Agnostic Causality-Driven Enhancement of Chemical Foundation Models on Downstream Tasks

Victor Shirasuna
IBM Research Brazil
São Paulo, SP, Brazil
victor.shirasuna@ibm.com

Eduardo Soares
IBM Research Brazil
Rio de Janeiro, RJ, Brazil
eduardo.soares@ibm.com

Emilio Vital Brazil
IBM Research Brazil
Rio de Janeiro, RJ, Brazil
evital@br.ibm.com

Karen Fiorella Aquino Gutierrez
IBM Research Brazil
São Paulo, SP, Brazil
fiorella.aquino@ibm.com

Renato Cerqueira
IBM Research Brazil
Rio de Janeiro, RJ, Brazil
rcerq@br.ibm.com

Dmitry Zubarev
IBM Research Almaden
San Jose, CA, USA
dmitry.zubarev@ibm.com

Kristin Schmidt
IBM Research Almaden
San Jose, CA, USA
schmidkr@us.ibm.com

Abstract

Recent advancements in large foundation models have revealed impressive capabilities in mastering complex chemical language representations. These models undergo a task-agnostic learning phase, characterized by pre-training on extensive unlabeled corpora followed by fine-tuning on specific downstream tasks. This methodology reduces reliance on labeled data, facilitating data acquisition and broadening the scope of chemical language representation. However, real-world scenarios often pose challenges due to domain shift, necessitating robust domain adaptation strategies to maintain performance levels across different contexts. To address this, we present a novel causal-based framework for feature selection and domain adaptation to enhance the performance of chemical foundation models on downstream tasks. Our approach employs a multi-stage feature selection method that identifies physico-chemical features based on their direct causal-effect over specific downstream properties. By employing Mordred descriptors and Markov blanket causal graphs, our approach provides insight into the causal relationships between features and target properties for prediction tasks. We evaluate our approach on various foundation model architectures and datasets, demonstrating consistent performance improvements, which showcases the robustness and the agnostic nature of our approach.

1 Introduction

Recent advancements in large foundation models have showcased impressive capabilities in learning complex chemical language representations [1]. The task-agnostic learning phase of these models are generally based on a two-step process: pre-training on large unlabeled corpora followed by fine-tuning on specific downstream tasks [2, 3, 4, 5]. This methodology significantly reduces dependence on labeled data, thereby streamlining data acquisition and expanding the boundaries of chemical language representation [6]. However, in real-world scenarios, these models often encounter challenges due to domain shift, wherein there's a notable misalignment between the target domain and the source

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

domain of their initial training [7]. This phenomenon underscores the critical need for robust domain adaptation strategies to maintain performance levels in different contexts [4]. The domain adaptation aims to embed domain knowledge into the foundation model [8].

Domain specific features can improve the efficiency of foundation model into specific downstream tasks [9, 10, 11]. Chemical descriptors as Mordred [12] and PADEL [13] and provides hundreds/thousands of physical and chemical information about molecules that carries predictive potential to machine learning models [14]. However, such molecular descriptors can carry highly redundant features [15]. This redundancy can be partially reduced by means of feature selection methods [15]. To select molecular domain features is laborious and time-consuming process that requires deeper expertise [16]. This limitation can be addressed with the aid of causal feature selection. Unlike traditional methods, causal feature selection aims to identify features that have a direct and meaningful impact on outcomes [17, 18]. It differentiates between correlation and causation, ensuring that the model's decisions are based on factors that influence the results [19, 20].

In this paper, we present a novel causal-based framework for feature selection and domain adaptation to enhance the performance of chemical foundation models on downstream tasks. Our approach employs a multi-stage feature selection based on the Markov Blanket method that selects physico-chemical features based on their direct causal-effect over a specific downstream property. The proposed approach used Mordred as physico-chemical descriptors and Markov blanket causal graphs [21] as inference algorithm to identify the causal relation between the features and the given property to predict on a fine-tuning task.

To demonstrate the robustness and versatility of our methodology, we investigated different chemical foundation models employing diverse architecture backbones, including transformers [22], graph-based models [23], and state-space models (Mamba) [24]. Our evaluation encompassed benchmark datasets from MoleculeNet [25] alongside real-world tasks focusing on biodegradability [26] and toxicity estimation [27]. Our results consistently demonstrate performance enhancements across diverse downstream tasks for the different foundation models when causal features are employed.

Results demonstrates that performance enhancements remained consistent for both frozen and fine-tuned models, covering a spectrum of classification and regression tasks. A detailed examination of the QM9 dataset showcases improvements across all quantum properties tasks for all tested foundation models. To ensure the reliability of our approach, we conducted experiments using 10 different seeds, with detailed results provided in the Supplementary Materials.

In summary, our paper introduces a promising avenue for future research in molecular property prediction. By leveraging high-dimensional embeddings derived from unlabeled data and incorporating domain-specific physico-chemical features with causal effects on downstream tasks, our approach offers a robust framework for advancing predictive modeling in various scientific domains.

2 Overview of the proposed approach

The schematic overview of our proposed approach, which leverages causal physico-chemical features to enhance language-based chemical foundation models for molecular prediction, is depicted in Fig. 1. Detailed insights into the multi-stage causal-based feature selection process for molecular properties are outlined in subsection 2.1. Additionally, subsection 2.2 elaborates on the integration process, wherein embeddings derived from the foundation models are incorporated with the physico-chemical features selected through causal analysis.

2.1 Multi-stage causal-based feature selection

The markov-blanket causal-based feature selection we propose follows a multi-stage approach and is divided into three modules (see Fig. 1). The first module is the responsible for the chemical descriptors extractor using Mordred calculation [12]. Mordred is an open-source library which produces 1826 physicochemical features splitted into 50 different modules, including both two- and three-dimensional descriptors [12].

Next, we have the causal feature selection per module. In this layer, we use a Markov Blanket based approach to identify and select physicochemical features that have a causal relationship with the

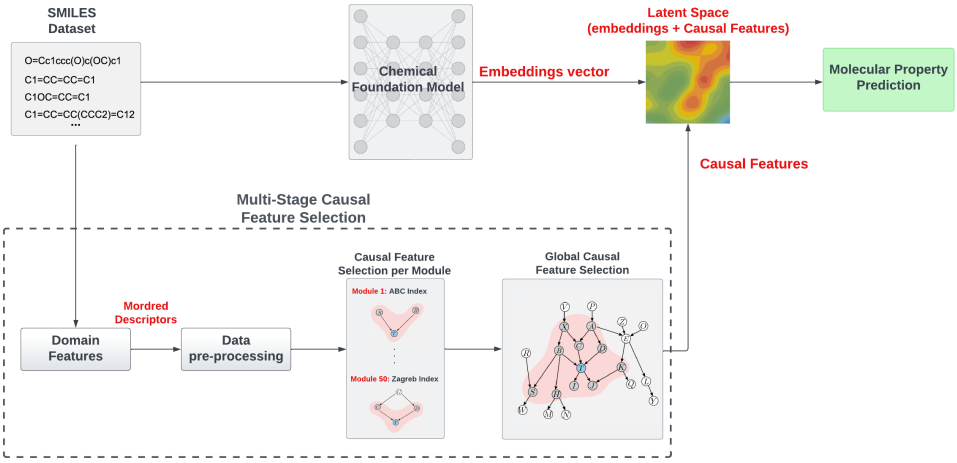

Figure 1: The diagram above depicts the proposed architecture of multi-stage causal-based feature selection process as an engine to improve molecular properties prediction based on chemical foundation models.

studied property. Differently from correlation analysis, causal-effect analysis delves deeper into the cause-and-effect relationship between a feature and a target variable [17].

Finally, we have the global causal feature selection block, which further refines the feature selection by considering eliminating redundancy and prioritizing predictive power.

2.1.1 Chemical descriptors extraction

In this paper, we use Mordred descriptors as physicochemical feature extractor of molecules. Mordred is an open-source library which allows the extraction more than 1800 physicochemical properties of molecules. Mordred has been widely adopted to machine learning models due to its predictive power. RDKit [28] was used to standardize the chemical language representations, mapping different strings belonging to the same molecule to a single molecule object.

Following the extraction of Mordred features from the molecules, we do pre-processing tasks as outlined below:

a) **Drop of features with constant values:** Features with constant values provide little or no relevant information to machine learning models and may introduce unnecessary noise. Thus, we removed such features during the data cleaning process.

b) **Drop of features with more than 50% of missing values:** Features with more than 50% of missing values may not provide meaningful insights as the missing values could mask important information.

c) **Drop of Autocorrelation features:** There are more than 600 Autocorrelation descriptors. According to [29], these Autocorrelation descriptors do not correspond to structural, physical, and chemical properties of the molecule. Therefore, these data are removed during the cleaning steps.

These pre-processing steps are central to eliminate redundant and noisy data. Enhance the dataset's quality and setting a solid foundation for causal analyses which is detailed in the next subsection.

2.1.2 Markov Blanket causal feature selection per descriptor module

In this paper, we consider the Markov Blanket (MB) algorithm to build the causal graph relation. The Markov Blanket algorithm theoretically selects an optimal set of features considering their performance both individually and in a group to estimate a target T property (see [30]). Therefore, with the MB(T), it is enough to know the distribution of T, and all the values of the other variables that do not belong to the MB(T) become irrelevant.

Next, in Fig 2, we present an example of a Markov Blanket graph that is used to represent the joint probability distribution of a set of features A, B, C, D, E, T in the form of a directed acyclic graph (DAG). The nodes of the graph represent the features, and the directed edges represent the conditional dependencies between these features.

The MB(T) consists of its parents (direct causes), children (direct effects), and spouses (other parents of this variable's children) of T [17]. In the graph, the variables inside the gray circle are in the MB of the target T and are used for their prediction.

In this study, we considered the Predictive Permutation Feature Selection (PPFS) algorithm [31], which employs a Markov Blanket graph for feature selection. PPFS stands out as a versatile method since it enables subset selection for datasets encompassing both categorical and continuous features, making it applicable to both classification and regression tasks. The PPFS algorithm is based on two phases: the growth and the shrink phase.

The growth phase of PPFS determines which features are individually important to predict the target variable. Let $X \in \mathbb{R}^{n \times d}$ be an input feature matrix. Let S be the set of features in X such that, $S = \{X_1, X_2, ... X_d\}$. Let Y be the target variable and MB(Y) be the Markov Blanket of Y. In order to determine individual importance, the predictive permutation independence (PPI) test checks for marginal independence by conditioning each feature, $X_i \in S$, on the empty set. The test returns a p-value which if less than some significance level α results in the addition of that feature into MB(Y) since the feature is important for predicting the target variable. This process is shown in Eq. 1.

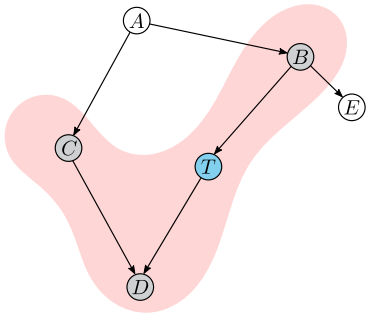

Figure 2: The figure illustrate a Markov Blanket Graph where the target T is direct affected by the feature B and causes feature D.

$$\text{MB}(Y) = \{X_i \in S \mid \text{PPI}(X_i \perp\!\!\!\perp Y \mid \emptyset) < \alpha, \quad \forall i \in \{1, 2, \ldots, d\} \tag{1}$$

Where PPI refers to the predictive permutation independence which is a non-parametric generalized conditional independence test [31]. The notation $\perp\!\!\!\perp$ denotes Conditional Independence (CI), which means that if a variable X is conditionally independent of Y given some variable Z, it can be written as $(X \perp\!\!\!\perp Y|Z)$. The null hypothesis of the CI test is represented by H_0. The null hypothesis of PPI is defined as,

$$H_0 : X_i \perp\!\!\!\perp Y \mid U \tag{2}$$

Where $U \subseteq S \backslash X_i$. Therefore, U can also be an empty set which is the case for marginal independence. The growth phase employs the PPI test d times, once for each feature resulting in a worst-case time complexity of $\mathcal{O}(d) * \mathcal{O}(nkdB)$. Since, the PPI test in the growth phase is conditioned on the empty set, the current worst-case time complexity of PPI is $\mathcal{O}(nkB)$ as $d = 1$. This results in $\mathcal{O}(dnkB)$ as the final worst-case time complexity of the growth phase. It is clear that, the growth phase is linear with respect to the number of features in the dataset. In the worst case, if all features together form the Markov blanket then $|\text{MB}(Y)| = d$, resulting in a polynomial worst-case time complexity of $\mathcal{O}(d^2)$.

The shrink phase is responsible for removing any false positives that might have been added to the MB during the growth phase. Each feature, $X_i \in \text{MB}(Y)$ is checked for conditional independence, using the PPI test, conditioned on the rest of the Markov Blanket, $\text{MB}(Y) \setminus \{X_i\}$. If the value from the PPI test is not significant then the feature, X_i is removed from MB(Y) and the Markov Blanket is updated as the feature is not important with respect to the remaining MB(Y). Once this happens, the entire shrink phase begins again and the process continues until no more features can be removed during a single iteration of the shrink phase.

The identified markov blankets are the minimal set of features that can predict the target variable [21]. Therefore, in the end of this process we choose the MB with the highest causal values for each module contained in the set of features descriptors.

2.1.3 Global causal feature selection

The global causal feature selection step consolidates the Markov Blanket features selected from each module in the preceding step. Its objective is to construct common Markov Blanket graph while filtering out redundant information and retaining only the most significant physicochemical features descriptors across modules.

Through this process, the proposed approach provides a concise set of causal features, along with their respective importance. This ensures not only the parsimony of the feature set but also enhances the interpretability and efficiency of the model.

2.2 Improving foundation models with causal physicochemical features

At this stage, chemical embeddings derived from the foundation models are combined with physicochemical features obtained from the causal multi-stage feature selection process. These chemical embeddings can be extracted from either frozen or fine-tuned models, and both alternatives are explored in this paper.

Let (x, y) denote a feature-target pair, where $x = (x_{\text{FM}}, x_{\text{causal}})$ represents the combination of features. Here, x_{FM} denotes features derived from the chemical foundation models, and x_{causal} refers to the physicochemical features selected during the multi-stage causal-based feature selection process. The size of each x_{FM} embedding may vary depending on the architectures they were built upon.

The resulting latent space can be utilized to train machine learning models for various downstream tasks. In this study, we employ the XGBoost algorithm for optimization. Further details can be found in the Supplementary Materials.

3 Experiments

To evaluate the effectiveness of our proposed methodology, we conducted experiments using a set of 10 datasets sourced from MoleculeNet [25], AI4PFAS [27], and biodegradability of compounds [26], as demonstrated in Table 1. Specifically, we evaluated 5 datasets for classification task and 5 datasets for regression tasks. To ensure an unbiased assessment, we maintained consistency with the original benchmark by adopting identical train/validation/test splits for all tasks [25]. We also conducted the experiments considered 10 different seeds for all the tests in other to guarantee the robustness of the approach. Details are provided in the Supplementary Materials.

Table 1: Evaluated datasets description

Dataset	Description	# compounds	# tasks	Metric
BBBP	Blood brain barrier penetration dataset	2039	1	ROC-AUC
HIV	Ability of small molecules to inhibit HIV replication	41127	1	ROC-AUC
BACE	Binding results for a set of inhibitors for β – secretase 1	1513	1	ROC-AUC
LDToxDB	oral rat LD_{50} toxicity measurements	13329	2	ROC-AUC/RMSE
"All-Public set"	Compounds biodegradability	2830	1	ROC-AUC
QM9	12 quantum mechanical calculations	133885	12	Average MAE
QM8	12 excited state properties of small molecules	21786	12	Average MAE
ESOL	Water solubility dataset	1128	1	RMSE
FreeSolv	Hydration free energy of small molecules in water	642	1	RMSE
Lipophilicity	Octanol/water distribution coefficient of molecules	4200	1	RMSE

To assess the agnostic nature of our methodology, we conducted the experiments considering foundation models which are based on different backbones including: Transformers-based models MoLFormer [22], and Mussurana [32], graph-based MHG-GNN [23], and the recently introduced state-space model (SSM), MoLMamba [33]. Experiments were conducted with fine-tuned and frozen models. Detailed results, hyper-parameters, training strategies, and the list of extracted causal features per task are provided in the supplementary materials due to limit of pages.

4 Results and Discussion

The Results and Discussion section of this paper is structured into two subsections. Firstly, we detail the outcomes of our methodology with we present the results with frozen models which are based

on different architectures to elucidate the agnostic claim of our causal enhancement of chemical foundation models. We also present our results with fine-tuned models to elucidate the potential of our methodology to improve state-of-the-art results.

4.1 Evaluation of frozen foundation models

In this subsection, we compare models from different architectures, including transformers, graphs, and SSMs, to evaluate their performance when causal feature are added. We employ XGBoost for downstream tasks as classification and regression, providing a detailed overview of the results obtained with the QM9 dataset, commonly used in quantum chemistry research. To ensure robustness, we utilize 10 different seeds for each scenario. Additionally, we conduct an ablation study comparing causal features derived from Mordred with the original feature set. Detailed results for each seed are available in the Supplementary Materials.

In Table 3, we present the results of our classification tasks analysis, focusing on the impact of causal features derived from Mordred descriptors. We found that the selected causal features led to performance improvements in 3 out of 5 tasks considered when compared to the original Mordred descriptors set. For instance, in the biodegradability task, just 5 features (refer to Fig. 3) enhanced the area under the ROC curve (AUC) from 84.0 to 88.84 ±0.3. Table 2 provides insight into the number of features selected for various tasks utilizing a multi-stage causal feature selection methodology. The original dataset comprised 1826 Mordred features, and through the application of this approach, dimensionality was effectively reduced. A significance level of p-value = 0.05 was utilized for the causal selector. The table showcases the diversity in feature selection across different datasets, highlighting the adaptability of the approach to individual task requirements. The full list of selected causal features per task are given in the Supplementary Materials.

Table 2: Number of features selected for each task during the multi-stage causal feature selection. For QM8 and QM9 the values reflects the average number of features per task. See the Supplementary Materials for more details.

Dataset	# Features
BBBP	5
HIV	12
BACE	6
LDToxDB	29
"All-Public set"	5
QM9	~ 396
QM8	~ 253
ESOL	402
FreeSolv	383
Lipophilicity	61

Furthermore, our investigation revealed that incorporating causal features over the embedding space improved the performance of transformers and SSM-based models in 4 out of 5 classification tasks. Despite the large embedding space of foundation models, only a few features were needed to achieve improved classification performance (see Fig. 3). In Fig. 4, we explore visually the changes in the latent space of the different foundation models when the causal features are added to it. In this case, we demonstrate the enhancement in class separation within the BBBP dataset upon the addition of 5 causal features. It is important to highlight that MHG-GNN model has an embedding space of size 2048, whereas the other models are equipped with an embedding space of size 768. This disparity in embedding space size may influence the impact of the causal features.

Table 4 reveals that causal features were pivotal in improving the prediction performance of foundation models in 5 out of 6 tasks considered. This performance improvements were consistent across different architectures and models. The inclusion of causal features in the MoLFormer embedding space improved the average MAE from 12.42 ± 0.1175 to 2.3568 ± 0.0213 in the challenging QM9 dataset, which involves 12 different tasks related to the quantum properties of molecules. This improvement for the QM9 dataset was consistent across all models considered in this experiment, as detailed in Table 5.

In Table 5, we explore the results for individual tasks within the QM9 benchmark dataset, aiming to uncover nuanced insights to each specific molecular property. The inclusion of causal features in the embedding space of chemical foundation models proves to be crucial for improving performance across all 12 quantum measures in the QM9 dataset for all the tested foundation models. Causal

Table 3: Results for the classification tasks. ↑ indicates where the causal features helped to improve the base model. In this case, we are using the AUC-ROC metric to evaluate the methods, so greater values indicates better performance.

Model	BBBP	HIV	BACE	LDToxDB	Biodegradability
Mordred Descriptors	72.90	79.20	86.70	66.00	84.00
Causal Features	90.36 ±0.4 ↑	79.68 ±0.87 ↑	84.95 ±0.57	57.44 ±0.2	88.84 ±0.3 ↑
MoLFormer	91.72 ±0.51	80.5 ±1.65	86.23 ±0.82	84.36 ±0.11	90.01 ±0.22
MoLFormer + Causal Features	91.91 ±0.61 ↑	81.01 ±1.06 ↑	86.8 ±0.47 ↑	84.25 ±0.08	90.88 ±0.31 ↑
Mussurana	91.45 ±0.78	81.64 ±1	85.22 ±0.63	85.17 ±0.11	90.32 ±0.32
Mussurana + Causal Features	91.94 ±0.56 ↑	81.79 ±1.22 ↑	86.03 ±0.82 ↑	85.12 ±0.13	91.86 ±0.36 ↑
MHG-GNN	93.33 ±0.57	84.69 ±0.72	88.28 ±0.82	86.26 ±0.07	90.49 ±0.33
MHG-GNN + Causal Features	93.35 ±0.38	85.13 ±0.95 ↑	88.12 ±0.46	86.05 ±0.12	90.9 ±0.24 ↑
MoLMamba	88 ±0.76	73.27 ±1.07	80.97 ±1.14	80.78 ±0.14	85.93 ±0.43
MoLMamba + Causal Features	91.55 ±0.5 ↑	77.15 ±0.71 ↑	82.62 ±1.6 ↑	80.6 ±0.08	88.58 ±0.36 ↑

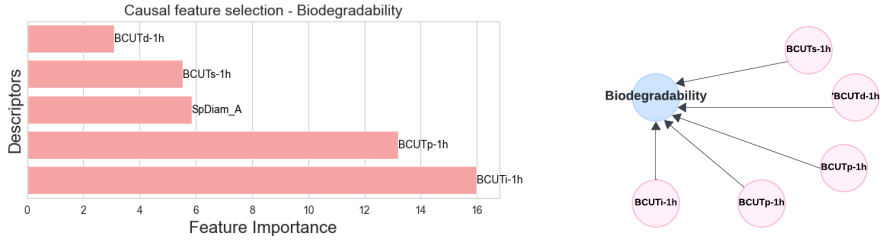

Figure 3: The figure illustrates the Markov Blanket feature importance and graph given by the causality feature selection process for the biodegradabilty task.

features enhance the performance of foundation models for all twelve properties of QM9, with a particular focus on enhancing the $\langle R^2 \rangle$ property.

These experiments underscore the potential benefits of causal features for accurately predicting molecular properties using chemical foundation models. Furthermore, they highlight the agnostic nature of our proposed approach, as it successfully enhances performance across foundation models of different types, including transformers, graphs, and SSM-based models. For a visual representation of the impact of including causal features on the performance of chemical foundation models, refer

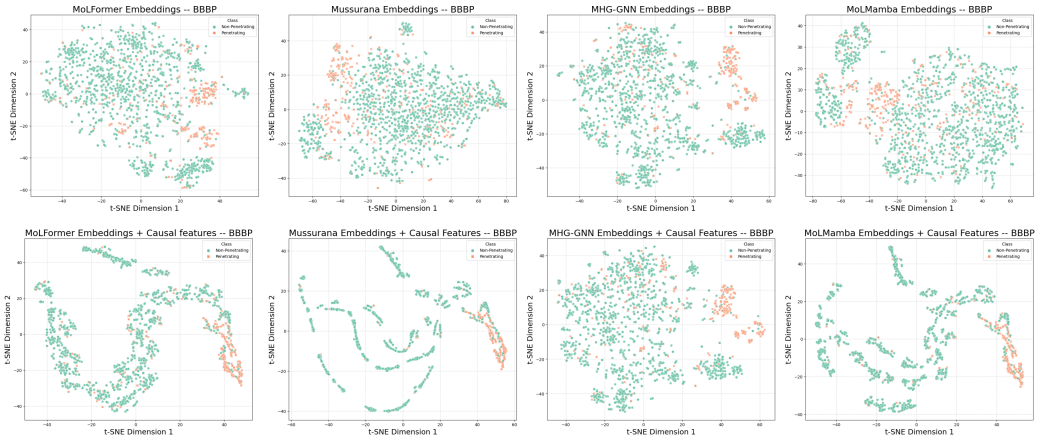

Figure 4: This figure illustrates how incorporating causal features into the embedding space affects various foundational models with different architectures. Specifically, it demonstrates the changes in the latent space concerning the BBBP dataset. The first row displays the original latent space for MoLFormer, Mussurana, MHG-GNN, and MoLMamba. In the second row, causal features are included in the embedding space of these models. For a clearer visualization, refer to the color version of the figure.

Table 4: Results for the regression tasks including base model and causal features. ↓ indicates where the causal features improved the results for regression tasks. For QM9 and QM8 datasets we considered the average MAE metric. For the others dataset RMSE is used. Therefore, lower numbers indicates better performance.

Model	ESOL	FreeSolv	Lipophilicity	LDToxDB	QM8	QM9
Mordred Descriptors	0.9900	1.7400	0.7990	0.5230	0.0150	4.3500
Causal Features	0.549 ±0.0233 ↓	1.1143 ±0.1057 ↓	0.5835 ±0.016 ↓	0.8863 ±0.0053	0.0132 ±0.0007 ↓	2.1829 ±0.0107 ↓
MoLFormer	0.9315 ±0.025	1.8551 ±0.0774	0.6704 ±0.0166	0.5892 ±0.0073	0.0205 ±0.0004	12.42 ±0.1175
MoLFormer + Causal Features	0.6079 ±0.0119 ↓	1.2087 ±0.0787 ↓	0.578 ±0.0198 ↓	0.5918 ±0.0096	0.0143 ±0.0006 ↓	2.3568 ±0.0213 ↓
Mussurana	0.7 ±0.0416	1.6545 ±0.0593	0.6495 ±0.0047	0.5639 ±0.0046	0.018 ±0.0006	7.7379 ±0.0485
Mussurana + Causal Features	0.6141 ±0.018 ↓	1.3333 ±0.0906 ↓	0.5757 ±0.0088 ↓	0.5668 ±0.0043	0.0144 ±0.0008 ↓	2.2197 ±0.0099 ↓
MHG-GNN	0.6452 ±0.0482	1.4247 ±0.0815	0.5864 ±0.0119	0.5209 ±0.0044	0.0135 ±0.0006	4.1537 ±0.0473
MHG-GNN + Causal Features	0.5433 ±0.0258 ↓	1.1787 ±0.0579 ↓	0.5508 ±0.0154 ↓	0.5261 ±0.0046	0.0126 ±0.0007 ↓	2.0911 ±0.0148 ↓
MoLMamba	0.9868 ±0.049	2.2197 ±0.0726	0.8103 ±0.0055	0.6447 ±0.0033	0.0215 ±0.0004	11.0362 ±0.0831
MoLMamba + Causal Features	0.6187 ±0.0202 ↓	1.2945 ±0.0966 ↓	0.6142 ±0.0132 ↓	0.646 ±0.0031	0.0141 ±0.0006 ↓	2.3079 ±0.0267 ↓

to Fig 5. This figure provides an illustration of how causal features contribute to improving the performance of foundation models on the QM9 dataset.

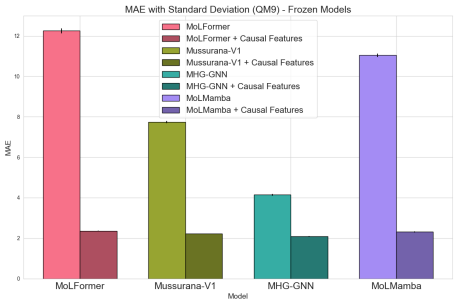

Figure 5: This figure illustrates the performance enhancement observed across different foundation models upon the inclusion of causal features, as demonstrated using the QM9 dataset.

Table 5: Detailed QM9 results comparing models with varying architectures, with and without causal features. ↓ indicates where causal features contributed to enhancing the base model's performance. Our evaluation employs the mean MAE metric across tasks, where lower values signify superior performance.

	Transformers-based				Graph-based		SSM-based	
Measure	MoLFormer	MoLFormer + Causal Features	Mussurana	Mussurana + Causal Features	MHG-GNN	MHG-GNN + Causal Features	MoLMamba	MoLMamba + Causal Features
α	2.8903	0.5272 ↓	1.6210	0.4986 ↓	0.8669	0.4538 ↓	2.2520	0.5119 ↓
C_v	1.1767	0.2117 ↓	0.6793	0.2138 ↓	0.3704	0.1948 ↓	1.0574	0.2165 ↓
G	13.11	0.0270 ↓	6.8212	0.0406 ↓	1.8261	0.0223 ↓	11.6328	0.0257 ↓
gap	0.0139	0.0061 ↓	0.0122	0.0076 ↓	0.0067	0.0063 ↓	0.0168	0.0069 ↓
H	12.8254	0.0244 ↓	6.9112	0.0431 ↓	1.8220	0.0236 ↓	11.4702	0.0275 ↓
ϵ_{homo}	0.0080	0.0046 ↓	0.0087	0.0063 ↓	0.0051	0.0050 ↓	0.0099	0.0052 ↓
ϵ_{lumo}	0.01256	0.0050 ↓	0.0104	0.0067 ↓	0.0052	0.0051 ↓	0.0156	0.0055 ↓
μ	0.7102	0.5360 ↓	0.6515	0.5330 ↓	0.5237	0.4833 ↓	0.7461	0.5453 ↓
$\langle R^2 \rangle$	94.70	26.83 ↓	62.49	25.20 ↓	40.80	23.83 ↓	82.15	26.26 ↓
U_0	12.979	0.0265 ↓	6.8063	0.0370 ↓	1.8252	0.0217 ↓	11.5451	0.0288 ↓
U	12.8618	0.0242 ↓	6.7895	0.0370 ↓	1.8286	0.0223 ↓	11.6206	0.0266 ↓
ZPVE	0.0086	0.0005 ↓	0.0060	0.0009 ↓	0.0017	0.0006 ↓	0.009	0.0005 ↓
Avg MAE	12.42	2.3568 ↓	7.7379	2.2197 ↓	4.1537	2.0911 ↓	11.0362	2.3079 ↓
Avg STD	±0.1175	±0.0213 ↓	±0.0485	±0.0099 ↓	±0.0473	±0.0148 ↓	±0.0831	±0.0267 ↓

In both the classification and regression experiments, we observed that causal features played a significant role in enhancing the performance of chemical foundation models, even if they are based on different backbones.

4.2 Evaluation of fine-tuned foundation models

Here, we present a comparative analysis of state-of-the-art benchmarks and the inclusion of causal features to improve performance of chemical foundation models. Table 6 displays the benchmark results from the literature for state-of-the-art models. In addition to these benchmarks, we also report the results of fine-tuned MoLFormer with the inclusion of causal features. To guarantee the

robustness of the approach, we evaluated our experiments considering 10 different seeds, results for each seed is presented on the Supplementary Materials.

