, affinities of molecules with mass under 500 Daltons were used, and the molecular graphs were rotated and scaled to augment the training data. The parameters of the best performing CNNs found using grid search are listed in Table 12.

Table 12 Hyperparameters for CNNs trained to predict Acetylcholinesterase Affinity

CNN label	Convolutional Layers	Fully Connected Layers	Mean CV RMSE	Test RMSE
1	[3, 3, 5, 10, 5, 10, 10]	[5]	1.433	1.440
2	[3, 5, 3, 3, 5, 10, 10]	[10]	1.436	1.562
3	[3, 5, 3, 10, 3, 5, 5]	[10]	1.437	1.583

The three best performing CNN models were trained for 100 epochs, and used to predict the affinities of the solvents. Correlations between the results of the different models are displayed in Figure 27 in the appendix. As can be seen, the highest correlations are between models with the same type of chemical encoding, particularly between the MACCS models. In addition, the models trained with MACCS keys had a lower test root mean squared error compared to similar models trained on ECFP4 encodings. This is likely due to the fact that MACCS keys are more condensed and are curated to include structures that have been demonstrated to affect the pharmacological profile of drugs.

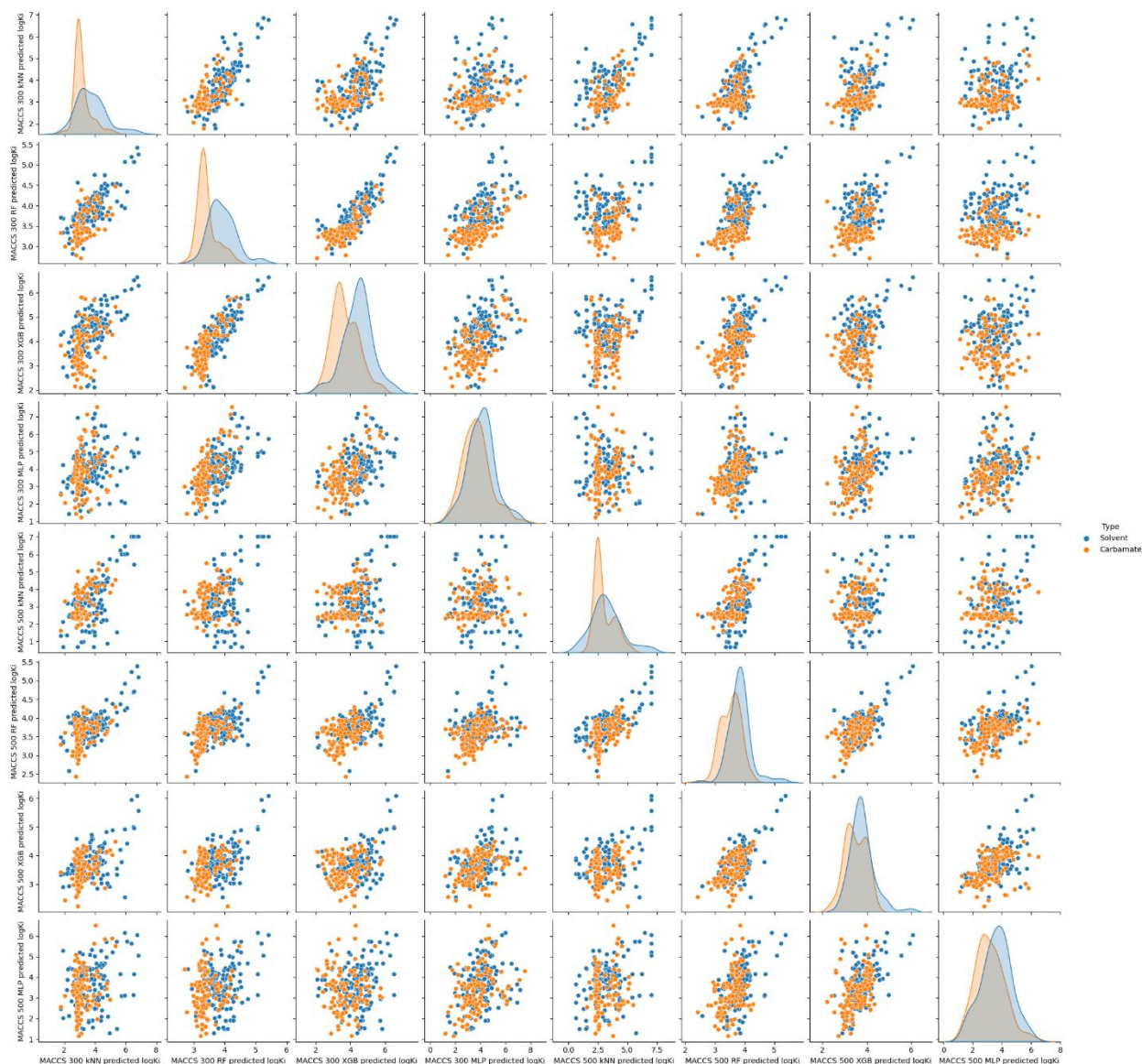


Figure 15 Pairplot of predicted solvent and carbamate AChE affinity ($\log K_i$ nM) made using MACCS key trained models

In all models, the range of $\log K_i$ values for both solvents and carbamates overlapped considerably, however, carbamates tended to have lower $\log K_i$ (Figure 15). The logarithmic difference predicted between the affinity of a solvent and its corresponding carbamate, using each model, was, on average, -0.126, meaning that the typical carbamate's affinity is 33% stronger. A histogram of all of the compiled solvent-carbamate differences from all of the used models is shown below in Figure 16.

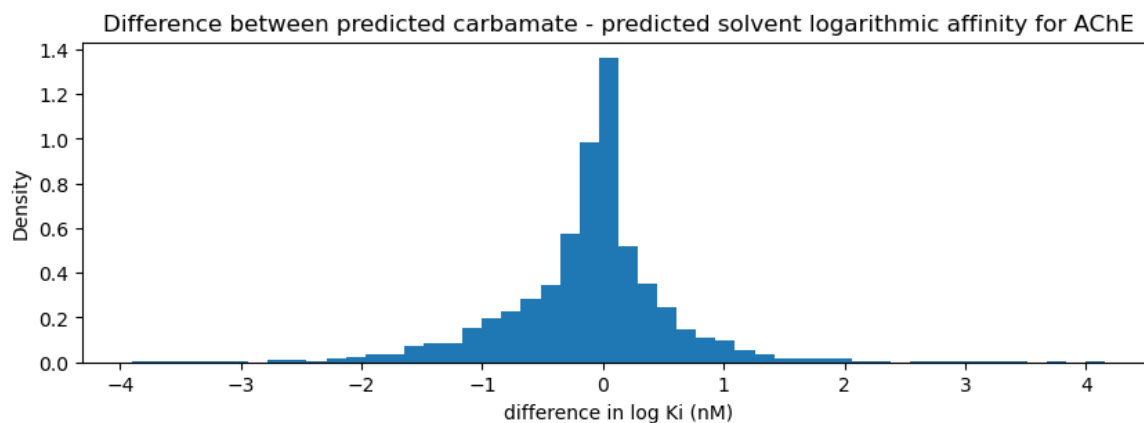


Figure 16 Histogram of the difference between the logarithmic affinity to AChE between solvents and carbamates

For both solvents and carbamates, non-cyclic amines tended to have stronger affinities to acetylcholinesterase than cyclic amines. The logarithmic mean of the affinity was 1.17 μM for non-cyclic amines and 3.02 μM for cyclic amines. A histogram of the logarithmic affinities for non-cyclic and cyclic amines is shown in Figure 17.

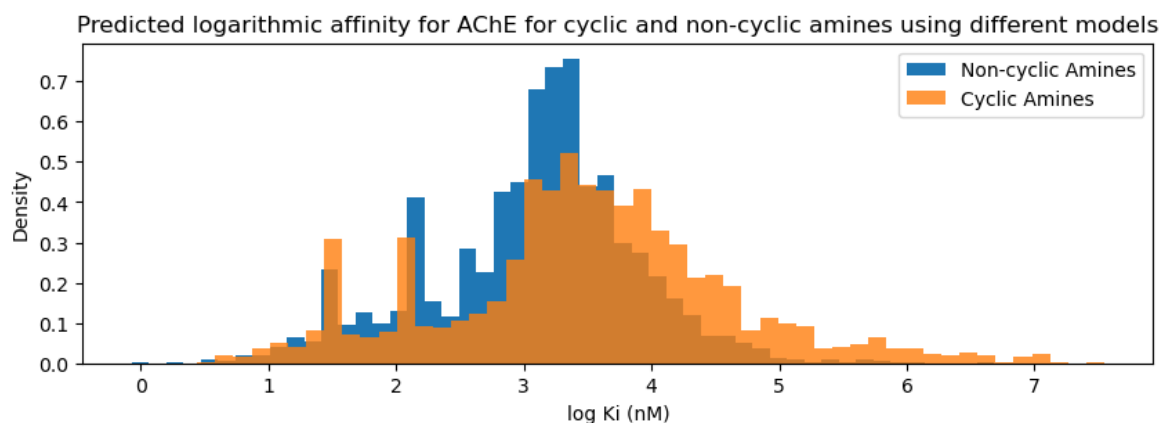


Figure 17 Histogram of the AChE affinities ($\log K_i$ nM) predicted by different models for both cyclic and non-cyclic amines

When sorted by the logarithmic mean of the predictions by different models, the 20 species with the strongest affinity were all non-cyclic and 15 were carbamates. The species with strongest \log mean predicted affinity was taurine carbamate, and then taurine, with values of 296 and 314 nM, respectively. On the other side all of the 19 chemicals with the least affinity were cyclic amines, with 15 of them being solvents. The species with the weakest \log mean predicted affinities were 4-methylimidazole and 2-methylimidazole, with affinities of 159 μM and 134 μM , respectively. Overall, non-cyclic amines with alkyl groups longer than ethyl tended to have the strongest affinities, while aromatic cyclic amines had the weakest affinities.

For every solvent, to minimize harm to wildlife, it is important to limit its concentration in the environment. However, the K_i for AChE must not be the only parameter affecting the acceptable concentration within the environment. There are other neuroactive structures within animals

that can lead to equal harm when a molecule interacts with them, such as nicotinic receptors. In addition, more lipophilic compounds are more likely to accumulate and reach higher concentrations in an organism and can enter the central nervous system more easily [269], causing more harm.

Nicotinic Affinity

For predicting the affinity of a ligand to two specific nicotinic receptors, nAChR $\alpha 4\beta 2$ and nAChR $\alpha 7$, 16 different fingerprint models were trained for each receptor. To find the ideal hyperparameters for kNN, RF, and neural network models, grid search and 5-fold cross validation were used. For XGBoost models, randomized search was used instead. Ideal parameters, and their test performance, are listed below in Table 13 and Table 14. The neural networks were activated using ReLU and were trained with Adam optimization.

Table 13 Hyperparameters for Fingerprint Models trained to predict nAChR $\alpha 4\beta 2$ affinity

Model	Fingerprint Scheme	Maximum Weight (Da)	# of neighbours/ estimators/ neurons	weights and distance metric /maximum depth/ layers	column subsample	row subsample	test RMSE
kNN	MACCS	300	13	1/d, Manhattan	-	-	1.0799
kNN	MACCS	500	7	1/d, Manhattan	-	-	0.9511
kNN	ECFP4	300	8	1/d, Manhattan	-	-	1.0526
kNN	ECFP4	500	4	1/d, Manhattan	-	-	0.9746
RF	MACCS	300	150	18	sqrt	-	1.0549
RF	MACCS	500	200	17	sqrt	-	0.8662
RF	ECFP4	300	200	19	sqrt	-	0.9739
RF	ECFP4	500	200	19	sqrt	-	0.9405
XGBoost	MACCS	300	100	11	0.25	0.75	1.0264
XGBoost	MACCS	500	175	7	0.5	1	0.8673
XGBoost	ECFP4	300	100	15	0.25	0.75	0.9612
XGBoost	ECFP4	500	75	16	0.25	1	0.8939
MLP	MACCS	300	(100,20)	2	-	-	1.0853
MLP	MACCS	500	(100,80,20)	3	-	-	1.0085
MLP	ECFP4	300	(80,60,20)	3	-	-	1.0375
MLP	ECFP4	500	(80,60,20)	3	-	-	0.9110

Table 14 Hyperparameters for Fingerprint Models trained to predict nAChR $\alpha 7$ affinity

Model	Fingerprint Scheme	Maximum Weight (Da)	# of neighbours or estimators	weights and distance metric /maximum depth/ layers	column subsample	row subsample	test RMSE
-------	--------------------	---------------------	-------------------------------	--	------------------	---------------	-----------

kNN	MACCS	300	4	1/d, Manhattan	-	-	0.9623
kNN	MACCS	500	9	1/d, Manhattan	-	-	0.8533
kNN	ECFP4	300	9	1/d, Manhattan	-	-	0.7846
kNN	ECFP4	500	8	-1/d, Manhattan	-	-	0.7743
RF	MACCS	300	125	16	log2	-	0.8617
RF	MACCS	500	200	14	sqrt	-	0.7896
RF	ECFP4	300	175	19	sqrt	-	0.7775
RF	ECFP4	500	200	19	sqrt	-	0.7570
XGBoost	MACCS	300	100	16	0.25	0.5	0.8683
XGBoost	MACCS	500	100	5	0.5	0.75	0.7853
XGBoost	ECFP4	300	100	13	1	0.25	0.7176
XGBoost	ECFP4	500	150	13	0.25	0.75	0.6865
MLP	MACCS	300	(100,40)	2	-	-	0.9297
MLP	MACCS	500	(100)	1	-	-	0.8133
MLP	ECFP4	300	(80,60,20)	3	-	-	0.7808
MLP	ECFP4	500	(80,60,20)	3	-	-	0.7907

Convolutional Neural Networks, with an input image size of 128x128 pixels, were also used to predict the affinities. Two architectures of CNNs were used, one type to solely predict a single receptor's affinity and another type to predict both affinities simultaneously. Since not all training molecules had information on both affinities, the loss function ignored the missing values for each data point. For each group, the three best performing CNNs, found using grid search with 5-fold cross validation, were selected for training, for 100 epochs, and testing. The parameters and test performances are listed below in Table 15.

