# AI Enabled Drug Design and Side Effect Prediction Powered by Multi-Objective Evolutionary Algorithms & Transformer Models

***Karl Grantham***

Submitted in partial fulfillment
of the requirements for the degree of

Master of Science

Department of Computer Science
Faculty of Mathematics and Science
Brock University
St. Catharines, Ontario

# Abstract

Due to the large search space and conflicting objectives, drug design and discovery is a difficult problem for which new machine learning (ML) approaches are required. Here, the problem is to invent a method by which new, therapeutically useful, compounds can be discovered; and to simultaneously avoid compounds which will fail clinical trials or pass unwanted effects onto the end patient. By extending current technologies as well as adding new ones, more design criteria can be included, and more promising novel drugs can be discovered. This work advances the field of computational drug design by (1) developing MOEA-DT, a non-deep learning application for multi-objective molecular optimization, which generates new molecules with high performance in a variety of design criteria; and (2) developing SEMTL-BERT, a side effect prediction algorithm which leverages the latest ML techniques and datasets to accomplish its task. Experiments performed show that MOEA-DT either matches or outperforms other similar methods, and that SEMTL-BERT can enhance predictive ability.

# Acknowledgements

This has been a year of ambition. I have recently attained a large portion of everything I have ever asked for, and its nothing like what I expected. Given this, I have a long list of people whom it is my good fortune to have in my life, and who played their own roles in my success. I would like to thank my family for their continued support financially, through the move-in process, and through assisting in house repair. My studies and therefore my future in research would be possible without their continued and unconditional support.

I would also like to thank Professor Yifeng Li for the opportunities he has given me, his inspiration and constant help with my research. His work has been essential in my post-graduate education.

I would also like to thank the members of the supervisory committee: Dr. Sheridan Houghten and Dr. Beatrice Ombuki-Berman, and the external examiner Dr. Divya Matta for taking the time out to discuss my thesis.

Finally, I would like to thank the members of the Brock biomedical data science (BMDS) lab. We rely on each other to share thoughts, ideas, and knowledge. Through our collaboration, we achieve much more than we could alone.

This research is supported by the Vector Scholarship in Artificial Intelligence, provided through the Vector Institute.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

## 1.1 Problem

Drug discovery is the process of identifying new, therapeutically useful compounds through the use of chemical and biological knowledge and/or computational drug design approaches. This is a difficult problem because the combinatorial explosion created by all valid chemical compounds creates a massive search space estimated to be in the order of $10^{60}$ [58]. In addition, approaches to drug design have many pharmacological properties over which to perform optimization. This changes the task of drug design from simple one-dimensional, single-property optimization to the selection of optimal "trade-offs" between all of the desired chemical properties. Further complicating the issue is the fact that drugs can cause unwanted adverse effects, which cause late-stage failures in drug development [32], and contribute to poor quality of life. With such a large search space, many properties to consider, and pitfalls to avoid, machine learning (ML) methods of drug discovery are necessary to reduce drug development costs, and improve product quality.

## 1.2 Impact

The prevention of adverse effects by the discovery of adverse effect-free, or minimal adverse effect drugs has the potential to improve quality of life for a large number of people. According to Lazarou et al.; [37] a meta-analysis of incidences of adverse drug reactions (ADRs) in the United States, 15% of hospital patients are harmed and 0.32% are killed by ADRs, making fatal ADRs between the fourth and sixth leading cause of death, whether they were admitted to the hospital as a result of the ADR,

or suffered the ADR while in the hospital. Another meta-analysis by Einarson et al. [17] of English-language studies covering only admissions to hospitals resulting from ADRs internationally, found that such reported admissions occurred at a proportion between 0.2% and 21.7% of patients, estimating 5.1% of patients. Either way, the minimization of ADRs will positively impact many people.

ADRs are also responsible for many late-stage failures in the drug development, which increases the time, cost, and financial risk. As a result, it may take 10 to 16 years [53] and US$800 million to US$ 1.8 billion [44] to discover and develop a new drug. Automated drug design with side effect prediction has the potential to significantly decrease the time and cost from end to end.

Finally, the adverse effects of drugs may only become apparent once the drug has passed clinical trials, leading to the potential withdrawal of the drug after several years on the market. Lasser et al. [35] found that of the 548 new chemical entities were approved between 1975-1999 by the FDA, 56(10.2%) acquired a new black box warning or were withdrawn, 45(8.2%) acquired 1 or more black box warnings and 16(2.9%) were withdrawn from the market. Using Kaplan-Meier analyses, they also found that the estimated probability of a new drug acquiring a new black box warning or being withdrawn from the market over 25 years was 20%. Therefore, the early stage prediction of side-effects might additionally warn drug producers of potential side-effects before they can be released to market, creating a safer environment for the producer and the end user.

## 1.3 Current Methods

In their article in *Nature*, Vamathevan et al. [81] cover current methods of ML in drug design. Here, a variety of methods have been used for a variety for application throughout the drug development pipeline. The earliest methods along the this pipeline aid in the identification and validation of drug targets. The idea here is to identify of a target protein for which its modulation would bring about a modulation in the disease. Thus, therapeutic benefit can be achieved through a chemical compound which fits the target. With modern biology becoming increasingly rich in data, data science and ML are becoming prominent aspects for this application. In such methods, a first step is the establishment of a causal relation between target and disease. ML is used here to process large datasets and make predictions on potential causality. Successful ML algorithms include a decision tree-based metaclassifier [8] for the prediction of druggable genomes associated with morbidity, and a deep neural

network (DNN) for the prediction of splicing patterns in individual tissues and differences of splicing patterns across tissues [38].

After the targets have been identified, small molecules which fit the target are then designed and optimized. This involves the discovery of new drug candidates which can block or activate the target protein, and the prediction of which drugs have the best therapeutic effect. Here, there is a variety of different ML approaches, but the most common approach is to use deep learning (DL). Multi-task DNNs, such the one from *DeepChem* [61] have been developed, and shown to be able to identify compounds of similar chemical structures to lead compounds [60]. Other methods such as the one by Olivecrona et al. [54], apply DL by using a reinforcement learning (RL) agent to tune a recurrent neural network (RNN) towards the generation of molecular structures with desired chemical properties.

Following the discovery of new chemical compounds, preclinical development begins in an attempt to improve clinical success rate. During this part of the pipeline, ML methods are used to discover predictive biomarkers, in order to better understand the effects of drugs and therefore find the right drug for the right patients. However, this has shown to be a difficult task for ML predictive models due to poor data quality, inability to select models, and reproducibility. However, some success has been found by using alternate input data, as well as indicators other than oncology. Tasaki et al. [72] used multi-omics data to achieve a better understanding of drug responses of patients with rheumatoid arthritis. Additionally, Rashid et al. [62] were able to use a variational autoencoder (VAE) to more efficiently differentiate between hidden tumor sub-populations in the latent feature space.

The final intersection with ML along the drug development pipeline is computational pathology. The task here is to describe how drugs interact with and affect tissues. Early attempts involved handcrafted algorithms for obtaining descriptive features in collaboration with pathologists. The advent of convolutional neural networks to pathology images was a significant step forward. Such models are beneficial because they can automatically learn features. Sharma et al. [64] successfully used CNNs to classify cancer based on immunohistochemical response.

## 1.4 Major Contribution

### 1.4.1 Molecular Optimization

A non-deep-learning algorithm for molecular optimization which leverages multi-objective optimization, as well as fragment and structure based drug design is presented, which finds new, high performance drug candidates. Its performance is experimentally shown to be similar or better than other significantly more complex DL models which use the same data for the same task.

### 1.4.2 Side Effect Prediction

A novel approach to side effect prediction in drug compounds is also presented, being is the one of the first to leverage a transformer architecture for the task. Its performance is then explored under a variety of different experimental conditions, and compared to other models for the same task. Finally, model performance is then compared to the baseline to show predictive ability.

# Chapter 2

# Related Work

In this chapter, prior work in the field of ML which directly relate to the methods presented in this paper is introduced. Sections are presented in such a way that algorithms, methods, and concepts are explored following those on which they are dependant. Thus, the reader will be presented with the background information immediately prior to the coverage of a topic. Coverage begins with work related to MOEA-DT, and reaches those associated with SEMTL-BERT in later sections.

## 2.1 Computational Drug Design

Early approaches to computational drug design involved high throughput screening (HTS), which performed quick searches of drug databases based on pre-established design criteria [71]. This soon became insufficient for more complex searches where the desired properties are more robust and less well-defined. Thus, machine learning became necessary to effectively search the chemical space. One popular approach for drug design is to involve deep learning (DL) [30]. Such methods are complex since they require the training of neural networks. Evolutionary algorithm (EA) is another popular method for drug discovery either separately or in combination with DL, as in *DEL: Deep Evolutionary Learning for for molecular design* [25].

## 2.2 Molecular Representation

Since computational methods can have a significant impact on the drug discovery process, a variety of different methods have been developed. Here, molecules are usually represented by molecular graphs (either 2D or 3D), or by **S**implified **M**olecular-

**I**nput **L**ine-**E**ntry **S**ystem (SMILES) strings [77]. A typical molecular graph represents atoms as vertices, and bonds as edges. SMILES strings however, encode an entire molecule as a string of characters. SMILES strings are particularly useful since it allows for excellent compression of information, as well as compatibility with natural language processing approaches. Issues stemming from the utilization of SMILES strings include many invalid strings within the search space, and many possible SMILES encodings for each chemical compound.

## 2.3 SMILES Syntax

In the often-cited papers, Weininger [77] [78] introduces the the SMILES language, and identifies the characters used in SMILES representation. Atom types are represented by their respective one-to-two letter abbreviation in the periodic table (chlorine as "Cl", for example). Hydrogens are usually left implicit, unless they explicitly form a part of a structure, as in chiral structures involving a hydrogen. In such cases, they too are listed with their symbol, "H". In the case of atoms possessing a charge, they are represented by the atom type, followed by the deficiency or sirplus of electrons, all within square brackets, as in "[Fe+2]". Atoms of an unknown type can be represented by an asterisk "*". Single bonds can be represented implicitly, or explicitly by the dash character "-". Double, and triple bonds are represented by the characters "=", and "#", respectively. Non-bonds between atoms are represented by a dot ".". Backwards and forwards directionality in double bonds are represented by backwards and forward slashes "\, /".

Atoms in a ring are listed between a pair of equal single numbers, which immediately follow an arbitrarily chosen starting atom. Multiple rings within one molecule are represented by the same format with different numbers. This creates a limitation of SMILES that 10 or more rings cannot be represented, since they would require more than one digit to identify the ring. Branching is represented by one branch contained within parentheses, followed by the other branch. This format can be leveraged for nesting branches. Aromaticity can be represented implicitly by the bonds, or explicitly by representing the atom in lower case letters, where they would otherwise be represented by the first (or only) letter capitalized.

The "@" character allows for partial chirality representation in SMILES. such structural features are represented locally, by the atoms immediate to these characters, rather than globally. Specification of chirality is done by the "@", signifying counter-

clockwise orientation and "@@", signifying clockwise orientation. Unless defaulted by the number of connected atoms, they are followed by the specific kind of chirality. Allene-likes are identified by "@AL1" or "@AL2", square-planar chiralities by "@SP1" to "@SP3", trigonal-bipyramidals by "@TB1" to "@TB20", and octahedrals by "@OH1" to "@OH30". In general, these structures are represented by the first listed non-chiral atom being considered the "from" atom, followed by the chiral center listed with the "@" or "@@" characters, followed by the rest of the atoms listed in order. In the case that the chiral center atom is listed first, the next listed atom is taken to be the "from" atom. An example a chiral structure represented in SMILES would be "OC(Cl)=[C@AL1]=C(C)F", which forms an allene-like structure around a carbon atom.

## 2.4 Fragment based Drug Design

Regardless of the representation used, a vast search space of chemical compounds is created. To deal with this, both computational and non-computational methods of drug design often include fragment-based drug design (FBDD) [20]. In this approach, the search space is redefined as all possible combinations of fragments, as opposed to combinations of individual atoms and bonds. Drug fragments are very small molecules which are a fraction of the size of a typical drug. Fragments are collected from databases of known drugs, or drug-like compounds. With a library of fragments, an ML model can be trained to combine the fragments in such a way that they create viable new compounds. FBDD is a significant help in ML methods of drug discovery specifically. Not only does it reduce the search space for chemical compounds, but it also significantly reduces the proportion of invalid chemical compounds in the search space. Due to the impact of FBDD, dozens of drugs have been discovered and reached clinical trials [20].

## 2.5 Method of Fragmentation

The automatic decomposition of molecules into fragments is not straightforward. Here, the task is to decide where to "cut" a molecule, such that the fragmentation works for a wide variety of structures, and the produced fragments represent of chemically significant pieces. Many algorithms exist for this task, for example, **R**etrosynthetic **C**ombinatorial **A**nalysis **P**rocedure (RECAP). Another method which

has been shown to be able to break apart more compounds than RECAP [11] is the **B**reaking of **R**etro-synthetically **I**nteresting **C**hemical **S**ubstructures (BRICS) algorithm [11]. Given this, BRICS was used for the fragmentation method employed in MOEA-DT. BRICS decomposes the given molecule according to a scheme that identifies strategic bonds based on retro-synthesis of the molecule, attaching "dummy" atoms to both sides of the cut bonds. Also included is a filter which prevents the generation of unwanted motifs or small fragments. All possible cuts are performed simultaneously, which avoids the generation of overlapping fragments.

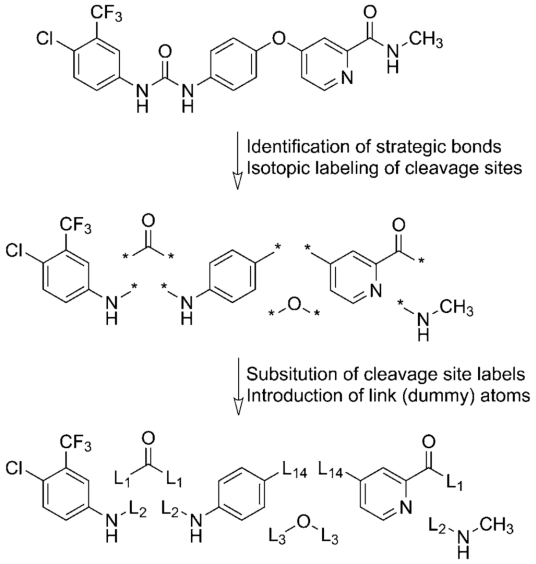



Figure 2.1: Illustration of the BRICS Algorithm [11]

## 2.6 Structure Based Drug Design & Multi-Targeting

Structure based drug design (SBDD) is another widely-used method both in computational and non-computational drug design. SBDD leverages manipulations of the 3D structure of a chemical to determine the goodness of fit into a predetermined pocket of a protein (called protein-ligand docking), which is a well-known indicator of suitability for achieving therapeutic effect. Means of determining docking scores through virtual docking like Autodock Vina [73] and QVina2 [2] exist to provide an interface through which ML algorithms can utilize SBDD as a design criterion, without need for a real-life study. They help by calculating, and determining the minimum the free energy of the molecule $\Delta G = min(C)$ through a weighted sum of atom pair interactions $h_{t_i t_j}$, throughout a number of different molecule conformations

$c_k \in C$ on a static protein structure:

$$\Delta G = min(C),\ c_k \in C = \sum_{i<j} h_{t_i t_j}(d_{ij})$$

Where $d_{ij}$ is the surface distance between the radii of atoms $i$ and $j$, $t_i, t_j$ are the types of atoms $i$ and $j$, respectively, and $h_{t_i t_j}$ is a weighted sum of steric interactions, hydrophobic interactions, and hydrogen bonding, according to the following scheme:

| Weight | Term |
|---|---|
| -0.0356 | $\text{gauss}_1$ |
| -0.00516 | $\text{gauss}_2$ |
| 0.840 | repulsion |
| -0.0351 | hydrophobic |
| -0.587 | hydrogen bonding |

Table 2.1: Weights of Interatomic Interactions in Vina BAS Calculations

In combination with other design criteria in a multi-objective algorithm, SBDD allows the algorithm to "tailor-fit" the produced molecules to a specific treatment pathway. Going further beyond, multiple proteins can be targeted in SBDD simultaneously through the same method. The net result of this multi-targeting approach is the ability to discover chemical compounds which have multiple pathways of therapeutic effect, and therefore either increase effectiveness of treatment, or reduce the needed dosage of drugs. A desirable trait of SBDD systems is that they can orient drug design towards a variety of different targets and therefore different conditions.

## 2.7 Evolutionary Algorithms for Drug Design

According to *From evolutionary computation to the evolution of things*, a review on applications of evolutionary computation methods by Eiben et al. [16] Evolutionary algorithms (EAs) have certain niche applications. This is the case in certain problems of design, parameter optimization, and other examples of black-box optimization. This is particularly the case when gradient information can only be approximated by sampling solutions.

When it comes to EAs for drug design, the task is to conceptualize various solutions to the molecular optimization problem as a population of individuals on which the

evolutionary operations of evaluation, selection, and variation are leveraged to find better performing novel drugs. To begin, such methods start with a population of molecules, and then apply this evolutionary pressure to improve that population relative to some fitness measure (usually chemical property scores). During evaluation, each individual (candidate molecule) is assessed as to its performance in those design criteria. In selection, some means of sampling the population is used to create a new population. This is typically biased in favour of high performance individuals based on the prior evaluation. The means of sampling and amount of bias vary for each implementation. Crossover and mutation are common ways of implementing variation. In crossover for drug design, offspring individuals (new molecules) usually are created by mixing the genomes (atoms and bonds) of parent individuals. In mutation, the genomes of random individuals are altered slightly (usually by changing an atom, bond, or fragment). This cycle is repeated over a pre-defined number of generations, or until a threshold of performance or improvement is reached. The net result is a population of more highly optimized individuals.

## 2.8 Multi-Objective Evolutionary Algorithms

Extending on earlier EAs, multi-objective optimization can be added to cope with problems for which there are several conflicting objectives. To this end, a set of optimal "trade-offs" (called Pareto optimal solutions) between the objectives, where one objective cannot be improved without worsening another is sought after. In such as set, no individual outperforms another in any measure without also underperforming in another one. This is called a non-dominated set. According to an overview of **M**ulti-**O**bjective **E**volutionary **A**lgorithms (MOEAs) by Coello Coello et al. [7], one of the most popular methods for implementing multi-objective optimization is evolutionary algorithms. As Goldberg et al. [24] points out however, optimization is not as simple as merely directly searching for the set of all Pareto optimal solutions. Instead, greater exploration of the solution space is necessary. To this end, an attempt is made to maintain diversity across generations. They show that this can be achieved by sampling equally from every non-dominated set in the population, called non-dominated sorting. Along the same idea of promoting diversity, density estimators like crowding distance began to be used as a means of estimating density in the population space and maintaining diverse solutions. Deb et al. [9] developed a means of comparing individuals on the bases of crowding distance to ensure that selection was guided towards a uniformly spread-out Pareto optimal front. Scalability

to larger numbers of objectives is a known limitation of MOEA. Coello Coello et al. identify three distinct problems:

1. *Deterioration of the search ability*: As the number of objectives increases, the proportion of non-dominated solutions in the population quickly increases. According to their sources, the number of non-dominated $k$-dimensional vectors on a set of n individuals is bounded by $O(\ln^{k-1} n)$, which implies that in many-objective problems, the selection carried out is guided almost entirely by the density estimator. Furthermore, Mostaghim and Schmeck [49] showed that a random search optimizer could achieve better results than then-current MOEAs for problems with 10 objectives or more.

2. *Curse of dimensionality*: As the number of objectives increases, the number of points needed to accurately represent the Pareto front increases exponentially. With $k$ objectives and resolution $r$, this necessary number of points is bounded by $O(kr^{k-1})$.

3. *Visualization of the Pareto front*: When the number of objectives exceeds 3, it becomes difficult to visualize the Pareto front because it would require more than 3 dimensions. Various techniques including heatmaps exist to solve this problem.

## 2.9 DEL: Deep Evolutionary Learning for Molecular Design

DEL [25] combines a deep generative model; a fragment variational autoencoder (FragVAE) with a multi-objective evolutionary algorithm to create a system of data-model co-evolution for molecular design which shows improved performance over baseline models for the same task, while demonstrating the ability to distinctively shift property distributions to more optimal benchmark scores. The benefit that DEL provides over existing methods is that 1) it uses evolutionary computation in the latent space, rather than evolving the parameters of the neural architectures, which tends to be more efficient w.r.t. processing time and memory space. 2) The single neural network is improved throughout processing by learning on the evolved data, making it an effective data augmentation strategy. 3) By carrying out evolutionary operations in the latent space, rather than the structural space, more efficient and smooth

exploration of the space is achieved. 4) Using multi-objective operations and non-dominated sorting with crowding distances helps identify diverse sets of competitive samples from which new offspring can be bred.

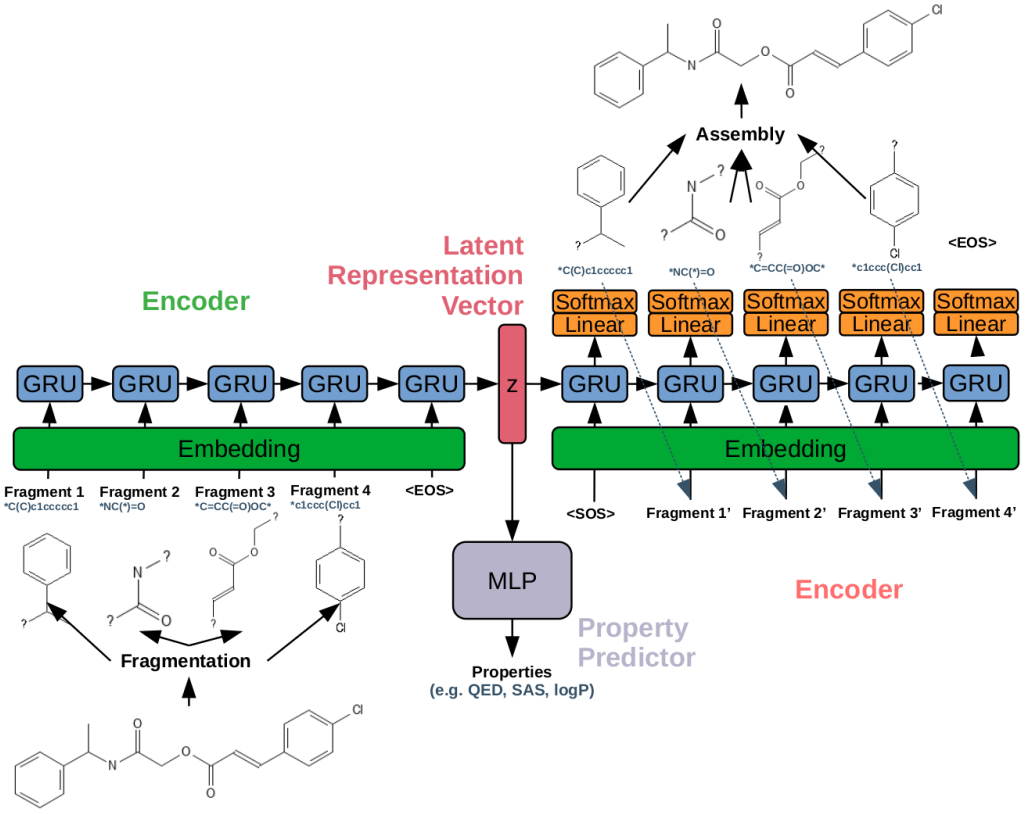



Figure 2.2: An overview of the DEL System [25]

## 2.10 Multi-Objective Drug Design Based on Graph-Fragment Molecular Representation and Deep Evolutionary Learning

Based on the work done in DEL, MODD [50] improves on it by the introduction of protein-ligand binding affinity scores (BAS). The introduction of BAS to the objectives in optimization allows the model to discover molecules which are chemically active towards targeted diseases, thus allowing a deeper search through pruning chemically inactive compounds. Another way in which MODD improves on DEL is by extending the overall method to another neural architecture, Junction-Tree Variational Autoencoder (JTVAE). JTVAE differs from the FragVAE in DEL by modeling

molecules as 2D graphs, rather than as SMILES. This allows for a more intuitive modeling and fragmentation approach.

## 2.11 Tokenization

Whether sequences of molecules, fragments, or natural language characters are at hand, tokenization is often a necessary step in ML applications. Tokenization is a term in natural language processing which describes the process of generating a series of words or subwords which make up the input or output space of an application. In tokenization, a set of tokens $S_T$ is generated through analysis of the expected input or output. A mapping (usually of natural numbers), which maps each token $t \in S_T$ to a unique number is then generated. This allows a vector or series of vectors to be received as input, or produced as output by the ML application. The vector representation is often necessary because it allows a model to apply continuous arithmetic to an otherwise discrete problem which could not be represented numerically. Furthermore, tokenization can be used to significantly reduce the search space of a problem just as in FBDD, by selecting tokens to represent large portions of the space, thereby compressing the information. The input data is separated into tokens, and then encoded through the mapping as input to the model, and then the output of the model is then decoded through its mapping into output tokens, which are then interpreted as the output of the model.

## 2.12 Attention Mechanisms

Attention mechanisms have been a popular approach to enhancing the performance of existing ML techniques, and providing the core concept of transformer models. Zhaoyang et al. [85] provide an overview of attention techniques and their applications. According to them, attention mechanisms exist to solve the problem of information overload by allocating computational resources to the most appropriate areas of the information space. Attention mechanisms have applications for just about every ML task, with shown improvements in image captioning, generation, text classification, machine translation, action recognition, image-based analysis, and speech recognition. Although they are quite useful, their lack of interpretability can cause ethical and practical problems.

Generally, the attention process can be split into 2 steps: 1) compute the attention distribution on the input information, and 2) compute the context vector according to

that distribution. To start, a neural network (NN) is used to encode the source data features as $K$, or the *key values*. Here, there can be many different representations for $K$, depending on the application. In addition to this, task-related features are encoded as $q$, or *query values*. Typically, $q$ consists as a vector, many vectors, or a matrix. With these, a NN is used to compute the correlation between $q$ and $K$ via a compatibility function $f$, which produces an energy score $e = f(q, K)$. Uniquely, in self-attention mechanisms, this function takes the form $e = f(K)$, without regard to $q$. Either way, the energy scores are then mapped onto attention weights $\alpha = g(e)$ by an attention distribution function $g$. Once the attention distribution function is computed, it can be combined with the vector $v$ or the *values vector*. Each element of this vector corresponds to exactly one element in $k$. Often this involves the values being identical, but in come applications they may be different representations of those values. Given the values and the attention distribution, the context vector $c = \phi(\{\alpha_i\}, \{v_i\})$ can then be computed. Here, $\phi$ is a function which returns a single vector, given a set of values and their corresponding weights. A common implementation is a weighted sum, where $c = \sum_{i=0}^{n} \alpha_i z_i$. In effect, a query and a set of key-value pairs are mapped to an output. This makes the result a sum of values which is weighted on the basis the computed compatibility between query and the corresponding key.

## 2.13 Multi-Head Attention & Transformer Models

Building off of work on attention mechanisms, Vaswani et al. [75] introduced the transformer model in the landmark paper *Attention Is All You Need*. This introduction was significant because the model relied entirely on the multi-head attention mechanism to determine dependency between input and output, rather than relying on recurrence, as did previous models like recurrent neural networks RNNs. One of the most significant aspects of their work is multi-head attention. This begins with packing the query, key, and value vectors into matrices query $Q$, key $K$,and value $V$, and then taking the a scaled dot-product attention:

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V$$

Where $d_k$ is the dimentionality of $K$. This attention function is then applied to linear projections of the queries, keys and values, using different, learned, linear projections. Each projection is called a *head*, which collectively constitute multi-head attention.

Multi-head attention is significant because it offers better performance than recent RNNs at a fraction of the computational cost, which can be further reduced through parallelization. Vaswani et al. go on to show high performance of their transformer model on a number of machine translation tasks.

## 2.14 BERT

An extension of the transformer model, Devlin et al. [12] introduced an encoder-only language representation model called BERT, which stands for **B**idirectional **E**ncoder **R**epresentations from **T**ransformers. BERT breaks learning down into pre-training on unlabeled data, and based on the model produced in that process, fine-tuning using labelled data. Pre-training involves masked sequence reconstruction, during which a random assortment of tokens in a sequence are removed and replaced with a mask token. The model is then tasked with predicting the correct token of the mask, based on its context. This allows the model to learn the context surrounding a token from both the ones that come before it and after it in the sequence, hence bidirectional. For supervised learning tasks like classification, the multi-head attention architecture of BERT can be appended to with a series of learned *classification heads*, which take special embedded classification tokens, and interpret them as class probabilities, like a neural network. As Figure 2.3 shows, a single pre-trained model can be re-used for multiple fine-tuned tasks.

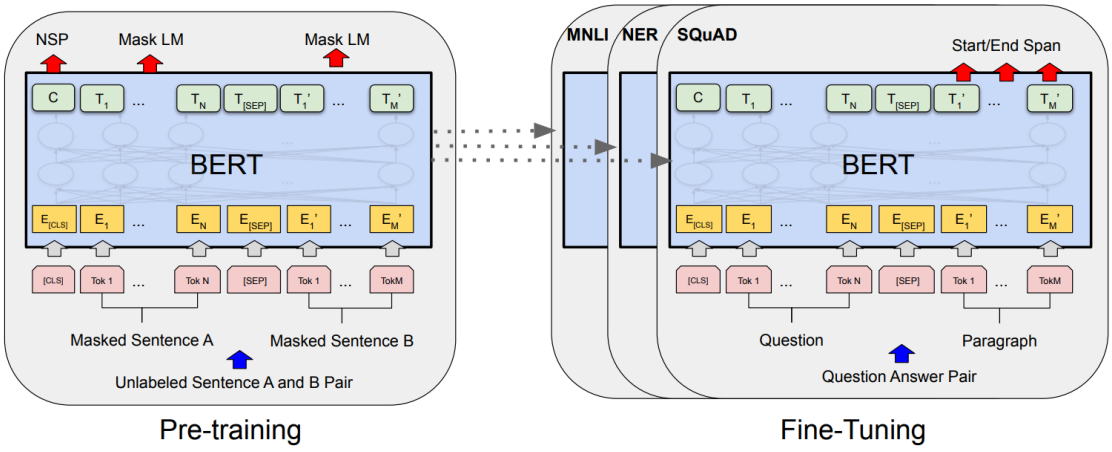



Figure 2.3: An Overview of the BERT System [12]

## 2.15 Transformer Models for Drug Design

Recent advancements in machine learning have yielded new state-of-the-art models for sequence learning like transformers, which makes them well-suited for tasks which use SMILES-based molecular representation. Wang et al [76] . developed a pre-trained transformer for de novo drug design which is pre-trained on the MOSES [59] dataset to generate drug-like compounds without regard to protein targets, and then is fine-tuned on target specific datasets in order to orient molecular generation towards target-specific compounds. Similar to the MOEA-DT half of this project, this model is designed to generate new drug compounds with high-performance property scores. In contrast to our approach, their method uses a **decoder-only** transformer, as opposed to our **encoder-only** SEMTL-BERT model. The main contributions of the work are that 1) the chemical space can be explored using the pre-trained model to generate high-performance, drug-like, molecules even without target-specific pre-training; 2) the fine-tuned model is capable of generating new molecules which correspond to their targets; and 3) the system is capable of being generalized to perform the same optimization process on different drug targets. This approach is limited however in that 1) the use of transformer models from end-to-end increases the number of learnable parameters relative to an MOEA approach; and 2) side effects are not considered.

## 2.16 Types of Classification

Classification is an umbrella term for a variety of problems in ML. Most problems fall into one of the following sub-types:

*Binary classification* is a prediction task for which an output of either 1 or 0 is expected for any input and a single class. An example of a binary classification class would be sorting pictures into ones which contain a car, and ones which do not.

*Multi-label classification* is a prediction task for which given $K$ classes, an output of any number between 0 and $K$ of class labels are expected for any input. An example of this type of problem would be determining which subset, out of a set of cars are present in a series of pictures.

## 2.17 MTL-BERT

Xiao-Chen et al. [84] created a multi-task encoder only Multi-Task Learning BERT *MTL-BERT* model for drug classification and regression. This model provided the basis for the side-effect prediction section of this system. The main contribution of their work is 1) the creation of a multi-task transformer model capable of prediction over multiple classification and regression tasks simultaneously; and 2) the discovery and analysis of correlations between several different classification and regression tasks. One limitation of their approach is that it was only designed to handle binary classification tasks, nor was it designed for multi-label problems.

## 2.18 Semantic Embedding

Semantic embedding is the process of encoding natural language into a latent representation, such that the underlying meaning of the language is preserved. Typically this involves encoding text into a real-valued vector, using a large pre-trained language model. The values contained within the vectors are expected to be encoded in such a way that mathematical similarity of the vectors corresponds to semantic similarity of the corresponding input text. Semantic embedding is useful for many NLP tasks, because it immediately translates natural language into a format which is useful as input into ML applications.

## 2.19 Cosine Similarity & Similarity Matrices

Given two vectors, $\vec{u}, \vec{v}$ their cosine similarity $S_c(\vec{u}, \vec{v})$can be determined by computing the following sum:

$$S_c(\vec{u}, \vec{v}) := \cos\theta = \frac{\vec{u} \cdot \vec{v}}{||\vec{u}||\,||\vec{v}||} = \frac{\sum_{i=1}^{n} \vec{u}_i \vec{v}_i}{\sqrt{\sum_{i=1}^{n} \vec{u}_i^2 \cdot \sum_{i=1}^{n} \vec{v}_i^2}}$$

Where $n$ is the length of the two vectors and $\vec{u}_i$ and $\vec{v}_i$ are the $i$th values of their respective vectors. The range of this similarity is [0,1], where 0 represents orthogonal vectors, indicating complete dissimilarity and 1 represents coincident vectors, indicating complete similarity. Given $k$ vectors, a cosine similarity matrix $S$ of size $k \times k$ can be computed, which identifies the cosine similarity between any vector $\vec{i}$ and $\vec{j}$ with the corresponding entry in the matrix $S_{i,j}$.

## 2.20 Spectral Clustering

Spectral clustering is a useful technique used in this work as part of a class reduction scheme. Roughgarden and Valiant [63] give an overview of the mathematical foundations of spectral graph theory, which is the basis of spectral clustering. According to them, spectral graph theory arises from the revelation of graph properties though studying the eigenvalues of the matrix representations of graphs. Eigenvalues are often determined by consulting the Laplacian Matrix $L$ of a graph $G = \{E, V\}$. There are many different kinds of Laplacians, but the simplest is $L_G$ which is defined as the difference of the diagonal degree matrix $D$, and the adjacency matrix $A$, $L_G = D - A$. Where, $D$ is the diagonal matrix in which the $i$th value along the main diagonal is the sum of the weights of the edges which connect to node $i$, and where each number in $A$, $A_{i,j}$ has the value of 1 if and only if there is an edge between node $i$ and $j$, and a value of 0 otherwise. Thus, $L_G$ has the following value:

$$L_G(i,j) = \left\{ \begin{array}{ll} deg(i) & \text{if } i = j \\ -1 & \text{if } (i,j) \in E \\ 0 & \text{otherwise.} \end{array} \right\} \tag{2.1}$$

Where $deg(i)$ is the degree of $i$, or $D_{i,i}$.
The calculation of a graph's Laplacian matrix is important because the eigenvalues of that matrix reveal information about the structure of that graph.
This can be shown by the following:
Suppose some vector $v$ is multiplied by $L_G$. The resulting product, $Lv$ can be represented by the following sum:

$$Lv_i = deg(i)v_i - \sum_{j:(i,j)\in E} v_j = \sum_{j:(i,j)\in E} (v_i - v_j)$$

Here, the $i$th element in the product is the sum of differences between $V_i$ and the indices of $v$ corresponding to the neighbours of $i$

Von Luxburg gives an in depth explanation of the spectral clustering algorithm. [43] Here, the goal of clustering is described as to compute $k$ clusters $CL_1, ..., CL_k$ based on a given similarity matrix $S$ such elements in the same cluster are similar to each other, and elements in different groups are dissimilar to each other. The algorithm for spectral clustering as von Luxburg describes can be seen in Algorithm 1 and 2.