Table 6: Results for the regression tasks including benchmark and MoLFormer with causal features. ↓ indicates where the causal features improved the base model for regression tasks. For QM9 and QM8 datasets we considered the avg MAE metric. For the others dataset RMSE is used, lower numbers indicates better performance. **Bold** indicates state-of-the-art results.

Method	Dataset					
	QM9	QM8	LDToxDB	ESOL	FreeSolv	Lipophilicity
GC [34]	4.35	0.0148	-	0.97	1.40	0.65
A-FP [35]	2.63	0.0282	-	0.50	0.74	0.58
$GROVER_{Large}$ [36]	-	-	-	0.89	2.27	0.82
Padel-DNN [37]	-	-	-	0.62	0.91	-
ChemRL-GEM [38]	-	-	-	0.80	1.88	0.66
ChemBERTa-2 [2]	-	-	-	0.89	-	0.80
SPMM [39]	-	-	-	0.82	1.90	0.69
Uni-Mol [40]	-	0.0156	-	0.79	1.48	0.60
MPNN [41]	3.18	0.0143	-	0.58	1.15	0.72
MoLFormer-XL [22]	1.59	0.0102	-	**0.28**	**0.23**	0.53
MoLFormer [22]	2.25	0.0111	0.62	0.28	0.26	0.62
MoLFormer + Causal Features	**1.47±0.0037** ↓	**0.0098±0.0001** ↓	**0.60±0.0019** ↓	0.58±0.0167	1.45±0.0606	**0.51±0.0071** ↓

The findings presented in Table 6 underscore the positive impact of integrating causal features into the foundation model. The integration of causal features into the foundation embedding space resulted in a performance superior to its state-of-the-art counterparts, including MoLFormer-XL. Despite being trained on a significantly smaller dataset of 100 million samples, compared to MoLFormer-XL's 1 billion samples, the inclusion of select causal features, tailored to the specific tasks (refer to Supplementary Materials for the complete list of causal features per task), enabled the smaller model to achieve state-of-the-art performance. This highlights the efficacy of causal features in enhancing the model's predictive capabilities. Moreover, it opens up opportunity to investigate the improvement on larger models when such domain features are added.

Table 7 also indicates that the inclusion of causal/domain features to the base MoLFormer model was important to improve the performance of the base model to classification tasks. The inclusion of causal features can also be beneficial to larger models.

Table 7: Results for the classification tasks. ↑ indicates where the causal features helped to improve the base model. In this case, we are using the AUC-ROC metric to evaluate the methods, so greater value is better.

Method	Dataset				
	BBBP	HIV	BACE	LDToxDB	Biodegradability
MoLFormer [22]	90.9	77.7	82.8	75.23	84.35
MoLFormer-XL [22]	93.7	82.2	88.2	–	–
MoLFormer + Causal Features	92.29±1.19 ↑	77.05±0.93	84.56 ±0.7 ↑	80.81±0.13 ↑	91.38±0.3 ↑

Our results shows highlight the importance of causal features to significantly improve the performance of chemical foundation models in predicting various chemical properties, regardless of the model's architecture. This approach offers a promising avenue for researchers to develop more powerful and interpretable tools for scientific discovery, with potential applications beyond chemistry.

5 Conclusion

In this paper, we have introduced a novel multi-stage causal feature selection approach aimed at enhancing the predictive performance of chemical foundation models. Our results underscore the effectiveness of this method, showcasing improvements across diverse model architectures, including transformers, graphs, and state-space-based models.

To ensure the reliability of our findings, we conducted comprehensive evaluations across multiple datasets, employing a methodology that encompassed 10 different seeds for each classification and regression task. Detailed results for each seed are available in the Supplementary Materials.

These findings offer a promising direction for enhancing the performance of foundation models for molecular property prediction. By integrating both chemical foundation models and causal physico-

chemical features, our approach has the potential to make significant contributions to drug discovery, materials science, and other fields that depend on accurate prediction of molecular properties.

References

[1] A. D. White, "The future of chemistry is language," *Nature Reviews Chemistry*, vol. 7, no. 7, pp. 457–458, 2023.

[2] W. Ahmad, E. Simon, S. Chithrananda, G. Grand, and B. Ramsundar, "Chemberta-2: Towards chemical foundation models," *arXiv preprint arXiv:2209.01712*, 2022.

[3] R. Irwin, S. Dimitriadis, J. He, and E. J. Bjerrum, "Chemformer: a pre-trained transformer for computational chemistry," *Machine Learning: Science and Technology*, vol. 3, no. 1, p. 015022, 2022.

[4] R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. S. Bernstein, J. Bohg, A. Bosselut, E. Brunskill *et al.*, "On the opportunities and risks of foundation models," *arXiv preprint arXiv:2108.07258*, 2021.

[5] E. Soares, F. Cipcigan, D. Zubarev, and E. V. Brazil, "A framework for toxic pfas replacement based on gflownet and chemical foundation model," in *NeurIPS 2023 AI for Science Workshop*, 2023.

[6] B. Zoph, G. Ghiasi, T.-Y. Lin, Y. Cui, H. Liu, E. D. Cubuk, and Q. Le, "Rethinking pre-training and self-training," *Advances in neural information processing systems*, vol. 33, pp. 3833–3845, 2020.

[7] L. Yang, H. Chen, Z. Li, X. Ding, and X. Wu, "Chatgpt is not enough: Enhancing large language models with knowledge graphs for fact-aware language modeling," *arXiv preprint arXiv:2306.11489*, 2023.

[8] B. Li, D. Hwang, Z. Huo, J. Bai, G. Prakash, T. N. Sainath, K. C. Sim, Y. Zhang, W. Han, T. Strohman *et al.*, "Efficient domain adaptation for speech foundation models," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[9] K. Yang, K. Swanson, W. Jin, C. Coley, P. Eiden, H. Gao, A. Guzman-Perez, T. Hopper, B. Kelley, M. Mathea *et al.*, "Analyzing learned molecular representations for property prediction," *Journal of chemical information and modeling*, vol. 59, no. 8, pp. 3370–3388, 2019.

[10] E. Soares, E. V. Brazil, K. F. A. Gutierrez, R. Cerqueira, D. Sanders, K. Schmidt, and D. Zubarev, "Beyond chemical language: A multimodal approach to enhance molecular property prediction," *arXiv preprint arXiv:2306.14919*, 2023.

[11] E. Soares, A. Kishimoto, E. V. Brazil, S. Takeda, H. Kajino, and R. Cerqueira, "Improving molecular properties prediction through latent space fusion," *arXiv preprint arXiv:2310.13802*, 2023.

[12] H. Moriwaki, Y.-S. Tian, N. Kawashita, and T. Takagi, "Mordred: a molecular descriptor calculator," *Journal of cheminformatics*, vol. 10, no. 1, pp. 1–14, 2018.

[13] C. W. Yap, "Padel-descriptor: An open source software to calculate molecular descriptors and fingerprints," *Journal of computational chemistry*, vol. 32, no. 7, pp. 1466–1474, 2011.

[14] H. Y. I. Lam, R. Pincket, H. Han, X. E. Ong, Z. Wang, J. Hinks, Y. Wei, W. Li, L. Zheng, and Y. Mu, "Application of variational graph encoders as an effective generalist algorithm in computer-aided drug design," *Nature Machine Intelligence*, vol. 5, no. 7, pp. 754–764, 2023.

[15] A. E. Comesana, T. T. Huntington, C. D. Scown, K. E. Niemeyer, and V. H. Rapp, "A systematic method for selecting molecular descriptors as features when training models for predicting physiochemical properties," *Fuel*, vol. 321, p. 123836, 2022.

[16] C. F. Aliferis, I. Tsamardinos, and A. Statnikov, "Hiton: a novel markov blanket algorithm for optimal variable selection," in *AMIA annual symposium proceedings*, vol. 2003. American Medical Informatics Association, 2003, p. 21.

[17] I. Guyon, C. Aliferis *et al.*, "Causal feature selection," in *Computational methods of feature selection*. Chapman and Hall/CRC, 2007, pp. 79–102.

[18] X. Wu, B. Jiang, K. Yu, H. Chen, and C. Miao, "Multi-label causal feature selection," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 04, 2020, pp. 6430–6437.

[19] X. Guo, K. Yu, L. Liu, F. Cao, and J. Li, "Causal feature selection with dual correction," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[20] S. Magliacane, T. Van Ommen, T. Claassen, S. Bongers, P. Versteeg, and J. M. Mooij, "Domain adaptation by using causal inference to predict invariant conditional distributions," *Advances in neural information processing systems*, vol. 31, 2018.

[21] C. F. Aliferis, A. Statnikov, I. Tsamardinos, S. Mani, and X. D. Koutsoukos, "Local causal and markov blanket induction for causal discovery and feature selection for classification part i: algorithms and empirical evaluation." *Journal of Machine Learning Research*, vol. 11, no. 1, 2010.

[22] J. Ross, B. Belgodere, V. Chenthamarakshan, I. Padhi, Y. Mroueh, and P. Das, "Large-scale chemical language representations capture molecular structure and properties," *Nature Machine Intelligence*, vol. 4, no. 12, pp. 1256–1264, 2022.

[23] A. Kishimoto, H. Kajino, M. Hirose, J. Fuchiwaki, I. Priyadarsini, L. Hamada, H. Shinohara, D. Nakano, and S. Takeda, "Mhg-gnn: Combination of molecular hypergraph grammar with graph neural network," *arXiv preprint arXiv:2309.16374*, 2023.

[24] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," *arXiv preprint arXiv:2312.00752*, 2023.

[25] Z. Wu, B. Ramsundar, E. N. Feinberg, J. Gomes, C. Geniesse, A. S. Pappu, K. Leswing, and V. Pande, "Moleculenet: a benchmark for molecular machine learning," *Chemical science*, vol. 9, no. 2, pp. 513–530, 2018.

[26] M. Lee and K. Min, "A comparative study of the performance for predicting biodegradability classification: the quantitative structure–activity relationship model vs the graph convolutional network," *ACS omega*, vol. 7, no. 4, pp. 3649–3655, 2022.

[27] J. Feinstein, G. Sivaraman, K. Picel, B. Peters, Á. Vázquez-Mayagoitia, A. Ramanathan, M. MacDonell, I. Foster, and E. Yan, "Uncertainty-informed deep transfer learning of perfluoroalkyl and polyfluoroalkyl substance toxicity," *Journal of chemical information and modeling*, vol. 61, no. 12, pp. 5793–5803, 2021.

[28] G. Landrum, "Rdkit documentation," *Release*, vol. 1, no. 1-79, p. 4, 2013.

[29] B. Hollas, "An analysis of the autocorrelation descriptor for molecules," *Journal of mathematical chemistry*, vol. 33, pp. 91–101, 2003.

[30] D. Koller and M. Sahami, "Toward optimal feature selection," Stanford InfoLab, Tech. Rep., 1996.

[31] A. Hassan, J. H. Paik, S. Khare, and S. A. Hassan, "Ppfs: Predictive permutation feature selection," *arXiv preprint arXiv:2110.10713*, 2021.

[32] E. Soares, V. Shirasuna, E. V. Brazil, R. Cerqueira, D. Zubarev, and K. Schmidt, "A large encoder-decoder family of foundation models for chemical language," *arXiv preprint arXiv:2407.20267*, 2024.

[33] E. A. Soares, V. Shirasuna, E. A. V. Brazil, R. F. de Gusmao Cerqueira, D. Zubarev, T. Callahan, and S. Capponi, "Molmamba: A large state-space-based foundation model for chemistry," in *American Chemical Society (ACS) Fall Meeting*, 2024.

[34] H. Altae-Tran, B. Ramsundar, A. S. Pappu, and V. Pande, "Low data drug discovery with one-shot learning," *ACS central science*, vol. 3, no. 4, pp. 283–293, 2017.

[35] Z. Xiong, D. Wang, X. Liu, F. Zhong, X. Wan, X. Li, Z. Li, X. Luo, K. Chen, H. Jiang *et al.*, "Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism," *Journal of medicinal chemistry*, vol. 63, no. 16, pp. 8749–8760, 2019.

[36] Y. Rong, Y. Bian, T. Xu, W. Xie, Y. Wei, W. Huang, and J. Huang, "Self-supervised graph transformer on large-scale molecular data," *Advances in Neural Information Processing Systems*, vol. 33, pp. 12 559–12 571, 2020.

[37] K. Zhang and H. Zhang, "Predicting solute descriptors for organic chemicals by a deep neural network (dnn) using basic chemical structures and a surrogate metric," *Environmental Science & Technology*, vol. 56, no. 3, pp. 2054–2064, 2022.

[38] X. Fang, L. Liu, J. Lei, D. He, S. Zhang, J. Zhou, F. Wang, H. Wu, and H. Wang, "Geometry-enhanced molecular representation learning for property prediction," *Nature Machine Intelligence*, vol. 4, no. 2, pp. 127–134, 2022.

[39] J. Chang and J. C. Ye, "Bidirectional generation of structure and properties through a single molecular foundation model," 2023.

[40] G. Zhou, Z. Gao, Q. Ding, H. Zheng, H. Xu, Z. Wei, L. Zhang, and G. Ke, "Uni-mol: a universal 3d molecular representation learning framework," 2023.

[41] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," in *International conference on machine learning*. PMLR, 2017, pp. 1263–1272.

[42] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, R. Mitchell, I. Cano, T. Zhou *et al.*, "Xgboost: extreme gradient boosting," *R package version 0.4-2*, vol. 1, no. 4, pp. 1–4, 2015.

[43] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 2623–2631.

A Supplementary Materials

Here, we provide detailed results for every experiment conducted in this paper. To ensure the robustness of our claims, we conducted experiments with 10 different seeds. For training models with frozen weights, we utilized XGBoost [42] as the learner and Optuna [43] for hyper-parameter optimization. All experiments with frozen models were conducted using a single NVIDIA V100 (32G) GPU. The multi-stage causal features selection took approximately **8 hours** to process all the datasets studied in this paper.

For fine-tuning MoLFormer, we employed a fully connected network with 2 layers. Table 8 provides a detailed overview of the hyper-parameters considered for fine-tuning $\text{Onyx}_{\text{SMILES}}$. We used a single NVIDIA V100 (32G) GPU for this task. Each experiment was run with 10 different seeds to ensure the robustness of the results.

Table 8: MoLFormer fine-tuning architecture specificity.

Hidden size	Attention heads	Layers	Dropout	Normalization
768	12	12	0.2	LayerNorm

Learning rate	# batch	# epochs	# tokens	# GPUs	Total params
3e-5	32	500	202	1 NVIDIA V100 (32G)	47M

A.1 Detailed results - frozen weights

Here, we detail the results for both classification and regression tasks conducted using features selected during the causal multi-stage process. Table 9 illustrates the results specifically for classification tasks, with ROC-AUC used as the evaluation metric.

Table 9: Classification results for 10 different seeds considering the causal features selected during the multi-stage process.

	ROC-AUC↑				
Seed	**BBBP**	**HIV**	**BACE**	**LDToxDB**	**Biodegradability**
0	90.40	79.30	84.41	57.42	88.93
10	90.64	79.81	85.43	56.93	88.85
20	89.70	79.73	84.69	57.45	89.03
30	90.41	79.82	84.47	57.43	88.35
40	91.07	78.07	85.68	57.60	89.08
50	90.29	81.12	85.54	57.66	89.11
60	89.88	80.86	84.69	57.42	88.63
70	90.71	79.66	85.57	57.58	89.16
80	90.27	78.95	84.95	57.49	88.93
90	90.19	79.45	84.09	57.44	88.36
Average	90.36	79.68	84.95	57.44	88.84
Std	0.40	0.87	0.57	0.20	0.30

Table 10 elucidates the results considering the ESOL, FreeSolv, Lipophilicity, LDToxDB, QM8, QM9 datasets. Here, we use 10 different seeds considering the Mordred descriptors selected during the causal multi-stage feature selector.

A.1.1 Detailed results - MoLFormer and Causal Features

We detail the results of classification and regression tasks using MoLFormer as the base model. Table 11 illustrates the results for the classification tasks for MoLFormer and MoLFormer + Causal features, considering 10 different seeds. ↑ indicates where causal features improved the performance of the base MoLFormer model. In 4 out of the 5 tested datasets, the inclusion of causal features improved the results of MoLFormer.

Table 10: Regression results for 10 different seeds considering the causal features selected during the multi-stage process.

	RMSE				MAE	
Seed	**ESOL**	**FreeSolv**	**Lipophilicity**	**LDToxDB**	**QM8**	**QM9**
0	0.5843	1.3000	0.6239	0.8845	0.0141	2.1800
10	0.5592	1.1214	0.5727	0.8882	0.0133	2.1954
20	0.5331	0.9681	0.5850	0.8928	0.0129	2.1700
30	0.5449	1.1121	0.5885	0.8807	0.0127	2.1791
40	0.5196	1.1369	0.5797	0.8952	0.0135	2.1841
50	0.5315	1.1626	0.5765	0.8829	0.0140	2.1789
60	0.5380	1.2299	0.5712	0.8786	0.0131	2.1879
70	0.5325	1.0927	0.5677	0.8833	0.0137	2.1961
80	0.5583	0.9548	0.5797	0.8871	0.0123	2.1641
90	0.5890	1.0640	0.5901	0.8896	0.0121	2.1937
Average	0.5490	1.1143	0.5835	0.8863	0.0132	2.1829
Std	0.0233	0.1057	0.0160	0.0053	0.0007	0.0107

Table 11: Results for 10 different seeds considering MoLFormer and causal features for classification tasks.

		ROC-AUC				
Seed	**Model**	**BBBP**	**HIV**	**BACE**	**LDToxDB**	**Biodegradability**
0	MoLFormer	91.33	81.18	86.39	84.41	89.83
10		92.22	80.59	86.85	84.39	90.10
20		92.24	80.96	85.49	84.12	90.45
30		91.60	81.01	85.88	84.30	89.83
40		92.56	78.96	85.55	84.31	89.84
50		91.62	77.27	86.85	84.36	90.08
60		91.11	83.27	87.82	84.55	90.13
70		91.51	81.94	85.14	84.30	90.13
80		91.94	80.31	86.56	84.44	89.98
90		91.02	79.55	85.75	84.38	89.71
Average		91.72	80.50	86.23	84.36	90.01
Std		0.51	1.65	0.82	0.11	0.22
0	MoLFormer + Causal Features	93.08	79.73	87.67	84.26	91.24
10		92.17	80.81	86.83	84.26	91.30
20		91.58	80.35	86.10	84.40	90.42
30		91.79	81.88	86.23	84.30	90.81
40		90.98	83.23	86.59	84.12	90.81
50		91.42	80.22	87.11	84.14	90.72
60		92.68	81.55	86.48	84.17	90.48
70		91.74	80.60	87.16	84.28	90.98
80		91.90	81.64	86.98	84.28	91.27
90		91.76	80.10	86.87	84.28	90.80
Average		**91.91**↑	**81.01**↑	**86.80**↑	84.25	**90.88**↑
Std		0.61	1.06	0.47	0.08	0.31

Table 12 illustrates the results for the regression tasks using MoLFormer and MoLFormer + Causal features, considering 10 different seeds. ↑ indicates where the causal features improved the performance of the base MoLFormer model. Causal features improved the results of the base MoLFormer in 5 out of the 6 datasets considered, as indicated by ↓. Specifically, in the QM9 dataset, which refers to the quantum properties of molecules, causal features were necessary to improve the results from 12.42 to 2.35.

A.1.2 Detailed results - Mussurana and Causal Features

Table 13 presents a comparison of results for molecular classification tasks between Mussurana and Mussurana + Causal features. To ensure result consistency, experiments were conducted with 10 different seeds. ↑ indicates where causal features improved the performance of Mussurana. In 4 out of the 5 tested datasets, the inclusion of causal features led to improvements in Mussurana base model results.

Similarly, Table 14 presents the results for molecular properties prediction, including quantum mechanics of molecules considered in the QM9 dataset. ↓ indicates where causal features were important for improving prediction performance. The table highlights the importance of including domain features to enhance the predictions of large foundation models. In the case of QM9, the performance improved from 7.73 to 2.21 with the inclusion of causal domain features.

A.1.3 Detailed results - MHG-GNN and Causal Features

Here, we consider the graph-based foundation model for molecules, MHG-GNN. Similar to other language-based foundation models, graph-based models also present an improvement in performance when causal domain

Table 12: Results for 10 different seeds considering MoLFormer and causal features for regression tasks

		RMSE				MAE	
Seed	**Model**	**ESOL**	**FreeSolv**	**Lipophilicity**	**LDToxDB**	**QM8**	**QM9**
0	MoLFormer	0.9099	1.8060	0.6818	0.5911	0.0205	12.4670
10		0.9286	1.9547	0.6870	0.5867	0.0207	12.3987
20		0.9212	1.8190	0.6583	0.5821	0.0199	12.6085
30		0.9061	1.7955	0.6559	0.5846	0.0202	12.4222
40		0.9392	1.8913	0.6896	0.5892	0.0211	12.4555
50		0.9405	1.9680	0.6537	0.5968	0.0204	12.2394
60		0.9598	1.7376	0.6637	0.5969	0.0210	12.3452
70		0.9470	1.9220	0.6540	0.5796	0.0207	12.5971
80		0.9716	1.8674	0.6630	0.5836	0.0206	12.3353
90		0.8914	1.7898	0.6969	0.6016	0.0198	12.3308
Average		0.9315	1.8551	0.6704	0.5892	0.0205	12.4200
Std		0.0250	0.0774	0.0166	0.0073	0.0004	0.1175
0	MoLFormer + Causal Features	0.6303	1.3044	0.6255	0.6023	0.0152	2.3731
10		0.5979	1.2341	0.5661	0.5803	0.0143	2.3412
20		0.6153	1.2266	0.5831	0.5827	0.0132	2.3524
30		0.5981	1.2509	0.5648	0.5813	0.0141	2.3680
40		0.6223	1.0164	0.5879	0.5931	0.0151	2.3782
50		0.5916	1.1872	0.5880	0.5926	0.0143	2.3196
60		0.6024	1.2184	0.5644	0.5987	0.0142	2.3349
70		0.6080	1.1921	0.5598	0.5819	0.0141	2.3886
80		0.6041	1.1774	0.5764	0.5999	0.0142	2.3485
90		0.6090	1.2798	0.5640	0.6053	0.0142	2.3631
Average		**0.6079**↓	**1.2087**↓	**0.5780**↓	0.5918	**0.0143**↓	**2.3568**↓
Std		0.0119	0.0787	0.0198	0.0096	0.0006	0.0213

Table 13: Results for 10 different seeds considering Mussurana and causal features for classification tasks

		ROC-AUC				
Seed	**Model**	**BBBP**	**HIV**	**BACE**	**LDToxDB**	**Biodegradability**
0	**Mussurana-V1**	91.12	82.70	85.31	85.22	90.33
10		91.55	82.47	85.56	85.19	90.82
20		91.72	82.43	85.20	85.28	89.88
30		91.33	80.85	84.78	85.26	90.54
40		92.39	80.31	84.28	85.13	90.26
50		90.35	81.09	84.45	85.20	89.72
60		91.08	81.25	85.93	85.10	90.16
70		92.92	80.90	85.71	84.89	90.43
80		91.50	81.07	84.83	85.21	90.43
90		90.49	83.32	86.15	85.18	90.47
Average		91.45	81.64	85.22	85.17	90.30
Std		0.78	1.00	0.63	0.11	0.32
0	**Mussurana-V1**	91.86	81.75	86.01	85.13	90.84
10		91.51	83.64	86.43	84.81	90.54
20		92.07	79.58	85.38	85.15	90.34
30		90.91	81.37	86.23	85.19	90.36
40		92.81	83.72	85.97	85.14	90.80
50		92.20	81.53	85.58	84.98	90.42
60		92.69	82.07	87.84	85.25	90.08
70		91.86	81.77	85.40	85.22	90.08
80		91.57	81.76	84.89	85.21	90.25
90		91.92	80.74	86.59	85.11	89.91
Average		**91.94**↑	**81.79**↑	**86.03**↑	85.12	**90.36**↑
Std		0.56	1.22	0.82	0.13	0.30

features are added to the original embedding space. Table 15 illustrates the results for the classification tasks. In this case, we observe performance improvements in 3 out of the 5 datasets tested.

We consider MHG-GNN for the prediction of molecular properties. Similar to language-based foundation models, the graph-based model also presents performance improvements when causal domain features are added to the original embedding space. Table 16 illustrates the prediction results for the different datasets considering 10 different seeds. In this case, we observe performance improvements in 5 out of the 6 datasets tested, including QM8 and QM9..

A.1.4 Detailed results - MoLMamba and Causal Features

Finally, we detail the results for the state-space molecular foundation model. Table 17 illustrates the results for the classification tasks for MoLMamba, both with and without the inclusion of causal domain features. In this

Table 14: Results for 10 different seeds considering Mussurana and causal features for regression tasks

		RMSE				MAE	
Seed	**Model**	**ESOL**	**FreeSolv**	**Lipophilicity**	**LDToxDB**	**QM8**	**QM9**
0	**Mussurana-V1**	0.7528	1.6526	0.6526	0.5722	0.0186	7.7058
10		0.6804	1.6651	0.6457	0.5647	0.0183	7.7341
20		0.6652	1.6906	0.6561	0.5628	0.0171	7.7171
30		0.6711	1.7309	0.6491	0.5608	0.0180	7.8124
40		0.6807	1.5840	0.6447	0.5620	0.0190	7.8022
50		0.6988	1.7138	0.6422	0.5668	0.0180	7.7055
60		0.6756	1.5975	0.6510	0.5573	0.0175	7.8033
70		0.7182	1.7051	0.6466	0.5583	0.0178	7.7082
80		0.6673	1.5539	0.6553	0.5655	0.0178	7.6881
90		0.7895	1.6512	0.6521	0.5683	0.0175	7.7021
Average		0.7000	1.6545	0.6495	0.5639	0.0180	7.7379
Std		0.0416	0.0593	0.0047	0.0046	0.0006	0.0485
0	**Mussurana-V1+ Causal Features**	0.6229	1.4083	0.5778	0.5746	0.0154	2.2121
10		0.5933	1.3928	0.5790	0.5685	0.0147	2.2192
20		0.6158	1.2665	0.5638	0.5650	0.0132	2.2189
30		0.6295	1.1609	0.5783	0.5630	0.0144	2.2233
40		0.6303	1.4330	0.5851	0.5654	0.0160	2.2318
50		0.6292	1.2606	0.5872	0.5731	0.0143	2.2091
60		0.5812	1.3127	0.5609	0.5687	0.0137	2.2113
70		0.6046	1.3637	0.5782	0.5633	0.0141	2.2397
80		0.6328	1.2905	0.5791	0.5633	0.0145	2.2203
90		0.6012	1.4437	0.5673	0.5627	0.0140	2.2112
Average		**0.6141**↓	**1.3333**↓	**0.5757**↓	0.5668	**0.0144**↓	**2.2197**↓
Std		0.0180	0.0906	0.0088	0.0043	0.0008	0.0099

Table 15: Results for 10 different seeds considering MHG-GNN and causal features for regression tasks

		ROC-AUC				
Seed	**Model**	**BBBP**	**HIV**	**BACE**	**LDToxDB**	**Biodegradability**
0	MHG-GNN	93.16	84.96	88.59	86.28	90.10
10		93.74	83.91	88.63	86.33	90.37
20		93.48	85.85	89.23	86.25	90.18
30		93.23	84.98	89.03	86.25	90.31
40		93.00	83.15	88.15	86.33	90.90
50		93.85	84.93	88.88	86.23	90.46
60		93.63	84.65	86.61	86.29	90.59
70		93.51	84.91	87.34	86.25	90.24
80		93.81	84.86	87.75	86.10	90.59
90		91.92	84.72	88.55	86.28	91.14
Average		93.33	84.69	88.28	86.26	90.49
Std		0.57	0.72	0.82	0.07	0.33
0	MHG-GNN + Causal Features	92.68	84.79	88.65	86.11	90.44
10		93.04	86.32	87.84	86.15	90.81
20		93.08	85.28	88.61	85.89	91.09
30		93.58	86.66	87.75	86.08	91.04
40		93.75	84.92	88.04	86.03	90.99
50		93.08	84.28	87.29	85.95	91.19
60		93.43	84.84	88.59	86.20	90.79
70		93.92	83.38	88.55	85.82	90.94
80		93.56	85.68	87.93	86.17	90.59
90		93.42	85.19	87.95	86.06	91.13
Average		**93.35**↑	**85.13**↑	88.12	86.05	**90.90**↑
Std		0.38	0.95	0.46	0.12	0.24

case, ↑ indicates where the causal features improved the performance of the base state-space model. In 4 out of the 5 datasets tested, the causal features improved the results of MoLMamba.

The results for molecular prediction tasks are presented in Table 18. These results indicate that causal domain features are important for performance improvement. For the QM9 dataset, the inclusion of causal domain features reduced the mean absolute error (MAE) from 11.03 to 2.30.

A.2 Detailed results - Fine-tuning

In this subsection, we detail the results for the fine-tuned version of MoLFormer, similarly to the results with frozen weights, to include causal domain features has been proved an efficient method to improve performance of foundation models. Table 19, illustrates the results of fine-tuned MoLFormer for classification tasks. Results indicates performance improvements in 4 out of 5 datasets studied when such domain features are included to the latent space.

Table 16: Results for 10 different seeds considering MHG-GNN and causal features for regression tasks

		RMSE				MAE	
Seed	**Name Model**	**ESOL**	**FreeSolv**	**Lipophilicity**	**LDToxDB**	**QM8**	**QM9**
0	MHG-GNN	0.6756	1.3299	0.5826	0.5146	0.0144	4.1583
10		0.5795	1.4935	0.5998	0.5182	0.0136	4.2037
20		0.6402	1.3765	0.5906	0.5179	0.0126	4.2175
30		0.6416	1.5780	0.5642	0.5294	0.0133	4.2011
40		0.6034	1.4884	0.5878	0.5225	0.0146	4.0553
50		0.5974	1.3661	0.5838	0.5168	0.0138	4.1329
60		0.7358	1.4552	0.5724	0.5198	0.0136	4.1449
70		0.6174	1.4133	0.5908	0.5231	0.0131	4.1574
80		0.6737	1.3170	0.5870	0.5254	0.0133	4.1226
90		0.6873	1.4286	0.6051	0.5214	0.0130	4.1435
Average		0.6452	1.4247	0.5864	0.5209	0.0135	4.1537
Std		0.0482	0.0815	0.0119	0.0044	0.0006	0.0473
0	MHG-GNN + Causal Features	0.5469	1.1599	0.5844	0.5376	0.0136	2.0980
10		0.5470	1.2835	0.5448	0.5217	0.0128	2.0944
20		0.5730	1.1372	0.5412	0.5242	0.0115	2.1089
30		0.5510	1.2474	0.5609	0.5270	0.0123	2.1059
40		0.5802	1.1802	0.5569	0.5264	0.0139	2.0948
50		0.5292	1.1727	0.5632	0.5245	0.0128	2.0646
60		0.4990	1.1343	0.5427	0.5205	0.0127	2.0723
70		0.5159	1.0797	0.5392	0.5267	0.0120	2.1039
80		0.5264	1.1983	0.5411	0.5264	0.0125	2.0896
90		0.5648	1.1941	0.5334	0.5256	0.0121	2.0789
Average		**0.5433**↓	**1.1787**↓	**0.5508**↓	0.5261	**0.0126**↓	**2.0911**↓
Std		0.0258	0.0579	0.0154	0.0046	0.0007	0.0148

Table 17: Results for 10 different seeds considering MoLMamba and causal features for classification tasks.