Table 15 Hyperparameters for CNNs trained to predict Nicotinic Affinity

CNN label	Convolutional Layers	Fully Connected Layers	Mean CV RMSE	Test nAChR α 4 β 2 RMSE	Test nAChR α 7 RMSE
Both 1	[3, 3, 5, 3, 10, 10, 10]	[5]	1.5794	1.6815	1.4258
Both 2	[3, 3, 5, 5, 10, 10, 10]	[10]	1.5822	1.6377	1.4009
Both 3	[3, 5, 5, 5, 5, 5, 10]	[10]	1.5822	1.7196	1.4058
α 4 β 2 - 1	[3, 5, 5, 10, 3, 3, 5]	[10]	1.6925	1.7350	-
α 4 β 2 - 2	[5, 3, 10, 5, 10, 10, 10]	[5]	1.7025	1.7068	-
α 4 β 2 - 3	[5, 5, 3, 3, 10, 10, 5]	[10,5]	1.7041	1.6112	-
α 7 - 1	[5, 3, 5, 5, 10, 5, 10]	[10]	1.3749	-	1.2855
α 7 - 2	[3, 3, 3, 3, 5, 5, 10]	[10]	1.3860	-	1.5522
α 7 - 3	[3, 3, 3, 3, 10, 5, 10]	[10]	1.3935	-	1.4754

The prediction correlations were positive between two sets trained from the same input data type, fingerprints or images, and small or negative between sets trained from different input types. For fingerprint trained models, the correlations between the predictions based on a single fingerprinting scheme were stronger than correlations between fingerprinting schemes.

Models using ECFP4 fingerprints had better test performance than equivalent models using MACCS keys. This suggests that the different groups learned that different patterns are important for predicting affinity. It also emphasizes the importance of input data type on ML model performance.

For nAChR $\alpha 7$, most fingerprint models predicted that cyclic amines tend to have a stronger affinity. However, for nAChR $\alpha 4\beta 2$, this pattern was not present. All of the models predicted that the cyclic and non-cyclic amines have overlapping affinity ranges for both nicotinic receptors. The histograms of log K_i are shown below in Figure 18 and Figure 19.

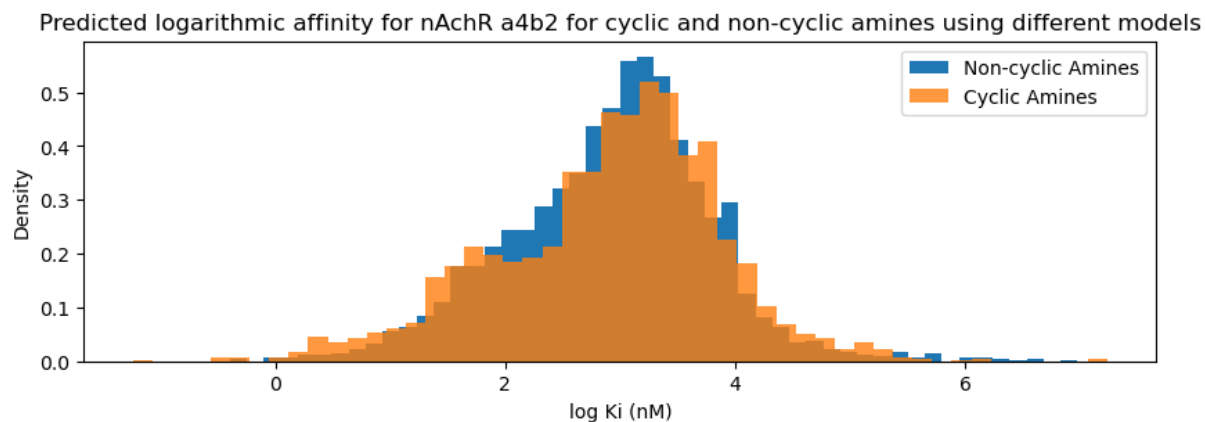


Figure 18 Histogram of the predicted logarithmic affinities (log k_i) for nAChR $\alpha 4\beta 2$

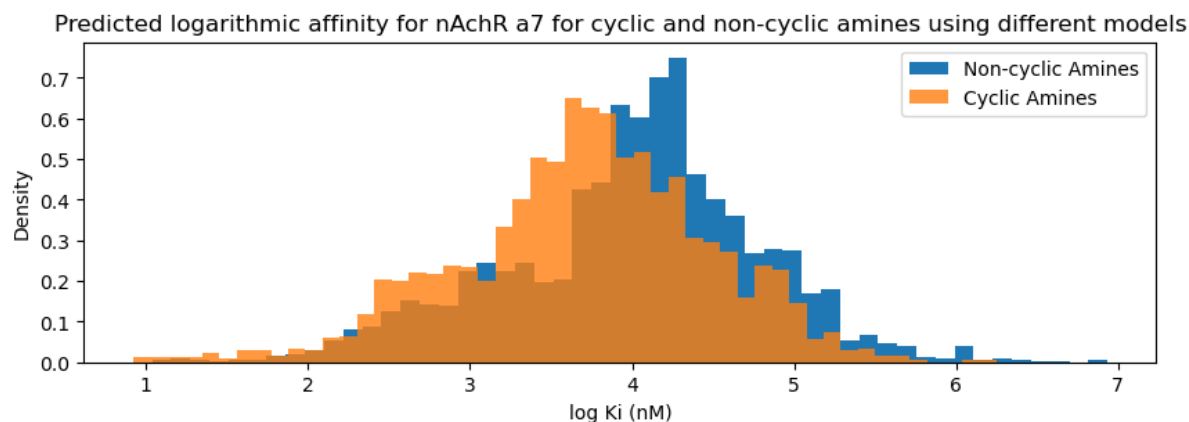


Figure 19 Histogram of the predicted logarithmic affinities (log k_i) for nAChR $\alpha 7$

The vast majority of predictions were that the carbamate of a solvent has a weaker affinity for each of the two nicotinic receptors investigated. For $\alpha 4\beta 2$ and $\alpha 7$, the carbamates had a logarithmic mean affinity that was 1.79 and 1.93 times weaker, respectively. A histogram of the logarithmic difference in affinities between a solvent and its carbamates for both $\alpha 4\beta 2$ and $\alpha 7$ is shown below in Figure 20.

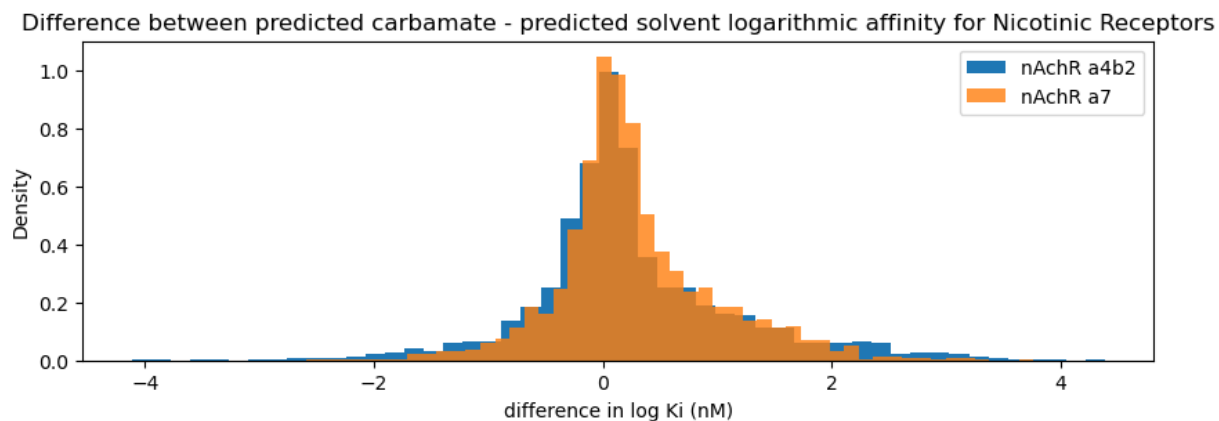


Figure 20 Histogram of the difference in nicotinic affinity for solvents and their respective carbamates

The range and distributions of the solvents and carbamates tended towards larger values of K_i , weaker affinities, compared to the distributions of the training data. This means that the species did not have the molecular substructures or properties associated with strong affinities to nicotinic receptors.

However, the tendency for species to have stronger affinities for $\alpha 4\beta 2$ than $\alpha 7$ was reflected in the predicted K_i values for the solvents and carbamates (Figure 21). Predictions for each chemical from equivalent models were subtracted from each other. The logarithmic mean and median affinities for $\alpha 4\beta 2$ were 9.73 and 9.77 times stronger, respectively, with a standard deviation of 0.89 logarithmic units, equivalent to 7.89 times. The differences in the distribution of K_i for the training data and the predictions is likely due to differences within the receptors themselves.

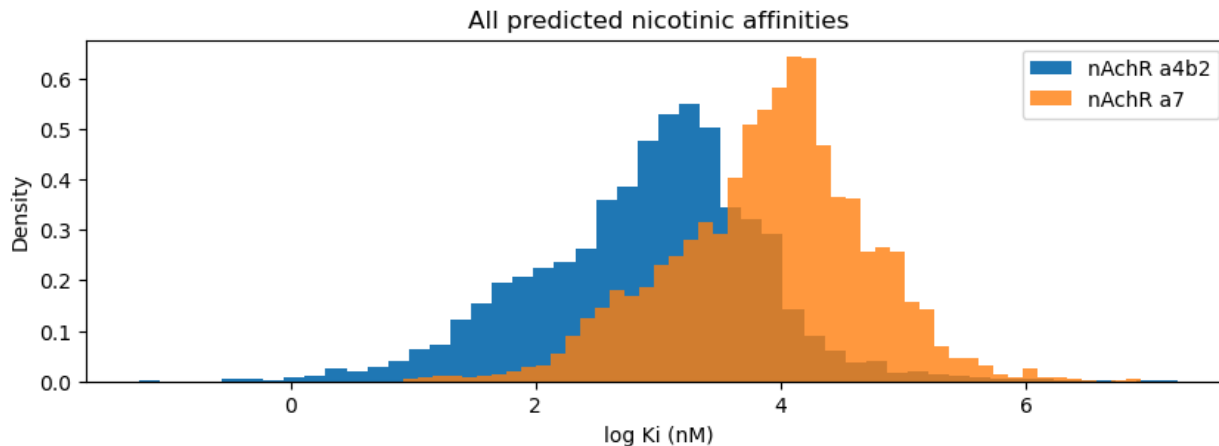


Figure 21 All predicted affinities for nAChR $\alpha 4\beta 2$ and nAChR $\alpha 7$

Both endogenous and exogenous ligands, such as nicotine [270], have stronger affinities to $\alpha 4\beta 2$ than to $\alpha 7$. Le Novere et al. [271] hypothesize that highly sensitive receptors, such as $\alpha 4\beta 2$, evolved to occupy a niche where they are exposed to lower concentrations of neurotransmitters, whereas receptors with lower sensitivity, such as $\alpha 7$, evolved to occupy spaces such as synaptic clefts, where they are exposed to higher concentrations of neurotransmitters.

Based on the mean predictions for each chemical and ML model type, fingerprint based or CNN, the chemicals with stronger predicted affinities were predominantly solvents with a cyclic amine group, and those with weaker affinities were predominantly non-cyclic amines. The chemical with the strongest predicted nAChR $\alpha 4\beta 2$ affinity, based on fingerprint and CNN models, were imidazole and N-(Tetrahydro-1H-pyrrolizin-7a(5H)-yl methyl) aniline, respectively. For nAChR $\alpha 7$ the chemicals with the strongest affinities based on fingerprint and CNN models were 1H-imidazo[4,5-b] pyridine and 1-Butyl-4-methylpyridinium, respectively.

CNN models predicted that for both receptors, the majority of the molecules with the 10 weakest affinities were amino acid carbamates. The CNNs concentrated predicted values for amino acids at the weakest end, whereas the fingerprint models scattered them throughout the range. This is evidence that the CNN models can detect carboxylate and carbamic acid groups. Another difference between the predictions of fingerprint and CNN models were that the fingerprint models predicted that the cyclic amines with lowest affinities were pyridines and imidazoles, when the CNN models predicted that the cyclic amines with lowest affinities were piperazines.

Based on the averages from all models, the compounds with the strongest $\alpha 4\beta 2$ affinity were imidazole, 1H-imidazo[4,5-b] pyridine, 2-methylimidazole, then 4 methylimidazole. For $\alpha 7$, they were 1H-imidazo[4,5-b] pyridine, 2-methylimidazole, tryptophan, then tryptophan carbamate.

Even if a substance does not have a strong affinity to nicotinic receptors, it may still hurt animals if it has a breakdown product that is a strong ligand. To avoid solvents potentially breaking down

into strong nAChR ligands, before potential utilization, amine solvents that include imidazole groups must be experimentally tested in conditions mimicking the environment and the concentrations of their breakdown products must be measured. If a significant portion of the breakdown products are potent nAChR ligands, the solvent should not be used.

Cross-Comparison of Lipophilicity and Neuroactivity

Neural tissues are predominantly composed of lipids[272]. When the predicted lipophilicities and neuroactivities, from different models, were compared to each other (Figure 22), there was no relationship between the lipophilicity and the neuroactivity. The correlation coefficients between predicted lipophilicities and log Ki for AChE, nAChR $\alpha 4\beta 2$, and nAChR $\alpha 7$ were 0.07, -0.06, and -0.03, respectively.

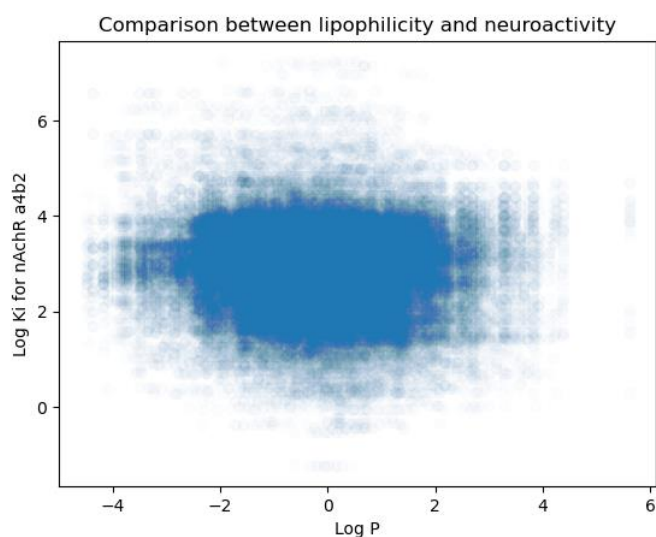


Figure 22 Comparison between lipophilicity and neuroactivity predictions for solvents and carbamates

The equilibrium concentration of a substance in fatty nervous tissue is the product of the concentration in the water-soluble phase and the lipid water partition coefficient. If a substance is lipophilic, at equilibrium, it's concentration in fatty tissues is larger than it's concentration in water. Since K_i is the concentration where a ligand occupies half of the binding site, for a binding site in fatty tissue, at equilibrium, the aqueous concentration for binding to half of the binding site is scaled by the partition coefficient. The octanol-water partition is commonly used as an estimate of the partition coefficient between biological membranes and the aqueous phase[273][274]. The aqueous concentration where a ligand binds to half of the binding site in fatty tissues is:

$$K_{aq} = \frac{K_i}{P_{ow}}$$

Equation 3 Inhibitory constant scaled by partition coefficient

$$\log K_{aq} = K_i - P_{ow}$$

As can be seen in Table 16, the solvents and carbamates that can reach a higher aqueous concentration before binding to half of acetylcholine sites in fatty tissue are predominantly amino acids. The hydrophilicity of amino acids lowered their neuroactive potential; for example, even though the log mean K_i for taurine was 314 nM, the K_{aq} was 59.3 mM.

