IDENTIFYING NOVEL CALORIC RESTRICTION MIMETICS USING GENERATIVE AI

A THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

M.Tech

BY

JEET NANSHI

MT21302

COMPUTATIONAL BIOLOGY (CB)

INDRAPRASTHA INSTITUTE OF INFORMATION TECHNOLOGY DELHI NEW

DELHI– 110020

20th January 2024

ACKNOWLEDGEMENT

I would like to express my sincere gratitude towards Dr Gaurav Ahuja, Associate Professor, IIIT, Delhi, for providing the opportunity to undertake this thesis work, his constant support and timely supervision. This thesis work would not be possible without the guidance from my mentor Sakshi Arora, PhD under Dr. Gaurav. Lastly, I would like to express my appreciation to Chemosensory Lab members, for support throughout this endeavor.

ABSRACT

In the contemporary era, anti-aging stands as a prominent research focus. Recognizing the impact of diet on aging, investigations into caloric restriction methods have emerged as a promising and effective approach in anti-aging research, aiming to extend longevity. Despite its potential benefits, a significant drawback of caloric restriction is the necessity for strict lifelong adherence. In response to this limitation, researchers have turned their attention to molecules that mimic the effects of caloric restriction, termed caloric restriction mimetics (CRMs). This study delves into the identification of novel Caloric Restriction Mimetics, employing a Generative AI approach that integrates computational drug discovery and an anti-aging probabilistic model. Specifically, a Graph Neural Network approach and anti-aging score optimization was utilized within the framework of Multi-Constraint Molecule Sampling for Molecule Optimization (MIMOSA). Given the scarcity of existing CRMs, this research generated novel CRMs. By prioritizing bioactivity as a key feature, highly promising novel CRMs were identified through screening, visualization and top-k-neighbour techniques from ZINC20. The outcomes of this analysis yield potential caloric restriction mimetic molecules, paving the way for future research and development in the field. Moreover, the study highlights the applicability of generative modeling approaches in cases where the availability of ground truth data for classical machine learning model building is limited. This research contributes to advancing our understanding of anti-aging strategies and underscores the significance of innovative methodologies in drug discovery within the realm of aging-related interventions.

Key Words: Caloric restriction mimetics, Graph neural networks, Novel CRMs, Autophagy, Gen AI.

TABLE OF COTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBRIVIATIONS

CR: Caloric restriction

CRM: Caloric Restriction Mimetics

CVD: Cardiovascular Diseases

AMPK: AMP-Activated Protein Kinase

mTOR: Mechanistic Target of Rapamycin

QOL: Quality of Life

2DG: 2-Deoxy-D-Glucose

SGLT2: Sodium-Glucose Cotransporter 2

ACLY: ATP citrate lyase

AcCoA: Acetyl Coenzyme A

HCA: Hydroxycitric acid

CPT1: Carnitine Palmitoyl Transferase 1

HUVEC: Human Umbilical Vein Endothelial Cells

EGCG: Epigallocatechin Gallate

KAT: Lysine Acetylases.

GAN: Generative Adversarial Networks

CCGVAE: Conditional Constrained Graph Variational Autoencoder

MIMOSA: Multi-Constraint Molecule Sampling for Molecule Optimization

SMILES: Simplified Molecular-Input Line-Entry System

GNN: Graph Neural Network

QED: Quantitative Estimate of Drug Likeness

1. Introduction

1.1) Influence of Diet on Aging.

“Food can heal and renew. Food can be your anti-aging medicine.” This quote by Deepak Chopra is not only true but highly accurate. The process of aging is a gradual time-dependent process where natural changes accumulating all the way to the years causes the body to not function properly. Even with the increase in life expectancy over the years the quality of life has still not improved in their later stages(Calder et al., 2018). Since childhood, we have been encouraged to consume green and healthy vegetables while maintaining a routine of physical activity. It is widely understood that the shape of certain foods is related to the organs they benefit upon consumption – for example, carrots for eyes, tomatoes for the heart, kidney beans for the kidneys, and so on. This not only enhances the biological state of the body but also stabilizes and reduces the risk of age-related diseases and obesity. Over the decades researches have understood the aging process and identified the 9 hallmarks of the aging, Altered Intercellular communication, Stem cell exhaustion, cellular senescence, Mitochondrial dysfunction, Deregulated nutrient-sensing, loss of proteostasis, Epigenetic alterations, telomere attrition and genomic instability (López-Otín et al., 2013).

Evidence supports the correlation between healthy eating habits and an improved quality of life in older Chinese populations. Consuming a diverse vegetable diet, incorporating green tea rich in catechins, and including foods high in antioxidants, fish, and eggs have shown beneficial effects on memory activity. Additionally, these dietary practices exhibit anti-inflammatory effects, contributing to the delay of telomere shortening and improvement of telomere length (Chou et al., 2019). Similarly, the Japanese diet, characterized by small portions of traditional foods with low fat content and high Omega-3 content, has been studied for its health benefits. Research indicates that Japanese individuals who incorporate elements of their traditional diet, and Mediterranean diet, experience a reduced incidence of cardiovascular diseases (CVD) when compared to other populations.

On the contrary, evidence suggests that Westernized diets, characterized by high-fat content, red meat, processed foods, and synthetic chemical additives, are associated with a higher risk of obesity, hyperinsulinemia, the production of reactive oxygen species leading to oxidative stress, type-2 diabetes, and more (Kopp, 2019). Therefore, it is advisable to avoid so-called "junk food" and maintain a well-balanced diet rich in green and nutritious components. This

highlights the significant role of diet as a crucial component in the aging process and its impact on aging mechanisms.

1.2) Caloric Restriction (CR) and effects.

In conclusion, recognizing diet as a critical component in the aging process, a non-pharmacological dietary intervention known as Caloric Restriction (CR) has garnered significant attention in research. This is attributed to its beneficial effects on extending longevity through improvements in metabolic adaptation, including the reduction of oxidative stress and decreased lipogenesis (Leong, 2018). CR, as its name suggests, involves a reduction of calorie intake by 20-40%, while ensuring complete intake of micronutrients and macronutrients. Studies on caloric restriction in various organisms, from yeast to primates, have demonstrated increased longevity and a reduction in reactive oxygen species (Choi et al., 2017)(Mattison et al., 2017). Human studies implementing CR have shown a decline in body weight, predominantly from fat mass(Most et al., 2017).

The benefits of CR have been observed through the activation of autophagy and modulation in protein acetylation, primarily. Under CR conditions, researchers have noted molecular effects that lead to a reduction in oxidative stress, modulation of nutrient signaling, thyroid axis activity, TNF alpha, cardiovascular disease (CVD) risk parameters, and mechanistic target of Rapamycin (mTOR). Simultaneously, there is an increase in AMP-activated protein kinase (AMPK), Sirtuins, and autophagy, along with the modulation of protein acetylation. These molecular-level effects manifest at the physiological level, resulting in improved Quality of Life (QOL), weight loss, metabolic adaptation, decreased core body temperature, blood pressure, inflammation markers, fasting insulin levels, and energy expenditure(Madeo et al., 2019).

While Caloric Restriction (CR) appears to be a promising and effective approach with tangible results, its main limitation lies in the difficulty of sustaining CR throughout one's entire life. The necessity to adhere to a specific and stringent dietary regimen poses challenges to long-term compliance.

1.3) Caloric Restriction Mimetics (CRMs) – types and mechanisms.

To address the challenges posed by the conventional CR method, researchers have identified a class of chemical compounds capable of replicating the effects observed in primary model organisms and humans undergoing CR. These compounds are referred to as Caloric Restriction

Mimetics (CRMs). CRMs effectively target a spectrum of physiological and metabolic effects similar to those achieved through traditional CR. CRMs can be categorized into two types: downstream-type CRMs and upstream-type CRMs (H. Shintani et al., 2018). The former impacts intracellular signaling systems, including AMPK activation, mTOR inhibition, Sirtuin activation, epigenetic controls, and redox balance. On the other hand, the latter influences energy metabolism, particularly glucose and lipid metabolism. However, the overall effects of CRMs are observed to converge towards autophagy, thereby delivering the beneficial outcomes associated with CR.

1.3.1) Glucose metabolism.

Molecules targeting glucose metabolism often concentrate on inhibiting glycolysis and reducing glucose throughout the body. Certain compounds within this category have demonstrated the ability to induce autophagy and decrease insulin levels.

D-Allulose, is a monosaccharide found as food component in wheat and plant, known to inhibit glucose metabolism, inhibits intracellular glycolysis, and disaccharide in the intestine. It has been seen to increase lifespan for nematode and mouse models at lowering the glucose levels, body weight and body fat along with change in energy metabolism and increasing the antioxidant enzyme, although the action of autophagy is yet unknown(T. Shintani et al., 2017).

2-Deoxy-D-Glucose (2DG), is a glucose analog know to inhibit glycolysis. It has shown positive effects towards reduced heart-rate acting cardioprotective and insulin levels. It also showed extension of lifespan in C.elegans, However it has showed certain toxic effects in rodents, thus it has showed both beneficial and negative effects on organisms (Ingram & Roth, 2021).

D-Glucosamine, an amino monosaccharide less commonly found in dietary supplements, serves as a precursor to lipids and glycolytic proteins. It exerts its effects by inhibiting glycolysis through hexokinase-1. Studies have demonstrated that D-Glucosamine has the potential to extend longevity and mitigate aging effects in nematodes and mice models, attributed to its ability to activate autophagy properties (T. Shintani et al., 2018).

In addition to mentioned glucose metabolic regulating CRMs, several other compounds influence glucose metabolism. These include Chitosan, Acarbose, Gymnemoside, Astragalin, as well as drugs classified as Sodium-Glucose Cotransporter 2 (SGLT2) inhibitors, such as empagliflozin, bexagliflozin, dapagliflozin, and canagliflozin (H. Shintani et al., 2018).

1.3.2) AcCoA Depleting Compounds.

Inhibiting ATP citrate lyase (ACLY) and HSP90 results in the depletion of acetyl coenzyme A (AcCoA). This depletion has been observed to induce autophagy and inhibit lipogenesis. Compounds like Hydroxycitric acid (HCA), SB-204990, radicicol, and perhexiline maleate are known for their ability to deplete AcCoA and exhibit these effects.

Hydroxycitric acid (HCA), an inhibitor of ATP citrate lyase (ACLY), is present in plant extracts, particularly in higher concentrations in Garcinia compared to Hibiscus. HCA has demonstrated the ability to induce autophagy in mouse organs, both in vitro and in vivo. While it exhibits antitumor properties, its impact on prolonging longevity has not been observed (Pietrocola et al., 2016).

SB-204990 and **Radicicol** both exhibit high-affinity inhibition of ATP citrate lyase (ACLY). Radicicol is a natural compound found in Diheterospora chlamydosporia and has demonstrated positive effects in mouse models with myocardial ischemia reperfusion (Griffin et al., 2004). While Radicicol's impact on autophagy is still under investigation, SB-204990, a synthetic compound has been observed to reduce tumors, cholesterol levels, and triglycerides by leveraging autophagy (Hatzivassiliou et al., 2005).

Perhexiline maleate acts by reducing beta-oxidation, depleting AcCoA, and inhibiting carnitine palmitoyl transferase 1 (CPT1). Additionally, it inhibits mTORC1 and induces autophagy by disrupting the flow of fatty acids to mitochondria (López-Otín et al., 2016).

1.3.3) Polyphenols.

Polyphenols, with their vast structural diversity, are crucial compounds found in plants, contributing to plant defense mechanisms and ecological functions. They are present in various foods and beverages, and their dietary significance is highlighted, recognizing the variability in content based on plant variety and processing methods. The potential health benefits of polyphenols, particularly as chemo-preventive agents, are alluded to in the context of their diverse roles and inclusion in the human diet.

Resveratrol, a stilbene, is widely distributed stilbene in the human diet, found in numerous plant species. Daily intake varies, with wines and grape products being significant sources. Lingonberry and the skin of red grapes, especially in Muscadine red wine, are highlighted for their high resveratrol content. Resveratrol had clinically shown to affect the autophagic

pathways by upregulating them in Human umbilical vein endothelial cells (HUVEC) (Berman et al., 2017).

Epicatechin and **EGCG**, intake levels are varied across countries, though flavan-3-ol intake is indicative of dietary patterns. The sources of epicatechin are fruits such as apples, berries and cocoa while EGCG in found abundantly in green tea. Treatment of EGCG on bovine aortic endothelial cells (BAEC) showed an autophagy flux induction via a CaMKKβ/AMPK-dependent mechanism (Liu & Li, 2019).

Quercetin, a flavanol, is known for its health-promoting properties and is found in foods such as onions, apples, and berries. The estimated intake levels of quercetin differ among populations, indicating variations in dietary habits. Quercetin effected the RAW264.7 macrophages by upregulating autophagy (H. Cao et al., 2019).

1.3.4) NAD+ Precursors.

NAD+ levels play a crucial role in regulating the activation of AMPK, glycolysis, TCA, and sirtuins. The decline in NAD+ levels with age suggests that replenishing these levels could be pivotal for therapeutic effects and extending longevity in humans. NAD+ precursors are present in both animal and plant-based foods.