		ROC-AUC				
Seed	**Name Model**	**BBBP**	**HIV**	**BACE**	**LDToxDB**	**Biodegradability**
0	MoLMamba	87.54	73.50	82.89	80.72	85.91
10		88.06	74.05	82.10	80.88	85.83
20		86.65	74.41	79.42	80.94	85.35
30		89.25	71.64	80.19	80.59	86.14
40		88.20	72.77	79.88	80.84	85.94
50		88.45	72.25	82.02	80.73	86.08
60		88.81	74.89	81.80	80.53	86.84
70		87.87	72.06	80.36	80.79	85.31
80		87.21	73.76	80.56	80.93	85.89
90		87.96	73.37	80.50	80.89	86.03
Average		88.00	73.27	80.97	80.78	85.93
Std		0.76	1.07	1.14	0.14	0.43
0	MoLMamba + Causal Features	91.77	76.39	82.01	80.48	88.61
10		90.80	76.19	84.52	80.58	88.09
20		91.46	78.38	82.91	80.54	88.82
30		92.07	77.59	84.24	80.68	88.50
40		91.65	77.43	80.74	80.59	88.30
50		91.45	77.59	82.69	80.52	88.50
60		92.26	76.78	85.11	80.72	88.91
70		90.77	76.88	80.72	80.66	89.01
80		91.31	77.74	82.25	80.61	89.01
90		91.96	76.48	80.96	80.64	88.04
Average		**91.55**↑	**77.15**↑	**82.62**↑	80.60	**88.58**↑
Std		0.50	0.71	1.60	0.08	0.36

Table 20 elucidates the results for molecular properties prediction using the fine-tuned version of MoLFormer with 10 different seeds. The symbol ↓ indicates where the inclusion of causal domain features helped to improve the performance of the base MoLFormer model. In 3 out of the 5 tasks studied, the inclusion of features that have a causal effect on the targets improved the prediction performance of the base model. These improved tasks also include QM8 and QM9, which refer to the quantum properties of the molecules.

A.3 List of selected causal features per dataset

Here, we detail the list of causal Mordred features selected for the different datasets considered in this study. To select these features, we used our proposed multi-stage causal features selector for molecular descriptors. The feature process for all the datasets took approximately 8 hours considering a single NVIDIA V100 (32G) GPU. Below, we describe the selected causal features for the classification tasks:

- **BBBP**: ["MDEO-11", "TopoPSA", "PEOE VSA10", "TopoPSA(NO)", "BCUTv-1l"]

Table 18: Results for 10 different seeds considering MoLMamba and causal features for regression tasks

		RMSE				MAE	
Seed	**Model**	**ESOL**	**FreeSolv**	**Lipophilicity**	**LDToxDB**	**QM8**	**QM9**
0	**MoLMamba**	0.9060	2.2730	0.8175	0.6463	0.0214	11.1499
10		1.0085	2.2964	0.8081	0.6400	0.0217	11.0439
20		0.9459	2.0578	0.7995	0.6481	0.0210	11.1640
30		0.9622	2.1700	0.8071	0.6454	0.0213	11.0003
40		0.9312	2.2495	0.8089	0.6439	0.0222	11.0976
50		1.0171	2.2787	0.8102	0.6486	0.0214	10.9006
60		1.0188	2.2014	0.8185	0.6482	0.0219	11.0521
70		0.9958	2.2690	0.8140	0.6404	0.0215	11.0025
80		1.0664	2.1680	0.8082	0.6408	0.0214	10.9570
90		1.0160	2.2331	0.8110	0.6455	0.0210	10.9941
Average		0.9868	2.2197	0.8103	0.6447	0.0215	11.0362
Std		0.0490	0.0726	0.0055	0.0033	0.0004	0.0831
0	**MoLMamba + Causal Features**	0.6116	1.1741	0.6177	0.6509	0.0150	2.3350
10		0.6301	1.4426	0.6151	0.6479	0.0140	2.3035
20		0.6108	1.3439	0.6087	0.6443	0.0131	2.2822
30		0.6029	1.1825	0.6056	0.6440	0.0139	2.3078
40		0.6490	1.3614	0.6088	0.6438	0.0151	2.2750
50		0.6256	1.2781	0.6028	0.6463	0.0142	2.2702
60		0.6454	1.2266	0.6214	0.6399	0.0139	2.3191
70		0.6204	1.2336	0.6470	0.6488	0.0139	2.3517
80		0.5808	1.4319	0.6127	0.6477	0.0141	2.3291
90		0.6102	1.2699	0.6019	0.6465	0.0140	2.3055
Average		**0.6187**↓	**1.2945**↓	**0.6142**↓	0.6460	**0.0141**↓	**2.3079**↓
Std		0.0202	0.0966	0.0132	0.0031	0.0006	0.0267

Table 19: Results for 10 different seeds considering MoLFormer (fine-tuned) and causal features for molecular classification tasks.

		ROC-AUC				
Seed	**Model**	BBBP	HIV	BACE	LDToxDB	Biodegradability
-	MoLFormer [22]	90.9	77.7	82.8	75.23	84.35
0	MoLFormer + Causal Features	92.11	78.28	84.28	80.97	91.16
10		92.48	77.69	85.49	80.89	90.98
20		92.90	75.69	84.14	80.66	91.46
30		93.46	76.19	84.10	80.84	91.09
40		90.16	76.34	83.98	80.59	91.65
50		94.22	77.72	84.86	80.82	91.80
60		90.69	77.80	83.61	80.90	91.22
70		92.38	76.46	84.58	80.83	91.67
80		92.28	76.32	85.86	80.69	91.10
90		92.18	78.00	84.70	80.95	91.66
Average		**92.29**↑	77.05	**84.56**↑	**80.81**↑	**91.38**↑
Std		1.19	0.93	0.70	0.13	0.30

- **HIV**: ["MDEC-33", "BCUTpe-1h", "GGI4", "GGI2", "MDEC-13", "SMR VSA1"]
- **Biodegradability**: ["BCUTi-1h", "BCUTp-1h", "SpDiam A", "BCUTs-1h", "BCUTd-1h"]
- **BACE**: ["MDEC-33", "BCUTpe-1h", "GGI4", "GGI2", "MDEC-13", "SMR VSA1"]
- **LDToxDB**: ["BCUTi-1h", "BCUTd-1h", "TopoPSA(NO)", "BCUTs-1l", "VR3 D", "SMR VSA1", "BCUTv-1h", "SlogP VSA2", "BCUTd-1l", "BCUTc-1h", "SMR VSA5", "SdsssP", "IC0", "BCUTm-1l", "Mv", "BCUTi-1l", "EState VSA1", "BCUTc-1l", "TopoPSA", "PEOE VSA8", "BCUTdv-1h", "Xch-7d", "MID h", "MDEC-23", "MDEC-33", "Xch-6dv", "BCUTm-1h", "BCUTv-1l", "AMID N"]

Causal domain features for the datasets ESOL, FreeSolv, and Lipophilicity are described below:

- **Lipophilicity**: ["nAcid", "nBase", "SpDiam A", "nS", "BCUTc-1h", "BCUTc-1l", "BCUTd-1h", "BCUTd-1l", "BCUTare-1l", "BCUTp-1h", "BCUTp-1l", "SpAbs Dzv", "SpMax Dzv", "LogEE Dzv", "VR3 Dzv", "SpDiam Dzp", "SM1 Dzi", "C3SP3", "Xch-7dv", "Xc-5d", "AXp-2d", "AXp-1dv", "Mp", "AETA beta", "ETA eta", "ETA eta L", "AETA eta BR", "ETA dEpsilon D", "nHBAcc", "nHBDon", "IC0", "BIC2", "CIC3", "ZMIC1", "ZMIC4", "PEOE VSA6", "PEOE VSA13", "SMR VSA2", "SMR VSA6", "SlogP VSA2", "SlogP VSA6", "MDEC-33", "MDEO-11", "MDEN-22", "MDEN-23", "AMID N", "AMID O", "MID X", "TMPC10", "piPC6", "piPC7", "piPC8", "SLogP", "TopoPSA(NO)", "TopoPSA", "GGI1", "JGI3", "Radius", "MWC06", "TSRW10", "mZagreb2"]
- **FreeSolv**: ["SdsssP", "NdsssP", "MATS6pe", "PEOE VSA4", "GATS1c", "ATSC2p", "GATS2i", "ATSC8dv", "GATS5v", "BCUTi-1l", "VSA EState5", "VR2 A", "AATS3dv", "AATS4dv", "AATS0Z", "AATS4Z", "AATS4m", "AATS0pe", "AATS4pe", "AATS5pe", "AATS4are", "ATSC5d", "ATSC1are", "AATSC0dv", "AATSC0Z", "AATSC4pe", "GATS1v", "GATS1se", "ATSC0s",

Table 20: Results for 10 different seeds considering MoLMamba and causal features for classification tasks.

		RMSE				MAE	
Seed	**Model**	ESOL	FreeSolv	Lipophilicity	LDToxDB	QM8	QM9
-	MoLFormer[22]	0.2798	0.2596	0.6492	0.6152	0.0111	2.25
0	MoLFormer + Causal Features	0.5897	1.4609	0.5249	0.6033	0.0098	1.4786
10		0.5868	1.4978	0.5134	0.5999	0.0098	1.4758
20		0.5778	1.4385	0.5174	0.6016	0.0097	1.4796
30		0.6000	1.4707	0.5272	0.6015	0.0097	1.4726
40		0.6142	1.3385	0.5210	0.6056	0.0098	1.4808
50		0.5759	1.4867	0.5130	0.6007	0.0098	1.4738
60		0.5647	1.4749	0.5114	0.6001	0.0100	1.4694
70		0.5730	1.3966	0.5130	0.6047	0.0097	1.4803
80		0.6107	1.4395	0.5262	0.6017	0.0097	1.4759
90		0.6012	1.5649	0.5307	0.6015	0.0098	1.4787
Average		0.5894	1.4569	**0.5198**↓	**0.6021**↓	**0.0098**↓	**1.4766**↓
Std		0.0167	0.0606	0.0071	0.0019	0.0001	0.0037

"ATSC1i", "AATSC5m", "AATSC1are", "GATS2d", "GATS5Z", "GATS1p", "AATS1d", "AATS3d", "GATS2se", "AATS0dv", "AATS4se", "GATS5pe", "ATS1s", "ATS1Z", "ATSC3dv", "AATSC1d", "AATSC6s", "GATS1dv", "BCUTc-1l", "BCUTs-1h", "BCUTv-1l", "BCUTpe-1l", "BCUTare-1l", "BCUTpe-1h", "BCUTare-1h", "BCUTse-1h", "SM1 Dzv", "SpMAD Dzp", "SM1 DzZ", "SM1 Dzm", "HybRatio", "FCSP3", "AXp-0dv", "AXp-1dv", "Xch-7d", "Mse", "Mpe", "Mare", "Mp", "Mi", "ETA beta ns", "ETA eta FL", "ETA epsilon 1", "ETA epsilon 5", "ETA dEpsilon D", "ETA dBeta", "ETA eta RL", "ETA dEpsilon C", "nHBDon", "VSA EState2", "VSA EState3", "VSA EState9", "AMID", "MID O", "MPC3", "nRing", "TopoPSA(NO)", "AMW", "WPol", "mZagreb2", "JGI9", "MATS5i", "GATS4se", "LogEE Dzse", "SpMax Dzpe", "LogEE Dzpe", "MATS4p", "MATS1c", "ATSC8i", "AATSC5p", "SsssN", "SpMAD A", "nC", "AATS4d", "AATS1are", "ATSC3s", "ATSC1v", "AATSC5d", "AATSC0m", "GATS1Z", "GATS1m", "ATS0p", "ATS8p", "GATS5m", "ATS8pe", "AATS1v", "ATS4m", "ATS0are", "AATS2s", "AATS1pe", "AATSC0c", "AATS4s", "AATS3Z", "BCUTc-1h", "BCUTp-1l", "SM1 Dzse", "SM1 Dzpe", "SM1 Dzare", "SM1 Dzp", "SpDiam Dzse", "SpAbs Dzpe", "SpAbs Dzp", "SpAbs Dzi", "SpDiam Dzi", "nBondsO", "C2SP3", "AXp-1d", "Xc-5dv", "MINsCH3", "ETA beta s", "AETA beta ns", "ETA eta F", "ETA dEpsilon A", "AETA dBeta", "ETA eta B", "SIC0", "SIC1", "PEOE VSA1", "SlogP VSA2", "AMID C", "Radius", "MW", "WPath", "C3SP3", "Xp-3dv", "GATS5se", "GATS5i", "nH", "ATS3s", "AATS1se", "ATSC7d", "ATSC0p", "AATSC1m", "AATSC0v", "GATS1s", "ATS4dv", "ATSC1Z", "ATSC1pe", "ATSC0i", "ATS8are", "MATS6are", "ATS5s", "ATS1se", "ATS1are", "AATS1s", "AATS0v", "AATS0s", "ATSC0dv", "AATSC5dv", "MATS4pe", "BCUTZ-1h", "BCUTm-1h", "BCUTm-1l", "BCUTse-1l", "VR2 Dzv", "VR2 Dzse", "VR2 Dzpe", "VR2 Dzp", "VR2 Dzi", "SpDiam Dzm", "SpAbs Dzse", "SpAD Dzse", "SpAD Dzpe", "VE1 Dzi", "VE3 Dzi", "SpAD Dzv", "VE2 Dzp", "NsssN", "AETA beta", "AETA beta s", "AETA eta F", "ETA epsilon 2", "ETA alpha", "ETA beta", "TIC0", "Kier3", "SMR VSA3", "EState VSA10", "AMID h", "AMID X", "MPC4", "MPC6", "TMPC10", "GGI6", "TSRW10", "NddssS", "NaasN", "SaaaC", "SMR VSA4", "nX", "AATS6p", "VSA EState7", "ATS2m", "ATS7p", "AATS4i", "SsF", "AATS6v", "GATS5dv", "ATSC7i", "ATS5Z", "ATS8v", "MATS5dv", "BCUTZ-1l", "Xp-0dv", "NsssCH", "piPC8", "VR1 A", "nHetero", "nO", "ATS1v", "AATS2dv", "ATSC0se", "ATSC0are", "AATSC1p", "MATS1v", "MATS1p", "AATSC1i", "GATS5are", "ATS0pe", "AATS5s", "ATSC1se", "ATS4d", "ATS1p", "AATS2are", "ATSC3se", "ATS3dv", "ATS0d", "AATS4v", "MATS5c", "GATS6d", "GATS4v", "BCUTdv-1h", "BCUTdv-1l", "BCUTd-1l", "SpMax Dzm", "SpAD Dzm", "LogEE Dzm", "VE1 Dzv", "VE3 Dzv", "VR2 Dzare", "VE1 Dzp", "VE3 Dzp", "VR2 DzZ", "VR2 Dzm", "VR3 Dzm", "VR1 Dzare", "VR1 Dzi", "VR1 DzZ", "VR1 Dzm", "nBonds", "nBondsA", "nBondsKD", "Sm", "Sare", "Si", "VE1 D", "VE3 D", "VR1 D", "VR2 D", "VE2 D", "VR3 D", "SpMAD D", "NsOH", "SaasC", "SsOH", "MINssCH2", "NssO", "ETA dAlpha B", "ETA epsilon 4", "nHBAcc", "BIC2", "IC2", "ZMIC0", "SMR VSA6", "PEOE VSA13", "piPC2", "bpol", "n6aRing", "TopoPSA", "Diameter", "SRW04", "Zagreb1", "NtN", "GATS4c", "ETA eta", "ATS6are", "nS", "GATS4p", "MATS5v", "ATSC8pe", "C1SP3", "NsCH3", "SpMax A", "nN", "ATS2Z", "AATS0d", "AATS3s", "AATS5v", "ATSC1m", "AATSC1pe", "AATSC4are", "AATSC0i", "MATS1i", "GATS1are", "GATS5p", "ATS6dv", "ATS7s", "ATS1pe", "MATS4dv", "MATS1s", "GATS3p", "ATS0dv", "ATS6pe", "ATS0i", "AATS0m", "AATS0are", "ATS3v", "ATS1i", "AATS5are", "ATSC0Z", "ATSC0v", "AATSC0d", "AATSC1s", "AATSC4i", "MATS5d", "ATS3are", "AATS5dv", "AATS2d", "ATSC8v", "MATS4are", "SpAD DzZ", "VE2 Dzm", "nBondsM", "C2SP2", "Sse", "Spe", "Sp", "Mv", "SdO", "AETA eta FL", "ETA shape y", "BIC5", "TIC1", "VSA EState8", "MID N", "AMID N", "piPC6", "TopoShapeIndex", "PetitjeanIndex", "nG12FaRing", "SdS", "NdS", "SdddsS", "C4SP3"]

- **ESOL**: ["SaasN", "SlogP VSA10", "SaaNH", "nS", "SpMAD DzZ", "SpMAD Dzm", "SM1 Dzse", "SpAD Dzp", "SM1 Dzi", "RNCG", "RPCG", "C1SP3", "Xp-0dv", "AXp-0dv", "AXp-4dv", "Sv", "Sp-

Diam D", "SdsssP", "AETA eta B", "CIC0", "CIC1", "ZMIC2", "LabuteASA", "PEOE VSA8", "PEOE VSA13", "PEOE VSA3", "VSA EState6", "TMPC10", "piPC8", "nRing", "nARing", "n5ARing", "naHRing", "SLogP", "GGI4", "GGI7", "JGI5", "MWC01", "MWC03", "MWC04", "MWC05", "SRW02", "Zagreb1", "Zagreb2", "mZagreb1", "mZagreb2", "VSA EState5", "nH", "BCUTs-1h", "SpMax Dzp", "LogEE Dzp", "SpAbs Dzp", "SpMAD Dzare", "SM1 Dzare", "AXp-7d", "AXp-1dv", "Spe", "Sare", "Mm", "MZ", "SsOH", "MAXsCH3", "MINssCH2", "NsCH3", "ETA shape y", "ETA epsilon 3", "ETA dEpsilon A", "SIC0", "BIC0", "IC5", "Kier1", "SlogP VSA4", "VSA EState8", "MID C", "piPC3", "piPC1", "RotRatio", "TopoPSA", "MWC08", "TMWC10", "SRW04", "NtN", "SsI", "nF", "NsF", "nAHRing", "MINsCH3", "SpDiam DzZ", "SpDiam Dzi", "SpMAD Dzi", "SpMAD Dzv", "C3SP3", "AXp-5dv", "Sse", "VE2 D", "MINaasC", "MIC1", "MIC4", "MID h", "AMID h", "AMID C", "MPC7", "n6aHRing", "SMR", "GGI8", "GGI9", "MWC02", "SRW06", "MWC06", "nG12FAHRing", "MAXssCH2", "StsC", "BCUTc-1l", "BCUTZ-1l", "BCUTv-1l", "BCUTse-1l", "BCUTp-1h", "BCUTp-1l", "SM1 DzZ", "SM1 Dzm", "SpAbs Dzse", "SM1 Dzpe", "Xc-6d", "Xp-3dv", "AXp-7dv", "SpAbs D", "SpMax D", "SpAD D", "LogEE D", "NdssC", "NddsN", "SdO", "SdS", "SsCl", "AETA eta F", "AETA eta FL", "ETA epsilon 5", "ETA dEpsilon B", "AETA eta BR", "AETA dBeta", "TIC5", "MIC0", "ZMIC4", "BIC1", "IC3", "SMR VSA6", "EState VSA8", "PEOE VSA12", "MDEC-12", "AMID", "AMID N", "MID X", "MPC4", "MPC10", "piPC10", "TpiPC10", "MPC6", "piPC2", "GGI5", "JGI1", "JGI6", "MWC10", "MW", "AMW", "WPol", "NddssS", "nI", "NsI", "SssNH", "NsssS", "nFaHRing", "SpMAD A", "nCl", "BCUTZ-1h", "BCUTm-1h", "BCUTm-1l", "BCUTpe-1l", "BCUTare-1l", "BCUTi-1h", "BCUTse-1h", "SpMax Dzm", "LogEE Dzm", "SpMax Dzse", "SpAD Dzse", "SpMAD Dzse", "LogEE Dzse", "SpMAD Dzpe", "nBondsKS", "AXp-0d", "Xp-1dv", "AXp-2dv", "AXp-6dv", "Xch-7dv", "Mare", "NaasC", "SssO", "ETA dEpsilon D", "nHBAcc", "MIC3", "ZMIC1", "BIC5", "Kier2", "PEOE VSA1", "EState VSA7", "PEOE VSA2", "MID N", "piPC5", "MPC9", "MPC3", "nG12FRing", "JGI4", "JGI3", "JGI9", "Radius", "MWC07", "TSRW10", "nBridgehead", "n10FARing", "NaaNH", "ABC", "SpAbs A", "SpAD A", "BCUTdv-1h", "BCUTs-1l", "BCUTi-1l", "BCUTare-1h", "SpMax DzZ", "LogEE DzZ", "VE2 DzZ", "VE2 Dzm", "SM1 Dzv", "SpAbs Dzpe", "SpMax Dzpe", "SpAD Dzpe", "LogEE Dzpe", "VE2 Dzare", "VE2 Dzi", "SpAD Dzv", "nBondsD", "C3SP2", "C4SP3", "Xp-4d", "AXp-3d", "AXp-5d", "Xp-6d", "AXp-3dv", "Xpc-5dv", "Xp-3d", "Sp", "Si", "Mv", "VR1 D", "NsNH2", "NdsssP", "SssCH2", "MAXaaCH", "MAXaasC", "SaasC", "SsssN", "ETA shape x", "AETA eta", "AETA eta L", "ETA eta RL", "ETA epsilon 4", "ETA dEpsilon C", "AETA beta", "BIC4", "MIC2", "MIC5", "IC1", "TIC1", "IC4", "PEOE VSA5", "PEOE VSA7", "SMR VSA5", "SlogP VSA5", "SMR VSA1", "VSA EState3", "PEOE VSA9", "MPC5", "piPC4", "nFRing", "JGI10", "SRW09", "MWC09", "Xch-4d", "Xch-4dv", "EState VSA10", "n5aRing", "EState VSA1", "NsssN", "nHRing", "nG12FHRing", "ABCGG", "VR2 A", "nAromAtom", "nAromBond", "nP", "SpAD DzZ", "SpAbs Dzm", "SpDiam Dzm", "VE2 Dzv", "VE2 Dzse", "VE2 Dzpe", "SpAbs Dzare", "SpMax Dzare", "SpDiam Dzse", "SpDiam Dzv", "nBondsS", "nBondsM", "HybRatio", "FCSP3", "Xch-5dv", "Xp-4dv", "Xch-7d", "SZ", "Mse", "Mp", "Sm", "Mpe", "SpMAD D", "SsCH3", "NsCl", "ETA beta", "ETA dBeta", "AETA beta ns", "ETA beta ns d", "ZMIC5", "SIC1", "ZMIC0", "CIC2", "Kier3", "EState VSA6", "PEOE VSA10", "SMR VSA4", "VSA EState7", "MDEC-33", "AMID X", "MPC8", "piPC9", "n6HRing", "naRing", "n10FHRing", "SRW07", "WPath", "NdNH", "SdNH", "n12FHRing", "n12FaRing", "n12FaHRing", "Xch-3dv", "NdssS", "Xch-3d", "SdssS", "n8FHRing", "n8FAHRing", "n3Ring", "n3ARing", "SRW03", "n8FRing", "n8FARing", "SlogP VSA11", "n5aHRing", "VR3 A", "nAtom", "nHetero", "nX", "BCUTdv-1l", "BCUTpe-1h", "SpAD Dzm", "VE2 Dzp", "SpAbs Dzi", "SpMax Dzi", "LogEE Dzi", "VR2 Dzi", "SpDiam Dzpe", "SpMAD Dzp", "nBondsKD", "Xch-5d", "Xc-5dv", "Xc-6dv", "NsOH", "SssssC", "ETA shape p", "AETA beta s", "ETA eta", "nHBDon", "IC0", "TIC2", "TIC4", "CIC4", "TIC0", "BIC3", "SlogP VSA6", "MDEC-23", "piPC7", "bpol", "n6Ring", "GGI10", "Diameter", "TopoShapeIndex", "PetitjeanIndex"]

For the QM8 and QM9 datasets, which have 12 tasks each, we detail each set of causal features per task. Below we describe the features for the **QM8 dataset**:

- **E1-CAM**: ["n8FHRing", "nBase", "SpDiam A", "SpMAD A", "LogEE A", "VE2 A", "VE3 A", "nAromAtom", "nAromBond", "nSpiro", "nBridgehead", "nH", "ATS3dv", "ATS4dv", "ATS2d", "ATS5d", "ATS2s", "ATS3s", "ATS1Z", "ATS2Z", "ATS4Z", "ATS2are", "AATS2dv", "AATS3dv", "AATS5dv", "AATS1d", "AATS3d", "AATS1s", "AATS2s", "AATS3s", "AATS5s", "AATS2Z", "AATS2v", "AATS3v", "AATS5v", "AATS3se", "AATS6se", "AATS3pe", "AATS1are", "AATS2are", "AATS3are", "AATS0p", "AATS1p", "AATS3p", "AATS0i", "AATS1i", "AATS2i", "ATSC0c", "ATSC1c", "ATSC2c", "ATSC3c", "ATSC4c", "ATSC7c", "ATSC0dv", "ATSC2dv", "ATSC3dv", "ATSC2d", "ATSC3d", "ATSC5d", "ATSC0s", "ATSC1s", "ATSC4s", "ATSC0Z", "ATSC1Z", "ATSC2Z", "ATSC6Z", "ATSC2m", "ATSC2se", "ATSC0pe", "ATSC1pe", "ATSC4pe", "ATSC0are", "ATSC1are", "ATSC2are", "ATSC1p", "ATSC0i", "ATSC1i", "AATSC0c", "AATSC2c", "AATSC3c", "AATSC5c", "AATSC6c", "AATSC1dv", "AATSC0d", "AATSC3d", "AATSC6d", "AATSC0s", "AATSC2s", "AATSC3s", "AATSC0Z", "AATSC1Z", "AATSC0m", "AATSC1m", "AATSC6m", "AATSC0se", "AATSC1se", "AATSC2se", "AATSC1pe", "AATSC2p", "AATSC0i", "AATSC1i",

"MATS1c", "MATS2c", "MATS3c", "MATS4c", "MATS5c", "MATS1dv", "MATS2dv", "MATS5dv", "MATS1d", "MATS6d", "MATS1s", "MATS2s", "MATS6s", "MATS1Z", "MATS1m", "MATS2m", "MATS1se", "MATS2se", "MATS3se", "MATS1pe", "MATS5pe", "MATS6are", "MATS1p", "MATS1i", "MATS3i", "GATS1c", "GATS2c", "GATS3c", "GATS5c", "GATS6c", "GATS1dv", "GATS5dv", "GATS1d", "GATS2d", "GATS3d", "GATS4d", "GATS1s", "GATS2s", "GATS3s", "GATS1Z", "GATS2m", "GATS4m", "GATS1v", "GATS1se", "GATS2se", "GATS1pe", "GATS2are", "GATS1p", "GATS1i", "GATS2i", "GATS5i", "BCUTc-1h", "BCUTc-1l", "BCUTdv-1h", "BCUTdv-1l", "BCUTd-1h", "BCUTd-1l", "BCUTs-1h", "BCUTs-1l", "BCUTZ-1h", "BCUTZ-1l", "BCUTm-1h", "BCUTm-1l", "BCUTv-1h", "BCUTv-1l", "BCUTse-1h", "BCUTse-1l", "BCUTpe-1h", "BCUTpe-1l", "BCUTare-1h", "BCUTare-1l", "BCUTp-1h", "BCUTp-1l", "BCUTi-1h", "BCUTi-1l", "VE1 Dzm", "VE2 Dzm", "VE3 Dzm", "VR2 Dzm", "SpAbs Dzv", "SpMax Dzv", "SpDiam Dzv", "VE1 Dzv", "VE3 Dzv", "VR2 Dzv", "VE3 Dzse", "SpAbs Dzpe", "SpMAD Dzpe", "VE1 Dzpe", "VE3 Dzpe", "SpAbs Dzare", "SpAD Dzare", "SpMAD Dzare", "LogEE Dzare", "VE1 Dzare", "VR2 Dzare", "SpMAD Dzp", "VE2 Dzp", "VR1 Dzp", "VR2 Dzp", "VR3 Dzp", "SpAbs Dzi", "SpMax Dzi", "SpDiam Dzi", "SpAD Dzi", "VE1 Dzi", "VE3 Dzi", "VR1 Dzi", "nBonds", "nBondsD", "nBondsM", "nBondsKS", "nBondsKD", "RNCG", "RPCG", "C2SP1", "C1SP2", "C2SP2", "C3SP2", "HybRatio", "FCSP3", "Xch-3d", "Xch-4d", "Xch-5d", "Xch-6d", "Xch-7d", "Xch-3dv", "Xch-4dv", "Xch-5dv", "Xch-6dv", "Xc-5d", "Xc-5dv", "Xpc-5d", "Xpc-6d", "Xpc-6dv", "Xp-1d", "Xp-3d", "Xp-5d", "Xp-6d", "Xp-7d", "AXp-2d", "AXp-3d", "AXp-4d", "AXp-5d", "AXp-0dv", "AXp-1dv", "AXp-2dv", "AXp-3dv", "AXp-4dv", "AXp-5dv", "AXp-6dv", "Sm", "Sare", "MZ", "Mse", "Mpe", "Mare", "Mp", "Mi", "VE1 D", "VE3 D", "VR1 D", "VR2 D", "VR3 D", "ETA alpha", "ETA shape p", "ETA beta", "AETA beta", "AETA beta s", "ETA beta ns", "AETA beta ns", "AETA eta", "ETA eta L", "AETA eta L", "ETA eta RL", "ETA eta F", "AETA eta F", "ETA eta FL", "AETA eta FL", "AETA eta BR", "ETA epsilon 4", "ETA epsilon 5", "ETA dEpsilon A", "ETA dEpsilon B", "ETA dEpsilon D", "ETA dBeta", "AETA dBeta", "ETA dPsi A", "nHBAcc", "nHBDon", "IC0", "TIC1", "TIC2", "TIC3", "TIC4", "TIC5", "SIC1", "SIC2", "SIC5", "BIC0", "BIC1", "BIC2", "BIC3", "BIC4", "BIC5", "CIC1", "MIC0", "MIC1", "ZMIC1", "ZMIC5", "Kier3", "PEOE VSA1", "PEOE VSA2", "PEOE VSA3", "PEOE VSA4", "PEOE VSA5", "PEOE VSA6", "PEOE VSA7", "PEOE VSA8", "PEOE VSA9", "PEOE VSA10", "PEOE VSA11", "PEOE VSA12", "PEOE VSA13", "SMR VSA1", "SMR VSA2", "SMR VSA3", "SMR VSA5", "SMR VSA6", "SMR VSA7", "SlogP VSA1", "SlogP VSA2", "SlogP VSA3", "SlogP VSA4", "SlogP VSA5", "SlogP VSA6", "SlogP VSA10", "SlogP VSA11", "MDEC-33", "MID", "AMID", "MID h", "AMID h", "AMID C", "MID N", "AMID N", "MID O", "AMID O", "MPC4", "piPC3", "piPC4", "piPC5", "piPC6", "piPC7", "TpiPC10", "n4Ring", "n4HRing", "n6HRing", "n5aRing", "n5aHRing", "n5ARing", "n3AHRing", "n4AHRing", "n5FRing", "n6FRing", "n7FRing", "n5FARing", "n6FARing", "n7FARing", "nFAHRing", "RotRatio", "SLogP", "SMR", "TopoPSA(NO)", "TopoPSA", "GGI2", "GGI3", "GGI4", "JGI2", "JGI3", "JGI4", "JGT10", "TopoShapeIndex", "PetitjeanIndex", "SRW05", "SRW06", "SRW07", "SRW08", "SRW09", "SRW10", "TSRW10", "MW", "AMW"]