**Algorithm 1** Algorithm 1: Spectral Clustering

```
Spectral_Clustering()
    Let sim(x_i, x_j) be the cosine similarity of points x_i and x_j
    Let W be the similarity matrix such that W_ij = sim(x_i, x_j)
    Let I be the identity matrix
    Let d_i be the following sum: sum_{j=1}^{n} W_ij
    Let D be the diagonal matrix with d_i values along the main diagonal

    Compute the normalized Laplacian Matrix L_sym given by:
    L_sym = I - D^{-1/2} S D^{-1/2}
    Compute the k smallest eigenvectors u_1, ..., u_k of L_sym
    Let U ∈ R^{n × k} be the matrix containing the vectors u_1, ..., u_k as columns.
    Form the matrix T ∈ R^{n × k} from U by normalizing the rows to norm 1,
    such that t_ij = u_ij / (sum_k u_ik^2)^{1/2}
    For i = 1,...,n, let y_i ∈ R^k be the vector corresponding to the ith row of T
    Cluster the points (yi)_{i=1,...,n} in R^k with the k-means algorithm into
    clusters C_1, ... , C_k
    Return: Clusters A_1, ..., A_k such that A_i = {j|y_j ∈ C_i }
```

**Algorithm 2** Algorithm 2: K-Means Clustering

```
Input:
    k := number of desired clusters
    x_1, ... , x_n || x_i ∈ R^n := set of points to cluster
K-Means()
    Initialize k cluster centroids randomly
    Loop Until Convergence:
        Assign each point x_i to the cluster of its nearest centroid
        by euclidean distance
        Update the position of all centroids to the average of the
        points assigned to them
```

## 2.21 Sci-BERT

Sci-BERT is a pre-trained large language model built for NLP tasks in the domain of scientific data created by Beltagy et al. [3]. Sci-BERT seeks to annotate data in the scientific domain using a pre-trained language model. Their model was able to achieve state-of-the-art performance on five separate NLP tasks: 1) Named Entity Recognition, 2) Participants, Interventions, Comparisons, and Outcomes (PICO) Extraction, 3) Text Classification, 4) Relation Classification, 5) Dependency Parsing. In pre-training, they produced an in-domain vocabulary, which could then be used by stacking task-specific architectures on top of the frozen embeddings.

## 2.22 Side-Effect Prediction

Side-effect prediction as an ML task is growing in popularity, with many recent contributions ranging from purely statistics-based approaches which cluster drugs according to their structure [14], to DL approaches which include chemical structure and/or gene expression information in their training data [74] [46].

### 2.22.1 DrugClust

Dimitri et al. [14] created DrugClust, an ML algorithm for side effect prediction which leverages clustering and Bayesian statistics to model the drug-side effect relation.

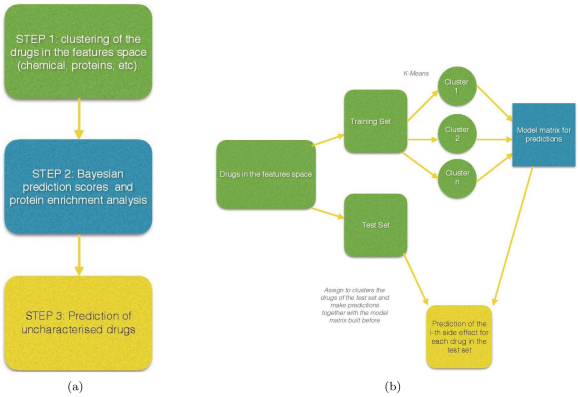



Figure 2.4: An overview of the DrugClust System [14]

Before training, the algorithm splits the data according to a 5-fold cross validation scheme. Training works by first clustering the training set according to Hamming

distances calculated between the matrices of the chemical structure and/or protein interaction profile. Three clustering algorithms were tested:

1. K-Means: A popular clustering method defined in Algorithm 2.
2. PAM: Partitioning Around Medoids. A variant of K-means which requires the centroids to be members of the given data (a mediod, by definition).
3. K-Seeds: A novel clustering algorithm which applies PAM, but maintains the same mediods throughout the clustering process without swapping with other datapoints.

Following clustering, the cluster-conditional side effect probabilities are calculated:

$$P(E_i|C_j) = \frac{P(C_j|E_i)\,P(E_i)}{P(C_j)}$$

Where $P(E_i|C_j)$, the probability of a drug possessing side effect $E_i$, given its corresponding structure cluster $C_j$ is calculated by leveraging Bayes theorem on:

1. The probability of having side effect $E_i$, regardless of the cluster, $P(E_i) = \frac{\text{\# of drugs with } E_i \text{ in the training set}}{\text{\# of drugs in the training set}}$
2. The probability of a drug belonging to structure cluster $C_j$, $P(C_j) = \frac{\text{\# of drugs in } C_j \text{ in the training set}}{\text{\# of drugs in the training set}}$
3. The probability that a drug belongs to structure cluster $C_j$, conditional on it having side effect $E_i$, $P(C_j|P(E_i) = \frac{\text{\# of drugs with } E_i \in C_j.}{\text{\# of drugs with } E_i.}$

Finally, side effects are predicted for drugs from the test set by first assigning them to a cluster, and then sampling their cluster-conditional side effect probabilities. In their experiments, they determined the optimal number of clusters to be 3, and also tested versions with 4 and 5 clusters. They then evaluate the performance of their model on benchmark approaches [48], [41], [83].

DrugClust is a simple approach to side effect prediction which shows improvement on baseline methods in some cases. Their method is limited however in that 1) the number of clusters used relative to the size of the dataset is quite low. To demonstrate this, in the dataset with the fewest compounds (658, by the one used by Mizutani et al. [48]) the average number of compounds per cluster would be 219. This represents a potentially quite diverse set of molecules from which cluster-conditional probabilities are sampled. This problem grows with the size and diversity of the data. 2) Due to

the fact that molecules are clustered on the basis of similarity measure, study into which structures have a causal relation with the side effects is limited. Such a model is also limited in its ability to generalize to novel structures for the same reason. Finally, 3) K-means is a clustering algorithm which has known limitations in stability [65]. Given the central role that this clustering algorithms play in this approach, lack of stability would be a fundamental problem.

### 2.22.2 DeepSide

Based on the work done in structure-based DL methods for side effect prediction [13], and related gene-expression DL methods [46], Uner et al. developed a series of neural network models which combine structure and gene expression for the task of multi-label side effect prediction [74]. Since the inclusion of gene expression data significantly changes the task, the methods of DeepSide which use that data will not be covered here. However, it's worth noting that the best performance reached in their experiments used solely chemical structure information anyway. Furthermore, the source code for the SMILESConv model could not be retrieved. Consequently, SEMTL-BERT was bench marked against the multi-layer perceptron (MLP) and the residual MLP (ResMLP). As in [46], their gene expression data consists of samples obtained from the LINCS L1000 dataset [70]. Besides the gene expression data, they obtain two datasets for chemical structure-only experiments: 1) corresponds to all experiments from LNCS L1000, and consists of 791 drugs which are either approved or in the experimental stage in the format of 166-bit MACCS chemical fingerprint matrix (formatting performed by OpenBabel [52]). 2) corresponds to only the highest quality experiments from LNCS L1000, and consists of 615 SMILES strings selected out of the prior 791, with SMILES strings extracted via RDKit [34]. In addition to this, drug side effect labels were obtained from the SIDER Database [33] for which filtering of side effects with fewer than 10 drugs leaves 1052 side effects. These side effects are then clustered according to the ADR ontology database (ADReCS) [6], which provides a hierarchical classification of side effects. Using this, they divide the side effects into 24 separate clusters based on the highest part of the hierarchy.

As part of their experiments, they test performance on many different neural architectures. The ones which were trained on chemical structures alone include: multi-layer perceptron (MLP), Residual Multi-Layer Perceptron (ResMLP), and SMILES Convolutional Network (SMILESConv). ResMLP is a modification of MLP such that the output of every intermediate fully connected layer is added to by the respective input

element-wise. This method exists to solve the vanishing gradient problem [26]. The SMILES convolutional network which performs a 1D convolutional operators for representation learning on SMILES strings. The network employed in their experiments contains 200 of these convolutional layers with kernel sizes ranging from 1 to 200. After passing through the convolutional layers, the vectors are concatenated together to form the input into the classification layers.

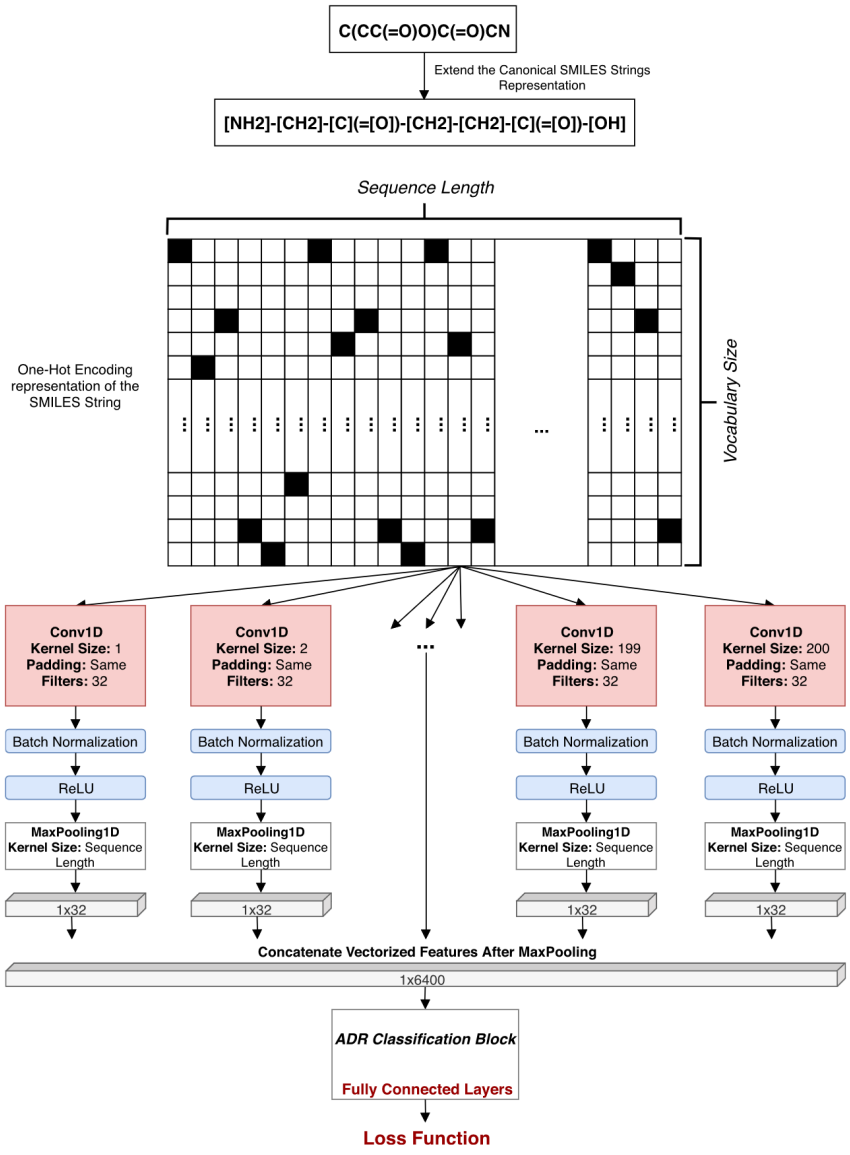



Figure 2.5: DeepSide SMILESConv Architecture [74]

As part of their experiments, MLP and ResMLP were tested on the first dataset of 791 molecules, whereas MLP and SMILESConv (with and without weight on binary

cross entropy loss) were tested on the second (615 molecules). The performance of each model is then compared on the basis of areas under the curve for both receiver operating characteristic and mean average precision. In the concluding paragraphs of their work, they discuss the performance of the models which utilize both chemical structure and gene expression data. Here, they assess the contribution of each feature to prediction by comparing the predictive performance of all models with and without the feature. They conclude from this study that all of the top 100 most important features are related to chemical structure, and a further 40 of the next 100 most important features are likewise.

DeepSide performs a thorough exploration of different methods and data for side effect prediction. By using the ADReCS database, they achieve leverage domain knowledge to establish a valid clustering method. They achieve high scores in performance metrics and also perform thorough analysis on the breakdown of their model's predictions. An important area which were not covered by their experiment include testing the latest ML techniques like attention mechanisms and transformer models. Another important area is the selection of the side effect clusters. The ADReCS database has a 4-level hierarchy which classifies side effects via an "is-a" relationship shown in Figure 2.6. For the work done in DeepSide, the highest level of the hierarchy (and therefore the greatest reduction in classes) was chosen. But it would also be interesting to explore 2nd and 3rd level hierarchies as a means of clustering the data. This would provide the benefit that the end predictions of the model would be more specific. However, it would also make the multi-label prediction task more difficult, as well as removing the means of classifying 1st tier side effects. The latter can be solved with the creation of new synthetic classes.

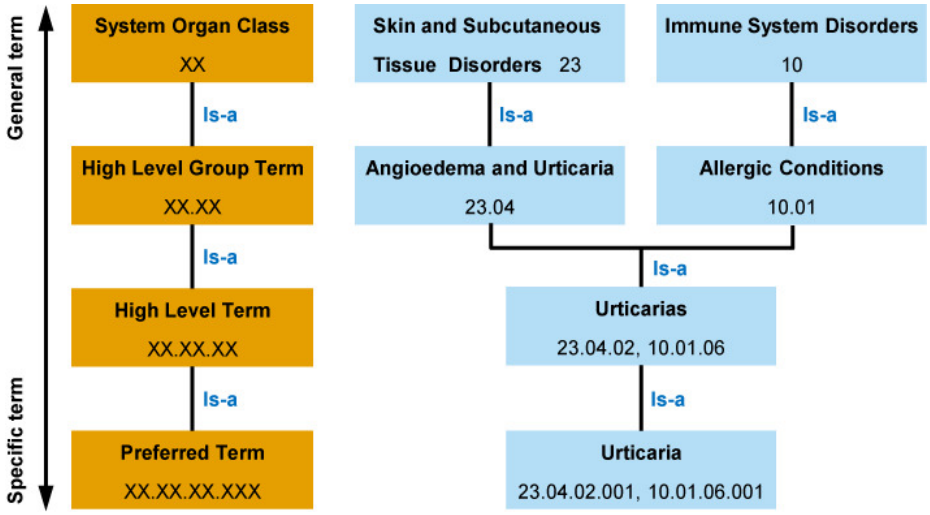



Figure 2.6: ADReCS Database Hierachry & Example [6]

# Chapter 3

# Methods

In this chapter, the details of the used ML techniques are discussed. The process of each algorithm is discussed in detail sequentailly. MOEA-DT is a molecular optimization algorithm which produces a set of high performance drug candidates. SEMTL-BERT is a side effect prediction encoder-only transformer model which uses learned association between drug structure and side effect to predict the potential side effects of new drugs. The models can be used separately, and are completely separate in terms of their architecture, training, task, and data. However, they can also be used together to produce a set of novel, high-performance, drug candidates, for which there are predicted side effects, according to the pipeline shown in Figure 3.1.

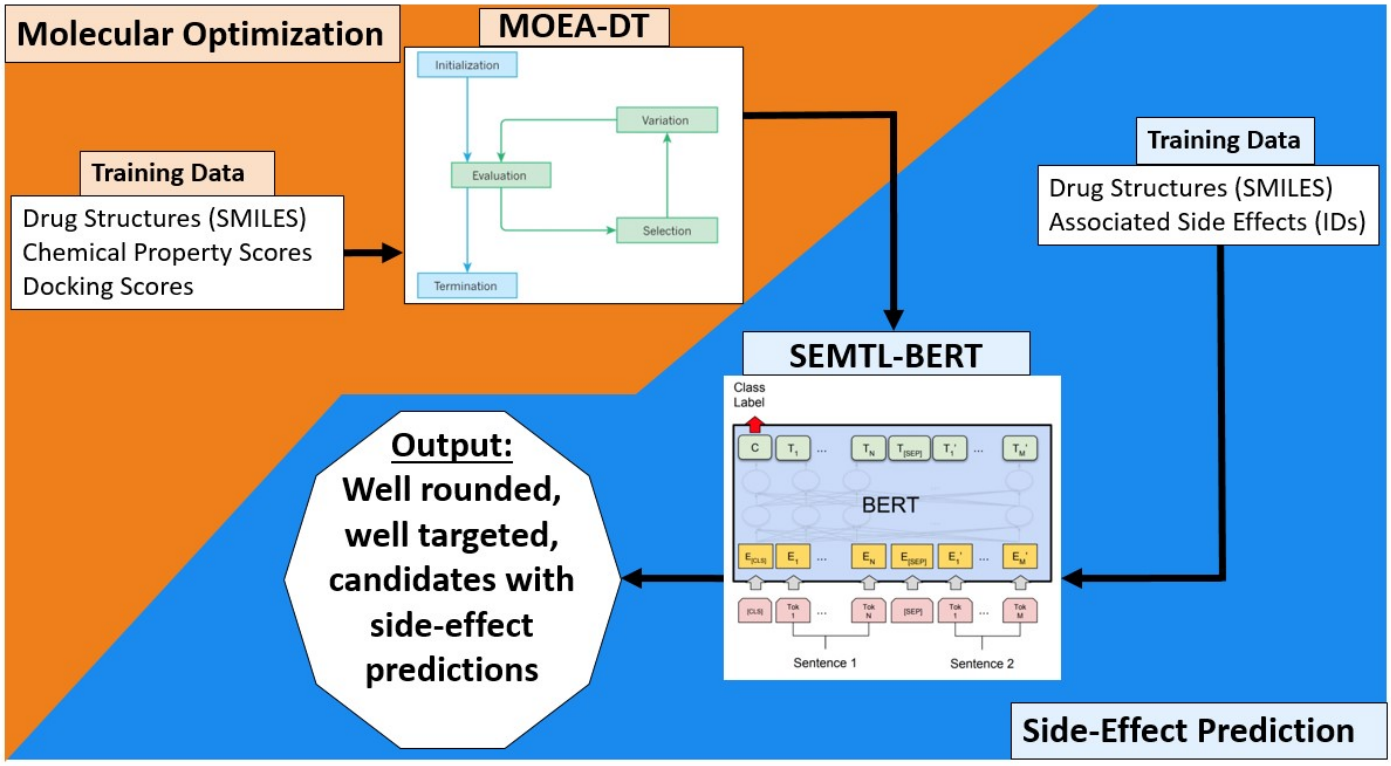



Figure 3.1: An Overview of the MOEA-DT & SEMTL-BERT System [12] [16]

## 3.1 MOEA-DT

Molecular optimization in drug design is the process of determining a set of candidates which show high performance with respect to certain criteria. To achieve this result, the **M**ulti-**O**bjective **E**volutionary **A**lgorithm, with **D**ual **T**argets begins with a population consisting of molecules sampled from the training data. This data is then pre-processed to identify fragments via the BRICS algorithm. These fragments are then added to a library for sampling later in the algorithm. From there, the process of selection, recombination, and mutation are carried out.

### 3.1.1 Selection of the Docking Target & Binding Sites

The proteins CA9 and GPX4 were selected as targets for the optimization of new molecules in MOEA-DT. In order for binding affinity calculations to be carried out by Qvina2 [2], a *bounding box* on the target protein is selected. This is a small 3D region on the surface of the protein which is user-specified, and needed in order to

constrain optimization.

#### CA9

CA9, or Human carbonic anhydrase IX, is a protein which upregulated in solid tumors undergoing hypoxia [21] [68]. This important because of the role that hypoxia plays in cancer. For example, hypoxic cells are particularly radioresistant, and correlates with resistance to chemotherapy. Without CA9, tumor survival under hypoxia is disturbed [55]. Since CA9 is one of the best-known genes associated with hypoxia, it's good therapeutic target for cancer, especially in combination with chemotherapy. The binding site on the CA9 protein was selected to be a box centered on [7.750, 14.556, 6.747] within the binding pocket, with height, width and depth of 20.

#### GPX4

Glutathione peroxidase 4 (GPX4) is a key enzyme in the prevention of oxidative stress of cells. It is also a regulator of ferroptosis (cell death by iron deficiency). [22] [82] Its role in ferroptosis can be leveraged as a form of programed cell death. Since inhibition of GPX4 is one way to cause ferropsis of the cell, and GPX4 is overexpressed in most cancer cells [86], its manipulation can have a therapeutic effect. The coordinates of the bounds of BAS calculation on GPX4 was chosen to be [9.67, 7.043, 4.609] with the same box size.

### 3.1.2 Selection

During selection, the non-domination rank and crowding distance of all individuals within the population is established according to the NSGA-II algorithm for multi-objective optimization [9], using the property scores and binding affinity as objectives. Here, an individual $z_1$ is said to dominate $z_2$ (denoted by $z_1 \succ z_2$ in $K$ objectives $f(z) = f_1(z), ..., f_K(z)$ if $\forall k \in 1, ..., K : f_k(z_1) \geq f_k(z_2)$ and $\exists k \in 1, ..., K : f_k(z_1) \leq f_k(z_2)$. In other words, $z_1$ is better than $z_2$ in at least one way, and not worse than it in any way. Using non-domination, all samples are sorted into Pareto fronts, where all individuals in a front dominate those of successive fronts, and do not dominate each other. To determine crowding distance of an individual $z_i$, the sum of the normalized distances between its nearest neighbour above (denoted by $z_a$), and below (denoted by $z_b$) is computed by the following:

$$d(z_i) = \sum_{k=1}^{K} d_k(z_i)$$

$$d_k(z_i) = \frac{f_k(z_a) - f_k(z_b)}{f_k^{max} - f_k^{min}}$$

Where $f_k^{max}$ and $f_k^{min}$ are the maximal and minimal values in the $k$-th objective.
Using these two metrics, an individual $z_1$ is selected over another individual $z_2$ if $z_1$ is a member of a prior front, or if they are both members of the same front, and $d(z_1) > d(z_2)$.
Selection is repeated for successive individuals until the selection rate is reached, remaining individuals carry on to the next generation.

### 3.1.3 Evolutionary Operations

Following the creation of a new population through selection, recombination and mutation are then performed. In recombination, random "parent" individuals $P_1, P_2$ are paired to produce "offspring" individuals $O_1, O_2$ to replace them. This is done by randomly selecting a crossover point $i|i \leq min(|P_1|, |P_2)$ between the fragments of the two parents, and then swapping the fragments of the two individuals about this crossover point, such that $O_1$ will have the fragments $F_1, ..., F_i$ of $P_1$, followed by all fragments $F_{i+1}$ and on from $P_2$. The opposite applies to $O_2$. If the new arrangement of fragments for both offspring do not result in a chemically valid arrangement, the parents are replaced with mutations of themselves instead.
Following crossover, every individual will have a user-defined chance of mutation during which the following is performed. A fragment $F_x$ from this individual is randomly selected, and replaced with a randomly selected fragment from those in the training data. The individual is the tested for chemical validity, and this process is reversed and repeated until the individual is chemically valid.

### 3.1.4 Property Objectives

Our approach optimizes the property scores: LogP, SAS, CA9 affinity, and GPX4 affinity, simultaneously. LogP is octonal/water partition coefficient which measures solubility of a compound in either water or lipids, which is important for measuring how well the drug will be absorbed by the body. High scores would mean a molecule is highly lipophilic and hydrophobic, and vice versa with low scores. The second property is synthetic accessibility score (SAS) which is a measure of the ease/cost of

synthesising the compound. Naturally, drugs which are easy to make are preferable to ones that are difficult to make. Finally there is protein-ligand binding affinity score (BAS) for each of the two drug targets. This is a measure of how well the compound fits to the target protein, and therefore how suitable the molecule is for the treatment of the particular disease or disorder. This further indicates the degree to which it may cause toxic off-target effects, simultaneously. [42, 19] LogP and SAS were calculated using packages provided by RDKit [34], and BAS was calculated using QVina2 [2].

### 3.1.5 Hypervolume

In a multi-objective problem, establishing relative performance between differing approaches can be complicated. This complexity derives from the fact that multiple parallel performance measures can conflict with one another, and are measured on entirely separate scales. This problem is compounded in cases where large sets of samples need to be compared, and true optimal samples are not known. Hypervolume (HV) addresses this limitation by producing one scalar value which represents the performance of entire sets of samples. Hypervolume is defined as the region in the target space which is bounded on one side by the set samples, and on the other by a pre-defined reference point. HV scores are defined by the following:

$$HV = Vol(\cup_{i=1}^{|S|} v_i)$$

Where $Vol$ is the n-dimensional volume created between a solution $v_i$ and the reference point, and $S$ is the set of all non-dominated samples. Note that dominated samples do not contribute to HV. Since by definition, there is another solution with same-or-higher performance in every dimension, such samples do not increase the region bounded by the entire population. Thus HV can be calculated only on the subset defined by non-dominated samples. Larger sets of non-dominated samples contain more diverse samples across a performance measures, which contributes to higher length of the bounded region in its respective dimension. Thus, HV gives us a scalar-valued metric which rewards models that produce large, diverse, and high-performance non-dominated sets of samples.

### 3.1.6 Similar Method Comparisons

For performance benchmarking, MOEA-DT was compared to other methods for a the same task of molecular optimization. All evolutionary methods were optimized

over 20 generations, with a population size of 20K, and using the same objectives.

#### AE+DEL

Three methods were tested which use autoencoders for evolutionary operations in the latent space, as part of a larger evolutionary algorithm. Each of these was based on the DEL [25] framework. The first is *FragVAE + DEL*, which is the same method presented in DEL, adapted for optimization over LogP, SAS, CA9, GPX4. The second, *JTVAE + DEL*, is a **J**unction **T**ree **V**ariational **A**uto**E**ncoder. Based on the work by Jin et al. [28], JTVAE uses a graph decomposition method to break input molecular graphs into component pieces using the BRICS algorithm, and creates a graph scaffold wich is encoded and decoded together with the molecular graph using the VAE. JTVAE was integrated with DEL in [25], where SMILES based representation was used. The third, **A**dversarial **A**uto**E**ncoder, was based of of the work in [45], and used in [29] and [1] for generation using molecular fingerprints. This method was similarly adapted to use SMILES representation for use in combination with the DEL framework.

#### MODRL

**M**ulti-**O**bjective **D**eep **R**einforcement **L**earning is a multi-objective molecular optimization framework based on DeepFMPO [69]. It uses a constrained deep reinforcement learning framework with modifications in the fragment space. Property scores of a given set of lead molecules are optimized using a linear combination of the four objectives into a scalar. During training, MODRL optimized a set of 748 lead molecules from DrugBank [80] which did not meet any of the screening criteria. Thus the MODRL agent was provided with a more difficult task, since criteria satisfying molecules were not involved in the training process. MODRL struggled to optimize over a larger training set due to lack of scalibility in compute cost because ligand docking needed to be performed for each action.

## 3.2 SEMTL-BERT for Side-Effect Prediction

In contrast to MOEA-DT, Side-Effect Multi-Task Learning in Bidirectional Encoder Representations from Transformers (SEMTL-BERT) seeks to predict potential side effects of drug candidates, rather than optimize novel molecules. The model is initially pre-trained to reconstruct inputted SMILES strings, and then fine-tuned for

the classification task. SEMTL-BERT predicts side effects by as a series of $n$ binary classification tasks, where $n$ is the number of potential side effects or side effect clusters in the data. Thus, for every input $x$, SEMTL-BERT produces vector of length $n$, where each value, (when put through a sigmoid filter) produces $p(y_i)$, which represents the model's probability prediction for molecule $x$ possessing the side-effect or side effect cluster $i$. This is then compared to the ground-truth of known side-effects $t_i(x)$ during training by using binary cross-entropy loss (BCE), which is defined as:

$$H = -\frac{1}{n}\sum_{i=1}^{n} t_i(x) \cdot \log(p(y_i)) + (1 - t_i(x)) \cdot \log(1 - p(y_i))$$

BCE is a desirable loss function for binary classification problems because it allows the model to get partial credit by using a continuous value (class probability prediction) rather than deriving loss from the discrete classifications themselves. BCE is also desirable for multi-label problems because it produces accepts a vector of class probabilities, which is an intuitive way of interpreting the problem.

### 3.2.1 Model Performance Metrics

The model's performance is measured at each epoch of training, based on the BCE loss, area under receiver operating characteristic curve (ROC AUC), F1 scores, and area under average precision curve (APR AUC). These scores are more desirable metrics than simple accuracy for applications in which there is an imbalance of classes. This is because a model which simply predicts all positives or all negatives will show skewed accuracy towards data which is unbalanced in its favour. F1, APR AUC, and to a lesser degree ROC AUC adjust themselves for imbalance in the data. All three metrics are defined on the basis of precision and recall. Precision $P$ is defined by the ratio between true positives $T_P$ and all positive predictions including false positives $F_P$. Recall $R$ is defined as the ratio between true positives $T_P$ and all correct positives, including false negatives $F_N$

$$P = \frac{T_P}{T_P + F_P} \quad R = \frac{T_P}{T_P + F_N}$$

ROC AUC score is defined as the area under the curve created by plotting the true positive rate (recall) on the y-axis, and the false positive rate on the x-axis.

$$ROCAUC = \int R \, d(\frac{F_P}{F_P + T_N})$$

F1 score is defined as the harmonic mean of precision and recall.

$$F1 = 2\frac{P \times R}{P + R}$$

Average precision score is defined by the weighted mean of precision achieved at each threshold $n$, with the increase in recall from the previous threshold used as the weight:

$$AP = \sum_n (R_n - R_{n-1}) P_n$$

Given that a scalar is required for scores in this framework, the average over the thresholds was taken.

### 3.2.2 Class Reduction by Spectral Clustering

Given a series of classes $(C_1, ..., C_n)$, and their corresponding natural language descriptions $(NL_1, ..., NL_n)$, a pre-trained Sci-BERT [3] model produced embedding vectors from the natural language description of the side effects. Following this, cosine similarity was computed between each embedding vector to produce a similarity matrix $S$, where each entry $s_{i,j}$ represents the calculated similarity of the semantic embeddings of the descriptions of the corresponding classes $(C_i, C_j)$. Using this similarity matrix, spectral clustering was then performed, producing a mapping from the class space to the cluster space. Then, this mapping was used to transform the data from a molecule associated with a series of side effects to a molecule associated with a series of side effect clusters. Thus, through establishing the clusters to which each molecule in the training set belongs, the number of total classes can be reduced to any arbitrary number. Using this, memory usage can be significantly reduced by having the model predict side effect clusters rather than the side effects themselves.

### 3.2.3 SMILES Augmentation

SMILES strings are encoded on the basis of a sequential enumeration of the atoms which make up the molecule. This means that there are multiple valid SMILES representations for any given molecule, each of which corresponds to a different order of enumeration of the atoms in that molecule. This property of SMILES strings can be leveraged as a means of data augmentation. Namely, datasets of SMILES strings can be iterated upon, producing multiple alternate SMILES representations for each molecule. This was done, creating an augmentation dataset of as many as

one thousand alternate SMILES representations for each sample in the dataset.

### 3.2.4 Stratified Data Splitting & Class Balancing

Some of the performance metrics used require at least one sample to possess a positive classification for each class or cluster. Therefore the data was split between training, validation, and testing such that such samples were added first, from most rare positive examples to least, followed by the remainder of the dataset. Following this, the alternate SMILES representations for each sample were appended to their respective dataset. In class-balanced experiments, this function was leveraged to reduce class imbalance by appending fewer alternate SMILES strings to samples which are associated with more common classes. In practice, the likelihood of each class occurring throughout the entire dataset was recorded. Each sample was then assigned a probability score $p_s$ of the average of the probabilities for each of its positive classes. Following this, the maximum number of alternate SMILES was limited by $1000 * (1 - p_s)$. Thus, class imbalance was reduced by reducing the number of samples with common classes.

### 3.2.5 Pre-Training and Tokenization

Before fine-tuning for the task of side effect prediction, the encoder section of the architecture is pre-trained to reconstruct the input data. Since pre-training does not rely on annotated data, a much larger dataset of SMILES strings was be used for pre-training. For tokenization, a SMILES-specific scheme was used rather than a character-by-character encoding.

# Chapter 4

# Experiments

In this chapter, the experiments performed are discussed. Firstly, the training data of each model is reviewed and the sources are discussed. Then, an overview of implementation requirements and hyperparameter settings is given. Following which, the results for each model are given. For MOEA-DT, a comparison is drawn between its performance and that of other models of a current work-in-progress study is shown. For SEMTL-BERT, a comparison between its performance metrics and those of a baseline model is shown, and an analysis of the class distributions is performed. Finally, a case study of molecules throughout both systems working together is shown.

## 4.1 Data

### 4.1.1 MOEA-DT

The training data for MOEA-DT consists of a variant of the ZINC250K [27] database, which was append with approximately 2K authentic drugs from the DrugBank [80] free-to-access dataset. ZINC is a free database of commercially-available compounds for virtual screening DrugBank Online is a bioinformatics and cheminformatics data resource which offers detailed chemical information on drug sequences and structures. The combined samples were then pre-processed by 1) separating them into fragments based on the BRICS algorithm [11], 2) calculating molecular properties and structure features using RDKit [34], and 3) removing any structures which contain only one fragment.

| (Structure Features Withheld) | | | | | | |
|---|---|---|---|---|---|---|
| SMILES | Fragments | Frags | logP | SAS | GPX4 | CA9 |
| CC(C)(C)c1ccc2occ(CC(=O)Nc3ccccc3F)c2c1 | *C(C)(C)C *NC(=O)Cc1coc2ccc(*)cc12 *c1ccccc1F | 3 | 5.05 | 2.08 | -6.6 | -7.9 |
| CC1CC(C)CC(Nc2cncc(−c3nnc n3C)c2)C1 | *C1CC(C)CC(C)C1 *Nc1cncc (−c2nncn2C)c1 | 2 | 3.11 | 3.43 | -6 | -7.1 |
| N#Cc1ccc(−c2ccc(OC(C(=O)N3 CCCC3)c3ccccc3)cc2)cc1 | *c1ccc(C#N)cc1 *c1ccc(OC(C(=O)N2CCCC2) c2ccccc2)cc1 | 2 | 4.97 | 2.47 | 0 | 0 |
| CCOC(=O)C1CCCN(C(=O)c2nc (−c3ccc(C)cc3)n3c2CCCCC3)C1 | *OCC *C(=O)C1CCCN(C(=O)c2nc (−c3ccc(C)cc3)n3c2CCCCC3)C1 | 2 | 4 | 2.82 | -5.6 | -7.3 |
| N#CC1=C(SCC(=O)Nc2cccc(Cl) c2)N=C([O−])C(C#N)C12CCCC C2 | *SC1=C(C#N)C2(CCCCC2)C (C#N)C([O-])=N1 *CC(=O)Nc1cccc(Cl)c1 | 2 | 3.61 | 4.04 | 0 | 0 |

Table 4.1: Sample MOEA-DT Training Data

### 4.1.2 SEMTL-BERT

Pre-training for SEMTL-BERT was run with two configurations: one dataset of approximately 1.7 million unlabeled molecules collected from the ChEMBL database [47] for the smaller architecture, and the other pre-training dataset consists of approximately 4 million unlabeled molecules containing data from ZINC-250K [27], ChEMBL, and MOSES [59]. ChEMBL is a manually curated database of bioactive molecules with drug-like properties. MOSES is a dataset based on ZINC, which has been filtered via medicinal chemistry filters (MCFs) and PAINS filters. The fine-tuning data consists molecules, annotated with associated side effect information, and was retrieved from DrugBank's premium database [80]. Fine-tuning data for SEMTL-BERT consists of 1786 unique samples with 11002 unique side effects.

| (Features Withheld) | | | |
|---|---|---|---|
| SMILES | Name | ID | SE |
| NC1=CC=CC2=C1CN(C1CCC (=O)NC1=O)C2=O | Lenalidomide | 210 | diverticulitis |
| CC(=O)NC1=CC=C(O)C=C1 | Acetaminophen | 1820 | anaphylaxis |
| ClC1=CC=CC(Cl)=C1NC1=NCCN1 | Clonidine | 12116 | decreased appetite |
| NC1=CC=NC=C1N | Amifampridine | 10699 | nausea |
| [Mg++].[Cl-].[Cl-] | Magnesium chloride | 17979 | headache |

Table 4.2: Sample DrugBank Data

| 100 Clusters | |
|---|---|
| SMILES | Cluster IDs |
| [H][C@]12[C@H](OC(=O)C3=CC=CC=C3)[C@]3(O)C[C@H](OC(=O)[C@H](O) [C@@H](NC(=O)C4=CC=CC=C4)C4=CC=CC=C4)C(C)=C([C@@H](OC(C)=O) C(=O)[C@]1(C)[C@@H](O)C[C@H]1OC[C@@]21OC(C)=O)C3(C)C | 98;14;28;99;82;69;39;46;94;90;70;66;89;44; 2;97;26;27;75;42;74;68;47;85;58;71;32;50;73; 61;84;38;8;91;12;65;80;54;18;41;5;77;19;67... |
| CCC1=CC(=CC=C1CN1CC(C1)C(O)=O)C(\C)=N\OCC1=CC=C(C2CCCCC2) C(=C1)C(F)(F)F | 98;14;74;24;69;90;46;89;28;47;78;2;91;41;54; 52;86;4;43;25;36;50;3;8 |
| CN(CC1=CN=C2N=C(N)N=C(N)C2=N1)C1=CC=C(C=C1)C(=O)N[C@@H] (CCC(O)=O)C(O)=O | 92;95;87;68;21;99;69;82;79;74;2;8;28;72;89; 51;85;80;27;32;97;25;90;47;42;86;35;15;63... |
| NC1=NC(=O)N(C=N1)[C@@H]1O[C@H](CO)[C@@H](O)[C@H]1O | 33;74;14;95;87;21;82;15;69;55;44;46;2;8;28... |
| CC(=O)N(O)CCCCCNC(=O)CCC(=O)N(O)CCCCCNC(=O)CCC(=O)N(O)CCC CCN | 47;68;99;69;82;39;74;28;2;89;25;75;24;42;87; 10;71;50;58;37;32;27;49;66;54;5;29;15;91;78... |

Table 4.3: Sample SEMTL-BERT Sample Fine-tuning Data

## 4.2 Hyperparameter Settings

For MOEA-DT the same hyperparameters were used as those in FragVAE, JTVAE, and AAE, which are ML methods being tested in a study which is a work in progress. For SEMTL-BERT, two different architectures were tested. Both were based on the ones in the original MTL-BERT [84] framework.

| Number of Generations | 20 |
|---|---|
| Population Size | 20K |
| Selection Rate | 95% |
| Mutation Rate | 1% |

Table 4.4: MOEA-DT Hyperparameters

| Number of Layers | 4 |
|---|---|
| Number of Heads | 4 |
| Expected Input Features | 512 |
| Batch Size | 100 |
| Learning Rate | 0.0005 |

Table 4.5: SEMTL-BERT Small Architecture Hyperparameters

| Number of Layers | 8 |
|---|---|
| Number of Heads | 8 |
| Expected Input Features | 1024 |
| Batch Size | 100 |
| Learning Rate | 0.0005 |

Table 4.6: SEMTL-BERT Large Architecture Hyperparameters

## 4.3 Implementation Requirements

The use of the MOEA-DT and SEMTL-BERT algorithms requires the use of the following libraries:

- Pytorch 1.13.1
- Botorch 0.8.5
- Numpy 1.23.5
- Scipy 1.8.1
- Scikit-learn 1.2.1
- Python 3.10.10
- tqdm 4.65.0

- Pandas 1.5.3
- Gensim 3.4.0
- Openbabel 3.1.1
- Rdkit 2022.9.5
- Qvina 2.1.0

## 4.4 MOEA-DT Results

### 4.4.1 Population Performance

The molecules produced by the six methods were compared on the basis of validity, novelty, and uniqueness. Validity is a measure of the proportion of the produced samples which are chemically valid, which was measured using RDKit [34]. Novelty is a measure of the proportion of the population which is not found in the training data. Uniqueness is the proportion of the population which is not a duplicate.

| Model | Scores | | |
|---|---|---|---|
| | Validity | Novelty | Uniqueness |
| FragVAE + DEL | 0.980 | 1 | 0.950 |
| JTVAE + DEL | 1 | 0.975 | 0.964 |
| AAE + DEL | 0.884 | 1 | 0.988 |
| MOEA-DT | 1 | 0.99815 | 0.9065 |
| MODRL+Constrain | 0.960 | 1 | 0.284 |
| MODRL + Scalarize | 0.950 | 1 | 0.184 |

Table 4.7: MOEA-DT Population Score Comparison

#### Hypervolumes & Virtual Screening

The produced molecules from each method were compared on the basis of the hypervolumes of the first Pareto rank of the final generation measured from a reference point of [-7.2893,-12.6058, 0, 0] (SAS, logP, CA9, GPX4). Additionally, screening criteria were developed w.r.t the four objectives: $SAS \leq 3, -0.4 \leq logP \leq 5.6, CA9 \leq -7.4$, and $GPX4 \leq -6.3$. The logP score was determined using Ghose filter rules [23]. The CA9 and GPX4 limits were determined using ligands obtained from the Protein Data Bank [4]. Finally, additional experiments were performed using FragVAE and JTVAE using only one target ligand.

| Model | Hypervolume | | |
|---|---|---|---|
| | Generation 1 | Generation 10 | Generation 20 |
| FragVAE (DEL) | 5525.16 | 5596.33 | 5622.85 |
| JTVAE | 5523.13 | 6243.53 | 6248.25 |
| AAE | 5654.34 | 6955.92 | 8320.32 |
| MOEA-DT | 5552.82 | 7209.37 | 8158.17 |
| | Initial Leads | MODRL+Scalarize | MODRL+Constrain |
| MODRL | 2964.52 | 3941.97 | 3379.53 |

Table 4.8: MOEA-DT Hypervolume Comparison

| Samples | CA9 | CA9 & GPX4 |
|---|---|---|
| FragVAE 1st Front | 42 | 178 |
| JTVAE 1st Front | 47 | 283 |
| AAE 1st Front | - | 920 |
| MOEA-DT 1st Front | - | 997 |
| FragVAE All Fronts | 2571 | 3548 |
| JTVAE All Fronts | 2893 | 6708 |
| AAE All Fronts | - | 8377 |
| MOEA-DT All Fronts | - | 8186 |
| MODRL+Scalarize | 461 | 347 |
| MODRL+Constrain | 488 | 392 |

Table 4.9: MOEA-DT Screening Criteria Comparison

## 4.5 SEMTL-BERT Experiment Setup

### 4.5.1 Pre-Training

Two separate architectures were tested for SEMTL-BERT, both of which were based upon the architectures in MTL-BERT [84]. One, with a smaller architecture, was trained on a set of 2 million molecules from the MTL-BERT repository (hereafter *small model*). The other, larger architecture was pre-tained on a set of 4 million molecules by Nicholas Aksamit (hereafter *large model*). Both were pre-trained on reconstruction cross entropy loss for 20 epochs.

