



**HAL**  
open science

# Screening of physical solvents for carbon capture through a combined reverse engineering machine learning -CAMD approach

Edoardo Parascandolo, Vincent Gerbaud, Sophie Thiebaud-Roux, Ivonne Rodriguez-Donis

## ► To cite this version:

Edoardo Parascandolo, Vincent Gerbaud, Sophie Thiebaud-Roux, Ivonne Rodriguez-Donis. Screening of physical solvents for carbon capture through a combined reverse engineering machine learning -CAMD approach. E2DT 2025 - 3rd International Conference on Energy, Environment & Digital Transition, AIDIC, David Bogle, Flavio Manenti, Piero Salatino, Oct 2025, Palerme, Italy. <hal-05348836>

**HAL Id: hal-05348836**

**<https://hal.science/hal-05348836v1>**

Submitted on 5 Nov 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License



# Screening of physical solvents for carbon capture through a combined reverse engineering – machine learning - CAMD approach

Edoardo Parascandolo<sup>a,\*</sup>, Vincent Gerbaud<sup>b</sup>, Sophie Thiebaud-Roux<sup>a</sup>, Ivonne Rodriguez-Donis<sup>a</sup>

<sup>a</sup> Laboratoire de Chimie Agro-Industrielle, Université de Toulouse, INRAE, INP, Toulouse, France

<sup>b</sup> Laboratoire de Génie Chimique, Université de Toulouse, CNRS, INP, UPS, Toulouse, France

[edoardo.parascandolo@toulouse-inp.fr](mailto:edoardo.parascandolo@toulouse-inp.fr)

Reducing CO<sub>2</sub> emissions is crucial to mitigating climate change. At the moment, the use of amines in Carbon Capture has major drawbacks. Although current physical solvents can partially overcome these limitations, they exhibit poor selectivity towards acid gases and require demanding operating conditions. In this work, a novel Computer Aided Molecular Design and Screening procedure is proposed as an alternative to trial-and-error approaches for identifying new solvent candidates. By leveraging predictive physical property models, tailored distribution functions, Health, Safety and Environment criteria and a Genetic Algorithm for the generation of molecular structures, the methodology is first validated on benchmark solvents to then proceed screening an extensive database of molecules. An unconventional alternative to current physical solvents is identified.

## 1. Introduction

With the global average temperature around 1.2°C above pre-industrial levels (IEA, 2023), consequences as extreme weather events and more have become frequent: CO<sub>2</sub> emissions remain the primary driver of global warming, making their reduction essential to deal with climate change. At the moment, Carbon Capture (CC) technologies represent a viable though promising technique. However, the most developed technology, that is chemical absorption with aqueous mixtures of blended amines, presents significant limitations, including toxicity, corrosion, and high energy demand for regeneration. The demand for developed alternatives in the recent future needs to be addressed. Due to the absence of any reaction and a higher sequestration capacity, physical absorption technologies allow for lower energy consumption and solvent recirculation rates (N.Borhani & Wang, 2019). Conversely, physical solvents can suffer from limitations in the selective removal of acid gases, consistent losses in the product and applicability restricted to specific operating conditions (Kidnay & Parrish, 2006).

Given the complicated pattern between properties and device performance, the selection of alternative solvents relies primarily on time-consuming and costly trial-and-error approaches. Recent studies in carbon capture have significantly advanced the development (Kim et al., 2024) and the accuracy (Li et al., 2019; Orlov et al., 2022) of computational frameworks for solvent screening. Despite the improvements, all proposed methodologies rely on a single or a limited set of fundamental properties, assessed independently during the selection phase. It has been emphasized that a more comprehensive evaluation is essential to fully capture the complexity of solvent effectiveness (Mota-Martinez et al., 2017). Besides, it is equally important not to tailor the evaluation on specific functional groups (e.g., amine groups), a limitation that persists in the most developed approaches. To address these challenges, solvent screening can be enhanced through Reverse Engineering leveraging Computer-Aided Molecular Design (CAMD). Initially, a set of fundamental physical and chemical properties aligned with the desired solvent functionalities is investigated for ranking purposes. The process then involves screening candidate solvents and designing novel molecular structures that better fulfill

these requirements. Unlike traditional methods, this approach enables the systematic exploration of a larger chemical domain through a weighted multi-objective evaluation of an extended set of properties, providing a more balanced and comprehensive selection criteria.