Nicotinic acid (NA), commonly known as Vitamin B3, can be obtained directly through the raw intake of tryptophan. In doses less than 1g, it has been observed to reduce LDL and hypercholesterolemia (Eisenberg et al., 2016).

Nicotinamide riboside (NR) and **Nicotinamide Mononucleotide (NMN)** are NAD+ precursors that can be obtained from the diet. In mice models, a dosage of 2g/day for both NR and NMN increased blood NAD+ levels in clinical trials, potentially inducing autophagy as well (Trammell et al., 2016).

1.3.5) Polyamines.

Polyamines, with a focus on spermidine and spermine, are acquired through various means, including dietary sources like wheat, soya, and nuts, as well as microbial synthesis in the gut and endogenous processes. Notably, dietary spermidine has been associated with positive outcomes for cardiovascular and neurological health, along with implications for longevity. The gut microbiome plays a role in contributing to the overall polyamine pool.

Spermidine has been observed to influence pathways involving KATs, AMPK, and NF-kB. Its effects have been characterized as cardioprotective and neuroprotective, contributing to its potential health-promoting properties (Eisenberg et al., 2016).

1.4) in-silico Chemical Synthesis Methods.

Traditional drug discovery processes involve substantial human intervention in screening a vast number of chemicals to identify a hit molecule. However, these methods are time-consuming and often incur significant costs. Recent technological advancements have paved the way for faster and more efficient drug discovery processes. These innovative approaches fall under the categories of generative methods or reinforcement learning methods. It's worth noting that these methods primarily concentrate on optimizing individual properties such as drug likeness, logP, and so on.

MolGAN, a likelihood-free generative model designed for graph-structured data, utilizes generative adversarial networks (GANs) and reinforcement learning. It stands out as a promising method for generating synthetic data that closely resembles the original dataset. This tool seamlessly integrates both implicit and explicit representations of chemical compounds (N. De Cao & Kipf, 2018). An enhanced iteration of MolGAN, known as Mol-Cycle GAN, has been developed. It generates data with a focus on a specific property by training on two datasets. **Mol-Cycle GAN** leverages the Junction Tree Variational Autoencoder (JTVAE) to optimize the latent space representation (Maziarka et al., 2020). Variational Autoencoder (VAE) has found application in various generative tools, with one notable example being the **Conditional Constrained Graph Variational Autoencoder** (CCGVAE). This approach integrates Gated Graph Recurrent Neural Networks within the encoder to pinpoint representations of atoms based on their valences. The decoder, on the other hand, generates molecules by adding bonds until a stopping criterion is met. Gradient descent is employed to optimize the chemical space, ensuring the incorporation of desired features (Rigoni et al., 2020).

EvoMol stands out as another effective generative method for producing new molecules, whether with or without the utilization of prior data. This approach employs an evolutionary algorithm and incorporates seven mutation functions at the atomic level, including append, change, remove, substitute, insert, cut, and move. EvoMol covers a diverse range of chemical considerations during the generation of synthetic compounds, encompassing aspects like HOMO and LUMO generation, integration of synthesizability score, and representation of the generated molecules through an evolutionary tree (Leguy et al., 2020).

Nevertheless, while many of these approaches focus on optimizing chemical space, most operate independently of prior data, meaning they are pre-trained with a specific dataset and don't involve re-training the models. To address this limitation, **MIMOSA** (Multi-Constraint Molecule Sampling for molecule optimization) comes into play. Unlike some other methods, MIMOSA undergoes a training period with the desired dataset and requires a seed input to generate structures similar to both input molecule and training dataset. The training process revolves around substructures, ensuring that the structural similarity of the newly generated compounds aligns with the seed molecules, thereby optimizing the desired property (Fu et al., 2020).

Thus, this study aims to identify novel potential Caloric Restriction Mimetics, employing a Generative AI approach that integrates computational drug discovery and an anti-aging probabilistic model. Specifically, a Graph Neural Network approach and anti-aging score optimization was utilized within the framework of MIMOSA.

2. Materials and Methods

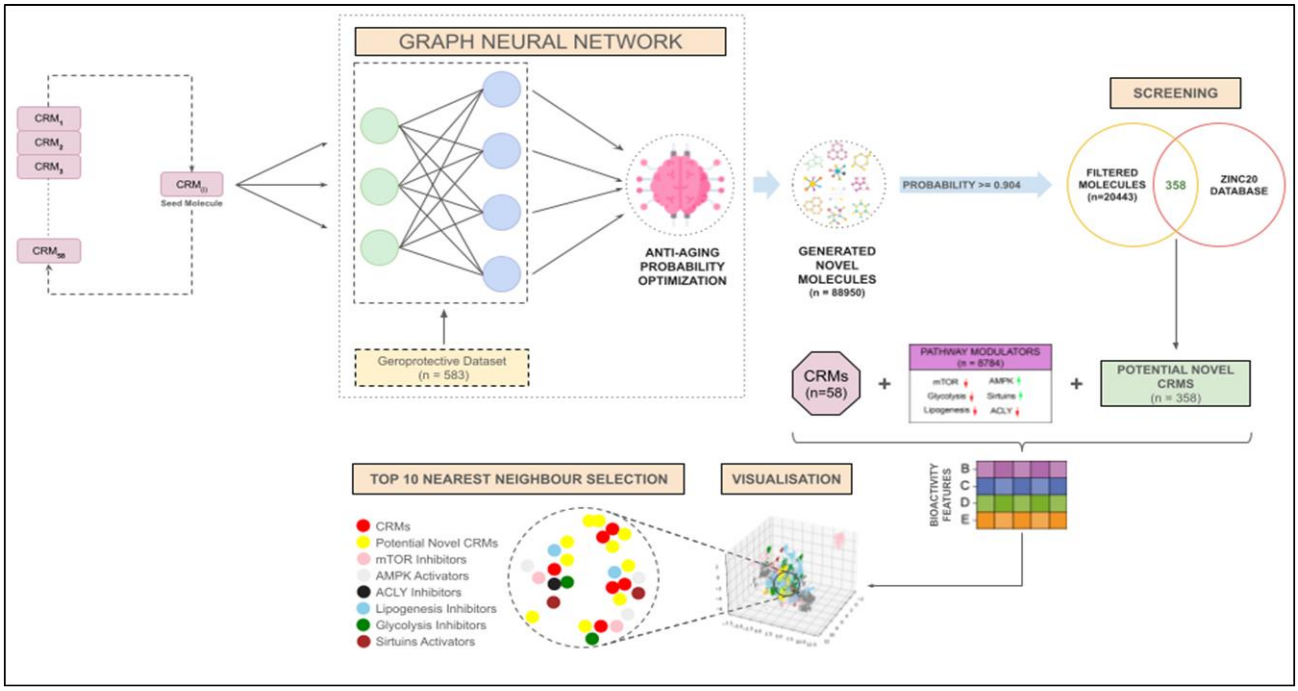


Figure 2.1. Utilization of a set of 58 caloric restriction mimetics as seed compounds, a novel molecule generation approach was adopted, leveraging graph neural networks with an optimization of Anti-Aging probability and set a cutoff of 0.904. The resulting compounds underwent screening against the ZINC20 database. Subsequently, the collected mimetics, pathway modulators, and novel potential compounds were integrated based on their bioactivity features. Visualization of these components was performed collectively, culminating in the utilization of K-nearest neighbours (KNN) to explore and identify potential novel drugs.

2.1) Data Collection and Feature Generation

2.1.1) known CRMs

Extensive mining was carried out on scientific databases and publications which yielded a total of 58 confirmed CRMs, of which some are approved CRMs and others tested in primitive model organisms viz S.cerevisiae, M.musculus, and C.elegans as CRMs and as potential for humans. The selected CRMs are known to have a caloric restriction effect influencing the reversal of aging.

2.1.2) Training Dataset for GNN

To generate new anti-aging drugs, a training dataset was finalised from 2 different sources which resulted in a diverse set of 583 molecules in total. 508 compounds from the DrugAge database and 75 compounds corresponding to Geroprotective compounds dataset. The compounds were stored in SMILES format and used as input for training process of the GNN. The training set has been chosen effectively to upgrade the model's performance and generating the novel molecules to optimized anti-aging properties.

2.1.3) Collection of Pathway Modulators

We identified six key pathways central to CRMs through an extensive review of existing literature. These pathways—AMPK signalling, mTOR signalling, ACLY signalling, Lipogenesis, Glycolysis, and SIRTUINs signalling—were selected based on their significance. To gather modulators affecting these pathways, we utilized PubChem Confirmatory Bioassays, focusing on understanding how CRMs influence these pathways by either activating or repressing them. The process involved a fine selection of assays meeting specific criteria. The assays were filtered based on the active compound count, excluding assays with zero counts, and targeted specific assays aligning with our objectives, while also excluding those aimed at analysing compound toxicity. Furthermore, any conflicts arising from ambiguous activity statuses within the bioassays were handled effectively ensuring the reliability and accuracy of the collected data.

2.1.4) Feature Generation

To enrich the molecular representation and encapsulate their biological essence, we harnessed bioactivity signatures. Leveraging bioactivity signatures aimed to infuse essential biological insights into our analysis, providing a nuanced understanding of molecular behaviour and potential therapeutic interactions.

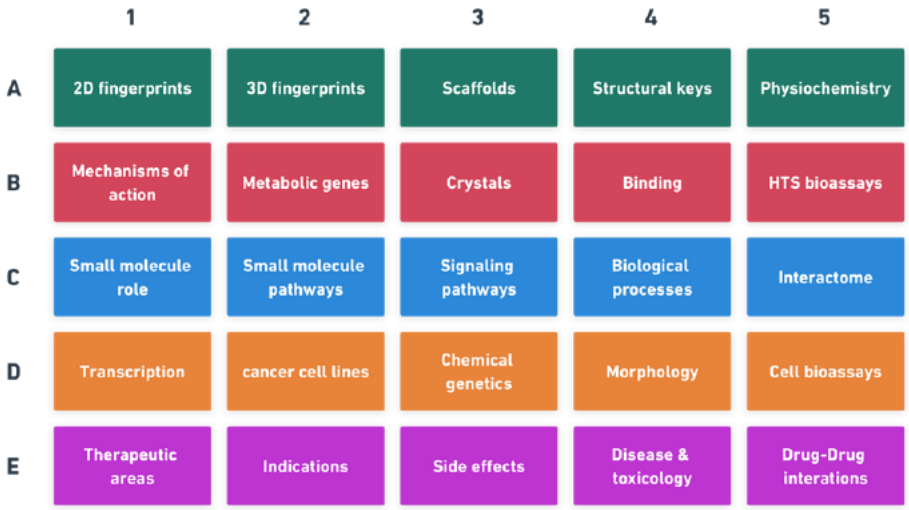


Figure 2.2. Feature space from Signaturizer. The 25-feature space representing both chemistry and biological features.

For feature extraction and preprocessing from the SMILES string dataset, a Python package 'Signaturizer' was utilized. A custom function was made to perform extraction of features and subsequent data refinement. SMILES strings were used as input, this function engendered descriptive features across five tiers of intricacy: A - chemistry, capturing structural elements; B - targets, highlighting molecular targets and their impact; C - networks, indicating potential biological pathways; D - cells, emphasizing cellular responses; and E - clinics, encompassing clinical implications. Each tier further subdivides into five sub-levels, resulting in a comprehensive array of 3200 feature spaces that encapsulate the molecules' 2D/3D structures and biological functional attributes.

2.2) Generative AI using Graph neural network

In crafting a generative AI framework, we initially utilized foundational code from an existing generative AI tool known as MIMOSA: Multi-constraint Molecule Sampling for Molecule Optimization. The existing generative AI used python packages ranging from basic pandas, numpy, seaborn, etc., to advanced packages as Keras, Tensorflow, pytorch, torch, signaturizer, RDKit, etc. However, this tool presented inefficiencies in its workflow due to the separation of modules handling distinct functionalities like training, optimization, and molecule generation. To overcome these limitations, we restructured and reorganized the codebase, implementing a systematic scheme. This restructuring involved customizing and reconfiguring modules to align with our specific requirements. The result was the development of a unified

pipeline that significantly reduced the need for manual intervention and produced a more streamlined, organized, and refined output

The enhanced mechanism led to a considerable improvement in the optimization process, streamlining the various stages and eliminating redundant steps. Additionally, we fine-tuned our generative AI to focus on optimizing anti-aging probabilities by integrating a sophisticated prediction algorithm. This algorithm, utilizing features derived from signaturizers, significantly improved our AI's efficiency and accuracy, aligning it more closely with our primary objective.

2.2.1) Validation of training data

The anti-aging dataset initially comprised 583 compounds represented in SMILES format. These SMILES strings were translated into graphical representations, enabling the extraction of substructures such as atoms, chains, and single-ring structures. The resulting output included all identified substructures across the compounds along with their respective frequency of occurrence.

To construct a robust training dataset for identifying potent anti-aging molecules, a stringent criterion was applied on substructure's occurrence crossing a threshold value of 10, and only those exceeding this threshold were considered valid substructures.

To ensure the generated molecules adhered closely to anti-aging criteria, a validation process for SMILES was implemented. The substructures meeting the threshold value were traced back in the compounds of the training dataset. This backtracking method facilitated the identification of validated SMILES strings. Subsequently, compounds containing these validated substructures were retained and employed in training the Graph Neural Network.