### 4.5.2 Class Balancing & Data Augmentation

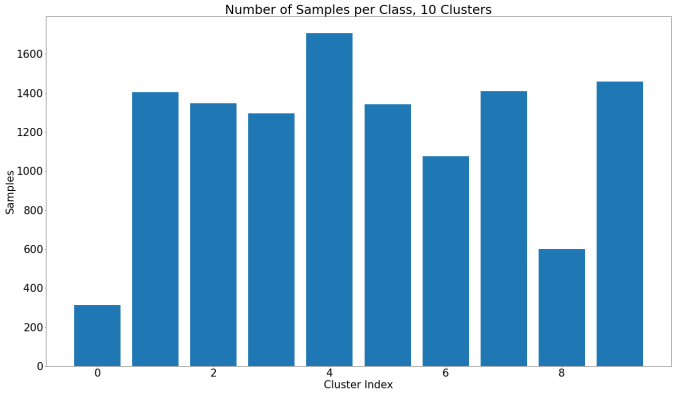
Up to 1K alternate SMILES representations per sample in the fine-tuning dataset were collected using RDKit [34], and then appended to the the set of their respective original molecule, following the split between training, testing, and validation sets. From here, two schemes were followed, one in which all alternate SMILES were appended (hereafter *Unbalanced Classes*), and one in which the amount of alternate SMILES was appended depended on the average class probabilities of the original sample's positive classes, (hereafter *Balanced Classes*).

### 4.5.3 Clusters

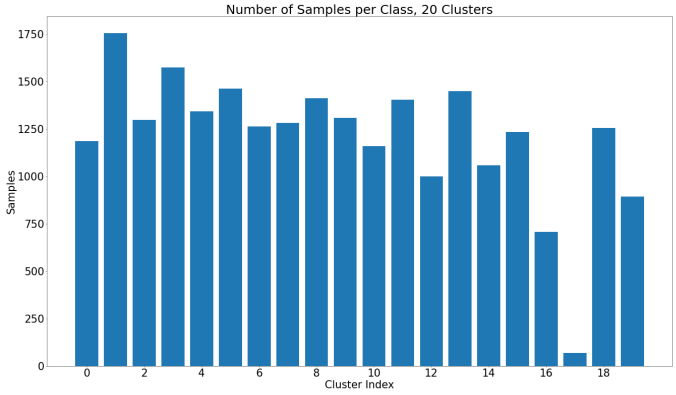
In this section, an analysis of the traits of each clustering arrangement shown in Figures 4.1 and 4.2 will be performed. The class distributions of original data (Figure 4.1e) show for each side effect, the number of samples in the entire training data for which that is an effect. Further, it highlights nausea as the most common side effect, with nearly 1.2K samples (approximately two thirds) of the molecules showing that particular effect. As one can see from the bar chart, the classes are highly unbalanced, with many of the classes belonging to only one sample.

Predictably, the balance between classes increases as the classes are packed into fewer clusters, with few notable exceptions. Namely, in Figures 4.1a, 4.1b, and 4.1c there exists 1-3 clusters for which there are significantly fewer samples than the rest. This increase in class balance likely aids in prediction performance, since the model doesn't need to learn so many edge cases where a class is seldom accurately predicted.

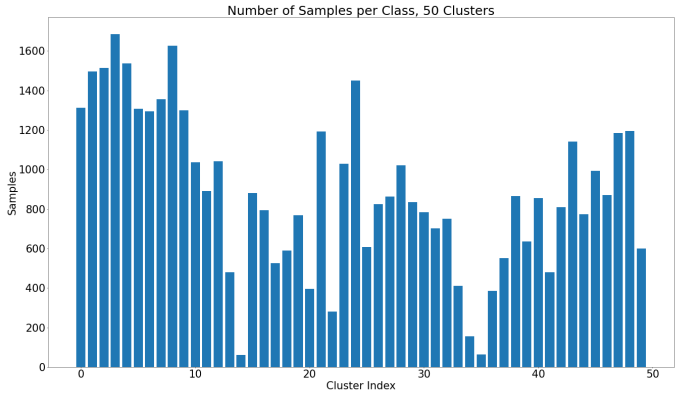
Figure 4.2 gives an overview of the expected number of positive classes per sample. This is an important metric, because lack of balance between positive and negative classes per sample can bias the model. As one can see from the first plot, samples in the original data have few positive classes relative to the total amount of classes. Thus performance of a hypothetical model may be hard to measure, as semi-accurate predictions could be attained simply by predicting every class as negative. In contrast, the arrangement of 10 clusters provides a landscape in which most classes are positive in most samples. Thus the same problem exists but in reverse. An appropriate medium between these two extremes is the 50 cluster arrangement, for which the average number of classes is 25.



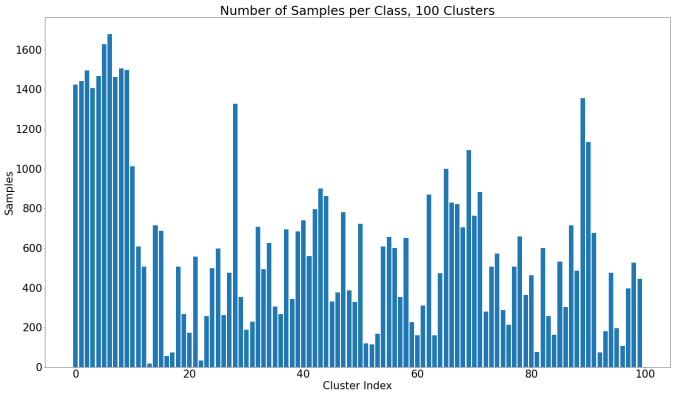

(a) 10 Clusters



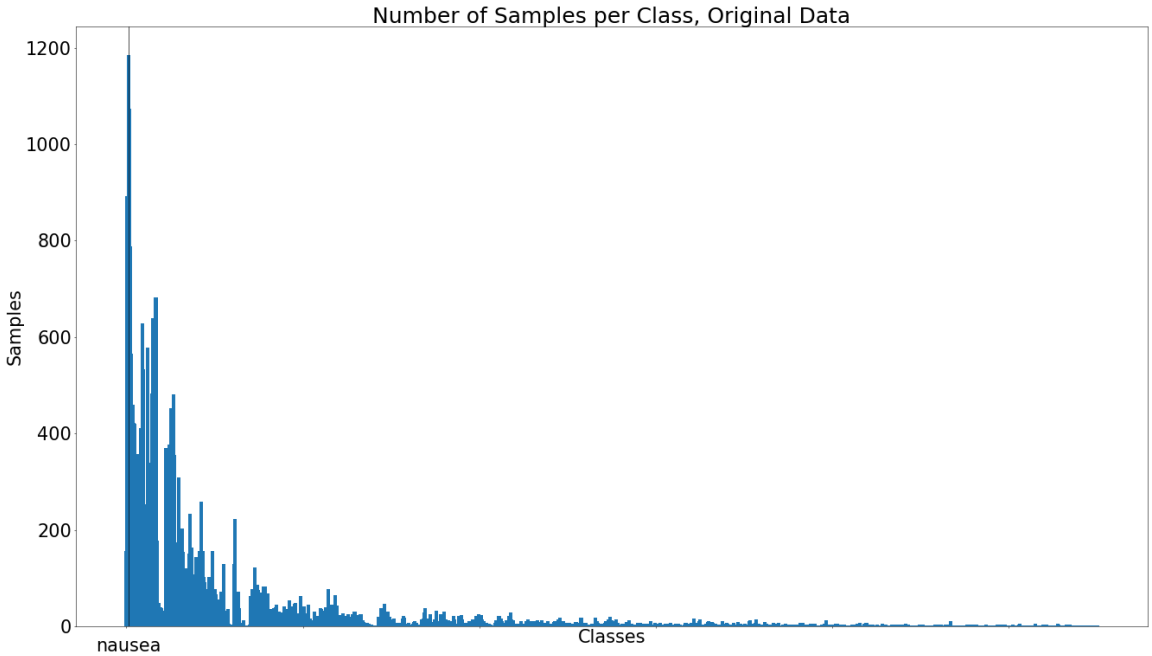

(b) 20 Clusters




(c) 50 Clusters




(d) 100 Clusters




(e) Original Data

Figure 4.1: SEMTL-BERT Class Distributions

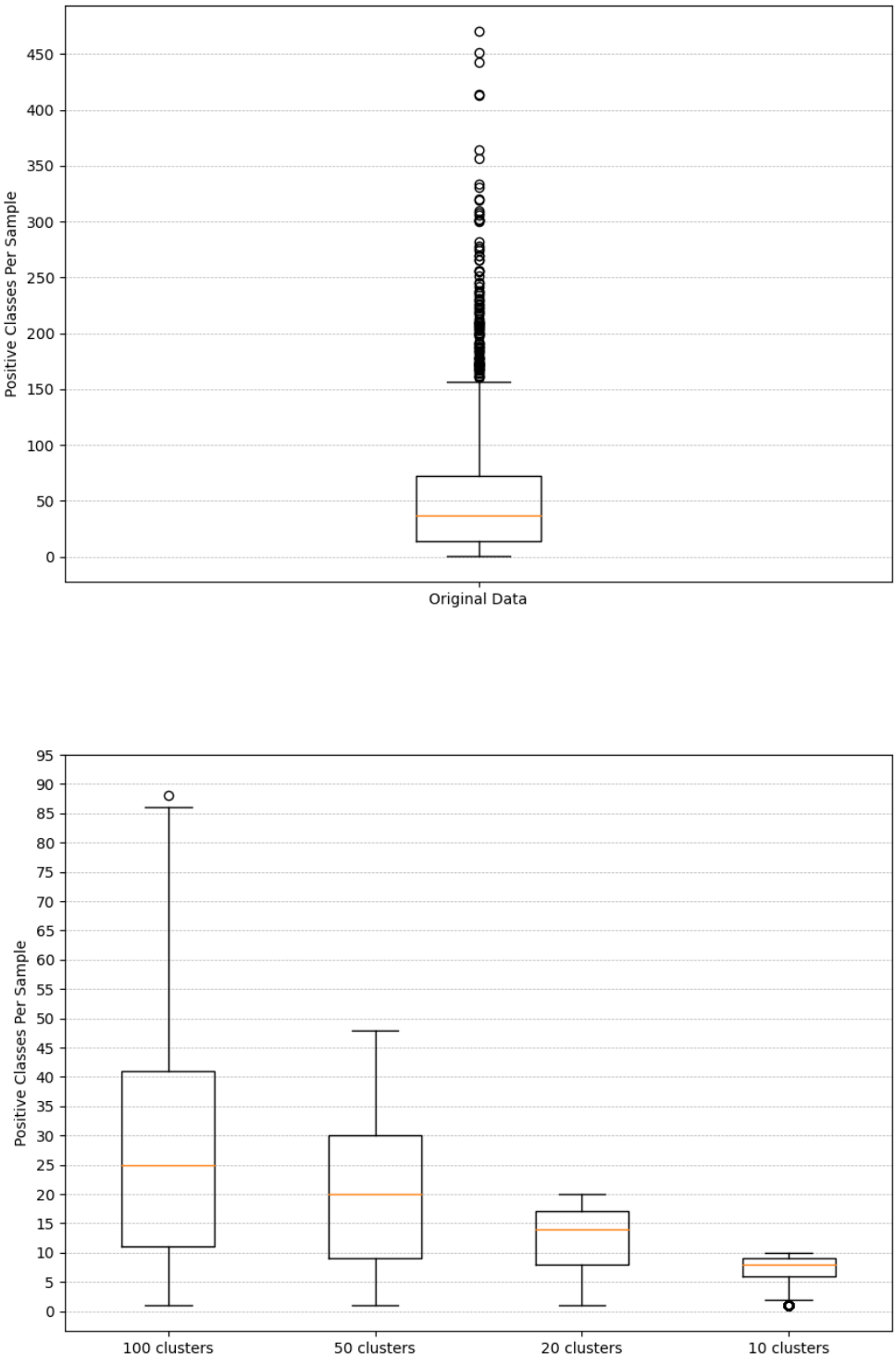
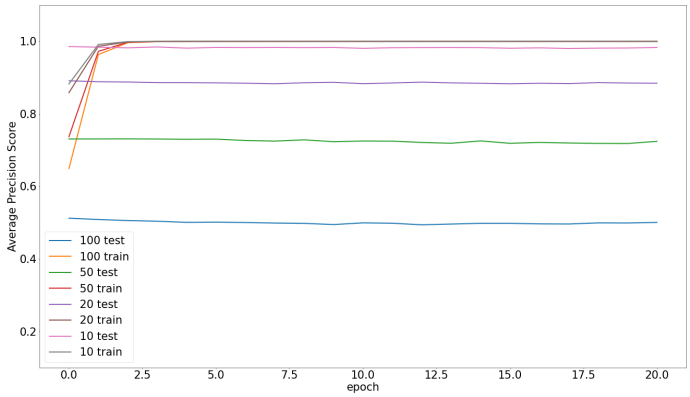



Figure 4.2: SEMTL-BERT Positive Classes Per Sample

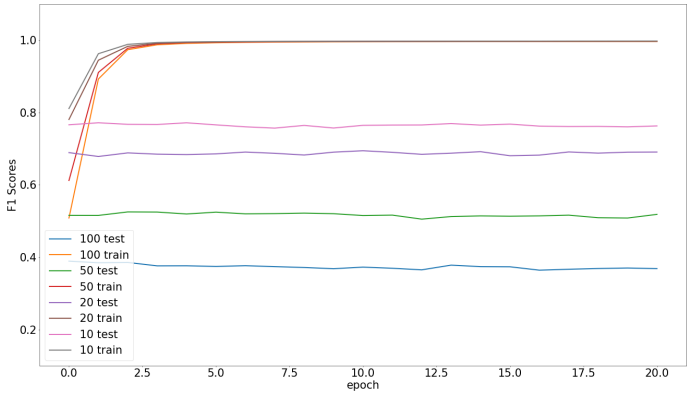
## 4.6 SEMTL-BERT Performance

### 4.6.1 Performance per Cluster Arrangement

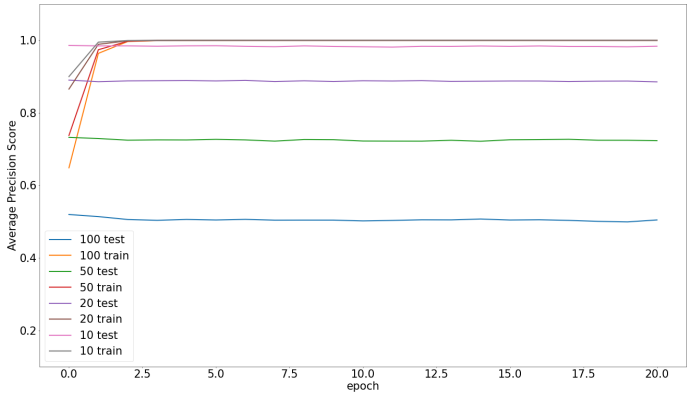
The English language descriptions of the side effects were fed through a pre-trained SCI-BERT model [3]. The final pooling layer of each sentence was used as a semantic embedding. From these vectors, a similarity matrix was created, such that each entry $S_{i,j}$ was the cosine similarity of embedded vectors $i$ and $j$. Spectral clustering was then performed using SciKit Learn, a Python package of useful ML tools [57]. Four clustering arrangements were then created with 10, 20, 50, and 100 clusters respectively. Each of these cluster arrangements were tested. Below are a series of line charts which show SEMTL-BERT performance over the 16 experiments.



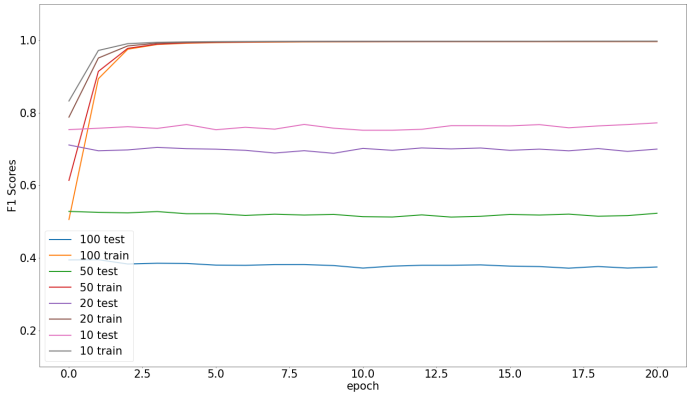

(a) Average Precision, Balanced Classes




(b) F1 Scores, Balanced Classes

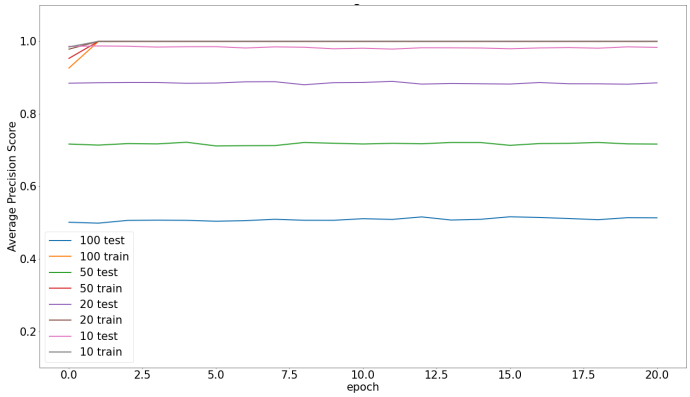



(c) Average Precision, Unbalanced Classes

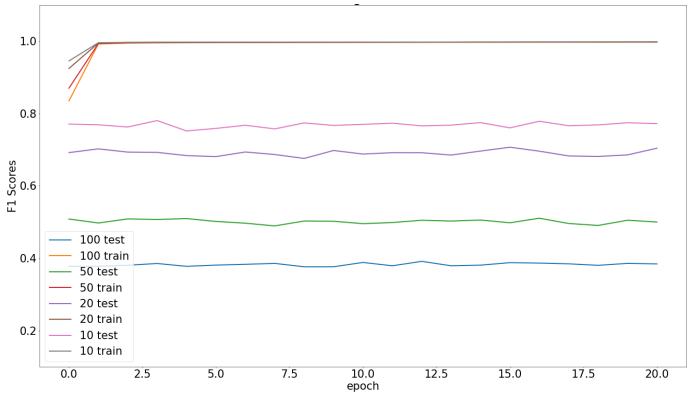



(d) F1 Scores, Unbalanced Classes

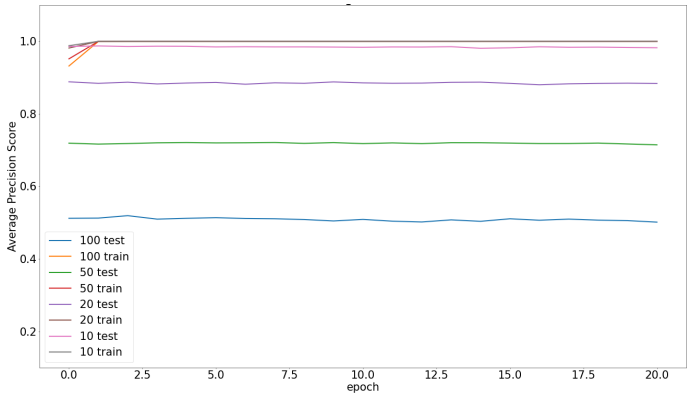
Figure 4.3: SEMTL-BERT Performance, Small Model



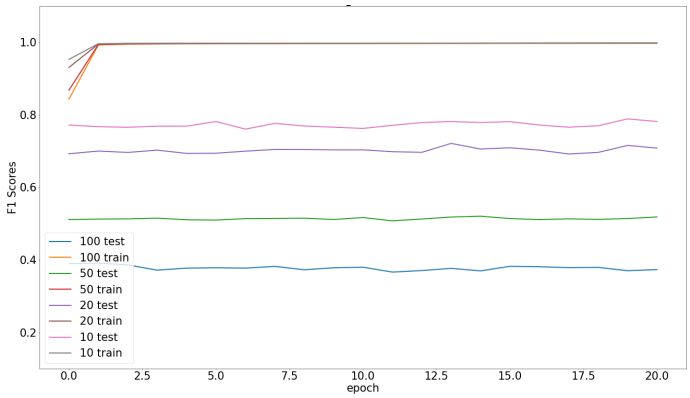

(a) Average Precision, Balanced Classes




(b) F1 Scores, Balanced Classes




(c) Average Precision, Unbalanced Classes

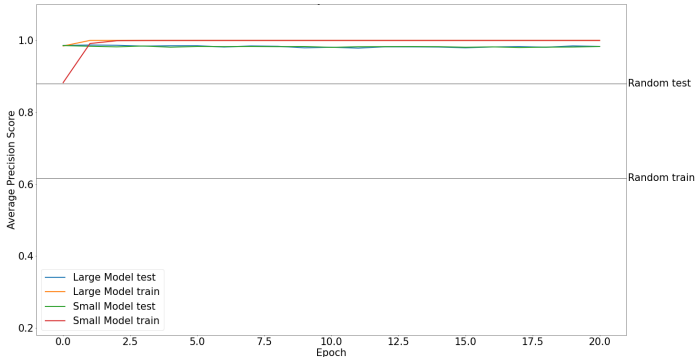



(d) F1 Scores, Unbalanced Classes

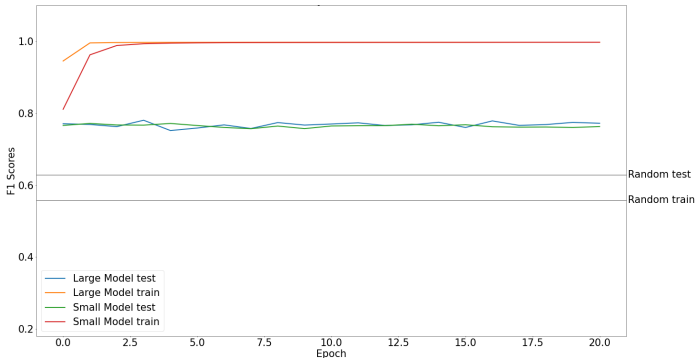
Figure 4.4: SEMTL-BERT Performance, Large Model

### 4.6.2 Random Baseline Comparison

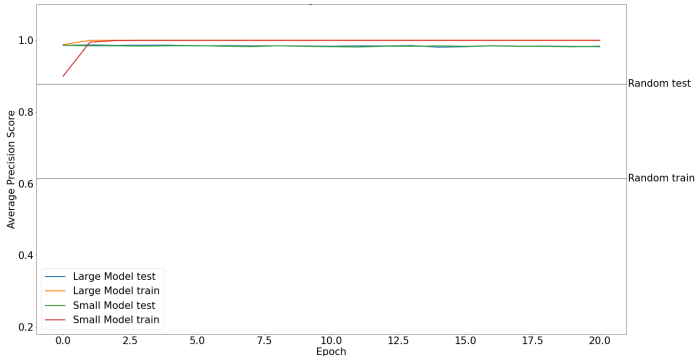
For each cluster and class balancing arrangement, a new, baseline model was created. Each model generates a random zero-one vector "guess" representing classifications for each cluster. These guesses were based on an evaluation of the class-wise training set probability for each cluster. Thus, given $K$ clusters, the random model produced a vector $R$ of length $K$ such each element $r_k \in R$ has the same probability of having the value of 1 as the cluster $C_k$ has class positivity rate in the training set. These random guesses were then averaged over 5 epochs, and compared to the performance of the model.



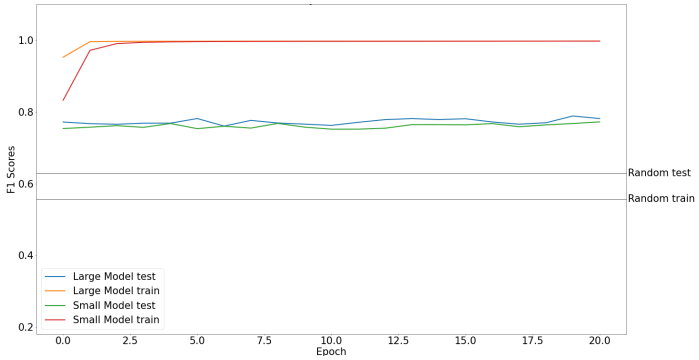

(a) Average Precision, Balanced Classes




(b) F1 Scores, Balanced Classes

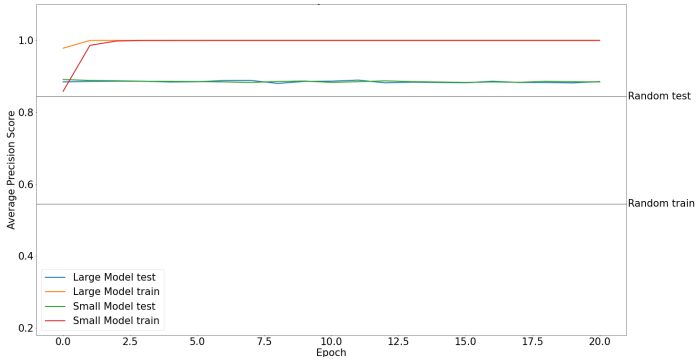



(c) Average Precision, Unbalanced Classes

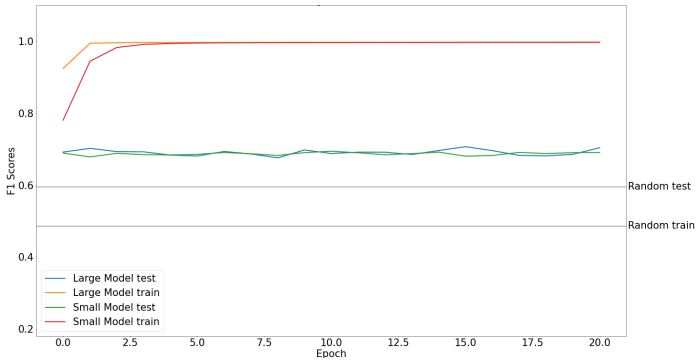



(d) F1 Scores, Unbalanced Classes

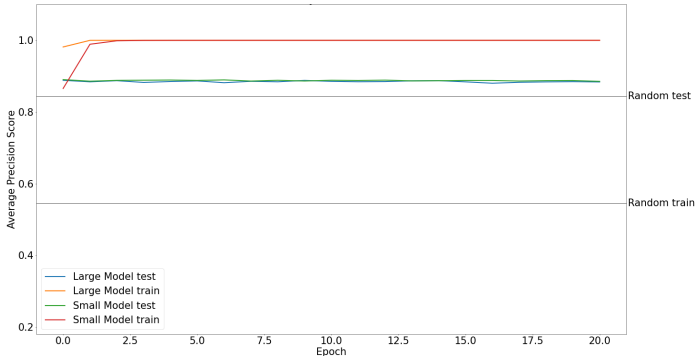
Figure 4.5: SEMTL-BERT vs. Baseline Model, 10 Clusters



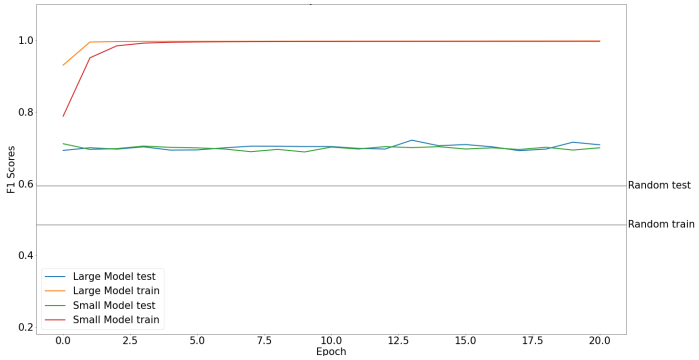

(a) Average Precision, Balanced Classes




(b) F1 Scores, Balanced Classes

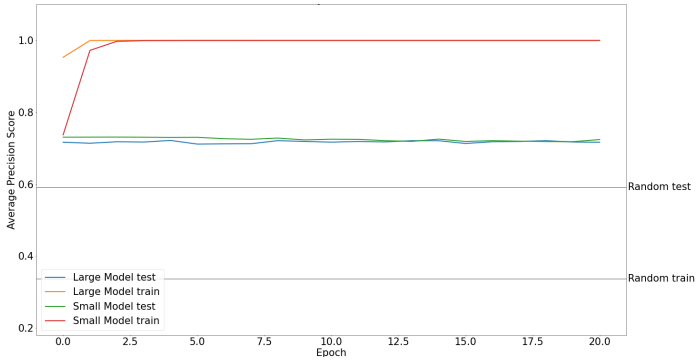



(c) Average Precision, Unbalanced Classes

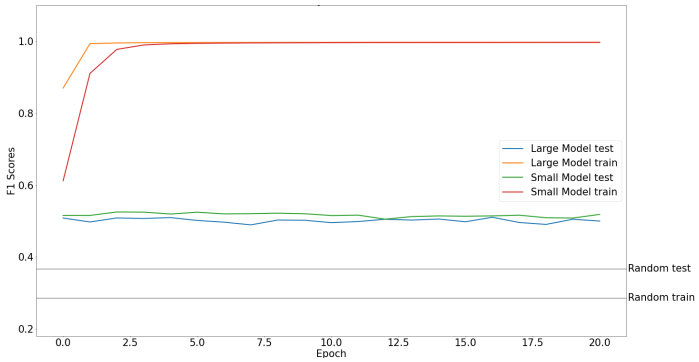



(d) F1 Scores, Unbalanced Classes

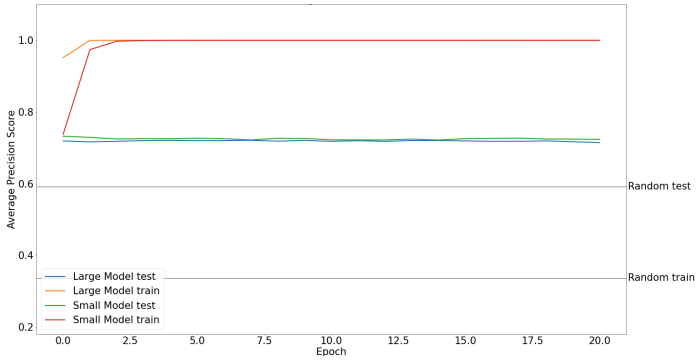
Figure 4.6: SEMTL-BERT vs. Baseline Model, 20 Clusters



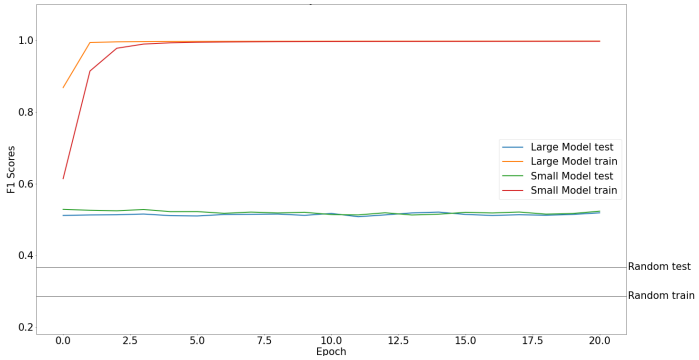

(a) Average Precision, Balanced Classes




(b) F1 Scores, Balanced Classes

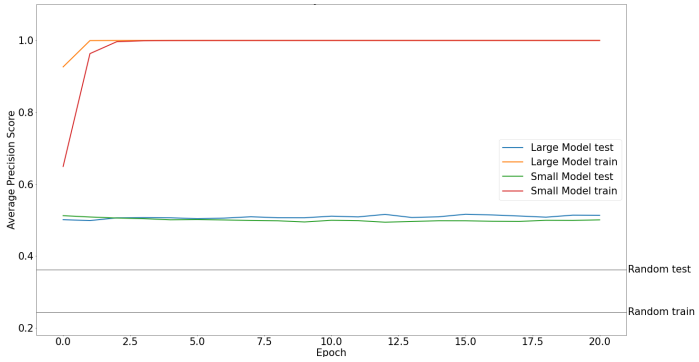



(c) Average Precision, Unbalanced Classes

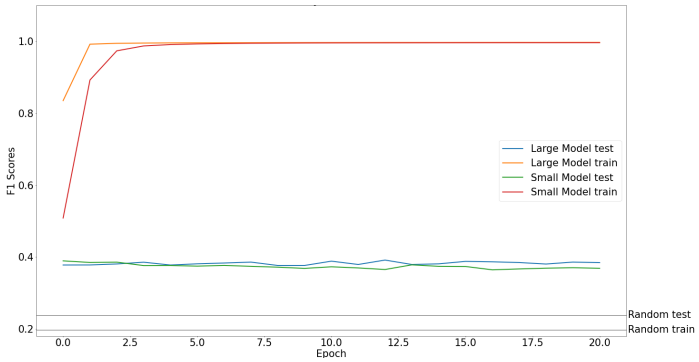



(d) F1 Scores, Unbalanced Classes

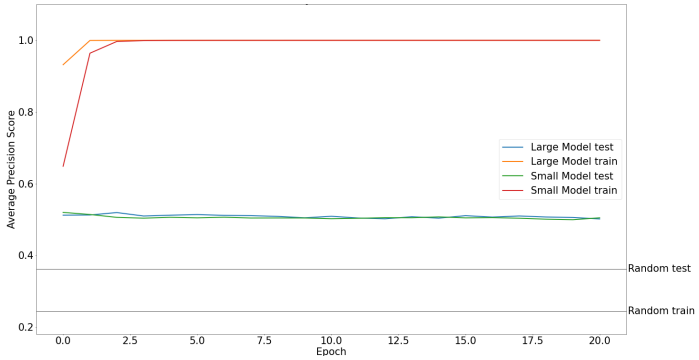
Figure 4.7: SEMTL-BERT vs. Baseline Model, 50 Clusters