In this study, the evaluation of properties was extended to all non-reactive molecules with CO<sub>2</sub> by employing different predictive models depending on the property of interest. Specifically, Group Contribution methods (GC), Corresponding State (CS) models and Graph Convolutional Neural Networks (GCNNs) were employed selectively. Health, Safety and Environment (HSE) criteria were used as a preliminary filter. The resulting molecules were then classified with a global performance function, which quantifies the deviation between estimated values and target requirements using Gaussian-type distributions. A genetic algorithm with elitism policy on the global performance function was then employed to modify the chemical structures of the highest-performing molecules.

## 2. Methodology

Reverse Engineering follows three sequential phases, which can be carried out employing two different strategies (Heintz et al., 2014), as illustrated in Figure 1. When the top-down approach is implemented, molecular structures are designed with structural constraints and then modified to better match desired functionalities, while in the bottom-up approach, target properties are computed for a large number of existing molecules while constraints are applied to the best-performing molecules to refine and optimize their properties. In the context of Carbon Capture with physical solvents, the bottom-up approach is first carried out on an extensive database in order to identify the chemical groups composing the best candidates. Following, the top-down approach is used for designing molecular structures which better meet target properties requirements.

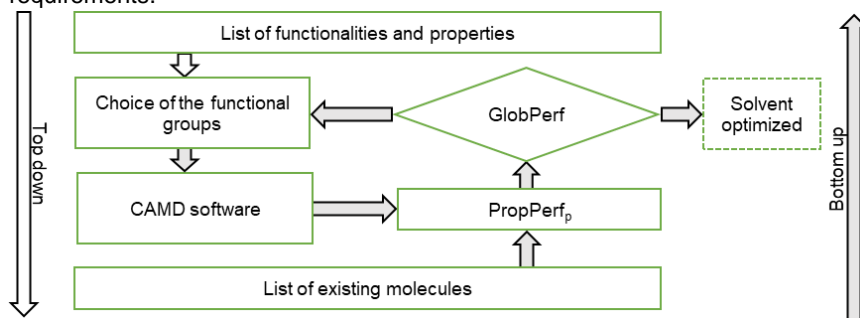


Figure 1: Reverse Engineering methodology, Top-Down and Bottom-Up procedures

### 2.1 Intelligence Phase

In the first phase, referred to as the Intelligence Phase, existing data of benchmark solvents (i.e., Methanol, NMP, Propylene Carbonate, Selexol™) (N.Borhani & Wang, 2019) are investigated to determine all the relevant physical properties and to establish their target values and weights. The selected properties (i.e., solubility and selectivity at 25 °C, viscosity, vapor pressure, surface tension, molar volume, boiling point, melting point and flash point) were identified to properly assess efficiency, operational feasibility, and process stability.

The most suitable predictive models available in the literature are then selected to estimate the target properties

with the exception of solubility and selectivity. Indeed, the estimation of Vapor-Liquid Equilibrium (VLE) between gas solutes and physical solvents with conventional methods requires the accurate description of the high-pressure deviations from ideality in the vapor pressure using an Equation of State (EOS) based approach. The Predictive Soave-Redlich-Kwong (PSRK) and Predictive Peng Robinson (PPR) EOS, although modeling the liquid phase through GC methods, still use critical properties and the acentric factor in the description of the vapor phase. Any inaccuracies in their estimation using GC predictive models instead of regressed values can significantly propagate through the calculations. For this reason, a GCNN (Heid et al., 2024; Yang et al., 2019) was developed to predict phase diagrams between physical solvents and CO<sub>2</sub> (for solubility purposes), H<sub>2</sub>S (for selectivity purposes). The two models use a Feed-Forward Network (FFN) to process the embeddings generated in the Message Passing (MP) phase from the SMILES of each molecule (Parascandolo, 2025). These embeddings are weighted according to scaled values of mole fraction and temperature of equilibrium, for the prediction of the corresponding equilibrium pressure. Therefore, the methodology remains fully predictive while preserving accuracy, ensuring reliable phase equilibrium

estimations without relying on experimentally regressed parameters. Selected predictive models are summarized in Table 1.