- **E1-CC2**: ["C2SP1", "nBase", "SpAbs A", "SpDiam A", "SpAD A", "LogEE A", "VE1 A", "VE2 A", "VE3 A", "VR1 A", "VR2 A", "nAromAtom", "nAromBond", "nSpiro", "nBridgehead", "nN", "nO", "ATS4dv", "ATS2d", "ATS0s", "ATS2s", "ATS3s", "ATS1Z", "ATS3m", "ATS4m", "ATS2pe", "ATS2are", "ATS2p", "ATS5i", "AATS2dv", "AATS5dv", "AATS0d", "AATS1d", "AATS2d", "AATS0s", "AATS1s", "AATS2s", "AATS3s", "AATS6s", "AATS0Z", "AATS3Z", "AATS5Z", "AATS0m", "AATS2m", "AATS2v", "AATS3v", "AATS5v", "AATS0se", "AATS1se", "AATS2se", "AATS3se", "AATS0pe", "AATS1pe", "AATS2pe", "AATS3pe", "AATS0are", "AATS2are", "AATS3are", "AATS0p", "AATS2p", "AATS3p", "AATS0i", "AATS1i", "AATS2i", "AATS5i", "ATSC0c", "ATSC1c", "ATSC3c", "ATSC4c", "ATSC1dv", "ATSC3dv", "ATSC2d", "ATSC3d", "ATSC0s", "ATSC1s", "ATSC2s", "ATSC5m", "ATSC4v", "ATSC5v", "ATSC0se", "ATSC1se", "ATSC2se", "ATSC4se", "ATSC0are", "ATSC1are", "ATSC5are", "ATSC1p", "ATSC1i", "AATSC1c", "AATSC2c", "AATSC2dv", "AATSC5dv", "AATSC0d", "AATSC3d", "AATSC5d", "AATSC0s", "AATSC1s", "AATSC2s", "AATSC3s", "AATSC0Z", "AATSC1Z", "AATSC2Z", "AATSC0m", "AATSC5m", "AATSC0se", "AATSC1se", "AATSC2se", "AATSC1pe", "AATSC2are", "AATSC0p", "AATSC1p", "MATS1c", "MATS2c", "MATS3c", "MATS2d", "MATS1s", "MATS3s", "MATS5s", "MATS1Z", "MATS1m", "MATS3v", "MATS6v", "MATS2se", "MATS3se", "MATS1p", "MATS6p", "MATS1i", "GATS1c", "GATS2c", "GATS6c", "GATS1dv", "GATS3dv", "GATS1d", "GATS2d", "GATS3d", "GATS4d", "GATS5d", "GATS1s", "GATS2s", "GATS5s", "GATS1m", "GATS2m", "GATS6m", "GATS1se", "GATS2se", "GATS5pe", "GATS1are", "GATS1p", "GATS5p", "GATS1i", "GATS3i", "BCUTc-1h", "BCUTc-1l", "BCUTdv-1h", "BCUTdv-1l", "BCUTd-1h", "BCUTd-1l", "BCUTs-1h", "BCUTs-1l", "BCUTZ-1h", "BCUTZ-1l", "BCUTm-1h", "BCUTm-1l", "BCUTv-1h", "BCUTv-1l", "BCUTse-1h", "BCUTse-1l", "BCUTpe-1h", "BCUTpe-1l", "BCUTare-1h", "BCUTare-1l", "BCUTp-1h", "BCUTp-1l", "BCUTi-1h", "BCUTi-1l", "SpAbs DzZ", "SpMax DzZ", "SpMAD DzZ", "LogEE DzZ", "SpMax Dzm", "SpMAD Dzm", "VE1 Dzm", "VE3 Dzm", "VR3 Dzm", "SpDiam Dzv", "VR1 Dzv", "VR3 Dzv", "SpMax Dzse", "SpDiam Dzse", "SpAD Dzse", "SpMAD Dzse", "LogEE Dzse", "SM1 Dzse", "VE1 Dzse", "VE3 Dzse", "VR3 Dzse", "SpMax Dzpe", "SpDiam

Dzpe", "LogEE Dzpe", "VE3 Dzpe", "VR1 Dzpe", "SpMax Dzare", "SpDiam Dzare", "SpAD Dzare", "LogEE Dzare", "SM1 Dzare", "VE1 Dzare", "VE3 Dzare", "VR3 Dzare", "SpAbs Dzp", "SpMax Dzp", "SpAD Dzp", "SpMAD Dzp", "VR1 Dzp", "VR2 Dzp", "VR3 Dzp", "SpMax Dzi", "SpDiam Dzi", "SpMAD Dzi", "LogEE Dzi", "VE1 Dzi", "VE3 Dzi", "VR3 Dzi", "nBondsD", "nBondsA", "nBondsM", "nBondsKD", "RNCG", "C1SP1", "C1SP2", "C2SP2", "HybRatio", "FCSP3", "Xch-3d", "Xch-5d", "Xch-3dv", "Xch-4dv", "Xch-6dv", "Xc-5d", "Xc-5dv", "Xc-6dv", "Xpc-5d", "Xpc-6d", "Xpc-4dv", "Xpc-6dv", "Xp-2d", "Xp-3d", "Xp-4d", "Xp-5d", "Xp-6d", "AXp-1d", "AXp-3d", "AXp-4d", "AXp-5d", "AXp-6d", "Xp-2dv", "Xp-4dv", "Xp-5dv", "Xp-6dv", "Xp-7dv", "AXp-3dv", "AXp-4dv", "AXp-5dv", "AXp-6dv", "Mpe", "SpAD D", "SpMAD D", "LogEE D", "VE1 D", "VE3 D", "VR1 D", "VR2 D", "VR3 D", "ETA beta", "AETA beta", "ETA beta ns", "AETA beta ns", "ETA beta ns d", "AETA beta ns d", "AETA eta", "ETA eta L", "AETA eta R", "ETA eta RL", "AETA eta F", "ETA eta FL", "AETA eta FL", "AETA eta BR", "ETA epsilon 4", "ETA epsilon 5", "ETA dEpsilon B", "ETA dBeta", "AETA dBeta", "ETA dPsi B", "nHBAcc", "nHBDon", "IC3", "IC4", "IC5", "TIC0", "TIC1", "TIC2", "TIC3", "TIC4", "TIC5", "SIC0", "SIC1", "SIC3", "SIC4", "SIC5", "CIC4", "MIC1", "MIC2", "MIC3", "ZMIC1", "Kier2", "Kier3", "PEOE VSA1", "PEOE VSA2", "PEOE VSA3", "PEOE VSA4", "PEOE VSA5", "PEOE VSA6", "PEOE VSA7", "PEOE VSA8", "PEOE VSA10", "PEOE VSA11", "PEOE VSA12", "PEOE VSA13", "SMR VSA1", "SMR VSA2", "SMR VSA3", "SMR VSA6", "SMR VSA7", "SMR VSA9", "SlogP VSA1", "SlogP VSA2", "SlogP VSA3", "SlogP VSA4", "SlogP VSA5", "SlogP VSA6", "SlogP VSA10", "SlogP VSA11", "MDEC-22", "MDEC-23", "MDEC-33", "AMID", "MID h", "AMID h", "AMID C", "MID N", "AMID N", "MID O", "AMID O", "MPC3", "MPC4", "MPC6", "piPC2", "piPC3", "piPC4", "piPC5", "piPC6", "TpiPC10", "apol", "bpol", "nRing", "n3Ring", "n4Ring", "n5Ring", "nHRing", "n4HRing", "n5HRing", "n5aRing", "n6aRing", "n5aHRing", "n6aHRing", "n4ARing", "n4AHRing", "n5AHRing", "n5FRing", "n6FRing", "n5FARing", "n6FARing", "RotRatio", "SLogP", "TopoPSA(NO)", "TopoPSA", "GGI2", "GGI3", "JGI2", "JGI3", "JGI4", "JGT10", "Radius", "TopoShapeIndex", "PetitjeanIndex", "MWC02", "MWC03", "MWC06", "MWC07", "MWC08", "MWC09", "MWC10", "TMWC10", "SRW03", "SRW04", "SRW05", "SRW06", "SRW07", "SRW08", "SRW09", "SRW10", "TSRW10"]

- **E1-PBE0**: ["ATS8i", "nBase", "SpMax A", "SpDiam A", "VE1 A", "VE3 A", "VR2 A", "nAromAtom", "nAromBond", "nBridgehead", "nHetero", "ATS3dv", "ATS4dv", "ATS1d", "ATS2s", "ATS2Z", "ATS3Z", "ATS6Z", "ATS2m", "ATS3m", "ATS4m", "ATS2v", "ATS3i", "AATS1dv", "AATS3dv", "AATS5dv", "AATS1d", "AATS3d", "AATS0s", "AATS1s", "AATS2s", "AATS3s", "AATS3Z", "AATS5m", "AATS1v", "AATS2v", "AATS3v", "AATS5v", "AATS1se", "AATS2se", "AATS3se", "AATS1pe", "AATS2pe", "AATS3pe", "AATS0are", "AATS2are", "AATS3are", "AATS0p", "AATS2p", "AATS0i", "AATS1i", "AATS2i", "AATS5i", "ATSC0c", "ATSC1c", "ATSC2c", "ATSC2dv", "ATSC3d", "ATSC0s", "ATSC2s", "ATSC1Z", "ATSC6m", "ATSC4v", "ATSC1se", "ATSC2se", "ATSC4se", "ATSC1pe", "ATSC2pe", "ATSC1are", "ATSC3are", "ATSC1p", "ATSC7p", "ATSC0i", "ATSC1i", "AATSC0c", "AATSC2c", "AATSC3c", "AATSC0dv", "AATSC1dv", "AATSC5dv", "AATSC0d", "AATSC3d", "AATSC0s", "AATSC2s", "AATSC3s", "AATSC5Z", "AATSC6Z", "AATSC0m", "AATSC2m", "AATSC3m", "AATSC0v", "AATSC1se", "AATSC2se", "AATSC5se", "AATSC0pe", "AATSC2pe", "AATSC0are", "AATSC1are", "AATSC3are", "AATSC0p", "AATSC5p", "AATSC1i", "AATSC2i", "MATS1c", "MATS2c", "MATS3c", "MATS2dv", "MATS5dv", "MATS1d", "MATS2d", "MATS1s", "MATS1v", "MATS2se", "MATS1pe", "MATS1p", "GATS1c", "GATS2c", "GATS2dv", "GATS6dv", "GATS1d", "GATS4d", "GATS1s", "GATS4s", "GATS5s", "GATS1Z", "GATS1m", "GATS2m", "GATS1v", "GATS1se", "GATS2se", "GATS1pe", "GATS1are", "GATS2are", "GATS1p", "GATS6p", "GATS1i", "BCUTc-1h", "BCUTc-1l", "BCUTdv-1h", "BCUTdv-1l", "BCUTd-1h", "BCUTd-1l", "BCUTs-1h", "BCUTs-1l", "BCUTZ-1h", "BCUTZ-1l", "BCUTm-1h", "BCUTm-1l", "BCUTv-1h", "BCUTv-1l", "BCUTse-1h", "BCUTse-1l", "BCUTpe-1h", "BCUTpe-1l", "BCUTare-1h", "BCUTare-1l", "BCUTp-1h", "BCUTp-1l", "BCUTi-1h", "BCUTi-1l", "SpDiam DzZ", "VE1 DzZ", "VE3 DzZ", "SpDiam Dzm", "VE1 Dzm", "VE3 Dzm", "SpDiam Dzv", "SM1 Dzv", "VR1 Dzv", "VR3 Dzv", "SpDiam Dzse", "SpMAD Dzse", "LogEE Dzse", "SM1 Dzse", "VE3 Dzse", "VR2 Dzse", "VR3 Dzse", "SpDiam Dzpe", "SM1 Dzpe", "VE2 Dzpe", "VR1 Dzpe", "VR3 Dzpe", "SpMax Dzare", "SpMAD Dzare", "VE3 Dzare", "VR1 Dzare", "VR2 Dzare", "VR3 Dzare", "SpMAD Dzp", "SM1 Dzp", "VE1 Dzp", "VE3 Dzp", "VR1 Dzp", "VR2 Dzp", "VR3 Dzp", "SpAbs Dzi", "SM1 Dzi", "VE1 Dzi", "VE2 Dzi", "VE3 Dzi", "VR1 Dzi", "VR2 Dzi", "VR3 Dzi", "nBondsS", "nBondsD", "nBondsA", "nBondsM", "nBondsKD", "RNCG", "RPCG", "C1SP1", "C2SP1", "C1SP2", "C2SP2", "C3SP2", "C2SP3", "HybRatio", "FCSP3", "Xch-3d", "Xch-4d", "Xch-5d", "Xch-6d", "Xch-3dv", "Xch-5dv", "Xch-6dv", "Xch-7dv", "Xc-5dv", "Xpc-5d", "Xpc-6d", "Xpc-4dv", "Xpc-5dv", "Xp-4d", "Xp-5d", "Xp-6d", "AXp-0d", "AXp-1d", "AXp-3d", "AXp-4d", "AXp-6d", "Xp-2dv", "Xp-4dv", "Xp-6dv", "AXp-0dv", "AXp-1dv", "AXp-2dv", "AXp-3dv", "AXp-4dv", "AXp-5dv", "AXp-6dv", "Si", "Mm", "Mpe", "Mare", "Mp", "SpAbs D", "SpMax D", "SpDiam D", "SpAD D", "SpMAD D", "LogEE D", "VE1 D", "VE3 D", "VR1 D", "VR2 D", "VR3 D", "ETA shape y", "AETA beta", "ETA beta s", "ETA beta ns", "AETA beta ns", "ETA beta ns d", "ETA eta", "AETA eta", "ETA eta L", "AETA eta L", "ETA eta R", "ETA eta F", "ETA eta FL", "AETA eta FL", "AETA eta BR", "ETA epsilon 3", "ETA epsilon

5", "ETA dEpsilon A", "ETA dEpsilon B", "ETA dBeta", "AETA dBeta", "ETA psi 1", "nHBAcc", "nHBDon", "IC0", "IC1", "IC2", "IC3", "IC4", "TIC1", "TIC4", "SIC1", "SIC4", "SIC5", "BIC1", "BIC3", "BIC4", "CIC0", "CIC1", "CIC5", "MIC0", "MIC1", "MIC2", "MIC3", "MIC4", "ZMIC0", "ZMIC1", "ZMIC2", "ZMIC5", "Kier3", "LabuteASA", "PEOE VSA1", "PEOE VSA2", "PEOE VSA3", "PEOE VSA4", "PEOE VSA5", "PEOE VSA6", "PEOE VSA7", "PEOE VSA8", "PEOE VSA9", "PEOE VSA10", "PEOE VSA11", "PEOE VSA12", "PEOE VSA13", "SMR VSA1", "SMR VSA2", "SMR VSA3", "SMR VSA4", "SMR VSA5", "SMR VSA6", "SMR VSA7", "SMR VSA9", "SlogP VSA1", "SlogP VSA2", "SlogP VSA3", "SlogP VSA4", "SlogP VSA5", "SlogP VSA6", "SlogP VSA10", "SlogP VSA11", "MDEC-22", "MDEC-23", "MDEC-33", "AMID", "MID h", "AMID h", "MID N", "AMID N", "MID O", "AMID O", "AMID X", "MPC4", "TMPC10", "piPC1", "piPC2", "piPC3", "piPC4", "piPC5", "TpiPC10", "apol", "bpol", "nRing", "n5Ring", "n4HRing", "n5aRing", "n6aRing", "n5aHRing", "n6aHRing", "nARing", "n4ARing", "nAHRing", "n4AHRing", "n5AHRing", "n5FRing", "n5FARing", "RotRatio", "SLogP", "TopoPSA(NO)", "TopoPSA", "GGI2", "GGI4", "GGI5", "JGI1", "JGI2", "JGI3", "JGI4", "JGT10", "Radius", "MWC03", "MWC04", "MWC05", "MWC07", "MWC08", "MWC09", "MWC10", "TMWC10", "SRW04", "SRW05", "SRW06", "SRW07", "SRW09", "SRW10", "TSRW10", "AMW", "WPol"]

- **E2-CAM**: ["nAcid", "nBase", "SpMax A", "SpDiam A", "VE3 A", "VR1 A", "VR2 A", "VR3 A", "nBridgehead", "ATS1dv", "ATS2dv", "ATS3dv", "ATS2d", "ATS3s", "ATS4s", "ATS1Z", "ATS3Z", "ATS4Z", "ATS0m", "ATS3se", "ATS1pe", "ATS3pe", "ATS1are", "ATS0i", "AATS1dv", "AATS2dv", "AATS3dv", "AATS4dv", "AATS5dv", "AATS3d", "AATS5d", "AATS0s", "AATS1s", "AATS2s", "AATS4s", "AATS5s", "AATS1Z", "AATS1m", "AATS2m", "AATS1v", "AATS3v", "AATS5v", "AATS2se", "AATS6se", "AATS2pe", "AATS1are", "AATS2are", "AATS3are", "AATS0p", "AATS2p", "AATS3p", "AATS0i", "AATS1i", "AATS2i", "AATS4i", "ATSC0c", "ATSC1c", "ATSC2c", "ATSC4c", "ATSC5c", "ATSC1dv", "ATSC2dv", "ATSC3d", "ATSC5d", "ATSC0s", "ATSC3s", "ATSC1m", "ATSC3v", "ATSC2se", "ATSC4se", "ATSC4are", "ATSC1p", "ATSC1i", "AATSC0c", "AATSC1c", "AATSC2c", "AATSC6dv", "AATSC0d", "AATSC2d", "AATSC3d", "AATSC5d", "AATSC0s", "AATSC1s", "AATSC3s", "AATSC0Z", "AATSC0m", "AATSC1m", "AATSC1v", "AATSC2se", "AATSC4se", "AATSC1pe", "AATSC5pe", "AATSC1are", "AATSC1p", "AATSC3p", "AATSC0i", "AATSC1i", "AATSC2i", "AATSC3i", "MATS1c", "MATS2c", "MATS3c", "MATS2dv", "MATS5dv", "MATS1d", "MATS3d", "MATS1s", "MATS3s", "MATS4s", "MATS1Z", "MATS1m", "MATS2v", "MATS1se", "MATS2se", "MATS5se", "MATS1are", "MATS5p", "MATS3i", "MATS5i", "GATS1c", "GATS2c", "GATS5c", "GATS1dv", "GATS1d", "GATS2d", "GATS4d", "GATS1s", "GATS2s", "GATS3s", "GATS4Z", "GATS4m", "GATS1v", "GATS2v", "GATS1se", "GATS2se", "GATS1pe", "GATS4pe", "GATS1are", "GATS2are", "GATS1p", "GATS2p", "GATS2i", "BCUTc-1h", "BCUTc-1l", "BCUTdv-1h", "BCUTdv-1l", "BCUTd-1h", "BCUTd-1l", "BCUTs-1h", "BCUTs-1l", "BCUTZ-1h", "BCUTZ-1l", "BCUTm-1h", "BCUTm-1l", "BCUTv-1h", "BCUTv-1l", "BCUTse-1h", "BCUTse-1l", "BCUTpe-1h", "BCUTpe-1l", "BCUTare-1h", "BCUTare-1l", "BCUTp-1h", "BCUTp-1l", "BCUTi-1h", "BCUTi-1l", "VE1 DzZ", "VE3 DzZ", "VR1 DzZ", "VR2 DzZ", "VR3 DzZ", "VE1 Dzm", "VE3 Dzm", "SpAbs Dzv", "SpMax Dzv", "SpDiam Dzv", "SpAD Dzv", "VR1 Dzv", "VR2 Dzv", "SpMax Dzse", "LogEE Dzse", "SM1 Dzse", "VE1 Dzse", "VE2 Dzse", "VE3 Dzse", "VR3 Dzse", "SpMAD Dzpe", "LogEE Dzpe", "VE1 Dzpe", "VE2 Dzpe", "VE3 Dzpe", "VR1 Dzpe", "VR2 Dzpe", "VR3 Dzpe", "SpAbs Dzare", "SpMax Dzare", "SpDiam Dzare", "SpMAD Dzare", "LogEE Dzare", "VE3 Dzare", "VR1 Dzare", "VR2 Dzare", "VR3 Dzare", "SpAbs Dzp", "SpMax Dzp", "SpDiam Dzp", "SpAD Dzp", "SpMAD Dzp", "LogEE Dzp", "SM1 Dzp", "VE1 Dzp", "VE2 Dzp", "VE3 Dzp", "VR1 Dzp", "VR2 Dzp", "VR3 Dzi", "nBondsO", "nBondsD", "nBondsM", "nBondsKD", "RNCG", "RPCG", "C1SP2", "C2SP2", "HybRatio", "FCSP3", "Xch-3d", "Xch-4d", "Xch-5d", "Xch-6d", "Xch-7d", "Xch-3dv", "Xch-4dv", "Xch-5dv", "Xch-6dv", "Xch-7dv", "Xc-3dv", "Xc-5dv", "Xpc-4d", "Xpc-5d", "Xpc-5dv", "Xp-1d", "Xp-2d", "Xp-3d", "Xp-4d", "Xp-6d", "AXp-1d", "AXp-3d", "AXp-4d", "AXp-5d", "AXp-6d", "Xp-0dv", "Xp-1dv", "Xp-2dv", "Xp-5dv", "AXp-0dv", "AXp-1dv", "AXp-4dv", "AXp-6dv", "Mv", "Mp", "Mi", "VE3 D", "VR1 D", "VR2 D", "VR3 D", "ETA shape p", "ETA shape y", "ETA beta", "AETA beta", "AETA beta s", "ETA beta ns", "ETA eta F", "AETA eta F", "ETA eta FL", "ETA eta B", "AETA eta B", "ETA dAlpha B", "ETA epsilon 1", "ETA epsilon 2", "ETA epsilon 3", "ETA epsilon 4", "ETA epsilon 5", "ETA dEpsilon B", "ETA dEpsilon C", "ETA dBeta", "AETA dBeta", "nHBDon", "IC0", "IC1", "IC4", "IC5", "TIC1", "TIC2", "TIC3", "TIC4", "TIC5", "BIC0", "BIC5", "CIC0", "CIC1", "MIC2", "MIC3", "ZMIC1", "ZMIC3", "Kier2", "Kier3", "LabuteASA", "PEOE VSA1", "PEOE VSA2", "PEOE VSA3", "PEOE VSA4", "PEOE VSA5", "PEOE VSA6", "PEOE VSA7", "PEOE VSA8", "PEOE VSA9", "PEOE VSA10", "PEOE VSA11", "PEOE VSA12", "PEOE VSA13", "SMR VSA2", "SMR VSA3", "SMR VSA4", "SMR VSA5", "SMR VSA6", "SMR VSA7", "SlogP VSA1", "SlogP VSA2", "SlogP VSA3", "SlogP VSA4", "SlogP VSA6", "SlogP VSA10", "MDEC-13", "MDEC-22", "MDEC-23", "MDEC-33", "AMID", "MID h", "AMID h", "AMID C", "MID N", "AMID N", "MID O", "MPC4", "MPC5", "piPC2", "piPC3", "piPC4", "piPC5", "piPC6", "apol", "bpol", "nRing", "n4Ring", "nHRing", "n4HRing", "n5HRing", "n6HRing", "n5aRing", "n5aHRing", "nARing", "n3ARing", "n4ARing", "nAHRing", "n4AHRing", "nFRing", "n5FRing", "n6FRing", "nFARing", "n5FARing", "n6FARing",

"SLogP", "SMR", "TopoPSA(NO)", "TopoPSA", "GGI2", "GGI3", "JGI2", "JGI3", "JGI4", "MWC03", "MWC04", "MWC05", "MWC06", "MWC07", "MWC08", "MWC09", "MWC10", "TMWC10", "SRW04", "SRW05", "SRW06", "SRW07", "SRW09", "TSRW10", "WPol", "Zagreb1", "Zagreb2", "mZagreb1"]

- **E2-CC2**: ["nBase", "SpDiam A", "VE3 A", "nSpiro", "nBridgehead", "nC", "ATS3dv", "ATS4dv", "ATS1d", "ATS0s", "ATS4m", "ATS1se", "ATS2se", "ATS3pe", "ATS3are", "AATS1dv", "AATS3dv", "AATS2s", "AATS3Z", "AATS2m", "AATS1v", "AATS5v", "AATS3se", "AATS4are", "AATS4p", "AATS0i", "AATS1i", "AATS4i", "ATSC0c", "ATSC1c", "ATSC2c", "ATSC3c", "ATSC3dv", "ATSC2d", "ATSC1s", "ATSC2s", "ATSC3s", "ATSC0se", "ATSC1se", "ATSC2se", "ATSC4se", "ATSC0pe", "ATSC1pe", "ATSC0are", "ATSC1p", "ATSC1i", "ATSC3i", "AATSC0c", "AATSC1c", "AATSC2c", "AATSC0dv", "AATSC2dv", "AATSC3dv", "AATSC6dv", "AATSC3d", "AATSC0s", "AATSC0Z", "AATSC1Z", "AATSC0m", "AATSC2se", "AATSC3se", "AATSC5se", "MATS1c", "MATS2c", "MATS2dv", "MATS4dv", "MATS1d", "MATS4d", "MATS5d", "MATS2se", "MATS6p", "GATS1c", "GATS2c", "GATS1dv", "GATS1d", "GATS1s", "GATS3s", "GATS1Z", "GATS4Z", "GATS1v", "GATS1se", "GATS4se", "GATS5se", "GATS1pe", "GATS1are", "GATS2are", "GATS1p", "GATS2i", "BCUTc-1h", "BCUTc-1l", "BCUTdv-1h", "BCUTdv-1l", "BCUTd-1h", "BCUTd-1l", "BCUTs-1h", "BCUTs-1l", "BCUTZ-1h", "BCUTZ-1l", "BCUTm-1h", "BCUTm-1l", "BCUTv-1h", "BCUTv-1l", "BCUTse-1h", "BCUTse-1l", "BCUTpe-1h", "BCUTpe-1l", "BCUTare-1h", "BCUTare-1l", "BCUTp-1h", "BCUTp-1l", "BCUTi-1h", "BCUTi-1l", "SpMax DzZ", "SpDiam DzZ", "SpAD DzZ", "SpMAD DzZ", "LogEE DzZ", "VE1 DzZ", "VE2 DzZ", "VE3 DzZ", "SpAbs Dzm", "SpMax Dzm", "SpDiam Dzm", "SpMAD Dzm", "LogEE Dzm", "VE2 Dzm", "SpMax Dzv", "LogEE Dzv", "VE1 Dzv", "VE3 Dzv", "VR3 Dzv", "SpDiam Dzse", "SM1 Dzse", "VE2 Dzse", "SpDiam Dzpe", "SpMAD Dzpe", "SM1 Dzpe", "SM1 Dzare", "VE2 Dzare", "SM1 Dzp", "VE1 Dzp", "VE3 Dzp", "VR1 Dzp", "VR2 Dzp", "VR3 Dzp", "SpAbs Dzi", "SpMax Dzi", "SpDiam Dzi", "SpMAD Dzi", "LogEE Dzi", "SM1 Dzi", "VE1 Dzi", "VE3 Dzi", "VR2 Dzi", "nBondsS", "nBondsD", "nBondsKD", "RNCG", "RPCG", "C2SP1", "C1SP2", "C2SP2", "C3SP2", "C1SP3", "HybRatio", "Xch-3d", "Xch-4d", "Xch-5d", "Xch-6d", "Xch-3dv", "Xch-4dv", "Xch-5dv", "Xc-6d", "Xc-5dv", "Xpc-5d", "Xpc-6d", "Xp-3d", "Xp-4d", "AXp-2d", "AXp-3d", "AXp-4d", "Xp-2dv", "Xp-3dv", "Xp-4dv", "AXp-0dv", "AXp-1dv", "AXp-2dv", "AXp-3dv", "AXp-5dv", "MZ", "Mm", "Mv", "Mse", "Mp", "SpMAD D", "VE1 D", "VE3 D", "ETA alpha", "AETA alpha", "ETA beta", "AETA beta", "AETA beta s", "ETA beta ns", "AETA beta ns", "ETA beta ns d", "AETA beta ns d", "AETA eta RL", "AETA eta F", "ETA eta FL", "AETA eta BR", "ETA dAlpha B", "ETA epsilon 1", "ETA epsilon 2", "ETA epsilon 4", "ETA epsilon 5", "ETA dEpsilon A", "ETA dEpsilon B", "ETA dEpsilon C", "AETA dBeta", "ETA psi 1", "ETA dPsi A", "nHBAcc", "nHBDon", "IC0", "IC1", "IC5", "TIC1", "TIC3", "TIC4", "TIC5", "SIC2", "SIC5", "BIC1", "BIC2", "BIC3", "BIC4", "BIC5", "CIC0", "MIC0", "ZMIC0", "ZMIC1", "ZMIC5", "Kier3", "LabuteASA", "PEOE VSA1", "PEOE VSA2", "PEOE VSA3", "PEOE VSA4", "PEOE VSA5", "PEOE VSA6", "PEOE VSA7", "PEOE VSA8", "PEOE VSA9", "PEOE VSA10", "PEOE VSA11", "PEOE VSA12", "SMR VSA2", "SMR VSA3", "SMR VSA6", "SMR VSA7", "SlogP VSA1", "SlogP VSA2", "SlogP VSA3", "SlogP VSA5", "SlogP VSA6", "SlogP VSA10", "SlogP VSA11", "MDEC-13", "MDEC-23", "MDEC-33", "MID h", "MID N", "AMID N", "MID O", "AMID O", "MPC4", "MPC5", "MPC6", "TMPC10", "piPC3", "piPC4", "piPC5", "piPC6", "TpiPC10", "n4Ring", "nHRing", "n3HRing", "n5aRing", "n6aRing", "n5aHRing", "n4ARing", "nAHRing", "n3AHRing", "n5AHRing", "n5FRing", "n7FRing", "n5FARing", "n7FARing", "nRot", "RotRatio", "SLogP", "TopoPSA(NO)", "TopoPSA", "GGI2", "JGI3", "JGT10", "MWC09", "SRW05", "SRW07", "SRW09", "TSRW10", "MW", "WPol", "mZagreb2"]
- **E2-PBE0**: ["n6aHRing", "nAcid", "nBase", "SpAbs A", "SpMax A", "SpDiam A", "SpAD A", "SpMAD A", "VR1 A", "VR2 A", "VR3 A", "nBridgehead", "nH", "ATS0dv", "ATS2dv", "ATS3dv", "ATS0s", "ATS2s", "ATS4s", "ATS3Z", "ATS4Z", "ATS0m", "ATS3m", "ATS2pe", "ATS2are", "ATS3are", "ATS3i", "AATS3dv", "AATS4dv", "AATS3d", "AATS0s", "AATS1s", "AATS2s", "AATS3Z", "AATS2m", "AATS3m", "AATS4m", "AATS3v", "AATS0se", "AATS4se", "AATS0pe", "AATS2pe", "AATS4pe", "AATS2are", "AATS4are", "AATS5are", "AATS2p", "AATS0i", "AATS1i", "AATS5i", "AATS6i", "ATSC0c", "ATSC1c", "ATSC2c", "ATSC3c", "ATSC4c", "ATSC0dv", "ATSC2dv", "ATSC3d", "ATSC4d", "ATSC5d", "ATSC0s", "ATSC1s", "ATSC3s", "ATSC5s", "ATSC2m", "ATSC0v", "ATSC2se", "ATSC4se", "ATSC1pe", "ATSC5pe", "ATSC2i", "ATSC3i", "AATSC0c", "AATSC1c", "AATSC2c", "AATSC0d", "AATSC3d", "AATSC6d", "AATSC0s", "AATSC1s", "AATSC2s", "AATSC3s", "AATSC1Z", "AATSC0m", "AATSC1m", "AATSC4m", "AATSC6m", "AATSC1v", "AATSC1se", "AATSC2se", "AATSC6se", "AATSC1pe", "AATSC4pe", "AATSC1are", "AATSC0i", "MATS1c", "MATS2c", "MATS4c", "MATS1dv", "MATS2dv", "MATS4dv", "MATS6dv", "MATS2d", "MATS3d", "MATS1s", "MATS2Z", "MATS6m", "MATS2se", "MATS5pe", "MATS1are", "MATS4are", "MATS4p", "MATS5p", "MATS1i", "MATS2i", "GATS1c", "GATS2c", "GATS1dv", "GATS1d", "GATS4d", "GATS1s", "GATS1Z", "GATS2m", "GATS5m", "GATS1v", "GATS2v", "GATS4v", "GATS1se", "GATS2se", "GATS1pe", "GATS2pe", "GATS1are", "GATS5are", "GATS1p", "GATS1i", "GATS2i", "GATS5i", "GATS6i", "BCUTc-1h", "BCUTc-1l", "BCUTdv-1h", "BCUTd-1h", "BCUTd-1l", "BCUTs-1h", "BCUTs-1l", "BCUTZ-1h", "BCUTZ-1l",