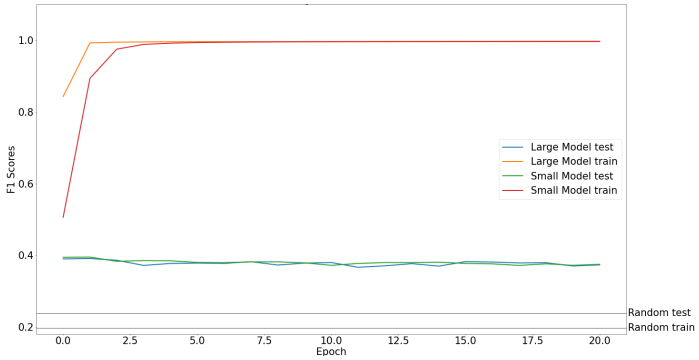

(a) Average Precision, Balanced Classes




(b) F1 Scores, Balanced Classes




(c) Average Precision, Unbalanced Classes




(d) F1 Scores, Unbalanced Classes

Figure 4.8: SEMTL-BERT vs. Baseline Model, 100 Clusters

### 4.6.3 SEMTL-BERT Prior Work Comparison

In this section, the performance metrics of SEMTL-BERT are compared with those of other methods of side effect prediction. Note that the methods presented here trained on different datasets, and the implementation of the performance metrics may have subtle differences. Both of these can have a large impact the net result.

| Liu Dataset [41] | | | | | |
|---|---|---|---|---|---|
| Model | # train | # test | ROC AUC | AP | Clusters |
| K-means | ≈ 666 | ≈ 166 | 0.896 | 0.405 | 4 |
| PAM | ≈ 666 | ≈ 166 | 0.895 | 0.400 | 4 |
| K-seeds | ≈ 666 | ≈ 166 | 0.895 | 0.402 | 4 |
| K-means | ≈ 666 | ≈ 166 | 0.895 | 0.408 | 5 |
| PAM | ≈ 666 | ≈ 166 | 0.895 | 0.400 | 5 |
| K-seeds | ≈ 666 | ≈ 166 | 0.895 | 0.405 | 5 |
| Mizutani Dataset [48] | | | | | |
| K-means | ≈ 526 | ≈ 132 | 0.891 | 0.410 | - |
| PAM | ≈ 526 | ≈ 132 | 0.889 | 0.406 | - |
| K-seeds | ≈ 526 | ≈ 132 | 0.889 | 0.406 | - |
| Zhang Dataset [83] | | | | | |
| K-means | 864 | 216 | 0.914 | 0.334 | - |
| PAM | 864 | 216 | 0.915 | 0.329 | - |
| K-seeds | 864 | 216 | 0.913 | 0.332 | - |
| DeepSide Dataset 1 [74] | | | | | |
| MLP (800 neurons) | 528 | 263 | 0.866 | 0.578 | 24 |
| MLP (2000 neurons) | 528 | 263 | 0.860 | 0.557 | 24 |
| ResMLP | 528 | 263 | 0.843 | 0.520 | 24 |
| DeepSide Dataset 2 [74] | | | | | |
| MLP | 410 | 205 | 0.849 | 0.577 | 24 |

Table 4.10: Performance Metrics in Comparable Side Effect Prediction Algorithms

#### Direct Comparison

The MLP models from DeepSide were adapted for use on the SEMTL-BERT database as a means of direct comparison. To achieve this, the SMILES strings from the original 1.7K samples were converted to their canonical representation, and then molecular

fingerprints were then generated using RDKit [34]. Since molecular fingerprints are unique for a molecule, neither the data augmentation nor the class balancing methods apply, and these steps were therefore skipped for DeepSide model experiments.
For SEMTL-BERT, the models are listed with their architecture size represented by "LM" for the large model, and "SM" for the small model, followed by "B" for balanced classes, and "NB" for unbalanced classes. SEMTL-BERT data used an 80-10-10 split for training, validation, and testing. Therefore the number of samples in the training set are listed as train set size + validation set size. The readings taken from SEMTL-BERT are all from the last epoch. The readings for DeepSide architectures are taken from the last epoch or the 150th, whichever is lower. This was done due to the general decay of some performance measures after this epoch, and is representative of performance of the DeepSide models.

| SEMTL-BERT Dataset | | | | | |
|---|---|---|---|---|---|
| Model | # train | # test | ROC AUC | AP | Clusters |
| LMB | 1413 + 177 | 177 | 0.680 | 0.983 | 10 |
| LMNB | 1413 + 177 | 177 | 0.670 | 0.982 | 10 |
| SMB | 1413 + 177 | 177 | 0.674 | 0.983 | 10 |
| SMNB | 1413 + 177 | 177 | 0.694 | 0.984 | 10 |
| MLP | 1413 + 177 | 177 | 0.491 | 0.974 | 10 |
| ResMLP | 1413 + 177 | 177 | 0.498 | 0.973 | 10 |
| LMB | 1413 + 177 | 177 | 0.594 | 0.886 | 20 |
| LMNB | 1413 + 177 | 177 | 0.608 | 0.884 | 20 |
| SMB | 1413 + 177 | 177 | 0.612 | 0.884 | 20 |
| SMNB | 1413 + 177 | 177 | 0.598 | 0.885 | 20 |
| MLP | 1413 + 177 | 177 | 0.634 | 0.887 | 20 |
| ResMLP | 1413 + 177 | 177 | 0.676 | 0.889 | 20 |
| LMB | 1413 + 177 | 177 | 0.575 | 0.717 | 50 |
| LMNB | 1413 + 177 | 177 | 0.567 | 0.715 | 50 |
| SMB | 1413 + 177 | 177 | 0.567 | 0.724 | 50 |
| SMNB | 1413 + 177 | 177 | 0.560 | 0.723 | 50 |
| MLP | 1413 + 177 | 177 | 0.570 | 0.726 | 50 |
| ResMLP | 1413 + 177 | 177 | 0.561 | 0.707 | 50 |
| LMB | 1413 + 177 | 177 | 0.591 | 0.513 | 100 |
| LMNB | 1413 + 177 | 177 | 0.569 | 0.501 | 100 |
| SMB | 1413 + 177 | 177 | 0.569 | 0.500 | 100 |
| SMNB | 1413 + 177 | 177 | 0.574 | 0.504 | 100 |
| MLP | 1413 + 177 | 177 | 0.589 | 0.521 | 100 |
| ResMLP | 1413 + 177 | 177 | 0.535 | 0.472 | 100 |

Table 4.11: Performance Metrics on SEMTL-BERT Data

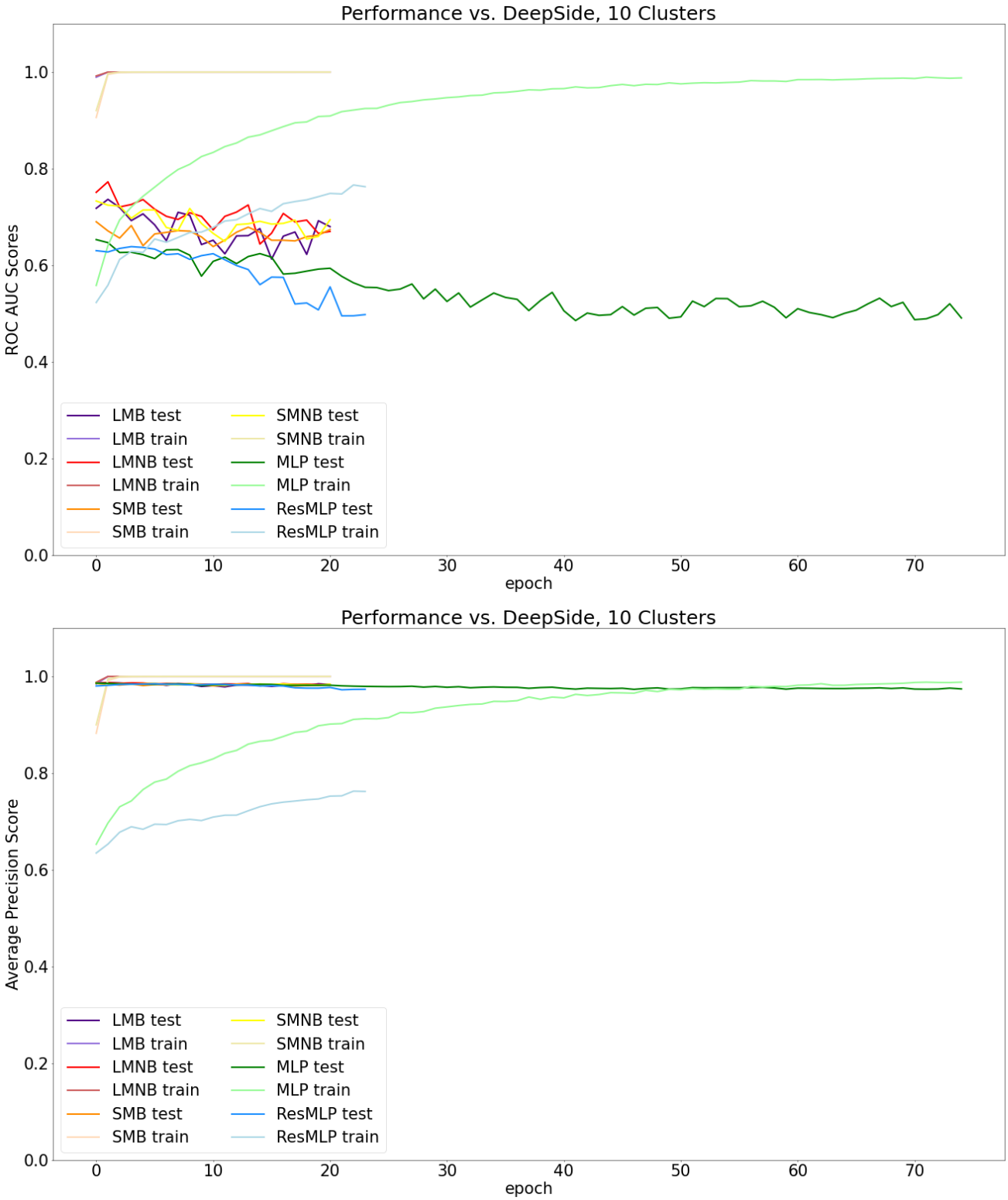



Figure 4.9: Performance of SEMTL-BERT and DeepSide Models, 10 Clusters

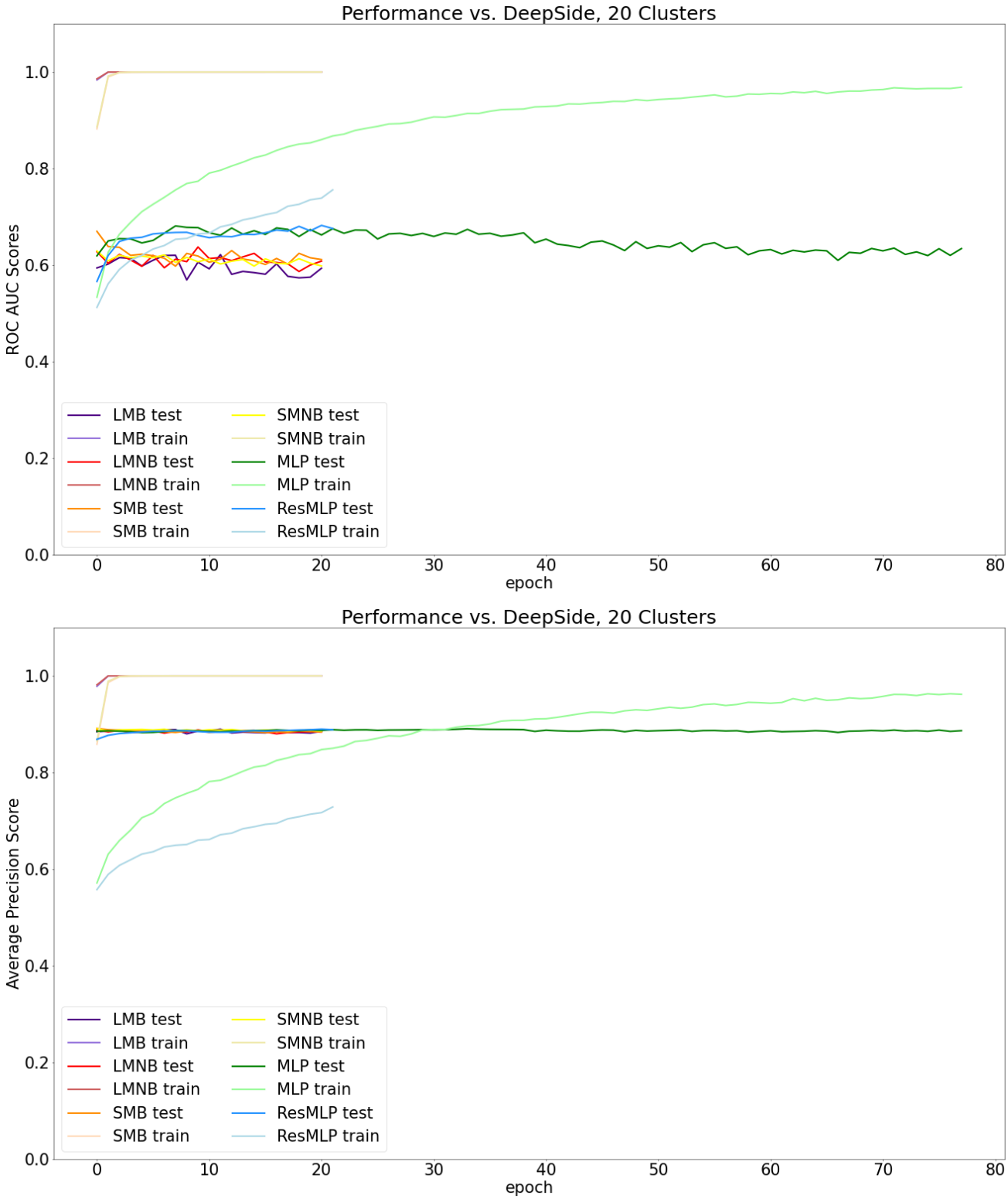



Figure 4.10: Performance of SEMTL-BERT and DeepSide Models, 20 Clusters

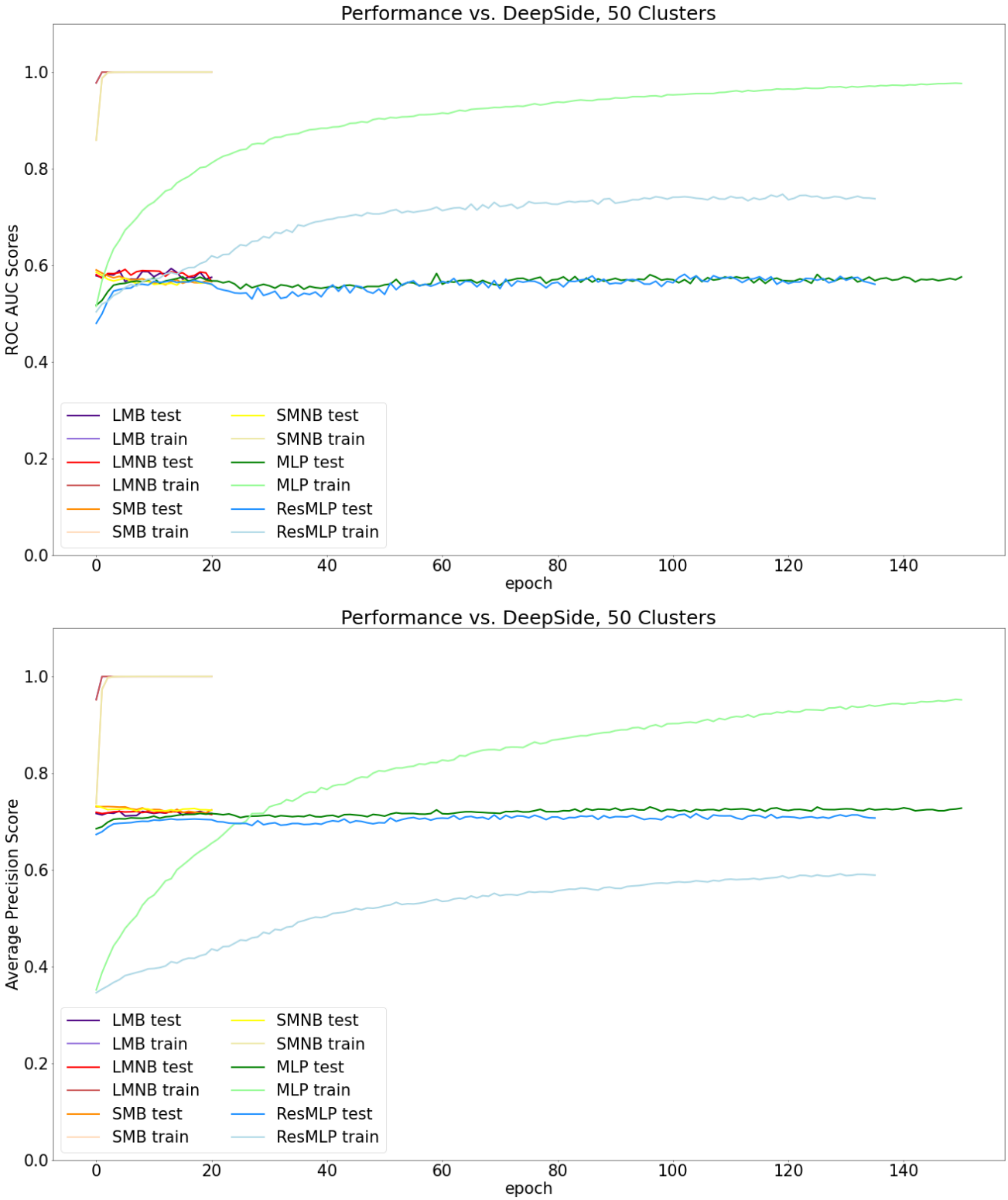



Figure 4.11: Performance of SEMTL-BERT and DeepSide Models, 50 Clusters

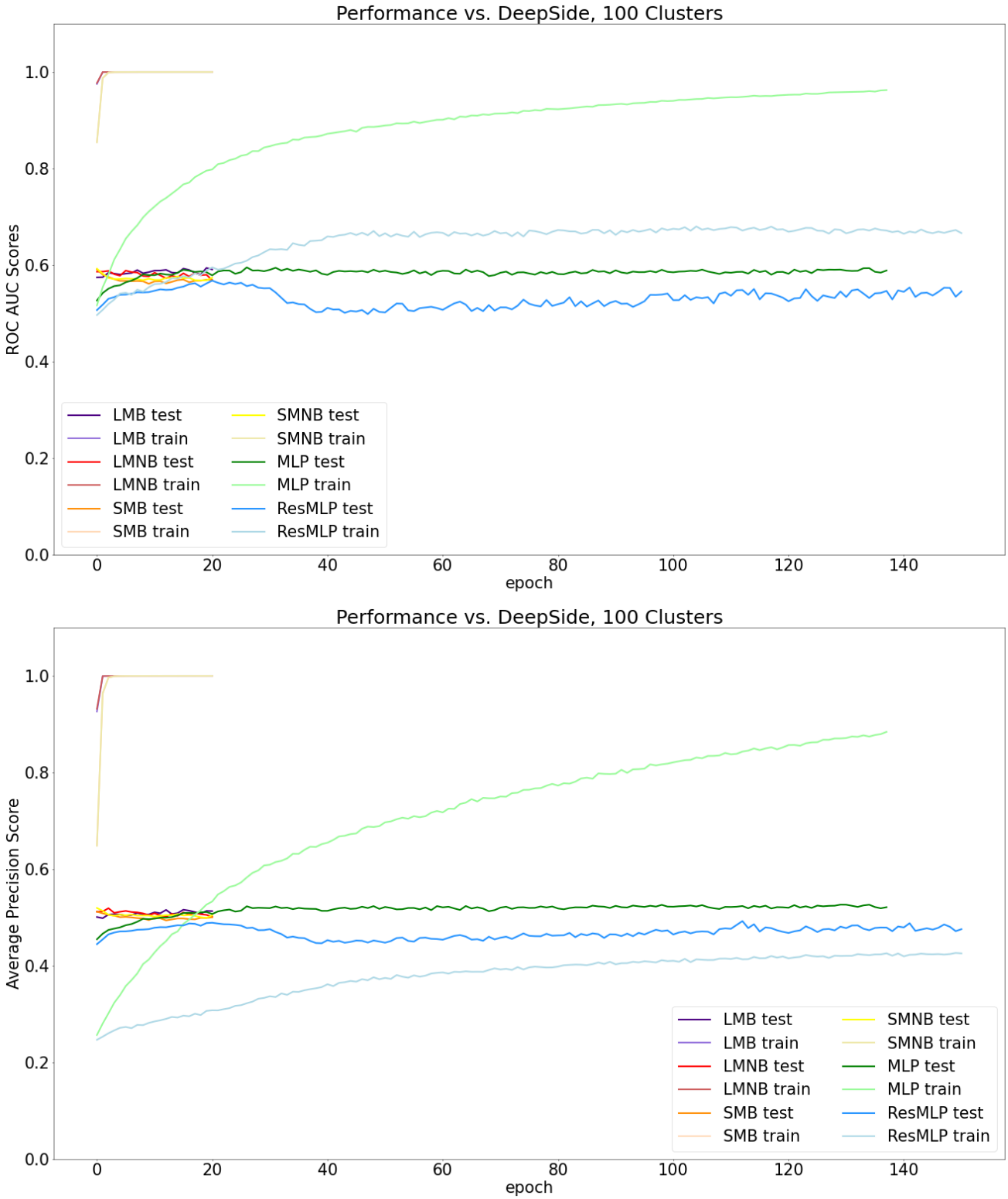


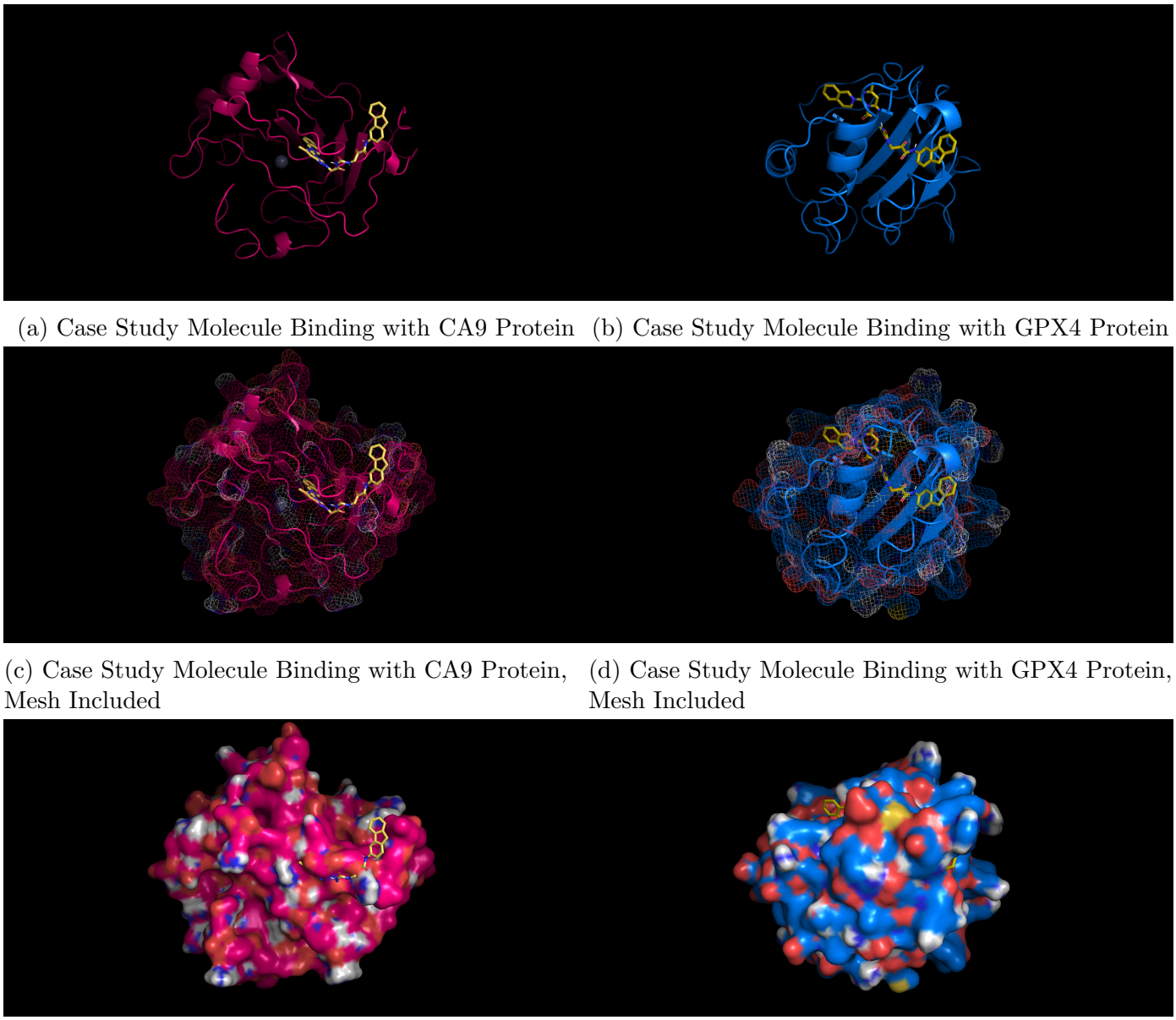

Figure 4.12: Performance of SEMTL-BERT and DeepSide Models, 100 Clusters

## 4.7 Case Study

This section presents the performance of molecules throughout the MOEA-DT and SEMTL-BERT system. Five, novel, high-performance molecules were selected from the first front of the final generation of MOEA-DT. Their structures and property scores were recorded, and a SEMTL-BERT model (large model, 50 clusters, balanced classes) was used to infer side effect predictions of those molecules. Using that, final side effect predictions were determined, via translation of the predicted clusters back into English language descriptions. Images of the molecule with the fewest side effect clusters virtually binding to the target proteins were captured using PyMol [66]. Below are the results of that study. Predicted side effects can be found in Appendix A

| (Structure Features Withheld) | | | | | |
|---|---|---|---|---|---|
| SMILES | Frags | logP | SAS | GPX4 | CA9 |
| NC(=O)NC(=O)c1ccccc1NC(=O)c1ccccc1 | 3 | 1.75 | 1.61 | -6.4 | -7.9 |
| O=C(NC(=O)c1cccc([N+](=O)[O-])c1)c1ccc(F)cc1 | 3 | 2.3 | 1.72 | -7.3 | -7.9 |
| Cc1cc2nn(CC(=O)NC(=O)NCC(=O)C(=O)Nc3ccc4c(c3)-c3ccccc3C4)c(=O)n2c(N2CCc3ccccc3C2)n1 | 4 | 2.4 | 3.2 | -10.4 | -11.3 |
| Cc1cc2nn(CC(=O)NCC(=O)NCc3ccc4c(c3)−c3ccccc3C4)c(=O)n2c(N2CCc3ccccc3C2)n1 | 3 | 2.77 | 2.93 | -11.1 | -8.5 |
| Cc1ccc(−c2cccc(CNC(=O)c3ccccc3)c2)cc1 | 3 | 4.59 | 1.47 | -7.1 | -8.2 |

Table 4.12: Case Study Molecule Property Scores



(a) Case Study Molecule Binding with CA9 Protein (b) Case Study Molecule Binding with GPX4 Protein

(c) Case Study Molecule Binding with CA9 Protein, Mesh Included (d) Case Study Molecule Binding with GPX4 Protein, Mesh Included

(e) Case Study Molecule Binding with CA9 Protein, Surface Included (f) Case Study Molecule Binding with GPX4 Protein, Surface Included

Figure 4.13: Case Study Molecule Binding with CA9 & GPX4 Proteins

# Chapter 5

# Conclusion

## 5.1 Discussion

In this research, MOEA-DT for molecular optimization was combined with SEMTL-BERT for side effect prediction to create a novel, flexible, drug design system. MOEA-DT samples a population of molecules from public datasets of drugs and drug-like compounds, then leverages evolutionary operations and multiple performance measures to optimize the population towards the creation of novel, high-performance compounds. These compounds were then compared to those of similar models as a means of benchmarking its performance. SEMTL-BERT pre-trains on a similar set, and then fintunes using side effect annotated data to enable the prediction of the side effects of novel molecules. The performance of SEMTL-BERT was measured across 16 experiments, measuring: three metrics, two architectures, two class balancing arrangements, and four clustering arrangements. Following which, the performance of each arrangement was compared against an educated guesses as well as other methods to establish predictive ability.

### 5.1.1 Performance

#### MOEA-DT

Given its high rank against deep learning models which are significantly more complex, the quality of the molecules produced by MOEA-DT is unintuitively high. Table 4.7 scores MOEA-DT amoung other methods in validity, novelty, and uniqueness. For MOEA-DT, the high validity score is not an indicator of performance, since every new generated molecule is tested for validity before being added to the population. Thus it maintains 100% validity inherently. Novelty however, is a surprising indicator of

high performance. It seems therefore that MOEA-DT adequately explores the search space to find improvements on the training data. The Uniqueness of MOEA-DT on the other hand, is the worst out of the non-reinforcement learning algorithms. This could be an indicator of early convergence, or it could be a non-significant difference. Further testing is necessary to determine for sure.

Table 4.8 outlines hypervolumes for the molecular optimization algorithms. Here too, MOEA-DT performs well. It produces the highest hypervolumes in generation 10 and the second highest in both other generations measured. Again, this could be an indicator of early convergence, but given the minute differences between its hypervolumes and those of AAE, it could very well be statistically insignificant.

Finally, table 4.9 shows the number of molecules produced by the models which satisfy the screening criteria. Again, MOEA-DT shows a close first place, or close second place next to AAE. Note that single-target runs for MOEA-DT and for AAE were not performed.

Overall, AAE and MOEA-DT seem to perform significantly better than the other methods tested. The differences in the performance between the two models is generally smaller than the gap between any other two models. Given that evolutionary computation approaches can produce unstable results, it seems uncertain which of the two models show overall higher performance, and more testing will be necessary to confirm any differences. Despite similar performance, the models are significantly different in design. In fact, MOEA-DT is the only model in this comparison which does not employ any kind of neural architecture. This is a significant distinction, because it means that MOEA-DT does not need training, nor does it need a complex system of weights and biases to optimize new molecules. In this respect, MOEA-DT offers similar or better performance at a fraction of the complexity and compute cost.

**SEMTL-BERT**

The performance of SEMTL-BERT on the test set as shown in Figure 4.3 and 4.4 is underwhelming. From the performance plots, it is clear that the improvement on test set performance is either none, or slightly negative across the epochs. Despite this, train set performance quickly converges to its maximum value, no matter the metric, class balancing, or number of clusters. This is a clear case of overfitting; and over the course of the experiments, no attempt to prevent it was successful. The most likely reason for overfitting is the inadequate amount of data. Given just over 1.7K samples to classify 11K classes is seemingly insufficient. Even when the number of classes was reduced significantly via clustering, the model is still tasked with interpreting

complex chemical interactions with very few samples. Another potential cause is the difficulty of inferring side effects based on SMILES strings. This is an environment in which the difference in molecular structure is not necessarily reflected in a significant difference in side effect, or SMILES notation. Another conclusion that can be drawn from SEMTL-BERT's performance is that most of its predictive ability is due to the pre-training step, rather than fine-tuning. This becomes apparent when considering how quickly SEMTL-BERT converges when compared to the DeepSide models. It seems therefore that SEMTL-BERT may be capable of so-called "few shot learning" in which a comprehensive pre-training step is all that is necessary for task-specific performance given few training examples.

The number of clusters seems to have a significant and consistent impact on performance. Regardless of all other parameters, the performance increases as the number of clusters decreases. This result is fairly intuitive, since a multi-label classification problem gets inherently easier as the number of classes decreases. The problem is, as the number of clusters decreases, the specificity of the model's prediction also decreases. This is because a constant number of classes are compressed into fewer clusters. Thus, the end user is left with a trade-off between specificity and accuracy of side effect predictions.

In terms of model architecture, the larger model's performance shown in Figure 4.4 shows clearly earlier convergence than the smaller model's performance in Figure 4.3. This too, is quite intuitive. A larger model, provided with more pre-training data provides a better ability to transfer that knowledge more quickly. What is surprising is that this does not translate into any increased performance on the testing set; both models end with minute differences in performance scores.

When compared against a random guess, however, SEMTL-BERT shows better performance. Regardless of the experiment or metric, SEMTL-BERT's performance on the test set is greater than that of the random model on either the train or test sets. This shows that SEMTL-BERT achieves at least some sample-dependant learning. In other words, SEMTL-BERT appears to be able to predict more than the class-wise distributions, instead managing more accurate predictions on a sample-by-sample basis, to at least some degree.

When it comes to indirect comparisons against other methods using different datasets, performance indicators are unclear. As stated in Chapter 4, there are slight differences in the implementation of of the metric calculation, which can affect the readings. Despite this, there are some key takeaways from the performance figures in Tables 4.10 and 4.11. Firstly, when looking at the performance of DrugClust models over

several different datasets, there appears to be an data-dependant inverse relationship between ROC and AP scores. Since the models are the same, and the clustering method is the same, this change must be due to the change in data. If this is the case, it may also be the case that the selection of data can have the same impact on models in general. With respect to performances of the models overall it's worth noting that ROC AUC metrics can be inflated due to unbalanced data. This is because a model which only classifies positive or only classifies negative can attain reasonably high scores in this metric, provided that the data is in its favour. Thus, ROC AUC fails to punish misclassifications of the minority class.
For the direct comparison between SEMTL-BERT and DeepSide models, neither one shows significantly better performance. Both models show overfitting on the data, and very little performance increase on the first few epochs, followed by performance decay over time. Note that SEMTL-BERT converges much quicker to the training set than DeepSide's MLP models. This is likely due to the pre-training step employed. As stated before, this is an example of "few-shot learning", and SEMTL-BERT shows the ability to converge on the data within one epoch, whereas the MLP models take more than 100, depending on the number of clusters.

#### Predicted Side Effects

Following the case study experiment, the produced side effect clusters were then decoded back into plain English descriptions. A list of the predicted side effects from the SEMTL-BERT model's inference on the molecules' structure can be found in Appendix A. It represents a prediction by SEMTL-BERT that the molecules will have one or many of the side effects within each block. From analysis of the side effects, one can see clear semantic similarities in the descriptions.

## 5.2 Limitations

### 5.2.1 Availability

Across both methods tested in this paper, a recurring problem is the lack of adequate hardware and data. Newer, more powerful machines are needed to run larger models and perform faster experiments. Data availability is a problem particularly with SEMTL-BERT, as the number of classes currently exceeds the number of samples.

### 5.2.2 MOEA-DT

As discussed previously, MOEAs have limitations for maximum number of objectives, and are also known to be unstable, necessitating large populations and many generations. Non-dominated sorting and protein ligand docking are also very computationally expensive. Combined, this causes MOEA-DT to require a large compute time investment. Improvements can be made in performing a sum of ranks, instead of non-dominated sorting, as well as reducing I/O operations during protein-ligand docking.

#### Multi-Objective vs. Many-Objective Problems

Multi-objective problems are those with 2-3 objectives. Any more objectives constitute a many-objective problem, for which different algorithms are needed. In discussion of the difficulty of finding non-dominated individuals in a population, Deb et al. [10] note the following difficulties as the number of objectives increases:

1. The proportion of the population which is non-dominated becomes larger, which slows the search process.
2. Evaluation of diversity becomes more computationally expensive.
3. The *trade-off* surface is more difficult to represent.
4. Performance metrics are more difficult to compute.
5. Visualization becomes more difficult.

Furthermore, since the MOEA-DT system is target-agnostic, the NSGA-III algorithm could also be used to extend MOEA performance to many more protein targets.