Table 1: Overview of the predictive models used for property estimation

Property	Reference	Method Type	Mean Error	Standard Deviation	Data
Melting Point	(Hukkerikar et al., 2012)	GC in 3 levels	17.65 °C	21.45°C	5183
Boiling Point	(Hukkerikar et al., 2012)	GC in 3 levels	5.96 °C	7.67 °C	3510
Flash Point	(Hukkerikar et al., 2012)	GC in 3 levels	8.38 °C	11.43 °C	512
Molar Volume	(Hukkerikar et al., 2012)	GC in 3 levels	2.1 mm <sup>3</sup> /kmol	4.2 mm <sup>3</sup> /kmol	1056
Vapor Pressure	(Riedel, 1954)	CS	N.A.	N.A.	N.A.
Surface Tension	(Conte et al., 2008)	GC in 3 levels	1.05 mN/m	1.47 mN/m	402
Viscosity	(Conte et al., 2008)	GC in 3 levels	0.37 cP	0.89 cP	430
Solubility	(Parascandolo, 2025)	GCNN	3.84 bar	4.77 bar	1091
Selectivity	(Parascandolo, 2025)	GCNN	10.64 bar	9.75 bar	163

All these insights are then used to construct individual performance functions (PropPerf<sub>p</sub>) which are then aggregated into a global one (GlobPerf) using either an arithmetic or geometric weighted mean. Gaussian type functions were used for evaluating the performance of single properties. These functions provide a smooth and systematic measure of how closely a predicted property aligns with its target value. The general formulation is given by:

$$\ln(\text{PropPerf}_p) = \ln(\text{val}) \cdot \left(\frac{P - x}{\text{tol}}\right)^2 \quad (1)$$

where  $x$  is the predicted value,  $P$  is the target value and  $\text{tol}$  accounts for property model uncertainty (Heintz et al., 2014). Upper and lower bounded Gaussian function were implemented to handle properties where exclusively negative or positive deviations from the target are penalized. The global performance function was then formulated as the sum of  $n_p$  individual property function multiplied for a factor  $w_p$ , where  $n_p$  is the total number of properties, as follows:

$$\text{GlobPerf} = \frac{\sum_{p=1}^{n_p} w_p \cdot \text{PropPerf}_p}{\sum_{p=1}^{n_p} w_p} \quad (2)$$

Values for  $\text{tol}$ ,  $\text{val}$  and  $P$  were respectively chosen taking into account the standard deviation of each model and benchmark solvents optimal properties. Unitary weighting factors were selected for all the properties with the exception of solubility and selectivity, that were set to 2.

## 2.2 Design Phase

In the next phase, starting with the bottom-up approach, the search was conducted over an in-house database of existing molecules containing more than 12,000 candidates. The CHEM21 criteria (Prat et al., 2016) were applied to exclude from the analysis compounds undesired from an HSE perspective. This selection guide uses a classification that follows Global Harmonized Systems (GHS) and European regulations to assign scores based on flammability, toxicity and environmental persistence, with a color-coded scale that categorizes solvents as recommended for the use (green), problematic (yellow) or hazardous (red). Due to occupational risks, severe environmental impact or regulatory concerns (e.g., carcinogenicity, high volatility, or bioaccumulation), all the solvents that were labeled as hazardous were systematically removed from consideration. Following, candidates showing the highest values of GlobPerf are selected as base chemical structures to proceed with the top-down approach. The selection of the most suitable modification requires the solution of a Mixed Integer Non-Linear Programming (MINLP) problem. The IBSS tool (Heintz et al., 2014) for CAMD was employed, where the targeted molecular optimization approach is implemented using a genetic algorithm with elitism policy, ensuring that the best-performing structures are retained across generations. For each generated molecule, the global performance function (GlobPerf) is evaluated guiding the search towards molecular structures maximizing the objective function.

## 2.3 Choice Phase

In the last phase, the best candidates in both the bottom-up and top-down lists undergo an experimental validation. For newly designed molecules, potential green synthetic routes are investigated to assess the feasibility of producing the selected molecules in a sustainable and scalable manner to then assess inherently

experimental properties, such as toxicity. For all the selected molecules, target properties are measured to validate the estimations obtained from predictive models.