For all binding sites, the cyclic amines had stronger binding potential than non-cyclic amines. This was the case even for acetylcholinesterase (Figure 23), where, in fat, the affinity was stronger for non-cyclic amines.

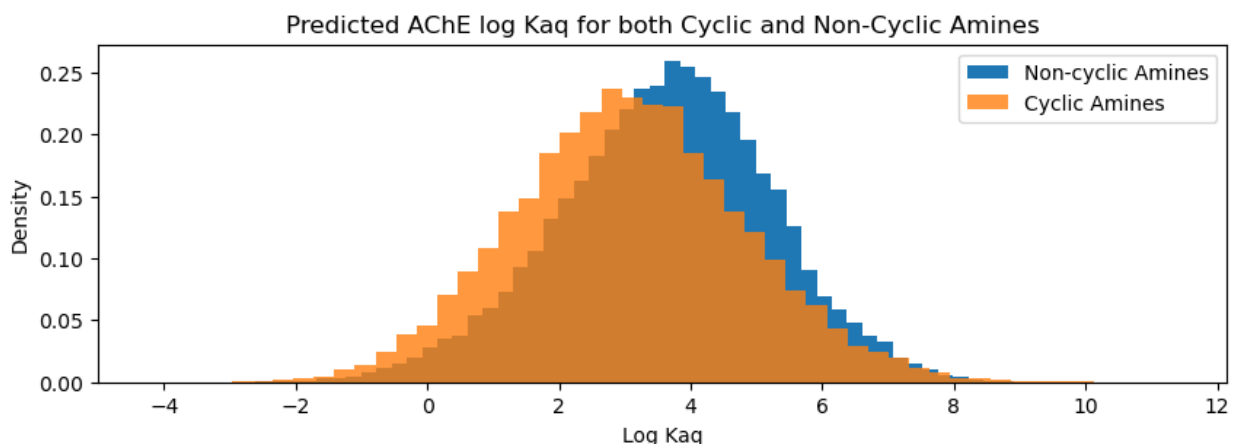


Figure 23 Log K_{aq} of solvents and carbamates for Acetylcholinesterase

For all binding sites, pyrrolizidine solvents and carbamates clustered around the strongest binding potentials. However, for Acetylcholinesterase, the next strongest binding potentials were for non-cyclic amines with strongly lipophilic groups attached, whereas for Nicotinic receptors it was for pyridines.

Table 16 Solvents predicted to be the most and the least neuroactive based on K_{aq}

AChE – Acetylcholinesterase					
Lowest K_{aq} – most neuroactive potential			Highest K_{aq} – least neuroactive potential		
Name	SMILES	Log K_{aq} prediction	Name	SMILES	Log K_{aq} prediction
Tributylamine	<chem>CCCN(CCCC)CCCC</chem>	-0.04±0.80	Arginine	<chem>N=C(N)NCCCC(N)C(=O)O</chem>	6.15±1.29
N-[2-(tetrahydro-1H-pyrrolizin-7a(5H)-yl)ethyl]aniline	<chem>c1ccc(NCCC23CCCN2CCC3)cc1</chem>	0.21±0.93	Aspartic acid	<chem>NC(CC(=O)O)C(=O)O</chem>	6.03±1.18

N-(Tetrahydro-1H-pyrrolizin-7a(5H)-ylmethyl)aniline	<chem>c1ccc(NCC23CCCN2CCC3)cc1</chem>	0.50±0.86	Arginine carbamate	<chem>N=C(N(C([O-])=O))NCCCC(N)C(=O)O</chem>	5.97±1.08
N-methylcyclohexanamine	<chem>CNC1CCCCC1</chem>	0.60±1.02	Histidine	<chem>NC(Cc1c[nH]cn1)C(=O)O</chem>	5.91±1.55
Hexylamine carbamate	<chem>CCCCCCN(C([O-])=O)</chem>	0.66±0.94	Glutamic acid	<chem>NC(CCC(=O)O)C(=O)O</chem>	5.91±1.13
nAChR α4β2 – Nicotinic Receptor					
Lowest Kaq – most neuroactive potential			Highest Kaq – least neuroactive potential		
Name	SMILES	Log Kaq prediction	Name	SMILES	Log Kaq prediction
Tributylamine	<chem>CCCCN(CCCC)CCCC</chem>	-0.74±0.97	Aspartic acid	<chem>NC(CC(=O)O)C(=O)O</chem>	6.07±1.19
N-(Tetrahydro-1H-pyrrolizin-7a(5H)-ylmethyl)aniline	<chem>c1ccc(NCC23CCCN2CCC3)cc1</chem>	-0.35±1.16	Glutamine	<chem>NC(=O)CCC(N)C(=O)O</chem>	5.97±1.32
N-[2-(tetrahydro-1H-pyrrolizin-7a(5H)-yl)ethyl]aniline	<chem>c1ccc(NCCC23CCCN2CCC3)cc1</chem>	-0.14±1.13	Asparagine	<chem>NC(=O)CC(N)C(=O)O</chem>	5.94±1.26
N-methylcyclohexanamine	<chem>CNC1CCCCC1</chem>	0.34±1.17	Arginine	<chem>N=C(N)NCCCC(N)C(=O)O</chem>	5.88±1.23
N-[2-(tetrahydro-1H-pyrrolizin-7a(5H)-yl)ethyl]aniline carbamate	<chem>c1ccc(N(C([O-])=O)CCC23CCCN2CCC3)cc1</chem>	0.44±1.09	Glutamic acid	<chem>NC(CCC(=O)O)C(=O)O</chem>	5.87±1.25
nAChR α7 – Nicotinic Receptor					
Lowest Kaq – most neuroactive potential			Highest Kaq – least neuroactive potential		
Name	SMILES	Log Kaq prediction	Name	SMILES	Log Kaq prediction
Tributylamine	<chem>CCCCN(CCCC)CCCC</chem>	0.71±1.08	Glutamine	<chem>NC(=O)CCC(N)C(=O)O</chem>	6.91±0.96
N-(Tetrahydro-1H-pyrrolizin-7a(5H)-ylmethyl)aniline	<chem>c1ccc(NCC23CCCN2CCC3)cc1</chem>	0.78±0.81	Arginine	<chem>N=C(N)NCCCC(N)C(=O)O</chem>	6.89±1.25
N-[2-(tetrahydro-1H-pyrrolizin-	<chem>c1ccc(NCCC23CCCN2CCC3)cc1</chem>	0.90±0.84	Aspartic acid	<chem>NC(CC(=O)O)C(=O)O</chem>	6.87±1.03

7a(5H)-yl)ethyl]aniline					
N-[2-(tetrahydro-1H-pyrrolizin-7a(5H)-yl)ethyl]aniline carbamate	<chem>c1ccc(N(C([O-])=O)CCC23CCCN2CCC3)cc1</chem>	1.22±0.99	Arginine carbamate	<chem>N=C(N(C([O-])=O))NCCCC(N)C(=O)O</chem>	6.84±1.15
N-(Tetrahydro-1H-pyrrolizin-7a(5H)-ylmethyl)aniline carbamate	<chem>c1ccc(N(C([O-])=O)CC23CCCN2CCC3)cc1</chem>	1.25±0.82	Asparagine	<chem>NC(=O)CC(N)C(=O)O</chem>	6.80±1.03

Conclusion

Industrial activities such as the combustion of fossil fuels for energy release carbon dioxide, as a byproduct, into the atmosphere. This has caused the CO₂ concentration in the atmosphere to rise from 280 ppm in the pre-industrial era to over 400 ppm. The increased quantity of this greenhouse gas has caused shifts in the global climate. To mitigate environmental disasters, it is urgent to reduce the rate at which CO₂ is released into the atmosphere.

One way to decarbonize industrial activities, such as power and heat generation, cement production, and steelmaking, is to capture the carbon dioxide before it is emitted. Existing infrastructure can be retrofitted with post-combustion carbon capture units. To minimize the amount of waste material that has to be handled, for storage, CO₂ should be separated from other gaseous constituents of the gas phase. The most mature method of gas separation is the use of amine absorbents.

Different amines have been proposed for use in absorption based post-combustion carbon capture. However, some of these amines, when they escape into the environment, may have deleterious effects. The goal of this study was to compare the potential environmental toxicity of 151 different amine solvents. Machine learning models were trained to predict four properties: volatility, lipophilicity, mutagenicity, and neuroactivity

Volatility indicates how readily amines can escape from the aqueous phase to the gaseous phase. Since the vapor pressure of solute is proportional to its concentration, the vapor pressures of pure substances was used to compare the relative volatilities of different amines. Machine learning models were trained on vapor pressures of pure liquid substances at different temperatures, and predictions were made for solvents and carbamates at temperatures from 5°C to 95°C. The presence of a carbamate group or any other polar group decreased the vapor pressure at any particular temperature. Overall, the most volatile solvents were composed of small molecules, and the least volatile solvents were composed of larger molecules.

Lipophilicity measures the equilibrium concentration of a solute in the lipid phase instead of in the aqueous phase. Lipophilic substances tend to be sequestered in lipids, such as cell membranes, and bioaccumulate. Toxicity which is not apparent in the short term can become visible in the long term. The most lipophilic compounds were those that contained a terminal alkyl chain larger than butyl, cyclohexyl, or phenyl. Amino acids were the least lipophilic solvents.

Two measures were used to predict potential mutagenicity, Ames mutagenicity, and the volatility of nitramines and nitrosamines. For the Ames test, multi-amines were classified as mutagens by more models. For the volatility of mutagenic nitramines and nitrosamines, nitrosamines, which are more mutagenic, had a higher predicted vapor pressure. Similar to the results of the vapor pressure predictions, heavier solvents were predicted to have less volatile nitrosamines.

For neuroactivity, the cholinergic signalling pathway was looked at, due to its central role in a vast array of animal species. Models were trained on the affinities of small organic models to acetylcholinesterase, and two nicotinic receptors, nAChR $\alpha 4\beta 2$ and nAChR $\alpha 7$. Non-cyclic amines had a stronger affinity to acetylcholinesterase than cyclic amines. In contrast, cyclic amines had a stronger affinity to nicotinic receptors than non-cyclic amines. Imidazoles had weak affinities for acetylcholinesterase, but the strongest affinities to nicotinic receptors. The solvents with the weakest nicotinic affinity were amino acids.

Since neural tissues are mainly composed of lipids, ligands that are lipophilic have a larger impact than similar affinity ligands that are not lipophilic. Instead of K_i , the concentration of ligand in a medium needed to bind to half of the receptors, K_{aq} was used, the equilibrium concentration in the aqueous phase needed to bind to half of the receptors in the lipid phase. This value was calculated using K_i and lipophilicity. For all receptors, the solvents with the largest K_{aq} , or the lowest potential neuroactivity were amino acids.

Overall, when released into the environment, due to their hydrophilicity, amino acids have the least potential to be toxic to living organisms. Further research should focus on comparing amino acids to determine which is the most feasible to use in carbon capture units. This includes testing the loading capacity and regeneration energy requirements.

References

- [1] G. Clark and D. Jacks, "Coal and the industrial revolution, 1700–1869," *European Review of Economic History*, vol. 11, (1), pp. 39–72, 2007. .
- [2] D. J. Hofmann, J. H. Butler and P. P. Tans, "A new look at atmospheric carbon dioxide," *Atmos. Environ.*, vol. 43, (12), pp. 2084–2086, 2009. .
- [3] H. Ritchie, M. Roser and P. Rosado, "CO₂ and greenhouse gas emissions," *Our World in Data*, 2020. .
- [4] M. D. Jones and A. Henderson-Sellers, "History of the greenhouse effect," *Prog. Phys. Geogr.*, vol. 14, (1), pp. 1–18, 1990. .
- [5] Q. Rayer, P. Pfleiderer and K. Haustein, "Global warming and extreme weather investment risks," *Ecological, Societal, and Technological Risks and the Financial Sector*, pp. 39–68, 2020. .
- [6] T. K. Matthews, R. L. Wilby and C. Murphy, "Communicating the deadly consequences of global warming for human heat stress," *Proceedings of the National Academy of Sciences*, vol. 114, (15), pp. 3861–3866, 2017. .
- [7] R. L. Peters, "Effects of global warming on forests," *For. Ecol. Manage.*, vol. 35, (1-2), pp. 13–33, 1990. .
- [8] L. Hughes, "Biological consequences of global warming: is the signal already apparent?" *Trends in Ecology & Evolution*, vol. 15, (2), pp. 56–61, 2000. .
- [9] P. R. Liu and A. E. Raftery, "Country-based rate of emissions reductions should increase by 80% beyond nationally determined contributions to meet the 2 C target," *Communications Earth & Environment*, vol. 2, (1), pp. 29, 2021. .
- [10] IEA, "Key world energy statistics 2021," Paris, 2021.
- [11] A. R. Ansari, "CFD Analysis of Aerodynamic Design of Tata Indica Car," *International Journal of Mechanical Engineering and Technology*, vol. 8, (3), pp. 344–355, 2017. .
- [12] C. Schmutzler *et al*, "Energy efficiency in automotive networks: Assessment and concepts," in *2010 International Conference on High Performance Computing & Simulation*, 2010, .
- [13] . *Electricity Mix* . Available: <https://ourworldindata.org/electricity-mix>.
- [14] M. D. Mathew, "Nuclear energy: A pathway towards mitigation of global warming," *Prog. Nuclear Energy*, vol. 143, pp. 104080, 2022. .

