In silico insights: QSAR modeling of TBK1 kinase inhibitors for enhanced drug discovery

Julian M. Ivanov,[1] Rumiana Tenchov,[1] Krittika Ralhan,[2] Kavita A. Iyer,[2] Shivangi Agarwal,[2] and Qiongqiong Angela Zhou[1]*

[1]CAS, A Division of the American Chemical Society, Columbus, Ohio 43210, United States

[2]ACS International India Pvt. Ltd., Pune 411044, India

*Corresponding author: qzhou@cas.org

Abstract

TBK1, or TANK-binding kinase 1, is an enzyme that functions as a serine/threonine protein kinase. It plays a crucial role in various cellular processes, including the innate immune response to viruses, cell proliferation, apoptosis, autophagy, and anti-tumor immunity. Dysregulation of TBK1 activity can lead to autoimmune diseases, neurodegenerative disorders, and cancer. Due to its central role in these critical pathways, TBK1 is a significant focus of research for therapeutic drug development.

In this paper, we explore data from the CAS Content Collection regarding TBK1 and its implication in a large assortment of diseases and disorders. With the demand for developing efficient TBK1 inhibitors been outlined, we focus on utilizing machine learning approach for developing predictive models for TBK1 inhibition, derived from the fragment-functional analysis descriptors. Using the extensive CAS Content Collection we assembled a training set of TBK1 inhibitors with experimentally measured IC50 values. We explored several machine learning techniques combined with various molecular descriptors to derive and select the best TBK1 inhibitor QSAR models. Certain significant structural alerts that potentially contribute to inhibition of TBK1 are outlined and discussed. The merit of the article stem from identifying the most adequate TBK1 QSAR models and subsequent successful development of advanced positive training data to facilitate and enhance drug discovery for an important therapeutic target such as TBK1 inhibitors, based on extensive, wide-ranging set of scientific information provided by the CAS Content Collection.

Key words: TBK1; kinase inhibitor; machine learning; QSAR modeling; molecular descriptor; cancer; autoimmune disease

Introduction

Kinases are enzymes that catalyze the transfer of a phosphate group from a high-energy molecule, such as adenosine triphosphate (ATP), to a specific substrate molecule, typically a protein, lipid, or carbohydrate.[1] This process is known as phosphorylation and plays a crucial role in cellular signaling pathways, regulating various cellular functions including metabolism, growth, differentiation, and cell death.[2] Kinases are essential for transmitting signals within cells and coordinating cellular responses to extracellular stimuli. They are involved in a wide range of physiological processes and are often targets for drug development in treating diseases such as cancer and inflammatory disorders.[3] Tank-binding kinase 1 (TBK1) is one such enzyme with kinase activity. Encoded by the TBK1 gene in humans, it is a pivotal serine/threonine kinase that orchestrates a variety of critical cellular processes, including innate immunity, inflammation, autophagy, and cell survival/proliferation. [4-12]

Since its discovery, TBK1 has emerged as a central node in the signaling pathways that underpin the defense mechanisms of the body against pathogens and maintain cellular homeostasis. Its ability to phosphorylate and activate key transcription factors, such as interferon regulatory factors (IRFs) and nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB), underscores its essential role in immune responses. [13-15]

The involvement of TBK1 in immune signaling begins with its activation by pattern recognition receptors (PRRs), which detect pathogen-associated molecular patterns (PAMPs) and damage-associated molecular patterns (DAMPs). [16-18] Upon activation, TBK1 phosphorylates downstream effectors to induce the expression of type I interferons and pro-inflammatory cytokines, which are crucial for antiviral defense and the modulation of inflammatory responses. Furthermore, TBK1 is integral to the autophagy pathway, where it regulates the selective degradation of ubiquitinated proteins and damaged organelles, thereby contributing to cellular quality control and stress responses. [19-21]

The significance of TBK1 extends beyond normal physiological functions; its dysregulation is implicated in a range of pathological conditions. Mutations and altered expression of TBK1 have been associated with neurodegenerative diseases such as amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD), highlighting its role in neuronal homeostasis. Additionally, TBK1's involvement in oncogenic pathways links it to cancer progression and survival, making it a potential target for therapeutic intervention. [22, 23]