### 3. Results and Discussions

The methodology was first validated on benchmark solvents to confirm its reliability in property estimation and scoring accuracy, ensuring that the computed GlobPerf values consistently reflect experimentally observed trends. The results demonstrate that Selexol™, NMP, and Propylene Carbonate are associated to consistently high scores due to their favorable physical properties for CO<sub>2</sub> absorption. In contrast, methanol receives unfavorable scores due to its inherent instability as a solvent under conventional operating conditions. Since methanol-based processes in CC are operated at subzero temperatures (N.Borhani & Wang, 2019) and the PropPerf<sub>p</sub> are estimated at 25 °C, the penalization in the scoring methodology appropriately reflects its limited suitability. Furthermore Selexol™ exhibit the highest solubility and selectivity towards H<sub>2</sub>S, whereas NMP and Propylene Carbonate are labeled as good and intermediate candidates. In terms of physical properties, NMP is severely penalized due to its high vapor pressure. All these observations are aligned with experimental evidence (N.Borhani & Wang, 2019). Table 2 collects the estimated values of PropPerf<sub>p</sub> for all the target property and the resulting aggregation in GlobPerf.

Table 2: Normalized PropPerf<sub>p</sub> and GlobPerf values for Benchmark Solvents employed in CO<sub>2</sub> Capture

Solvent	B.P.	M.P.	F.P.	Mol.V.	Vap.P.	Surf.Ten.	Visc.	Sol.	Sel.	GlobPerf
Selexol™	1	0.974	1	1	1	0.996	1	1	1	0.997
NMP	1	1	1	1	0.615	0.999	1	0.8	0.9	0.912
Propylene Carbonate	1	1	1	0.998	1	1	1	0.4	0.788	0.852
Methanol	0.000038	1	0.262	0.996	0	1	1	0.331	0.769	0.587

#### 3.1 Bottom-Up Results

The procedure has been further applied to the molecules identified from the screened database introduced in section 2.2. Interestingly, a dimer representing the monomeric backbone of Selexol™ was found to exhibit comparable performance in terms of solubility and selectivity (Figure 2). However, its high vapor pressure significantly limits its applicability, as excessive volatility leads to increased solvent losses and operational instability. Due to this limitation, Selexol™ excludes dimers and only incorporates oligomers starting from n = 3 (with n number of monomers repeated in the chain) (Ashkanani et al., 2020) to ensure adequate thermal and physical stability. Nevertheless, the CHEM21 guide classifies the dimer as "recommended", highlighting its lower environmental impact compared to other solvent alternatives. This classification is primarily driven by its lower potential for bioaccumulation. In contrast, long-chain molecules are penalized due to their increased persistence in the environment, leading to a less favorable score.

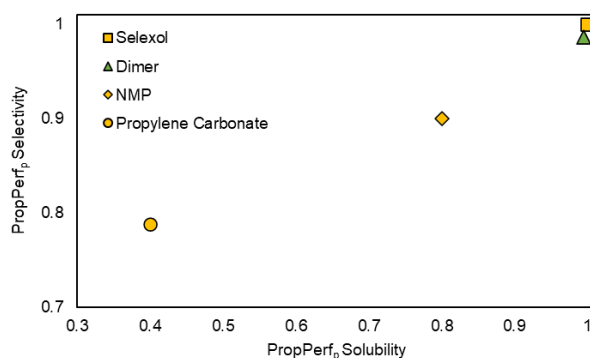


Figure 2: PropPerf<sub>p</sub> values of solubility and selectivity for benchmark solvents and the dimer. Green and yellow colors indicate respectively recommended or problematic solvents according to CHEM21 selection guide

#### 3.2 Top-Down Results

The dimeric unit was selected as the core molecular scaffold for searching new structures optimizing the required properties. To achieve this, different chemical bricks were systematically introduced at specific

positions to generate structures covering most of existing chemicals available in the market (Figure 3). The approach allowed for the aggregation of up to three groups in series for each site, enabling an adequate diversity while preserving favorable physical properties.

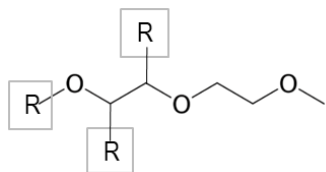


Figure 3: Molecular scaffold used for generating the population

The result was the identification of a molecule featuring additional functional groups beyond the ether backbone of the dimer that allows for the retention of high selectivity and solubility values (Figure 4) while simultaneously exhibiting a negligible vapor pressure compared to the dimer.