- [15] . *Nuclear power*. Available: <https://www.opg.com/powering-ontario/our-generation/nuclear/>.
- [16] Y. Zhou *et al*, "Uranium recovery from sandstone-type uranium deposit by acid in-situ leaching-an example from the Kujieertai," *Hydrometallurgy*, vol. 191, pp. 105209, 2020. .
- [17] R. Wufuer *et al*, "Bioremediation of Uranium-and Nitrate-Contaminated Groundwater after the In Situ Leach Mining of Uranium," *Water*, vol. 13, (22), pp. 3188, 2021. .
- [18] Z. Lu and Z. Liu, "Pollution characteristics and risk assessment of uranium and heavy metals of agricultural soil around the uranium tailing reservoir in Southern China," *J. Radioanal. Nucl.*, vol. 318, (2), pp. 923–933, 2018. .
- [19] M. V. Ramana, "Technical and social problems of nuclear waste," *Wiley Interdisciplinary Reviews: Energy and Environment*, vol. 7, (4), pp. e289, 2018. .
- [20] G. Steinhauser, A. Brandl and T. E. Johnson, "Comparison of the Chernobyl and Fukushima nuclear accidents: a review of the environmental impacts," *Sci. Total Environ.*, vol. 470, pp. 800–817, 2014. .
- [21] . *Hydropower*. Available: https://www.ctg.com.cn/en/our_business/hydropower88/index.html.
- [22] J. Wu *et al*, "Three-Gorges dam--experiment in habitat fragmentation?" *Science*, vol. 300, (5623), pp. 1239–1240, 2003. .
- [23] P. Xie, "Three-Gorges Dam: risk to ancient fish," *Science*, vol. 302, (5648), pp. 1149–1151, 2003. .
- [24] G. L. Volpato *et al*, "Fish ladders select fish traits on migration--still a growing problem for natural fish populations," *Mar. Freshwat. Behav. Physiol.*, vol. 42, (5), pp. 307–313, 2009. .
- [25] L. Bambace *et al*, "Mitigation and recovery of methane emissions from tropical hydroelectric dams," *Energy*, vol. 32, (6), pp. 1038–1046, 2007. .
- [26] A. Kemenes, B. R. Forsberg and J. M. Melack, "Methane release below a tropical hydroelectric dam," *Geophys. Res. Lett.*, vol. 34, (12), 2007. .
- [27] S. Li *et al*, "Methane and CO₂ emissions from China's hydroelectric reservoirs: a new quantitative synthesis," *Environmental Science and Pollution Research*, vol. 22, pp. 5325–5339, 2015. .
- [28] J. Giles, "Methane quashes green credentials of hydropower," *Nature*, vol. 444, (7119), pp. 524–526, 2006. .

- [29] Energy Institute, "Statistical Review of World Energy," 2023. .
- [30] P. Higgins and A. Foley, "The evolution of offshore wind power in the United Kingdom," *Renewable and Sustainable Energy Reviews*, vol. 37, pp. 599–612, 2014. .
- [31] D. Micallef and A. Rezaeiha, "Floating offshore wind turbine aerodynamics: Trends and future challenges," *Renewable and Sustainable Energy Reviews*, vol. 152, pp. 111696, 2021. .
- [32] H. Demolli *et al*, "Wind power forecasting based on daily wind speed data using machine learning algorithms," *Energy Conversion and Management*, vol. 198, pp. 111823, 2019. .
- [33] Z. Huang and Z. S. Chalabi, "Use of time-series analysis to model and forecast wind speed," *J. Wind Eng. Ind. Aerodyn.*, vol. 56, (2-3), pp. 311–322, 1995. .
- [34] J. D. Farmer and F. Lafond, "How predictable is technological progress?" *Research Policy*, vol. 45, (3), pp. 647–665, 2016. .
- [35] C. Government of Canada, "NEB—Provincial and Territorial Energy Profiles—Ontario," .
- [36] S. Yan *et al*, "Comparative analysis of CO₂ separation from flue gas by membrane gas absorption technology and chemical absorption technology in China," *Energy Conversion and Management*, vol. 49, (11), pp. 3188–3197, 2008. .
- [37] X. Fan *et al*, "Establishment of refined sintering flue gas recirculation patterns for gas pollutant reduction and waste heat recycling," *J. Clean. Prod.*, vol. 235, pp. 1549–1558, 2019. .
- [38] Á A. Ramírez-Santos, C. Castel and E. Favre, "Utilization of blast furnace flue gas: opportunities and challenges for polymeric membrane gas separation processes," *J. Membr. Sci.*, vol. 526, pp. 191–204, 2017. .
- [39] A. Bosoaga, O. Masek and J. E. Oakey, "CO₂ capture technologies for cement industry," *Energy Procedia*, vol. 1, (1), pp. 133–140, 2009. .
- [40] J. A. Lara-Gil, C. Senés-Guerrero and A. Pacheco, "Cement flue gas as a potential source of nutrients during CO₂ mitigation by microalgae," *Algal Research*, vol. 17, pp. 285–292, 2016. .
- [41] L. Dion, M. Lefsrud and V. Orsat, "Review of CO₂ recovery methods from the exhaust gas of biomass heating systems for safe enrichment in greenhouses," *Biomass Bioenergy*, vol. 35, (8), pp. 3422–3432, 2011. .
- [42] Y. Jin *et al*, "Supercritical CO₂ extraction of essential oils from *Chamaecyparis obtusa*," *Natural Product Communications*, vol. 5, (3), pp. 1934578X1000500324, 2010. .

- [43] I. Zizovic *et al*, "Supercritical carbon dioxide extraction of essential oils from plants with secretory ducts: Mathematical modelling on the micro-scale," *The Journal of Supercritical Fluids*, vol. 39, (3), pp. 338–346, 2007. .
- [44] S. Dabral and T. Schaub, "The use of carbon dioxide (CO₂) as a building block in organic synthesis from an industrial perspective," *Advanced Synthesis & Catalysis*, vol. 361, (2), pp. 223–246, 2019. .
- [45] Y. Zhang, G. Wu and D. J. Darensbourg, "CO₂-based block copolymers: present and future designs," *Trends in Chemistry*, vol. 2, (8), pp. 750–763, 2020. .
- [46] W. Seifritz, "CO₂ disposal by means of silicates," *Nature*, vol. 345, (6275), pp. 486, 1990. .
- [47] Z. Liu and J. Zhao, "Contribution of carbonate rock weathering to the atmospheric CO₂ sink," *Environ. Geol.*, vol. 39, pp. 1053–1058, 2000. .
- [48] A. Sanna *et al*, "A review of mineral carbonation technologies to sequester CO₂," *Chem. Soc. Rev.*, vol. 43, (23), pp. 8049–8080, 2014. .
- [49] P. Kelemen *et al*, "An overview of the status and challenges of CO₂ storage in minerals and geological formations," *Frontiers in Climate*, vol. 1, pp. 482595, 2019. .
- [50] National Academies of Sciences *et al*, "Negative emissions technologies and reliable sequestration: a research agenda," 2019. .
- [51] National Academies of Sciences, Engineering, and Medicine, "sequestration of supercritical CO₂ in deep sedimentary geological formations," *Negative Emissions Technologies and Reliable Sequestration: A Research Agenda*, pp. 273–281, 2019. .
- [52] Z. Tao and A. Clarens, "Estimating the carbon sequestration capacity of shale formations using methane production rates," *Environ. Sci. Technol.*, vol. 47, (19), pp. 11318–11325, 2013. .
- [53] M. Godec *et al*, "Enhanced gas recovery and CO₂ storage in gas shales: a summary review of its status and potential," *Energy Procedia*, vol. 63, pp. 5849–5857, 2014. .
- [54] E. Mohagheghian, H. Hassanzadeh and Z. Chen, "CO₂ sequestration coupled with enhanced gas recovery in shale gas reservoirs," *Journal of CO₂ Utilization*, vol. 34, pp. 646–655, 2019. .
- [55] Y. Artanto *et al*, "Performance of MEA and amine-blends in the CSIRO PCC pilot plant at Loy Yang Power in Australia," *Fuel*, vol. 101, pp. 264–275, 2012. .
- [56] C. Song *et al*, "Tri-reforming of methane over Ni catalysts for CO₂ conversion to syngas with desired H₂/CO ratios using flue gas of power plants without CO₂ separation," *Stud.Surf.Sci.Catal*, vol. 153, pp. 315–322, 2004. .

- [57] N. Otsuka, "1.18 - fireside corrosion," in *Shreir's Corrosion*, B. Cottis *et al*, Ed. 2010, Available: <https://www.sciencedirect.com/science/article/pii/B978044452787500192X>. DOI: <https://doi.org/10.1016/B978-044452787-5.00192-X>.
- [58] R. Zevenhoven and P. Kilpinen, "Control of pollutants in flue gases and fuel gases," 2001. Available: <https://www.osti.gov/etdeweb/biblio/20238538>.
- [59] H. Sijbesma *et al*, "Flue gas dehydration using polymer membranes," *J. Membr. Sci.*, vol. 313, (1-2), pp. 263–276, 2008. .
- [60] J. Pérez-Calvo *et al*, "A methodology for the heuristic optimization of solvent-based CO₂ capture processes when applied to new flue gas compositions: A case study of the Chilled Ammonia Process for capture in cement plants," *Chemical Engineering Science: X*, vol. 8, pp. 100074, 2020. .
- [61] U. Arachchige, M. Mohsin and M. C. Melaen, "Optimization of post combustion carbon capture process-solvent selection," *International Journal of Energy and Environment*, vol. 3, (6), pp. 861–870, 2012. .
- [62] G. Vunjak-Novakovic *et al*, "Air-Lift Bioreactors for Algal Growth on Flue Gas: Mathematical Modeling and Pilot-Plant Studies," *Ind Eng Chem Res*, vol. 44, (16), pp. 6154–6163, 2005. . DOI: 10.1021/ie049099z.
- [63] X. Xu *et al*, "Adsorption separation of carbon dioxide from flue gas of natural gas-fired boiler by a novel nanoporous "molecular basket" adsorbent," *Fuel Process Technol*, vol. 86, (14), pp. 1457–1472, 2005. Available: <https://www.sciencedirect.com.proxy.lib.uwaterloo.ca/science/article/pii/S0378382005000159>. DOI: <https://doi-org.proxy.lib.uwaterloo.ca/10.1016/j.fuproc.2005.01.002>.
- [64] U. Desideri *et al*, "MCFC-based CO₂ capture system for small scale CHP plants," *Int J Hydrogen Energy*, vol. 37, (24), pp. 19295–19303, 2012. .
- [65] B. Rajh *et al*, "Advanced modelling and testing of a 13 MW.sub.th waste wood-fired grate boiler with recycled flue gas," *Energy Conversion and Management*, vol. 125, pp. 230, 2016. . DOI: 10.1016/j.enconman.2016.02.036.
- [66] A. I. Calvo *et al*, "Characterization of operating conditions of two residential wood combustion appliances," *Fuel Process Technol*, vol. 126, pp. 222–232, 2014. .
- [67] Á A. Ramírez-Santos, C. Castel and E. Favre, "Utilization of blast furnace flue gas: opportunities and challenges for polymeric membrane gas separation processes," *J. Membr. Sci.*, vol. 526, pp. 191–204, 2017. .

- [68] M. T. Ho, A. Bustamante and D. E. Wiley, "Comparison of CO₂ capture economics for iron and steel mills," *International Journal of Greenhouse Gas Control*, vol. 19, pp. 145–159, 2013. .
- [69] Linde. . *Air Separation Plants: History and Technological Progress in the Course of Time*. Available: https://www.linde-engineering.com/en/images/Air-separation-plants-history-and-technological-progress-2019_tcm19-457349.pdf.
- [70] D. Clodic and M. Younes, "A new method for CO₂ capture: Frosting CO₂ at atmospheric pressure," in *Greenhouse Gas Control Technologies-6th International Conference*, 2003, .
- [71] C. Font-Palma, D. Cann and C. Udemu, "Review of cryogenic carbon capture innovations and their potential applications," *C*, vol. 7, (3), pp. 58, 2021. .
- [72] D. Berstad, R. Anantharaman and P. Neksa, "Low-temperature CO₂ capture technologies—Applications and potential," *Int. J. Refrig.*, vol. 36, (5), pp. 1403–1416, 2013. .
- [73] P. Willson *et al*, "Evaluation of the performance and economic viability of a novel low temperature carbon capture process," *International Journal of Greenhouse Gas Control*, vol. 86, pp. 1–9, 2019. .
- [74] M. J. Jensen *et al*, "Prediction and validation of external cooling loop cryogenic carbon capture (CCC-ECL) for full-scale coal-fired power plant retrofit," *International Journal of Greenhouse Gas Control*, vol. 42, pp. 200–212, 2015. .
- [75] G. L. Weibel and C. K. Ober, "An overview of supercritical CO₂ applications in microelectronics processing," *Microelectronic Engineering*, vol. 65, (1-2), pp. 145–152, 2003. .
- [76] G. Brunner, "Supercritical fluids: technology and application to food processing," *J. Food Eng.*, vol. 67, (1-2), pp. 21–33, 2005. .
- [77] M. J. Tuinier *et al*, "Cryogenic CO₂ capture using dynamically operated packed beds," *Chemical Engineering Science*, vol. 65, (1), pp. 114–119, 2010. .
- [78] J. Hilz *et al*, "Long-term pilot testing of the carbonate looping process in 1 MWth scale," *Fuel*, vol. 210, pp. 892–899, 2017. .
- [79] A. Martínez *et al*, "Energy penalty reduction in the calcium looping cycle," *International Journal of Greenhouse Gas Control*, vol. 7, pp. 74–81, 2012. .
- [80] C. Cormos, "Techno-economic assessment of calcium and magnesium-based sorbents for post-combustion CO₂ capture applied in fossil-fueled power plants," *Fuel*, vol. 298, pp. 120794, 2021. .