#### Fragmentation

A limitation of the ZINC dataset used is that due to the way in which the molecule is fragmented, many of the molecules contained in it consist of 1-2 fragments. The low granularity helps with chemical validity, but also significantly narrows the search space. Furthermore, fragments of this size make it difficult to add or subtract fragments, causing near-uniform fragment quantities in the produced molecules.

### 5.2.3 SEMTL-BERT

The main limitation of performance in SEMTL-BERT is over-fitting. Data availability is likely a large factor in this problem. More side effect annotated data is necessary for further testing.

#### Clustering & Representation of the Problem

While representing the multi-label classification problem using all 11K original classes is intractable using current hardware, data, and the MTL-Bert framework, the clustering approach fundamentally limits the predictive ability of the system. Namely, as the number of clusters decreases, the specificity of SEMTL-BERT's predictions decreases. This is because there is no way of determining which side effect(s) is actually predicted out of a given cluster; only that one or many of them are predicted.
Furthermore, clustering by semantic embedding is not necessarily a valid approach. Instead, clusters can be produced by leveraging domain knowledge to group side effects together based on their pharmacokinetic similarities, rather than side effect descriptions, as in DeepSide [74]. This would create a more direct correlation between drug structure and side effect cluster. An example of the limitation of clustering by semantic embedding would be a side effect cluster containing both *grade 3 vomiting* and *grades 1-4 dry skin*, which may not be associated conditions w.r.t. bodily functions, but are associated in a semantic embedding because they are both conditions of a specific grade. Thus, a semantic embedding can produce counter intuitive results which may not be useful for establishing the relationship between structure and side effect.

#### Hardware

Early experiments for SEMTL-BERT were delayed due to limited memory. In fact, the initial motivation behind side effect clustering itself was the inability to create a model which predicted each class individually. With modern ML techniques requiring more and more powerful machines, hardware availability will become more of a problem with time.

### 5.2.4 Statistical Significance

Given the amount of time available, it was not feasible to run the experiments a sufficient number of times to perform statistical analysis. Especially in the case of

MOEA-DT and its comparisons to other models, performance could be more accurately established using the average of several runs, rather than individual runs. Performing several runs is a necessary and important step, since findings require the addition degrees of confidence in order to determine the reliability and reproducibility of the results. This is a fundamental feature of scientific research.

## 5.3 Future Work

### 5.3.1 Reproducibility

Statistical analysis of the results must be performed to determine degree of confidence in the results. To this end, five runs of each experiment will be performed, and the mean and standard deviation will be recorded. Following this, p-values will be determined by the corresponding confidence intervals. The results of this analysis will then be attached to their respective sections.

### 5.3.2 Optimizations

Given that on current hardware, MOEA-DT takes 40 days to run, and SEMTL-BERT takes 1 day (depending on number of clusters), improvements to both software and hardware will be performed to enable the quicker completion of the necessary experiments. On the hardware side, the experiments will be performed on a new server with 8 NVIDIA A100 GPUs and 112 CPU cores. On the software side, MOEA-DT will be improved with I/O optimizations, and the removal of CPU limitations. Options for improving SEMTL-BERT will include the reduction of number of experiments and the removal of data augmentation in the pre-processing stage.
The efficiency of protein-ligand docking through the use of newer docking programs like QVina-w [51], which enables search over the entire protein, without a pre-specified binding site. Additionally, VinaGPU2.0 [15] contains GPU enabled versions of QVina-w, Qvina2, and Autodock Vina, which will significantly speed up the runs necessary to show reproducibility of the methods.

### 5.3.3 MOEA-DT

#### Surrogate Models

A surrogate model is a separate ML application which estimates sample performance. Surrogate models can be used to replace exact calculation tools, which are computationally expensive. Thus overall compute time could be reduced.

#### Drug Targets

Since MOEA-DT is target-agnostic, the same application can be used with more and different drug targets to achieve similar results for the development of drugs for different diseases.

#### Many-Objective Optimization Problems

The NSGA-II [9] algorithm used may be inadequate for a many-optimization problem like the one presented here. Better performance may be achieved using the newer NSGA-III [10] algorithm, which was designed for the purpose of handling many-objective problems of up to 15 objectives.

### 5.3.4 New Fragmentation Approaches & Molecular Representations

Alternate fragment approaches may show increased performance in molecular diversity. One such approach is to simply reduce the size of the fragments, thus having more fragments per molecule. This would likely increase diversity of molecules during evolutionary operations, but may have a negative impact on chemical validity. Another approach is the so called *hybrid representation* in which special fragments are introduced which are mutated on the character level. This is an interesting idea for which the impact is not well known. Hypothetically, it would have a much larger search space, without as much of a negative impact on compute cost or chemical validity as a purely character-based approach would.
Another approach entirely would be opting for a graph-based approach, like that in JTVAE. This would be significantly more computationally complex, but it may offer improved performance.

### 5.3.5 SEMTL-BERT

#### NLP-Enhanced Side Effect Predictions

Side Effect Prediction achieved through an NLP task, rather than a multi-label classification one. Here, the task would be to predict the English-language descriptions of the side effects of the molecule. Such a model would not need task-specific architecture, but it would be difficult to compare its performance to the existing SEMTL-BERT.

#### Further Class Balancing

Given that the performance of SEMTL-BERT is equal when the classes are balanced and unbalanced, it is possible that the method employed did not sufficiently balance the classes to affect performance. Another, more effective class balancing algorithm could be used instead, to explore SEMTL-BERT's behaviour in different circumstances.

#### Byte Pair Encoding & SMILES Pair Encoding

Byte pair encoding (BPE) is a method of tokenization, which empirically determines a set of tokens on the basis of subword frequency in the dataset. BPE achieves this by first determining the frequency with which candidate tokens appear within a dataset, and then iteratively creating new tokens out of the most common pairings of tokens. After many iterations, the most common subwords are represented as a single token, whereas rare subwords are broken down into smaller subword tokens. X, Li & D, Fourches [40] developed a version of the BPE algorithm, which was designed to encode SMILES strings specifically. Their algorithm combines the function of BPE with the addition of encoding certain chemically meaningful substructures. This algorithm could be used to tokenize the data in a chemically meaningful way, thus offering a potential means of performance improvement.

#### Improved Clustering

Semantic embedding followed by spectral clustering is far from the only means of clustering the data. Clustering medical information is a research problem in itself. Resources like the ADReCS database can be used to cluster the side effects by using medical domain knowledge, and thereby potentially improve performance.

#### K-Fold Splitting

Due to the unbalanced nature of SEMTL-BERT's data, as well as the need to partition the data while maintaining at least one positive example of each class, splitting the data randomly via a k-fold configuration was not possible. Manual partition of the dataset will be explored as an option going forward.

### 5.3.6 Benchmarking

To improve further analysis and determine relative performance, both systems should be compared to a wider variety of similar applications. For MOEA-DT, the performance should be measured against MOSES [59]. For SEMTL-BERT, more direct experiments will need to be performed to fully compare against the methods covered in Table 4.10. This will involve getting SEMTL-BERT to operate on the datasets of these other methods, and with their metric calculations.

# Bibliography


[1] Abouchekeir S., Vu A., Mukaidaisi M., Grantham K., Tchagang A., Li Y., Adversarial deep evolutionary learning for drug design, Biosystems, Volume 222, Pages 104790, 2022, ISSN 0303-2647, https://doi.org/10.1016/j.biosystems.2022.104790.

[2] Alhossary A., Stephanus D. H., Mu Y., Kwoh C., Fast, accurate, and reliable molecular docking with QuickVina 2, Bioinformatics, Volume 31, Issue 13, Pages 2214–2216, July 2015, DOI:10.1093/bioinformatics/btv082

[3] Beltagy I., Lo K., Cohan A., SciBERT: A pretrained language model for scientific text, Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Pages 3615-3620, Association for Computational Linguistics, 2019, DOI: 10.18653/v1/D19-1371

[4] Berman, H. M. et al., The protein data bank. Nucleic Acids Research, Volume 28, Issue 1, Pages 235–242, 2000

[5] Bickerton G. R. J.,Paolini G. V. , Besnard J., et al., Quantifying the chemical beauty of drugs. Nature Chemistry, Volume 4, Pages 90–98, 2012.

[6] Cai M.C., Xu Q., Pan Y.J., Pan W., Ji N., Li Y.B., Jin H.J., Liu K., Ji Z.L., ADReCS: an ontology database for aiding standardization and hierarchical classification of adverse drug reaction terms. Nucleic Acids Research, Volume 43, Pages D907–D913. https://doi.org/10.1093/nar/gku1066

[7] Coello Coello C.A., González Brambila S., Figueroa Gambo, J., et al., Evolutionary multiobjective optimization: open research areas and some challenges lying ahead. Complex Intelligent Systems, Volume 6, Pages 221–236, 2020. https://doi.org/10.1007/s40747-019-0113-4

[8] Costa P. R., Acencio M. L., Lemke N., A machine learning approach for genome-wide prediction of morbid and druggable human genes based on systems-level data. BMC Genomics, Volume 11, Pages S9–S9, 2010

[9] Deb K., Pratap A., Agarwal S., Meyarivan T., A fast and elitist multiobjective genetic algorithm: NSGA-II, in IEEE Transactions on Evolutionary Computation, Volume 6, Number 2, Pages 182-197, April 2002, doi: 10.1109/4235.996017.

[10] Deb K., Jain H., An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, Part I: solving problems with box constraints, in IEEE Transactions on Evolutionary Computation, Volume 18, Number 4, Pages 577-601, Aug. 2014, doi: 10.1109/TEVC.2013.2281535.

[11] Degen J., Wegscheid-Gerlach C., Zaliani A., Rarey M., On the art of compiling and using 'drug-like' chemical fragment spaces. ChemMedChem, Volume 3, Issue 10, Pages 1503-1507, 2008.

[12] Devlin J., Chang M.W., Lee K., Toutanova K., BERT: Pre-training of deep bidirectional transformers for language understanding, Proceedings of NAACL-HLT 2019, pages 4171–4186 Minneapolis, Minnesota, June 2 - June 7, 20192019

[13] Dey S., Luo H., Fokoue A., et al., Predicting adverse drug reactions through interpretable deep learning framework. BMC Bioinformatics, Volume 19. Issue Suppl 21, Article Number 476, 2018. https://doi.org/10.1186/s12859-018-2544-0

[14] Dimitri G. M., Lió P., DrugClust: A machine learning approach for drugs side effects prediction, Computational Biology and Chemistry, Volume 68, Pages 204-210, 2017, ISSN 1476-9271, https://doi.org/10.1016/j.compbiolchem.2017.03.008. (https://www.sciencedirect.com/science/article/pii/S1476927116302195)

[15] Ding J.,Tang S., Zheming M., Lingyue W., Qinqin H., Haifeng H., Ming L., Jiansheng W., Vina-GPU 2.0: Further accelerating AutoDock Vina and its derivatives with graphics processing units, Journal of Chemical Information and Modeling, Volume 63, Issue 7, Pages 1982-1998, 2023, 10.1021/acs.jcim.2c01504

[16] Eiben A., Smith J., From evolutionary computation to the evolution of things. Nature, Volume 521, Pages 476–482, 2015, https://doi.org/10.1038/nature14544

[17] Einarson T. R., Drug-related hospital admissions. Annals of Pharmacotherapy, Volume 27, Issues 7-8, Pages 832-840, 1993.

[18] Elton D. C., Boukouvalas Z., Fuge M. D., Chung P. W., Deep learning for molecular design—a review of the state of the art, Molecular Systems Design & Engineering, Pages 828—849, 2019, 10.1039/c9me00039a4

[19] Engel T., Gasteiger J., Chemoinformatics: Basic Concepts and Methods. Wiley VCH, 2018.

[20] Erlanson, D. A., Introduction to fragment-based drug discovery. Fragment-Based Drug Discovery and X-Ray Crystallography, Volume 317, Pages 1-32, 2012

[21] Koruza, K., Lafumat, B., Nyblom, M., Mahon, B. P., Knecht, W., McKenna, R., Fisher, S. Z., Acta Crystallographica Section D, Volume 75, Part 10, Pages 895-903, 2019

[22] Friedmann Angeli J., Schneider M., Proneth B., et al., Inactivation of the ferroptosis regulator Gpx4 triggers acute renal failure in mice. Nature Cell Biology, Volume 16, Pages 1180–1191, 2014, https://doi.org/10.1038/ncb3064

[23] Ghose A., Viswanadhan V., Wendoloski J., A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. a qualitative and quantitative characterization of known drug databases. Journal of Combinatorial Chemistry, Volume 1, Issue 1, , Pages 55–68, 1999

[24] Goldberg, D. E. Genetic algorithms in search, optimization, and machine learning. Chapter 5: Advanced Operators and Techniques in Genetic Search. Addison-Wesley Longman Publishing Co., Inc., 1989.

[25] Grantham K., Mukaidaisi M., Ooi H. K. , Ghaemi M. S., Tchagang A., Li Y., Deep evolutionary learning for molecular design, in IEEE Computational Intelligence Magazine, Volume 17, Number 2, Pages 14-28, May 2022, doi: 10.1109/MCI.2022.3155308.

[26] He K., Zhang X., Ren S., Sun J., Deep residual learning for image recognition, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Pages 770–778, 2016.

[27] Irwin J.J., Shoichet B.K., ZINC - a free database of commercially available compounds for virtual screening. Journal of Chemical Information and Modeling, Volume 45, Pages 177-82. doi: 10.1021/ci049714+. PMID: 15667143; PMCID: PMC1360656.

[28] Jin, W., et al., Junction tree variational autoencoder for molecular graph generation. in International Conference on Machine Learning, Pages 2323–2332, 2018.

[29] Kadurin, A., et al., druGAN: An advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico. Molecular Pharmaceutics, Volume 14, Pages 3098–3104, 2017.

[30] Kim, J., Park, S., Min, D., Kim, W., Comprehensive survey of recent drug discovery using deep learning. International Journal of Molecular Sciences, Volume 22, Issue 18, Article Number 9983, https://doi.org/10.3390/ijms22189983

[31] Kim S., et al., PubChem substance and compound databases, Nucleic Acids Resaerch, Volume 44, Number D1, Pages D1202–D1213, 2015

[32] Kola I., Landis J., Can the pharmaceutical industry reduce attrition rates?. Nature Reviews Drug Discovery, Volume 3, Issue 8, Pages 711-716, 2004

[33] Kuhn M., Campillos M., Letunic I., Jensen L.J., Bork P., A side effect resource to capture phenotypic effects of drugs. Molecular Systems Biology, Volume 6, Article Number 343, 2010.

[34] G. Landrum, RDKit: Open-source cheminformatics, 2006, Available: http://www.rdkit.org

[35] Lasser K.E., Allen P.D., Woolhandler S.J., Himmelstein D.U., Wolfe S.M., Bor D.H., Timing of new black box warnings and withdrawals for prescription medications. Journal of the American Medical Association. Volume 287, Issue 17, Pages 2215–2220, 2002 doi:10.1001/jama.287.17.2215

[36] Laurie A.T., Jackson R.M., Methods for the prediction of protein-ligand binding sites for structure-based drug design and virtual ligand screening. Current Protein & Peptide Science, Volume 7, Issue 5, Pages 395–406, 2006 doi: 10.2174/138920306778559386. PMID: 17073692.

[37] Lazarou J., Pomeranz B.H., Corey P.N., Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. Journal of the American Medical Association, Volume 279, Issue 15, Pages 1200–1205, 1998 https://doi.org/10.1001/jama.279.15.1200

[38] Leung M,K,, Xiong H.Y., Lee L.J., Frey B.J., Deep learning of the tissue-regulated splicing code. Bioinformatics (Oxford, England), Volume 30, Issue 12, Pages i121–i129, 2014, https://doi.org/10.1093/bioinformatics/btu277

[39] Lewell, X.Q., Judd, D., Watson, S., Hann, M., RECAP: Retrosynthetic combinatorial analysis procedure: A powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry, Journal of Chemical Information and Computer Sciences, American Chemical Society, Volume 38, Issue 3, Pages 511-522, 1998, https://doi.org/10.1021/ci970429i

[40] Li X., Fourches D., SMILES pair encoding: a data-driven substructure tokenization algorithm for deep learning, Journal of Chemical Information and Modeling, Volume 61, Pages 1560-1569, 2021, DOI: 10.1021/acs.jcim.0c01127

[41] Liu M., Wu Y., Chen Y., Sun J., Zhao Z., Chen X.W., Matheny M.E., Xu H., Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs. Journal of the American Medical Informatics Association, Volume 9, Issue e1, Pages e28–e35. https://doi.org/10.1136/amiajnl-2011-000699

[42] Lipinski C. A., Lombardo F., Dominy B. W., and Feeney P. J., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews, Volume 46, Issues 1-3, Pages 3–26. https://doi.org/10.1016/s0169-409x(00)00129-0

[43] von Luxburg U., A tutorial on spectral clustering. Statistics and Computing, Volume 17, Pages 395–416, 2007. https://doi.org/10.1007/s11222-007-9033-z

[44] Macalino S.J.Y., Gosu V., Hong S., et al., Role of computer-aided drug design in modern drug discovery. Archives of Pharmacal Research, Volume 38, Issue 9, Pages 1686–1701, https://doi.org/10.1007/s12272-015-0640-5

[45] Makhzani A., et al., Adversarial autoencoders. In International Conference on Learning Representations, 2016.

[46] McDermott M., Wang J., Zhao W.N., Sheridan S., Szolovits P., Kohane I., Haggarty S., Perlis R., Deep learning benchmarks on L1000 gene expression data,IEEE/ACM Transactions on Computational Biology and Bioinformatics, Volume 17, Pages 1846-1857, 2022, 10.1109/TCBB.2019.291006

[47] Mendez D., Gaulton A., Bento A.P., Chambers J., De Veij M., Félix E., Magariños M.P., Mosquera J.F., Mutowo P., Nowotka M., Gordillo-Marañón M., Hunter F., Junco L., Mugumbate G., Rodriguez-Lopez M., Atkinson F., Bosc N., Radoux C.J., Segura-Cabrera A., Hersey A., Leach A.R. ChEMBL: towards direct

deposition of bioassay data. Nucleic Acids Research, Volume 47, Issue D1, Pages D930-D940, 2019

[48] Mizutani S., Pauwels E., Stove V., Goto S., Yamanishi Y., Relating drug-protein interaction network with drug side effects Bioinformatics, Oxford University Press, Volume 28, Issue 18, Pages i522-i528, 2012.

[49] Mostaghim, S., Schmeck, H. Distance Based Ranking in Many-Objective Particle Swarm Optimization. in Parallel Problem Solving from Nature, 2008, https://doi.org/10.1007/978-3-540-87700-4_75

[50] Mukaidaisi M., Vu A., Grantham K., Tchagang A., Li Y., Multi-objective drug design based on graph-fragment molecular representation and deep evolutionary learning. Frontiers in Pharmacology, Volume 13, 2022, doi: 10.3389/fphar.2022.920747

[51] Hassan N., Alhossary A., Mu Y., Chee-Keong K., Protein-L;igand blind docking Using QuickVina-W with inter-process spatio-temporal integration. Nature Scientific Reports, Volume 7, Issue 1, 2017. DOI:10.1038/s41598-017-15571-7

[52] O'Boyle N. M., Banck M., James C. A., Morley C., Vandermeersch T., Hutchison G. R., Open Babel: An open chemical toolbox, Journal of Cheminformatics, Volume 3, Issue 1, Article Number 33, 2011.

[53] Oduguwa A., Tiwari A., Roy R., Bessant C., An overview of soft computing techniques used in the drug discovery process, in Applied Soft Computing Technologies: The Challenge of Complexity. Editors A. Abraham, B. de Baets, M. Köppen, and B. Nickolay, Advances in Soft Computing, Pages 465–480, 2006

[54] Olivecrona M., Blaschke T., Engkvist O., et al., Molecular de-novo design through deep reinforcement learning. Journal of Cheminformatics, Volume 9, Article Number 48, 2017. https://doi.org/10.1186/s13321-017-0235-x

[55] Pastorekova S., Ratcliffe P.J., Pastorek J., Molecular mechanisms of carbonic anhydrase IX-mediated pH regulation under hypoxia. BJU international, Volume 101 Issue Suppl 4, Pages 8–15. https://doi.org/10.1111 j.1464-410X.2008.07642.xdoi:10.1111/j.1464-410X.2008.07642.x. PMID 18430116. S2CID 8780292.

[56] Pauwels E., Stoven V., Yamanishi Y., Predicting drug side-effect profiles: a chemical fragment-based approach BMC Bioinformatics, Volume 12, Issue 1, Article Number 169, 2011.

[57] Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay E., Scikit-learn: machine learning in Python, Journal of Machine Learning Research, Volume 12, Pages 2825-2830, 2011

[58] Podda M., Bacciu D., Micheli A. A deep generative model for fragment-based molecule generation, in International Conference on Artificial Intelligence and Statistics, Pages 2240–2250, 2020.

[59] Polykovskiy D., Zhebrak A. Sanchez-Lengeling B., Golovanov S., Tatanov O., Belyaev S., Kurbanov R., Artamonov A., Aladinskiy V., Veselov M., Kadurin A., Johansson S., Chen H., Nikolenko S., Aspuru-Guzik A., Zhavoronkov A., Molecular ets (MOSES): a benchmarking platform for molecular generation models. Frontiers in Pharmacology. Volume 11, 2020. DOI:10.3389/fphar.2020.565644. ISSN:1663-9812

[60] Ramsundar B., Liu B., Wu Z. Verras A., Tudor M., Sheridan R., Pande V., Is multitask deep learning practical for pharma?, Journal of Chemical Information and Modeling, Volume 57, Pages 2068-2076, 2017, 10.1021/acs.jcim.7b00146

[61] Ramsundar B, Eastman P., Walters P., Pande V., Leswing K., Wu Z., Deep learning for the life sciences, O'Reilly Media, 2019

[62] Rashid S., Shah S., Bar-Joseph Z., Pandya R.D., Variational autoencoder for unmasking tumor heterogeneity from single cell genomic data. Bioinformatics. Volume 37, Issue 11, Pages 1535–1543, June 2021

[63] Roughgarden T., Valiant G., CS168: The Modern Algorithmic Toolbox Lectures #11: Spectral graph theory, I, Stanford University, May 8, 2023.

[64] Sharma H., Zerbe N., Klempert I., Hellwich O., Hufnagl P., Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology. Computerized Medical Imaging and Graphics : the Official Journal of the Computerized Medical Imaging Society, Volume 61, Pages 2–13. https://doi.org/10.1016/j.compmedimag.2017.06.001

[65] Shraddha S., Naganna S. A review on K-means data clustering spproach. International Journal of Information & Computation Technology. Volume 4, Number 17, Pages 1847-1860, 2014, ISSN 0974-2239

[66] The PyMOL molecular graphics system, Version 1.8, Schrödinger, LLC., Schrödinger, LLC,2010-08-19 17:29:55, The PyMOL Molecular Graphics System, Version 1.8, 2015

[67] Sousa T., Correia J., Pereira V., Rocha M., Combining multi-objective evolutionary algorithms with deep generative models towards focused molecular design. In Applications of Evolutionary Computation: 24th International Conference, Springer International Publishing, Proceedings 24, Pages 81-96, April 7–9, 2021.

[68] Span P., Bussink J., Manders P., Beex L., Sweep C., Carbonic anhydrase-9 expression levels and prognosis in human breast cancer: association with treatment outcome. British journal of cancer, Volume 89, Issue 2, Pages 271–276. https://doi.org/10.1038/sj.bjc.6601122

[69] Ståhl, N., et al., Deep reinforcement learning for multiparameter optimization in de novo drug design. Journal of Chemical Information and Modeling, Volume 59, Issue 7, Pages 3166–3176, 2019.

[70] Subramanian A., et al., A next generation connectivity map: L1000 platform and the first 1,000,000 profiles, Cell, Volume 171, Number 6, Pages 1437–1452, 2017.

[71] Szymański P., Markowicz M., Mikiciuk-Olasik E., Adaptation of high-throughput screening in drug discovery-toxicological screening tests. International Journal of Molecular Sciences, Volume 13, Issue 1, Pages 427–452, 2012, https://doi.org/10.3390/ijms13010427

[72] Tasaki S., Suzuki K., Kassai Y., Takeshita M., Murota A., Kondo Y., Ando T., Nakayama Y., Okuzono Y., Takiguchi M., Kurisu R., Miyazaki T., Yoshimoto K., Yasuoka H., Yamaoka K., Morita R., Yoshimura A., Toyoshiba H., Takeuchi T., Multi-omics monitoring of drug response in rheumatoid arthritis in pursuit of molecular remission, Nature Communications, Volume 9, Issue 2755, 2018

[73] Trott O., Olson A.J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithread-

ing. Journal of Computational Chemistry, Volume 31, Issue 2, Pages 455–461. https://doi.org/10.1002/jcc.21334

[74] Uner O.C., Cinbis R.G., A. E. C., DeepSide: a deep learning framework for drug side effect prediction, IEEE/ACM Transactions on Computational Biology and Bioinformatics, Volume 20, Issue 1, Pages 330–339, 2023, https://doi.org/10.1109/TCBB.2022.3141103

[75] Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A., Kaiser Ł., Polosukhin I., 2017. Attention is all you need. in Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17).

[76] Wang W., Wang Y., Zhao H., Sciabola S., A transformer-based generative model for de Novo molecular design. arXiv.org. 2022, Retrieved January 3, 2023, from https://doi.org/10.48550/arXiv.2210.08749

[77] Weininger, D., SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. Journal of Chemical Information and Modeling, Volume 28, Pages 31–36. 1988. doi:10.1021/ci00057a005

[78] Weininger, D., SMILES a language for molecules and reactions. in Handbook of Chemoinformatics, 2005. https://doi.org/10.1002/9783527618279.ch5

[79] Wigh D.S., Goodman J.M., Lapkin A.A., A review of molecular representation in the age of machine learning. Wiley Interdisciplinary Reviews: Computational Molecular Science, Volume 12, Issue 5, Pages e1603.

[80] Wishart D.S., Feunang Y.D., Guo A.C., Lo E.J., Marcu A., Grant J.R., Sajed T., Johnson D., Li C., Sayeeda Z., Assempour N., Iynkkaran I., Liu Y., Maciejewski A., Gale N., Wilson A., Chin L., Cummings R., Le D., Pon A., Knox C., Wilson M., DrugBank 5.0: a major update to the DrugBank database for 2018, Nucleic Acids Research, Volume 46, Issue D1, Pages D1074–D1082, 2017, doi: 10.1093/nar/gkx1037.

[81] Vamathevan J., Clark D., Czodrowski P., et al., Applications of machine learning in drug discovery and development. Nature Reviews Drug Discovery, Volume 18, Pages 463–477, 2019. https://doi.org/10.1038/s41573-019-0024-5

[82] Yang W.S., SriRamaratnam R., Welsch M., Shimada K., Skouta R., Viswanathan V., Cheah J., Clemons P., Shamji A., Clish C., Brown L., Girotti A., Cornish

Vasanthi., Schreiber S., Stockwell B., Regulation of ferroptotic cancer cell death by GPX4, Cell, Volume 156, Issues 1–2, 2014, Pages 317-331, ISSN 0092-8674, https://doi.org/10.1016/j.cell.2013.12.010.

[83] Zhang W., Liu F., Luo L., Zhang J., Predicting drug side effects by multi-label learning and ensemble learning BMC Bioinformatics, Volume 16, Issue 1, Article Number 365, 2015

[84] Xiao-Chen Z., Cheng-Kun W., Jia-Cai Y., Xiang-Xiang Z., Can-Qun Y., Ai-Ping L., Ting-Jun H., Dong-Sheng C., Pushing the boundaries of molecular property prediction for drug discovery with multitask learning BERT enhanced by SMILES enumeration, Research, Volume 2022, Article ID 0004, 2022 DOI:10.34133/research.0004

[85] Niu Z., Zhong G., Yu H., A review on the attention mechanism of deep learning, Neurocomputing, Volume 452, Pages 48-62, 2021, ISSN: 0925-2312

[86] Zuo S., Yu J., Pan H. et al., Novel insights on targeting ferroptosis in cancer therapy. Biomarker Research, Volume 8, Article Number 50, 2020. https://doi.org/10.1186/s40364-020-00229-w

# Appendix A

# Case Study Side Effect Predictions

## A.1 Molecule 1

SMILES: NC(=O)NC(=O)c1ccccc1NC(=O)c1ccccc1

catalepsy; snoring; sneezing; negative cardiac inotropic effect; bullous conditions; pathological gambling disorder; minimal change disease; morbilliform eruptions; gross bleeding/perforation; bullous reactions; seizure-like activity; heavy head/tired; no light reflex; any noncerebral bleeding; unusual dreams; generalized itching; reactive hypotension; unusual and sometimes aggressive behaviour; any bleeding reaction; antiperistaltic effects; pure cell aplasia; heavy headedness; axillary sweating; all grades edema; all grades nausea; all grades abdominal pain; all grades constipation; all grades vomiting; all grades dyspnea; all grades cough; all grades pleural effusion; all grades decreased appetite; all grades pneumonia; all grades decreased albumin; all grades increased creatinine; all grades elevations in serum alanine aminotransferase (alt); all grades increased ast; all grades decreased sodium; all grades increased potassium; all grades increased amylase; all grades decreased lymphocytes; all grades decreased leukocytes; all grades increased lipase; all grades diarrhoea; all grades pyrexia; all grades peripheral neuropathy; all grades dizziness; all grades upper respiratory tract infection; all grades cataract; all grades vision blurred; all grades decreased platelets; all grades decreased neutrophils; all grades decreased phosphate; all grades decreased calcium; all grades increased bun; all grades decreased potassium; all grades decreased magnesium; all grades increased alt; all grades increased bilirubin; all grades urinary tract infection; all grades taste disorder; all grades mental status changes; all grades hypotension; all grades hemorrhage; all grades increased blood glucose; all grades elevated creatine kinase; all grades peripheral sensory neuropathy; all grades headache; all grades myalgia; all grades arthralgia; all grades back pain; all grades

pain in extremity; all grades rash; all grades lymphopenia; all grades hypocalcemia; all grades hypophosphatemia; all grades influenza; all grades hypertension; all grades visual impairment; all grades weight increased; all grades gastroesophageal reflux disease; all grades dyspepsia; all grades dry skin; all grades alopecia; all grades upper respiratory tract infections; all grades rhinorrhea; all grades ocular toxicity; all grades prolonged qtc interval; all grades red blood cell count decreased; all grades increased alanine aminotransferase; all grades increased aspartate aminotransferase; all grades stomatitis; all grades paronychia; all grades pruritus; all grades increased triglycerides; all grades increased creatine kinase; all grades increased serum uric acid; all grades calcium decreased; all grades increased serum creatinine; all grades increased cholesterol; all grades increased serum lipase; all grades decreased serum calcium; all grades increases in total serum bilirubin; all grades hypothyroidism; all grades thrombocytopenia; all grades hypomagnesemia; all grades insomnia; all grades hypoglycemia; all grades peripheral edema; all grades increased blood creatinine; all grades proteinuria; any grade neutropenia; any grade anemia; any grade thrombocytopenia; any grade diarrhea; any grade constipation; any grade nausea; any grade fatigue; any grade pyrexia; any grade peripheral edema; any grade asthenia; any grade upper respiratory tract infection; any grade bronchitis; any grade viral upper respiratory tract infection; any grade pneumonia; any grade hypokalemia; any grade hypocalcemia; any grade hyperglycemia; any grade muscle spasms; any grade back pain; any grade peripheral neuropathy; any grade insomnia; any grade dyspnea; any grade cough; any grade rash; any grade hypertension; any grade thromboembolic events; any grade vomiting; any grade headache; any grade respiratory tract infection; all grades infusion-related reaction; all grades respiratory tract infection; all grades bronchitis; all grades chills; all grades muscle spasms; all grades chest pain; all grades epistaxis; any grade abdominal pain; any grade edema; any grade decreased appetite; any grade arthralgia; any grade musculoskeletal pain; any grade mucositis; any grade transaminase elevation; any grade lower respiratory tract infection; any grade lymphocytosis; any grade leukopenia; any grade lymphopenia; any grade increased alt; any grade ast increased; any grade lipase increased; any grade hypophosphatemia; any grade increased alp; any grade elevated serum amylase; any grade hyponatremia; any grade hyperkalemia; any grade hypoalbuminemia; any grade increased creatinine; any grade decreased weight; any grade decreased phosphate; any grade increased amylase; any grade alkaline phosphatase increased; all grades diarrhea/colitis; all grades dry mouth; all grades acne; all grades photosensitivity reaction; all grades nail abnormalities; all grades prolonged qt interval; all grades corneal abnormalities; all grades

blurred vision; all grades depression

pyuria; pseudomonas infections; shock, cardiogenic; hematemesis; hematuria; polycythemia; vasculitis, leukocytoclastic, cutaneous; hemoptysis; shock, hemorrhagic; serum sickness; neutropenic infection; neutropenic sepsis; nodal arrhythmia; detachment of retinal pigment epithelium; cycloplegia; intra-abdominal haemorrhage; serum sickness-like reaction; cholangiocarcinoma; hepatitis infectious; cholangitis; immunosuppression; pyelonephritis; rhinitis; empyema; candiduria; systemic candida; enterobacter sepsis; transplant dysfunction; incision site haemorrhage; arteriovenous fistula thrombosis; arteriovenous fistula site haemorrhage; arterial rupture; pleural fibrosis; arteritis; visceral arterial ischaemia; surgery; bacteremia; candidiasis; euthyroid goitre; esophagitis; postoperative pain; cmv colitis; embolization; mediastinal disorders; metastases; pleural disorders; septic shock; hematology; postoperative sedation; postoperative complications; polyomavirus infections; aneurysms; postoperative wound infection; streptococcal infections; intracranial pressure; blood in stool; hematochezia; esophageal perforation; hematospermia; viral tonsillitis; neovascularization; viral pneumonia; systemic fungal infections; cmv; coagulation management; neutropenic fever; surgical intervention; surgical procedures; viral hepatitis b; postoperative inflammatory response; arterial hypotension; ischemic lesions; postoperative pancreatitis; hemothorax; microscopic hematuria; vascular disorders; neutropenic enterocolitis; postoperative infections; intraoperative blood loss; systemic infection; necrotizing pancreatitis; hematologic changes; blood platelet transfusion; postoperative anaemia; streptococcal pneumonia; arterial insufficiency; viral respiratory tract infection; vascular insufficiency; fibula fracture; arterial occlusion; hypotension drug-induced; residual paralysis; sepsis and septic shock; vasculitis, cerebral; abdominal distress; residual urine volume; rectal hemorrhage; serum calcium increased; embolus; circulatory depression; persistent anesthesia; infusion vein complication; perforated duodenal ulcer; total protein lab abnormality; microscopic pyuria; gross hematuria; perforated uterus; required platelet transfusions; reddened sclera; bleeding/hematoma; loss of mental acuity; intrahepatic cholestatic jaundice; blood glucose increased; circulatory; visceral disease; advanced hiv disease; abdominal pain (localized); intraoperative muscle movement; treatment site infection; febrile response; perioperative hyperkalemia; postoperative euphoria; cmv infection/syndrome; serum iron decreased; varicella zoster/herpes zoster; serum lipase ¿2.0xuln; total germinal aplasia; urine incontinence; variable deceleration; abdominal pain/distress; cmv tissue

invasive disease; pleural pain; shock-like syndrome with hepatic involvement; rhinitis tests; infection mycotic; hematuria with mesna; neutropic fever; nodal; pleural effusion with pericarditis; blood insulin increased; pleuritic chest pain; blood chloride decreased; blood bilirubin decreased; permanent neurologic damage; blood alkaline phosphatase decreased; embolism and infarction; infused vein infection; streptococcal bacteremia; perforations; systemic acidosis; blood parathyroid hormone increased; blood cell transfusion; intraarticular hemorrhage; hematuria aggravated; arterial ischemia; intravascular volume contraction; pleural puncture; macroscopic hematuria; bleeding events; hemodynamic disturbances; hematopoietic supression; significant hypotensive episodes; orthostasis; angioedema-like swelling; disseminated herpes simplex; new onset diabetes; new onset seizure; abdominal/gastrointestinal pain; blood estrogen decreased; blood gonadotrophin increased; incision site complications; intraabdominal fluid collection; blood creatinine decreased; blood creatinine; perforated gastric ulcer; blood cholesterol abnormal; blood triglycerides abnormal; vascular aneurysm; hematologic disease and disorders; vascular access site thrombosis; vasculitis, systemic; shock, hypovolemic; local anaesthesia therapy; transplanted organ rejection; blood glucose fluctuations; advanced bone age; abdominal pain (grade $\geq$3); bleeding requiring hospitalization; biopsy-confirmed hepatic necrosis; flare up; new onset ventricular fibrillation; treatment interrupted; postoperative agitation in children; new-onset primary melanoma; hematoma, subdural