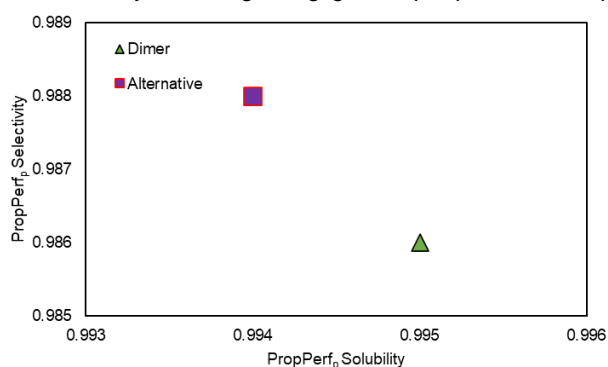


Figure 4: PropPerf<sub>p</sub> values of solubility and selectivity for the dimer and the alternative solvent

Additionally, the other properties of interest are similar in their values to the already ideal ones of the dimer, as a result of the targeted optimization of the research algorithm. A more in-depth investigation provided detailed insights regarding its safety aspects. Toxicological data reported in literature revealed that the molecule exhibits negligible risk phrases (H302, H315), indicating a low hazard potential in terms of health and environmental impact. Furthermore, being a short-chain molecule with lower boiling point, its reduced bioaccumulation potential and lower environmental persistence make it a more sustainable alternative to Selexol™.

Table 3: Physical properties for the dimer and the alternative solvent

Solvent	B.P. (°C)	M.P. (°C)	F.P. (°C)	Mol. V. (cm <sup>3</sup> /mol)	Vap. P. (mmHg)	Surf. Ten. (mN/m)	Visc. (cP)
Dimer	162	-64	70	142	1.57	23.9	1.09
Alternative	231.4	-15.4	96.4	172.1	0.0376	31.9	2.9

#### 4. Conclusions

This study presents an advanced methodology for screening and designing physical solvents for CO<sub>2</sub> capture, integrating reverse engineering, predictive property models and CAMD techniques. The approach was validated against benchmark solvents, demonstrating its accuracy in predicting key physicochemical properties and ensuring alignment with experimentally observed trends. The computed performance function values correctly ranked Selexol™ as the best-performing solvent, followed by NMP, propylene carbonate, and methanol, which was appropriately penalized due to its high volatility under standard operating conditions. The procedure was then extended to an in-house database, leading to the identification of a dimeric unit structurally related to Selexol™ but not considered due to its excessive volatility. The top-down phase led to the design of an optimized molecular structure featuring an additional functional group beyond the ether backbone, which successfully retained similar properties while exhibiting a drastically reduced vapor pressure.

Furthermore, the optimized solvent displayed a favorable safety profile, with negligible risk phrases and a lower environmental persistence compared to longer-chain oligomers. These results highlight the potential of the proposed framework to systematically explore the chemical space and optimize solvent design. Future work will focus on refining both the intelligence and design phases for expanding the search space to chemical absorbents.

## Acknowledgments

This research was supported by State funding administered by the National Research Agency under the France 2030 initiative, reference ANR-22-PESP-0008, and through the “Investissements d’Avenir” program, reference ANR-18-EURE-0021.