- [81] J. Blamey *et al*, "The calcium looping cycle for large-scale CO₂ capture," *Progress in Energy and Combustion Science*, vol. 36, (2), pp. 260–279, 2010. .
- [82] C. Chou and C. Chen, "Carbon dioxide recovery by vacuum swing adsorption," *Separation and Purification Technology*, vol. 39, (1), pp. 51–65, 2004. . DOI: 10.1016/j.seppur.2003.12.009.
- [83] L. Jiang *et al*, "Comparative analysis on temperature swing adsorption cycle for carbon capture by using internal heat/mass recovery," *Appl. Therm. Eng.*, vol. 169, pp. 114973, 2020. .
- [84] M. Bui *et al*, "Carbon capture and storage (CCS): the way forward," *Energy & Environmental Science*, vol. 11, (5), pp. 1062–1176, 2018. .
- [85] A. E. Ogungbenro *et al*, "Activated carbon from date seeds for CO₂ capture applications," *Energy Procedia*, vol. 114, pp. 2313–2321, 2017. .
- [86] M. Song *et al*, "The pyrolysis of multi-component municipal solid waste in fixed bed reactor for activated carbon production," *J. Anal. Appl. Pyrolysis*, vol. 109, pp. 278–282, 2014. .
- [87] M. Idrees, V. Rangari and S. Jeelani, "Sustainable packaging waste-derived activated carbon for carbon dioxide capture," *Journal of CO₂ Utilization*, vol. 26, pp. 380–387, 2018. .
- [88] X. Tang *et al*, "Comparison and application of different component municipal solid wastes based carbon on adsorption of carbon dioxide," *International Journal of Green Energy*, vol. 14, (2), pp. 135–140, 2017. .
- [89] H. Zhou *et al*, "The application of adsorption to remove aromatic hydrocarbons from flue gas," in *2010 4th International Conference on Bioinformatics and Biomedical Engineering*, 2010, .
- [90] M. Keramati and A. A. Ghoreyshi, "Improving CO₂ adsorption onto activated carbon through functionalization by chitosan and triethylenetetramine," *Physica E: Low-Dimensional Systems and Nanostructures*, vol. 57, pp. 161–168, 2014. .
- [91] R. Ben-Mansour *et al*, "Carbon capture by physical adsorption: Materials, experimental investigations and numerical modeling and simulations – A review," *Appl. Energy*, vol. 161, pp. 225–255, 2016. Available:
<https://www.sciencedirect.com/science/article/pii/S0306261915012386>. DOI:
<https://doi.org/10.1016/j.apenergy.2015.10.011>.
- [92] N. Prasetya *et al*, "A review on emerging organic-containing microporous material membranes for carbon capture and separation," *Chem. Eng. J.*, vol. 391, pp. 123575, 2020. .
- [93] J. Xiao and J. Wei, "Diffusion mechanism of hydrocarbons in zeolites—I. Theory," *Chemical Engineering Science*, vol. 47, (5), pp. 1123–1141, 1992. . DOI: 10.1016/0009-2509(92)80236-6.

- [94] E. Nagy, "Chapter 18—Membrane Gas Separation," *Basic Equations of Mass Transport through a Membrane Layer, 2nd Ed.*; Nagy E., Ed, pp. 457–481, 2019. .
- [95] M. Rao and S. Sircar, "Nanoporous carbon membranes for separation of gas mixtures by selective surface flow," *J. Membr. Sci.*, vol. 85, (3), pp. 253–264, 1993. .
- [96] Y. Han and W. W. Ho, "Recent advances in polymeric membranes for CO₂ capture," *Chin. J. Chem. Eng.*, vol. 26, (11), pp. 2238–2254, 2018. .
- [97] P. M. Budd *et al*, "Gas separation membranes from polymers of intrinsic microporosity," *J. Membr. Sci.*, vol. 251, (1-2), pp. 263–269, 2005. .
- [98] Y. Han, D. Wu and W. S. W. Ho, "Nanotube-reinforced facilitated transport membrane for CO₂/N₂ separation with vacuum operation," *J. Membr. Sci.*, vol. 567, pp. 261–271, 2018. . DOI: 10.1016/j.memsci.2018.08.061.
- [99] R. Casadei *et al*, "Polyvinylamine membranes containing graphene-based nanofillers for carbon capture applications," *Membranes*, vol. 9, (9), pp. 119, 2019. .
- [100] G. He *et al*, "High-permeance polymer-functionalized single-layer graphene membranes that surpass the postcombustion carbon capture target," *Energy & Environmental Science*, vol. 12, (11), pp. 3305–3312, 2019. .
- [101] M. Liu *et al*, "Postcombustion Carbon Capture Using Thin-Film Composite Membranes," *Accounts Chem Res*, vol. 52, (7), pp. 1905–1914, 2019. . DOI: 10.1021/acs.accounts.9b00111.
- [102] J. K. Stolaroff *et al*, "Microencapsulation of advanced solvents for carbon capture," *Faraday Discuss.*, vol. 192, pp. 271–281, 2016. .
- [103] L. M. Robeson, "The upper bound revisited," *J. Membr. Sci.*, vol. 320, (1-2), pp. 390–400, 2008. .
- [104] C. Song *et al*, "Cryogenic-based CO₂ capture technologies: State-of-the-art developments and current challenges," *Renewable & Sustainable Energy Reviews*, vol. 101, pp. 265, 2019. . DOI: 10.1016/j.rser.2018.11.018.
- [105] R. R. Bottoms, "Separating acidic gases," *US1783901*, 1930. .
- [106] P. Mores, N. Scenna and S. Mussati, "CO₂ capture using monoethanolamine (MEA) aqueous solution: Modeling and optimization of the solvent regeneration and CO₂ desorption process," *Energy*, vol. 45, (1), pp. 1042–1058, 2012. .

- [107] F. Ghiasi *et al*, "Machine learning models for absorption-based post-combustion carbon capture," in *Carbon Capture, Utilization, and Storage Technologies: Towards More Sustainable Cities* Anonymous 2024, .
- [108] P. Luis, "Use of monoethanolamine (MEA) for CO₂ capture in a global scenario: Consequences and alternatives," *Desalination*, vol. 380, pp. 93–99, 2016. .
- [109] K. Li *et al*, "Systematic study of aqueous monoethanolamine (MEA)-based CO₂ capture process: Techno-economic assessment of the MEA process and its improvements," *Appl. Energy*, vol. 165, pp. 648–659, 2016. .
- [110] T. Spietz *et al*, "Reduction of amines emission and their volatile degradation products," *Chemik*, vol. 69, (10), pp. 625–634, 2015. .
- [111] A. S. Al-Dughaiter, A. A. Ibrahim and W. A. Al-Masry, "Investigating droplet separation efficiency in wire-mesh mist eliminators in bubble column," *Journal of Saudi Chemical Society*, vol. 14, (4), pp. 331–339, 2010. .
- [112] J. Mertens *et al*, "A wet electrostatic precipitator (WESP) as countermeasure to mist formation in amine based carbon capture," *International Journal of Greenhouse Gas Control*, vol. 31, pp. 175–181, 2014. .
- [113] K. Zoannou, D. J. Sapsford and A. J. Griffiths, "Thermal degradation of monoethanolamine and its effect on CO₂ capture capacity," *International Journal of Greenhouse Gas Control*, vol. 17, pp. 423–430, 2013. .
- [114] K. B. Fischer *et al*, "MEA and piperazine corrosion of carbon steel and stainless steel," *Energy Procedia*, vol. 114, pp. 1751–1764, 2017. .
- [115] G. T. Rochelle, "Thermal degradation of amines for CO₂ capture," *Current Opinion in Chemical Engineering*, vol. 1, (2), pp. 183–190, 2012. .
- [116] S. YAMADA *et al*, "Effect of piperidine and related alicyclic amines on nicotinic and muscarinic agonist binding sites in the mammalian brain," *J. Pharmacobio-Dyn.*, vol. 9, (7), pp. 620–625, 1986. .
- [117] Y. Kase *et al*, "Pharmacological studies on alicyclic amines (II) Central actions of piperidine, pyrrolidine and piperazine," *The Japanese Journal of Pharmacology*, vol. 19, (2), pp. 300–314, 1969. .
- [118] C. J. Nielsen *et al*, "Atmospheric chemistry of 2-aminoethanol (MEA)," *Energy Procedia*, vol. 4, pp. 2245–2252, 2011. .

- [119] S. A. Mazari, P. Alaba and I. M. Saeed, "Formation and elimination of nitrosamines and nitramines in freshwaters involved in post-combustion carbon capture process," *Journal of Environmental Chemical Engineering*, vol. 7, (3), pp. 103111, 2019. .
- [120] B. D. Reh *et al*, "O 6-methylguanine DNA adducts associated with occupational nitrosamine exposure," *Carcinogenesis*, vol. 21, (1), pp. 29–33, 2000. .
- [121] F. Gholami *et al*, "Technologies for the nitrogen oxides reduction from flue gas: A review," *Sci. Total Environ.*, vol. 714, pp. 136712, 2020. .
- [122] T. Spietz *et al*, "Nitrosamines and nitramines in Carbon Capture plants," *Environmental Protection and Natural Resources*, vol. 28, (4), pp. 43–50, 2017. .
- [123] C. J. Nielsen, H. Herrmann and C. Weller, "Atmospheric chemistry and environmental impact of the use of amines in carbon capture and storage (CCS)," *Chem. Soc. Rev.*, vol. 41, (19), pp. 6684–6704, 2012. .
- [124] E. Gjernes, L. I. Helgesen and Y. Maree, "Health and environmental impact of amine based post combustion CO₂ capture," *Energy Procedia*, vol. 37, pp. 735–742, 2013. .
- [125] F. Arab and C. N. Mulligan, "An eco-friendly method for heavy metal removal from mine tailings," *Environmental Science and Pollution Research*, vol. 25, pp. 16202–16216, 2018. .
- [126] Y. Gong *et al*, "A review of oil, dispersed oil and sediment interactions in the aquatic environment: influence on the fate, transport and remediation of oil spills," *Mar. Pollut. Bull.*, vol. 79, (1-2), pp. 16–33, 2014. .
- [127] R. C. Prince, "Oil spill dispersants: boon or bane?" *Environ. Sci. Technol.*, vol. 49, (11), pp. 6376–6384, 2015. .
- [128] I. Eide-Haugmo *et al*, "Environmental impact of amines," *Energy Procedia*, vol. 1, (1), pp. 1297–1304, 2009. .
- [129] G. Chen, "Xenobiotic metabolism and disposition," in *An Introduction to Interdisciplinary Toxicology* Anonymous 2020, .
- [130] S. Manzetti, E. R. van der Spoel and D. van der Spoel, "Chemical properties, environmental fate, and degradation of seven classes of pollutants," *Chem. Res. Toxicol.*, vol. 27, (5), pp. 713–737, 2014. .
- [131] H. Goerke *et al*, "Increasing levels and biomagnification of persistent organic pollutants (POPs) in Antarctic biota," *Mar. Pollut. Bull.*, vol. 48, (3-4), pp. 295–302, 2004. .

- [132] K. Han *et al*, "Current status and challenges of the ammonia-based CO₂ capture technologies toward commercialization," *International Journal of Greenhouse Gas Control*, vol. 14, pp. 270–281, 2013. .
- [133] B. Lv *et al*, "Mechanisms of CO₂ capture into monoethanolamine solution with different CO₂ loading during the absorption/desorption processes," *Environ. Sci. Technol.*, vol. 49, (17), pp. 10728–10735, 2015. .
- [134] F. A. Chowdhury *et al*, "CO₂ capture by tertiary amine absorbents: a performance comparison study," *Ind Eng Chem Res*, vol. 52, (24), pp. 8323–8331, 2013. .
- [135] L. Li *et al*, "Amine blends using concentrated piperazine," *Energy Procedia*, vol. 37, pp. 353–369, 2013. .
- [136] I. M. Saeed *et al*, "Opportunities and challenges in the development of monoethanolamine and its blends for post-combustion CO₂ capture," *International Journal of Greenhouse Gas Control*, vol. 79, pp. 212–233, 2018. .
- [137] S. A. Mazari *et al*, "Thermal degradation kinetics of morpholine for carbon dioxide capture," *Journal of Environmental Chemical Engineering*, vol. 8, (3), pp. 103814, 2020. .
- [138] M. Nainar and A. Veawab, "Corrosion in CO₂ Capture Process Using Blended Monoethanolamine and Piperazine," *Industrial & Engineering Chemistry Research.*, vol. 48, (20), pp. 9299–9306, 2009. . DOI: 10.1021/ie801802a.
- [139] B. H. Hansen *et al*, "Molecular effects of diethanolamine exposure on *Calanus finmarchicus* (Crustacea: Copepoda)," *Aquatic Toxicology*, vol. 99, (2), pp. 212–222, 2010. .
- [140] S. Xu *et al*, "Kinetics of the reaction of carbon dioxide with 2-amino-2-methyl-1-propanol solutions," *Chemical Engineering Science*, vol. 51, (6), pp. 841–850, 1996. .
- [141] S. Y. W. Chai, L. H. Ngu and B. S. How, "Review of carbon capture absorbents for CO₂ utilization," *Greenhouse Gases: Science and Technology*, vol. 12, (3), pp. 394–427, 2022. .
- [142] Z. Xu, S. Wang and C. Chen, "CO₂ absorption by biphasic solvents: Mixtures of 1, 4-Butanediamine and 2-(Diethylamino)-ethanol," *International Journal of Greenhouse Gas Control*, vol. 16, pp. 107–115, 2013. .
- [143] P. Muchan *et al*, "Assessment of the relationship between degradation and emission activities of carbon capture amines based on their chemical structures," in *Proceedings of the 15th Greenhouse Gas Control Technologies Conference, 2021*, .
- [144] R. J. Hook, "An investigation of some sterically hindered amines as potential carbon dioxide scrubbing compounds," *Ind Eng Chem Res*, vol. 36, (5), pp. 1779–1790, 1997. .

- [145] D. Guo *et al*, "Amino acids as carbon capture solvents: chemical kinetics and mechanism of the glycine CO₂ reaction," *Energy Fuels*, vol. 27, (7), pp. 3898–3904, 2013. .
- [146] P. S. Kumar *et al*, "Kinetics of the reaction of CO₂ with aqueous potassium salt of taurine and glycine," *AIChE J.*, vol. 49, (1), pp. 203–213, 2003. .
- [147] X. Wang *et al*, "Amino acid-functionalized ionic liquid solid sorbents for post-combustion carbon capture," *ACS Applied Materials & Interfaces*, vol. 5, (17), pp. 8670–8677, 2013. .
- [148] Z. Zhang *et al*, "Effectiveness of amino acid salt solutions in capturing CO₂: A review," *Renewable and Sustainable Energy Reviews*, vol. 98, pp. 179–188, 2018. .
- [149] G. Hu *et al*, "Carbon dioxide absorption into promoted potassium carbonate solutions: A review," *International Journal of Greenhouse Gas Control*, vol. 53, pp. 28–40, 2016. .
- [150] E. S. Fernandez and E. L. Goetheer, "DECAB: Process development of a phase change absorption process," *Energy Procedia*, vol. 4, pp. 868–875, 2011. .
- [151] E. Sanchez-Fernandez *et al*, "Analysis of Process Configurations for CO₂ Capture by Precipitating Amino Acid Solvents," *Industrial & Engineering Chemistry Research.*, vol. 53, (6), pp. 2348–2361, 2014. . DOI: 10.1021/ie402323r.
- [152] K. Fukumoto, M. Yoshizawa and H. Ohno, "Room temperature ionic liquids from 20 natural amino acids," *J. Am. Chem. Soc.*, vol. 127, (8), pp. 2398–2399, 2005. .
- [153] B. Jaffary, L. Jaafari and R. Idem, "CO₂ capture performance comparisons of polyamines at practical concentrations for use as activators for methyldiethanolamine for natural gas sweetening," *Energy Fuels*, vol. 35, (9), pp. 8081–8094, 2021. .
- [154] L. Jaafari, B. Jaffary and R. Idem, "Screening study for selecting new activators for activating MDEA for natural gas sweetening," *Separation and Purification Technology*, vol. 199, pp. 320–330, 2018. .
- [155] E. Alper, "Kinetics of reactions of carbon dioxide with diglycolamine and morpholine," *The Chemical Engineering Journal.*, vol. 44, (2), pp. 107–111, 1990. . DOI: 10.1016/0300-9467(90)80063-I.
- [156] C. Wang *et al*, "Highly efficient CO₂ capture by tunable alkanolamine-based ionic liquids with multidentate cation coordination," *Chemical Communications.*, vol. 48, (52), pp. 6526, 2012. . DOI: 10.1039/C2CC32365F.
- [157] J. Zhang, "Study on CO₂ Capture Using Thermomorphic Biphasic Solvents with Energy-Efficient Regeneration," *Study on CO₂ Capture using Thermomorphic Biphasic Solvents with Energy Efficient Regeneration*, 2014. .