2.2.2) Graph Neural Network Training

Valid SMILES that successfully passed the validation check were harnessed in the training of graph neural networks. The dataset underwent a train-test split, with a 10% allocation for validation and the remaining 90% dedicated to training. The training procedure for the graph neural network relies on three key inputs: an adjacency matrix delineating edges, a node matrix encapsulating each individual node, and atom indices. This triad of inputs forms the foundation for the comprehensive training regimen of the graph neural network.

The process involved transforming substructures and ring structures into indices by cross-referencing them with a vocabulary file. This resulted in the creation of a comprehensive list containing indices for all identified substructures and ring structures. Next, atom symbols in

all SMILES that did not match those present in the ring structures underwent a similar transformation into indices, using the vocabulary file for comparison. Subsequently, these two lists of indices were meticulously interconnected to map them through the creation of a dictionary.

The creation of adjacency matrix implemented by a null matrix. The process involved iterating through each bond in the molecules and assessing whether the bond was not part of a ring. If the bond did not belong to a ring, the algorithm retrieved the indices of the two atoms involved in the bond and set the corresponding entries of the null matrix to 1, indicating the connection between respective two atoms. Furthur, another iteration took place over the ring structures to identify any common atoms between the two ring structures. If a common atom was found, an entry of 1 was recorded in the matrix to signify the connection between atoms involved in the identified ring structures. Lastly, the algorithm conducted a check on the total number of non-zero entries in the matrix, specifically focusing on diagonal entries. This final step helped assess the overall connectivity and relationships between atoms, ensuring a comprehensive representation of the molecular structure in the matrix.

An index list representing the leaf nodes in the adjacency matrix was generated, and a leaf node was randomly selected for concealment. The variable "label" was subsequently assigned the index of the concealed node. Following this, a node matrix was established, and an iterative process was applied to the leaf node indices list. Upon a match between an index and the concealed node, a value of 1 was inserted into the last column of the matrix. This modified molecular graph served as the training data for two Graph Neural Networks (GNNs). These GNNs underwent training to acquire an embedding vector hv for each node v ∈ V, leveraging the information from both node and edge features within the graph.

$$\mathbf{h}_v^{(k)} = \mathrm{ReLU}\left(\mathrm{MLP}\left(\mathrm{CONCAT}\left(\sum_{u \in \mathcal{N}(v) \cup \{v\}} \mathbf{h}_u^{(k-1)}, \sum_{e=(u,v): u \in \mathcal{N}(v)} \mathbf{g}_e^{(k-1)}\right)\right)\right) \quad \text{Eq (1)}$$

N(v) represents the set of all neighbors of node v. The initial node embedding is denoted as $h_v^{(0)}$, corresponding to the feature vector fv. After undergoing K layers of Graph Neural

Network (GNN) processing, the resulting final node embedding is represented as $h_v^{(k)}$ for the given node v. Here the layers are k = 1, …., K. The CONCAT function is to concatenate the 2 vector outputs from the GNNs. Here we chose K= 5.

2.2.3) Molecular Generation, Termination and Optimization.

The two employed Graph Neural Networks (GNNs) include mGNN and bGNN, serving distinct purposes in the predictive modelling process. mGNN functions as a multi-classification neural network, predicting the substructure type of a masked node. The training process utilizes entropy loss as the guiding criterion. For each masked node, mGNN employs a softmax function on hv(K) to predict an output vector containing the probabilities of all possible substructures.

$$\hat{\mathbf{y}}_v = \text{Softmax}\left(\text{FC}(\mathbf{h}_v^{(K)})\right) \quad \text{Eq (2)}$$

On the other hand, bGNN is a binary classification neural network designed for predicting molecule topology. It comprises a fully connected network with 50 neurons as hidden layers and 1 layer as the output layer. This network assesses the expansion potential of leaf nodes. Using a sigmoid function, it assigns a value of 0 to leaf nodes (indicating no expansion) and 1 to non-leaf nodes (suggesting a chance of further expansion). The training of bGNN is guided by binary cross-entropy loss. For each leaf node, bGNN employs a sigmoidal function on hv(K) to predict an output probability that the node will expand or become permanent.

$$\hat{z}_v = \text{Sigmoid}\left(\text{FC}(\mathbf{h}_v^{(K)})\right) \quad \text{Eq (3)}$$

The process of generating molecules involves two operations that modify substructures with the assistance of both GNNs. These operations—replace, add, or delete—affect the substructure of input molecules. In the replacement operation, a new structure is sampled from the multinomial distribution of substructures, and the replacement is executed only if the chosen substructure is selected. The addition function considered the combined output from both bGNN and mGNN, with the former assessing whether the node has the probability to expand and the latter strategically determining which substructure to add for the maximum likelihood of the molecule. The deletion operation is straightforward; if the node is considered unsatisfactory, it is removed from the molecule. These three operations collectively contribute

to molecule generation, utilizing a combination of probabilistic sampling and strategic decision-making based on the GNN outputs.

The three operations yield distinct sets of sampled molecules, with each set transformed into a probability distribution denoted as $\{1, w_r\}$, $\{1, w_a\}$, and $\{1, w_d\}$, where w_r, w_a, and w_d represent the weights for the replace, add, and delete operations as given by Eq (4), Eq (5), and Eq (6) respectively

$$w_r = \frac{p_X(Y') \cdot [\text{mGNN}(Y, v)]_{s'_v}}{p_X(Y) \cdot [\text{mGNN}(Y, v)]_{s_v}}, \quad \text{Eq (4)}$$

$$w_a = \frac{p_X(Y') \cdot \text{bGNN}(Y, u) \cdot [\text{mGNN}(Y', v)]_{s_v}}{p_X(Y) \cdot (1 - \text{bGNN}(Y, u))} \quad \text{Eq (5)}$$

$$w_d = \frac{p_X(Y') \cdot \big(1 - \text{bGNN}(Y', u)\big)}{p_X(Y) \cdot \text{bGNN}(Y', u) \cdot [\text{mGNN}(Y, v)]_{s_v}} \quad \text{Eq (6)}$$

The set of molecules is selected through a Markov Chain Monte Carlo (MCMC) sampling method. Once the weights are determined, the MCMC employs a transition kernel to choose the set of molecules in a way that satisfies the condition ɤ1 + ɤ2 + ɤ3 = 1, with the additional constraint that ɤ2 equals ɤ3, for any arbitrary values of ɤ1, ɤ2, and ɤ3 belonging to the set of real numbers.

$$Y' \sim \begin{cases} S_{\text{replace}}(Y' \mid Y), & \text{prob } \gamma_1, \text{accept w. } \min\{1, w_r\} \\ S_{\text{add}}(Y' \mid Y), & \text{prob. } \gamma_2, \text{accept w. } \min\{1, w_a\} \\ S_{\text{delete}}(Y' \mid Y), & \text{prob. } \gamma_3, \text{accept w. } \min\{1, w_d\} \end{cases} \quad \text{Eq (7)}$$

The optimization of the newly generated molecules also involves the MCMC algorithm, which selects the top K molecules from the distribution of the next iteration. The optimization of molecule generation depends on the activity score we aim to optimize. The existing framework offers four scores for optimization: QED (Quantitative Estimate of Drug Likeness), DRD (Dopamine Receptor), PLogP (Penalized LogP), and a cumulative score of all mentioned scores. In this study, we modified the score optimization using an anti-aging probabilistic model.

The generation parameter was configured for 15 iterations, for quantification of novel molecules generated. The population size was set to 1500, indicating the number of individuals

within each generation of the evolutionary algorithm. The remaining parameters were maintained at their default values, ensuring a standardized and consistent experimental setup

2.3) ZINC20 Database screening

The compounds meeting the threshold of 0.904 or higher underwent a transformation into RDKit canonical smiles and were subsequently subjected to string matching within the ZINC20 database, housing an expansive collection of roughly 2 billion compounds. The ZINC20 database was processed to form a comprehensive dataset containing only the SMILES and ZincID of the compound. The conversion of qualifying smiles to RDKit canonical format was expedited due to time constraints, considering that ZINC20 compounds inherently maintain RDKit canonical representation. This approach allowed for efficient comparison and retrieval of potential matches within the extensive ZINC20 compound library.

2.4) Visualisations and Top-10 Nearest Neighbour

Post the extraction of bioactivity features from known 58 CRMs, pathway modulators, and the identified novel molecules post-screening with ZINC20, the dataset underwent preprocessing steps. This involved the elimination of duplicate entries within the pathway modulators and across the entire dataset. Subsequently, UMAP, a dimensionality reduction technique, was employed to obtain embeddings. Leveraging the visualization capabilities of Matplotlib, we created visually compelling representations to highlight and identify distinct categories within the dataset. To further investigate the relationships, a K-Nearest Neighbours (KNN) approach was utilized to uncover the top 10 neighbours for each CRM. The annotations within the bioactivity neighbour maps were done using Adobe Illustrator.

3. Results

3.1) Data Collection and Feature Generation

3.1.1) Existing CRMs

A comprehensive review of the literature identified 58 particularly promising Cancer Therapeutic Response Modifiers (CRMs). In-depth analyses were conducted on the mode of action and significant pathways associated with these CRMs. The mechanisms of action revealed that certain pathways were upregulated, including Autophagy, Sirtuins, AMPK (AMP-activated protein kinase), and KATs/HATs (Lysine acetylation and Histone acetylation). Conversely, other pathways were found to be downregulated, such as mTOR (mammalian target of rapamycin), ACLY (ATP-citrate lyase), lipid metabolism, and glucose metabolism. The compounds selected are mentioned in below table.

CRMs	**TARGET(S)**	**PubChemID**
Aspirin	AMPK, EP300, COX-1, COX-2, mTOR, NF-kB	2244
Salicylic acid	AMPK, EP300, COX-1, COX-2, mTOR, NF-kB	338
Hydroxycitric acid	ACLY	123908
Resveratrol	KDACs (SIRT1), AMPK, NF-kB	445154
Spermidine	KATs (EP300), mTOR, AMPK	1102
1,2,3-Benzenetricarboxylate	citrate transport protein	520281
Acipimox	niacin receptor 1	5310993
Berberine	SIRT1	2353
Caffeic acid	AMPK and sirtuins	689043
Catechin	pleiotropic, exact mechanism unknown	9064
Curcumin	AMPK, mTOR, and EP300	969516
Epicatechin	pleiotropic, exact mechanism unknown	72276
Gallic acid	AMPK and HATs	370
Metformin	AMPK, mTOR, HATs, and KDACs (sirtuins)	4091
Myricetin	SIRT1	5281672
NAD+	KDACs (sirtuins), AMPK	5892
Nicotinamide	KDACs (sirtuins)	936

Nicotinamide mononucleotide	KDACs (sirtuins)	14180
Nicotinamide riboside	KDACs (sirtuins)	439924
Perhexiline maleate	Carnitine O-palmitoyl transferase 1, mTOR	5284439
Piceatannol	SIRT1	667639
Quercetin	SIRT1	5280343
Rapamycin	mTOR	5284616
SRT1720	SIRT1	25232708
Uk5099	Mitochondrial pyruvate carrier	6438504
4,4'- Dimethoxychalcone	GATA transcription factors	129704062
A-485	EP300	118958122
Acarbose	a-glucosidase inhibitor	41774
Anacardic acid	EP300	167551
C646	EP300	1285941
Cape	AMPK	5281787
Cpi-613	pyruvate dehydrogenase	24770514
Garcinol	EP300	5281560
Glucosamine	hexokinase and mTOR	439213
Radicicol	ACLY and HSP90	6323491
SB-204990	ACLY	10340264
SRT2104	SIRT1	25108829
Oxaloacetic acid	(sirtuins) and AMPK	970
Chitosan	Glucose metabolism via improved lipid metabolism	71853
2-Deoxy-D-glucose	Glycolysis reduction, lipid metabolism, AMPK	108223
D-allulose	Glycolysis reduction, lipid metabolism, AMPK	441036
Empagliflozin	SGLT2 inhibitor	11949646
Canagliflozin	SGLT2 inhibitor	24812758
Dapagliflozin	SGLT2 inhibitor	9887712
Bexagliflozin	SGLT2 inhibitor	25195624
Rosiglitazone	PPAR receptors	77999

Gymnemoside	Glucose absorption	6442215
Iodoacetate	GAPDH inhibitor	5240
Mannoheptulase	Glycolytic hexokinase inhibitor	12600
Epigallocatechin gallate	SASP, autophagy, AMPK, mTOR, HATs, and KDACs	65064
Apigenin	SASP reduction, induced autophagy	5280443
Amentoflavone	Dimer of Apigenin, induced autophagy	5281600
Astraglin	Glycolysis inhibitor	5282102
Chrysin	Glycolysis inhibitor	5281607
Tryptophan	NAD+ Precursors	6305
Nicotinic acid	NAD+ Precursors	938
Chalcones	GATA inhibitors	162517550
Trehalose	DAF-16 transcription factor, autophagy	7427

Table 3.1. Known CRMs collection table. The CRMs column contains the known 58 CRM. The TARGET(S) column contains, pathways each CRM influences to get the CR effect. The PubChemID column contains the PubChem IDs for each of the corresponding CRMs.