Given its central role in multiple signaling pathways, TBK1 represents a critical juncture in the regulation of immune responses, inflammation, and cell survival. This paper aims to provide a comprehensive overview of TBK1's structural features, its regulatory mechanisms, and its diverse functional roles in health and disease. By elucidating the molecular intricacies of TBK1, we can better understand its contributions to cellular homeostasis and its potential as a therapeutic target in various disease contexts. Its intricate functions make it an intriguing target for further research and therapeutic interventions. [22, 24-27]

The objective of this article is to utilize machine learning approach for developing valuable predictive models for TBK1 inhibition – a topic of utmost importance as an attractive

target for drug development. We explored data from the CAS Content Collection [28], the world's largest human expert-curated collection of scientific data, regarding TBK1 and its implication in a large assortment of diseases and disorders. With the demand for developing efficient TBK1 inhibitors outlined, we further focused on developing predictive analytics for TBK1 inhibition, derived from the fragment-functional analysis descriptors. For such fragment-based methodology, the molecular descriptors are the structural alerts obtained by splitting the chemical structures of the training set into all possible sub-fragments. Based on data from the CAS Content Collection, we assembled a training set known TBK1 inhibitors whose IC50 values have been determined experimentally. We explored several machine learning techniques combined with various molecular descriptors to derive and select the best TBK1 QSAR models. We also considered various aspects of TBK1 structure and potential inhibitors. Certain significant structural alerts potentially important for TBK1 inhibition have been outlined and discussed. The novelty and merit of the article stem from identifying the most adequate TBK1 QSAR models and subsequent successful development of advanced positive training data to facilitate and enhance drug discovery for an important therapeutic target such as TBK1 inhibitors, based on extensive, wide-ranging set of scientific information provided by the CAS Content Collection.

TBK1 overview and landscape of research progress

Importance of TBK1 in cellular processes

TBK1, or TANK-binding kinase 1, is a multifunctional protein kinase that plays a crucial role in various cellular processes. [14, 29-32]

- TBK1 is a key regulator of the innate immune response to viral and bacterial infections. It is activated upon detection of viral nucleic acids by pattern recognition receptors (PRRs) in the cytoplasm. Activated TBK1 phosphorylates the transcription factor IRF3 (interferon regulatory factor 3), leading to its dimerization and translocation to the nucleus. IRF3 induces the expression of type I interferons and other antiviral genes, which help to limit viral replication and spread. [33, 34]

- TBK1 is involved in the regulation of inflammatory responses. [7, 8] It can activate the transcription factor NF-κB (nuclear factor kappa-light-chain-enhancer of activated B cells) by phosphorylating IκBα (inhibitor of NF-κB alpha), leading to its degradation and the subsequent release of NF-κB. NF-κB translocates to the nucleus and induces the expression of pro-inflammatory cytokines, chemokines, and other inflammatory mediators.

- TBK1 plays a role in the regulation of autophagy, a cellular process involved in the degradation and recycling of damaged organelles and proteins. [9, 10] TBK1 phosphorylates autophagy-related proteins such as ULK1 (Unc-51 like autophagy activating kinase 1) and OPTN (optineurin), promoting autophagosome formation and maturation.

- TBK1 signaling contributes to cell survival and proliferation in various contexts. It can activate the AKT (protein kinase B) pathway, which promotes cell survival and growth, and it regulates the expression of anti-apoptotic genes. [35, 36]

– TBK1 is involved in the regulation of metabolic processes, including glucose metabolism and lipid homeostasis. [37-39] It can modulate insulin signaling pathways and influence the expression of genes involved in metabolism.

– TBK1 has been implicated in the cellular response to DNA damage. [27, 40] It can phosphorylate and activate the DNA repair protein BRCA1 (breast cancer type 1 susceptibility protein), contributing to the repair of DNA double-strand breaks.

TBK 1 structure

TBK1 is a non-canonical IKK kinase, which phosphorylates the nuclear factor kB. Consisting of 729 amino acids, TBK1 includes four main domains: an N-terminal kinase domain (KD; 1-307), a ubiquitin-like domain (ULD; 308-384), and two coiled-coil domains (CCD1; 407-657 and CCD2; 659-713).[41]