- [158] Y. Shen *et al*, "Two-stage interaction performance of CO₂ absorption into biphasic solvents: Mechanism analysis, quantum calculation and energy consumption," *Applied Energy*, vol. 260, pp. 114343, 2020. . DOI: 10.1016/j.apenergy.2019.114343.
- [159] G. Shavaliyeva *et al*, "Environmental, health and safety assessment of post-combustion CO₂ capture processes with phase-change solvents," *Sustainable Production and Consumption*, vol. 25, pp. 60–76, 2021. .
- [160] R. Vijayraghavan *et al*, "Diamino protic ionic liquids for CO₂ capture," *Physical Chemistry Chemical Physics* ;, vol. 15, (46), pp. 19994, 2013. . DOI: 10.1039/C3CP54082K.
- [161] G. Rochelle *et al*, "Aqueous piperazine as the new standard for CO₂ capture technology," *Chem. Eng. J.*, vol. 171, (3), pp. 725–733, 2011. .
- [162] A. V. Rayer *et al*, "High-pressure solubility of carbon dioxide (CO₂) in aqueous 1-methyl piperazine solution," *Journal of Chemical & Engineering Data*, vol. 59, (11), pp. 3610–3623, 2014. .
- [163] P. Luis, "Use of monoethanolamine (MEA) for CO₂ capture in a global scenario: Consequences and alternatives," *Desalination*, vol. 380, pp. 93–99, 2016. .
- [164] A. A. Orlov *et al*, "Computational screening methodology identifies effective solvents for CO₂ capture," *Communications Chemistry*, vol. 5, (1), pp. 37, 2022. .
- [165] S. Balchandani and R. Singh, "COSMO-RS Analysis of CO₂ Solubility in N-Methyldiethanolamine, Sulfolane, and 1-Butyl-3-methyl-imidazolium Acetate Activated by 2-Methylpiperazine for Postcombustion Carbon Capture," *ACS Omega*, vol. 6, (1), pp. 747–761, 2020. .
- [166] K. Akinpelumi, C. Saha and G. T. Rochelle, "Piperazine aerosol mitigation for post-combustion carbon capture," *International Journal of Greenhouse Gas Control*, vol. 91, pp. 102845, 2019. .
- [167] R. J. Martin, "Modes of action of anthelmintic drugs," *The Veterinary Journal*, vol. 154, (1), pp. 11–34, 1997. .
- [168] R. F. Squires and E. Saederup, "Mono N-aryl ethylenediamine and piperazine derivatives are GABA A receptor blockers: Implications for psychiatry," *Neurochem. Res.*, vol. 18, pp. 787–793, 1993. .
- [169] T. H. Askary and M. Abd-Elgawad, "Beneficial nematodes in agroecosystems: A global perspective." in *Biocontrol Agents: Entomopathogenic and Slug Parasitic Nematodes* Anonymous 2017, .

- [170] R. J. Martin, "γ-Aminobutyric acid-and piperazine-activated single-channel currents from *Ascaris suum* body muscle," *Br. J. Pharmacol.*, vol. 84, (2), pp. 445–461, 1985. .
- [171] Y. Coulier *et al*, "Thermodynamic modeling and experimental study of CO₂ dissolution in new absorbents for post-combustion CO₂ capture processes," *ACS Sustainable Chemistry & Engineering*, vol. 6, (1), pp. 918–926, 2018. .
- [172] L. Raynal *et al*, "The DMX™ process: an original solution for lowering the cost of post-combustion carbon capture," *Energy Procedia*, vol. 4, pp. 779–786, 2011. .
- [173] S. YAMADA *et al*, "Effect of piperidine and related alicyclic amines on nicotinic and muscarinic agonist binding sites in the mammalian brain," *J. Pharmacobio-Dyn.*, vol. 9, (7), pp. 620–625, 1986. .
- [174] N. S. Millar and I. Denholm, "Nicotinic acetylcholine receptors: targets for commercially important insecticides," *Invertebr. Neurosci.*, vol. 7, pp. 53–66, 2007. .
- [175] K. A. N. P. Bandara *et al*, "Insecticidal piperidine alkaloid from *Microcos paniculata* stem bark," *Phytochemistry*, vol. 54, (1), pp. 29–32, 2000. Available: <https://www.sciencedirect.com/science/article/pii/S003194220000025X>. DOI: 10.1016/S0031-9422(00)00025-X.
- [176] C. Ding *et al*, "Synthesis and insecticidal activity of novel piperidine thiazole compounds," *Chinese Journal of Organic Chemistry*, vol. 39, (3), pp. 836, 2019. .
- [177] B. T. Green *et al*, "Piperidine alkaloids: human and food animal teratogens," *Food and Chemical Toxicology*, vol. 50, (6), pp. 2049–2055, 2012. .
- [178] T. Miyata *et al*, "Effects of intracerebral administration of piperidine on EEG and behavior," *Life Sciences.*, vol. 15, (6), pp. 1135–1152, 1974. . DOI: 10.1016/S0024-3205(74)80010-X.
- [179] A. García-Abuín *et al*, "CO₂ capture by pyrrolidine: reaction mechanism and mass transfer," *AIChE J.*, vol. 60, (3), pp. 1098–1106, 2014. .
- [180] E. Skylogianni *et al*, "Carbon Capture Demonstration at Irving Oil Whitegate Refinery," Available at SSRN 4277476, 2022. .
- [181] J. M. Hanusch *et al*, "Pyrrolizidines for direct air capture and CO₂ conversion," *Chemical Communications*, vol. 55, (7), pp. 949–952, 2019. .
- [182] H. Wiedenfeld and J. Edgar, "Toxicity of pyrrolizidine alkaloids to humans and ruminants," *Phytochemistry Reviews*, vol. 10, pp. 137–151, 2011. .

- [183] S. Abdelfatah *et al*, "Pyrrolizidine alkaloids cause cell cycle and DNA damage repair defects as analyzed by transcriptomics in cytochrome P450 3A4-overexpressing HepG2 clone 9 cells," *Cell Biol. Toxicol.*, pp. 1–21, 2021. .
- [184] T. R. Rajalakshmi *et al*, "DNA adducts-chemical addons." *Journal of Pharmacy & Bioallied Sciences*, vol. 7, 2015. .
- [185] X. Zhang *et al*, "Carbon capture with ionic liquids: overview and progress," *Energy & Environmental Science*, vol. 5, (5), pp. 6668–6681, 2012. .
- [186] S. Carda–Broch, A. Berthod and D. W. Armstrong, "Solvent properties of the 1-butyl-3-methylimidazolium hexafluorophosphate ionic liquid," *Analytical and Bioanalytical Chemistry*, vol. 375, pp. 191–199, 2003. .
- [187] G. Cui, J. Wang and S. Zhang, "Active chemisorption sites in functionalized ionic liquids for carbon capture," *Chemical Society Reviews.*, vol. 45, (15), pp. 4307–4339, 2016. . DOI: 10.1039/C5CS00462D.
- [188] W. D. Amith, J. C. Araque and C. J. Margulis, "Ether tails make a large difference for the structural dynamics of imidazolium-based ionic liquids," *Journal of Ionic Liquids*, vol. 2, (1), pp. 100012, 2022. .
- [189] P. Stepnowski, W. Mroziak and J. Nichthauser, "Adsorption of alkylimidazolium and alkylpyridinium ionic liquids onto natural soils," *Environ. Sci. Technol.*, vol. 41, (2), pp. 511–516, 2007. .
- [190] A. Oskarsson and M. C. Wright, "Ionic Liquids: New Emerging Pollutants, Similarities with Perfluorinated Alkyl Substances (PFASs)," *Environmental Science & Technology*, vol. 53, (18), pp. 10539–10541, 2019. Available: <http://dx.doi.org/10.1021/acs.est.9b04778>. DOI: 10.1021/acs.est.9b04778.
- [191] T. P. T. Pham, C. Cho and Y. Yun, "Environmental fate and toxicity of ionic liquids: a review," *Water Res.*, vol. 44, (2), pp. 352–372, 2010. .
- [192] P. Stepnowski and A. Zaleska, "Comparison of different advanced oxidation processes for the degradation of room temperature ionic liquids," *J. Photochem. Photobiol. A.*, vol. 170, (1), pp. 45–50, 2005. .
- [193] J. Ranke *et al*, "Design of sustainable chemical products the example of ionic liquids," *Chem. Rev.*, vol. 107, (6), pp. 2183–2206, 2007. .
- [194] A. R. Shaikh *et al*, "Selective absorption of H₂S and CO₂ by azole based protic ionic liquids: A combined density functional theory and molecular dynamics study," *Journal of Molecular Liquids*, vol. 367, pp. 120558, 2022. .

- [195] K. Tomizaki *et al*, "Heats of reaction and vapor– liquid equilibria of novel chemical absorbents for absorption/recovery of pressurized carbon dioxide in integrated coal gasification combined cycle– carbon capture and storage process," *Ind Eng Chem Res*, vol. 49, (3), pp. 1214–1221, 2010. .
- [196] N. Dai and W. A. Mitch, "Effects of flue gas compositions on nitrosamine and nitramine formation in postcombustion CO₂ capture systems," *Environ. Sci. Technol.*, vol. 48, (13), pp. 7519–7526, 2014. .
- [197] K. Liu *et al*, "Absorption of Carbon Dioxide in Aqueous Morpholine Solutions," *Industrial & Engineering Chemistry Research.*, vol. 52, (45), pp. 15932–15938, 2013. . DOI: 10.1021/ie402570u.
- [198] Z. Chen, "A review of pre-combustion carbon capture technology," in *2022 7th International Conference on Social Sciences and Economic Development (ICSSSED 2022)*, 2022, .
- [199] X. Luo *et al*, "Significant Improvements in CO₂ Capture by Pyridine-Containing Anion-Functionalized Ionic Liquids through Multiple-Site Cooperative Interactions," *Angewandte Chemie.*, vol. 126, (27), pp. 7173–7177, 2014. . DOI: 10.1002/ange.201400957.
- [200] G. Cui *et al*, "Tuning the structure of pyridinolite-based functional ionic liquids for highly efficient SO₂ absorption," *Fuel*, vol. 303, pp. 121311, 2021. . DOI: 10.1016/j.fuel.2021.121311.
- [201] S. P. Kelley *et al*, "Structural analysis of mono-substituted N-butyl-pyridinium salts: in search of ionic liquids," *Journal of Coordination Chemistry*, vol. 74, (1-3), pp. 117–128, 2021. .
- [202] S. Stolte *et al*, "Effects of different head groups and functionalised side chains on the aquatic toxicity of ionic liquids," *Green Chem.*, vol. 9, (11), pp. 1170–1179, 2007. .
- [203] M. Karl *et al*, "Worst case scenario study to assess the environmental impact of amine emissions from a CO₂ capture plant," *International Journal of Greenhouse Gas Control*, vol. 5, (3), pp. 439–447, 2011. .
- [204] L. Lepori, E. Matteoli and P. Gianni, "Vapor pressure and its temperature dependence of 28 organic compounds: cyclic amines, cyclic ethers, and cyclic and open chain secondary alcohols," *Journal of Chemical & Engineering Data*, vol. 62, (1), pp. 194–203, 2017. .
- [205] M. Covarrubias-Cervantes *et al*, "Saturated vapour pressure of aroma compounds at various temperatures," *Food Chem.*, vol. 85, (2), pp. 221–229, 2004. .
- [206] D. R. Lide, *CRC Handbook of Chemistry and Physics*. 200485.

- [207] I. Mokbel *et al*, "Experimental vapor pressures of 2-phenylethylamine, benzylamine, triethylamine, and cis-2, 6-dimethylpiperidine in the range between 0.2 Pa and 75 kPa," *Journal of Chemical & Engineering Data*, vol. 54, (3), pp. 819–822, 2009. .
- [208] S. P. Verevkin *et al*, "Vapor pressures and enthalpies of vaporization of a series of low-volatile alkanolamines," *Journal of Chemical & Engineering Data*, vol. 56, (12), pp. 4400–4406, 2011. .
- [209] K. Klepáčová *et al*, "Vapor pressures of several commercially used alkanolamines," *Journal of Chemical & Engineering Data*, vol. 56, (5), pp. 2242–2248, 2011. .
- [210] B. P. Soares *et al*, "Vapor pressures and thermophysical properties of selected ethanolamines," *Fluid Phase Equilib.*, vol. 473, pp. 245–254, 2018. .
- [211] S. P. Verevkin and Y. Chernyak, "Vapor pressure and enthalpy of vaporization of aliphatic propanediamines," *The Journal of Chemical Thermodynamics*, vol. 47, pp. 328–334, 2012. .
- [212] M. Thornton *et al*, "The vaporization enthalpy and vapor pressure of (d)-amphetamine and of several primary amines used as standards at T/K= 298 as evaluated by correlation gas chromatography and transpiration," *Journal of Chemical & Engineering Data*, vol. 58, (7), pp. 2018–2027, 2013. .
- [213] E. D. Vieira, J. P. Torres and O. Malm, "DDT environmental persistence from its use in a vector control program: a case study," *Environ. Res.*, vol. 86, (2), pp. 174–182, 2001. .
- [214] M. La Merrill *et al*, "Toxicological function of adipose tissue: focus on persistent organic pollutants," *Environ. Health Perspect.*, vol. 121, (2), pp. 162–169, 2013. .
- [215] S. Martel *et al*, "Large, chemically diverse dataset of log P measurements for benchmarking studies," *European Journal of Pharmaceutical Sciences*, vol. 48, (1-2), pp. 21–29, 2013. .
- [216] J. Sangster, "Octanol-water partition coefficients of simple organic compounds," *Journal of Physical and Chemical Reference Data*, vol. 18, (3), pp. 1111–1229, 1989. .
- [217] K. Mansouri *et al*, "An automated curation procedure for addressing chemical errors and inconsistencies in public datasets used in QSAR modelling," *SAR QSAR Environ. Res.*, vol. 27, (11), pp. 911–937, 2016. .
- [218] (Aug 5). *Ames Mutagenicity Dataset for Multi-Task Learning*. Available: <https://data.mendeley.com/datasets/ktc6gbfsbh>. DOI: 10.17632/ktc6gbfsbh.2.