"BCUTm-1h", "BCUTm-1l", "BCUTv-1h", "BCUTv-1l", "BCUTse-1h", "BCUTse-1l", "BCUTpe-1h", "BCUTpe-1l", "BCUTare-1h", "BCUTare-1l", "BCUTp-1h", "BCUTp-1l", "BCUTi-1h", "BCUTi-1l", "SpAbs DzZ", "SpDiam DzZ", "SpAD DzZ", "SpMAD DzZ", "SM1 DzZ", "VE2 DzZ", "VR2 DzZ", "VR3 DzZ", "SpAbs Dzm", "SpMax Dzm", "SpDiam Dzm", "SpAD Dzm", "SpMAD Dzm", "LogEE Dzm", "SM1 Dzm", "VE1 Dzm", "VE2 Dzm", "VE3 Dzm", "VR1 Dzm", "VR2 Dzm", "SpMax Dzv", "SpDiam Dzv", "LogEE Dzv", "VE1 Dzv", "VE2 Dzv", "VE3 Dzv", "VR2 Dzv", "SpAbs Dzse", "SpMax Dzse", "SpDiam Dzse", "SpAD Dzse", "SpMAD Dzse", "LogEE Dzse", "VR3 Dzse", "SpDiam Dzpe", "SM1 Dzpe", "VE2 Dzpe", "VR3 Dzpe", "VR1 Dzare", "VR2 Dzare", "VR3 Dzare", "SpAbs Dzp", "SpMax Dzp", "SpDiam Dzp", "SpMAD Dzp", "LogEE Dzp", "VE1 Dzp", "SpMax Dzi", "SpDiam Dzi", "SpAD Dzi", "SpMAD Dzi", "LogEE Dzi", "VE2 Dzi", "nBondsD", "nBondsM", "nBondsKS", "nBondsKD", "RNCG", "RPCG", "C1SP2", "C2SP2", "C3SP2", "HybRatio", "FCSP3", "Xch-3d", "Xch-4d", "Xch-5d", "Xch-6d", "Xch-7d", "Xch-3dv", "Xch-4dv", "Xch-5dv", "Xch-6dv", "Xc-4d", "Xc-3dv", "Xpc-4d", "Xpc-5d", "Xpc-6d", "Xpc-5dv", "Xp-0d", "Xp-1d", "Xp-2d", "Xp-4d", "Xp-6d", "AXp-1d", "AXp-2d", "AXp-3d", "AXp-4d", "Xp-3dv", "Xp-5dv", "AXp-0dv", "AXp-1dv", "AXp-2dv", "AXp-3dv", "AXp-4dv", "AXp-5dv", "AXp-6dv", "Sv", "Mm", "Mi", "VR1 D", "VR2 D", "VR3 D", "ETA alpha", "AETA alpha", "ETA shape p", "ETA shape y", "ETA beta", "AETA beta", "AETA beta s", "ETA beta ns", "AETA beta ns", "AETA eta L", "AETA eta RL", "AETA eta F", "ETA eta FL", "ETA eta B", "AETA eta B", "AETA eta BR", "ETA epsilon 1", "ETA epsilon 2", "ETA epsilon 4", "ETA epsilon 5", "ETA dEpsilon A", "ETA dEpsilon B", "ETA dEpsilon C", "ETA dBeta", "AETA dBeta", "ETA psi 1", "ETA dPsi A", "IC1", "IC2", "IC3", "IC5", "TIC3", "TIC4", "TIC5", "SIC1", "SIC2", "BIC1", "BIC2", "BIC5", "CIC1", "CIC2", "CIC3", "MIC0", "MIC1", "MIC3", "MIC5", "ZMIC1", "ZMIC3", "Kier3", "PEOE VSA1", "PEOE VSA2", "PEOE VSA3", "PEOE VSA4", "PEOE VSA5", "PEOE VSA6", "PEOE VSA7", "PEOE VSA8", "PEOE VSA9", "PEOE VSA10", "PEOE VSA11", "PEOE VSA12", "PEOE VSA13", "SMR VSA1", "SMR VSA2", "SMR VSA3", "SMR VSA5", "SMR VSA6", "SMR VSA7", "SMR VSA9", "SlogP VSA1", "SlogP VSA2", "SlogP VSA3", "SlogP VSA4", "SlogP VSA6", "SlogP VSA10", "SlogP VSA11", "MDEC-22", "MDEC-23", "MDEC-33", "AMID", "MID h", "AMID h", "AMID C", "MID N", "AMID N", "MID O", "MID X", "MPC2", "MPC4", "MPC6", "piPC2", "piPC3", "piPC4", "piPC5", "piPC6", "TpiPC10", "bpol", "nRing", "n4Ring", "nHRing", "n4HRing", "n5HRing", "n5aRing", "n5aHRing", "nARing", "n3ARing", "n4ARing", "n4AHRing", "n5FRing", "n5FARing", "SLogP", "SMR", "TopoPSA(NO)", "TopoPSA", "GGI2", "GGI3", "GGI4", "JGI2", "JGI3", "JGI4", "Radius", "SRW05", "SRW06", "SRW07", "SRW09", "SRW10", "TSRW10"]

- **f1-CAM**: ["C4SP3", "VE2 A", "nBridgehead", "ATS4d", "ATS0s", "ATS1Z", "ATS8Z", "ATS1m", "ATS2se", "ATS6pe", "AATS4dv", "AATS0s", "AATS2s", "AATS0p", "AATS1p", "AATS2p", "AATS3p", "AATS1i", "ATSC1c", "ATSC3c", "ATSC4c", "ATSC5c", "ATSC4d", "ATSC0s", "ATSC2v", "ATSC4v", "ATSC1p", "AATSC0c", "AATSC1c", "AATSC0s", "AATSC6v", "AATSC1pe", "MATS1c", "MATS2c", "MATS2s", "MATS5se", "MATS1pe", "MATS1are", "MATS6are", "MATS6p", "MATS5i", "GATS5c", "GATS1dv", "GATS2dv", "GATS1d", "GATS2s", "GATS1se", "GATS5se", "GATS3pe", "GATS2are", "BCUTc-1h", "BCUTc-1l", "BCUTdv-1h", "BCUTd-1h", "BCUTd-1l", "BCUTs-1h", "BCUTs-1l", "BCUTm-1h", "BCUTm-1l", "BCUTv-1h", "BCUTv-1l", "BCUTse-1h", "BCUTse-1l", "BCUTpe-1h", "BCUTpe-1l", "BCUTare-1h", "BCUTare-1l", "BCUTp-1h", "BCUTp-1l", "BCUTi-1h", "VE1 DzZ", "VE2 DzZ", "VE3 DzZ", "VE1 Dzm", "VE3 Dzm", "VE1 Dzv", "VR1 Dzv", "VR2 Dzv", "VE1 Dzse", "VE2 Dzse", "VE3 Dzse", "SpAD Dzpe", "VE1 Dzpe", "VE3 Dzpe", "VE1 Dzare", "VE3 Dzare", "VE2 Dzi", "nBondsS", "nBondsD", "nBondsKD", "RNCG", "RPCG", "C1SP2", "C2SP2", "C3SP2", "HybRatio", "FCSP3", "Xch-5dv", "Xp-5d", "AXp-3d", "Xp-2dv", "Xp-5dv", "Xp-6dv", "AXp-1dv", "AXp-4dv", "AXp-6dv", "Mse", "Mpe", "Mi", "VE2 D", "ETA shape y", "ETA beta", "AETA beta", "AETA beta s", "ETA beta ns", "AETA beta ns", "ETA beta ns d", "AETA beta ns d", "AETA eta", "AETA eta F", "ETA eta FL", "ETA epsilon 2", "ETA epsilon 5", "ETA dEpsilon B", "ETA dEpsilon C", "ETA dBeta", "AETA dBeta", "IC2", "TIC0", "TIC1", "TIC2", "TIC3", "TIC4", "SIC1", "BIC1", "BIC3", "BIC4", "BIC5", "CIC2", "ZMIC0", "ZMIC2", "Kier3", "PEOE VSA1", "PEOE VSA2", "PEOE VSA3", "PEOE VSA4", "PEOE VSA7", "PEOE VSA8", "PEOE VSA9", "PEOE VSA10", "SMR VSA3", "SMR VSA4", "SMR VSA5", "SMR VSA7", "SlogP VSA2", "SlogP VSA4", "SlogP VSA5", "SlogP VSA6", "SlogP VSA10", "MDEC-23", "MID h", "AMID h", "AMID C", "MID N", "AMID N", "piPC4", "piPC5", "n3HRing", "nARing", "n4ARing", "n5ARing", "n7ARing", "nAHRing", "n3AHRing", "nFARing", "nFAHRing", "SMR", "TopoPSA(NO)", "TopoPSA", "SRW10", "TSRW10", "AMW"]

- **f1-CC2**: ["nBridgehead", "SpDiam A", "ATS2pe", "AATS2dv", "AATS4are", "AATS1p", "AATS2p", "AATS4p", "AATS5i", "ATSC0c", "ATSC2c", "ATSC3c", "ATSC3d", "ATSC0s", "ATSC4s", "ATSC1Z", "ATSC4Z", "AATSC1d", "AATSC3Z", "AATSC2pe", "AATSC0i", "AATSC1i", "MATS1se", "MATS2pe", "MATS4are", "GATS5c", "GATS2dv", "GATS1d", "GATS2d", "GATS2s", "GATS1se", "GATS2se", "GATS1pe", "GATS2are", "BCUTc-1h", "BCUTc-1l", "BCUTdv-1h", "BCUTd-1h", "BCUTd-1l", "BCUTs-1h", "BCUTs-1l", "BCUTZ-1h", "BCUTv-1h", "BCUTv-1l", "BCUTse-1h", "BCUTse-1l", "BCUTpe-1h", "BCUTpe-1l", "BCUTare-1h", "BCUTare-1l", "BCUTp-1h", "BCUTp-1l", "BCUTi-1h", "BCUTi-1l", "LogEE Dzv", "SpAD Dzpe", "VE2 Dzpe", "VR2 Dzpe",

"VR3 Dzpe", "VR1 Dzare", "LogEE Dzp", "VR1 Dzp", "VR2 Dzp", "VR3 Dzp", "SpDiam Dzi", "VE1 Dzi", "VE2 Dzi", "VE3 Dzi", "nBondsS", "nBondsM", "nBondsKD", "RNCG", "RPCG", "C1SP2", "C2SP2", "C3SP2", "HybRatio", "FCSP3", "Xch-5dv", "Xch-7dv", "Xc-4d", "Xc-3dv", "Xpc-5dv", "Xp-4d", "Xp-7d", "AXp-0d", "AXp-3dv", "AXp-5dv", "AXp-6dv", "Sm", "Mp", "SpMAD D", "ETA beta", "ETA beta ns d", "AETA beta ns d", "AETA eta L", "ETA eta R", "AETA eta R", "AETA eta F", "ETA eta FL", "AETA eta FL", "ETA epsilon 5", "AETA dBeta", "IC1", "TIC1", "TIC2", "SIC2", "PEOE VSA2", "PEOE VSA3", "PEOE VSA4", "PEOE VSA5", "PEOE VSA7", "PEOE VSA9", "PEOE VSA11", "PEOE VSA12", "SMR VSA1", "SMR VSA2", "SMR VSA4", "SMR VSA7", "SlogP VSA2", "SlogP VSA3", "SlogP VSA6", "SlogP VSA8", "SlogP VSA10", "AMID", "MID h", "MID C", "AMID C", "MID N", "AMID N", "piPC1", "piPC4", "piPC5", "piPC6", "apol", "n6Ring", "n4HRing", "n6aRing", "n6aHRing", "n3AHRing", "n4AHRing", "nFRing", "nFHRing", "nFARing", "n8FAHRing", "nRot", "RotRatio", "SLogP", "JGT10", "MWC03", "TMWC10", "SRW07", "SRW09", "TSRW10", "Zagreb1", "Zagreb2"]

- **f1-PBE0**: ["VE1 A", "VE2 A", "VE3 A", "nAromAtom", "nAromBond", "nBridgehead", "ATS3s", "ATS5se", "AATS2dv", "AATS1s", "AATS2s", "AATS2Z", "AATS4m", "AATS0v", "AATS4se", "AATS1pe", "AATS4are", "AATS0i", "AATS2i", "ATSC0c", "ATSC1c", "ATSC2c", "ATSC3c", "ATSC4c", "ATSC2d", "ATSC3d", "ATSC0s", "ATSC2Z", "ATSC0se", "AATSC4c", "AATSC6m", "AATSC5v", "AATSC0se", "AATSC1se", "AATSC3p", "AATSC6p", "AATSC1i", "MATS6c", "MATS2dv", "MATS2d", "MATS3d", "MATS6s", "MATS1se", "MATS6i", "GATS1c", "GATS5c", "GATS1dv", "GATS2d", "GATS3d", "GATS2s", "GATS1Z", "GATS5v", "GATS2se", "BCUTc-1h", "BCUTc-1l", "BCUTd-1h", "BCUTd-1l", "BCUTs-1h", "BCUTs-1l", "BCUTZ-1h", "BCUTm-1h", "BCUTv-1h", "BCUTv-1l", "BCUTse-1h", "BCUTpe-1l", "BCUTp-1h", "BCUTi-1l", "VE1 DzZ", "SM1 Dzm", "VE2 Dzm", "SpAbs Dzv", "SpMax Dzv", "VR2 Dzv", "VR3 Dzv", "VE3 Dzare", "VR1 Dzare", "VR2 Dzare", "SpAD Dzp", "LogEE Dzp", "VE1 Dzp", "SpDiam Dzi", "VE1 Dzi", "VR1 Dzi", "VR2 Dzi", "VR3 Dzi", "nBondsO", "nBondsD", "nBondsT", "nBondsA", "nBondsM", "nBondsKD", "RPCG", "C1SP2", "C2SP2", "C3SP2", "FCSP3", "Xch-4dv", "Xch-7dv", "Xpc-6dv", "Xp-5d", "AXp-3d", "AXp-5d", "Xp-2dv", "VE2 D", "AETA beta", "ETA beta ns", "AETA beta ns", "ETA eta L", "AETA eta L", "ETA eta F", "AETA eta F", "ETA eta FL", "AETA eta FL", "AETA dBeta", "nHBDon", "IC1", "IC2", "TIC1", "TIC2", "SIC0", "SIC2", "BIC1", "MIC2", "ZMIC2", "PEOE VSA1", "PEOE VSA3", "PEOE VSA4", "PEOE VSA6", "PEOE VSA7", "PEOE VSA9", "PEOE VSA10", "SMR VSA3", "SMR VSA4", "SMR VSA7", "SlogP VSA2", "SlogP VSA3", "SlogP VSA5", "SlogP VSA6", "SlogP VSA8", "SlogP VSA10", "SlogP VSA11", "AMID", "MID h", "MID N", "AMID N", "MPC4", "piPC2", "piPC4", "piPC6", "piPC7", "TpiPC10", "bpol", "n5aRing", "n5aHRing", "n3ARing", "n5AHRing", "nFRing", "nFHRing", "nFARing", "nFAHRing", "SMR", "GGI1", "JGI3", "JGT10", "SRW03", "SRW09", "MW", "AMW", "mZagreb2"]
- **f2-CAM**: ["VR1 A", "VR2 A", "ATS3m", "AATS2dv", "AATS1d", "AATS2d", "AATS5d", "AATS0s", "AATS2v", "AATS0se", "AATS4se", "AATS5se", "AATS1are", "AATS3are", "AATS5are", "AATS0p", "AATS1p", "AATS2p", "AATS2i", "ATSC0c", "ATSC2c", "ATSC3c", "ATSC3dv", "ATSC4dv", "ATSC7d", "ATSC2s", "ATSC2se", "ATSC1p", "ATSC1i", "AATSC1c", "AATSC2c", "AATSC1v", "AATSC6pe", "AATSC1are", "AATSC2are", "AATSC0p", "MATS1c", "MATS2c", "MATS1dv", "MATS3d", "MATS4d", "MATS2s", "MATS1m", "MATS5m", "MATS2se", "MATS5se", "MATS1pe", "MATS1i", "GATS1c", "GATS2dv", "GATS1s", "GATS2s", "GATS2se", "GATS3se", "GATS1are", "GATS1i", "BCUTc-1h", "BCUTc-1l", "BCUTdv-1h", "BCUTd-1l", "BCUTs-1h", "BCUTs-1l", "BCUTZ-1h", "BCUTZ-1l", "BCUTv-1h", "BCUTse-1h", "BCUTpe-1h", "BCUTare-1h", "BCUTare-1l", "BCUTp-1h", "BCUTp-1l", "BCUTi-1h", "BCUTi-1l", "VE1 DzZ", "VE2 DzZ", "VE3 DzZ", "SM1 Dzm", "VE1 Dzm", "VE2 Dzm", "VE3 Dzm", "LogEE Dzv", "VE1 Dzv", "VE2 Dzv", "VE3 Dzv", "SpMax Dzpe", "VR1 Dzpe", "SM1 Dzare", "VR3 Dzare", "VE1 Dzp", "VE2 Dzp", "VE3 Dzp", "VR1 Dzp", "VR2 Dzp", "VR3 Dzp", "SpMAD Dzi", "nBondsD", "nBondsKD", "RNCG", "C1SP2", "C2SP2", "C3SP2", "HybRatio", "Xch-3d", "Xch-3dv", "Xch-4dv", "Xc-5dv", "Xpc-4d", "Xpc-4dv", "AXp-4d", "AXp-5d", "Xp-3dv", "Xp-4dv", "Xp-5dv", "AXp-1dv", "AXp-2dv", "AXp-3dv", "AXp-4dv", "AXp-5dv", "AXp-6dv", "AETA beta", "ETA beta ns", "AETA beta ns", "ETA eta F", "ETA eta FL", "ETA epsilon 5", "ETA dEpsilon A", "AETA dBeta", "nHBAcc", "IC1", "TIC0", "TIC1", "SIC1", "SIC2", "BIC3", "CIC1", "CIC2", "CIC3", "MIC1", "MIC4", "ZMIC1", "ZMIC4", "ZMIC5", "PEOE VSA2", "PEOE VSA3", "PEOE VSA5", "PEOE VSA6", "PEOE VSA7", "PEOE VSA8", "PEOE VSA9", "PEOE VSA10", "PEOE VSA12", "PEOE VSA13", "SMR VSA2", "SMR VSA5", "SMR VSA7", "SlogP VSA1", "SlogP VSA2", "SlogP VSA5", "SlogP VSA6", "MDEC-13", "MDEC-23", "MDEC-33", "MID h", "AMID h", "AMID C", "MID N", "AMID N", "MID O", "AMID O", "piPC3", "piPC4", "piPC5", "TpiPC10", "nRing", "n3Ring", "n5Ring", "n5HRing", "n6ARing", "n5AHRing", "SLogP", "SMR", "TopoPSA(NO)", "TopoPSA", "GGI3", "GGI4", "JGI2", "JGT10", "MWC03", "MWC04", "MWC05", "MWC06", "MWC09", "MWC10", "TMWC10", "SRW07", "SRW09", "TSRW10", "Zagreb2"]
- **f2-CC2**: ["SpDiam A", "VE1 A", "ATS1s", "ATS7s", "ATS4m", "ATS5m", "ATS6are", "ATS3i", "AATS2dv", "AATS0s", "AATS2s", "AATS3s", "AATS2m", "AATS3se", "AATS2i", "ATSC2c", "ATSC3c", "ATSC4c", "ATSC3d", "ATSC0s", "ATSC1s", "ATSC1m", "ATSC2se", "ATSC1i",

"AATSC0c", "AATSC2c", "AATSC3c", "AATSC6c", "AATSC3d", "AATSC1se", "AATSC1pe", "AATSC2pe", "AATSC6p", "AATSC5i", "MATS1c", "MATS5dv", "MATS1d", "MATS1s", "MATS2s", "MATS1se", "MATS2se", "MATS4se", "MATS1pe", "MATS2are", "MATS1i", "GATS1c", "GATS2c", "GATS3c", "GATS6c", "GATS2s", "GATS1m", "GATS1v", "GATS2se", "GATS3se", "GATS1pe", "GATS5pe", "GATS5p", "GATS1i", "BCUTc-1h", "BCUTc-1l", "BCUTdv-1h", "BCUTd-1h", "BCUTd-1l", "BCUTs-1h", "BCUTZ-1h", "BCUTm-1h", "BCUTm-1l", "BCUTv-1l", "BCUTse-1h", "BCUTse-1l", "BCUTpe-1h", "BCUTpe-1l", "BCUTare-1h", "BCUTare-1l", "BCUTp-1h", "BCUTp-1l", "BCUTi-1h", "BCUTi-1l", "SpMAD Dzm", "VE1 Dzm", "VE3 Dzm", "VR2 Dzm", "VR3 Dzm", "SpMax Dzv", "SpDiam Dzv", "SpAD Dzv", "SM1 Dzv", "VE2 Dzv", "VR3 Dzv", "SpDiam Dzse", "VR3 Dzse", "VE2 Dzpe", "VR1 Dzpe", "VR2 Dzpe", "VR3 Dzpe", "SpMax Dzare", "SpMAD Dzare", "VR1 Dzp", "SpAbs Dzi", "SpMax Dzi", "SpDiam Dzi", "SpAD Dzi", "SpMAD Dzi", "LogEE Dzi", "VE2 Dzi", "nBondsD", "nBondsM", "nBondsKD", "RNCG", "RPCG", "C1SP2", "C2SP3", "HybRatio", "FCSP3", "Xch-4d", "Xch-3dv", "Xch-4dv", "Xch-5dv", "Xc-3dv", "Xpc-4dv", "Xpc-5dv", "Xp-4d", "AXp-3dv", "AXp-5dv", "Mi", "SpDiam D", "VE1 D", "VE2 D", "VE3 D", "VR3 D", "AETA alpha", "AETA beta s", "ETA beta ns", "AETA eta", "ETA eta L", "AETA eta L", "ETA eta F", "AETA eta F", "ETA eta FL", "AETA eta FL", "ETA dAlpha B", "ETA epsilon 2", "ETA dEpsilon A", "ETA dEpsilon B", "ETA dEpsilon C", "ETA dEpsilon D", "IC1", "IC4", "TIC1", "BIC2", "MIC1", "ZMIC1", "PEOE VSA1", "PEOE VSA2", "PEOE VSA6", "PEOE VSA8", "PEOE VSA9", "PEOE VSA10", "PEOE VSA11", "PEOE VSA12", "PEOE VSA13", "SMR VSA2", "SMR VSA5", "SMR VSA6", "SMR VSA7", "SlogP VSA1", "SlogP VSA2", "SlogP VSA4", "SlogP VSA5", "SlogP VSA6", "SlogP VSA11", "MDEC-12", "MDEC-13", "MDEC-22", "AMID", "MID C", "MID N", "AMID N", "MID O", "AMID O", "MPC3", "MPC4", "MPC7", "piPC3", "piPC4", "piPC5", "piPC6", "TpiPC10", "n6Ring", "n3HRing", "n6HRing", "nARing", "n3AHRing", "n5AHRing", "n5FHRing", "SMR", "TopoPSA(NO)", "TopoPSA", "GGI1", "JGI2", "JGI3", "JGI5", "SRW05", "SRW09", "SRW10"]

- **f2-PBE0**: ["SpMax A", "ATS3Z", "ATS4Z", "ATS0se", "ATS1se", "ATS2se", "ATS1pe", "ATS1p", "ATS2p", "ATS3p", "AATS2dv", "AATS3dv", "AATS0d", "AATS0s", "AATS2s", "AATS0m", "AATS2m", "AATS0v", "AATS1v", "AATS2v", "AATS0p", "AATS2p", "AATS0i", "AATS2i", "ATSC1c", "ATSC2c", "ATSC3c", "ATSC2dv", "ATSC7d", "ATSC2s", "ATSC1se", "ATSC1pe", "ATSC3are", "ATSC1p", "ATSC1i", "AATSC0c", "AATSC4dv", "AATSC0s", "AATSC1s", "AATSC2s", "AATSC3Z", "AATSC0m", "AATSC1se", "AATSC2se", "AATSC3se", "AATSC1pe", "AATSC5pe", "AATSC3are", "AATSC1p", "AATSC4p", "AATSC5p", "AATSC0i", "AATSC1i", "AATSC2i", "AATSC3i", "MATS2c", "MATS1Z", "MATS1m", "MATS1pe", "MATS5p", "MATS6i", "GATS1c", "GATS3c", "GATS6c", "GATS1s", "GATS2s", "GATS3s", "GATS4s", "GATS5m", "GATS3se", "GATS2pe", "GATS1are", "GATS2are", "GATS3are", "BCUTc-1h", "BCUTc-1l", "BCUTdv-1h", "BCUTdv-1l", "BCUTd-1l", "BCUTs-1h", "BCUTs-1l", "BCUTZ-1h", "BCUTm-1h", "BCUTm-1l", "BCUTv-1h", "BCUTv-1l", "BCUTse-1h", "BCUTse-1l", "BCUTpe-1h", "BCUTpe-1l", "BCUTare-1h", "BCUTare-1l", "BCUTp-1h", "BCUTp-1l", "BCUTi-1h", "SM1 Dzv", "VE3 Dzv", "SpAD Dzse", "SpMAD Dzse", "LogEE Dzse", "SpMax Dzare", "LogEE Dzare", "VE1 Dzare", "VE3 Dzare", "VE1 Dzp", "VR2 Dzp", "VR3 Dzp", "SpMax Dzi", "SpDiam Dzi", "SpAD Dzi", "SpMAD Dzi", "LogEE Dzi", "VR1 Dzi", "VR3 Dzi", "nBondsS", "nBondsD", "nBondsT", "RNCG", "RPCG", "C2SP1", "C1SP2", "C2SP2", "C3SP2", "HybRatio", "FCSP3", "Xch-3d", "Xch-3dv", "Xch-4dv", "Xp-1d", "Xp-1dv", "Xp-4dv", "AXp-1dv", "AXp-3dv", "AXp-5dv", "Mi", "SpMAD D", "VR1 D", "VR3 D", "AETA alpha", "ETA shape p", "ETA shape y", "ETA beta ns", "AETA beta ns", "AETA beta ns d", "AETA eta L", "ETA eta F", "ETA eta FL", "ETA dAlpha B", "ETA epsilon 4", "ETA dEpsilon B", "ETA dEpsilon C", "AETA dBeta", "IC1", "IC2", "IC5", "TIC1", "SIC1", "SIC2", "SIC4", "BIC2", "MIC1", "MIC2", "MIC3", "ZMIC1", "Kier3", "PEOE VSA1", "PEOE VSA2", "PEOE VSA9", "PEOE VSA12", "SMR VSA2", "SMR VSA3", "SMR VSA7", "SlogP VSA5", "SlogP VSA6", "SlogP VSA10", "MDEC-22", "MDEC-23", "MDEC-33", "AMID h", "MID C", "MID N", "piPC1", "piPC4", "piPC6", "TpiPC10", "n3Ring", "n4Ring", "nHRing", "n5HRing", "n6ARing", "nAHRing", "n5AHRing", "n8FHRing", "nRot", "SLogP", "SMR", "TopoPSA(NO)", "TopoPSA", "JGI2", "JGT10", "MWC03", "MWC04", "MWC05", "MWC06", "MWC07", "MWC08", "MWC09", "MWC10", "TMWC10", "SRW05", "SRW06", "SRW07", "TSRW10"]