The KD is responsible for the catalytic activity of TBK1. It adopts a typical protein kinase fold and contains the active site for phosphorylation reactions. The ULD, located adjacent to the KD, plays a role in regulating TBK1 activity and interactions. The CCD1, also referred to as scaffold dimerization domain (SDD), harbors a leucine zipper domain (LZ; 499-527) and a helix-loop-helix domain (HLH; 591-632), both of which mediate dimerization of TBK1 molecules. It forms extensive interactions with both the KD and ULD, contributing to the overall stability of the TBK1 structure. [42] The CCD2 at the C-terminus holds an adaptor-binding motif which assists the interaction of TBK1 with adaptor proteins, such as TANK, NAK–associated protein, TBKBP1, or optineurin. [42-46] TBK1 forms an intimate dimer through extensive interactions between the SDDs, KDs, and ULDs of two TBK1 molecules. This dimerization is crucial for TBK1 function and regulation. [30, 41, 43, 47-49]

The various domains of TBK1 allow it to participate in multiple cellular processes, including: (i) immune response: TBK1 phosphorylates and activates IRF3 and IRF7, leading to the production of type I interferons in response to viral infection; (ii) NF-κB pathway: TBK1 can activate the NF-κB pathway, promoting the expression of pro-inflammatory cytokines; (iii) autophagy: by phosphorylating autophagy-related proteins, TBK1 regulates the degradation of cellular components, which is important for cellular homeostasis and defense against pathogens; (iv) cell proliferation and survival: TBK1 is involved in pathways that control cell growth and survival, linking it to cancer biology. [41, 43, 48, 49]

Journal publications related to TBK1 have grown rapidly and consistently over the last two decades, nearly doubling between 2020 and 2023 as seen from data leveraged from the CAS Content Collection [28]. This active and rapid increase is indicative of interest in TBK1 from the scientific community. Growth in patent publications on the other hand has been slow, speaking to potential difficulties in targeting TBK1 (elaborated briefly in later sections). The journal-to-patent ratio was >12X for the year 2023 (Figure 1).

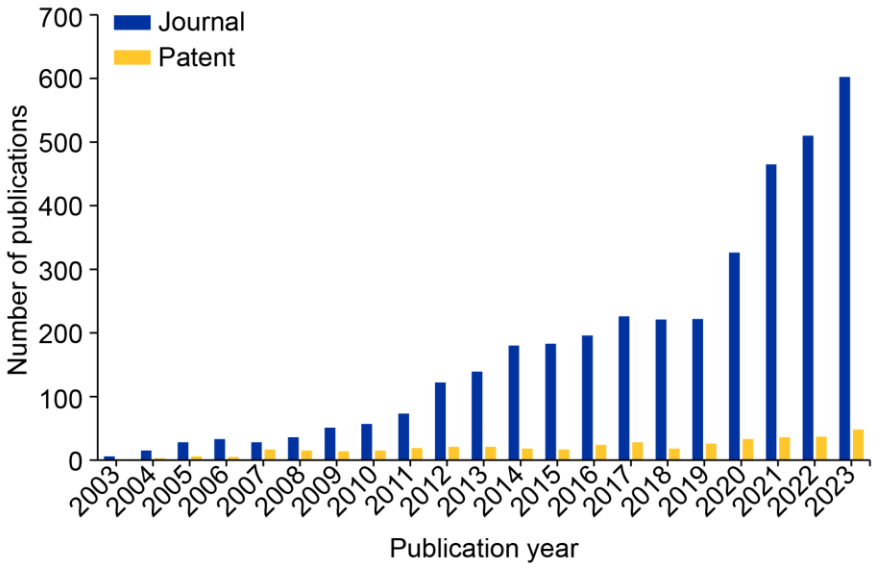


Figure 1. Publications related to TBK1 from the CAS Content Collection for the period 2003-2023.

Patenting activity is dominated by corporate players as compared to academics (Figure 2). Merck, MetaProteomics, and Arvinas have the highest number of patent applications among commercial entities, while Max-Planck Institute, the Korea Institute of Science and Technology (KIST), and Purdue University are the leaders among the academic organizations.

Patent assignees (Commercial)	Number of patents (2003-2023)
Merck	16
MetaProteomics	11
Arvinas	7
Flagship Pioneering Inno	6
Novartis	5
Domainex	5
Gilead Sciences	4
Bayer	4
Asuragen	3
Shenzhen Chipscreen Biosciences	3

Patent assignees (Non-commercial)	Number of patents (2003-2023)
Max-Planck	9
KIST	7
Purdue	5
Osaka University	5
University of Michigan	5
University of California	5
Massachusetts Institute of Technology	4
University of Washington	3
Children's hospital of Philadelphia	3
Kyoto University	3

Figure 2. Leading patent assignees in the field of TBK1-related research based on data from CAS Content Collection for the period 2003-2023. Patent assignees have been separated into two categories – commercial (left panel) and non-commercial (right panel).