Certain compounds exhibit specificity to individual pathways, while others target multiple pathways simultaneously. These drugs play a crucial role in elucidating the predominant pathway influenced by caloric restriction mimetics, providing valuable insights into the comprehensive effects on cellular processes

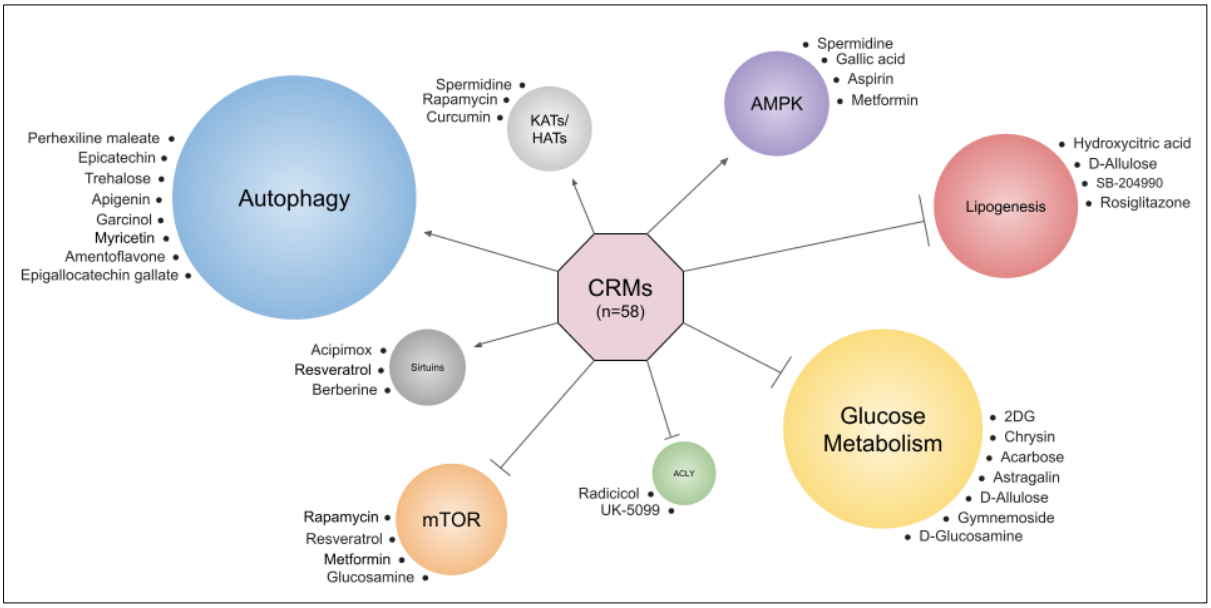


Figure 3.1. Selected 58 CRMs illustrating the mode of action and highlighting the pathways they influence. The representation involves illustrating the upregulation and the downregulation of pathways, with the size of the circle corresponding to the significance of the mechanism employed by CRMs to achieve the desired caloric restriction effects.

3.1.2) Collection of pathway modulators

The initial counts of bioassays which were the raw count, for the pathways mTOR, AMPK, Glycolysis, Lipogenesis, Sirtuins, and ACLY were found to be 2826, 545, 445, 234, 115, and 20 bioassays respectively. After processing of the bioassays, the final counts of bioassays for each pathway become 51, 58, 91, 34, 27, and 12, respectively. The total compounds obtained from the bioassays for these pathways are 55769, 12638, 1643, 724423, 204, and 45, respectively.

These compounds were further categorized into three classes: Class-1 (inhibitors), Class 0 (inconclusive/inactive compounds of bioassays), and Class 1 (activators). The numbers of inhibitors for each pathway were found to be 4,601, 120, 1,108, 3,782, 3, and 32, respectively. Notably, no activators were identified for mTOR, Glycolysis, ACLY, and Lipogenesis, however there are activators for AMPK and sirtuins which were found to be 1549 and 151 respectively. Based on the mode of action we selected compounds acting as inhibitors for mTOR, Glycolysis, ACLY and Lipogenesis and acting as activators for AMPK and Sirtuins. Consequently, after gathering all respective activators and inhibitors, a dataset comprising 8,784 compounds was established for subsequent mapping and visualization analysis.

Pathways	Initial Count for Bioassay	Final Count of Bioassay	Total compounds	Class -1 (Inhibitors)	Class 0 (inactive/ inconclusive)	Class 1 (Activators)
mTOR	2826	51	55,769	4,601	50,153	0
AMPK	545	58	12,638	120	10,967	1,549
Glycolysis	445	91	1,643	1,108	535	0
Lipogenesis	234	34	724,423	3,782	714,792	0
Sirtuins	115	27	204	3	50	151
ACLY	20	12	45	32	13	0

Table 3.2. Bioassay analysis for each pathway selected. The initial count Bioassay column contains the count of all bioassays found for the corresponding pathway. The final count bioassay denotes the count of bioassays selected after the processing of the initial count. The sum of compounds obtained from bioassays are given in the total compound's column. All the compounds for each pathway were spilt into 3 classes based on their mode of action on their particular pathway.

3.2) Optimization of Graph Neural Network models and Anti-aging Probabilities.

The model integrated two neural networks, namely mGNN and bGNN, designed for multi-class classification and binary classification, respectively. The optimization of anti-aging property was successfully done by incorporating an anti-aging probabilistic model into the framework of GNN. This optimization of anti-aging property served as the foundation for generating new molecules with enhanced bioactivity.

Training the GNNs required valid SMILES based on substructures that exceeded a set threshold of occurrence as 10. The initial input for training comprised 583 molecules and after substructure extraction, 358 valid molecules were obtained. The valid training dataset was used for training process where 90% of compound SMILES (approx. 346 molecules) were used for training and other 10% of compound SMILES (approx. 34 molecules) were used for validation. 5 epochs were done resulting in 5 validation loss values for the GNNs. The 5-validation loss as ascending with epoch 0 were found to be 3.44279, 3.06645, 2.74437, 2.29467 and 1.92443. The model selected for compound generation utilized the lowest validation loss of 1.92443, capturing intricate details of the compounds and generating molecules closely resembling the parent molecules in both molecular structures and bioactivity.

For the optimization of anti-aging probabilities, a binary classification model was employed, assigning labels 0 and 1 for aging and anti-aging, respectively. The anti-aging probability served as the primary property for optimization. Each generation of novel molecules underwent a Markov Chain Monte Carlo (MCMC) selection process. This process selected a batch of molecules to achieve the optimization of anti-aging probability for that generation, thereby creating more accurate and optimized molecular structures. Due to MCMC selection process each 58 seed CRMs have varied generation number.

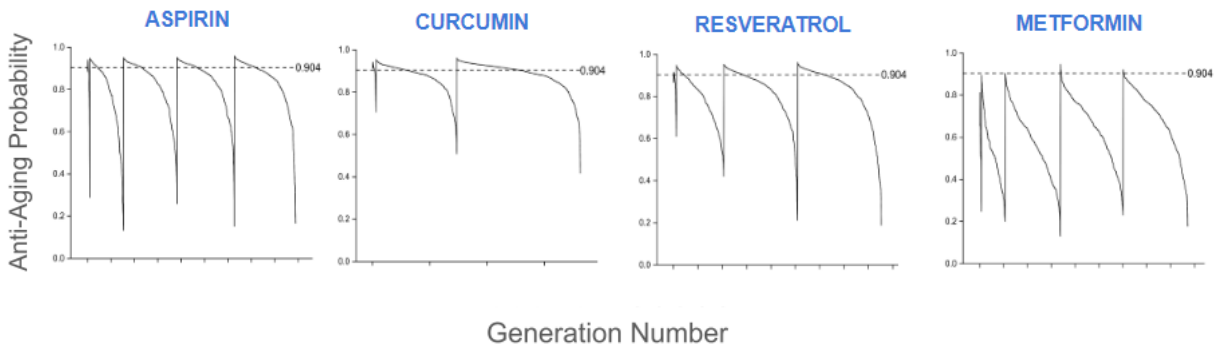


Figure 3.2 Anti-Aging probability optimization plots for aspirin, curcumin, resveratrol and metformin. The x axis depicts the generation number of each novel set of population, and y axis depicts the anti-aging probabilities of each novel compound generated in respective generation.

3.3) Generated Novel compounds

The 58 literature-confirmed CRMs served as seed molecules, initiating the generation of a total of 88,950 novel potential CRMs. However, there were exceptions where no generation occurred. It is hypothesized that this might be attributed to the training of substructures, affecting the likelihood of a node to be expanded or not, as well as the operations of add, delete, and subtract for specific nodes. 15 CRMs, Rapamycin, Iodoacetate, Radicicol, Berberine, A-485, SRT1720, UK5099, C646, Canagliflozin, SRT2104, Rosiglitazone, Acipimox, Empagliflozin, NAD+, and Bexagliflozin, did not generate any novel populations during this process.

The rest 46 compounds Caffeic acid, Salicylic acid, Astraglin, 2-Deoxy-D-glucose, Chrysin, CPI-613, Apigenin, Chalcones, Myricetin, CAPE, Resveratrol, SRT501, Gallic acid, Mannoheptulase, Metformin, Tryptophan, Garcinol, Perhexiline maleate (first conjugate), Aspirin, Nicotinic acid, Curcumin, Perhexiline maleate (second conjugate), Trehalose, Oxaloacetic acid, Nicotinamide riboside, Quercetin, Piceatannol, Amentoflavone, Hydroxycitric acid, 1,2,3-Benzenetricarboxylate, Nicotinamide, Dapagliflozin, 4,4'-Dimethoxychalcone, Anacardic acid, D-Allulose, Glucosamine, Catechin, Epicatechin, Spermidine, Nicotinamide mononucleotide, Epigallocatechin gallate, EGCG, SB-204990, Gymnemoside, Acarbose, Chitosan effectively generated the 1502, 1554, 1555, 1581, 1596, 1602, 1611, 1615, 1616, 1637, 1703, 1703, 1713, 1735, 1743, 1749, 1770, 1773, 1775, 1782, 1812, 1849, 1872, 1874, 1893, 1893, 1921, 1929, 1944, 1983, 1994, 1998, 2018, 2028, 2033, 2064, 2096, 2096, 2117, 2241, 2301, 2301, 2395, 2723, 2835, 3425 novel compounds

respectively. The minimum compounds generated were seen in Caffeic acid with generating 1502 novel compounds and maximum seen in Chitosan generating 3425 novel compounds.

The anti-aging probability distribution for each set of novel molecules by there respective

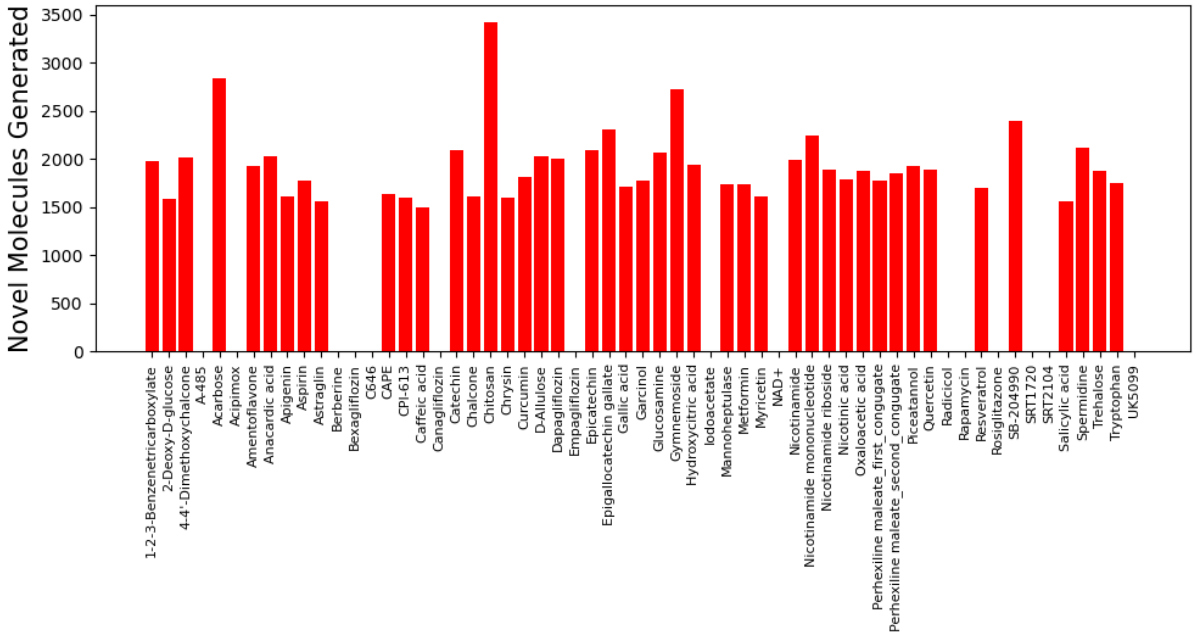


Figure 3.3. Bar plot representing the count of novel compound generated by respective molecules (58 CRMs). The y axis represents count of novel molecules generated and x axis depicts CRMs as seeds. The gaps in plot represents no generation by the compound.

parent molecules are represented as a box plot. The highest anti-aging probability for a novel molecule was found to be 0.966905, Dapagliflozin being the parent molecule. The lowest anti-aging probability for a novel molecule was found to be 0.90443, Chitosan being the parent molecule.