- [219] A. G. Karczmar, J. Lindstrom and A. Christopoulos, "History of research on nicotinic and muscarinic cholinergic receptors," *Exploring the Vertebrate Central Cholinergic Nervous System*, pp. 151–162, 2007. .
- [220] M. K. Gilson *et al*, "BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology," *Nucleic Acids Res.*, vol. 44, (D1), pp. D1045–D1053, 2016. .
- [221] C. N. Cavasotto and V. Scardino, "Machine Learning Toxicity Prediction: Latest Advances by Toxicity End Point," *ACS Omega*, 2022. .
- [222] . *SCM - Accelerate your chemistry & materials research*. Available: <https://www.scm.com/>.
- [223] M. Orio, D. A. Pantazis and F. Neese, "Density functional theory," *Photosynthesis Res.*, vol. 102, pp. 443–453, 2009. .
- [224] A. Klamt and F. Eckert, "COSMO-RS: a novel and efficient method for the a priori prediction of thermophysical data of liquids," *Fluid Phase Equilib.*, vol. 172, (1), pp. 43–72, 2000. .
- [225] S. Lin *et al*, "Prediction of vapor pressures and enthalpies of vaporization using a COSMO solvation model," *The Journal of Physical Chemistry A*, vol. 108, (36), pp. 7429–7439, 2004. .
- [226] B. Schröder, M. Fulem and M. A. Martins, "Vapor pressure predictions of multi-functional oxygen-containing organic compounds with COSMO-RS," *Atmos. Environ.*, vol. 133, pp. 135–144, 2016. .
- [227] J. Warnau, K. Wichmann and J. Reinisch, "COSMO-RS predictions of logP in the SAMPL7 blind challenge," *J. Comput. Aided Mol. Des.*, vol. 35, (7), pp. 813–818, 2021. .
- [228] E. Olsen and F. Nielsen, "Predicting Vapour Pressures of Organic Compounds from Their Chemical Structure for Classification According to the VOC Directive and Risk Assessment in General." *Molecules*, vol. 6, (4), pp. 370–389, 2001. .
- [229] F. Gharagheizi *et al*, "Determination of vapor pressure of chemical compounds: a group contribution model for an extremely large database," *Ind Eng Chem Res*, vol. 51, (20), pp. 7119–7125, 2012. .
- [230] S. Prasad and B. R. Brooks, "A deep learning approach for the blind logP prediction in SAMPL6 challenge," *J. Comput. Aided Mol. Des.*, vol. 34, pp. 535–542, 2020. .
- [231] C. Isert *et al*, "Machine Learning for Fast, Quantum Mechanics-Based Approximation of Drug Lipophilicity," *ACS Omega*, 2023. .

- [232] M. J. Martínez *et al*, "Multitask Deep Neural Networks for Ames Mutagenicity Prediction," *Journal of Chemical Information and Modeling*, vol. 62, (24), pp. 6342–6351, 2022. .
- [233] H. Moriwaki *et al*, "Mordred: a molecular descriptor calculator," *Journal of Cheminformatics*, vol. 10, (1), pp. 1–14, 2018. .
- [234] R. Benigni, "Towards quantitative read across: Prediction of Ames mutagenicity in a large database," *Regulatory Toxicology and Pharmacology*, vol. 108, pp. 104434, 2019. .
- [235] S. D. Dimitrov *et al*, "QSAR Toolbox–workflow and major functionalities," *SAR QSAR Environ. Res.*, vol. 27, (3), pp. 203–219, 2016. .
- [236] B. Sharma *et al*, "Accurate clinical toxicity prediction using multi-task deep neural nets and contrastive molecular explanations," *Scientific Reports*, vol. 13, (1), pp. 4908, 2023. .
- [237] R. Huang *et al*, "Tox21Challenge to build predictive models of nuclear receptor and stress response pathways as mediated by exposure to environmental chemicals and drugs," *Frontiers in Environmental Science*, vol. 3, pp. 85, 2016. .
- [238] Z. Wu *et al*, "MoleculeNet: a benchmark for molecular machine learning," *Chemical Science*, vol. 9, (2), pp. 513–530, 2018. .
- [239] D. V. Sweet, V. P. Anderson and J. Fang, "An overview of the Registry of Toxic Effects of Chemical Substances (RTECS): Critical information on chemical hazards," *Chemical Health and Safety*, vol. 6, (6), pp. 12–16, 1999. .
- [240] G. B. Goh *et al*, "Chemception: a deep neural network with minimal chemistry knowledge matches the performance of expert-developed QSAR/QSPR models," *arXiv Preprint arXiv:1706.06689*, 2017. .
- [241] G. B. Goh *et al*, "How much chemistry does a deep neural network need to know to make accurate predictions?" in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2018, .
- [242] T. Shi *et al*, "Molecular image-based convolutional neural network for the prediction of ADMET properties," *Chemometrics Intellig. Lab. Syst.*, vol. 194, pp. 103853, 2019. .
- [243] M. Fernandez *et al*, "Toxic colors: the use of deep learning for predicting toxicity of compounds merely from their graphic images," *Journal of Chemical Information and Modeling*, vol. 58, (8), pp. 1533–1543, 2018. .
- [244] R. Huang *et al*, "Tox21Challenge to build predictive models of nuclear receptor and stress response pathways as mediated by exposure to environmental chemicals and drugs," *Frontiers in Environmental Science*, vol. 3, pp. 85, 2016. .

- [245] I. Wallach, M. Dzamba and A. Heifets, "AtomNet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery," *arXiv Preprint arXiv:1510.02855*, 2015. .
- [246] D. K. Duvenaud *et al*, "Convolutional networks on graphs for learning molecular fingerprints," *Advances in Neural Information Processing Systems*, vol. 28, 2015. .
- [247] N. Ulrich, K. Goss and A. Ebert, "Exploring the octanol–water partition coefficient dataset using deep learning techniques and data augmentation," *Communications Chemistry*, vol. 4, (1), pp. 90, 2021. .
- [248] I. V. Tetko *et al*, "Virtual computational chemistry laboratory–design and description," *J. Comput. Aided Mol. Des.*, vol. 19, pp. 453–463, 2005. .
- [249] U. EPA, "Estimation programs interface suite™ for Microsoft® windows, v 4.11," *United States Environmental Protection Agency, Washington, DC, USA*, 2012. .
- [250] J. Chen *et al*, "Chemical toxicity prediction based on semi-supervised learning and graph convolutional neural network," *Journal of Cheminformatics*, vol. 13, (1), pp. 1–16, 2021. .
- [251] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv Preprint arXiv:1609.02907*, 2016. .
- [252] (March 1). *RDKit: Open-Source Cheminformatics and Machine Learning*. Available: rdkit.org.
- [253] D. Rogers and M. Hahn, "Extended-connectivity fingerprints," *Journal of Chemical Information and Modeling*, vol. 50, (5), pp. 742–754, 2010. .
- [254] A. Cereto-Massagué *et al*, "Molecular fingerprint similarity search in virtual screening," *Methods*, vol. 71, pp. 58–63, 2015. .
- [255] Y. LeCun *et al*, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, (4), pp. 541–551, 1989. .
- [256] S. S. Dhaliwal, A. Nahid and R. Abbas, "Effective intrusion detection system using XGBoost," *Information*, vol. 9, (7), pp. 149, 2018. .
- [257] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 2016, .
- [258] G. TensorFlow, "Large-scale machine learning on heterogeneous systems," *Google Research*, vol. 10, pp. s15326985ep4001, 2015. .

- [259] S. Velasco, F. L. Román and J. A. White, "On the Clausius–Clapeyron vapor pressure equation," *J. Chem. Educ.*, vol. 86, (1), pp. 106, 2009. .
- [260] N. Kwak *et al*, "A study of the CO₂ capture pilot plant by amine absorption," *Energy*, vol. 47, (1), pp. 41–46, 2012. .
- [261] S. A. Wildman and G. M. Crippen, "Prediction of physicochemical parameters by atomic contributions," *J. Chem. Inf. Comput. Sci.*, vol. 39, (5), pp. 868–873, 1999. .
- [262] P. Gee *et al*, "Comparison of responses of base-specific Salmonella tester strains with the traditional strains for identifying mutagens: the results of a validation study," *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, vol. 412, (2), pp. 115–130, 1998. .
- [263] D. E. Levin *et al*, "A new Salmonella tester strain (TA102) with AXT base pairs at the site of mutation detects oxidative mutagens." *Proceedings of the National Academy of Sciences*, vol. 79, (23), pp. 7445–7449, 1982. .
- [264] M. J. Prival and E. Zeiger, "Chemicals mutagenic in Salmonella typhimurium strain TA1535 but not in TA100," *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, vol. 412, (3), pp. 251–260, 1998. .
- [265] E. D. Wagner *et al*, "Comparative in vitro toxicity of nitrosamines and nitramines associated with amine-based carbon capture and storage," *Environ. Sci. Technol.*, vol. 48, (14), pp. 8203–8211, 2014. .
- [266] Y. Jiao *et al*, "Massive expansion and diversity of nicotinic acetylcholine receptors in lophotrochozoans," *BMC Genomics*, vol. 20, pp. 1–15, 2019. .
- [267] L. Pezzementi and A. Chatonnet, "Evolution of cholinesterases in the animal kingdom," *Chem. Biol. Interact.*, vol. 187, (1-3), pp. 27–33, 2010. .
- [268] Y. H. Kim and S. H. Lee, "Which acetylcholinesterase functions as the main catalytic enzyme in the Class Insecta?" *Insect Biochem. Mol. Biol.*, vol. 43, (1), pp. 47–53, 2013. .
- [269] B. Testa *et al*, "The influence of lipophilicity on the pharmacokinetic behavior of drugs: Concepts and examples," *Perspect. Drug Discov. Des.*, vol. 19, pp. 179–211, 2000. .
- [270] A. S. Lewis and M. R. Picciotto, "High-affinity nicotinic acetylcholine receptor expression and trafficking abnormalities in psychiatric illness," *Psychopharmacology (Berl.)*, vol. 229, pp. 477–485, 2013. .

[271] N. Le Novère, P. Corringier and J. Changeux, "The diversity of subunit composition in nAChRs: evolutionary origins, physiologic and pharmacologic consequences," *J. Neurobiol.*, vol. 53, (4), pp. 447–456, 2002. .

[272] C. Chang, D. Ke and J. Chen, "Essential fatty acids and human brain," *Acta Neurol Taiwan*, vol. 18, (4), pp. 231–241, 2009. .

[273] M. Moo-Young, *Comprehensive Biotechnology*. 2019.

[274] W. R. Roy, "Environmental impact of solvents," *Handb.Solvents*, vol. 2, pp. 361–385, 2014.

Appendix

	MACCS model1	MACCS model2	MACCS model3	MACCS model4	ECFP4 model1	ECFP4 model2	ECFP4 model3	ECFP4 model4	CNN model1	CNN model2	CNN model3	CNN model4	Num Atoms model1	Num Atoms model2	Num Atoms model3	Num Atoms linear	Weight model1	Weight model2	Weight model3	Weight linear
MACCS model1	1	0.62961	0.58371	0.33337	0.32193	0.18641	0.02607	0.38472	0.24076	0.16977	0.21138	0.10213	0.06012	0.25546	0.37818	0.29545	0.46105	0.46165	0.43698	0.37293
MACCS model2	0.62961	1	0.53822	0.37358	0.18095	0.19481	0.14856	0.41016	0.3441	0.20592	0.23336	0.10264	0.15798	0.36736	0.53501	0.38193	0.53351	0.53182	0.54097	0.48271
MACCS model3	0.58371	0.53822	1	0.50484	0.49606	0.22781	0.13016	0.41682	0.13563	0.33731	0.03109	0.00405	0.32221	0.42882	0.58585	0.58789	0.68667	0.68847	0.65026	0.63333
MACCS model4	0.33337	0.37358	0.50484	1	0.41095	0.1802	0.04781	0.2883	0.03695	0.07751	-0.0048	-0.1082	-0.1129	-0.0044	0.2248	0.10341	0.25231	0.2525	0.20021	0.19593
ECFP4 model1	0.32193	0.18095	0.49606	0.41095	1	0.21008	0.12522	0.45827	-0.034	0.11247	0.0186	-0.1143	0.0123	0.04796	0.13839	0.19171	0.30289	0.30077	0.25968	0.22528
ECFP4 model2	0.18641	0.19481	0.22781	0.1802	0.21008	1	0.31645	0.35902	0.08393	0.14481	0.03682	0.00163	0.16889	0.20004	0.32491	0.2456	0.28806	0.28868	0.28485	0.27335
ECFP4 model3	0.02607	0.14856	0.13016	0.04781	0.12522	0.31645	1	0.341	0.12243	0.08096	0.09154	0.10008	-0.0281	0.1354	0.20855	0.09852	0.13861	0.13855	0.12035	0.15438
ECFP4 model4	0.38472	0.41016	0.41682	0.2883	0.45827	0.35902	0.341	1	0.17156	0.2133	0.18605	0.14867	0.01556	0.18465	0.41115	0.22697	0.35345	0.35542	0.34111	0.33176
CNN model1	0.24076	0.3441	0.13563	0.03695	-0.034	0.08393	0.12243	0.17156	1	0.30725	0.64715	0.49375	0.12607	0.1927	0.159	0.12285	0.15639	0.1559	0.13312	0.15227
CNN model2	0.16977	0.20592	0.33731	0.07751	0.11247	0.14481	0.08096	0.2133	0.30725	1	0.20322	0.10407	0.23673	0.17218	0.23859	0.40304	0.37281	0.37651	0.33413	0.38167
CNN model3	0.21138	0.23336	0.03109	-0.0048	0.0186	0.03682	0.09154	0.18605	0.64715	0.20322	1	0.51322	0.00468	0.07877	0.00521	-0.0241	0.0097	0.00795	-0.0031	-0.0057
CNN model4	0.10213	0.10264	0.00405	-0.1082	-0.1143	0.00163	0.10008	0.14867	0.49375	0.10407	0.51322	1	0.13985	0.18586	0.05966	0.0174	-0.0039	-0.0002	-0.0228	0.01665
Num Atoms model1	0.06012	0.15798	0.32221	-0.1129	0.0123	0.16889	-0.0281	0.01556	0.12607	0.23673	0.00468	0.13985	1	0.75954	0.56685	0.72659	0.5863	0.58949	0.57834	0.58003
Num Atoms model2	0.25546	0.36736	0.42882	-0.0044	0.04796	0.20004	0.1354	0.18465	0.1927	0.17218	0.07877	0.18586	0.75954	1	0.72339	0.6781	0.64313	0.64258	0.63978	0.62269
Num Atoms model3	0.37818	0.53501	0.58585	0.2248	0.13839	0.32491	0.20855	0.41115	0.159	0.23859	0.00521	0.05966	0.56685	0.72339	1	0.79367	0.84549	0.84948	0.80746	0.86127
Num Atoms linear	0.29545	0.38193	0.58789	0.10341	0.19171	0.2456	0.09852	0.22697	0.12285	0.40304	-0.0241	0.0174	0.72659	0.6781	0.79367	1	0.91561	0.92219	0.83204	0.96264
Weight model1	0.46105	0.53351	0.68667	0.25231	0.30289	0.28806	0.13861	0.35345	0.15639	0.37281	0.0097	-0.0039	0.5863	0.64313	0.84549	0.91561	1	0.998	0.95498	0.96066
Weight model2	0.46165	0.53182	0.68847	0.2525	0.30077	0.28868	0.13855	0.35542	0.1559	0.37651	0.00795	-0.0002	0.58949	0.64258	0.84948	0.92219	0.998	1	0.95413	0.96615
Weight model3	0.43698	0.54097	0.65026	0.20021	0.25968	0.28485	0.12035	0.34111	0.13312	0.33413	-0.0031	-0.0228	0.57834	0.63978	0.80746	0.83204	0.95498	0.95413	1	0.88671
Weight linear	0.37293	0.48271	0.63333	0.19593	0.22528	0.27335	0.15438	0.33176	0.15227	0.38167	-0.0057	0.01665	0.58003	0.62269	0.86127	0.96264	0.96066	0.96615	0.88671	1