bronchopneumonia; pneumonia, pneumococcal; tracheitis; ventricular flutter; pulmonary infarction; pneumatosis cystoides intestinalis; pulmonary eosinophilia; respiratory paralysis; respiratory distress syndrome, adult; pneumothorax; cough; ventricular hypokinesia; grey syndrome neonatal; endocarditis; chest discomfort; ventricular asystole; throat irritation; fever neonatal; reye's syndrome; bronchospasm; bronchiolitis; pneumonia respiratory syncytial viral; laryngitis; pulmonary sepsis; pneumonia; bronchitis; pneumonia fungal; respiratory moniliasis; bronchopulmonary aspergillosis; lung infection pseudomonal; pneumonia legionella; escherichia sepsis; pulmonary toxicity; pulmonary oil microembolism; epidural lipomatosis; chondrolysis; spondylitis; breath holding; laryngospasm; respiratory tract congestion; respiratory tract irritation; respiration abnormal; bronchial irritation; sputum increased; sputum discoloured; productive cough; cough decreased; respiratory depth increased; pulmonary arteriopathy; pulmonary microemboli; pulmonary alveolar haemorrhage; lung consolidation; pneumonitis; wheezing; bronchospasm paradoxical; bronchial obstruction;

bronchiectasis; erythrodermic psoriasis; bronchoconstriction; fever; bronchial airway hyperreactivity; nose edema; breathing abnormalities; pulmonary embolism; respiratory distress syndrome; pulmonary fibrosis; chest pain; respiratory insufficiency; pleural effusions; airway obstruction; bronchioloalveolar carcinoma; bronchial asthma; respiratory failure; respiratory distress; respiratory depression; influenza-like illness; respiratory disorders; respiratory diseases; pulmonary arterial hypertension (pah); pulmonary obstruction; pulmonary edemas; right heart failure; respiratory symptoms; pulmonary veno occlusive disease; pulmonary mass; lung disorder; pulmonary infections; ventilatory depression; bronchopleural fistula; sputum; pulmonary hypertension (ph); cough, acute; respiratory alkalosis; respiratory system abnormalities; pulmonary inflammation; respiratory distress syndrome, infant; pulmonary alveolar proteinosis (pap); ventricular failure, right; pulmonary haemorrhage; lung, carcinoma; respiratory tract hemorrhage; pulmonary congestion; lung cancers; lung function changes; chest disorders; pulmonary diseases, obstructive; respiratory muscle paralysis; ventricular tachycardia (vt); airway complication of anesthesia; lung edema; pulmonary infiltrates; pulmonary events; chest tightness; chest congestion; breath odor; pulmonary interstitial infiltrates; dyspnea aggravated; throat discomfort; lung edema disorder; respiratory inadeqaucy; sweat gland; bronchospasm aggravated; pulmonary problem; nervousness/irritability; throat inflammation; chest pain/discomfort; nose pain; escherichia bacteremia; bronchial constriction; throbbing headache; pulmonary hypersensitivity reactions; allergic-type respiratory symptoms; chest pain/pressure/angina; respiratory flu; pulmonary infiltration with eosinophilia; expiratory wheezing; bronchial wheezing; fever and headache; chest wall rigidity; chest wall pain; respiratory acidosis during weaning; fever or influenza-like illness; ecg change consistent with myocardial ischemia; vagolysis; respiratory and cns depression; chest symptoms; throat/neck symptoms; respiratory side effects; perlèche; bronchiolar constriction; fetal anticonvulsant syndrome; pulmonary hypotension; meniere's syndrome; pulmonary-clinical; pulmonary allergy; chest flushing; innapropiate secretion of antidiuretic hormone; allergic pericarditis; delirium hallucinations; pulmonary vascular occlusion; cough suppression; cough experience post-inhalation; intramuscular hemorrhage; throat pressure; asthma-related death; respiratory impairment; respiratory symptom-related adverse reactions; swellin of lower legs; respiratory wheezing; respiratory arrest (disorder); pulmonary changes; pulmonary symptoms; mechanical irritation at application site; mechanical cardiac dysfunction; bronchial anastomotic dehiscence; throat infections; erythroleukemia; chest mass; thoracic spine pain; asthmatic type dyspnea; respiratory rate increased; fluid overload/retention; respiratory

acidosis/alkalosis; respiratory arrest/failure; pulmonary infarct; endocarditis caused by staphylococcus aureus; pneumonia (grade ≥3); pulmonary reactions; pleural effusion (grade ≥3); respiratory function impaired; bronchorrhea; pulmonary hypertension, primary; pulmonary fibrosis interstitial; lung hemorrhage; dyspepsia/heartburn; respiratory rate decreased; expiratory reserve volume decreased; chest infections

increased tendency to bruise; increased insulin requirement; altered visual depth perception; change of bowel habit; lower gastrointestinal haemorrhage; decreased appetite; increased appetite; decreased activity; hypertrophy; epididymitis; decreased vibratory sense; decreased interest; increased upper airway secretion; increased viscosity of bronchial secretion; hypoxia; impaired glucose tolerance; elevated temperature; change in bone mineral density; altered mental status; decreased left ventricular function; impaired balance; lower respiratory infection; increased secretion of gastrin; inflammation of mouth; loss of appetite; altered cardiac rhythm; increased inr; increased sweating; decreased hematocrit; decreased bone mineral density; increased thirst; increased alkaline phosphatase; increased heart rate; increased cough; changes in blood pressure; increased serum bicarbonate; increased lactate dehydrogenase; increased lipase; increased ast; increased salivation; decreased orgasm; increased alp; increased blood pressure; decreased corneal sensitivity; increased bleeding; increased spontaneous bleeding; increased dreaming; increased hair loss; increased discomfort; increased drug level; decreased hemoglobin; decreased pupillary reflex; change in amount of cervical secretion; changes in heart rate; decrease in vital capacity; decrease in total lung volume; loss of perineal sensation; loss of sphincter control; change in cervical erosion and secretion; reduced tolerance to carbohydrates; decreased deep tendon reflex; decreased carnitine concentrations; decreased spermatozoa motility; changes in av conduction; increased venous streaking; increased serum iron; effects on blood clotting mechanisms; increased bleeding time; decreased systolic blood pressure; increased irritability; altered taste; increased eosinophils; decreased intraocular pressure/ocular hypotony; diminished concentration; increased ventricular arrhythmias/pvcs; changes in the oral mucous membrane; changes in the laryngeal mucous membrane; changes in the pharyngeal mucous membrane; changes in menstrual flow; changes in cervical erosion; changes in cervical secretions; decreased serum calcium; loss of coordination; increased muscle spasticity; decreased cognition; increased body temperature; increased glucocorticoids; increased phosphorus; increased magnesium; decreased magnesium; increased haptoglobin; increased neutrophils; decreased high

occult blood; increased calcium; increased chloride; decreased chloride; increased bands; increased lymphocytes; decreased haptoglobin; decreased lymphocytes; increased rbcs; increased globulin; increased total protein; increased serum albumin; increased iron; impaired motor skills; increased serum sodium; decreased serum testosterone; excessive appetite; changes in nail pigmentation; increased depth of respiration; altered color perception; decreased leukocytes; increased conjugated bilirubin; increased blood urea; decreased night vision; increased alkaline phosphate; loss of deep tendon reflexes; decrease of visual acuity; increased gamma-glutamyltransferase; increased urine glucose; decreased polymorphonuclear neutrophils; increased urine ph; increased polymorphonuclear neutrophils; change in blood glucose levels; increased skin pigmentation; loss of nails; increased peristalsis; increased bronchial secretions; increased thephylline levels; increased deep tendon reflexes; decreased resistance to infection; alterations in skin pigmentation; increased platelet count; decreased secretions from mucus membranes; decreased sexual activity; increased sensitivity to alcohol; loss of muscle mass; increased creatinine excretion; increased angina; increased libido in women; increased menstrual bleeding; alterations in uterine contractions during labor; increased pigmentation of the iris; reduced vitamin d; reduced vitamin e; reduced vitamin k; altered menses; elevated lh; increased npn; decreased serum magnesium; abnormal healing; decreased potency; increased sexual urges; exudate on skin; increase in mean arterial pressure; decrease in mean arterial pressure; increased systolic and diastolic blood pressure; increased susceptibility to infection; alterations in temperature; changed visual acuity; decreased glucocortiocoid levels; decreased salivation; increased reticulocyte counts; decreased reticulocyte counts; reduced oral intake; decreased breathing sounds; increased drug effect; decreases in serum potassium concentrations; increased excretion of heavy metals; increased muscle tone and cramping; increased risk of infection; increased risk of opportunistic infection; changes in renal function; changes in phosphorus; changes in bicarbonate; changes in calcium; changes in hemoglobin; changes in hematocrit; changes in wbcs; changes in ldh; changes in total serum bilirubin; changes in serum proteins; changes in sgot; changes in cholesterol; changes in creatinine clearance; increased urinary glucose; increased ocular fibrin; elevated cpk-mb level; change in sleep habits; increased vaginal discharge; increased nitrogen; decrease in memory; increased risk of cv thrombotic event; increased risk of stroke; increased risk of myocardial infarction; increased redness; decreased gastrointestinal motility and ileus; decreased heat tolerance; increased fibrinolysis; increased albumin; increased creatine; increase in menstrual flow; decreased menstrual flow; decreased position sense; increased tracheobronchial secretions; in-

creased gastric and intestinal secretions; increased tremor; increased bradykinesia; increased apraxia; decreases in hemoglobin and hematocrit; decreased gastrointestinal motility; inflammation localized; decreases in blood pressure; increased defecation; decrease accommodation; increased coagulation times; decreased igg; decreased urinary output; increase in bronchial and lachrymal secretions; increased agitation; increased depression; change in urinary frequency; increased secretions; increased angioedema; diminished taste; increased drowsiness; decreased environmental awareness; decreased blood cortisol; altered behavior; increased mortality; increase in diastolic blood pressure; increased qrs duration; increased glucose in urine; elevation in glucose; altered hearing; loss of stimulatory effect; increased airway pressure; changes in platelet counts; changes in neutrophil counts; changes in white blood cell counts; changes in liver function; changes in renal function tests; changes in transaminases levels; changes in bilirubin levels; changes in alkaline phosphatase levels; increasing disorientation; inflammation at infusion site; increased urinary urgency; decreased urinary urgency; increased oral; increased and decreased wbc; decreased alertness; increased middle ear pressure; increase in tsh; decrease in tsh; reduction of voluntary movements; increase in libido; increased fibrinogen; decreased response to adrenocorticotropic hormone; increased; increased bleeding events; increase in hepatic fat; increased systolic blood pressure; increased frequency of seizures; increased incidence of events; reduction in peak expiratory flow; increased risk of death; increased risk for breakthrough seizures; increase in plasma prolactin levels; decreases in platelet count; increases in total cholesterol; increased cholecystectomy rate; increases in serum total cholesterol; decreases in serum triglycerides; increase in cervical mucus; altered coagulation tests; altered thyroid function tests; decrease in pco2; increase in serum chloride; increases in total serum bilirubin; decreases in total protein; decreased protein; increased risk of myocardial ischemia; increased risk of bone fracture; increase in non high density lipoprotein cholesterol; increase in hemoglobin; increased risk of cancer; increase in mean corpuscular volume; increase in pulse rate; decreased right ventricular function; decreased coagulation factors; increased creatine phosphatase; reduced ability to regulate body temperature; increased circulatory osmotic load; increased risk of neoplasia; decreased serum chloride; decreased serum amylase activity; increase in clotting time; decrease in hemoglobin concentration; loss of peripheral sensation; changes to hair or nails; increased blood monocytes; decrease in serum albumin; increased state of depression; changes to hair; increased bleeding of ocular tissue; decreased acuity; increased anxiety; increase in pbi; increases in appetite and weight; increased skin friability; increased calcium secretion; increased mean

corpuscular volume; increased blood levels; changes in sensorium; decreased urine stream; diminished gastrointestinal motility; decreased secretion in salivary gland; decreased secretion in sweat glands; decreased secretion in pharynx; decreased secretion in bronchi and nasal passages; decreased penile sensation; reduced tolerance to cold; increased sedation; increased risk for hemorrhagic complications; reduction in hemoglobin; decreases in heart rate; reduction in serum potassium; reduction in uric acid; decreased; decreases in serum bicarbonate levels; increased creatine and creatinine excretion; increased length of eyelashes; decreased length of eyelashes; decreased cd4+ cell count; changes in vaginal bleeding; decreased general strength and energy; increased systemic drug absorption; decreased serum sodium levels; decreased serum potassium levels; increased post-void residual volume; elevated acid phosphatase; decreased blood glucose; increased risk of nephrolithiasis; changes in distal tubules and glomeruli; increased micturition; decreased dark adaptation; increased hypotension; increase in serum toxicity of varenicline; increased alanine aminotransaminase; elevated total neutrophils; reduced efficacy with estrogen containing contraceptives; increase methemoglobin; increase in intracranial pressure (icp); increased risk for prostate cancer; alteration in mood; increased urinary calcium excretion; changes in high density lipoprotein levels; decreased appetite (grade $\geq$3); increased creatinine (grade 3 or 4); significantly disabling bleeding; inflammation pelvic; decreased food intake; elevated creatine kinase $\geq$3xuln; increased inr (grade $\geq$3); decreased appetite (grade 3-4); increased glucose (grade $\geq$ 3); decreased appetite (grade 3); increased spermatozoa motility; reduced fvc; increased serum urate; decreased mental focus; decreased therapeutic efficacy; decreased androstenedione; decreased breast milk production; increased cortisol-binding globulin; increased resistance to cardioversion; reduced absorption of fat-soluble vitamins

blastomycosis; typhlitis; stevens-johnson syndrome; acute generalized exanthematous pustulosis; encephalocele; colles' fracture; neutrophilic dermatosis; acute febrile neutrophilic dermatosis; hodgkin's disease nodular sclerosis; blast crisis in myelogenous leukaemia; strabismus; anterior chamber inflammation; early satiety; enteritis; malaise; acute hepatic failure; measles; acute promyelocytic leukaemia differentiation syndrome; injury; acute phosphate nephropathy; sudden onset of sleep; restlessness; reversible ischaemic neurological deficit; reversible cerebral vasoconstriction syndrome; negativism; compulsions; aspiration; acute enterocolitis; acute rhinitis; acute nonspecific tenosynovitis; acute angle-closure glaucoma; acute manic episode; trau-

mas; atypical mycobacterial infections; encephalopathies; acute respiratory failure; graves' disease; neurological disorders; aspiration pneumonia; acute gouty arthritis; acute hepatitis; avitaminosis; restless legs syndrome (rls); acute coronary syndrome (acs); acute urinary retention; plasmodium infections; necrotizing fasciitis; plasmodium falciparum infection; acute promyelocytic leukemia; acute cholecystitis; neurological impairments; acute liver injury; viral pharyngitis; acute exacerbation of psychosis; delayed gastric emptying; rapid eye movement sleep behavior disorder; injury of liver; acute circulatory failure; acute congestive heart failure; acute respiratory distress syndrome (ards); acute tonsillitis; acute pulmonary edema; acute cardiac ischemia; acute intermittent porphyria (aip); acute phase reaction; ruptured achilles tendon; acute respiratory distress; acute sinusitis; acute optic neuritis; acute tracheobronchitis; acute interstitial nephritis; acute urticaria; depigmentation; acute heart failure (ahf); acute exacerbation of chronic obstructive pulmonary disease; acute kidney injury (aki); raised lipids; aspiration pneumonitis; acute anterior uveitis; arrest; acute renal failure (arf); acute coronary events; moniliasis; malaise and fatigue; reversible neutropenia; acute tubular necrosis; accidental trauma; acute hypersensitivity reactions; reversible azotemia; irreversible renal insufficiency; fulminant hepatic necrosis; temporary loss of sense of smell; reversible hearing loss; acute hepatic necrosis; acute iritis; acute hemolytic anemia; fulminant hepatitis; reversible alopecia; reversible cholestatic jaundice; reversible agranulocytosis; acute labyrinthitis; monilia infection; acute fulminant liver failure; acute eye pain; reversible thrombocytopenia; acute non-thrombocytopenic purpura; neurological deterioration; reversible posterior leukoencephalopathy; reversible hpa axis suppression; acute dystonic reaction; acute ergot toxicity; reversible posterior leukoencephalopathy syndrome; atypical urinary bladder epithelial cells; neurological adverse reactions; cns neurologic disorder/cns toxicity; progressive immunosuppression; reversible interstitial nephritis; acute brain syndrome; acute elevated iop; severe tissue necrosis; reversible changes in liver function tests; delayed urine flow; impaired mental ability; acute and prolonged hypotensive episodes; acute pulmonary hypersensitivity reaction; impaired accommodation; defibrination; reversible hyperactivity; acute anxiety reaction; reversible jaundice; acute organic brain syndrome; reversible mental confusion; reversible blurred vision; delayed catamenia; acute psychotic reactions; viral gastrointestinal infections; bacterial reproductive infections; new onset hypertension; progressive liver damage; neurological side effects; delayed menstrual period; viral ear/nose/throat infections; reversible corneal toxicity; reversible interstitial pneumonitis; acute tubulopathy; reversible acute renal failure; reversible intracranial pressure increased; acute dyskinesia; malaise/lethargy;

acute oxalate nephropathy; reversible nephrotic syndrome; monilial dermatitis; impaired sleep quality; acute muscular paralysis; progressive loss of brainstem reflexes; irreversible renal failure; fatal angioedema in larynx; acute withdrawal symptoms; bizarre breathing patterns; acute renal failure possibly requiring dialysis; fulminating pneumonia; reversible elevation of serum calcium; abscess formation; acute dystonia; suspected foodborne fatal salmonella sepsis; acute elevated intraocular pressure; acute diffuse infiltrative pulmonary disease; cerebral hematomas; elevated lipase ¿ 3xuln; acute exacerbation of hepatitis b; cerebral-vascular disease with mitral valve prolapse; malaise relapse; delayed weight gain; sudden death cardiorespiratory arrest; neurological events; irreversible retinal damage; reversible corneal opacities; progressive pigmentation of skin; progressive pigmentation of conjunctiva; delayed orgasm; atypical trauma; early diarrhea; acute hyperexcited states; onset of new neurological symptoms; acute anaphylactic reaction with bronchospasm; fluctuations in blood pressure; reversible hyperuricemia; acute coronary insufficiency; progressive cerebral hypoxia; acute transient myopia; acute cytolytic hepatitis; acute hepatic injury; lymophocele; rapid heartbeat; reversible acute renal insufficiency; hearing disability; acute narrow angle glaucoma; injury to nerves adjacent to injection site; acute necrosis of proximal tubules; acute eosinophilic pneumonia; delayed myelosuppression; neurological toxicity; reversible bone growth inhibition; early abortion; acute myopia; acute otitis media (aom); acute graft-versus-host disease (gvhd); neurological changes; rapid suffocation; delayed hemolysis; acute ventricular pause; wart viral; acute glaucoma; acute arachnoiditis; myelin degeneration; acute gout attack; acute opioid withdrawal (disorder); delayed recovery

contusions; eczema, dyshidrotic; anaphylactoid syndrome of pregnancy; porphyria non-acute; infusion related reaction; infusion site erythema; infusion site induration; infusion site pain; infusion site swelling; infusion site phlebitis; infusion site pruritus; infusion site reaction; infusion site infection; infusion site bruising; infusion site rash; infusion site cellulitis; infusion site irritation; infusion site necrosis; infusion site thrombosis; infusion site urticaria; infusion site discomfort; catheter site haemorrhage; anaphylactoid shock; hypnagogic hallucination; dermatitis exfoliative generalised; methotrexate toxicity; exacerbation of asthma; porphyrias, acute; rashes, eruptions and exanthems nec; autoimmune hemolytic anemia; compartment syndromes; migraine with aura; eczema herpeticum; complex regional pain syndrome type ii; exacerbation of chronic obstructive pulmonary disease; stress fractures; med-

ication overuse headache; disease exacerbation; corticotroph adenoma; exacerbation of copd; ige-mediated hypersensitivity; filtering blebs; thromboembolic diseases; disorder of the urea cycle; catheter related complications; initial insomnia; disorder of ejaculation; eczema, dermatitis; catheter thrombosis; generalized muscle spasm; worsening of sleep apnea; infusion site burning; exacerbation of dyspnea; exacerbation of pre-existing diabetes mellitus; hypersensivitiy syndrome; elevation of gastric hcl; exacerbation of systemic lupus erythematosus; onset of pseudomembranous colitis symptoms; worsening of psoriasis; reactivation of herpes zoster; exacerbation of psoriasis; exacerbation of convulsions; exacerbation of porphyric symptoms; worsening of angina pectoris; worsening of arterial insufficiency; temporary blurred vision; impairment of performance of routine activities; temporary unilateral loss of vision; worsening of congestive heart failure; elevation in liver function tests; infusion related pain; exacerbation of inflammatory bowel disease; activation of mania; generalized exfoliative erythroderma; exacerbation of psychotic symptoms; minimal decreased rbcs; generalized spasms; reactivation of latent infections; exacerbation of psychosis; vein pigmentation; exacerbation of recurrent herpes labialis; exacerbation of angina; exacerbation of cough; stimulation of urinary bladder with spontaneous voiding; substernal chest pain; asthma symptoms; activation of pre-existing peptic ulcer; thromboembolic disorders; exacerbation of epilepsy; joint and muscle stiffness; retention of serum electrolytes; inhibition of gonadotropin secretion; gustatory sense diminished; severe sclerosis of the skin and subcutaneous tissues; impaired adaptation to dark; eczematous eruptions; excess mucus or phlegm; activation of mania/hypomania; migraine aggravated; generalized burning; protrusion of the tongue; exacerbation of heart failure; anaphylactic reactions with injection; elevation in serum levels of skeletal muscle enzymes; exacerbated bradycardia in sick sinus syndrome; worsening of rosacea; elevation of cerebrospinal fluid pressure; exacerbation of psychoses; worsening of preexisting hypertension; worsening of heart failure; contusions and hematomas; worsening of urinary retention; exacerbation of headache; activation of systemic lupus erythematosus; puffy eyes; risks with concomitant use of antiretroviral drugs; signs or symptoms of urinary tract irritation; exacerbation of preexisting ulcer disease; low systolic blood pressure; worsening of the conjunctivitis; signs and symptoms of eye allergy; generalized numbness; exacerbation of arthritis; activation of latent iritis; exacerbation of peptic ulcer disease; thrombophlebitis of the leg; exacerbation of symptoms of myasthenia gravis; elevation in prothrombin; eczematoid eruption; weakness in the legs; symptoms of hypocorticism; worsening of the depression; cyptosporidiosis; reduction of left ventricular ejection factor; activa-

tion of latent horner's syndrome; eczema/rash/urticaria; worsening of organic brain syndrome; worsening of underlying illness; exacerbation of hyperphosphatemia; worsening of diabetes mellitus; disrupted body temperature regulation; exacerbation of joint symptoms; reactivation of hepatitis b virus infection; activation of mania or hypomania; infusion site thrombophlebitis; existing infections worsened; exacerbation of kearns-sayre syndrome; anaphylaxis/angioedema; weakness of the legs; exacerbation of chorea; exacerbation of preexisting pulmonary infection; worsening of narrow angle glaucoma; exacerbation of congestive heart failure; new onset diabetic macular edema; worsening of diabetic macular edema; exacerbation of parkinsonian symptoms; exacerbation of pruritus; generalized hives; generalized warmth; prolonged activated partial thromboplastin time; potentiation of antihypertensive effect; persistent or severe hand-and-foot syndrome; exacerbation of raynaud's syndrome; exacerbation of hepatitis; exacerbation of porphyria; exacerbation of motor and phonic tics; exacerbation of viral ocular infections; asthma attack; psoriasis (pso); exacerbation of mood disorders; exacerbation of chronic hepatitis b; worsening of pre-existing narcolepsy; hyperglycemia (grade 3 or 4); partial permanent deafness; exacerbation of cutaneous lupus erythematosus; symptoms of hypercorticism; glucocorticoid related adverse effects; deterioration, clinical; stimulation caused by adrenergic beta-2 agonist effect; exacerbation of endometriosis; exacerbation of migraine; exacerbation of systemic lupus erythematosus (sle); exacerbation of hepatic hemangioma; exacerbation of, pre-existing lesions; exacerbated psychotic disorders; anaphylactic reactions with hypotension; exacerbation of reflux; exacerbation of chronic lung disease; ild/pneumonitis

endomyocardial fibrosis; mucositis; blepharospasm; blepharoptosis; scleritis; endometriosis; periostitis; periarthritis; fibromyalgia; myoglobinuria; fibroadenoma; sciatica; myokymia; myotonia; flatulence; mucocele; retroperitoneal fibrosis; ecthyma; pyoderma gangrenosum; pseudolymphoma; pleuropericarditis; pericarditis; pericardial rub; intracardiac thrombus; extrasystoles; epidermolysis; supernumerary nipple; talipes; scleral haemorrhage; scleral hyperaemia; endophthalmitis; punctate keratitis; giant papillary conjunctivitis; lacrimal disorder; lenticular opacities; retroperitoneal haemorrhage; encapsulating peritoneal sclerosis; palatal oedema; mucocutaneous haemorrhage; mucous membrane disorder; inflammation; fibrosis; nodular regenerative hyperplasia; glomerulonephritis; scleroderma; circumoral oedema; cytomegalovirus syndrome; cystitis; myopathy toxic; epicondylitis; tenosynovitis; tendinous contrac-

ture; rhabdomyolysis; bursitis; fibroadenoma of breast; myoclonus; extrapyramidal disorder; lacunar infarction; pseudodementia; cystitis-like symptom; hydronephrosis; mesangioproliferative glomerulonephritis; cystitis ulcerative; endometrial thickening; polymenorrhoea; atelectasis; papule; superficial ocular infections; tendinitis; exfoliative erythroderma; mucocutaneous candida infection; superficial punctate keratitis; blepharoconjunctivitis; periarteritis nodosa; lacrimation disorders; cytopenias; subarachnoid hemorrhage; angioneurotic edema; fibrocystic disease of breast; fibrocystic changes of breast; perioral dermatitis; granulomatosis with polyangiitis; vasospastic angina; bullous keratopathy; extrauterine pregnancy; scleral thinning; cytomegalovirus disease; erosive gastritis; pap test abnormalities; lymphocytic lymphomas; deliberate overdose; dermabrasion; demyelinating polyneuropathy; cytomegalovirus viraemia; endometrial polyps; lichen planus (lp); nodules; lacrimation; pseudopheochromocytoma; hydrocephaly; intraventricular conduction delay; keloids scars; extravasation injury; peri rectal abscess; thrombus formation during pci; circumoral pallor; moniliasis genital; myoglobin increased; galactorrhea; periorbital edema; ectopic calcification; purpuric rash; extraocular palsy; purpuric dermatitis; discoloration of thyroid gland; lacrimation abnormal; subconjunctival hemorrhage; deep and superficial venous thrombosis; subacute or chronic pneumonitis; scleroderma-like skin changes; tenesmus; pseudomembraneous colitis; retrobulbar neuritis; centrilobular necrosis; extravasation necrosis at injection site; superficial punctate keratopathy; stromal edema; subepithelial corneal lesions; inability to concentrate; superficial keratitis; cyst nos; periportal hepatic fibrosis; myopathy/increased creatinine phosphokinase; meningeal signs; perisinusoidal fibrosis; cytolytic hepatitis; lacrimal duct stenosis; suppuration; extrasystoles ventricular; vesiculation and bullae formation; vesiculobullous reaction; fibrin decrease; fixed drug eruptions; defective oogenesis; defective spermatogenesis; mucocutaneous eruptions; bullous lichen planus; pseudoparkinsonism; mucositis/stomatitis; subacute painful myopathy; mucocutaneous infection; fibrotic complications; fibrocystic breast; extrapulmonary pneumocystosis; mucocutaneous abnormalities; folliculosis; purpura/hematoma; endocervical ectropion; retrocollis; diffuse epithelial keratitis; periorbital pigmentation; papanicolaou smear suspicious; intramenstrual bleeding; small bowel mesenteric ischemia; incoherent speech; subdued temperament; periarteritic vasculitis; vasculitic toxicity; diffuse pulmonary infiltrates; discoloration of secretions; supraorbital pain; pemphigoid reaction; mucous membrane abnormality; precordial chest pain; pral ulcers; perivascular infiltration; extradural hematoma; subacute stent thrombosis; intercurrent infections; malodor; extrapyramidal phenomena; periungual erythema; an-

gioedema in mucous membranes; pemphigoid-like lesion; discoloration of body fluids; lumbalgia/backache; sclerosing of the skin; myoclonic jerks; peridiverticular abscess; myopathy/myositis; myopathy cases; myositis cases; discoloration; vasculitis necrotizing; discoloration of toes; pleomorphic rashes; purpuric nephritis; granulomatous interstitial nephritis; lenticular and corneal deposits; intramyocardial lesions; pericardial effusion/hemorrhage/tamponade; cytogenic abnormality; pericaridal effusion; periesophageal abscess; sclera hyperemia; capsular opacity; mixed hepatocellular liver injury; pseudoaneurysms at puncture site; perifollicular erythema; perifollicular edema; false low creatine phosphokinase; unspecified circulatory system disorder; mucous membranes reaction; lacrimal gland dysfunction; hyaline casts in urine; turbinate edema; extravasation with burning pain; pemphigus-like reactions; extravasation necrosis; mucocutaneous edema; pericaridal tamponade; pleomorphic skin eruptions; discoloration of urine; extravasation at injection site; punctate subepithelial; circumoralparesthesia; angioneurotic edema-type reactions; conjuctival hemorrhage; eosinophilic infiltration; intrahepatic biliary sclerosis; extrahepatic biliary sclerosis; tingling of fingers; may enter human milk; lenticular deposits; intranasal hypoesthesia; reduced ability to recognize pregnancy; metahemoglobinemia; platelet count (¡50,000/mm3); mucous membrane pigmentation disorder; pseudo obstruction; pseudopemphigoid; subacute myelopathy; subacute peripheral sensory neuropathy; subacute alopecia; subacute peripheral edema; subacute lymphedema; subacute loss of energy; extraprostatic necrosis; perifollicular hyperkeratosis; subcostal recession; conjuctival hyperemia; cystatin c increase

tinea pedis; venous thromboembolism; capillary leak syndrome; tics; dyspepsia; vitiligo; purpura; dizziness postural; bradycardia foetal; thyroiditis; thyroid disorder; vitritis; mastitis; neuromuscular toxicity; retinoic acid syndrome; neuromuscular block prolonged; hypotonia; bradyphrenia; convulsions local; cholinergic syndrome; bradykinesia; cerebral artery embolism; cerebral artery occlusion; disinhibition; neurosis; nervousness; venous thrombosis limb; spasms; convulsions; tinea infections; nausea and vomiting; venous insufficiency; erectile dysfunction; metastatic breast cancer; thrombocytopenic purpura; cerebral ischemia; neuromuscular blockade; thyroid nodules; thyroid dysfunction; thromboembolic stroke; temporomandibular joint dysfunction; spasmodic dysphonia; back pain lower back; apoplexy; pituitary adenoma; aches; neuromuscular disorders; thyroid adenoma; cerebral thrombosis; cerebral infarctions; copper deficiency; metastatic lung cancer; grand mal status epilepticus; pi-

tuitary neoplasms; thyroid cancers; cerebral microangiopathy; thyroiditis hashimoto; cerebral vein thrombosis; non convulsive status epilepticus; tumor flare; twitching; thyroid function abnormalities; tenderness of the skin; grand mal convulsions; prothrombin decreased; prothrombin lab abnormality; hypotensive crisis; reflex tachycardia; neuromuscular excitability; dryness of the nasal mucosa; thought disorder; marrow hypoplasia; adrenal cortex insufficiency; thyroid stimulating hormone increased; bradypnea; marked hypertension; hypotensive episode; venous pressure increased; taxia; spasm of vesicle sphincters; venous irritation; venous leak; tonic spasm of the masticatory muscles; magnesium ¡1.2meq/l; paralysis flaccid; lid reactions; magnesium loss; lid scales; vagal syndrome; weakness of legs; generalized ache; thyroid hormone level altered; venous phlebitis; aplastic anemia and pancytopenia; cerebral function deficiency; wasting syndrome; spasm of the neck muscles; hypotensive reactions; angioedema of the tongue; hypotensive collapse; weakness of hands or feet; hypotensive shock; weakness and paralysis of lower extremity; neuromuscular symptoms; venous phlebitis from injection site; reversible myocardial hypertrophy; active hepatic disease; nerve deafness; weakness and atrophy of proximal muscle groups; no effect; rhythm nodal; pituitary tumor benign; reflex bradycardia; developement of pubic hair; thrombocytopenia (grade $\geq$3); tumefactive multiple sclerosis (ms); weakness of the lower limbs; venous embolism; venous pressure decreased; side pain; melanin binding in ocular drug delivery

salpingitis; takotsubo cardiomyopathy; colic; aerophagy; anaemia heinz body; haemoconcentration; low set ears; orbital oedema; meibomianitis; madarosis; mesenteric vascular insufficiency; mesenteric arterial occlusion; haemorrhagic erosive gastritis; peritonitis; thirst; cold sweat; crying; haemorrhagic urticaria; perineal abscess; overgrowth fungal; poisoning; sleep terror; slow speech; masked facies; belligerence; fear of open spaces; fear; tension; haemorrhage urinary tract; pollakiuria; haemorrhage subcutaneous; sticky skin; shock; cold; redness; detached; dryness; warmth; gout attack; sting of the eye; asthenic conditions; cataracts; tension headache; sore throat; cold sore; heat stroke; back strain; still births; apraxias; delayed recovery from anaesthesia; tissue damage; hpv-related carcinoma; worrying; dry socket syndrome; pelvic symptoms; paradoxical anxiety; stinging of the skin; hangover effect; high energy; stinging or burning; sprue-like enteropathy; bigeminy; cold or flu syndrome; integument and mucus membrane toxicity; excessive hypotension; bigeminal rhythms; stinging sensation in eyes upon instillation; itchy eyes and lids; u waves; proarrhythmia; fearfulness;

cold sensations in hands and feet; bad taste following instillation; reddening; mind racing; fever and/or chills; lowered t-waves; dryness of the pharynx; staring episode; shallow breathing; dry scaly skin; sgpt; stinging of lips; pelvic cramping; sting on injection; reduced folic acid; loose or frequent stools; shallow respirations; cataract (not specified); bad dreams; cold feet; enteritis at all levels; stinging at application site; itching at wart site; dry and breaking hair; marked diuresis; cold or numbness; heat rash; shock-like coma; hives/ itching; migratory polyarthralgia; dry mouth/nose; punctate epithelial staining; grossly obvious gingivitis; patch non-adhesion; high urine wbc/hpf; high sgot; sense of pelvic pressure; warm feeling in vagina; tightness; warm sensations; warm/cold sensations; mesenteric arterial thrombosis; tightness and rigidity; cold sensations; dryness of hair and scalp; bad taste in mouth; dryness around the eye; lowered blood pressure; slow urination; dryness of the mucous membranes; stinging sensation; stinging of skin; thirst disturbance; serum phenytoin alteration; ideation; high pitched cry; low absolute neutrophil count; bad taste/metallic taste; erratic blood pressure; acne-like rash; foot tapping; coldness of the extremities; reactive hypertension; shortening of refractory periods; thirst and dehydration; sticky sensation; low hemoglobin counts; grimace; watery itchy eyes; hemorrhage at puncture site; dissection of the coronary vessels; flattening of t wave; sloughing; warm sensation over body; high bun; dryness of the oropharynx; proness to falling; haemorrhagic eruption; reactivation of psychotic processes; excessive skin wrinkling; dryness of the oral mucous membranes; cold skin; proneness to falling; dry mouth/nasal stuffiness; low plasma cortisol; dry skin non-application site; mesenteric embolism; libido decreased-male; pco2 changes; dryness of paranasal area; dryness of inrtaorbital area; warm autoimmune hemolytic anemia (waiha); gangrenous bowel bleeding; leaks, anastomotic; headache (grade $\geq$3); dry skin (grade $\geq$3); back pain (grade 3); seeing half of an object; dry, cracked skin; lowered blood urea; death caused by therapeutic agent toxicity; bigeminal beats