The narrowest range of anti-aging probabilities was observed for Gymnemoside as the parent molecule, with the range spanning from the lowest probability of 0.84372 to the highest of 0.94843. In contrast, the widest range of anti-aging probabilities was identified for Perhexiline maleate (first conjugate) as the parent molecule, showcasing a range from the lowest probability of 0.0726 to the highest of 0.93649.

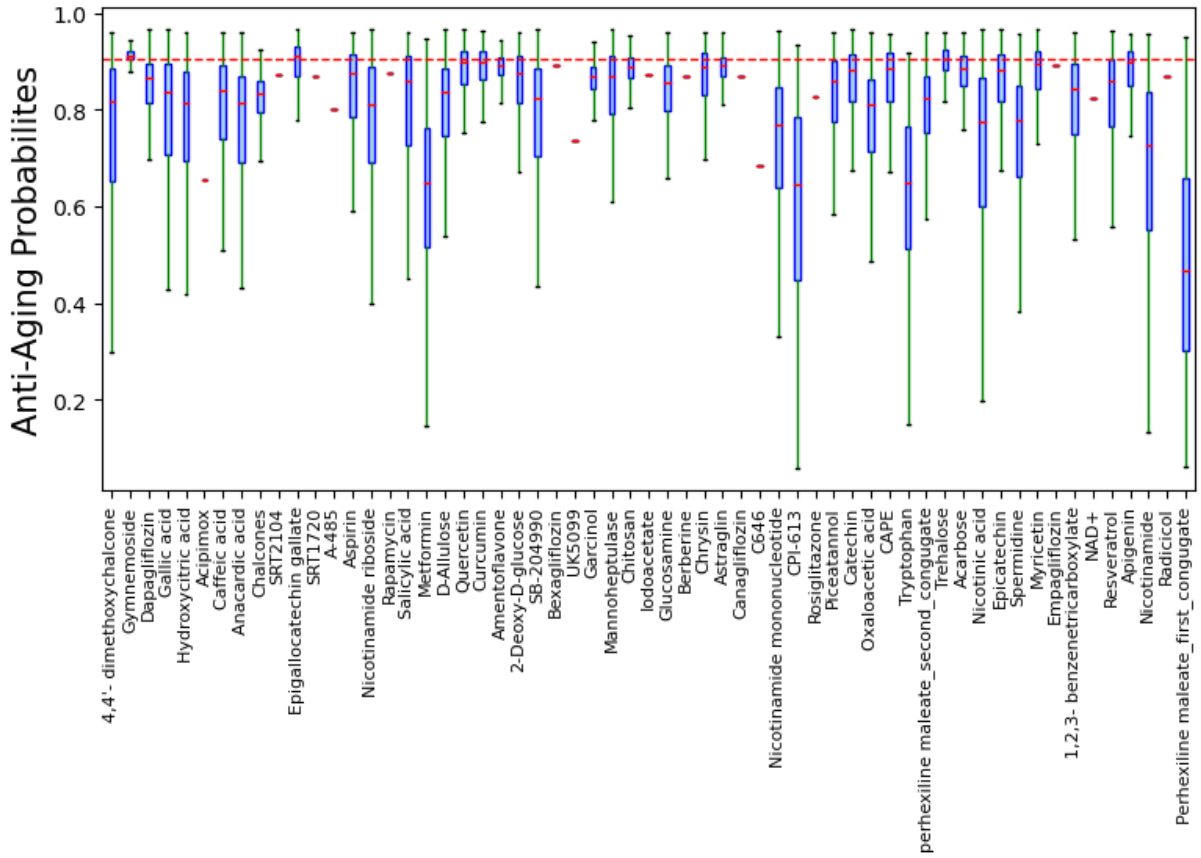


Figure 3.4. Box plot representing the score range for all novel set of compounds generated by respective parent molecules (58 CRMs). The red dotted line represents the anti-aging probability cutoff.

3.3.1) Selection of Novel Potential CRMs

The initial set of 58 literature-confirmed Caloric Restriction Mimetics (CRMs) served as seed molecules, leading to the generation of a total of 88,950 novel potential CRMs. However, after applying a cutoff anti-aging probability of 0.904, only 20,443 novel potential CRMs met the criteria.

The selection of the 0.904 cutoff was determined through hyperparameter tuning and cross-validation of the anti-aging probabilistic model. This model, being a binary classification model, underwent a 5-fold cross-validation process iterated 10 times. The selected cutoff score was the one that minimized false positives across all iterations. The resulting averaged score from this process was 0.904, establishing the threshold for subsequent evaluations.

The set of novel potential Caloric Restriction Mimetics (CRMs) that met the cutoff criteria of 0.904 consists of 43 parent CRMs. Below is the list of each parent CRM along with the associated novel potential compounds that qualified the cutoff, Aspirin-581, Curcumin-821,

Epicatechin-691, EGCG-1274, Gallic acid-347, Myricetin-725, Nicotinamide-137, Nicotinamide mononucleotide-134, Nicotinamide riboside-336, Piceatannol-425, Salicylic acid-465, Quercetin-807, 4,4'- Dimethoxychalcone-328, Acarbose-891, Anacardic acid-162, CAPE-584, CPI-613-38, Hydroxycitric acid-264, Garcinol-193, Glucosamine-341, SRT501-440, SB-204990-368, Oxaloacetic acid-148, Chitosan-1053, 2-Deoxy-D-glucose-491, D-Allulose-311, Dapagliflozin-330, Gymnemoside-2001, Mannoheptulase-503, Spermidine-121, Apigenin-696, Amentoflavone-639, Astraglin-463, Chrysin-518, Tryptophan-6, Nicotinic acid-194, Chalcones-45, Trehalose-992, Perhexiline maleate_first_congugate-12, Perhexiline maleate_second_congugate-179, 1,2,3- Benzenetricarboxylate-391, Metformin-12, Caffeic acid-292, Catechin-691.

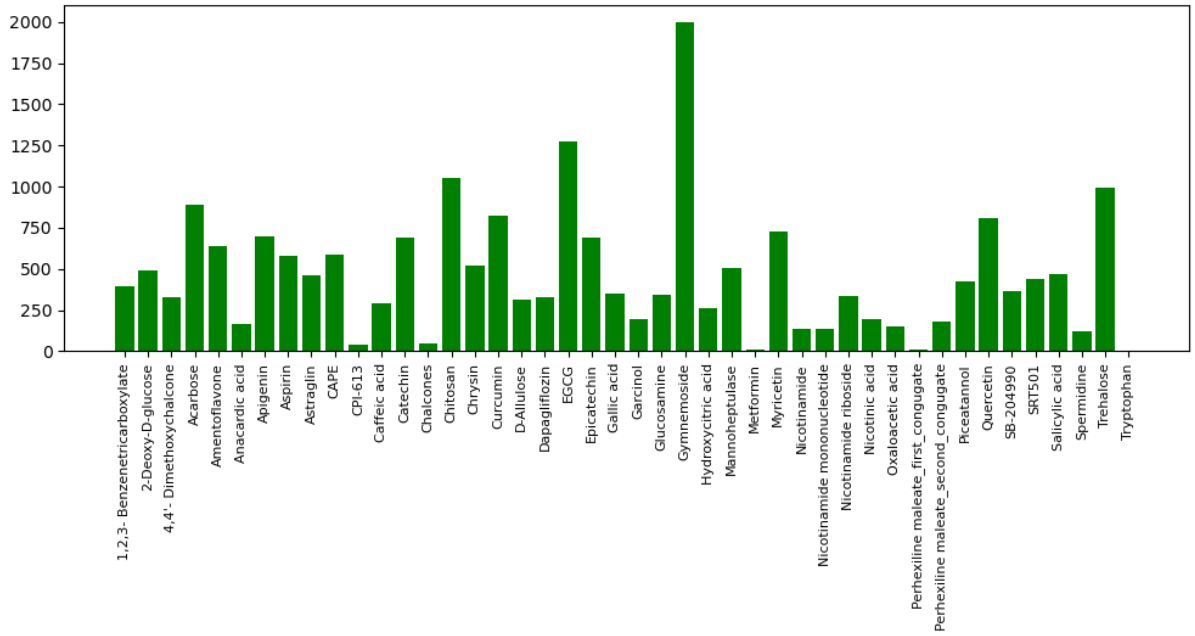


Figure 3.5. Bar plot representing the count of potential novel CRMs generated by respective molecules (43 CRMs) after the implementation of cutoff (0.904). The y axis represents count of potential novel CRMs generated and x axis depicts CRMs as seeds.

Among the parent compounds, Tryptophan showed the most limited number of qualifying novel compounds of 6 novel compounds, meeting the anti-aging probability cutoff. In contrast, Gymnemoside displayed the highest count, with 2001 novel compounds meeting the specified criteria. This information sheds light on the varied impact of anti-aging probability within the context of different parent compounds.

3.4) ZINC20 Database Screening

The ZINC20 Database, comprising around 2 billion compounds, underwent screening against the 20,443 qualified novel potential Caloric Restriction Mimetics (CRMs). In this screening process, 358 compounds were identified and matched within the ZINC20 database.

Further statistical analysis was conducted on these 358 screened compounds to associate them with their respective parent molecules. The results revealed that 26 screened compounds originated from Aspirin, 4 from Curcumin, 10 from Epicatechin, 11 from Gallic acid, 9 from Myricetin, 2 from Piceatannol, 97 from Salicylic acid, 44 from Quercetin, 20 from CAPE, 5 from Resveratrol, 1 from SB-204990, 4 from 2-Deoxy-D-glucose, 1 from Mannoheptulase, 6 from Epigallocatechin gallate, 3 from Spermidine, 24 from Apigenin, 5 from Astragalin,

57 from Chrysin, 7 from Nicotinic acid, 7 from Perhexiline maleate (second conjugate), 1 from 1,2,3-Benzenetricarboxylate, 2 from Caffeic acid, 12 from Catechin. The screening statistics unveiled a range from a minimum of 1 novel compound for 1,2,3-Benzenetricarboxylate to a maximum of 97 for Salicylic acid.

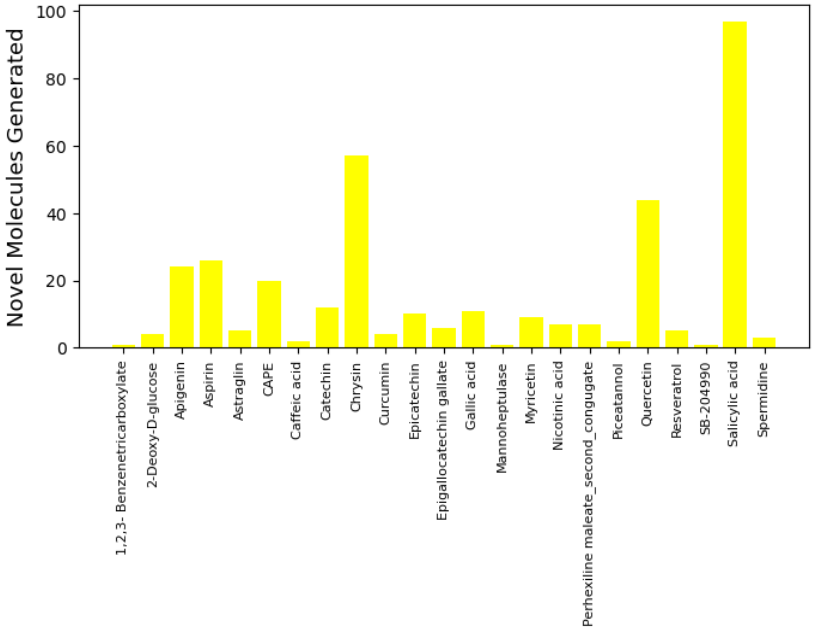


Figure 3.6. Bar plot representing the count of screened novel molecules from respective parent CRMs (23 CRMs). The x axis represents each parent CRMs and the y axis represents the count of screened novel molecules generated by respective CRMs and screened ton ZINC20 database.

Following the t-SNE analysis of 358 highly potential novel compounds screened alongside their 23 parent Caloric Restriction Mimetics (CRMs), utilizing the signaturizer features derived from the components ABCDE, a distinct separation emerged. The compounds and their parent CRMs manifested as four significantly different clusters, achieving a silhouette score of 0.658. This clustering provides valuable insights into the structural and bioactivity similarities among the screened and qualified novel compounds in comparison to their 24 parent CRMs.

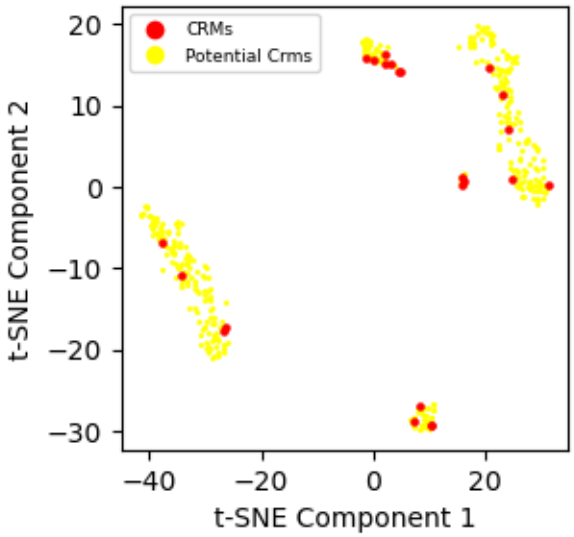


Figure 3.7. t-SNE visualization depicts the distribution of 24 parent Caloric Restriction Mimetics (CRMs) and 358 screened potential novel compounds. The x-axis corresponds to the first component of the embedding from the t-SNE, while the y-axis represents the second component. The data exhibits a distinct separation into four clusters, achieving a silhouette score of 0.658. The yellow points on the graph represent the 358 screened potential novel CRMs, and the red dots signify the 24 parent CRMs.