Detailed causal features for each task that compose **QM9 dataset** is detailed below:

- μ: ["nF", "nX", "nAcid", "nBase", "nBridgehead", "nHetero", "ATS3dv", "ATS4dv", "ATS4s", "ATS3m", "ATS5m", "ATS6m", "AATS1dv", "AATS3s", "AATS4s", "AATS5s", "AATS6s", "AATS4se", "AATS4pe", "AATS3are", "AATS6are", "AATS5i", "AATS6i", "ATSC0c", "ATSC2c", "ATSC4c", "ATSC5c", "ATSC6c", "ATSC7c", "ATSC2dv", "ATSC5dv", "ATSC7dv", "ATSC8dv", "ATSC0s", "ATSC3s", "ATSC5s", "ATSC6s", "ATSC7s", "ATSC8s", "ATSC5v", "ATSC2se", "ATSC8se", "ATSC6pe", "ATSC6are", "ATSC7p", "ATSC4i", "AATSC4c", "AATSC5c", "AATSC6c", "AATSC2dv", "AATSC4dv", "AATSC5dv", "AATSC6dv", "AATSC1s", "AATSC2s", "AATSC3s", "AATSC5s", "AATSC6s", "AATSC1are", "MATS1c", "MATS4c", "MATS3dv", "MATS5dv", "MATS1s", "MATS6s", "MATS3m", "MATS3pe", "MATS6are", "MATS2i", "GATS1c", "GATS4c",

"GATS5c", "GATS1dv", "GATS2dv", "GATS3dv", "GATS1d", "GATS1s", "GATS2s", "GATS3s", "GATS4s", "GATS5s", "GATS6s", "GATS2Z", "GATS5Z", "GATS2m", "GATS3m", "GATS1v", "GATS5v", "GATS2se", "GATS5se", "GATS6se", "GATS3pe", "GATS3are", "GATS1i", "BCUTc-1h", "BCUTc-1l", "BCUTdv-1h", "BCUTd-1l", "BCUTs-1h", "BCUTs-1l", "BCUTZ-1h", "BCUTZ-1l", "BCUTm-1h", "BCUTv-1l", "BCUTse-1h", "BCUTpe-1h", "BCUTpe-1l", "BCUTare-1h", "BCUTare-1l", "BCUTp-1h", "BCUTp-1l", "BCUTi-1h", "SM1 DzZ", "SM1 Dzm", "VE1 Dzm", "VE3 Dzm", "SpMAD Dzv", "SM1 Dzv", "VE1 Dzv", "VR1 Dzv", "SM1 Dzse", "SM1 Dzpe", "SM1 Dzare", "SpMax Dzp", "SpDiam Dzp", "SpAD Dzp", "SpMAD Dzp", "LogEE Dzp", "SM1 Dzp", "SM1 Dzi", "VE1 Dzi", "VE2 Dzi", "VE3 Dzi", "nBondsD", "nBondsA", "nBondsM", "nBondsKD", "RNCG", "RPCG", "C1SP2", "C2SP2", "HybRatio", "FCSP3", "Xch-3d", "Xch-4d", "Xch-3dv", "Xch-4dv", "Xc-4d", "Xc-3dv", "Xc-4dv", "Xc-6dv", "Xpc-4d", "Xpc-5dv", "AXp-5d", "AXp-6d", "AXp-0dv", "AXp-5dv", "AXp-6dv", "AXp-7dv", "Sp", "MZ", "Mm", "Mi", "AETA alpha", "ETA shape y", "ETA beta ns", "AETA beta ns", "ETA eta L", "AETA eta L", "ETA eta RL", "AETA eta F", "AETA eta FL", "AETA eta BR", "ETA dAlpha B", "ETA epsilon 1", "ETA epsilon 2", "ETA epsilon 4", "ETA epsilon 5", "ETA dEpsilon C", "ETA dEpsilon D", "AETA dBeta", "ETA psi 1", "ETA dPsi A", "ETA dPsi B", "nHBAcc", "IC0", "IC1", "IC2", "TIC4", "TIC5", "SIC5", "CIC0", "CIC4", "MIC2", "MIC3", "MIC4", "MIC5", "ZMIC2", "Kier3", "PEOE VSA1", "PEOE VSA2", "PEOE VSA3", "PEOE VSA4", "PEOE VSA5", "PEOE VSA6", "PEOE VSA8", "PEOE VSA9", "PEOE VSA10", "PEOE VSA11", "PEOE VSA12", "PEOE VSA13", "SMR VSA1", "SMR VSA2", "SMR VSA3", "SMR VSA7", "SlogP VSA1", "SlogP VSA2", "SlogP VSA5", "SlogP VSA6", "SlogP VSA10", "SlogP VSA11", "MDEC-12", "MDEC-22", "AMID", "MID h", "AMID h", "AMID C", "MID N", "AMID N", "MID O", "MID X", "AMID X", "MPC5", "piPC2", "piPC3", "piPC4", "piPC5", "piPC6", "piPC7", "piPC8", "nRing", "n5Ring", "n6Ring", "nHRing", "n3HRing", "n4HRing", "n5HRing", "n6HRing", "n7HRing", "n5aRing", "n5aHRing", "n6aHRing", "nARing", "n6ARing", "nAHRing", "n3AHRing", "n4AHRing", "n5AHRing", "n6AHRing", "nFHRing", "n8FARing", "SLogP", "SMR", "TopoPSA(NO)", "TopoPSA", "GGI5", "JGI2", "JGI3", "JGI5", "JGI6", "SRW07", "SRW08", "SRW10", "TSRW10", "mZagreb1", "mZagreb2"]

- α: ["n7AHRing", "n7FAHRing", "nAcid", "nBase", "SpDiam A", "VR3 A", "nAromAtom", "nAromBond", "nSpiro", "nBridgehead", "ATS4dv", "ATS1d", "ATS2d", "ATS3d", "ATS4d", "ATS6d", "ATS7d", "ATS8d", "ATS6s", "ATS2Z", "ATS3m", "ATS6m", "ATS7m", "ATS0v", "ATS1v", "ATS4v", "ATS5v", "ATS7v", "ATS4se", "ATS5se", "ATS2pe", "ATS6pe", "ATS7pe", "ATS8pe", "ATS5are", "ATS0p", "ATS4p", "ATS6p", "ATS7p", "ATS1i", "ATS5i", "AATS0d", "AATS1d", "AATS4d", "AATS6d", "AATS1s", "AATS0v", "AATS1v", "AATS2v", "AATS3v", "AATS1se", "AATS2se", "AATS1pe", "AATS2pe", "AATS1are", "AATS5p", "AATS4i", "ATSC0c", "ATSC1c", "ATSC2c", "ATSC7c", "ATSC1d", "ATSC6d", "ATSC0s", "ATSC2s", "ATSC1Z", "ATSC7Z", "ATSC0m", "ATSC1m", "ATSC7m", "ATSC1v", "ATSC8v", "ATSC0se", "ATSC7p", "ATSC0i", "ATSC6i", "ATSC7i", "AATSC0c", "AATSC3dv", "AATSC4d", "AATSC6d", "AATSC0s", "AATSC3s", "AATSC0v", "AATSC1v", "AATSC0se", "AATSC1se", "AATSC0are", "AATSC4are", "MATS3d", "MATS5d", "MATS6d", "MATS2s", "MATS1v", "MATS4pe", "MATS1p", "GATS1c", "GATS2c", "GATS1dv", "GATS1d", "GATS3s", "GATS1m", "GATS1v", "GATS2se", "GATS5pe", "GATS1are", "GATS1p", "GATS1i", "BCUTc-1h", "BCUTc-1l", "BCUTdv-1h", "BCUTdv-1l", "BCUTd-1h", "BCUTd-1l", "BCUTs-1h", "BCUTs-1l", "BCUTZ-1h", "BCUTZ-1l", "BCUTm-1h", "BCUTm-1l", "BCUTv-1h", "BCUTv-1l", "BCUTse-1h", "BCUTse-1l", "BCUTpe-1l", "BCUTare-1h", "BCUTare-1l", "BCUTp-1h", "BCUTp-1l", "BCUTi-1h", "BCUTi-1l", "SpAbs DzZ", "SpMax DzZ", "SpAD DzZ", "LogEE DzZ", "VE1 DzZ", "VE2 DzZ", "VE3 DzZ", "SpAbs Dzm", "SpMax Dzm", "SpDiam Dzm", "SpAD Dzm", "LogEE Dzm", "SM1 Dzm", "VE2 Dzm", "SpAbs Dzv", "SpMax Dzv", "VE2 Dzv", "SpAbs Dzse", "SpAD Dzse", "SpMAD Dzse", "VE2 Dzse", "VE2 Dzpe", "VE1 Dzare", "VE3 Dzare", "LogEE Dzp", "SM1 Dzp", "SpDiam Dzi", "VE2 Dzi", "nBondsD", "nBondsM", "nBondsKD", "RNCG", "C2SP1", "C1SP2", "C2SP2", "C3SP2", "C1SP3", "HybRatio", "FCSP3", "Xch-3d", "Xch-4d", "Xch-5d", "Xch-6d", "Xch-7d", "Xch-3dv", "Xch-4dv", "Xch-5dv", "Xch-7dv", "Xc-4d", "Xc-5d", "Xc-6d", "Xc-4dv", "Xp-5d", "AXp-2d", "AXp-5d", "AXp-6d", "AXp-7d", "Xp-0dv", "Xp-1dv", "Xp-3dv", "Xp-7dv", "AXp-6dv", "AXp-7dv", "Sm", "Sv", "Mse", "Mp", "Mi", "VE1 D", "VE2 D", "VE3 D", "ETA alpha", "ETA shape y", "AETA beta", "AETA beta s", "AETA beta ns d", "AETA eta", "AETA eta RL", "ETA eta FL", "AETA eta FL", "ETA eta BR", "AETA eta BR", "ETA epsilon 3", "ETA epsilon 4", "ETA epsilon 5", "ETA dEpsilon A", "ETA dEpsilon B", "ETA dEpsilon C", "nHBDon", "TIC3", "SIC3", "BIC3", "BIC5", "MIC1", "MIC5", "ZMIC2", "LabuteASA", "PEOE VSA1", "PEOE VSA2", "PEOE VSA3", "PEOE VSA5", "PEOE VSA6", "PEOE VSA7", "PEOE VSA8", "PEOE VSA9", "PEOE VSA10", "PEOE VSA11", "PEOE VSA12", "PEOE VSA13", "SMR VSA1", "SMR VSA2", "SMR VSA3", "SMR VSA4", "SMR VSA5", "SMR VSA6", "SMR VSA7", "SMR VSA9", "SlogP VSA1", "SlogP VSA2", "SlogP VSA3", "SlogP VSA5", "SlogP VSA6", "SlogP VSA8", "SlogP VSA10", "SlogP VSA11", "MDEC-12", "MID h", "AMID h", "MID C", "MID N", "AMID N", "MID O", "AMID O", "MPC5", "piPC1", "piPC2", "piPC3", "piPC4", "piPC5", "piPC6", "piPC7", "TpiPC10", "apol", "nRing", "n4Ring", "n5Ring", "n6Ring", "n7Ring", "n4HRing", "n5HRing", "naRing", "n5aRing", "n6aRing", "naHRing",

"n5aHRing", "n6aHRing", "nARing", "n4ARing", "n6ARing", "n4AHRing", "n5AHRing", "n5FRing", "n7FRing", "n6FHRing", "n7FARing", "n6FAHRing", "nRot", "RotRatio", "SLogP", "SMR", "GGI7", "JGI3", "JGI4", "JGT10", "TopoShapeIndex", "PetitjeanIndex", "SRW06", "SRW07", "SRW10", "MW", "mZagreb1", "mZagreb2"]

- ϵ_{homo}: ["C2SP1", "n8FARing", "n9FAHRing", "GGI6", "nAcid", "nBase", "SpDiam A", "LogEE A", "VE1 A", "nAromAtom", "nAromBond", "nSpiro", "nBridgehead", "ATS0dv", "ATS3dv", "ATS4dv", "ATS5dv", "ATS2d", "ATS0s", "ATS1s", "ATS3s", "ATS4s", "ATS2Z", "ATS7v", "ATS0are", "ATS2are", "ATS5p", "ATS1i", "AATS3dv", "AATS4dv", "AATS0d", "AATS1d", "AATS3d", "AATS4d", "AATS0s", "AATS1s", "AATS2s", "AATS4s", "AATS3m", "AATS0v", "AATS1v", "AATS2se", "AATS1pe", "AATS2pe", "AATS3pe", "AATS2p", "AATS1i", "AATS2i", "ATSC0c", "ATSC1c", "ATSC2c", "ATSC3c", "ATSC4c", "ATSC5c", "ATSC6c", "ATSC7c", "ATSC4d", "ATSC0s", "ATSC1s", "ATSC3s", "ATSC5Z", "ATSC0m", "ATSC1m", "ATSC3m", "ATSC1se", "ATSC0are", "ATSC1are", "ATSC6are", "ATSC0p", "ATSC1i", "ATSC3i", "AATSC0c", "AATSC1c", "AATSC2c", "AATSC0d", "AATSC1d", "AATSC3d", "AATSC0s", "AATSC1s", "AATSC2s", "AATSC3s", "AATSC5s", "AATSC1Z", "AATSC0m", "AATSC1m", "AATSC2m", "AATSC1pe", "AATSC0p", "AATSC0i", "MATS1c", "MATS3c", "MATS2dv", "MATS2d", "MATS3d", "MATS1s", "MATS2s", "MATS1Z", "MATS1v", "MATS1pe", "MATS1p", "MATS2p", "MATS3i", "GATS1c", "GATS2c", "GATS3c", "GATS4c", "GATS6c", "GATS1dv", "GATS3dv", "GATS1d", "GATS3d", "GATS4d", "GATS1s", "GATS2s", "GATS3s", "GATS4s", "GATS1Z", "GATS1m", "GATS3m", "GATS1v", "GATS1se", "GATS2se", "GATS5se", "GATS1pe", "GATS2pe", "GATS1are", "GATS1p", "GATS2p", "GATS1i", "BCUTc-1h", "BCUTc-1l", "BCUTdv-1h", "BCUTdv-1l", "BCUTd-1h", "BCUTd-1l", "BCUTs-1h", "BCUTs-1l", "BCUTZ-1h", "BCUTZ-1l", "BCUTm-1h", "BCUTm-1l", "BCUTv-1h", "BCUTv-1l", "BCUTse-1h", "BCUTse-1l", "BCUTpe-1h", "BCUTpe-1l", "BCUTare-1h", "BCUTare-1l", "BCUTp-1h", "BCUTp-1l", "BCUTi-1h", "BCUTi-1l", "SpAbs DzZ", "SpMax DzZ", "SpAD DzZ", "SpMAD DzZ", "LogEE DzZ", "VE2 DzZ", "VR1 DzZ", "VR2 DzZ", "VR3 DzZ", "SpAbs Dzm", "SpMax Dzm", "VE2 Dzm", "SpAbs Dzv", "SpAD Dzv", "VR3 Dzv", "SpAbs Dzse", "VE1 Dzpe", "VE3 Dzpe", "SpAbs Dzare", "SpMax Dzare", "SpDiam Dzare", "SpAD Dzare", "SpMAD Dzare", "LogEE Dzare", "SpDiam Dzp", "SpMax Dzi", "SpDiam Dzi", "LogEE Dzi", "SM1 Dzi", "VR2 Dzi", "nBondsD", "nBondsA", "nBondsM", "nBondsKS", "nBondsKD", "RNCG", "RPCG", "C1SP2", "C2SP2", "C3SP2", "HybRatio", "FCSP3", "Xch-3d", "Xch-4d", "Xch-5d", "Xch-7d", "Xch-3dv", "Xch-4dv", "Xch-5dv", "Xch-6dv", "Xch-7dv", "Xc-5d", "Xc-6d", "Xc-4dv", "Xc-5dv", "Xpc-5d", "Xp-3d", "Xp-4d", "Xp-5d", "Xp-7d", "AXp-1d", "AXp-3d", "AXp-4d", "AXp-5d", "AXp-7d", "Xp-2dv", "Xp-5dv", "Xp-7dv", "AXp-5dv", "AXp-6dv", "Si", "Mp", "SpDiam D", "SpMAD D", "VE1 D", "VE3 D", "ETA alpha", "ETA shape p", "ETA shape y", "AETA beta", "ETA beta ns", "AETA beta ns", "ETA beta ns d", "AETA beta ns d", "AETA eta L", "AETA eta RL", "ETA eta F", "AETA eta F", "ETA eta FL", "ETA eta B", "AETA eta B", "ETA eta BR", "AETA eta BR", "ETA dAlpha B", "ETA epsilon 4", "ETA epsilon 5", "ETA dEpsilon B", "ETA dEpsilon C", "ETA dEpsilon D", "AETA dBeta", "nHBAcc", "nHBDon", "IC1", "IC2", "TIC1", "TIC5", "BIC1", "BIC2", "BIC5", "MIC0", "MIC2", "MIC3", "MIC4", "MIC5", "ZMIC1", "ZMIC4", "PEOE VSA1", "PEOE VSA2", "PEOE VSA3", "PEOE VSA4", "PEOE VSA5", "PEOE VSA6", "PEOE VSA7", "PEOE VSA8", "PEOE VSA9", "PEOE VSA10", "PEOE VSA11", "PEOE VSA12", "PEOE VSA13", "SMR VSA1", "SMR VSA2", "SMR VSA3", "SMR VSA4", "SMR VSA6", "SMR VSA7", "SMR VSA9", "SlogP VSA1", "SlogP VSA2", "SlogP VSA3", "SlogP VSA4", "SlogP VSA5", "SlogP VSA6", "SlogP VSA8", "SlogP VSA10", "SlogP VSA11", "MDEC-12", "MDEC-13", "MDEC-22", "MDEC-33", "AMID", "MID h", "AMID h", "MID C", "AMID C", "MID N", "AMID N", "MID O", "AMID O", "MID X", "AMID X", "MPC5", "MPC6", "TMPC10", "piPC1", "piPC2", "piPC3", "piPC4", "piPC5", "piPC6", "piPC7", "piPC8", "TpiPC10", "nRing", "n3Ring", "n4Ring", "n5Ring", "n6Ring", "nHRing", "n3HRing", "n4HRing", "n5HRing", "n6HRing", "naRing", "n5aRing", "n6aRing", "n5aHRing", "n6aHRing", "nARing", "n3ARing", "n4ARing", "n5ARing", "n6ARing", "nAHRing", "n3AHRing", "n4AHRing", "n5AHRing", "n6AHRing", "n5FRing", "n6FRing", "n7FRing", "nFHRing", "n7FHRing", "n5FARing", "n7FARing", "n7FAHRing", "nRot", "RotRatio", "SLogP", "TopoPSA(NO)", "TopoPSA", "GGI2", "GGI3", "GGI4", "GGI5", "JGI1", "JGI2", "JGI3", "JGI4", "JGT10", "TopoShapeIndex", "PetitjeanIndex", "MWC08", "SRW05", "SRW07", "SRW08", "SRW09", "TSRW10", "WPath", "mZagreb2"]
- ϵ_{lumo}: ["C4SP3", "AMID X", "n7ARing", "nFHRing", "n5FHRing", "n6FHRing", "nBase", "SpMax A", "SpDiam A", "SpMAD A", "LogEE A", "VE1 A", "VE2 A", "VE3 A", "VR1 A", "VR2 A", "VR3 A", "nAromAtom", "nAromBond", "nSpiro", "nBridgehead", "ATS2dv", "ATS3dv", "ATS4dv", "ATS5dv", "ATS2d", "ATS0s", "ATS1s", "ATS2s", "ATS3s", "ATS4s", "ATS5s", "ATS3Z", "ATS5Z", "ATS1m", "ATS3m", "ATS1v", "ATS1se", "ATS1pe", "ATS2pe", "ATS3pe", "ATS0are", "ATS1are", "ATS2are", "ATS0p", "ATS5p", "ATS1i", "ATS2i", "ATS6i", "AATS0dv", "AATS1dv", "AATS3dv", "AATS5dv", "AATS6dv", "AATS0d", "AATS1d", "AATS2d", "AATS3d", "AATS1s", "AATS2s", "AATS3s", "AATS5s", "AATS6s", "AATS0Z", "AATS1Z", "AATS3Z", "AATS4Z", "AATS5Z", "AATS0m", "AATS1m", "AATS3m", "AATS5m", "AATS0v", "AATS1v", "AATS2v", "AATS3v", "AATS4v", "AATS3se", "AATS5se", "AATS1pe", "AATS2pe", "AATS3pe", "AATS6pe", "AATS0p",

"AATS1p", "AATS4p", "AATS5p", "AATS0i", "AATS6i", "ATSC0c", "ATSC1c", "ATSC2c", "ATSC3c", "ATSC6c", "ATSC7c", "ATSC2dv", "ATSC6dv", "ATSC7dv", "ATSC3d", "ATSC4d", "ATSC7d", "ATSC0s", "ATSC2s", "ATSC3s", "ATSC5s", "ATSC0Z", "ATSC5Z", "ATSC1m", "ATSC0se", "ATSC1se", "ATSC2se", "ATSC4se", "ATSC0pe", "ATSC1pe", "ATSC3pe", "ATSC1p", "ATSC0i", "ATSC1i", "ATSC3i", "AATSC0c", "AATSC1c", "AATSC2c", "AATSC3c", "AATSC4c", "AATSC6c", "AATSC0dv", "AATSC1dv", "AATSC3dv", "AATSC4d", "AATSC5d", "AATSC6d", "AATSC0s", "AATSC1s", "AATSC2s", "AATSC3s", "AATSC4s", "AATSC0Z", "AATSC1Z", "AATSC0m", "AATSC0v", "AATSC1v", "AATSC3v", "AATSC2se", "AATSC5se", "AATSC0pe", "AATSC2pe", "AATSC1are", "AATSC2are", "AATSC1p", "AATSC5p", "AATSC6p", "AATSC1i", "AATSC6i", "MATS1c", "MATS2c", "MATS3c", "MATS4c", "MATS5c", "MATS1dv", "MATS2dv", "MATS3dv", "MATS1d", "MATS3d", "MATS2s", "MATS3s", "MATS6s", "MATS1m", "MATS1v", "MATS4v", "MATS2se", "MATS3se", "MATS4se", "MATS1are", "MATS2are", "MATS6are", "MATS1p", "MATS3p", "MATS1i", "GATS1c", "GATS2c", "GATS3c", "GATS5c", "GATS1dv", "GATS2dv", "GATS3dv", "GATS5dv", "GATS1d", "GATS2d", "GATS4d", "GATS5d", "GATS6d", "GATS1s", "GATS2s", "GATS3s", "GATS4s", "GATS1Z", "GATS1m", "GATS2m", "GATS3m", "GATS4m", "GATS1v", "GATS3v", "GATS1se", "GATS2se", "GATS5se", "GATS1pe", "GATS2pe", "GATS1are", "GATS2are", "GATS5p", "GATS1i", "BCUTc-1h", "BCUTc-1l", "BCUTdv-1h", "BCUTdv-1l", "BCUTd-1h", "BCUTd-1l", "BCUTs-1h", "BCUTs-1l", "BCUTZ-1h", "BCUTZ-1l", "BCUTm-1h", "BCUTm-1l", "BCUTv-1h", "BCUTv-1l", "BCUTse-1h", "BCUTse-1l", "BCUTpe-1h", "BCUTpe-1l", "BCUTare-1h", "BCUTare-1l", "BCUTp-1h", "BCUTp-1l", "BCUTi-1h", "BCUTi-1l", "SpMAD DzZ", "SM1 DzZ", "SpAbs Dzm", "SpAD Dzm", "SpMAD Dzm", "SM1 Dzm", "VE3 Dzm", "VR3 Dzm", "SpDiam Dzv", "VR1 Dzv", "VR3 Dzv", "SpAbs Dzse", "SpAD Dzse", "SpMAD Dzse", "LogEE Dzse", "VE1 Dzse", "VE3 Dzse", "SpMAD Dzpe", "LogEE Dzpe", "SpAbs Dzare", "SpDiam Dzare", "SpMAD Dzare", "VE1 Dzare", "VE2 Dzare", "VE3 Dzare", "VR2 Dzare", "VR3 Dzare", "SpAbs Dzp", "SpMax Dzp", "SpDiam Dzp", "SpAD Dzp", "SpMAD Dzp", "LogEE Dzp", "SM1 Dzp", "VE1 Dzp", "VR3 Dzp", "SM1 Dzi", "VE1 Dzi", "VE2 Dzi", "VE3 Dzi", "VR3 Dzi", "nBondsD", "nBondsT", "nBondsM", "nBondsKS", "nBondsKD", "RNCG", "RPCG", "C2SP1", "C1SP2", "C2SP2", "C3SP2", "C1SP3", "HybRatio", "FCSP3", "Xch-3d", "Xch-5d", "Xch-6d", "Xch-7d", "Xch-3dv", "Xch-4dv", "Xch-5dv", "Xch-6dv", "Xc-5d", "Xc-6d", "Xc-3dv", "Xc-4dv", "Xc-5dv", "Xpc-4d", "Xpc-6d", "Xpc-4dv", "Xp-0d", "Xp-3d", "Xp-4d", "Xp-5d", "Xp-6d", "Xp-7d", "AXp-2d", "AXp-3d", "AXp-4d", "AXp-5d", "AXp-6d", "AXp-7d", "Xp-1dv", "Xp-2dv", "Xp-5dv", "Xp-6dv", "Xp-7dv", "AXp-1dv", "AXp-2dv", "AXp-3dv", "AXp-4dv", "AXp-5dv", "AXp-6dv", "AXp-7dv", "SZ", "Sv", "Spe", "Sare", "Si", "MZ", "Mm", "Mv", "Mpe", "Mp", "Mi", "SpAD D", "VE1 D", "VE2 D", "VE3 D", "VR1 D", "VR2 D", "VR3 D", "ETA alpha", "AETA alpha", "ETA shape y", "ETA beta", "AETA beta", "ETA beta s", "AETA beta s", "ETA beta ns", "AETA beta ns", "ETA beta ns d", "AETA beta ns d", "ETA eta", "ETA eta L", "AETA eta L", "AETA eta R", "AETA eta RL", "ETA eta F", "AETA eta F", "ETA eta FL", "AETA eta FL", "ETA eta BR", "AETA eta BR", "ETA epsilon 1", "ETA epsilon 2", "ETA epsilon 4", "ETA epsilon 5", "ETA dEpsilon B", "ETA dEpsilon C", "ETA dEpsilon D", "ETA dBeta", "AETA dBeta", "nHBAcc", "nHBDon", "IC0", "IC1", "IC4", "IC5", "TIC1", "TIC2", "TIC4", "TIC5", "SIC0", "SIC1", "SIC4", "BIC1", "BIC2", "BIC3", "BIC4", "BIC5", "CIC0", "CIC1", "MIC0", "MIC1", "MIC2", "ZMIC0", "ZMIC1", "ZMIC3", "ZMIC4", "ZMIC5", "Kier3", "LabuteASA", "PEOE VSA1", "PEOE VSA2", "PEOE VSA3", "PEOE VSA4", "PEOE VSA5", "PEOE VSA6", "PEOE VSA7", "PEOE VSA8", "PEOE VSA9", "PEOE VSA10", "PEOE VSA11", "PEOE VSA12", "PEOE VSA13", "SMR VSA1", "SMR VSA2", "SMR VSA3", "SMR VSA4", "SMR VSA5", "SMR VSA6", "SMR VSA7", "SMR VSA9", "SlogP VSA1", "SlogP VSA2", "SlogP VSA3", "SlogP VSA4", "SlogP VSA5", "SlogP VSA6", "SlogP VSA8", "SlogP VSA10", "SlogP VSA11", "MDEC-12", "MDEC-13", "MDEC-22", "MDEC-23", "MDEC-33", "MID", "AMID", "MID h", "MID C", "AMID C", "MID N", "AMID N", "MID O", "AMID O", "MPC5", "MPC6", "MPC7", "MPC8", "TMPC10", "piPC2", "piPC3", "piPC4", "piPC5", "piPC6", "piPC7", "piPC8", "TpiPC10", "apol", "bpol", "nRing", "n4Ring", "n5Ring", "n7Ring", "nHRing", "n3HRing", "n4HRing", "n5HRing", "n8HRing", "naRing", "n5aRing", "n5aHRing", "n6aHRing", "nARing", "n4ARing", "n5ARing", "n6ARing", "nAHRing", "n3AHRing", "n4AHRing", "n5AHRing", "n8AHRing", "n5FRing", "n6FRing", "n7FRing", "n9FRing", "n7FHRing", "n9FHRing", "n5FARing", "n6FARing", "n7FARing", "n9FARing", "n6FAHRing", "n7FAHRing", "n9FAHRing", "RotRatio", "SLogP", "SMR", "TopoPSA(NO)", "TopoPSA", "GGI2", "GGI3", "GGI4", "JGI1", "JGI3", "JGI4", "JGT10", "Diameter", "TopoShapeIndex", "PetitjeanIndex", "MWC08", "MWC10", "SRW05", "SRW07", "SRW08", "SRW09", "TSRW10", "AMW", "WPol", "mZagreb1"]

- *gap*: ["MID X", "AMID X", "n7HRing", "n6FAHRing", "nAcid", "nBase", "SpDiam A", "VE1 A", "VE2 A", "VE3 A", "VR3 A", "nAromAtom", "nAromBond", "nSpiro", "nBridgehead", "ATS3dv", "ATS4dv", "ATS5dv", "ATS6s", "ATS2Z", "ATS3Z", "ATS0m", "ATS3m", "ATS5m", "ATS6m", "ATS1v", "ATS3v", "ATS0se", "ATS2se", "ATS3se", "ATS4se", "ATS0are", "ATS5are", "ATS2i", "AATS3dv", "AATS5dv", "AATS3d", "AATS6d", "AATS0s", "AATS1s", "AATS2s", "AATS3s", "AATS5s", "AATS6s", "AATS0Z", "AATS1Z", "AATS2Z", "AATS4Z", "AATS0m", "AATS1m", "AATS2m", "AATS0v", "AATS5v", "AATS2se", "AATS4pe", "AATS6pe", "AATS1are", "AATS2are",