## A.2 Molecule 2

SMILES: O=C(NC(=O)c1cccc([N+](=O)[O-])c1)c1ccc(F)cc1

catalepsy; snoring; sneezing; negative cardiac inotropic effect; bullous conditions; pathological gambling disorder; minimal change disease; morbilliform eruptions; gross

bleeding/perforation; bullous reactions; seizure-like activity; heavy head/tired; no light reflex; any noncerebral bleeding; unusual dreams; generalized itching; reactive hypotension; unusual and sometimes aggressive behaviour; any bleeding reaction; antiperistaltic effects; pure cell aplasia; heavy headedness; axillary sweating; all grades edema; all grades nausea; all grades abdominal pain; all grades constipation; all grades vomiting; all grades dyspnea; all grades cough; all grades pleural effusion; all grades decreased appetite; all grades pneumonia; all grades decreased albumin; all grades increased creatinine; all grades elevations in serum alanine aminotransferase (alt); all grades increased ast; all grades decreased sodium; all grades increased potassium; all grades increased amylase; all grades decreased lymphocytes; all grades decreased leukocytes; all grades increased lipase; all grades diarrhoea; all grades pyrexia; all grades peripheral neuropathy; all grades dizziness; all grades upper respiratory tract infection; all grades cataract; all grades vision blurred; all grades decreased platelets; all grades decreased neutrophils; all grades decreased phosphate; all grades decreased calcium; all grades increased bun; all grades decreased potassium; all grades decreased magnesium; all grades increased alt; all grades increased bilirubin; all grades urinary tract infection; all grades taste disorder; all grades mental status changes; all grades hypotension; all grades hemorrhage; all grades increased blood glucose; all grades elevated creatine kinase; all grades peripheral sensory neuropathy; all grades headache; all grades myalgia; all grades arthralgia; all grades back pain; all grades pain in extremity; all grades rash; all grades lymphopenia; all grades hypocalcemia; all grades hypophosphatemia; all grades influenza; all grades hypertension; all grades visual impairment; all grades weight increased; all grades gastroesophageal reflux disease; all grades dyspepsia; all grades dry skin; all grades alopecia; all grades upper respiratory tract infections; all grades rhinorrhea; all grades ocular toxicity; all grades prolonged qtc interval; all grades red blood cell count decreased; all grades increased alanine aminotransferase; all grades increased aspartate aminotransferase; all grades stomatitis; all grades paronychia; all grades pruritus; all grades increased triglycerides; all grades increased creatine kinase; all grades increased serum uric acid; all grades calcium decreased; all grades increased serum creatinine; all grades increased cholesterol; all grades increased serum lipase; all grades decreased serum calcium; all grades increases in total serum bilirubin; all grades hypothyroidism; all grades thrombocytopenia; all grades hypomagnesemia; all grades insomnia; all grades hypoglycemia; all grades peripheral edema; all grades increased blood creatinine; all grades proteinuria; any grade neutropenia; any grade anemia; any grade thrombocytopenia; any grade diarrhea; any grade constipation; any grade nausea; any grade

fatigue; any grade pyrexia; any grade peripheral edema; any grade asthenia; any grade upper respiratory tract infection; any grade bronchitis; any grade viral upper respiratory tract infection; any grade pneumonia; any grade hypokalemia; any grade hypocalcemia; any grade hyperglycemia; any grade muscle spasms; any grade back pain; any grade peripheral neuropathy; any grade insomnia; any grade dyspnea; any grade cough; any grade rash; any grade hypertension; any grade thromboembolic events; any grade vomiting; any grade headache; any grade respiratory tract infection; all grades infusion-related reaction; all grades respiratory tract infection; all grades bronchitis; all grades chills; all grades muscle spasms; all grades chest pain; all grades epistaxis; any grade abdominal pain; any grade edema; any grade decreased appetite; any grade arthralgia; any grade musculoskeletal pain; any grade mucositis; any grade transaminase elevation; any grade lower respiratory tract infection; any grade lymphocytosis; any grade leukopenia; any grade lymphopenia; any grade increased alt; any grade ast increased; any grade lipase increased; any grade hypophosphatemia; any grade increased alp; any grade elevated serum amylase; any grade hyponatremia; any grade hyperkalemia; any grade hypoalbuminemia; any grade increased creatinine; any grade decreased weight; any grade decreased phosphate; any grade increased amylase; any grade alkaline phosphatase increased; all grades diarrhea/colitis; all grades dry mouth; all grades acne; all grades photosensitivity reaction; all grades nail abnormalities; all grades prolonged qt interval; all grades corneal abnormalities; all grades blurred vision; all grades depression

anencephaly; necrolytic migratory erythema; dermatitis, phototoxic; large intestine polyp; pancreatitis haemorrhagic; hangover; pyoderma; sleep talking; frustration; retrograde ejaculation; bladder dysfunction; excessive weight gain; pancreatic acinar cell carcinoma; red man syndrome; neoplasms, malignant; high triglyceride level; fundic gland polyps; atypical subtrochanteric fracture; primary biliary cholangitis; excessive perspiration; local inflammation at injection site; aggravated convulsions; increased risks for incranial bleeding; increased incidence of gallbladder disease; local irritation at implant site; increased incidence of forceps delivery; necrotizing angiitis; excessive thirst; acetonuria; increased frequency of sinus pauses/bradycardia; increased severity of atrial flutter/chest pain; excessive volume depletion; increased risk for and severity of pneumonias; involuntary motor activity; unexpected pregnancy; retinal and petechial hemorrhages; accidental and intentional overdose; excessive rise in blood pressure; elastosis perforans serpinginosa; platelet count ¡75,000 cells/mcl; intense urge

to gamble; negative t-waves; elevation of intracholedochal pressure; elevated amylase grade 4 (¿5xuln); elevated alt grade 3 (¿5-10xuln); elevated alt grade 4 (¿10xuln); elevated cholesterol grade 2 (¿300-400 mg/dl); elevated cholesterol grade 3 (¿400-500 mg/dl); elevated cholesterol grade 4 (¿500 mg/dl); elevated triglycerides grade 3 (¿750-1200 mg/dl); increased risk of asthma-related death; increased risk of suicidal thoughts or behaviour; calcium loss; fluid disturbances; necrotizing colitis; elevations in creatinine levels; microvesicular steatosis; 3 second r-r interval; festination; orthostatic change in blood pressure; accommodative change; disorientation and delirium; infiltration of unintended structures or cavities; phlebitic symptoms; failure to gain weight; aggressiveness; long-term paralysis; nitrogen balance negative due to protein catabolism; total cholesterol increased; intense urges to spend money; increased risk of supine hypertension; elastosis perforans serpiginosa; increased number of eyelashes; fatal or life-threatening bleeding; excessive bleeding at injection site; painful enlargement of breast; fullness sensation; negative calcium balance; growth potentiation of benign meningioma; retrolental fibroplasia; fluid accumulation; elevated ast, alt or alkaline phosphatase grade 3-4; elevated ast, alt or alkaline phosphatase grade $\geq$1; elevations of serum cholesterol; increased risk for and severity of infections; elevations in serum alanine aminotransferase (alt); inadequate response to treatment; carcinoma of prostate (disorder); neutrophil count ¡1000 cells/mm3; high phosphate levels; partial transitory deafness; elevations in liver enzymes or bilirubin; platelet transfusion refractoriness (ptr); neutrophil count 500-1000 cells/mm3; neutrophil count ¡500 cells/mm3; grade 3, grade 4 fatigue; grade 3, grade 4, grade 5 nausea; grade 3, grade 4, grade 5 hypotension; grade 3, grade 4, grade 5 constipation; grade 3, grade 4, grade 5 hypoxia; grade 3, grade 4 depressed level of consciousness; grade 3, grade 4 basal cell carcinoma; suppurative parotitis; grade 3, grade 4 squamous cell carcinoma of skin; grade 3 all grades elevations in serum alanine aminotransferase (alt); grade 4 all grades elevations in serum alanine aminotransferase (alt); grade 2, grade 3, grade 4 nausea; grade 3, grade 4 cognitive effects; lithium toxicities; urinary tract toxicity caused by radiations; nodal extrasystole; perioperative myoclonus

bronchopneumonia; pneumonia, pneumococcal; tracheitis; ventricular flutter; pulmonary infarction; pneumatosis cystoides intestinalis; pulmonary eosinophilia; respiratory paralysis; respiratory distress syndrome, adult; pneumothorax; cough; ventricular hypokinesia; grey syndrome neonatal; endocarditis; chest discomfort; ventricular asystole; throat irritation; fever neonatal; reye's syndrome; bronchospasm; bronchioli-

tis; pneumonia respiratory syncytial viral; laryngitis; pulmonary sepsis; pneumonia; bronchitis; pneumonia fungal; respiratory moniliasis; bronchopulmonary aspergillosis; lung infection pseudomonal; pneumonia legionella; escherichia sepsis; pulmonary toxicity; pulmonary oil microembolism; epidural lipomatosis; chondrolysis; spondylitis; breath holding; laryngospasm; respiratory tract congestion; respiratory tract irritation; respiration abnormal; bronchial irritation; sputum increased; sputum discoloured; productive cough; cough decreased; respiratory depth increased; pulmonary arteriopathy; pulmonary microemboli; pulmonary alveolar haemorrhage; lung consolidation; pneumonitis; wheezing; bronchospasm paradoxical; bronchial obstruction; bronchiectasis; erythrodermic psoriasis; bronchoconstriction; fever; bronchial airway hyperreactivity; nose edema; breathing abnormalities; pulmonary embolism; respiratory distress syndrome; pulmonary fibrosis; chest pain; respiratory insufficiency; pleural effusions; airway obstruction; bronchioloalveolar carcinoma; bronchial asthma; respiratory failure; respiratory distress; respiratory depression; influenza-like illness; respiratory disorders; respiratory diseases; pulmonary arterial hypertension (pah); pulmonary obstruction; pulmonary edemas; right heart failure; respiratory symptoms; pulmonary veno occlusive disease; pulmonary mass; lung disorder; pulmonary infections; ventilatory depression; bronchopleural fistula; sputum; pulmonary hypertension (ph); cough, acute; respiratory alkalosis; respiratory system abnormalities; pulmonary inflammation; respiratory distress syndrome, infant; pulmonary alveolar proteinosis (pap); ventricular failure, right; pulmonary haemorrhage; lung, carcinoma; respiratory tract hemorrhage; pulmonary congestion; lung cancers; lung function changes; chest disorders; pulmonary diseases, obstructive; respiratory muscle paralysis; ventricular tachycardia (vt); airway complication of anesthesia; lung edema; pulmonary infiltrates; pulmonary events; chest tightness; chest congestion; breath odor; pulmonary interstitial infiltrates; dyspnea aggravated; throat discomfort; lung edema disorder; respiratory inadeqaucy; sweat gland; bronchospasm aggravated; pulmonary problem; nervousness/irritability; throat inflammation; chest pain/discomfort; nose pain; escherichia bacteremia; bronchial constriction; throbbing headache; pulmonary hypersensitivity reactions; allergic-type respiratory symptoms; chest pain/pressure/angina; respiratory flu; pulmonary infiltration with eosinophilia; expiratory wheezing; bronchial wheezing; fever and headache; chest wall rigidity; chest wall pain; respiratory acidosis during weaning; fever or influenza-like illness; ecg change consistent with myocardial ischemia; vagolysis; respiratory and cns depression; chest symptoms; throat/neck symptoms; respiratory side effects; perlèche; bronchiolar constriction; fetal anticonvulsant syndrome; pulmonary hypotension; meniere's syn-

drome; pulmonary-clinical; pulmonary allergy; chest flushing; innapropiate secretion of antidiuretic hormone; allergic pericarditis; delirium hallucinations; pulmonary vascular occlusion; cough suppression; cough experience post-inhalation; intramuscular hemorrhage; throat pressure; asthma-related death; respiratory impairment; respiratory symptom-related adverse reactions; swellin of lower legs; respiratory wheezing; respiratory arrest (disorder); pulmonary changes; pulmonary symptoms; mechanical irritation at application site; mechanical cardiac dysfunction; bronchial anastomotic dehiscence; throat infections; erythroleukemia; chest mass; thoracic spine pain; asthmatic type dyspnea; respiratory rate increased; fluid overload/retention; respiratory acidosis/alkalosis; respiratory arrest/failure; pulmonary infarct; endocarditis caused by staphylococcus aureus; pneumonia (grade $\geq$3); pulmonary reactions; pleural effusion (grade $\geq$3); respiratory function impaired; bronchorrhea; pulmonary hypertension, primary; pulmonary fibrosis interstitial; lung hemorrhage; dyspepsia/heartburn; respiratory rate decreased; expiratory reserve volume decreased; chest infections

blastomycosis; typhlitis; stevens-johnson syndrome; acute generalized exanthematous pustulosis; encephalocele; colles' fracture; neutrophilic dermatosis; acute febrile neutrophilic dermatosis; hodgkin's disease nodular sclerosis; blast crisis in myelogenous leukaemia; strabismus; anterior chamber inflammation; early satiety; enteritis; malaise; acute hepatic failure; measles; acute promyelocytic leukaemia differentiation syndrome; injury; acute phosphate nephropathy; sudden onset of sleep; restlessness; reversible ischaemic neurological deficit; reversible cerebral vasoconstriction syndrome; negativism; compulsions; aspiration; acute enterocolitis; acute rhinitis; acute nonspecific tenosynovitis; acute angle-closure glaucoma; acute manic episode; traumas; atypical mycobacterial infections; encephalopathies; acute respiratory failure; graves' disease; neurological disorders; aspiration pneumonia; acute gouty arthritis; acute hepatitis; avitaminosis; restless legs syndrome (rls); acute coronary syndrome (acs); acute urinary retention; plasmodium infections; necrotizing fasciitis; plasmodium falciparum infection; acute promyelocytic leukemia; acute cholecystitis; neurological impairments; acute liver injury; viral pharyngitis; acute exacerbation of psychosis; delayed gastric emptying; rapid eye movement sleep behavior disorder; injury of liver; acute circulatory failure; acute congestive heart failure; acute respiratory distress syndrome (ards); acute tonsillitis; acute pulmonary edema; acute cardiac ischemia; acute intermittent porphyria (aip); acute phase reaction; ruptured achilles tendon; acute respiratory distress; acute sinusitis; acute optic neuritis; acute tracheobronchi-

tis; acute interstitial nephritis; acute urticaria; depigmentation; acute heart failure (ahf); acute exacerbation of chronic obstructive pulmonary disease; acute kidney injury (aki); raised lipids; aspiration pneumonitis; acute anterior uveitis; arrest; acute renal failure (arf); acute coronary events; moniliasis; malaise and fatigue; reversible neutropenia; acute tubular necrosis; accidental trauma; acute hypersensitivity reactions; reversible azotemia; irreversible renal insufficiency; fulminant hepatic necrosis; temporary loss of sense of smell; reversible hearing loss; acute hepatic necrosis; acute iritis; acute hemolytic anemia; fulminant hepatitis; reversible alopecia; reversible cholestatic jaundice; reversible agranulocytosis; acute labyrinthitis; monilia infection; acute fulminant liver failure; acute eye pain; reversible thrombocytopenia; acute non-thrombocytopenic purpura; neurological deterioration; reversible posterior leukoencephalopathy; reversible hpa axis suppression; acute dystonic reaction; acute ergot toxicity; reversible posterior leukoencephalopathy syndrome; atypical urinary bladder epithelial cells; neurological adverse reactions; cns neurologic disorder/cns toxicity; progressive immunosuppression; reversible interstitial nephritis; acute brain syndrome; acute elevated iop; severe tissue necrosis; reversible changes in liver function tests; delayed urine flow; impaired mental ability; acute and prolonged hypotensive episodes; acute pulmonary hypersensitivity reaction; impaired accommodation; defibrination; reversible hyperactivity; acute anxiety reaction; reversible jaundice; acute organic brain syndrome; reversible mental confusion; reversible blurred vision; delayed catamenia; acute psychotic reactions; viral gastrointestinal infections; bacterial reproductive infections; new onset hypertension; progressive liver damage; neurological side effects; delayed menstrual period; viral ear/nose/throat infections; reversible corneal toxicity; reversible interstitial pneumonitis; acute tubulopathy; reversible acute renal failure; reversible intracranial pressure increased; acute dyskinesia; malaise/lethargy; acute oxalate nephropathy; reversible nephrotic syndrome; monilial dermatitis; impaired sleep quality; acute muscular paralysis; progressive loss of brainstem reflexes; irreversible renal failure; fatal angioedema in larynx; acute withdrawal symptoms; bizarre breathing patterns; acute renal failure possibly requiring dialysis; fulminating pneumonia; reversible elevation of serum calcium; abscess formation; acute dystonia; suspected foodborne fatal salmonella sepsis; acute elevated intraocular pressure; acute diffuse infiltrative pulmonary disease; cerebral hematomas; elevated lipase ¿ 3xuln; acute exacerbation of hepatitis b; cerebral-vascular disease with mitral valve prolapse; malaise relapse; delayed weight gain; sudden death cardiorespiratory arrest; neurological events; irreversible retinal damage; reversible corneal opacities; progressive pigmentation of skin; progressive pigmentation of conjunctiva; delayed orgasm; atypical

trauma; early diarrhea; acute hyperexcited states; onset of new neurological symptoms; acute anaphylactic reaction with bronchospasm; fluctuations in blood pressure; reversible hyperuricemia; acute coronary insufficiency; progressive cerebral hypoxia; acute transient myopia; acute cytolytic hepatitis; acute hepatic injury; lymophocele; rapid heartbeat; reversible acute renal insufficiency; hearing disability; acute narrow angle glaucoma; injury to nerves adjacent to injection site; acute necrosis of proximal tubules; acute eosinophilic pneumonia; delayed myelosuppression; neurological toxicity; reversible bone growth inhibition; early abortion; acute myopia; acute otitis media (aom); acute graft-versus-host disease (gvhd); neurological changes; rapid suffocation; delayed hemolysis; acute ventricular pause; wart viral; acute glaucoma; acute arachnoiditis; myelin degeneration; acute gout attack; acute opioid withdrawal (disorder); delayed recovery

contusions; eczema, dyshidrotic; anaphylactoid syndrome of pregnancy; porphyria non-acute; infusion related reaction; infusion site erythema; infusion site induration; infusion site pain; infusion site swelling; infusion site phlebitis; infusion site pruritus; infusion site reaction; infusion site infection; infusion site bruising; infusion site rash; infusion site cellulitis; infusion site irritation; infusion site necrosis; infusion site thrombosis; infusion site urticaria; infusion site discomfort; catheter site haemorrhage; anaphylactoid shock; hypnagogic hallucination; dermatitis exfoliative generalised; methotrexate toxicity; exacerbation of asthma; porphyrias, acute; rashes, eruptions and exanthems nec; autoimmune hemolytic anemia; compartment syndromes; migraine with aura; eczema herpeticum; complex regional pain syndrome type ii; exacerbation of chronic obstructive pulmonary disease; stress fractures; medication overuse headache; disease exacerbation; corticotroph adenoma; exacerbation of copd; ige-mediated hypersensitivity; filtering blebs; thromboembolic diseases; disorder of the urea cycle; catheter related complications; initial insomnia; disorder of ejaculation; eczema, dermatitis; catheter thrombosis; generalized muscle spasm; worsening of sleep apnea; infusion site burning; exacerbation of dyspnea; exacerbation of pre-existing diabetes mellitus; hypersensivitiy syndrome; elevation of gastric hcl; exacerbation of systemic lupus erythematosus; onset of pseudomembranous colitis symptoms; worsening of psoriasis; reactivation of herpes zoster; exacerbation of psoriasis; exacerbation of convulsions; exacerbation of porphyric symptoms; worsening of angina pectoris; worsening of arterial insufficiency; temporary blurred vision; impairment of performance of routine activities; temporary unilateral loss of

vision; worsening of congestive heart failure; elevation in liver function tests; infusion related pain; exacerbation of inflammatory bowel disease; activation of mania; generalized exfoliative erythroderma; exacerbation of psychotic symptoms; minimal decreased rbcs; generalized spasms; reactivation of latent infections; exacerbation of psychosis; vein pigmentation; exacerbation of recurrent herpes labialis; exacerbation of angina; exacerbation of cough; stimulation of urinary bladder with spontaneous voiding; substernal chest pain; asthma symptoms; activation of pre-existing peptic ulcer; thromboembolic disorders; exacerbation of epilepsy; joint and muscle stiffness; retention of serum electrolytes; inhibition of gonadotropin secretion; gustatory sense diminished; severe sclerosis of the skin and subcutaneous tissues; impaired adaptation to dark; eczematous eruptions; excess mucus or phlegm; activation of mania/hypomania; migraine aggravated; generalized burning; protrusion of the tongue; exacerbation of heart failure; anaphylactic reactions with injection; elevation in serum levels of skeletal muscle enzymes; exacerbated bradycardia in sick sinus syndrome; worsening of rosacea; elevation of cerebrospinal fluid pressure; exacerbation of psychoses; worsening of preexisting hypertension; worsening of heart failure; contusions and hematomas; worsening of urinary retention; exacerbation of headache; activation of systemic lupus erythematosus; puffy eyes; risks with concomitant use of antiretroviral drugs; signs or symptoms of urinary tract irritation; exacerbation of preexisting ulcer disease; low systolic blood pressure; worsening of the conjunctivitis; signs and symptoms of eye allergy; generalized numbness; exacerbation of arthritis; activation of latent iritis; exacerbation of peptic ulcer disease; thrombophlebitis of the leg; exacerbation of symptoms of myasthenia gravis; elevation in prothrombin; eczematoid eruption; weakness in the legs; symptoms of hypocorticism; worsening of the depression; cyptosporidiosis; reduction of left ventricular ejection factor; activation of latent horner's syndrome; eczema/rash/urticaria; worsening of organic brain syndrome; worsening of underlying illness; exacerbation of hyperphosphatemia; worsening of diabetes mellitus; disrupted body temperature regulation; exacerbation of joint symptoms; reactivation of hepatitis b virus infection; activation of mania or hypomania; infusion site thrombophlebitis; existing infections worsened; exacerbation of kearns-sayre syndrome; anaphylaxis/angioedema; weakness of the legs; exacerbation of chorea; exacerbation of preexisting pulmonary infection; worsening of narrow angle glaucoma; exacerbation of congestive heart failure; new onset diabetic macular edema; worsening of diabetic macular edema; exacerbation of parkinsonian symptoms; exacerbation of pruritus; generalized hives; generalized warmth; prolonged activated partial thromboplastin time; potentiation of antihypertensive effect; persistent or se-

vere hand-and-foot syndrome; exacerbation of raynaud's syndrome; exacerbation of hepatitis; exacerbation of porphyria; exacerbation of motor and phonic tics; exacerbation of viral ocular infections; asthma attack; psoriasis (pso); exacerbation of mood disorders; exacerbation of chronic hepatitis b; worsening of pre-existing narcolepsy; hyperglycemia (grade 3 or 4); partial permanent deafness; exacerbation of cutaneous lupus erythematosus; symptoms of hypercorticism; glucocorticoid related adverse effects; deterioration, clinical; stimulation caused by adrenergic beta-2 agonist effect; exacerbation of endometriosis; exacerbation of migraine; exacerbation of systemic lupus erythematosus (sle); exacerbation of hepatic hemangioma; exacerbation of, pre-existing lesions; exacerbated psychotic disorders; anaphylactic reactions with hypotension; exacerbation of reflux; exacerbation of chronic lung disease; ild/pneumonitis

tinea pedis; venous thromboembolism; capillary leak syndrome; tics; dyspepsia; vitiligo; purpura; dizziness postural; bradycardia foetal; thyroiditis; thyroid disorder; vitritis; mastitis; neuromuscular toxicity; retinoic acid syndrome; neuromuscular block prolonged; hypotonia; bradyphrenia; convulsions local; cholinergic syndrome; bradykinesia; cerebral artery embolism; cerebral artery occlusion; disinhibition; neurosis; nervousness; venous thrombosis limb; spasms; convulsions; tinea infections; nausea and vomiting; venous insufficiency; erectile dysfunction; metastatic breast cancer; thrombocytopenic purpura; cerebral ischemia; neuromuscular blockade; thyroid nodules; thyroid dysfunction; thromboembolic stroke; temporomandibular joint dysfunction; spasmodic dysphonia; back pain lower back; apoplexy; pituitary adenoma; aches; neuromuscular disorders; thyroid adenoma; cerebral thrombosis; cerebral infarctions; copper deficiency; metastatic lung cancer; grand mal status epilepticus; pituitary neoplasms; thyroid cancers; cerebral microangiopathy; thyroiditis hashimoto; cerebral vein thrombosis; non convulsive status epilepticus; tumor flare; twitching; thyroid function abnormalities; tenderness of the skin; grand mal convulsions; prothrombin decreased; prothrombin lab abnormality; hypotensive crisis; reflex tachycardia; neuromuscular excitability; dryness of the nasal mucosa; thought disorder; marrow hypoplasia; adrenal cortex insufficiency; thyroid stimulating hormone increased; bradypnea; marked hypertension; hypotensive episode; venous pressure increased; taxia; spasm of vesicle sphincters; venous irritation; venous leak; tonic spasm of the masticatory muscles; magnesium ¡1.2meq/l; paralysis flaccid; lid reactions; magnesium loss; lid scales; vagal syndrome; weakness of legs; generalized ache; thyroid

hormone level altered; venous phlebitis; aplastic anemia and pancytopenia; cerebral function deficiency; wasting syndrome; spasm of the neck muscles; hypotensive reactions; angioedema of the tongue; hypotensive collapse; weakness of hands or feet; hypotensive shock; weakness and paralysis of lower extremity; neuromuscular symptoms; venous phlebitis from injection site; reversible myocardial hypertrophy; active hepatic disease; nerve deafness; weakness and atrophy of proximal muscle groups; no effect; rhythm nodal; pituitary tumor benign; reflex bradycardia; developement of pubic hair; thrombocytopenia (grade $\geq$3); tumefactive multiple sclerosis (ms); weakness of the lower limbs; venous embolism; venous pressure decreased; side pain; melanin binding in ocular drug delivery

## A.3 Molecule 3

SMILES: Cc1cc2nn(CC(=O)NC(=O)NCC(=O)C(=O)Nc3ccc4c(c3)-c3ccccc3C4)c(=O)n2c(N2CCc
atrial flutter; atrial fibrillation; ventricular fibrillation; mitral valve insufficiency; eclampsia; albuminuria; echolalia; heartburn; infarction; cardiotoxicity; heart injuries; diaper rash; cardiac tamponade; cardiomegaly; myocardial rupture; atrial rupture; myocardial haemorrhage; coronary artery dissection; coronary artery thrombosis; coronary artery embolism; carditis; atrial thrombosis; cardiac fibrillation; cardiac flutter; exercise tolerance decreased; catheter site pain; metabolic encephalopathy; adrenergic syndrome; athetosis; hypertensive encephalopathy; apathy; bilirubinuria; postmenopausal haemorrhage; vasoconstriction; hypertensive crisis; thrombophlebitis; arterial thrombosis; vasospasm; arteriosclerosis; digoxin toxicity; cardiac murmur; cardiac conduction disorders; cerebrovascular embolism and thrombosis; cerebrovascular venous and sinus thrombosis; pruritus nec; coronary artery disorders; vascular hypertensive disorders; myocardial ischemia; angina pectoris; myocardial infarction; vascular diseases; heart failure; cerebrovascular accident; ventricular arrhythmia; cardiovascular; ischemic heart disease; ischemic stroke; embolism and thrombosis; cardiac arrest; ventricular dysfunction; cardiac complications; vasodilation; coronary artery stenosis; heart block; cardiovascular abnormalities; major depressive episode; valvular heart disease; coronary artery insufficiency; cardiovascular complications; coronary occlusions; metabolic disorders; cerebrovascular insufficiency; major bleeding; cardiac; cardiovascular risk; cardiac failure; arterial thromboembolism; lipid disorders;

myocardial depression; ventricular hypertrophy; cardiovascular mortality; risk of cardiac arrhythmias; arterial thromboembolic events; non cirrhotic portal hypertension; atherosclerosis risk; coronary artery occlusion; metabolic and nutritional complications; pacing; atrial arrhythmia; cardiovascular events; cardiac disorder; coronary artery atherosclerosis; coronary artery vasospasm; calcium deficiency; cardiac decompensation; metabolic and nutritional disorders; ventricular tachyarrhythmias; cardiac rhythm disorders; cardiovascular disorders; stenosis of bile duct; aspirin exacerbated asthma; cardiovascular disease (cvd); cardiac amyloidosis; lipid disorders and lipid measurement; major adverse cardiac events; ischemic cerebrovascular accident; ecg changes; anginal pain; vessel puncture site pain; major adverse cardiovascular events; cardiac function impaired; supraventricular arrhythmias; cardiac dysfunction; cerebrovascular diseases; metabolic disturbances; diaper dermatitis; hemodynamics instability; cardiorespiratory arrest; cardiac output decreased; ejection fraction decreased; exertional dyspnea; myocardial reinfarction; myocardial ischemia or infarction; hypertensive reaction; hypertensive episodes; st segment changes; ischemic events; new onset diabetes mellitus; ventricular ectopic activity; cardiac valvulopathy; maculopapular and erythematous rashes; arrhythmia aggravated; minor hematuria; coronary artery bypass graft hemorrhage; cardiac conduction disturbances; cardiovascular collapse; arterial events; arteriospasm; antidiuretic effects; ischemic injury; convulsive events; angina/angina-like pain; myocardial ischemia/infarction; systolic ejection murmur; cerebrovascular reactions; angina/myocardial infarction; ecg pattern changes; cardiac abnormalities; thrombophlebitis deep; exertional hypotension; heart rate elevated/abnormal; elevation of serum triglyceride levels; vasodilation (usually flushing); rebound hypertension; angina-like chest pain; major noncerebral bleeding; cardiovascular depression following ect; kawasaki-like syndrome; electrocardiogram st segment depression; cardio-respiratory collapse; electrocardiogram qrs complex abnormal; myocardial hypertrophy; postmenopausal vaginal bleeding; cardiovascular side effects; cardiac conduction abnormalities; cardiovascular thrombotic events; metabolic effects; vasomotor response; new or increased angina pectoris; cardiac irregularities; thrombophlebitis leg; cerebrovascular spasm; calcium taste; cardiac-clinical; myocardial insufficiency; fatal outcomes have been reported; ischemic cardiovascular events; st elevated; cardiac rhythm abnormalities; autonomic instability; elevation of aldosterone; cardiac-related fluid retention; electrocardiogram pr prolongation; myocardial infarction/ cerebral infarction; lowering of estrogen; ecg and eeg changes; ischemic cerebrovascular events; lipoprotein cholesterol; cardiac failure (congestive and acute); qt interval abnormal; timi major bleeding requiring surgical interven-

tion; timi major bleeding requiring transfusion ≥4 units; timi minor bleeding; timi major bleeding requiring reoperation; timi major bleeding requiring transfusion ≥5 units; thrombotic/thromboembolic complications; rhythm abnormalities; ventricular conduction abnormalities; st-t elevation; major bleeding event; ischemic necrosis; angioedema of the face; angioedema of the larynx; cardiac vascular occlusion; cardiovascular occlusion; pacemaker intervention; overt bleeding; vasomotor reaction; urticarial reaction; electrocardiogram t wave abnormal; cerebrovascular hemorrhage; lowering of blood glucose; tachycardia atrial; angioedema of glottis; embolism-limb; angina requiring surgery; coronary heart disease mortality; obstruction of bile duct; myocardial toxicity; cardiac left ventricular function; cardiovascular reactions; monocyte count decreased; arterial limb thrombosis; arterial stenosis limb; risk of increased blood pressure; cardiac/cardiopulmonary arrest/failure; thrombosis portal vein; cardiac accessory conduction pathway disorders; cardiac rhythm irregularities; cardiovascular toxicity; vascular events; cardiovascular depression; ventricular ectopic beat; hypotension/postural hypotension; atrial tachyarrhythmias; infarction, brain; adrenocortical unresponsiveness; ventricular pause; angina attacks; thrombosis pulmonary; risk of fetal toxicity; non-hemolytic anemia; heart valve regurgitation

## A.4 Molecule 4

SMILES: : Cc1cc2nn(CC(=O)NCC(=O)NCc3ccc4c(c3)-c3ccccc3C4)c(=O)n2c(N2CCc3ccccc3C2)n
contusions; eczema, dyshidrotic; anaphylactoid syndrome of pregnancy; porphyria non-acute; infusion related reaction; infusion site erythema; infusion site induration; infusion site pain; infusion site swelling; infusion site phlebitis; infusion site pruritus; infusion site reaction; infusion site infection; infusion site bruising; infusion site rash; infusion site cellulitis; infusion site irritation; infusion site necrosis; infusion site thrombosis; infusion site urticaria; infusion site discomfort; catheter site haemorrhage; anaphylactoid shock; hypnagogic hallucination; dermatitis exfoliative generalised; methotrexate toxicity; exacerbation of asthma; porphyrias, acute; rashes, eruptions and exanthems nec; autoimmune hemolytic anemia; compartment syndromes; migraine with aura; eczema herpeticum; complex regional pain syndrome type ii; exacerbation of chronic obstructive pulmonary disease; stress fractures; medication overuse headache; disease exacerbation; corticotroph adenoma; exacerbation

of copd; ige-mediated hypersensitivity; filtering blebs; thromboembolic diseases; disorder of the urea cycle; catheter related complications; initial insomnia; disorder of ejaculation; eczema, dermatitis; catheter thrombosis; generalized muscle spasm; worsening of sleep apnea; infusion site burning; exacerbation of dyspnea; exacerbation of pre-existing diabetes mellitus; hypersensivitiy syndrome; elevation of gastric hcl; exacerbation of systemic lupus erythematosus; onset of pseudomembranous colitis symptoms; worsening of psoriasis; reactivation of herpes zoster; exacerbation of psoriasis; exacerbation of convulsions; exacerbation of porphyric symptoms; worsening of angina pectoris; worsening of arterial insufficiency; temporary blurred vision; impairment of performance of routine activities; temporary unilateral loss of vision; worsening of congestive heart failure; elevation in liver function tests; infusion related pain; exacerbation of inflammatory bowel disease; activation of mania; generalized exfoliative erythroderma; exacerbation of psychotic symptoms; minimal decreased rbcs; generalized spasms; reactivation of latent infections; exacerbation of psychosis; vein pigmentation; exacerbation of recurrent herpes labialis; exacerbation of angina; exacerbation of cough; stimulation of urinary bladder with spontaneous voiding; substernal chest pain; asthma symptoms; activation of pre-existing peptic ulcer; thromboembolic disorders; exacerbation of epilepsy; joint and muscle stiffness; retention of serum electrolytes; inhibition of gonadotropin secretion; gustatory sense diminished; severe sclerosis of the skin and subcutaneous tissues; impaired adaptation to dark; eczematous eruptions; excess mucus or phlegm; activation of mania/hypomania; migraine aggravated; generalized burning; protrusion of the tongue; exacerbation of heart failure; anaphylactic reactions with injection; elevation in serum levels of skeletal muscle enzymes; exacerbated bradycardia in sick sinus syndrome; worsening of rosacea; elevation of cerebrospinal fluid pressure; exacerbation of psychoses; worsening of preexisting hypertension; worsening of heart failure; contusions and hematomas; worsening of urinary retention; exacerbation of headache; activation of systemic lupus erythematosus; puffy eyes; risks with concomitant use of antiretroviral drugs; signs or symptoms of urinary tract irritation; exacerbation of preexisting ulcer disease; low systolic blood pressure; worsening of the conjunctivitis; signs and symptoms of eye allergy; generalized numbness; exacerbation of arthritis; activation of latent iritis; exacerbation of peptic ulcer disease; thrombophlebitis of the leg; exacerbation of symptoms of myasthenia gravis; elevation in prothrombin; eczematoid eruption; weakness in the legs; symptoms of hypocorticism; worsening of the depression; cyptosporidiosis; reduction of left ventricular ejection factor; activation of latent horner's syndrome; eczema/rash/urticaria; worsening of organic brain

syndrome; worsening of underlying illness; exacerbation of hyperphosphatemia; worsening of diabetes mellitus; disrupted body temperature regulation; exacerbation of joint symptoms; reactivation of hepatitis b virus infection; activation of mania or hypomania; infusion site thrombophlebitis; existing infections worsened; exacerbation of kearns-sayre syndrome; anaphylaxis/angioedema; weakness of the legs; exacerbation of chorea; exacerbation of preexisting pulmonary infection; worsening of narrow angle glaucoma; exacerbation of congestive heart failure; new onset diabetic macular edema; worsening of diabetic macular edema; exacerbation of parkinsonian symptoms; exacerbation of pruritus; generalized hives; generalized warmth; prolonged activated partial thromboplastin time; potentiation of antihypertensive effect; persistent or severe hand-and-foot syndrome; exacerbation of raynaud's syndrome; exacerbation of hepatitis; exacerbation of porphyria; exacerbation of motor and phonic tics; exacerbation of viral ocular infections; asthma attack; psoriasis (pso); exacerbation of mood disorders; exacerbation of chronic hepatitis b; worsening of pre-existing narcolepsy; hyperglycemia (grade 3 or 4); partial permanent deafness; exacerbation of cutaneous lupus erythematosus; symptoms of hypercorticism; glucocorticoid related adverse effects; deterioration, clinical; stimulation caused by adrenergic beta-2 agonist effect; exacerbation of endometriosis; exacerbation of migraine; exacerbation of systemic lupus erythematosus (sle); exacerbation of hepatic hemangioma; exacerbation of, pre-existing lesions; exacerbated psychotic disorders; anaphylactic reactions with hypotension; exacerbation of reflux; exacerbation of chronic lung disease; ild/pneumonitis

atrial flutter; atrial fibrillation; ventricular fibrillation; mitral valve insufficiency; eclampsia; albuminuria; echolalia; heartburn; infarction; cardiotoxicity; heart injuries; diaper rash; cardiac tamponade; cardiomegaly; myocardial rupture; atrial rupture; myocardial haemorrhage; coronary artery dissection; coronary artery thrombosis; coronary artery embolism; carditis; atrial thrombosis; cardiac fibrillation; cardiac flutter; exercise tolerance decreased; catheter site pain; metabolic encephalopathy; adrenergic syndrome; athetosis; hypertensive encephalopathy; apathy; bilirubinuria; postmenopausal haemorrhage; vasoconstriction; hypertensive crisis; thrombophlebitis; arterial thrombosis; vasospasm; arteriosclerosis; digoxin toxicity; cardiac murmur; cardiac conduction disorders; cerebrovascular embolism and thrombosis; cerebrovascular venous and sinus thrombosis; pruritus nec; coronary artery disorders; vascular hypertensive disorders; myocardial ischemia; angina pectoris; myocardial infarction; vascu-

lar diseases; heart failure; cerebrovascular accident; ventricular arrhythmia; cardiovascular; ischemic heart disease; ischemic stroke; embolism and thrombosis; cardiac arrest; ventricular dysfunction; cardiac complications; vasodilation; coronary artery stenosis; heart block; cardiovascular abnormalities; major depressive episode; valvular heart disease; coronary artery insufficiency; cardiovascular complications; coronary occlusions; metabolic disorders; cerebrovascular insufficiency; major bleeding; cardiac; cardiovascular risk; cardiac failure; arterial thromboembolism; lipid disorders; myocardial depression; ventricular hypertrophy; cardiovascular mortality; risk of cardiac arrhythmias; arterial thromboembolic events; non cirrhotic portal hypertension; atherosclerosis risk; coronary artery occlusion; metabolic and nutritional complications; pacing; atrial arrhythmia; cardiovascular events; cardiac disorder; coronary artery atherosclerosis; coronary artery vasospasm; calcium deficiency; cardiac decompensation; metabolic and nutritional disorders; ventricular tachyarrhythmias; cardiac rhythm disorders; cardiovascular disorders; stenosis of bile duct; aspirin exacerbated asthma; cardiovascular disease (cvd); cardiac amyloidosis; lipid disorders and lipid measurement; major adverse cardiac events; ischemic cerebrovascular accident; ecg changes; anginal pain; vessel puncture site pain; major adverse cardiovascular events; cardiac function impaired; supraventricular arrhythmias; cardiac dysfunction; cerebrovascular diseases; metabolic disturbances; diaper dermatitis; hemodynamics instability; cardiorespiratory arrest; cardiac output decreased; ejection fraction decreased; exertional dyspnea; myocardial reinfarction; myocardial ischemia or infarction; hypertensive reaction; hypertensive episodes; st segment changes; ischemic events; new onset diabetes mellitus; ventricular ectopic activity; cardiac valvulopathy; maculopapular and erythematous rashes; arrhythmia aggravated; minor hematuria; coronary artery bypass graft hemorrhage; cardiac conduction disturbances; cardiovascular collapse; arterial events; arteriospasm; antidiuretic effects; ischemic injury; convulsive events; angina/angina-like pain; myocardial ischemia/infarction; systolic ejection murmur; cerebrovascular reactions; angina/myocardial infarction; ecg pattern changes; cardiac abnormalities; thrombophlebitis deep; exertional hypotension; heart rate elevated/abnormal; elevation of serum triglyceride levels; vasodilation (usually flushing); rebound hypertension; angina-like chest pain; major noncerebral bleeding; cardiovascular depression following ect; kawasaki-like syndrome; electrocardiogram st segment depression; cardio-respiratory collapse; electrocardiogram qrs complex abnormal; myocardial hypertrophy; postmenopausal vaginal bleeding; cardiovascular side effects; cardiac conduction abnormalities; cardiovascular thrombotic events; metabolic effects; vasomotor response; new or increased angina pectoris; cardiac irreg-

ularities; thrombophlebitis leg; cerebrovascular spasm; calcium taste; cardiac-clinical; myocardial insufficiency; fatal outcomes have been reported; ischemic cardiovascular events; st elevated; cardiac rhythm abnormalities; autonomic instability; elevation of aldosterone; cardiac-related fluid retention; electrocardiogram pr prolongation; myocardial infarction/ cerebral infarction; lowering of estrogen; ecg and eeg changes; ischemic cerebrovascular events; lipoprotein cholesterol; cardiac failure (congestive and acute); qt interval abnormal; timi major bleeding requiring surgical intervention; timi major bleeding requiring transfusion $\geq$4 units; timi minor bleeding; timi major bleeding requiring reoperation; timi major bleeding requiring transfusion $\geq$5 units; thrombotic/thromboembolic complications; rhythm abnormalities; ventricular conduction abnormalities; st-t elevation; major bleeding event; ischemic necrosis; angioedema of the face; angioedema of the larynx; cardiac vascular occlusion; cardiovascular occlusion; pacemaker intervention; overt bleeding; vasomotor reaction; urticarial reaction; electrocardiogram t wave abnormal; cerebrovascular hemorrhage; lowering of blood glucose; tachycardia atrial; angioedema of glottis; embolism-limb; angina requiring surgery; coronary heart disease mortality; obstruction of bile duct; myocardial toxicity; cardiac left ventricular function; cardiovascular reactions; monocyte count decreased; arterial limb thrombosis; arterial stenosis limb; risk of increased blood pressure; cardiac/cardiopulmonary arrest/failure; thrombosis portal vein; cardiac accessory conduction pathway disorders; cardiac rhythm irregularities; cardiovascular toxicity; vascular events; cardiovascular depression; ventricular ectopic beat; hypotension/postural hypotension; atrial tachyarrhythmias; infarction, brain; adrenocortical unresponsiveness; ventricular pause; angina attacks; thrombosis pulmonary; risk of fetal toxicity; non-hemolytic anemia; heart valve regurgitation