Using data from CAS Content Collection, we determined co-occurrences of TBK1 with other protein targets as well as diseases (shown in the Sankey graph in Figure 3). Below we discuss briefly the role of TBK1 in diseases it co-occurs with in our data.

Implication of TBK1 in diseases

Due to its role in immune regulation, inflammation, autophagy, and cellular stress responses, dysregulation or dysfunction of TBK1 is implicated in various diseases.

- TBK1 has been associated with **autoimmune diseases** such as systemic lupus erythematosus (SLE), where dysregulated immune responses contribute to tissue damage and inflammation. TBK1 may play a role in the activation of immune cells and the production of inflammatory cytokines in autoimmune conditions. [31, 50, 51] TBK1 is also responsible for regulating the immune response in multiple sclerosis, rheumatoid arthritis, inflammatory bowel disease, and psoriasis among others

- There is growing evidence linking TBK1 dysfunction to **neurodegenerative disorders** such as amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). Mutations in the TBK1 gene have been identified in individuals with ALS-FTD spectrum disorders, suggesting a potential role for TBK1 in the pathogenesis of these diseases. [15, 52]

- TBK1 is involved in the host immune response to **viral infections** by inducing the production of type I interferons and other antiviral proteins. Dysregulation of TBK1 signaling pathways may impact the ability of the immune system to control viral infections, leading to increased susceptibility to viral diseases. [53-55]

- TBK1 has been implicated in **cancer** development and progression in various ways. It can promote tumor growth, survival, and metastasis by enhancing cell proliferation, inhibiting apoptosis, and modulating the tumor microenvironment. Dysregulated TBK1 signaling has been observed in several types of cancer, including lung cancer, breast cancer, and melanoma. [34, 56, 57] For instance, TBK1 has been shown to activate the NF-κB pathway, which is generally involved in inflammation, cell proliferation, and resistance to apoptosis in cells. This can eventually lead to tumor growth and proliferation. TBK1 phosphorylates STING, which in turn recruits IRF3 for phosphorylation by TBK1. Phosphorylated IRF3 dimerizes and then enters the nucleus, where it functions with NF-kB to turn on the expression of type I interferons and other immunomodulatory molecules. TBK1 activates several signaling pathways such as AKT-mTOR and MYC that are responsible for tumor cell survival. TBK1 also plays an important role in the IRF3 pathway that is important in generating an antiviral response, any disruption in these pathways is linked to tumorigenesis. In addition, TBK1 is also known to activate transcription factors responsible for epithelial-mesenchymal transition (EMT) which can lead to cancer metastasis. It can also influence the tumor microenvironment by modulating immune cell activity and causing cytokine release which can cause tumor progression. It is also implicated in resistance to different cancer therapies as it can inhibit apoptosis and promote cancer cell survival. TBK1 is also involved in crosstalk between other cancer-linked pathways in the cell such as KRAS, PI3K, and EGFR pathway. Due to all these factors any mutation inTBK1 gene can be linked to cancer.

- TBK1 may also play a role in metabolic regulation and energy homeostasis. Dysregulation of TBK1 activity has been linked to **metabolic disorders** such as obesity and insulin resistance, although the precise mechanisms remain to be fully elucidated. [38, 39, 58]