3.5) Selection of Promising ZINC20 Screened Novel Potential CRMs

3.5.1) Visualization

The bioactivity features for known 58 CRMs, 8784 pathway modulators, and the 358 identified novel molecules post-screening with ZINC20 were extracted. The bioactivity features contain the A - chemistry, capturing structural elements; B - targets, highlighting molecular targets and their impact; C - networks, indicating potential biological pathways; D - cells, emphasizing cellular responses; and E - clinics, encompassing clinical implications. The three datasets were combined for the understanding of the similarity between all compounds based on bioactivity space.

In this analysis, the A – chemistry component was intentionally excluded. This decision stemmed from closely monitoring the structural similarity between the parent Caloric Restriction Mimetics (CRMs) and the novel CRMs generated using the Graph Neural Network (GNN).

From Figure 3.8, it is evident that all 58 parent CRMs, along with the 358 highly potential ZINC20 screened novel CRMs and most pathway regulators, are closely clustered together. Unlike forming separate clusters, they converge into the same region. This observation suggests that the known 58 CRMs and the 358 highly potential ZINC30 screened novel CRMs share a similar bioactivity space. Consequently, it validates the bioactivity of the 358 novel CRMs.

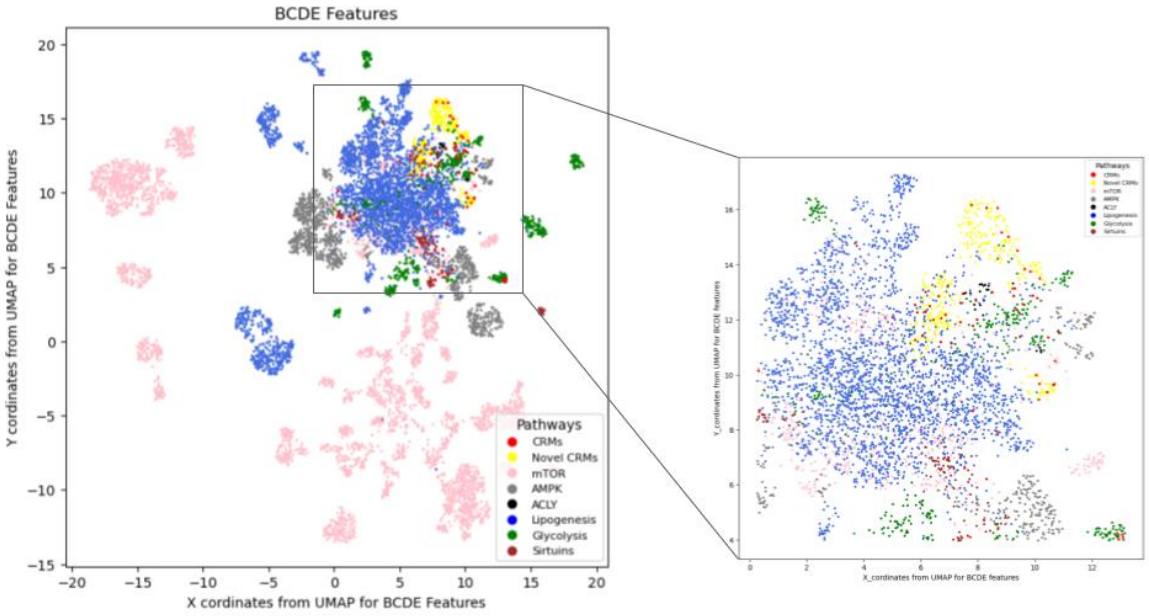


Figure 3.8, left, presents a visualization of 58 known Caloric Restriction Mimetics (CRMs), 358 screened CRMs, and 8784 pathway modulators based on bioactivity features using UMAP embeddings. Right, a focused view of the notable cluster where all the compounds of interest are present is shown. The x-axis corresponds to the first embedding from the UMAP, while the y-axis represents the second embedding. This visualization provides insights into the distribution and clustering of compounds based on their bioactivity features, emphasizing a specific cluster containing the compounds of interest.

The notable view of the cluster from figure 3.8 (right) we see that the 58 parent CRMs are spread across all the pathway modulators however there are certain CRMs which are not surrounded by all pathway modulators but are specific to some. The 358 CRMs are arranged around the parent CRMs along with certain pathway modulators.

3.5.2) Top 10 Neighbours for ZINC30 screened novel potential CRMs and Pathway Modulators

3.5.2.1) ZINC20 Screened Molecules

On counting the neighbours of parent CRMs, we found that Berberine, NAD+, Nicotinamide mononucleotide, Perhexiline maleate, Rapamycin, SRT1720, UK5099, A-485, Acarbose, Anacardic acid, C646, CPI-613, Garcinol, Radicicol, SRT2104, Chitosan, D-Allulose, Empagliflozin, Canagliflozin, Dapagliflozin, Bexagliflozin, Gymnemoside, Mannoheptulase has 0 neighbours from 358 screened molecules. Nicotinamide riboside and Nicotinic acid has 2 neighbours form 358 screened molecules. 3 neighbours were seen in Nicotineamide and Trehalose. 4 neighbours were seen for Hydroxycitric acid, Glucosamine, SB-204990, Rosiglitazone. 2-Deoxy-D-Glucose was the only 1 surrounded with 5 neighbours. 6 neighbours were seen for 1,2,3-benzenetricarboxylate, Acipimox, Metformin, Tryptophan. 7 neighbours were seen for Salicylic acid, Resveratrol, EGCG, 4,4'-dimethoxychalcone, and Chalcones. 8 neighbours were seen for Spermidine, Piceatannol, and Oxaloacetic acid. 9 neighbours were seen for Aspirin, Caffeic acid, Catechin, Epicatechin, Gallic acid, Myricetin, CAPE, Iodoacetate, and Astragalin, while the largest set of neighours of zinc20 screened compounds as 10 was found for Curcumin, Quercetin, Apigenin, Amentoflavone, and Chrysin.

3.5.2.2) Lipogenesis Modulators

Checking the neighbours of parent CRMs as lipogenesis pathway modulators we found that no neighbors were found for Resveratrol, Caffeic acid, Catechin, Curcumin, Epicatechin, Myricetin, NAD+, Nicotinamide mononucleotide, Quercetin, Rapamycin, SRT1720, CAPE, SRT2104, Empagliflozin, Canagliflozin, Dapagliflozin, Bexagliflozin, Apigenin, Amentoflavone, Astragalin, and Chrysin. For Aspirin, Spermidine, EGCG, Gallic acid, Piceatannol, Anacardic acid, CPI-613, Glucosamine, 2-Deoxy-D-glucose, and Iodoacetate, the number of neighbours seen is 1. For Hydroxycitric acid, Nicotinamide riboside, Garcinol, Oxaloacetic acid, and Trehalose, the number of neighbours seen is 2. For Salicylic acid, Acipimox, Metformin, 4,4'-Dimethoxychalcone, A-485, Acarbose, SB-204990, Chitosan, D-Allulose, Mannoheptulase, Tryptophan, and Chalcones, the number of neighbours seen is 3. For 1,2,3-Benzenetricarboxylate, Radicicol, and Gymnemoside, the number of neighbours seen is 4. Rosiglitazone is the only 1 which showed 5 neighbours of lipogenesis modulators. Further Berberine, Nicotinamide, and Nicotinic acid each had 7 neighbors. However, Perhexiline

maleate, C646, and UK5099 showed only compounds showing the highest neighbours of lipogenesis as 8,9, and 10 respectively.

3.5.2.3) Glycolysis Modulators.

Upon analysis of neighbours for glycolysis modulators in the cluster we found that almost 35 parent CRMs (Aspirin, Salicylic acid, Resveratrol, Spermidine, 1,2,3-Benzenetricarboxylate, Acipimox, Caffeic acid, Catechin, Curcumin, Epicatechin, Gallic acid, Metformin, Myricetin, Nicotinamide, Perhexiline maleate, Piceatannol, Quercetin, SRT1720, UK5099, 4,4'-Dimethoxychalcone, A-485, C646, CAPE, Glucosamine, SRT2104, Oxaloacetic acid, Rosiglitazone, Iodoacetate, Apigenin, Amentoflavone, Chrysin, Tryptophan, Nicotinic acid, Chalcones, and Trehalose) has no neighbours for glycolysis modulators. Hydroxycitric acid, Rapamycin, SB-204990, D-Allulose, and Astraglin have 1 neighbour. Berberine, EGCG, Nicotinamide riboside, 2-Deoxy-D-glucose, and Mannoheptulase have 2 neighbours. Acarbose, Anacardic acid, CPI-613, Garcinol, and Chitosan have 6 neighbours. NAD+, Dapagliflozin, and Bexagliflozin have 8 neighbours. Nicotinamide mononucleotide and Empagliflozin have 9 neighbours. However, 3 parent CRMs Radicicol, Canagliflozin, and Gymnemoside has 4,7 and 3 glycolysis modulators as neighbours.

3.5.2.4) AMPK Modulators.

Only 3 parent CRMs, namely C646, SRT2104, and Gymnemoside, have surrounding neighbors that are AMPK modulators, with 1, 6, and 2 AMPK modulators as neighbors, respectively. The remaining 55 CRMs do not have any AMPK modulators as neighbors in their surroundings.

3.5.2.5) mTOR Modulators.

Regarding mTOR modulators, we found that Berberine, CPI-613, Garcinol, Glucosamine, SRT2104, Rosiglitazone, and Trehalose each have 1 neighbor. Resveratrol, Perhexiline maleate, Radicicol, and SB-204990 each have 2 neighbors. Rapamycin and A-485 have 8 and 3 mTOR modulators as neighbors, respectively. However, the remaining 45 parent CRMs did not have any mTOR modulators as neighbors.

3.5.2.6) SIRTUINs Modulators.

For sirtuin modulators, we observed that NAD+, Nicotinamide riboside, Glucosamine, and Trehalose have only 1 sirtuin modulator as a neighbor. SRT1720 was found to have all 10 neighbors as sirtuin modulators. A-485 and SRT2104 have 4 and 3 modulators as neighbors,

respectively. On the contrary, the remaining 51 parent CRMs did not show a bioactivity relation to sirtuin modulators, as they did not have any sirtuin modulators as neighbours in their vicinity.

3.5.2.7) ACLY Modulators.

Among the ACLY modulators, NAD+, rapamycin, glucosamine, 2-Deoxy-D-glucose, and Trehalose each have 1 neighbour. Anacardic acid, Hydroxycitric acid, D-Allulose, and Mannoheptulase have 2, 3, 6, and 5 neighbours, respectively. The remaining 49 parent CRMs did not exhibit any ACLY modulators as neighbours.

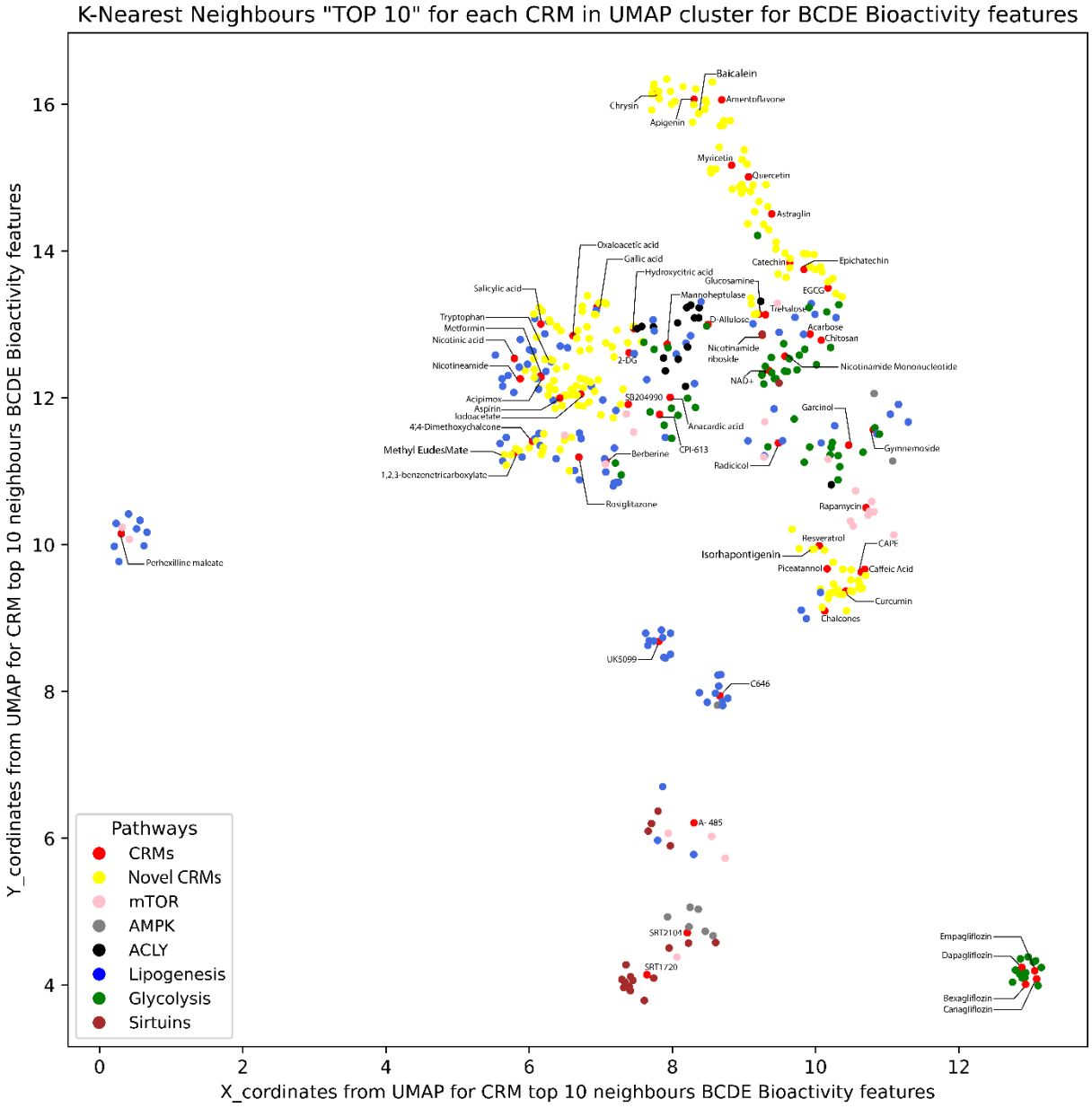


Figure 3.9. Top 10 Neighbours of each CRM along with the selected compounds.