## A.5 Molecule 5

SMILES: Cc1ccc(-c2cccc(CNC(=O)c3ccccc3)c2)cc1

bronchopneumonia; pneumonia, pneumococcal; tracheitis; ventricular flutter; pulmonary infarction; pneumatosis cystoides intestinalis; pulmonary eosinophilia; respiratory paralysis; respiratory distress syndrome, adult; pneumothorax; cough; ventricular hypokinesia; grey syndrome neonatal; endocarditis; chest discomfort; ventricular asystole; throat irritation; fever neonatal; reye's syndrome; bronchospasm; bronchioli-

tis; pneumonia respiratory syncytial viral; laryngitis; pulmonary sepsis; pneumonia; bronchitis; pneumonia fungal; respiratory moniliasis; bronchopulmonary aspergillosis; lung infection pseudomonal; pneumonia legionella; escherichia sepsis; pulmonary toxicity; pulmonary oil microembolism; epidural lipomatosis; chondrolysis; spondylitis; breath holding; laryngospasm; respiratory tract congestion; respiratory tract irritation; respiration abnormal; bronchial irritation; sputum increased; sputum discoloured; productive cough; cough decreased; respiratory depth increased; pulmonary arteriopathy; pulmonary microemboli; pulmonary alveolar haemorrhage; lung consolidation; pneumonitis; wheezing; bronchospasm paradoxical; bronchial obstruction; bronchiectasis; erythrodermic psoriasis; bronchoconstriction; fever; bronchial airway hyperreactivity; nose edema; breathing abnormalities; pulmonary embolism; respiratory distress syndrome; pulmonary fibrosis; chest pain; respiratory insufficiency; pleural effusions; airway obstruction; bronchioloalveolar carcinoma; bronchial asthma; respiratory failure; respiratory distress; respiratory depression; influenza-like illness; respiratory disorders; respiratory diseases; pulmonary arterial hypertension (pah); pulmonary obstruction; pulmonary edemas; right heart failure; respiratory symptoms; pulmonary veno occlusive disease; pulmonary mass; lung disorder; pulmonary infections; ventilatory depression; bronchopleural fistula; sputum; pulmonary hypertension (ph); cough, acute; respiratory alkalosis; respiratory system abnormalities; pulmonary inflammation; respiratory distress syndrome, infant; pulmonary alveolar proteinosis (pap); ventricular failure, right; pulmonary haemorrhage; lung, carcinoma; respiratory tract hemorrhage; pulmonary congestion; lung cancers; lung function changes; chest disorders; pulmonary diseases, obstructive; respiratory muscle paralysis; ventricular tachycardia (vt); airway complication of anesthesia; lung edema; pulmonary infiltrates; pulmonary events; chest tightness; chest congestion; breath odor; pulmonary interstitial infiltrates; dyspnea aggravated; throat discomfort; lung edema disorder; respiratory inadeqaucy; sweat gland; bronchospasm aggravated; pulmonary problem; nervousness/irritability; throat inflammation; chest pain/discomfort; nose pain; escherichia bacteremia; bronchial constriction; throbbing headache; pulmonary hypersensitivity reactions; allergic-type respiratory symptoms; chest pain/pressure/angina; respiratory flu; pulmonary infiltration with eosinophilia; expiratory wheezing; bronchial wheezing; fever and headache; chest wall rigidity; chest wall pain; respiratory acidosis during weaning; fever or influenza-like illness; ecg change consistent with myocardial ischemia; vagolysis; respiratory and cns depression; chest symptoms; throat/neck symptoms; respiratory side effects; perlèche; bronchiolar constriction; fetal anticonvulsant syndrome; pulmonary hypotension; meniere's syn-

drome; pulmonary-clinical; pulmonary allergy; chest flushing; innapropiate secretion of antidiuretic hormone; allergic pericarditis; delirium hallucinations; pulmonary vascular occlusion; cough suppression; cough experience post-inhalation; intramuscular hemorrhage; throat pressure; asthma-related death; respiratory impairment; respiratory symptom-related adverse reactions; swellin of lower legs; respiratory wheezing; respiratory arrest (disorder); pulmonary changes; pulmonary symptoms; mechanical irritation at application site; mechanical cardiac dysfunction; bronchial anastomotic dehiscence; throat infections; erythroleukemia; chest mass; thoracic spine pain; asthmatic type dyspnea; respiratory rate increased; fluid overload/retention; respiratory acidosis/alkalosis; respiratory arrest/failure; pulmonary infarct; endocarditis caused by staphylococcus aureus; pneumonia (grade $\geq$3); pulmonary reactions; pleural effusion (grade $\geq$3); respiratory function impaired; bronchorrhea; pulmonary hypertension, primary; pulmonary fibrosis interstitial; lung hemorrhage; dyspepsia/heartburn; respiratory rate decreased; expiratory reserve volume decreased; chest infections

contusions; eczema, dyshidrotic; anaphylactoid syndrome of pregnancy; porphyria non-acute; infusion related reaction; infusion site erythema; infusion site induration; infusion site pain; infusion site swelling; infusion site phlebitis; infusion site pruritus; infusion site reaction; infusion site infection; infusion site bruising; infusion site rash; infusion site cellulitis; infusion site irritation; infusion site necrosis; infusion site thrombosis; infusion site urticaria; infusion site discomfort; catheter site haemorrhage; anaphylactoid shock; hypnagogic hallucination; dermatitis exfoliative generalised; methotrexate toxicity; exacerbation of asthma; porphyrias, acute; rashes, eruptions and exanthems nec; autoimmune hemolytic anemia; compartment syndromes; migraine with aura; eczema herpeticum; complex regional pain syndrome type ii; exacerbation of chronic obstructive pulmonary disease; stress fractures; medication overuse headache; disease exacerbation; corticotroph adenoma; exacerbation of copd; ige-mediated hypersensitivity; filtering blebs; thromboembolic diseases; disorder of the urea cycle; catheter related complications; initial insomnia; disorder of ejaculation; eczema, dermatitis; catheter thrombosis; generalized muscle spasm; worsening of sleep apnea; infusion site burning; exacerbation of dyspnea; exacerbation of pre-existing diabetes mellitus; hypersensivitiy syndrome; elevation of gastric hcl; exacerbation of systemic lupus erythematosus; onset of pseudomembranous colitis symptoms; worsening of psoriasis; reactivation of herpes zoster; exacerbation of psoriasis; exacerbation of convulsions; exacerbation of porphyric symptoms; wors-

ening of angina pectoris; worsening of arterial insufficiency; temporary blurred vision; impairment of performance of routine activities; temporary unilateral loss of vision; worsening of congestive heart failure; elevation in liver function tests; infusion related pain; exacerbation of inflammatory bowel disease; activation of mania; generalized exfoliative erythroderma; exacerbation of psychotic symptoms; minimal decreased rbcs; generalized spasms; reactivation of latent infections; exacerbation of psychosis; vein pigmentation; exacerbation of recurrent herpes labialis; exacerbation of angina; exacerbation of cough; stimulation of urinary bladder with spontaneous voiding; substernal chest pain; asthma symptoms; activation of pre-existing peptic ulcer; thromboembolic disorders; exacerbation of epilepsy; joint and muscle stiffness; retention of serum electrolytes; inhibition of gonadotropin secretion; gustatory sense diminished; severe sclerosis of the skin and subcutaneous tissues; impaired adaptation to dark; eczematous eruptions; excess mucus or phlegm; activation of mania/hypomania; migraine aggravated; generalized burning; protrusion of the tongue; exacerbation of heart failure; anaphylactic reactions with injection; elevation in serum levels of skeletal muscle enzymes; exacerbated bradycardia in sick sinus syndrome; worsening of rosacea; elevation of cerebrospinal fluid pressure; exacerbation of psychoses; worsening of preexisting hypertension; worsening of heart failure; contusions and hematomas; worsening of urinary retention; exacerbation of headache; activation of systemic lupus erythematosus; puffy eyes; risks with concomitant use of antiretroviral drugs; signs or symptoms of urinary tract irritation; exacerbation of preexisting ulcer disease; low systolic blood pressure; worsening of the conjunctivitis; signs and symptoms of eye allergy; generalized numbness; exacerbation of arthritis; activation of latent iritis; exacerbation of peptic ulcer disease; thrombophlebitis of the leg; exacerbation of symptoms of myasthenia gravis; elevation in prothrombin; eczematoid eruption; weakness in the legs; symptoms of hypocorticism; worsening of the depression; cyptosporidiosis; reduction of left ventricular ejection factor; activation of latent horner's syndrome; eczema/rash/urticaria; worsening of organic brain syndrome; worsening of underlying illness; exacerbation of hyperphosphatemia; worsening of diabetes mellitus; disrupted body temperature regulation; exacerbation of joint symptoms; reactivation of hepatitis b virus infection; activation of mania or hypomania; infusion site thrombophlebitis; existing infections worsened; exacerbation of kearns-sayre syndrome; anaphylaxis/angioedema; weakness of the legs; exacerbation of chorea; exacerbation of preexisting pulmonary infection; worsening of narrow angle glaucoma; exacerbation of congestive heart failure; new onset diabetic macular edema; worsening of diabetic macular edema; exacerbation of parkinsonian symptoms; ex-

acerbation of pruritus; generalized hives; generalized warmth; prolonged activated partial thromboplastin time; potentiation of antihypertensive effect; persistent or severe hand-and-foot syndrome; exacerbation of raynaud's syndrome; exacerbation of hepatitis; exacerbation of porphyria; exacerbation of motor and phonic tics; exacerbation of viral ocular infections; asthma attack; psoriasis (pso); exacerbation of mood disorders; exacerbation of chronic hepatitis b; worsening of pre-existing narcolepsy; hyperglycemia (grade 3 or 4); partial permanent deafness; exacerbation of cutaneous lupus erythematosus; symptoms of hypercorticism; glucocorticoid related adverse effects; deterioration, clinical; stimulation caused by adrenergic beta-2 agonist effect; exacerbation of endometriosis; exacerbation of migraine; exacerbation of systemic lupus erythematosus (sle); exacerbation of hepatic hemangioma; exacerbation of, pre-existing lesions; exacerbated psychotic disorders; anaphylactic reactions with hypotension; exacerbation of reflux; exacerbation of chronic lung disease; ild/pneumonitis

molluscum contagiosum; hypertrophy, right ventricular; tachycardia, ectopic junctional; fecal impaction; diabetic ketoacidosis; embolism, amniotic fluid; anemia, sideroblastic; linear iga bullous dermatosis; osteomalacia; genu valgum; vocal cord paralysis; cataplexy; meningitis, aseptic; cluster headache; tetany; refeeding syndrome; strongyloidiasis; encopresis; gagging; hot flashes; earache; chills; pituitary apoplexy; fat necrosis; wounds, gunshot; burns, chemical; fractures, compression; heat exhaustion; nails, ingrown; phobic disorders; bulimia nervosa; ductus arteriosus premature closure; dyspnoea paroxysmal nocturnal; craniosynostosis; limb reduction defect; imperforate oesophagus; ear malformation; ear pruritus; ear discomfort; ear congestion; ear canal erythema; ototoxicity; ear haemorrhage; ear disorder; acoustic neuritis; vertigo; insulin autoimmune syndrome; menstruation delayed; diabetic coma; retinal depigmentation; oculogyric crisis; oculomucocutaneous syndrome; eyelids pruritus; peritoneal cloudy effluent; lip haemorrhage; neonatal intestinal perforation; oesophagitis ulcerative; oesophageal ulcer; bruxism; oesophageal pain; impaired gastric emptying; oesophageal disorder; colpocele; hunger; fatigue; catheter site related reaction; catheter site erythema; instillation site pain; mixed liver injury; panniculitis; pharyngeal oedema; mycobacterium avium complex infection; penile infection; cervicitis; bullous impetigo; citrate toxicity; frostbite; road traffic accident; wound haemorrhage; eschar; joint hyperextension; electrolyte depletion; lipohypertrophy; lipodystrophy acquired; costochondritis; cogwheel rigidity; osteosclerosis; coccydy-

nia; fallopian tube cyst; spinal cord disorder; formication; postictal state; areflexia; parkinsonian rest tremor; osmotic demyelination syndrome; blighted ovum; ovarian disorder; cervix disorder; acquired phimosis; penile oedema; nipple swelling; breast discomfort; nipple pain; breast swelling; breast discharge; breast enlargement; choking; rosacea; ingrown hair; hair texture abnormal; hair growth abnormal; hair colour changes; anhidrosis; acne fulminans; acne; eczema asteatotic; ischaemia; phlebosclerosis; agitated psychotic state; thromboembolism; phlegm; upset stomach; wound site; mixed manic depressive episode; circulatory collapse and shock; acute radiation sickness; calcium crystalluria; labyrinthine disorder; pancreatic enzymes increased; anaplastic large cell lymphomas t- and null-cell types; lactation disorders; joint dislocations; lipodystrophies; menstruation with decreased bleeding; menstruation with increased bleeding; nail and nail bed conditions (excl infections and infestations); potassium imbalance; breast disorders; gastrointestinal neoplasms malignant and unspecified; paraproteinemias; pure red cell aplasia; tuberculosis, mycobacterium infection; postherpetic neuralgia; nervous system malformations; hip fracture; weight changes; barrett's esophagus; vasovagal syncope; learning disorders; mitochondrial diseases; wound infections; condylomata acuminata; mitochondrial toxicity; hip replacement surgery; urolithiasis; jarisch herxheimer reaction; breast pain; glycemic control; knee replacement; leukemias; nephrogenic diabetes insipidus; wound dehiscence; osteonecrosis of the knee; glucose tolerance; organ dysfunction; circulatory collapse; lobar pneumonia; necrotizing enterocolitis (nec); discoid lupus erythematosus (dle); breast cysts; melancholic depression; foot drop; gait instability; tolerance; wound complications; eyelid diseases; back injuries; osteonecrosis of the jaw; hip dysplasia; bladder spasms; falls, accidental; perioperative blood loss; menstrual spotting; spinal cord injuries (sci); dilutional hyponatremia; acneiform rash; keratosis pilaris (kp); gouty arthritis; allergic rhinitis (ar); parkinson's disease (pd); immobility; muscle spasms; gait or balance disorder problems; central nervous system infections; mammary tumor; fluid over-load; ophthalmologic complications; diabetic macular edema (dme); cauda equina syndrome; facial paresis; muscle tightness; androgenetic alopecia (aga); balance disorders; foot pain; post procedural discharge; duchenne's muscular dystrophy (dmd); pneumocystis jirovecii pneumonia; clostridial infection; fluid imbalance; ureteral spasm; bullous pemphigoid (bp); hiatus hernia; atelectasis, postoperative; henoch schönlein purpura; allergic conjunctivitis (ac); achalasia, esophageal; gait, shuffling; sleepwalking; nail toxicity; organ failure, multiple; olfactory dysfunction; venous thrombosis deep (limbs); sudden unexpected death in epilepsy (sudep); eustachian tube dysfunction (etd); osteopenia (disorder); adynamic

bone disease; lactate blood increase; nail changes; blurred vision; diaphoresis; chills or fever; parkinsonian symptoms; pharyngolaryngeal pain; intermenstrual bleeding; micturition frequency; neck rigidity; obsessions; nasal stinging; calming; coordination disturbance; spontaneous posttraumatic bleeding; bullous eruptions; prolonged bone marrow hypercellularity; reactivated tuberculosis; reversible oligospermia; motor loss; wound healing abnormalities; bullous skin reactions; groin bleeding; groin hematoma; bruising at implant site; change in corneal curvature; hair texture changes; hyperglycinemia; mixed hepatitis; orofacial dyskinesia; petechiael hemorrhages; acneiform skin rash; facial and lip edema; thromboembolic phenomena; goiter production; fast eeg activity; vein distension; breast exisional biopsy; staggering gait; tightness in urinary bladder; bloated feeling; vaso-vagal syndrome; eczematoid dermatitis; glucosuria; unifocal premature ventricular contractions; fixed pupillary dilatation; orgasmic dysfunction; ear tightness; phobic neurosis; laryngitis/hoarseness; positive direct coombs tests; ear pain/discharge/erythema/swelling; strongyloidiasis hyperinfection; nail disorders; increase in serum urea nitrogen (bun) or creatinine; embryo-fetal toxicity; bladder contracture; increased microhematuria; increased urine specific gravity; prothrombin time prolongation; breast lump; colic pain; ears plugged; kneecap pain; wound drainage increased; itching tongue; increase in size of uterine leiomyomata; changes in cervical ectropion; nipple discharge; changes in breast size; back or hairy tongue; increased or decreased number of spermatozoa; increase in cephalin flocculation; virilization; eyelash changes; equilibrium disorders; tight feeling in head; significant prolongation of the qt interval; tracheal/tracheobronchial calcification; vasovagal episode; ears blocked; decreased systolic and diastolic blood pressure; chills occurring with fever; aseptic peritonitis; painful knees; increased frequency of urination; increased activated partial thromboplastin; ear, nose, and throat infections; orthostatic hypotension, hypotension, and/or syncope; microcytic anemia; trauma (various physical injuries); testes pain; uterine fibroids enlarged; weeping; tingling at the application site; asthma aggravated; dilatation of the pupil; tightness of chest and wheezing; unspecified auricular disorders; unspecified liver disorder; shortened partial thromboplastin time; shininess; spasm of the right ventricle infundibulum; penile fibrosis; penile rash; central nervous system disturbances; vasomotor collapse; musculoskeletal spasms; facial and finger puffiness; osmolality increased; changes in color vision; fatigability; neuromuscular (extrapyramidal) reactions; motor restlessness; aching of the tongue; adynamic ileus; mammary hyperplasia; ggt ¿10xuln; nipple tenderness; changes in co2 content; changes in prothrombin time; changes in differential counts; changes in sgpt; browache; schizophrenic/schizophreniform behaviour; fleeting joint

pain; runny eyes; loin pain; fat-soluble vitamin deficiency; valvulopathy; ear signs and symptoms; wounds and lacerations; unspecified oropharyngeal plaques; pituitary unresponsiveness; breathing disorders; quantitative red cell or hemoglobin defects; hypothalamus/pituitary hypofunction; acne and folliculitis; breast symptoms; virilizing changes; ear fullness; lacrimal dysfunction; tightness in the chest and throat; postprandial lightheadedness; total catheter events; defecation urgency; oral cavity discoloration; borborygmi; acne, generalized; eczematous rash; fullness of gi tract; hunger abnormal; focusing disturbances; trilineage bone marrow hypoplasia; phayngolaryngeal pain; bullous rashes; cochlear ototoxicity; fatty or oily stool; itching of skin; nail discoloration; transient rise in ldh; vortex keratopathy; spontaneous ecchymosis; bruising and local swelling at injection site; cochlear damage; central nervous system toxicity; oxyhemoglobin desaturation; nail bed changes; neuroconstipation; respiration depression; bruising easily; buffalo hump; ear infection-not otherwise specified; nail bed discoloration; ear debris; oral/pharyngeal edema; breast fibroadenosis; vertigo/dizziness; central nervous system bleeding; preexisting concomitant skin infection does not improve; need for coronary revascularization; skipped beats; psoriaform dermatitis; hand tremors; muscle hyperirritability; clonic movements of whole limbs; transient increases in ldh; vein induration; eczematoid rash; fecal compaction; apocrine gland disorder; nail infestation; balance difficulty; glucose decreased; ecchyhmosis; joint sprains; immediate wheal; dowsiness; timi major bleeding requiring inotropes; foamy urine; underactivity of the thyroid; limbal and conjunctival hyperemia; prostatic acid phosphatase increased; gynecomastia aggravated; neuroocular lesion; constriction of the chest; scales on lash; vasovagal reactions; increases in serum pancreatic amylase test; hair graying; 7th nerve paralysis; phosphorous decreased; reverse posterior leukoencephalopathy syndrome; bullous lesions; osmotic nephrosis of proximal tubular cells; ear symptoms; decreased estimated creatinine clearance; fall in pulse rate; oxygen desaturation; spinal nerve paralysis; insensitivity of the eye; oesophageal lesion; osteoarthritis aggravated; intact matrices in the feces; embedment; pressure in the chest; severe hyoernatremic dehydration; decreased serum total proteins; breast leakage; phosphate intoxication; marked perspiration; persistent sneezing; ear irritation; adenocarcinoma of the chest; rolling of the eyes; decrease in thyroid-stimulating hormone (tsh); electrocardiogram st-t segment abnormal; glucose urine present; bun and creatinine elevations; venous thrombosis (disorder); cold and clammy skin; wheal and flare over the vein with iv injection; phlebitis following iv injection; disorientation for time; disorientation for place; aching and sore throat; lichen planus-like eruptions; balance disturbances; changes in amount of cervical;

changes in salivation; petit mal convulsions; polyphagia; graft versus host disease (acute and chronic); new onset status epilepticus; chills/cold; sids; breast issues; decrease in serum high-density lipoprotein (hdl); costovertebral pain; pain in urethra; elevation or depression of blood sugar levels; electrocardiogram st-t change; ammonia increased; digital vasospasm; decreased sleep requirement; intensification of atrioventricular block; growth suppression of fetus; ischemic and hemorrhagic events; evidenced by paresthesia; increased requirements for oral hypoglycemic agents; muscle, joint pains; wound-related reactions; pharyngitis/ nasopharyngitis; reduced analgesic effect; dyspepsia and gastritis; strong smelling urine; eating, binge; labyrinth disorders; osteodynia; petechiae and ecchymoses; decreased carbohydrate and glucose tolerance; mottling of skin; response supression; pediatric safety and effectiveness not established; methemoglobinemia, acquired; fluid and electrolyte imbalance; "hot spells"; electrocardiogram repolarisation abnormality; tardive dyskinesia (td); lipase ¿=3.0 times uln; fasted cholesterol ¿=300 mg/dl; fasted triglycerides ¿500 mg/dl; change in lipids from baseline; decreased thyroxine-binding globulin; prostate disorders; graft versus host disease (cgvhd); estradiol increased; ear precipitate (residue); valve stenoses, aortic; muscle atrophy, proximal; acute exacerbation of asthma; prolonged anovulation; back pain (grade $\geq$3); increased amylase (grade 3 or 4); increased alkaline phosphatase (grade 3 or 4); bruising (grade 1); obtundation; creatine kinase ¿10x uln; phosphate decreased (grade 2-4); ulcers, leg; decreased leukocytes (grade $\geq$3); increased bilirubin (grade $\geq$3); increased ast (grade $\geq$3); decreased potassium (grade $\geq$3); decreased albumin (grade $\geq$3); decreased magnesium (grade $\geq$3); decreased sodium (grade $\geq$3); increased alt (grade $\geq$3); decreased lymphocytes (grade $\geq$ 3); fracture of pelvis (disorder); fishbane reaction; osteoporosis (grade 3-4); osteopenia (grade 3-4); breast pain (grade 3-4); endometrial proliferation disorder (grade 3-4); musculoskeletal stiffness (grade 3-4); metabolic disorders (grade 3-4); phosphorous decreased (grade 3); phosphorous decreased (grade 4); adherence to vaginal wall; carbohydrate and lipid effects; bleaching of hair pigment; ggt increased ¿5x uln; osteonecrosis of the ankle; osteonecrosis of the wrist; osteonecrosis of the femur; circulatory disturbance; parkinson's disease aggravated; gall bladder pain; increased sex hormone binding globulin; increased thyroxine-binding globulin; change of the ocular lens; tactile disturbance; burns of the skin; hyperglycaemic hyperosmolar nonketotic syndrome; estradiol decreased; aphthous ulcer of the buccal mucosa; telangiectasias of the face; dermatitis of the eyelid; osteonecrosis of the hip; allergic reaction of the lung; ulceration of the anus

bradycardia; lactose intolerance; brachydactyly; reflex, abnormal; pellagra; anticholinergic syndrome; genitalia external ambiguous; oscillopsia; pupils unequal; heterophoria; foreign body in eye; sensitivity of teeth; teeth brittle; glossodynia; buccoglossal syndrome; hypertrophy of tongue papillae; faeces soft; paradoxical drug reaction; sensation of foreign body; discomfort; polyserositis; pallor; mazzotti reaction; polymyositis; parainfluenzae virus infection; carcinogenicity; cinchonism; stoma site irritation; excoriation; hypotonia neonatal; anterograde amnesia; tonic clonic movements; clonic convulsion; opisthotonus; clumsiness; placental disorder; dependence; claustrophobia; cylindruria; oedema genital; hypoxemia; scaly conditions; stomatitis and ulceration; investigations; veno-occlusive disease; hypersensitivity vasculitis; irregular heart rate; epiphora; staphylococcal sepsis; antisocial behavior; angle-closure glaucoma; phenytoin toxicity; alteration of cognitive function; pathological fractures; irritation; varicella-zoster virus; accommodation disorders; reactions; systemic inflammatory response syndrome (sirs); obstipation; bradyarrhythmias; sensitization; sicca syndrome; intrahepatic, cholestasis; variegate porphyria; transaminitis; clamminess; irritability or anxiety or nervousness; reflexes decreased; congestion of upper airway; paradoxical excitation; lightheadedness; aggravated multiple sclerosis; hypalgesia; intoxicated feeling; extracardiac fibrotic reactions; enlargement of the lips; general fatigue; reactivation of herpes simplex; excitation; hypersensitive skin reactions; reflexes increased; aggravation of porphyria; irreversible bone marrow failure; aggravation of angioedema; aggravation of hereditary angioedema; aggravation of varicose veins; intolerance to contact lenses; chlolelithiasis; irreversible blindness; vacuolation of erythrocytes; hypporeflexia; paradoxical excitement; dilatation of the colon; paradoxical sinus bradycardia; coordination difficulties; stimulation; general edema; heaviness in limbs; aggravated; lack of effect; slightly clouded sensorium; trembling sensation; staphylococcal bacteremia; tremulousness; variations in heart rate; aggravation of disseminated lupus erythematosus; aggravation of hypertension; aggravation of coronary artery disease; microscopic deposits in the urine; genital pruritus; photosensitivity skin reaction; transitory deafness; sore or dry mucous membranes; psychic disturbances; soreness of gum; retention of chloride; retention of potassium; retention of calcium; retention of inorganic phosphates; hypersensitivity/allergic reactions; paradoxical reactions; argumentativeness; temporary stinging; local tissue necrosis; excessive gas; irritation at application site; crusting of skin; collapse; prolonged musculoskeletal block; crusting; induration at the site of im injection; basophilia and hyperhistaminemia; enamel hypoplasia; slight swelling of the breast; drug idiosyncra-

cies; photosensitive rash; brittle fingernails; clonus on flexing foot; temporary infertility; slight elevation in bun; abnormality of cerebrospinal fluid proteins; paradoxical exacerbation of psychotic symptoms; reactivated peptic ulcer; hypersensitivity syndrome; exanthem; photosensitivity allergic reaction; underventilation due to cephalad extension; aggravation of angina pectoris; hypnagogic effects; abnormalities of urine specific gravity; symptoms of upper respiratory tract infection; violet papules; junctional rhythm; disturbances of cardiorhythm; sensitization reaction; drunkenness; reactions decreased; general discomfort; difficulties in balance; enlargement of lump in breast; irrigation of unintended structures or cavities; accidental contamination; prolonged menstruation; reversal of analgesia; reversible mental depression; tonic movements; clonic movements; superimposed ear infection; difficulties with visual accomodation; concomitant skin infection develops; pathological fracture of long bones; psychic changes; hypersensitivity hepatitis; redistribution of body fat; accumulation of body fat; slight increases in postoperative serum glucose; disturbances in consciousness; neutropenia-related death; arterial vascular occlusive events; sensation of heaviness; hypercholesterolemia/hyperlipidemia; genital pruritus male; taste aversion; discomfort of nasal cavity; reversible leukopenia; underventilation; arterial occlusive event; displacement of arterial plaques; citomegalovirus infection; aggravate pre-existing diabetes; slight deafness; heaviness in extremities; heaviness of legs; exanthemas nec; stomatitis/pharyngitis; histamine release; subjective cardiac rhythm disturbances; bullae at application site; slight reductions in platelet counts; aggravation of psychoses; sensation of pain; aggravation of symptoms of dementia; adams-stokes attacks; hypothesia; prolongation of sedation; muzziness; aggravation of parkinsonism; aggravated hostility; abnormality of cerebrospinal fluid; aggravation of infections; psychic disorders; halo around lights; temperature instability; inadvertent staining of vitreous face; concomitant conditions; anticholinergic adverse reaction; harm; photosensitive dermatitis; investigations (grade 3-4); pupillary disorders; extra-renal hyperazotemia; temporary discontinuation of therapy; retention phosphate; signs and symptoms glucocorticoid related adverse effects; aggravated allergies; viscid plugs; weak peripheral pulses; pupils poorly reactive to light

salpingitis; takotsubo cardiomyopathy; colic; aerophagy; anaemia heinz body; haemoconcentration; low set ears; orbital oedema; meibomianitis; madarosis; mesenteric vascular insufficiency; mesenteric arterial occlusion; haemorrhagic erosive gastritis; peritonitis; thirst; cold sweat; crying; haemorrhagic urticaria; perineal abscess; over-

growth fungal; poisoning; sleep terror; slow speech; masked facies; belligerence; fear of open spaces; fear; tension; haemorrhage urinary tract; pollakiuria; haemorrhage subcutaneous; sticky skin; shock; cold; redness; detached; dryness; warmth; gout attack; sting of the eye; asthenic conditions; cataracts; tension headache; sore throat; cold sore; heat stroke; back strain; still births; apraxias; delayed recovery from anaesthesia; tissue damage; hpv-related carcinoma; worrying; dry socket syndrome; pelvic symptoms; paradoxical anxiety; stinging of the skin; hangover effect; high energy; stinging or burning; sprue-like enteropathy; bigeminy; cold or flu syndrome; integument and mucus membrane toxicity; excessive hypotension; bigeminal rhythms; stinging sensation in eyes upon instillation; itchy eyes and lids; u waves; proarrhythmia; fearfulness; cold sensations in hands and feet; bad taste following instillation; reddening; mind racing; fever and/or chills; lowered t-waves; dryness of the pharynx; staring episode; shallow breathing; dry scaly skin; sgpt; stinging of lips; pelvic cramping; sting on injection; reduced folic acid; loose or frequent stools; shallow respirations; cataract (not specified); bad dreams; cold feet; enteritis at all levels; stinging at application site; itching at wart site; dry and breaking hair; marked diuresis; cold or numbness; heat rash; shock-like coma; hives/ itching; migratory polyarthralgia; dry mouth/nose; punctate epithelial staining; grossly obvious gingivitis; patch non-adhesion; high urine wbc/hpf; high sgot; sense of pelvic pressure; warm feeling in vagina; tightness; warm sensations; warm/cold sensations; mesenteric arterial thrombosis; tightness and rigidity; cold sensations; dryness of hair and scalp; bad taste in mouth; dryness around the eye; lowered blood pressure; slow urination; dryness of the mucous membranes; stinging sensation; stinging of skin; thirst disturbance; serum phenytoin alteration; ideation; high pitched cry; low absolute neutrophil count; bad taste/metallic taste; erratic blood pressure; acne-like rash; foot tapping; coldness of the extremities; reactive hypertension; shortening of refractory periods; thirst and dehydration; sticky sensation; low hemoglobin counts; grimace; watery itchy eyes; hemorrhage at puncture site; dissection of the coronary vessels; flattening of t wave; sloughing; warm sensation over body; high bun; dryness of the oropharynx; proness to falling; haemorrhagic eruption; reactivation of psychotic processes; excessive skin wrinkling; dryness of the oral mucous membranes; cold skin; proneness to falling; dry mouth/nasal stuffiness; low plasma cortisol; dry skin non-application site; mesenteric embolism; libido decreased-male; pco2 changes; dryness of paranasal area; dryness of inrtaorbital area; warm autoimmune hemolytic anemia (waiha); gangrenous bowel bleeding; leaks, anastomotic; headache (grade $\geq$3); dry skin (grade $\geq$3); back pain (grade 3); seeing half of an object; dry, cracked skin; lowered blood urea; death caused by therapeutic agent

toxicity; bigeminal beats