"AATS3are", "AATS2p", "AATS1i", "AATS2i", "AATS6i", "ATSC0c", "ATSC1c", "ATSC2c", "ATSC3c", "ATSC4c", "ATSC5c", "ATSC6c", "ATSC2dv", "ATSC3dv", "ATSC4dv", "ATSC0d", "ATSC3d", "ATSC4d", "ATSC0s", "ATSC2s", "ATSC3s", "ATSC4s", "ATSC6m", "ATSC0se", "ATSC2se", "ATSC3se", "ATSC4se", "ATSC6se", "ATSC7se", "ATSC0pe", "ATSC4are", "ATSC0i", "ATSC1i", "ATSC3i", "AATSC2c", "AATSC5c", "AATSC6c", "AATSC2dv", "AATSC6dv", "AATSC1s", "AATSC2s", "AATSC3s", "AATSC0m", "AATSC1se", "AATSC2se", "AATSC0pe", "AATSC0are", "AATSC1are", "AATSC0p", "AATSC5p", "AATSC1i", "MATS1c", "MATS2c", "MATS3c", "MATS2dv", "MATS2d", "MATS3d", "MATS4d", "MATS3s", "MATS5s", "MATS6s", "MATS1m", "MATS1v", "MATS1se", "MATS2se", "MATS1pe", "MATS6pe", "MATS1are", "MATS1p", "GATS1c", "GATS2c", "GATS3c", "GATS5c", "GATS1dv", "GATS3dv", "GATS1d", "GATS2s", "GATS3s", "GATS4s", "GATS1Z", "GATS1v", "GATS1se", "GATS2se", "GATS2pe", "GATS3pe", "GATS1p", "GATS5p", "GATS1i", "GATS2i", "GATS3i", "GATS5i", "BCUTc-1h", "BCUTc-1l", "BCUTdv-1h", "BCUTdv-1l", "BCUTd-1h", "BCUTd-1l", "BCUTs-1h", "BCUTs-1l", "BCUTZ-1h", "BCUTZ-1l", "BCUTm-1h", "BCUTm-1l", "BCUTv-1h", "BCUTv-1l", "BCUTse-1h", "BCUTse-1l", "BCUTpe-1h", "BCUTpe-1l", "BCUTare-1h", "BCUTare-1l", "BCUTp-1h", "BCUTp-1l", "BCUTi-1h", "BCUTi-1l", "VE1 DzZ", "VE3 DzZ", "VE2 Dzm", "SpMax Dzv", "VE1 Dzv", "VE2 Dzv", "VE3 Dzv", "VE2 Dzse", "VE1 Dzpe", "VE2 Dzpe", "VE3 Dzpe", "VE3 Dzare", "SpDiam Dzp", "SM1 Dzp", "VE2 Dzp", "SpDiam Dzi", "SpAD Dzi", "VE3 Dzi", "VR1 Dzi", "nBondsD", "nBondsA", "nBondsM", "nBondsKD", "RNCG", "RPCG", "C1SP1", "C2SP1", "C1SP2", "C2SP2", "C3SP2", "C4SP3", "HybRatio", "FCSP3", "Xch-3d", "Xch-4d", "Xch-5d", "Xch-6d", "Xch-3dv", "Xch-4dv", "Xch-5dv", "Xch-6dv", "Xc-4d", "Xc-5d", "Xc-6d", "Xc-4dv", "Xc-6dv", "Xpc-4d", "Xpc-5d", "Xp-1d", "Xp-3d", "Xp-5d", "Xp-6d", "Xp-7d", "AXp-0d", "AXp-1d", "AXp-2d", "AXp-3d", "AXp-4d", "AXp-5d", "AXp-6d", "AXp-7d", "AXp-0dv", "AXp-2dv", "AXp-3dv", "AXp-5dv", "AXp-6dv", "AXp-7dv", "Sm", "Mv", "SpMAD D", "VE1 D", "VE3 D", "VR2 D", "ETA beta", "AETA beta", "ETA beta s", "AETA beta s", "ETA beta ns", "AETA beta ns", "ETA beta ns d", "AETA beta ns d", "AETA eta L", "ETA eta R", "AETA eta R", "AETA eta RL", "AETA eta F", "ETA eta FL", "ETA eta BR", "AETA eta BR", "ETA epsilon 2", "ETA epsilon 5", "ETA dEpsilon A", "ETA dEpsilon B", "ETA dBeta", "AETA dBeta", "ETA dPsi A", "nHBAcc", "IC0", "IC1", "IC4", "TIC1", "TIC3", "TIC4", "TIC5", "SIC0", "BIC1", "BIC2", "CIC1", "CIC5", "MIC1", "MIC3", "ZMIC0", "ZMIC1", "Kier1", "Kier3", "PEOE VSA1", "PEOE VSA2", "PEOE VSA3", "PEOE VSA4", "PEOE VSA5", "PEOE VSA6", "PEOE VSA7", "PEOE VSA8", "PEOE VSA9", "PEOE VSA10", "PEOE VSA11", "PEOE VSA12", "PEOE VSA13", "SMR VSA2", "SMR VSA3", "SMR VSA4", "SMR VSA5", "SMR VSA6", "SMR VSA7", "SMR VSA9", "SlogP VSA1", "SlogP VSA2", "SlogP VSA3", "SlogP VSA4", "SlogP VSA5", "SlogP VSA6", "SlogP VSA8", "SlogP VSA10", "SlogP VSA11", "MDEC-22", "MDEC-23", "MDEC-33", "MID", "AMID", "MID h", "AMID h", "AMID C", "MID N", "AMID N", "MID O", "AMID O", "MPC3", "MPC4", "MPC5", "MPC6", "MPC7", "MPC8", "TMPC10", "piPC1", "piPC2", "piPC3", "piPC4", "piPC5", "piPC6", "piPC7", "piPC8", "TpiPC10", "bpol", "nRing", "n4Ring", "n5Ring", "n6Ring", "n7Ring", "nHRing", "n3HRing", "n4HRing", "n5HRing", "naRing", "n5aRing", "n6aRing", "n5aHRing", "n6aHRing", "nARing", "n3ARing", "n4ARing", "n5ARing", "n6ARing", "n7ARing", "nAHRing", "n3AHRing", "n4AHRing", "n5AHRing", "n5FRing", "n6FRing", "n7FRing", "n5FHRing", "n7FHRing", "nFARing", "n5FARing", "n6FARing", "n7FARing", "n5FAHRing", "n7FAHRing", "nRot", "RotRatio", "SLogP", "TopoPSA(NO)", "TopoPSA", "GGI1", "GGI2", "GGI3", "GGI4", "JGI1", "JGI2", "JGI3", "JGI4", "Radius", "TopoShapeIndex", "PetitjeanIndex", "MWC07", "MWC08", "MWC09", "MWC10", "SRW03", "SRW04", "SRW05", "SRW06", "SRW07", "SRW08", "SRW09", "SRW10", "TSRW10", "WPol", "Zagreb1", "Zagreb2", "mZagreb1", "mZagreb2"]

- $\langle R^2 \rangle$: ["MID X", "n8ARing", "n7AHRing", "n6FRing", "n7FHRing", "n6FARing", "n7FARing", "n7FAHRing", "SpAbs A", "SpMax A", "SpDiam A", "SpAD A", "SpMAD A", "VE2 A", "VE3 A", "VR1 A", "VR2 A", "VR3 A", "nSpiro", "nBridgehead", "ATS4dv", "ATS5dv", "ATS6dv", "ATS1s", "ATS2s", "ATS6s", "ATS7s", "ATS4Z", "ATS5Z", "ATS6Z", "ATS7Z", "ATS8Z", "ATS2m", "ATS3m", "ATS4m", "ATS5m", "ATS6m", "ATS7m", "ATS4v", "ATS5v", "ATS6v", "ATS7v", "ATS2p", "ATS8i", "AATS0d", "AATS2d", "AATS5d", "AATS5m", "AATS2v", "AATS4v", "AATS5v", "AATS2pe", "AATS2are", "AATS2p", "AATS2i", "AATS3i", "ATSC3c", "ATSC4c", "ATSC7c", "ATSC8c", "ATSC5dv", "ATSC6dv", "ATSC3d", "ATSC5d", "ATSC7d", "ATSC3s", "ATSC7s", "ATSC3Z", "ATSC2v", "ATSC8v", "ATSC4se", "ATSC7pe", "ATSC1p", "ATSC3p", "ATSC6p", "ATSC1i", "ATSC3i", "ATSC4i", "ATSC7i", "AATSC1c", "AATSC3c", "AATSC4c", "AATSC6c", "AATSC0d", "AATSC2d", "AATSC3d", "AATSC5d", "AATSC6d", "AATSC3s", "AATSC4s", "AATSC3are", "AATSC4are", "AATSC2p", "AATSC1i", "MATS3c", "MATS1dv", "MATS2dv", "MATS2d", "MATS4d", "MATS6d", "MATS1s", "MATS6m", "MATS4v", "MATS5v", "MATS5p", "GATS2c", "GATS1dv", "GATS2d", "GATS3d", "GATS4d", "GATS5d", "GATS6d", "GATS4s", "GATS1Z", "GATS4Z", "GATS2v", "GATS6v", "GATS2se", "GATS6se", "GATS1pe", "GATS3pe", "GATS6are", "GATS3i", "BCUTc-1h", "BCUTdv-1h", "BCUTdv-1l", "BCUTd-1h", "BCUTd-1l", "BCUTs-1l", "BCUTZ-1l", "BCUTm-1l", "BCUTv-1h", "BCUTv-1l", "BCUTse-1h", "BCUTse-1l", "BCUTpe-1l", "BCUTare-1l", "BCUTp-1h", "BCUTp-1l", "BCUTi-1l", "SpMax DzZ", "SpDiam

DzZ", "SpMAD DzZ", "LogEE DzZ", "SM1 DzZ", "VE2 DzZ", "VE3 DzZ", "VR3 DzZ", "SpDiam Dzm", "SpMAD Dzm", "LogEE Dzm", "VE1 Dzm", "VE2 Dzm", "VE3 Dzm", "SpAbs Dzv", "SpMax Dzv", "SpDiam Dzv", "SpAD Dzv", "SpMAD Dzv", "LogEE Dzv", "VE1 Dzv", "VE2 Dzv", "VE3 Dzv", "VR1 Dzv", "SpAbs Dzse", "SpMax Dzse", "SpDiam Dzse", "SpMAD Dzse", "LogEE Dzse", "VE2 Dzse", "SpAbs Dzpe", "SpMax Dzpe", "SpDiam Dzpe", "SpMAD Dzpe", "LogEE Dzpe", "VR2 Dzpe", "SpAbs Dzare", "SpMax Dzare", "SpDiam Dzare", "SpAD Dzare", "SpMAD Dzare", "LogEE Dzare", "VE1 Dzare", "VE2 Dzare", "VE3 Dzare", "SpDiam Dzp", "LogEE Dzp", "VE1 Dzp", "VE2 Dzp", "VE3 Dzp", "SpMax Dzi", "SpDiam Dzi", "SpAD Dzi", "SpMAD Dzi", "LogEE Dzi", "VE1 Dzi", "VE3 Dzi", "VR1 Dzi", "C2SP2", "C3SP2", "C1SP3", "C2SP3", "C4SP3", "FCSP3", "Xch-3d", "Xch-4d", "Xch-5d", "Xch-6d", "Xch-7d", "Xch-3dv", "Xch-4dv", "Xch-5dv", "Xch-6dv", "Xch-7dv", "Xc-3d", "Xc-4d", "Xc-5d", "Xc-6d", "Xc-5dv", "Xc-6dv", "Xpc-4d", "Xpc-5d", "Xpc-4dv", "Xpc-5dv", "Xp-0d", "Xp-3d", "Xp-4d", "Xp-5d", "Xp-6d", "Xp-7d", "AXp-0d", "AXp-1d", "AXp-2d", "AXp-3d", "AXp-4d", "AXp-5d", "AXp-6d", "AXp-7d", "Xp-0dv", "Xp-1dv", "Xp-2dv", "Xp-4dv", "Xp-5dv", "Xp-6dv", "AXp-3dv", "AXp-4dv", "AXp-5dv", "AXp-6dv", "AXp-7dv", "SZ", "SpAbs D", "SpMax D", "SpDiam D", "SpAD D", "SpMAD D", "LogEE D", "VE1 D", "VE2 D", "VE3 D", "AETA alpha", "ETA shape p", "ETA shape y", "ETA shape x", "ETA beta s", "AETA beta s", "ETA beta ns", "AETA beta ns", "ETA eta R", "AETA eta R", "ETA eta RL", "AETA eta RL", "ETA eta F", "AETA eta F", "AETA eta FL", "ETA eta B", "AETA eta B", "ETA eta BR", "AETA eta BR", "ETA dAlpha B", "ETA epsilon 2", "ETA epsilon 4", "ETA dEpsilon D", "ETA dBeta", "AETA dBeta", "nHBDon", "TIC5", "BIC0", "BIC2", "BIC4", "BIC5", "MIC1", "ZMIC0", "ZMIC1", "ZMIC3", "ZMIC5", "Kier3", "PEOE VSA2", "PEOE VSA3", "PEOE VSA4", "PEOE VSA5", "PEOE VSA6", "PEOE VSA9", "PEOE VSA12", "PEOE VSA13", "SMR VSA3", "SMR VSA6", "SMR VSA7", "SlogP VSA1", "SlogP VSA2", "SlogP VSA3", "SlogP VSA5", "SlogP VSA6", "SlogP VSA8", "SlogP VSA10", "MDEC-12", "MDEC-13", "MDEC-22", "MDEC-23", "MDEC-33", "AMID", "MID h", "AMID h", "MID N", "AMID N", "AMID O", "MPC4", "MPC5", "MPC7", "MPC8", "TMPC10", "piPC2", "piPC3", "piPC4", "piPC5", "piPC6", "piPC7", "piPC8", "TpiPC10", "nRing", "n4Ring", "n5Ring", "n4HRing", "n6HRing", "n5aHRing", "n4ARing", "n6ARing", "nAHRing", "n4AHRing", "n5AHRing", "n6AHRing", "n5FRing", "nRot", "RotRatio", "SMR", "GGI1", "GGI2", "GGI3", "GGI4", "GGI5", "JGI1", "JGI2", "JGI3", "JGI4", "JGI5", "JGT10", "Radius", "TopoShapeIndex", "PetitjeanIndex", "MWC03", "MWC05", "MWC06", "MWC07", "MWC08", "MWC09", "MWC10", "TMWC10", "SRW04", "SRW05", "SRW06", "SRW07", "SRW08", "SRW09", "SRW10", "TSRW10", "MW", "WPath", "Zagreb1", "Zagreb2", "mZagreb1", "mZagreb2"]

- $ZPVE$: ["nF", "nX", "n6HRing", "n7ARing", "n6AHRing", "n8FHRing", "n7FARing", "JGI6", "SRW03", "nAcid", "nBase", "SpAbs A", "SpDiam A", "SpAD A", "SpMAD A", "LogEE A", "VE1 A", "VE2 A", "VE3 A", "VR1 A", "VR2 A", "VR3 A", "nAtom", "nHeavyAtom", "nSpiro", "nBridgehead", "nH", "nN", "ATS1dv", "ATS4dv", "ATS5dv", "ATS1d", "ATS2d", "ATS3d", "ATS4d", "ATS0s", "ATS1s", "ATS2s", "ATS0Z", "ATS1Z", "ATS2Z", "ATS4Z", "ATS0m", "ATS1m", "ATS2m", "ATS4m", "ATS0v", "ATS1v", "ATS2v", "ATS0se", "ATS1se", "ATS2se", "ATS3se", "ATS5se", "ATS0pe", "ATS1pe", "ATS2pe", "ATS0are", "ATS1are", "ATS2are", "ATS0p", "ATS1p", "ATS2p", "ATS0i", "ATS1i", "ATS2i", "ATS3i", "AATS0dv", "AATS1dv", "AATS2dv", "AATS3dv", "AATS0d", "AATS0s", "AATS1s", "AATS2s", "AATS3s", "AATS4s", "AATS0Z", "AATS1Z", "AATS0m", "AATS1m", "AATS2m", "AATS0v", "AATS1v", "AATS3v", "AATS0se", "AATS1se", "AATS2se", "AATS3se", "AATS0pe", "AATS2pe", "AATS3pe", "AATS0are", "AATS2are", "AATS3are", "AATS0p", "AATS2p", "AATS3p", "AATS5p", "AATS0i", "AATS1i", "AATS2i", "AATS3i", "AATS5i", "AATS6i", "ATSC0c", "ATSC1c", "ATSC3c", "ATSC4c", "ATSC0dv", "ATSC2dv", "ATSC0d", "ATSC1d", "ATSC4d", "ATSC0s", "ATSC0Z", "ATSC1Z", "ATSC0m", "ATSC1m", "ATSC2m", "ATSC5m", "ATSC0v", "ATSC1v", "ATSC0se", "ATSC0pe", "ATSC1pe", "ATSC0are", "ATSC0p", "ATSC1p", "ATSC2p", "ATSC0i", "ATSC1i", "ATSC3i", "AATSC2c", "AATSC4c", "AATSC0dv", "AATSC1dv", "AATSC2dv", "AATSC0d", "AATSC1d", "AATSC2d", "AATSC3d", "AATSC0s", "AATSC0Z", "AATSC1Z", "AATSC0m", "AATSC1m", "AATSC0v", "AATSC0se", "AATSC1se", "AATSC2se", "AATSC6se", "AATSC0pe", "AATSC1pe", "AATSC2pe", "AATSC0are", "AATSC1are", "AATSC2are", "AATSC0p", "AATSC2p", "AATSC0i", "AATSC1i", "MATS1c", "MATS3c", "MATS6d", "MATS2s", "MATS1Z", "MATS2m", "MATS1v", "MATS2v", "MATS1se", "MATS2se", "MATS3se", "MATS1pe", "MATS1are", "MATS1p", "MATS2p", "MATS1i", "MATS5i", "MATS6i", "GATS1c", "GATS1dv", "GATS6d", "GATS1s", "GATS2s", "GATS4s", "GATS1Z", "GATS2Z", "GATS3Z", "GATS1m", "GATS2m", "GATS4m", "GATS1v", "GATS1se", "GATS2se", "GATS3pe", "GATS1are", "GATS5p", "GATS1i", "GATS2i", "BCUTc-1h", "BCUTc-1l", "BCUTdv-1h", "BCUTdv-1l", "BCUTd-1l", "BCUTs-1h", "BCUTs-1l", "BCUTZ-1h", "BCUTZ-1l", "BCUTm-1h", "BCUTm-1l", "BCUTv-1h", "BCUTv-1l", "BCUTse-1h", "BCUTse-1l", "BCUTpe-1h", "BCUTpe-1l", "BCUTare-1h", "BCUTare-1l", "BCUTp-1h", "BCUTp-1l", "BCUTi-1h", "BCUTi-1l", "SpAbs DzZ", "SpMax DzZ", "SpDiam DzZ", "SpAD DzZ", "SpMAD DzZ", "LogEE DzZ", "VE1 DzZ", "VE2 DzZ", "VE3 DzZ", "VR2 DzZ", "VR3 DzZ", "SpAbs Dzm", "SpMax Dzm", "SpDiam Dzm", "SpAD Dzm", "SpMAD Dzm", "LogEE Dzm", "VE1 Dzm", "VE2 Dzm", "VE3 Dzm", "VR2 Dzm", "VR3 Dzm", "LogEE Dzv", "SM1 Dzv", "VE1 Dzv", "VE2 Dzv",

"VE3 Dzv", "VR1 Dzv", "VR2 Dzv", "VR3 Dzv", "SpMax Dzse", "SM1 Dzse", "VE1 Dzse", "VE2 Dzse", "VE3 Dzse", "VR2 Dzse", "VE1 Dzpe", "VE2 Dzpe", "VE3 Dzpe", "VR2 Dzpe", "VR3 Dzpe", "LogEE Dzare", "VE1 Dzare", "VE2 Dzare", "VE3 Dzare", "VR1 Dzare", "VR2 Dzare", "VR3 Dzare", "SpDiam Dzp", "SpMAD Dzp", "VE1 Dzp", "VE2 Dzp", "VE3 Dzp", "VR1 Dzp", "VR3 Dzp", "SpAbs Dzi", "SpMax Dzi", "LogEE Dzi", "SM1 Dzi", "VE1 Dzi", "VE2 Dzi", "VE3 Dzi", "VR1 Dzi", "VR2 Dzi", "VR3 Dzi", "nBonds", "nBondsS", "nBondsD", "nBondsA", "nBondsM", "nBondsKS", "nBondsKD", "RNCG", "RPCG", "C1SP2", "C2SP2", "C3SP2", "HybRatio", "FCSP3", "Xch-3d", "Xch-4d", "Xch-5d", "Xch-6d", "Xch-3dv", "Xch-4dv", "Xch-5dv", "Xch-6dv", "Xch-7dv", "Xc-3d", "Xc-4d", "Xc-5d", "Xc-4dv", "Xc-5dv", "Xpc-4d", "Xpc-5d", "Xp-0d", "Xp-1d", "Xp-2d", "Xp-3d", "Xp-4d", "Xp-5d", "Xp-6d", "Xp-7d", "AXp-0d", "AXp-2d", "AXp-3d", "AXp-4d", "AXp-6d", "Xp-0dv", "Xp-1dv", "Xp-2dv", "Xp-3dv", "Xp-4dv", "Xp-5dv", "Xp-7dv", "AXp-3dv", "AXp-5dv", "SZ", "Sm", "Sv", "Sse", "Spe", "Sare", "Sp", "Si", "MZ", "Mm", "Mv", "Mse", "Mpe", "Mare", "Mi", "SpAbs D", "SpMax D", "SpDiam D", "SpAD D", "SpMAD D", "LogEE D", "VE1 D", "VE2 D", "VE3 D", "VR1 D", "VR2 D", "VR3 D", "ETA alpha", "AETA alpha", "ETA shape p", "ETA shape y", "ETA shape x", "ETA beta", "AETA beta", "ETA beta s", "AETA beta s", "AETA beta ns", "ETA eta", "AETA eta", "ETA eta L", "AETA eta L", "ETA eta R", "ETA eta RL", "AETA eta RL", "ETA eta F", "ETA eta B", "AETA eta B", "ETA dAlpha B", "ETA epsilon 1", "ETA epsilon 2", "ETA epsilon 4", "ETA epsilon 5", "ETA dEpsilon A", "ETA dEpsilon B", "ETA dEpsilon C", "ETA dEpsilon D", "ETA dBeta", "AETA dBeta", "ETA dPsi A", "IC0", "IC2", "IC3", "IC4", "IC5", "TIC0", "TIC2", "SIC0", "SIC2", "SIC5", "BIC0", "BIC2", "BIC3", "BIC4", "CIC0", "CIC1", "CIC3", "CIC4", "CIC5", "MIC0", "MIC1", "MIC2", "MIC3", "MIC4", "MIC5", "ZMIC0", "ZMIC1", "ZMIC2", "ZMIC3", "ZMIC4", "Kier3", "LabuteASA", "PEOE VSA2", "PEOE VSA3", "PEOE VSA4", "PEOE VSA5", "PEOE VSA6", "PEOE VSA7", "PEOE VSA8", "PEOE VSA9", "PEOE VSA10", "PEOE VSA11", "PEOE VSA12", "PEOE VSA13", "SMR VSA1", "SMR VSA3", "SMR VSA4", "SMR VSA5", "SMR VSA6", "SMR VSA7", "SMR VSA9", "SlogP VSA1", "SlogP VSA2", "SlogP VSA3", "SlogP VSA4", "SlogP VSA5", "SlogP VSA6", "SlogP VSA8", "SlogP VSA10", "SlogP VSA11", "MDEC-12", "MDEC-13", "MDEC-23", "MDEC-33", "MID", "AMID", "MID h", "AMID h", "AMID C", "MID N", "AMID N", "MID O", "MPC5", "TMPC10", "piPC2", "piPC3", "piPC4", "piPC5", "piPC6", "piPC7", "piPC8", "TpiPC10", "apol", "bpol", "nRing", "n3Ring", "n4Ring", "n5Ring", "n6Ring", "nHRing", "n3HRing", "n4HRing", "n5HRing", "n7HRing", "nARing", "n3ARing", "n4ARing", "n5ARing", "n6ARing", "nAHRing", "n3AHRing", "n4AHRing", "n5AHRing", "n5FRing", "n6FRing", "n5FARing", "n6FARing", "n8FARing", "n9FARing", "RotRatio", "SLogP", "SMR", "GGI4", "GGI5", "JGI1", "JGI2", "JGI3", "JGI4", "JGI5", "JGT10", "SRW05", "SRW07", "SRW08", "SRW09", "TSRW10", "MW", "AMW", "WPol", "mZagreb1"]

- U: ["ATSC8c", "nBase", "LogEE A", "VE1 A", "VE2 A", "VE3 A", "VR3 A", "nAromAtom", "nAromBond", "nAtom", "nHeavyAtom", "nSpiro", "nBridgehead", "nHetero", "nH", "nC", "nN", "nO", "nF", "nX", "ATS0dv", "ATS1dv", "ATS2dv", "ATS3d", "ATS4d", "ATS5d", "ATS0s", "ATS1s", "ATS2s", "ATS3s", "ATS4s", "ATS5s", "ATS6s", "ATS7s", "ATS8s", "ATS0Z", "ATS1Z", "ATS0m", "ATS0v", "ATS1v", "ATS2v", "ATS4v", "ATS5v", "ATS0se", "ATS1se", "ATS2se", "ATS3se", "ATS4se", "ATS5se", "ATS0pe", "ATS1pe", "ATS2pe", "ATS3pe", "ATS4pe", "ATS5pe", "ATS0are", "ATS1are", "ATS2are", "ATS3are", "ATS4are", "ATS5are", "ATS0p", "ATS1p", "ATS2p", "ATS3p", "ATS4p", "ATS5p", "ATS0i", "ATS1i", "ATS2i", "ATS3i", "ATS4i", "ATS5i", "AATS0dv", "AATS1dv", "AATS2dv", "AATS3dv", "AATS1s", "AATS2s", "AATS0Z", "AATS1Z", "AATS2Z", "AATS0m", "AATS1m", "AATS2m", "AATS0v", "AATS1v", "AATS0se", "AATS1se", "AATS2se", "AATS4se", "AATS0pe", "AATS1pe", "AATS2pe", "AATS4pe", "AATS0are", "AATS1are", "AATS2are", "AATS4are", "AATS0p", "AATS2p", "AATS0i", "AATS1i", "AATS4i", "ATSC0s", "ATSC6s", "ATSC7s", "ATSC8s", "ATSC0Z", "ATSC0m", "ATSC0v", "ATSC0se", "ATSC0pe", "ATSC0are", "ATSC0p", "ATSC0i", "AATSC0c", "AATSC0dv", "AATSC0Z", "AATSC0m", "AATSC0v", "AATSC0se", "AATSC1se", "AATSC0pe", "AATSC1pe", "AATSC0are", "AATSC1are", "AATSC0p", "AATSC0i", "GATS1s", "GATS2s", "GATS1Z", "GATS1m", "GATS1se", "GATS1pe", "BCUTZ-1h", "BCUTm-1h", "BCUTv-1l", "BCUTse-1h", "BCUTpe-1h", "BCUTare-1h", "BCUTi-1h", "SpAbs DzZ", "SpMax DzZ", "SpDiam DzZ", "SpAD DzZ", "LogEE DzZ", "SM1 DzZ", "VE1 DzZ", "VE2 DzZ", "VE3 DzZ", "VR1 DzZ", "VR3 DzZ", "SpAbs Dzm", "SpMax Dzm", "SpDiam Dzm", "SpAD Dzm", "LogEE Dzm", "SM1 Dzm", "VE1 Dzm", "VE2 Dzm", "VE3 Dzm", "VR1 Dzm", "VR3 Dzm", "SpAbs Dzv", "SpMax Dzv", "SpDiam Dzv", "SpAD Dzv", "LogEE Dzv", "SM1 Dzv", "VE1 Dzv", "VE2 Dzv", "VE3 Dzv", "VR1 Dzv", "VR3 Dzv", "SpAbs Dzse", "SpMax Dzse", "SpDiam Dzse", "SpAD Dzse", "SpMAD Dzse", "LogEE Dzse", "SM1 Dzse", "VE1 Dzse", "VE2 Dzse", "VE3 Dzse", "VR1 Dzse", "VR3 Dzse", "SpAbs Dzpe", "SpMax Dzpe", "SpDiam Dzpe", "SpAD Dzpe", "LogEE Dzpe", "SM1 Dzpe", "VE1 Dzpe", "VE2 Dzpe", "VE3 Dzpe", "VR1 Dzpe", "VR3 Dzpe", "SpMax Dzare", "SpDiam Dzare", "SpMAD Dzare", "LogEE Dzare", "SM1 Dzare", "VE1 Dzare", "VE2 Dzare", "VE3 Dzare", "VR1 Dzare", "VR3 Dzare", "SpAbs Dzp", "SpMax Dzp", "SpAD Dzp", "SpMAD Dzp", "LogEE Dzp", "SM1 Dzp", "VE1 Dzp", "VE2 Dzp", "VE3 Dzp", "VR1 Dzp", "VR3 Dzp", "SpAbs Dzi", "SpDiam Dzi", "SpAD Dzi", "LogEE Dzi", "SM1 Dzi", "VE1 Dzi", "VE2 Dzi", "VE3 Dzi", "VR1 Dzi", "VR3 Dzi", "nBonds", "nBondsS", "nBondsA", "nBondsM",

"nBondsKS", "C2SP1", "C1SP3", "HybRatio", "FCSP3", "Xch-6dv", "Xch-7dv", "Xpc-6dv", "Xp-0d", "Xp-1d", "AXp-1d", "AXp-5d", "Xp-0dv", "Xp-1dv", "Xp-2dv", "AXp-0dv", "AXp-1dv", "AXp-2dv", "AXp-3dv", "AXp-4dv", "AXp-5dv", "SZ", "Sm", "Sv", "Sse", "Spe", "Sare", "Sp", "Si", "MZ", "Mm", "Mv", "Mse", "Mpe", "Mare", "Mp", "Mi", "SpAbs D", "SpMax D", "SpAD D", "LogEE D", "VE1 D", "VE2 D", "VE3 D", "VR1 D", "VR3 D", "ETA alpha", "AETA alpha", "ETA beta", "AETA beta", "AETA beta s", "ETA beta ns", "AETA beta ns", "AETA beta ns d", "ETA eta", "AETA eta", "ETA eta L", "ETA eta R", "ETA eta RL", "ETA eta F", "ETA eta FL", "ETA dAlpha B", "ETA epsilon 1", "ETA epsilon 2", "ETA epsilon 4", "ETA epsilon 5", "ETA dEpsilon A", "ETA dEpsilon C", "ETA dBeta", "ETA psi 1", "ETA dPsi A", "nHBAcc", "IC0", "TIC0", "SIC0", "SIC1", "BIC0", "BIC1", "CIC0", "CIC1", "MIC0", "MIC1", "MIC3", "MIC4", "MIC5", "ZMIC0", "ZMIC1", "ZMIC4", "ZMIC5", "Kier1", "LabuteASA", "PEOE VSA3", "PEOE VSA7", "SMR VSA1", "SMR VSA7", "SlogP VSA5", "MID", "MID h", "AMID h", "MID C", "AMID C", "MID N", "AMID N", "MID O", "AMID O", "MID X", "MPC6", "TMPC10", "piPC1", "piPC6", "TpiPC10", "apol", "bpol", "n4Ring", "n3HRing", "n3ARing", "n4ARing", "n3AHRing", "n7FRing", "nFHRing", "n7FARing", "n8FARing", "SLogP", "SMR", "TopoPSA(NO)", "TopoPSA", "Radius", "SRW03", "MW", "AMW", "WPath"]