3.5.3) Identification of Promising CRMs

We conducted a comprehensive literature analysis for each of the 358 screened molecules, assessing whether they exhibited indications of anti-aging properties and the strength of the supporting evidence. Consequently, 201 compounds were identified as toxic, causing irritation to species, or displaying a significant deviation from anti-aging properties, rendering them unsuitable for further consideration, especially in biological applications. On the other hand, 107 compounds demonstrated potential anti-aging, immunosuppressive, or neuroprotective effects at low to moderate levels. While these compounds were retained for secondary consideration, a more focused set of 50 compounds exhibited a high impact on anti-aging properties, supported by robust experimental evidence. These 50 compounds emerged as promising candidates with the potential to manifest a caloric restriction effect during experimentation, thereby contributing to longevity.

After an extensive literature survey and analysing the neighbours of all the CRMs, we have identified four promising compounds with a high likelihood of exhibiting a caloric restriction effect, potentially contributing to longevity and serving as novel caloric restriction mimetics. Baicalein (5,6,7-trihydroxy-2-phenyl-4H-1-benzopyran-4-one) is one selected compound having an anti-aging probability of 0.917 and is structurally similar to its parent CRM Chrysin and in terms of bioactivity Apigenin as neighbour. Methyl EudesMate (Methyl 3,4,5-trimethoxybenzoate) has an anti-aging probability of 0.924 and is structuraly similar to its

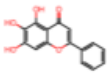
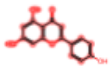
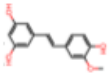
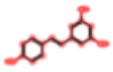
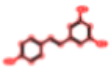
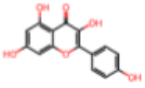
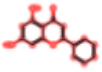
SELECTED PROMISING CRMs	PARENT CRMs	NEIGHBOUR CRMs
 Baicalein (AAP : 0.917)	 Chrysin	 Apigenin
 Methyl EudesMate (AAP : 0.924)	 Gallic acid	 1,2,3-benzenetricarboxylate
 Isorhapontigenin (AAP : 0.927)	 Resveratrol	 Resveratrol
 Kaempferol (AAP : 0.952)	 Chrysin	NONE

Table 3.3. Collection of Identified potential CRMs. The first column has promising compounds along with their anti-aging probability. The other columns represent the parent CRMs they were generated from and the neighbour CRMs which has a similarity in bioactivity. These compounds, strategically chosen for their high anti-aging probabilities, structural resemblances, and bioactivity correlations, are promising candidates for experimental investigations as potential novel caloric restriction mimetics.

parent CRM Gallic acid and bioactively with 1,2,3- benzenetricarboxylate. The next compound selected has an anti-aging probability of 0.927 and the common name is Isorhapontigenin ((E)-5-[2-(4-Hydroxy-3-methoxyphenyl)ethenyl]-1,3-benzenediol). Isorhapontigenin was generated from resveratrol and is also bioactively similar to resveratrol as well. The last compound we selected is Kaempferol (3,4′,5,7-tetrahydroxyflavone) it has anti-aging probability of 0.952 and is structurally similar to Chrysin. However, Kaempferol was not found in any of the top neighbours counted for CRMs. These compounds have been strategically chosen for their anti-aging probabilities, structural resemblances to parent CRMs, and bioactivity relationships with neighbouring compounds. They represent promising candidates for further experimental investigations into their caloric restriction effects and potential as novel caloric restriction mimetics.

4. Discussion

This study undertook a thorough investigation into potential Caloric Restriction Mimetics using a Generative AI methodology that includes Graph Neural Networks with an anti-aging probabilistic model. The Graph Neural Network architecture was reconstructed by integrating a pre-existing frame work from a tool MIMOSA with an anti-aging probabilistic model {mimosa ref}. It was selected above other generative AI tools like EvoMol (Leguy et al., 2020)and MolGAN (N. De Cao & Kipf, 2018) due to its pre-trained model, along with the combinatorial trainings of the model, and the optimization capacity of molecule generation which lacked in other mentioned tools.

Recognizing the limitations posed by the relatively small pool of known Caloric Restriction Mimetics (CRMs), comprising 58 compounds, conventional methods for building a model to identify CRM efficacy for a given compound are deemed ineffective and inefficient. To address this challenge, our study employed a novel approach utilizing Generative AI (GenAI) with bioactivity as a central feature. The 58 CRMs in question are a combination of two mechanisms, namely the intracellular signalling system and another mechanism targeting the energy metabolism system (H. Shintani et al., 2018). Each of these selected compounds falls into the class of glycolysis inhibitors, TCA modulators, NAD+ precursors, polyamines, polyphenols, and flavonoids (Hofer et al., 2021). In our investigation, these 58 known CRMs served as the foundation for generating 88,950 synthetic molecules. These synthesized compounds were designed to exhibit structural similarities to their parent CRMs while possessing inherent anti-aging properties.

To refine our selection process and identify the most promising CRMs with anti-aging potential, we implemented a stringent anti-aging probability cutoff of 0.904. This yielded 20,443 qualified molecules, still a substantial number for practical synthesis. To address this challenge, we employed a ZINC20 database screening, resulting in the identification of 358 novel potential CRMs worthy of further analysis. Recognizing the need for a more focused approach, we employed clustering visualization methods based on the modulators of pathways targeted by CRMs. This strategic step aimed to reduce the number of candidates while ensuring they promising novel CRMs to show a similar bioactivity. Additionally, we conducted a neighbourhood analysis for each parent CRM, further narrowing down the selection process. This involved assessing the richness of novel compounds in terms of pathway modulators within their respective neighbourhoods, considering the parent CRM as a crucial factor.

Following these steps, we meticulously selected four promising candidates for experimental analysis. This selection was based not only on the abundance of pathway modulators in the neighbourhood but also on the identity of the parent CRM and bioactivity as prime feature. To enhance the precision of our selection, a comprehensive literature survey was conducted to gather insights into the anti-aging potential of each compound. This meticulous approach ensures a more informed and accurate selection process, minimizing the risk of false positives and contributing to the robustness of our experimental design.

The selected compound Baicalein (5,6,7-trihydroxy-2-phenyl-4H-1-benzopyran-4-one) is a flavonoid extracted from Scutellaria species mostly predominantly in eastern countries such as Russia, Japan, China and Korea. Baicalein has been receiving significant interest for properties such as antioxidant, neuroprotective, anticancer, anti-ulcerative colitis, anti-inflammatory, antiviral, antidiabetic, eye-protective, cardioprotective, antithrombotic, and hepatoprotective properties (Dinda et al., 2017). Studies have highlighted the therapeutic potential of Baicalein, particularly in type-2 diabetic and obese mice. Administered at a dosage of 400mg/kg/d alongside a high-fat diet, Baicalein demonstrated efficacy in activating AMPK, effectively suppressing fatty acid synthesis, reducing cholesterol levels by downregulating SREBP-1c, and increasing PPAR-alpha (Pu et al., 2012). Baicalein has exhibited the ability to induce autophagy in various cancers through the AKT/mTOR pathway, impacting autophagic flux, and modulating Beclin-1 expression (Aryal et al., 2014). Thus, baicalein is chosen to be one of the top compounds for laboratory analysis.

Methyl EudesMate (Methyl 3,4,5-trimethoxybenzoate) is a naturally occurring compound identified in various organisms, including Buxus natalensis, Eucalyptus aggregata, and other sources. The synthesis of Methyl EudesMate has traditionally involved conventional methods of chemical synthesis and was identified to be a potential antiproliferative agent for certain types of tumour cells (Murty et al., 2016). However, in our study, we have identified this compound as having potential anti-aging properties based on its high anti-aging probability. The compound was generated using a less exhaustive molecular generation technique compared to the approach mentioned in Murty et al., 2016.

Isorhapontigenin, a resveratrol analog, is naturally present in the Chinese herb Gnetum cleistostachyum and in the seeds of Aiphanes aculeata. Its presence in grapes further establishes it as a potential source of antioxidants (Fernández-Marín et al., 2012). This compound has demonstrated anti-inflammatory properties and the ability to suppress the PI3k/Akt pathway,

potentially blocking mTOR (Yeo et al., 2017). Additionally, Isorhapontigenin has been found to activate SIRT-1, a protein associated with autophagy (Subedi et al., 2019). The influence of Isorhapontigenin on the mTOR and SIRT pathways aligns with the neighbour mapping for CRMs, further emphasizing its potential role in anti-aging mechanisms.

Lastly, Kaempferol is a naturally occurring flavonoid found and isolated from grapefruit, lotus ucrainicus, Delphinium, and other plant sources. It has demonstrated potential antimicrobial and antioxidant activity towards gram-positive bacteria and has also exhibited anti-SARS-CoV2 properties when extracted from Pistacia lentiscus (Selim et al., 2022) (Sun et al., 2023). Kaempferol and its derivatives have further demonstrated anti-fungal and anti-protozoal activities (Periferakis et al., 2022). Kaempferol has been observed to reduce reactive oxygen species (ROS) levels by activating HO-1 (Alam et al., 2020). Additionally, it has been implicated in influencing autophagy and endoplasmic reticulum (ER) stress (Ashrafizadeh et al., 2020). Consequently, Kaempferol is considered a potential CRM.

The selected compounds will undergo in-vitro and in-vivo testing, utilizing the model organism Saccharomyces cerevisiae. The bioassays will encompass various parameters, including a cell growth measurement assay employing turbidimetric analysis at 600nm, measurement of ROS (Reactive Oxygen Species) generation using DHET and H2DCF-DA fluorescence, and analysis of mitochondrial structure using the Rhodamine B fluorescence assay (Maslanka et al., 2017) (Arlia-Ciommo et al., 2018). To assess the chronological lifespan of yeast, a Cellular Lifespan (CLS) assay can be conducted. Furthermore, the concentration of glucose and lipid can be determined through glucose Calorimetric Assay, and fluorescence microscopy using BODIPY dye can provide insights into cellular aspects (Choi et al., 2017). These assays collectively aim to elucidate the bioactive effects of the selected compounds.

These chosen compounds' well-documented therapeutic effects, coupled with its diverse range of properties, However, it is crucial to note that these compounds have not been tested specifically for their efficacy as CRMs, thus, making them a promising subject for further exploration and experimentation in the context of caloric restriction mimetics and anti-aging research. Thus, the innovative methodology not only expands the dataset for analysis but also enables the exploration of a diverse chemical space to identify potential CRMs. By employing Generative AI and incorporating relevant features, our study contributed significantly to the identification and characterization of novel Caloric Restriction Mimetics with potential anti-aging effects.