Figure 24 Correlations between different models for logarithm of vapor pressure

	MACCS kNN predicted logP	MACCS RF predicted logP	MACCS XGB predicted logP	MACCS MLP predicted logP	Morgan kNN predicted logP	Morgan RF predicted logP	Morgan XGB predicted logP	Morgan MLP predicted logP	CNN 1 predicted logP	CNN 2 predicted logP	CNN 3 predicted logP	rdkit computed logP
MACCS kNN predicted logP	1	0.921715	0.888185	0.838558	0.704496	0.814018	0.792186	0.708551	0.700969	0.677225	0.625484	0.640323
MACCS RF predicted logP	0.921715	1	0.93934	0.858564	0.767814	0.893982	0.862013	0.791332	0.750316	0.743851	0.681038	0.661554
MACCS XGB predicted logP	0.888185	0.93934	1	0.915938	0.810961	0.89218	0.904705	0.800086	0.811962	0.81576	0.730698	0.765751
MACCS MLP predicted logP	0.838558	0.858564	0.915938	1	0.748923	0.787924	0.799013	0.693479	0.760765	0.747191	0.659358	0.749963
Morgan kNN predicted logP	0.704496	0.767814	0.810961	0.748923	1	0.804465	0.801829	0.697986	0.697197	0.680608	0.611802	0.709919
Morgan RF predicted logP	0.814018	0.893982	0.89218	0.787924	0.804465	1	0.920815	0.869774	0.74817	0.733199	0.700219	0.604368
Morgan XGB predicted logP	0.792186	0.862013	0.904705	0.799013	0.801829	0.920815	1	0.849583	0.77732	0.759132	0.709581	0.712061
Morgan MLP predicted logP	0.708551	0.791332	0.800086	0.693479	0.697986	0.869774	0.849583	1	0.689288	0.680273	0.666164	0.540501
CNN 1 predicted logP	0.700969	0.750316	0.811962	0.760765	0.697197	0.74817	0.77732	0.689288	1	0.849808	0.777092	0.775619
CNN 2 predicted logP	0.677225	0.743851	0.81576	0.747191	0.680608	0.733199	0.759132	0.680273	0.849808	1	0.755173	0.787624
CNN 3 predicted logP	0.625484	0.681038	0.730698	0.659358	0.611802	0.700219	0.709581	0.666164	0.777092	0.755173	1	0.65958
rdkit computed logP	0.640323	0.661554	0.765751	0.749963	0.709919	0.604368	0.712061	0.540501	0.775619	0.787624	0.65958	1

Figure 25 Correlations between different models for lipophilicity

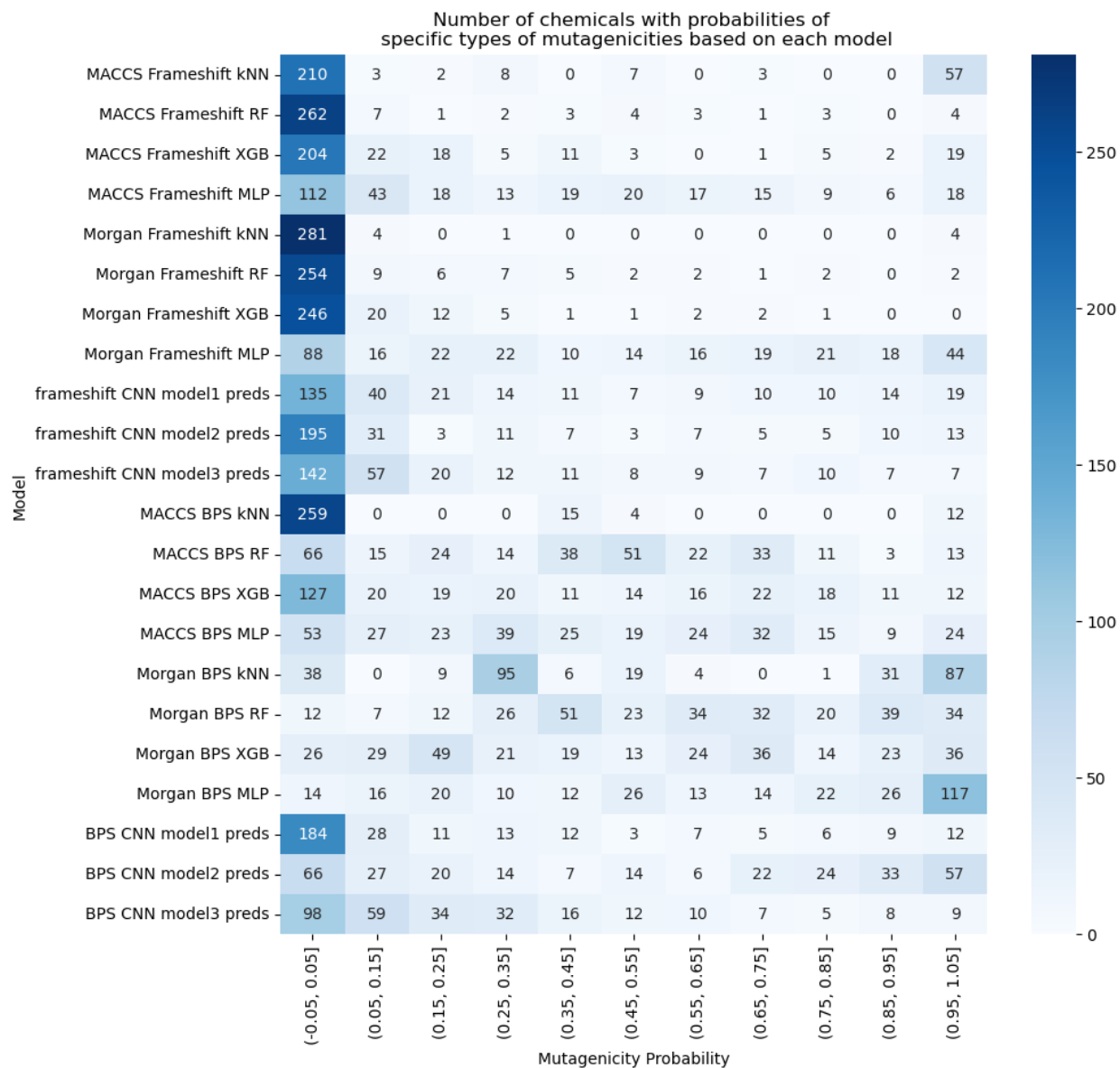


Figure 26 Counts of solvents and carbamates in each bin for predicted mutagenicity probability

	MACCS 300 kNN predicted logK _i	MACCS 300 RF predicted logK _i	MACCS 300 XGB predicted logK _i	MACCS 300 MLP predicted logK _i	Morgan 300 kNN predicted logK _i	Morgan 300 RF predicted logK _i	Morgan 300 XGB predicted logK _i	Morgan 300 MLP predicted logK _i	MACCS 500 kNN predicted logK _i	MACCS 500 RF predicted logK _i	MACCS 500 XGB predicted logK _i	MACCS 500 MLP predicted logK _i	Morgan 500 kNN predicted logK _i	Morgan 500 RF predicted logK _i	Morgan 500 XGB predicted logK _i	Morgan 500 MLP predicted logK _i	CNN model 1 AChE preds	CNN model 2 AChE preds	CNN model 3 AChE preds
MACCS 300 kNN predicted logK _i	1	0.763666	0.622087	0.320805	0.436141	0.417155	0.448136	0.047766	0.617115	0.599936	0.487724	0.292953	-0.025226	0.208637	0.3469	0.10164	0.15937	0.515898	0.427916
MACCS 300 RF predicted logK _i	0.763666	1	0.825648	0.563888	0.290136	0.424215	0.41984	-0.030969	0.368831	0.618533	0.497505	0.269318	-0.033012	0.222943	0.425224	0.065691	0.126306	0.485155	0.355255
MACCS 300 XGB predicted logK _i	0.622087	0.825648	1	0.492169	0.268871	0.348107	0.3749	-0.091823	0.350697	0.555812	0.336855	0.15801	-0.069637	0.172641	0.313916	-0.019259	0.179386	0.424804	0.349596
MACCS 300 MLP predicted logK _i	0.320805	0.563888	0.492169	1	0.05533	0.195921	0.211367	0.007182	0.054604	0.363569	0.398691	0.369099	-0.023415	0.177214	0.243832	0.089249	0.034305	0.216151	0.191578
Morgan 300 kNN predicted logK _i	0.436141	0.290136	0.268871	0.05533	1	0.530144	0.577621	0.421996	0.354414	0.292353	0.351521	0.243962	0.168531	0.275228	0.236136	0.303793	0.229478	0.431917	0.395365
Morgan 300 RF predicted logK _i	0.417155	0.424215	0.348107	0.195921	0.530144	1	0.820487	0.64019	0.181292	0.278616	0.41208	0.398345	-0.043524	0.489831	0.312997	0.532294	0.191904	0.279704	0.185915
Morgan 300 XGB predicted logK _i	0.448136	0.41984	0.3749	0.211367	0.577621	0.820487	1	0.534593	0.285729	0.314643	0.414209	0.365359	-0.011176	0.378705	0.306173	0.361018	0.233355	0.364764	0.259714
Morgan 300 MLP predicted logK _i	0.047766	-0.030969	-0.091823	0.007182	0.421996	0.64019	0.534593	1	-0.058848	-0.047615	0.197376	0.395625	0.082267	0.144347	-0.068684	0.671683	0.064126	-0.037959	-0.037493
MACCS 500 kNN predicted logK _i	0.617115	0.368531	0.350697	0.054604	0.354414	0.181292	0.285729	-0.058848	1	0.625979	0.34881	0.301425	-0.036674	0.222687	0.288102	0.037913	0.215494	0.399527	0.453005
MACCS 500 RF predicted logK _i	0.599936	0.618533	0.555812	0.363569	0.292353	0.278616	0.314643	-0.047615	0.625979	1	0.633448	0.516104	0.042813	0.351967	0.463918	0.127966	0.18182	0.349457	0.43525
MACCS 500 XGB predicted logK _i	0.487724	0.497505	0.336855	0.398691	0.351521	0.41208	0.414209	0.197376	0.34881	0.633448	1	0.552217	-0.01724	0.320204	0.40661	0.192674	0.109976	0.300806	0.39632
MACCS 500 MLP predicted logK _i	0.292953	0.269318	0.15801	0.369099	0.243962	0.398345	0.365359	0.395625	0.301425	0.516104	0.552217	1	-0.023281	0.262876	0.226307	0.426362	0.108858	0.106192	0.255276
Morgan 500 kNN predicted logK _i	-0.025226	-0.033012	-0.069637	-0.023415	0.168531	-0.043524	-0.011176	0.062267	-0.036674	0.042813	-0.01724	-0.023281	1	0.059203	0.021264	0.104033	-0.008217	-0.097937	-0.131057
Morgan 500 RF predicted logK _i	0.208637	0.222943	0.172641	0.177214	0.275228	0.489831	0.378705	0.144347	0.222687	0.351967	0.320204	0.262876	0.059203	1	0.695845	0.405768	0.101937	0.252965	0.319916
Morgan 500 XGB predicted logK _i	0.3469	0.425224	0.313916	0.243832	0.236136	0.312997	0.306173	-0.068684	0.288102	0.463918	0.40661	0.226307	0.021264	0.695845	1	0.25104	0.10968	0.375685	0.393759
Morgan 500 MLP predicted logK _i	0.10164	0.065691	-0.019259	0.069249	0.303793	0.532294	0.361018	0.671683	0.037913	0.127966	0.192674	0.426362	0.104033	0.405768	0.25104	1	-0.020412	0.038592	0.102447
CNN model 1 AChE preds	0.15937	0.126306	0.179386	0.034305	0.229478	0.191904	0.233355	0.064126	0.215494	0.18182	0.109976	0.108858	-0.008217	0.101937	0.10968	-0.020412	1	0.290124	0.292728
CNN model 2 AChE preds	0.515898	0.485155	0.424804	0.216151	0.431917	0.279704	0.364764	-0.037959	0.399527	0.349457	0.300806	0.106192	-0.097937	0.252965	0.375685	0.038592	0.290124	1	0.740051
CNN model 3 AChE preds	0.427916	0.355255	0.349596	0.191578	0.395365	0.185915	0.259714	-0.037493	0.453005	0.43525	0.39632	0.255276	-0.131057	0.319916	0.393759	0.102447	0.292728	0.740051	1

Figure 27 Correlations between different models for AChE affinity