- H: ["n6ARing", "n7AHRing", "n7FHRing", "n6FAHRing", "LogEE A", "VE2 A", "VE3 A", "VR1 A", "VR3 A", "nAromAtom", "nAromBond", "nAtom", "nHeavyAtom", "nSpiro", "nBridgehead", "nHetero", "nH", "nC", "nN", "nO", "nF", "nX", "ATS0dv", "ATS1dv", "ATS2dv", "ATS3d", "ATS4d", "ATS5d", "ATS0s", "ATS1s", "ATS2s", "ATS3s", "ATS4s", "ATS5s", "ATS6s", "ATS7s", "ATS8s", "ATS0Z", "ATS0m", "ATS1m", "ATS0v", "ATS1v", "ATS2v", "ATS3v", "ATS4v", "ATS0se", "ATS1se", "ATS2se", "ATS3se", "ATS4se", "ATS5se", "ATS0pe", "ATS1pe", "ATS2pe", "ATS3pe", "ATS4pe", "ATS5pe", "ATS0are", "ATS1are", "ATS2are", "ATS3are", "ATS4are", "ATS5are", "ATS0p", "ATS1p", "ATS2p", "ATS3p", "ATS4p", "ATS5p", "ATS0i", "ATS1i", "ATS2i", "ATS3i", "ATS4i", "ATS5i", "AATS0dv", "AATS1dv", "AATS2dv", "AATS3dv", "AATS4dv", "AATS1s", "AATS2s", "AATS0Z", "AATS1Z", "AATS2Z", "AATS0m", "AATS1m", "AATS2m", "AATS0v", "AATS1v", "AATS0se", "AATS1se", "AATS2se", "AATS4se", "AATS0pe", "AATS1pe", "AATS2pe", "AATS4pe", "AATS0are", "AATS1are", "AATS2are", "AATS4are", "AATS0p", "AATS2p", "AATS4p", "AATS0i", "AATS1i", "ATSC0s", "ATSC3s", "ATSC6s", "ATSC7s", "ATSC8s", "ATSC0Z", "ATSC0m", "ATSC0v", "ATSC0se", "ATSC1se", "ATSC0pe", "ATSC0are", "ATSC0p", "ATSC0i", "AATSC0c", "AATSC0dv", "AATSC0Z", "AATSC0m", "AATSC0v", "AATSC0se", "AATSC1se", "AATSC0pe", "AATSC1pe", "AATSC0are", "AATSC1are", "AATSC0p", "AATSC0i", "GATS1s", "GATS2s", "GATS1Z", "GATS1m", "GATS1se", "GATS1pe", "BCUTs-1h", "BCUTZ-1h", "BCUTm-1h", "BCUTv-1l", "BCUTse-1h", "BCUTpe-1h", "BCUTare-1h", "BCUTp-1l", "BCUTi-1h", "SpAbs DzZ", "SpMax DzZ", "SpAD DzZ", "LogEE DzZ", "SM1 DzZ", "VE1 DzZ", "VE2 DzZ", "VE3 DzZ", "VR1 DzZ", "VR3 DzZ", "SpAbs Dzm", "SpMax Dzm", "SpAD Dzm", "SpMAD Dzm", "LogEE Dzm", "SM1 Dzm", "VE1 Dzm", "VE2 Dzm", "VE3 Dzm", "VR1 Dzm", "VR3 Dzm", "SpAbs Dzv", "SpMax Dzv", "SpDiam Dzv", "SpAD Dzv", "LogEE Dzv", "SM1 Dzv", "VE1 Dzv", "VE2 Dzv", "VE3 Dzv", "VR1 Dzv", "VR3 Dzv", "SpAbs Dzse", "SpMax Dzse", "SpAD Dzse", "SpMAD Dzse", "LogEE Dzse", "SM1 Dzse", "VE1 Dzse", "VE2 Dzse", "VE3 Dzse", "VR1 Dzse", "VR3 Dzse", "SpAbs Dzpe", "SpMax Dzpe", "SpAD Dzpe", "SpMAD Dzpe", "LogEE Dzpe", "SM1 Dzpe", "VE1 Dzpe", "VE2 Dzpe", "VE3 Dzpe", "VR1 Dzpe", "VR3 Dzpe", "SpAbs Dzare", "SpMax Dzare", "SpAD Dzare", "SpMAD Dzare", "LogEE Dzare", "SM1 Dzare", "VE1 Dzare", "VE2 Dzare", "VE3 Dzare", "VR1 Dzare", "VR3 Dzare", "SpMax Dzp", "SpDiam Dzp", "LogEE Dzp", "SM1 Dzp", "VE1 Dzp", "VE2 Dzp", "VE3 Dzp", "VR1 Dzp", "VR3 Dzp", "SpAbs Dzi", "SpMax Dzi", "SpDiam Dzi", "SpAD Dzi", "LogEE Dzi", "SM1 Dzi", "VE1 Dzi", "VE2 Dzi", "VE3 Dzi", "VR1 Dzi", "VR3 Dzi", "nBonds", "nBondsS", "nBondsA", "nBondsM", "nBondsKS", "C2SP2", "HybRatio", "FCSP3", "Xch-7d", "Xch-4dv", "Xpc-4dv", "Xpc-5dv", "Xpc-6dv", "Xp-0d", "Xp-1d", "Xp-0dv", "Xp-1dv", "Xp-2dv", "Xp-5dv", "AXp-0dv", "AXp-1dv", "AXp-2dv", "AXp-3dv", "AXp-4dv", "AXp-5dv", "SZ", "Sm", "Sv", "Sse", "Spe", "Sare", "Sp", "Si", "MZ", "Mm", "Mv", "Mse", "Mpe", "Mare", "Mp", "Mi", "SpAbs D", "SpMax D", "SpAD D", "LogEE D", "VE1 D", "VE2 D", "VE3 D", "VR1 D", "VR3 D", "ETA alpha", "AETA alpha", "ETA beta", "AETA beta", "AETA beta s", "ETA beta ns", "AETA beta ns", "AETA beta ns d", "ETA eta", "AETA eta", "ETA eta L", "ETA eta R", "ETA eta RL", "ETA eta F", "AETA eta F", "ETA eta FL", "ETA dAlpha B", "ETA epsilon 1", "ETA epsilon 2", "ETA epsilon 4", "ETA epsilon 5", "ETA dEpsilon A", "ETA dEpsilon B", "ETA dEpsilon C", "ETA dBeta", "ETA psi 1", "ETA dPsi A", "nHBAcc", "IC0", "TIC0", "SIC0", "SIC1", "BIC0", "BIC1", "CIC0", "CIC1", "MIC0", "MIC1", "MIC2", "MIC3", "MIC4", "MIC5", "ZMIC0", "ZMIC1", "ZMIC3", "ZMIC4", "ZMIC5", "Kier1", "LabuteASA", "PEOE VSA3", "PEOE VSA8", "SMR VSA1", "SlogP VSA5", "MDEC-23", "MID", "MID h", "AMID h", "MID C", "AMID C", "MID N", "AMID N", "MID O", "AMID O", "MID X", "MPC5", "MPC6", "TMPC10", "piPC1", "piPC6", "TpiPC10", "apol", "bpol", "nHRing", "naRing", "naHRing", "n4ARing", "n7FRing", "n6FARing", "n7FARing", "SLogP", "SMR", "TopoPSA(NO)", "TopoPSA", "JGI4", "MW", "AMW", "WPath"]
- G: ["ATS8v", "ATS8se", "ATS8pe", "ATS8are", "ATS8p", "n7FARing", "n6FAHRing", "nBase", "LogEE A", "VE2 A", "VE3 A", "VR1 A", "VR3 A", "nAtom", "nHeavyAtom", "nSpiro", "nBridgehead", "nHetero", "nH", "nC", "nN", "nO", "nF", "nX", "ATS0dv", "ATS1dv", "ATS2dv", "ATS3d",

"ATS4d", "ATS5d", "ATS0s", "ATS1s", "ATS2s", "ATS3s", "ATS4s", "ATS5s", "ATS8s", "ATS0Z", "ATS0m", "ATS0v", "ATS1v", "ATS2v", "ATS3v", "ATS4v", "ATS0se", "ATS1se", "ATS2se", "ATS3se", "ATS4se", "ATS5se", "ATS0pe", "ATS1pe", "ATS2pe", "ATS3pe", "ATS4pe", "ATS5pe", "ATS0are", "ATS1are", "ATS2are", "ATS3are", "ATS4are", "ATS5are", "ATS0p", "ATS1p", "ATS2p", "ATS3p", "ATS4p", "ATS5p", "ATS0i", "ATS1i", "ATS2i", "ATS3i", "ATS4i", "ATS5i", "AATS0dv", "AATS1dv", "AATS2dv", "AATS3dv", "AATS1s", "AATS2s", "AATS0Z", "AATS1Z", "AATS2Z", "AATS4Z", "AATS0m", "AATS1m", "AATS2m", "AATS4m", "AATS0v", "AATS1v", "AATS0se", "AATS1se", "AATS2se", "AATS3se", "AATS4se", "AATS0pe", "AATS1pe", "AATS2pe", "AATS3pe", "AATS4pe", "AATS0are", "AATS1are", "AATS2are", "AATS3are", "AATS4are", "AATS0p", "AATS1p", "AATS2p", "AATS0i", "AATS1i", "ATSC0s", "ATSC5s", "ATSC6s", "ATSC8s", "ATSC0Z", "ATSC0m", "ATSC0v", "ATSC0se", "ATSC1se", "ATSC0pe", "ATSC0are", "ATSC0p", "ATSC0i", "AATSC0c", "AATSC1c", "AATSC0dv", "AATSC5d", "AATSC4s", "AATSC0Z", "AATSC0m", "AATSC0v", "AATSC0se", "AATSC1se", "AATSC0pe", "AATSC0are", "AATSC1are", "AATSC0p", "AATSC0i", "GATS1d", "GATS1s", "GATS2s", "GATS1Z", "GATS1m", "GATS1pe", "GATS1are", "BCUTdv-1h", "BCUTs-1h", "BCUTZ-1h", "BCUTm-1h", "BCUTv-1l", "BCUTse-1h", "BCUTpe-1h", "BCUTpe-1l", "BCUTare-1h", "BCUTi-1h", "SpAbs DzZ", "SpMax DzZ", "SpDiam DzZ", "SpAD DzZ", "SpMAD DzZ", "LogEE DzZ", "SM1 DzZ", "VE1 DzZ", "VE2 DzZ", "VE3 DzZ", "VR1 DzZ", "VR2 DzZ", "VR3 DzZ", "SpAbs Dzm", "SpMax Dzm", "SpDiam Dzm", "SpAD Dzm", "LogEE Dzm", "SM1 Dzm", "VE1 Dzm", "VE2 Dzm", "VE3 Dzm", "VR1 Dzm", "VR2 Dzm", "VR3 Dzm", "SpAbs Dzv", "SpMax Dzv", "SpDiam Dzv", "SpAD Dzv", "LogEE Dzv", "SM1 Dzv", "VE1 Dzv", "VE2 Dzv", "VE3 Dzv", "VR1 Dzv", "VR3 Dzv", "SpAbs Dzse", "SpMax Dzse", "SpDiam Dzse", "SpAD Dzse", "LogEE Dzse", "SM1 Dzse", "VE1 Dzse", "VE2 Dzse", "VE3 Dzse", "VR2 Dzse", "VR3 Dzse", "SpMax Dzpe", "SpDiam Dzpe", "SpAD Dzpe", "LogEE Dzpe", "SM1 Dzpe", "VE1 Dzpe", "VE2 Dzpe", "VE3 Dzpe", "VR1 Dzpe", "VR2 Dzpe", "VR3 Dzpe", "SpAbs Dzare", "SpMax Dzare", "SpDiam Dzare", "SpAD Dzare", "SpMAD Dzare", "LogEE Dzare", "SM1 Dzare", "VE1 Dzare", "VE2 Dzare", "VE3 Dzare", "VR2 Dzare", "VR3 Dzare", "SpDiam Dzp", "SM1 Dzp", "VE1 Dzp", "VE2 Dzp", "VE3 Dzp", "VR1 Dzp", "VR2 Dzp", "VR3 Dzp", "SpAbs Dzi", "SpMax Dzi", "SpAD Dzi", "LogEE Dzi", "SM1 Dzi", "VE1 Dzi", "VE2 Dzi", "VE3 Dzi", "VR1 Dzi", "VR3 Dzi", "nBonds", "nBondsS", "nBondsM", "nBondsKS", "C2SP2", "HybRatio", "FCSP3", "Xch-6d", "Xch-7d", "Xch-4dv", "Xch-6dv", "Xpc-6d", "Xp-0d", "Xp-1d", "AXp-5d", "AXp-6d", "AXp-7d", "Xp-0dv", "Xp-1dv", "Xp-2dv", "AXp-0dv", "AXp-1dv", "AXp-2dv", "AXp-3dv", "AXp-4dv", "AXp-5dv", "AXp-6dv", "SZ", "Sm", "Sv", "Sse", "Spe", "Sare", "Sp", "Si", "MZ", "Mm", "Mv", "Mse", "Mpe", "Mare", "Mp", "Mi", "SpAbs D", "SpMax D", "SpAD D", "LogEE D", "VE1 D", "VE2 D", "VE3 D", "VR1 D", "VR3 D", "ETA alpha", "AETA alpha", "ETA shape x", "ETA beta", "AETA beta", "AETA beta s", "ETA beta ns", "AETA beta ns", "AETA beta ns d", "ETA eta", "AETA eta", "ETA eta L", "ETA eta R", "ETA eta RL", "ETA eta F", "AETA eta F", "ETA eta FL", "ETA eta BR", "ETA dAlpha B", "ETA epsilon 1", "ETA epsilon 2", "ETA epsilon 4", "ETA epsilon 5", "ETA dEpsilon A", "ETA dEpsilon B", "ETA dEpsilon C", "ETA dBeta", "ETA psi 1", "ETA dPsi A", "nHBAcc", "IC0", "TIC0", "SIC0", "SIC1", "BIC0", "BIC1", "BIC4", "BIC5", "CIC0", "CIC1", "MIC0", "MIC1", "MIC2", "MIC3", "MIC4", "MIC5", "ZMIC0", "Kier1", "LabuteASA", "PEOE VSA3", "PEOE VSA10", "SMR VSA3", "SMR VSA7", "SlogP VSA5", "MID", "MID h", "AMID h", "MID C", "AMID C", "MID N", "AMID N", "MID O", "AMID O", "MID X", "MPC5", "TMPC10", "piPC1", "TpiPC10", "apol", "bpol", "nRing", "nHRing", "n3HRing", "naRing", "naHRing", "n7AHRing", "nFRing", "n5FRing", "n8FRing", "nFARing", "n5FARing", "SLogP", "SMR", "TopoPSA(NO)", "TopoPSA", "TopoShapeIndex", "PetitjeanIndex", "SRW06", "SRW09", "SRW10", "MW", "AMW", "WPath", "mZagreb1"]

- U_0: ["ATS7dv", "ATSC8v", "n8FARing", "n6FAHRing", "nBase", "LogEE A", "VE2 A", "VE3 A", "VR1 A", "VR3 A", "nAromAtom", "nAromBond", "nAtom", "nHeavyAtom", "nSpiro", "nBridgehead", "nHetero", "nH", "nC", "nN", "nO", "nF", "nX", "ATS0dv", "ATS1dv", "ATS2dv", "ATS3d", "ATS4d", "ATS5d", "ATS0s", "ATS1s", "ATS2s", "ATS3s", "ATS4s", "ATS5s", "ATS6s", "ATS7s", "ATS8s", "ATS0Z", "ATS0m", "ATS0v", "ATS1v", "ATS2v", "ATS4v", "ATS0se", "ATS1se", "ATS2se", "ATS3se", "ATS4se", "ATS5se", "ATS0pe", "ATS1pe", "ATS2pe", "ATS3pe", "ATS4pe", "ATS5pe", "ATS0are", "ATS1are", "ATS2are", "ATS3are", "ATS4are", "ATS5are", "ATS0p", "ATS1p", "ATS2p", "ATS3p", "ATS4p", "ATS5p", "ATS0i", "ATS1i", "ATS2i", "ATS3i", "ATS4i", "ATS5i", "AATS0dv", "AATS1dv", "AATS2dv", "AATS3dv", "AATS1s", "AATS2s", "AATS0Z", "AATS1Z", "AATS2Z", "AATS0m", "AATS1m", "AATS2m", "AATS0v", "AATS1v", "AATS0se", "AATS1se", "AATS2se", "AATS4se", "AATS0pe", "AATS1pe", "AATS2pe", "AATS4pe", "AATS0are", "AATS1are", "AATS2are", "AATS4are", "AATS0p", "AATS1p", "AATS2p", "AATS0i", "AATS1i", "ATSC0dv", "ATSC0s", "ATSC5s", "ATSC7s", "ATSC8s", "ATSC0Z", "ATSC0m", "ATSC0v", "ATSC0se", "ATSC1se", "ATSC0pe", "ATSC0are", "ATSC0p", "ATSC0i", "AATSC0c", "AATSC1c", "AATSC0dv", "AATSC0Z", "AATSC0m", "AATSC0v", "AATSC0se", "AATSC0pe", "AATSC0are", "AATSC0p", "AATSC0i", "GATS1s", "GATS2s", "GATS1Z", "GATS1m", "BCUTc-1l", "BCUTs-1h", "BCUTZ-1h", "BCUTm-1h", "BCUTv-1l", "BCUTse-1h", "BCUTpe-1h", "BCUTare-1h", "BCUTi-1h", "SpAbs DzZ", "SpMax DzZ", "SpAD DzZ", "SpMAD DzZ", "Lo-

gEE DzZ", "SM1 DzZ", "VE1 DzZ", "VE2 DzZ", "VE3 DzZ", "VR1 DzZ", "VR3 DzZ", "SpAbs Dzm", "SpMax Dzm", "SpAD Dzm", "SpMAD Dzm", "LogEE Dzm", "SM1 Dzm", "VE1 Dzm", "VE2 Dzm", "VE3 Dzm", "VR3 Dzm", "SpAbs Dzv", "SpMax Dzv", "SpDiam Dzv", "SpAD Dzv", "SpMAD Dzv", "LogEE Dzv", "SM1 Dzv", "VE1 Dzv", "VE2 Dzv", "VE3 Dzv", "VR1 Dzv", "VR2 Dzv", "VR3 Dzv", "SpAbs Dzse", "SpMax Dzse", "SpAD Dzse", "SpMAD Dzse", "LogEE Dzse", "SM1 Dzse", "VE1 Dzse", "VE2 Dzse", "VE3 Dzse", "VR1 Dzse", "VR3 Dzse", "SpAbs Dzpe", "SpMax Dzpe", "SpDiam Dzpe", "SpAD Dzpe", "LogEE Dzpe", "SM1 Dzpe", "VE1 Dzpe", "VE2 Dzpe", "VE3 Dzpe", "VR3 Dzpe", "SpMax Dzare", "SpAD Dzare", "SpMAD Dzare", "LogEE Dzare", "SM1 Dzare", "VE1 Dzare", "VE2 Dzare", "VE3 Dzare", "VR3 Dzare", "SpAbs Dzp", "SpMax Dzp", "SpDiam Dzp", "SM1 Dzp", "VE1 Dzp", "VE2 Dzp", "VE3 Dzp", "VR1 Dzp", "VR3 Dzp", "SpAbs Dzi", "SpMax Dzi", "SpDiam Dzi", "SpAD Dzi", "LogEE Dzi", "SM1 Dzi", "VE1 Dzi", "VE2 Dzi", "VE3 Dzi", "VR1 Dzi", "VR2 Dzi", "VR3 Dzi", "nBonds", "nBondsS", "nBondsA", "nBondsM", "nBondsKS", "HybRatio", "FCSP3", "Xch-6d", "Xp-0d", "Xp-1d", "AXp-7d", "Xp-0dv", "Xp-1dv", "Xp-2dv", "AXp-0dv", "AXp-1dv", "AXp-2dv", "AXp-3dv", "AXp-4dv", "AXp-5dv", "AXp-6dv", "SZ", "Sm", "Sv", "Sse", "Spe", "Sare", "Sp", "Si", "MZ", "Mm", "Mv", "Mse", "Mpe", "Mare", "Mp", "Mi", "VE1 D", "VE2 D", "VE3 D", "VR1 D", "VR3 D", "ETA alpha", "AETA alpha", "ETA beta", "AETA beta", "AETA beta s", "ETA beta ns", "AETA beta ns", "ETA beta ns d", "AETA beta ns d", "ETA eta", "AETA eta", "ETA eta L", "ETA eta RL", "ETA eta F", "AETA eta F", "ETA eta FL", "ETA eta BR", "ETA dAlpha B", "ETA epsilon 1", "ETA epsilon 2", "ETA epsilon 4", "ETA epsilon 5", "ETA dEpsilon A", "ETA dEpsilon B", "ETA dEpsilon C", "ETA dBeta", "ETA psi 1", "ETA dPsi A", "nHBAcc", "IC0", "TIC0", "SIC0", "SIC1", "BIC0", "BIC1", "BIC2", "BIC4", "BIC5", "CIC0", "CIC1", "MIC0", "MIC1", "MIC2", "MIC3", "MIC4", "MIC5", "ZMIC0", "Kier1", "Kier3", "LabuteASA", "PEOE VSA3", "PEOE VSA12", "SMR VSA1", "SMR VSA3", "SlogP VSA5", "MDEC-23", "MID", "MID h", "AMID h", "MID C", "AMID C", "MID N", "AMID N", "MID O", "AMID O", "MPC3", "piPC1", "piPC6", "TpiPC10", "apol", "bpol", "naRing", "naHRing", "n7FRing", "nFARing", "n6FARing", "n7FARing", "SLogP", "SMR", "TopoPSA(NO)", "TopoPSA", "SRW06", "SRW08", "MW", "AMW", "WPath", "WPol"]

- C_v: ["n6FHRing", "n7FHRing", "n8FaRing", "n8FaHRing", "n5FAHRing", "JGI7", "nAcid", "nBase", "SpAbs A", "SpMax A", "SpDiam A", "SpAD A", "SpMAD A", "LogEE A", "VE1 A", "VE2 A", "VE3 A", "VR1 A", "VR2 A", "VR3 A", "nAromAtom", "nAromBond", "nAtom", "nSpiro", "nBridgehead", "nH", "ATS2dv", "ATS3dv", "ATS5dv", "ATS0d", "ATS1d", "ATS2d", "ATS4d", "ATS7d", "ATS0s", "ATS1s", "ATS2s", "ATS3s", "ATS4s", "ATS5s", "ATS2Z", "ATS3Z", "ATS4Z", "ATS5Z", "ATS7Z", "ATS0m", "ATS2m", "ATS3m", "ATS5m", "ATS6m", "ATS0v", "ATS4v", "ATS5v", "ATS6v", "ATS7v", "ATS0se", "ATS1se", "ATS2se", "ATS7se", "ATS0pe", "ATS1pe", "ATS2pe", "ATS5pe", "ATS0are", "ATS1are", "ATS4are", "ATS5are", "ATS7are", "ATS0p", "ATS5p", "ATS7p", "ATS0i", "ATS1i", "ATS3i", "ATS6i", "ATS7i", "AATS1dv", "AATS0d", "AATS1d", "AATS4d", "AATS5d", "AATS1Z", "AATS6Z", "AATS1m", "AATS0v", "AATS3v", "AATS4v", "AATS6v", "AATS0p", "AATS4p", "AATS5p", "AATS0i", "AATS2i", "AATS3i", "AATS4i", "AATS5i", "ATSC0c", "ATSC3c", "ATSC4c", "ATSC6c", "ATSC0dv", "ATSC1dv", "ATSC0d", "ATSC3d", "ATSC4d", "ATSC8d", "ATSC0s", "ATSC1s", "ATSC2s", "ATSC2Z", "ATSC0m", "ATSC1m", "ATSC2m", "ATSC0v", "ATSC6v", "ATSC1se", "ATSC0pe", "ATSC0p", "ATSC1p", "ATSC0i", "ATSC1i", "AATSC0c", "AATSC3c", "AATSC4c", "AATSC6c", "AATSC0d", "AATSC4d", "AATSC6d", "AATSC0s", "AATSC2s", "AATSC3s", "AATSC1Z", "AATSC2v", "AATSC1p", "AATSC1i", "AATSC4i", "MATS1c", "MATS2c", "MATS2dv", "MATS1d", "MATS5d", "MATS6d", "MATS1s", "MATS2s", "MATS1m", "MATS2m", "MATS6m", "MATS1v", "MATS1pe", "MATS4are", "MATS1p", "MATS3p", "MATS4p", "MATS5p", "MATS1i", "GATS1dv", "GATS4dv", "GATS2d", "GATS4d", "GATS6d", "GATS1s", "GATS2s", "GATS3s", "GATS1Z", "GATS4Z", "GATS1m", "GATS2m", "GATS1v", "GATS2v", "GATS1se", "GATS2se", "GATS3se", "GATS4se", "GATS1pe", "GATS3pe", "GATS4pe", "GATS1p", "GATS1i", "BCUTc-1h", "BCUTc-1l", "BCUTdv-1h", "BCUTdv-1l", "BCUTd-1h", "BCUTd-1l", "BCUTs-1h", "BCUTm-1h", "BCUTm-1l", "BCUTv-1h", "BCUTv-1l", "BCUTse-1l", "BCUTpe-1h", "BCUTare-1h", "BCUTp-1h", "BCUTp-1l", "BCUTi-1h", "BCUTi-1l", "SpMax DzZ", "SpDiam DzZ", "LogEE DzZ", "VE1 DzZ", "VE2 DzZ", "VE3 DzZ", "SpAbs Dzm", "SpMax Dzm", "SpDiam Dzm", "LogEE Dzm", "SpAD Dzv", "SpMAD Dzv", "VE1 Dzv", "VE2 Dzv", "VE3 Dzv", "SpMax Dzse", "LogEE Dzse", "VE1 Dzse", "SpMax Dzpe", "LogEE Dzpe", "SM1 Dzpe", "SpAbs Dzare", "SpMax Dzare", "SpDiam Dzare", "SpAD Dzare", "LogEE Dzare", "VR1 Dzare", "VR2 Dzare", "VR3 Dzare", "VE1 Dzp", "VE2 Dzp", "VE3 Dzp", "VR1 Dzp", "VR2 Dzp", "SpAbs Dzi", "SpMax Dzi", "SpMAD Dzi", "LogEE Dzi", "VE2 Dzi", "VR2 Dzi", "VR3 Dzi", "nBondsS", "nBondsD", "nBondsA", "nBondsM", "RNCG", "RPCG", "C1SP2", "C2SP2", "C3SP2", "HybRatio", "Xch-3d", "Xch-4d", "Xch-5d", "Xch-7d", "Xch-3dv", "Xch-4dv", "Xch-5dv", "Xch-6dv", "Xch-7dv", "Xc-4d", "Xc-6d", "Xc-4dv", "Xc-5dv", "Xpc-4d", "Xpc-5d", "Xpc-6d", "Xpc-4dv", "Xpc-5dv", "Xpc-6dv", "Xp-0d", "Xp-1d", "Xp-3d", "Xp-4d", "Xp-5d", "Xp-6d", "Xp-7d", "AXp-0d", "AXp-1d", "AXp-2d", "AXp-5d", "Xp-0dv", "Xp-1dv", "Xp-2dv", "Xp-7dv", "AXp-2dv", "AXp-6dv", "Sv", "Sse", "Spe", "Sare", "Sp", "Si", "Mv", "Mp", "Mi", "SpAD D", "SpMAD D", "LogEE D", "VE1 D", "VE2 D", "VE3 D", "ETA shape p", "ETA

shape y", "ETA shape x", "ETA beta", "ETA beta s", "AETA beta s", "AETA beta ns", "AETA beta ns d", "ETA eta", "AETA eta", "ETA eta R", "AETA eta R", "ETA eta RL", "AETA eta RL", "ETA eta B", "AETA eta B", "ETA eta BR", "AETA eta BR", "ETA epsilon 5", "ETA dEpsilon B", "ETA dEpsilon D", "AETA dBeta", "nHBAcc", "nHBDon", "TIC0", "TIC1", "TIC5", "BIC4", "CIC1", "MIC2", "ZMIC0", "Kier1", "Kier2", "LabuteASA", "PEOE VSA1", "PEOE VSA2", "PEOE VSA3", "PEOE VSA4", "PEOE VSA5", "PEOE VSA6", "PEOE VSA9", "PEOE VSA11", "PEOE VSA12", "PEOE VSA13", "SMR VSA1", "SMR VSA2", "SMR VSA3", "SMR VSA4", "SMR VSA5", "SMR VSA6", "SMR VSA7", "SMR VSA9", "SlogP VSA1", "SlogP VSA2", "SlogP VSA3", "SlogP VSA4", "SlogP VSA5", "SlogP VSA6", "SlogP VSA8", "SlogP VSA10", "SlogP VSA11", "MDEC-12", "MDEC-13", "MDEC-22", "MDEC-23", "MDEC-33", "MID", "AMID", "AMID h", "MID C", "AMID C", "AMID N", "AMID O", "MPC2", "MPC3", "MPC4", "MPC5", "MPC8", "TMPC10", "piPC1", "piPC2", "piPC3", "piPC4", "piPC6", "piPC7", "piPC8", "TpiPC10", "apol", "bpol", "n3Ring", "n4Ring", "n5Ring", "n4HRing", "n5aRing", "n6aRing", "n5aHRing", "nARing", "n3ARing", "n4ARing", "n5ARing", "n6ARing", "n7ARing", "nAHRing", "n3AHRing", "n4AHRing", "n5AHRing", "n6FRing", "n7FRing", "n8FRing", "n6FARing", "n7FARing", "n7FAHRing", "nRot", "RotRatio", "SLogP", "SMR", "GGI1", "GGI2", "GGI3", "GGI4", "JGI1", "JGI2", "JGI4", "JGT10", "Diameter", "Radius", "TopoShapeIndex", "PetitjeanIndex", "MWC02", "MWC03", "MWC05", "MWC08", "MWC09", "TMWC10", "SRW04", "SRW05", "SRW06", "SRW07", "SRW08", "SRW09", "SRW10", "TSRW10", "WPol", "mZagreb1", "mZagreb2"]

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: Code, weights, and data will be provided.

   Guidelines:

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [Yes]

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

13. **New Assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

14. **Crowdsourcing and Research with Human Subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]