## References

- Ashkanani, H. E., Wang, R., Shi, W., Siefert, N. S., Thompson, R. L., Smith, K., Steckel, J. A., Gamwo, I. K., Hopkinson, D., Resnik, K., & Morsi, B. I. (2020). Levelized Cost of CO<sub>2</sub> Captured Using Five Physical Solvents in Pre-combustion Applications. *International Journal of Greenhouse Gas Control*, *101*, 103135. <https://doi.org/10.1016/J.IJGGC.2020.103135>
- Conte, E., Martinho, A., Matos, H. A., & Gani, R. (2008). Combined Group-Contribution and Atom Connectivity Index-Based Methods for Estimation of Surface Tension and Viscosity. *Industrial & Engineering Chemistry Research*, *47*(20), 7940–7954. <https://doi.org/10.1021/ie071572w>
- Heid, E., Greenman, K. P., Chung, Y., Li, S.-C., Graff, D. E., Vermeire, F. H., Wu, H., Green, W. H., & McGill, C. J. (2024). Chemprop: A Machine Learning Package for Chemical Property Prediction. *Journal of Chemical Information and Modeling*, *64*(1), 9–17. <https://doi.org/10.1021/acs.jcim.3c01250>
- Heintz, J., Belaud, J. P., Pandya, N., Teles Dos Santos, M., & Gerbaud, V. (2014). Computer aided product design tool for sustainable product development. *Computers & Chemical Engineering*, *71*, 362–376. <https://doi.org/10.1016/J.COMPCHEMENG.2014.09.009>
- Hukkerikar, A. S., Sarup, B., Ten Kate, A., Abildskov, J., Sin, G., & Gani, R. (2012). Group-contribution+ (GC+) based estimation of properties of pure components: Improved property estimation and uncertainty analysis. *Fluid Phase Equilibria*, *321*, 25–43. <https://doi.org/10.1016/J.FLUID.2012.02.010>
- IEA. (2023). *World Energy Outlook 2023*. <https://www.iea.org/reports/world-energy-outlook-2023>
- Kidnay, A. J., & Parrish, W. R. (2006). *Fundamentals of Natural Gas Processing*. CRC Press. <https://books.google.fr/books?id=IWrlBQAAQBAJ>
- Kim, J., Kim, K., Lim, H., Kang, J. H., Park, H. S., Park, J., & Song, H. (2024). Structural investigation of aqueous amine solutions for CO<sub>2</sub> capture: CO<sub>2</sub> loading, cyclic capacity, absorption–desorption rate, and pK<sub>a</sub>. *Journal of Environmental Chemical Engineering*, *12*(3), 112664. <https://doi.org/10.1016/J.JECE.2024.112664>
- Li, H., Yan, D., Zhang, Z., & Lichtfouse, E. (2019). Prediction of CO<sub>2</sub> absorption by physical solvents using a chemoinformatics-based machine learning model. *Environmental Chemistry Letters*, *17*(3), 1397–1404. <https://doi.org/10.1007/s10311-019-00874-0>
- Mota-Martinez, M. T., Hallett, J. P., & Mac Dowell, N. (2017). Solvent selection and design for CO<sub>2</sub> capture – how we might have been missing the point. *Sustainable Energy Fuels*, *1*(10), 2078–2090. <https://doi.org/10.1039/C7SE00404D>
- N.Borhani, T., & Wang, M. (2019). Role of solvents in CO<sub>2</sub> capture processes: The review of selection and design methods. *Renewable and Sustainable Energy Reviews*, *114*, 109299. <https://doi.org/10.1016/J.RSER.2019.109299>
- Orlov, A. A., Valtz, A., Coquelet, C., Rozanska, X., Wimmer, E., Marcou, G., Horvath, D., Poulain, B., Varnek, A., & de Meyer, F. (2022). Computational screening methodology identifies effective solvents for CO<sub>2</sub> capture. *Communications Chemistry*, *5*(1), 37. <https://doi.org/10.1038/s42004-022-00654-y>
- Parascandolo, E. (2025). *edparasca/CO2\_physical\_solvents\_models: pre-release*. Zenodo. <https://doi.org/10.5281/zenodo.14875951>
- Prat, D., Wells, A., Hayler, J., Sneddon, H., McElroy, C. R., Abou-Shehadeh, S., & Dunn, P. J. (2016). CHEM21 selection guide of classical- and less classical-solvents. *Green Chem.*, *18*(1), 288–296. <https://doi.org/10.1039/C5GC01008J>
- Riedel, L. (1954). Eine neue universelle Dampfdruckformel Untersuchungen über eine Erweiterung des Theorems der übereinstimmenden Zustände. Teil I. *Chemie Ingenieur Technik*, *26*(2), 83–89. <https://doi.org/https://doi.org/10.1002/cite.330260206>
- Yang, K., Swanson, K., Jin, W., Coley, C., Eiden, P., Gao, H., Guzman-Perez, A., Hopper, T., Kelley, B., Mathea, M., Palmer, A., Settels, V., Jaakkola, T., Jensen, K., & Barzilay, R. (2019). Analyzing Learned

Molecular Representations for Property Prediction. *Journal of Chemical Information and Modeling*, 59(8), 3370–3388. <https://doi.org/10.1021/acs.jcim.9b00237>