5. References

Alam, W., Khan, H., Shah, M. A., Cauli, O., & Saso, L. (2020). Kaempferol as a Dietary Anti-Inflammatory Agent: Current Therapeutic Standing. *Molecules*, *25*(18), 4073. https://doi.org/10.3390/molecules25184073

Arlia-Ciommo, A., Leonov, A., Beach, A., Richard, V. R., Bourque, S. D., Burstein, M. T., Kyryakov, P., Gomez-Perez, A., Koupaki, O., Feldman, R., & Titorenko, V. I. (2018). Caloric restriction delays yeast chronological aging by remodeling carbohydrate and lipid metabolism, altering peroxisomal and mitochondrial functionalities, and postponing the onsets of apoptotic and liponecrotic modes of regulated cell death. *Oncotarget*, *9*(22), 16163–16184. https://doi.org/10.18632/oncotarget.24604

Aryal, P., Kim, K., Park, P.-H., Ham, S., Cho, J., & Song, K. (2014). Baicalein induces autophagic cell death through AMPK/ULK1 activation and downregulation of mTORC1 complex components in human cancer cells. *The FEBS Journal*, *281*(20), 4644–4658. https://doi.org/10.1111/febs.12969

Ashrafizadeh, M., Tavakol, S., Ahmadi, Z., Roomiani, S., Mohammadinejad, R., & Samarghandian, S. (2020). Therapeutic effects of kaempferol affecting autophagy and endoplasmic reticulum stress. *Phytotherapy Research : PTR*, *34*(5), 911–923. https://doi.org/10.1002/ptr.6577

Berman, A. Y., Motechin, R. A., Wiesenfeld, M. Y., & Holz, M. K. (2017). The therapeutic potential of resveratrol: a review of clinical trials. *NPJ Precision Oncology*, *1*. https://doi.org/10.1038/s41698-017-0038-6

Calder, P. C., Carding, S. R., Christopher, G., Kuh, D., Langley-Evans, S. C., & McNulty, H. (2018). A holistic approach to healthy ageing: how can people live longer, healthier lives? *Journal of Human Nutrition and Dietetics : The Official Journal of the British Dietetic Association*, *31*(4), 439–450. https://doi.org/10.1111/jhn.12566

Cao, H., Jia, Q., Yan, L., Chen, C., Xing, S., & Shen, D. (2019). Quercetin Suppresses the Progression of Atherosclerosis by Regulating MST1-Mediated Autophagy in ox-LDL-Induced RAW264.7 Macrophage Foam Cells. *International Journal of Molecular Sciences*, *20*(23). https://doi.org/10.3390/ijms20236093

Choi, K.-M., Hong, S.-J., van Deursen, J. M., Kim, S., Kim, K. H., & Lee, C.-K. (2017). Caloric Restriction and Rapamycin Differentially Alter Energy Metabolism in Yeast. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, *73*(1), 29–38. https://doi.org/10.1093/gerona/glx024

Chou, Y.-C., Lee, M.-S., Chiou, J.-M., Chen, T.-F., Chen, Y.-C., & Chen, J.-H. (2019). Association of Diet Quality and Vegetable Variety with the Risk of Cognitive Decline in Chinese Older Adults. *Nutrients*, *11*(7). https://doi.org/10.3390/nu11071666

De Cao, N., & Kipf, T. (2018). *MolGAN: An implicit generative model for small molecular graphs*.

Dinda, B., Dinda, S., DasSharma, S., Banik, R., Chakraborty, A., & Dinda, M. (2017). Therapeutic potentials of baicalin and its aglycone, baicalein against inflammatory

disorders. *European Journal of Medicinal Chemistry*, *131*, 68–80. https://doi.org/10.1016/j.ejmech.2017.03.004

Eisenberg, T., Abdellatif, M., Schroeder, S., Primessnig, U., Stekovic, S., Pendl, T., Harger, A., Schipke, J., Zimmermann, A., Schmidt, A., Tong, M., Ruckenstuhl, C., Dammbrueck, C., Gross, A. S., Herbst, V., Magnes, C., Trausinger, G., Narath, S., Meinitzer, A., … Madeo, F. (2016). Cardioprotection and lifespan extension by the natural polyamine spermidine. *Nature Medicine*, *22*(12), 1428–1438. https://doi.org/10.1038/nm.4222

Fernández-Marín, M. I., Guerrero, R. F., García-Parrilla, M. C., Puertas, B., Richard, T., Rodriguez-Werner, M. A., Winterhalter, P., Monti, J.-P., & Cantos-Villar, E. (2012). Isorhapontigenin: A novel bioactive stilbene from wine grapes. *Food Chemistry*, *135*(3), 1353–1359. https://doi.org/10.1016/j.foodchem.2012.05.086

Fu, T., Xiao, C., Li, X., Glass, L. M., & Sun, J. (2020). *MIMOSA: Multi-constraint Molecule Sampling for Molecule Optimization*.

Griffin, T. M., Valdez, T. V, & Mestril, R. (2004). Radicicol activates heat shock protein expression and cardioprotection in neonatal rat cardiomyocytes. *American Journal of Physiology. Heart and Circulatory Physiology*, *287*(3), H1081-8. https://doi.org/10.1152/ajpheart.00921.2003

Hatzivassiliou, G., Zhao, F., Bauer, D. E., Andreadis, C., Shaw, A. N., Dhanak, D., Hingorani, S. R., Tuveson, D. A., & Thompson, C. B. (2005). ATP citrate lyase inhibition can suppress tumor cell growth. *Cancer Cell*, *8*(4), 311–321. https://doi.org/10.1016/j.ccr.2005.09.008

Hofer, S. J., Davinelli, S., Bergmann, M., Scapagnini, G., & Madeo, F. (2021). Caloric Restriction Mimetics in Nutrition and Clinical Trials. *Frontiers in Nutrition*, *8*. https://doi.org/10.3389/fnut.2021.717343

Ingram, D. K., & Roth, G. S. (2021). Glycolytic inhibition: an effective strategy for developing calorie restriction mimetics. *GeroScience*, *43*(3), 1159–1169. https://doi.org/10.1007/s11357-020-00298-7

Kopp, W. (2019). How Western Diet And Lifestyle Drive The Pandemic Of Obesity And Civilization Diseases. *Diabetes, Metabolic Syndrome and Obesity : Targets and Therapy*, *12*, 2221–2236. https://doi.org/10.2147/DMSO.S216791

Leguy, J., Cauchy, T., Glavatskikh, M., Duval, B., & Da Mota, B. (2020). EvoMol: a flexible and interpretable evolutionary algorithm for unbiased de novo molecular generation. *Journal of Cheminformatics*, *12*(1), 55. https://doi.org/10.1186/s13321-020-00458-z

Leong, I. (2018). Sustained caloric restriction in health. *Nature Reviews Endocrinology*, *14*(6), 322–322. https://doi.org/10.1038/s41574-018-0008-2

Liu, J., & Li, L. (2019). Targeting Autophagy for the Treatment of Alzheimer's Disease: Challenges and Opportunities. *Frontiers in Molecular Neuroscience*, *12*. https://doi.org/10.3389/fnmol.2019.00203

López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2013). The Hallmarks of Aging. *Cell*, *153*(6), 1194–1217. https://doi.org/10.1016/j.cell.2013.05.039

López-Otín, C., Galluzzi, L., Freije, J. M. P., Madeo, F., & Kroemer, G. (2016). Metabolic Control of Longevity. *Cell*, *166*(4), 802–821. https://doi.org/10.1016/j.cell.2016.07.031

Madeo, F., Carmona-Gutierrez, D., Hofer, S. J., & Kroemer, G. (2019). Caloric Restriction Mimetics against Age-Associated Disease: Targets, Mechanisms, and Therapeutic Potential. In *Cell Metabolism* (Vol. 29, Issue 3, pp. 592–610). Cell Press. https://doi.org/10.1016/j.cmet.2019.01.018

Maslanka, R., Kwolek-Mirek, M., & Zadrag-Tecza, R. (2017). Consequences of calorie restriction and calorie excess for the physiological parameters of the yeast Saccharomyces cerevisiae cells. *FEMS Yeast Research*, *17*(8). https://doi.org/10.1093/femsyr/fox087

Mattison, J. A., Colman, R. J., Beasley, T. M., Allison, D. B., Kemnitz, J. W., Roth, G. S., Ingram, D. K., Weindruch, R., de Cabo, R., & Anderson, R. M. (2017). Caloric restriction improves health and survival of rhesus monkeys. *Nature Communications*, *8*(1), 14063. https://doi.org/10.1038/ncomms14063

Maziarka, Ł., Pocha, A., Kaczmarczyk, J., Rataj, K., Danel, T., & Warchoł, M. (2020). Mol-CycleGAN: a generative model for molecular optimization. *Journal of Cheminformatics*, *12*(1), 2. https://doi.org/10.1186/s13321-019-0404-1

Most, J., Tosti, V., Redman, L. M., & Fontana, L. (2017). Calorie restriction in humans: An update. *Ageing Research Reviews*, *39*, 36–45. https://doi.org/10.1016/j.arr.2016.08.005

Murty, M. S. R., Penthala, R., Polepalli, S., & Jain, N. (2016). Synthesis and biological evaluation of novel resveratrol-oxadiazole hybrid heterocycles as potential antiproliferative agents. *Medicinal Chemistry Research*, *25*(4), 627–643. https://doi.org/10.1007/s00044-016-1514-1

Periferakis, A., Periferakis, K., Badarau, I. A., Petran, E. M., Popa, D. C., Caruntu, A., Costache, R. S., Scheau, C., Caruntu, C., & Costache, D. O. (2022). Kaempferol: Antimicrobial Properties, Sources, Clinical, and Traditional Applications. *International Journal of Molecular Sciences*, *23*(23), 15054. https://doi.org/10.3390/ijms232315054

Pietrocola, F., Pol, J., Vacchelli, E., Rao, S., Enot, D. P., Baracco, E. E., Levesque, S., Castoldi, F., Jacquelot, N., Yamazaki, T., Senovilla, L., Marino, G., Aranda, F., Durand, S., Sica, V., Chery, A., Lachkar, S., Sigl, V., Bloy, N., … Kroemer, G. (2016). Caloric Restriction Mimetics Enhance Anticancer Immunosurveillance. *Cancer Cell*, *30*(1), 147–160. https://doi.org/10.1016/j.ccell.2016.05.016

Pu, P., Wang, X.-A., Salim, M., Zhu, L.-H., Wang, L., Chen, K.-J., Xiao, J.-F., Deng, W., Shi, H.-W., Jiang, H., & Li, H.-L. (2012). Baicalein, a natural product, selectively activating AMPKα(2) and ameliorates metabolic disorder in diet-induced mice. *Molecular and Cellular Endocrinology*, *362*(1–2), 128–138. https://doi.org/10.1016/j.mce.2012.06.002

Rigoni, D., Navarin, N., & Sperduti, A. (2020). *Conditional Constrained Graph Variational Autoencoders for Molecule Design*.

Selim, S., Almuhayawi, M. S., Alharbi, M. T., Al Jaouni, S. K., Alharthi, A., Abdel-Wahab, B. A., Ibrahim, M. A. R., Alsuhaibani, A. M., Warrad, M., & Rashed, K. (2022). Insights into the Antimicrobial, Antioxidant, Anti-SARS-CoV-2 and Cytotoxic Activities of Pistacia lentiscus Bark and Phytochemical Profile; In Silico and In Vitro Study. *Antioxidants*, *11*(5), 930. https://doi.org/10.3390/antiox11050930

Shintani, H., Shintani, T., Ashida, H., & Sato, M. (2018). Calorie Restriction Mimetics: Upstream-Type Compounds for Modulating Glucose Metabolism. *Nutrients*, *10*(12). https://doi.org/10.3390/nu10121821

Shintani, T., Kosuge, Y., & Ashida, H. (2018). Glucosamine Extends the Lifespan of Caenorhabditis elegans via Autophagy Induction. *Journal of Applied Glycoscience*, *65*(3), 37–43. https://doi.org/10.5458/jag.jag.JAG-2018_002

Shintani, T., Sakoguchi, H., Yoshihara, A., Izumori, K., & Sato, M. (2017). d-Allulose, a stereoisomer of d-fructose, extends Caenorhabditis elegans lifespan through a dietary restriction mechanism: A new candidate dietary restriction mimetic. *Biochemical and Biophysical Research Communications*, *493*(4), 1528–1533. https://doi.org/10.1016/j.bbrc.2017.09.147

Subedi, L., Teli, M. K., Lee, J. H., Gaire, B. P., Kim, M., & Kim, S. Y. (2019). A Stilbenoid Isorhapontigenin as a Potential Anti-Cancer Agent against Breast Cancer through Inhibiting Sphingosine Kinases/Tubulin Stabilization. *Cancers*, *11*(12), 1947. https://doi.org/10.3390/cancers11121947

Sun, Y., Tao, Q., Cao, Y., Yang, T., Zhang, L., Luo, Y., & Wang, L. (2023). Kaempferol has potential anti-coronavirus disease 2019 (<scp>COVID</scp> -19) targets based on bioinformatics analyses and pharmacological effects on endotoxin-induced cytokine storm. *Phytotherapy Research*, *37*(6), 2290–2304. https://doi.org/10.1002/ptr.7740

Trammell, S. A. J., Schmidt, M. S., Weidemann, B. J., Redpath, P., Jaksch, F., Dellinger, R. W., Li, Z., Abel, E. D., Migaud, M. E., & Brenner, C. (2016). Nicotinamide riboside is uniquely and orally bioavailable in mice and humans. *Nature Communications*, *7*(1), 12948. https://doi.org/10.1038/ncomms12948

Yeo, S. C. M., Fenwick, P. S., Barnes, P. J., Lin, H. S., & Donnelly, L. E. (2017). Isorhapontigenin, a bioavailable dietary polyphenol, suppresses airway epithelial cell inflammation through a corticosteroid-independent mechanism. *British Journal of Pharmacology*, *174*(13), 2043–2059. https://doi.org/10.1111/bph.13803