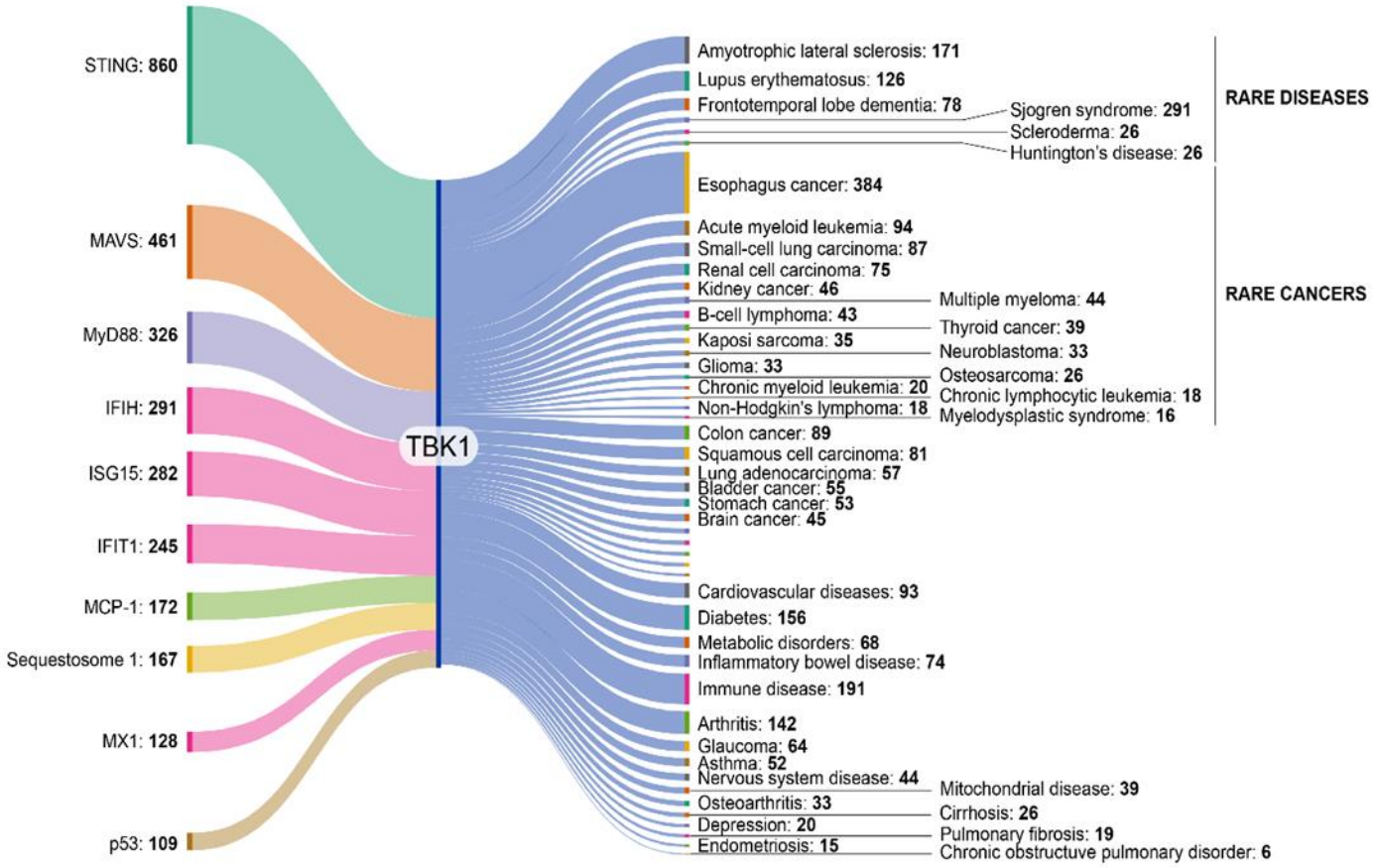


Figure 3. Sankey graph showing co-occurrences of TBK1 with other proteins (left column) and diseases (right column) based on CAS indexing. Data includes patent and journal publications from the CAS Content Collection for the period 2003-2023.

The need of TBK1 inhibitors

Because of its implication in a large assortment of diseases and disorders, TBK1 is a promising therapeutic target for the development of drugs aimed at modulating immune responses and treating various diseases. [59-61] TBK1 inhibitors are compounds that can selectively block the activity of TBK1, potentially offering therapeutic benefits in various diseases where TBK1 is dysregulated or overactive. Research into TBK1 inhibitors has gained considerable attention due to the role of TBK1 in immunity, inflammation, and other cellular processes, as well as its implication in diseases such as cancer, autoimmune disorders, and neurodegenerative conditions. [14, 15, 34] Several small molecule inhibitors targeting TBK1 have been developed and studied in preclinical and early clinical research. These inhibitors typically work by binding to specific regions of TBK1 and interfering with its kinase activity, thus preventing its downstream signaling and biological effects. [25, 26, 59, 62, 63]

One of the primary motivations for developing TBK1 inhibitors is their potential as anti-cancer agents. TBK1 has been implicated in promoting tumor growth and progression in certain types of cancer by enhancing cell survival, proliferation, and metastasis. Inhibiting TBK1 activity could

potentially suppress these cancer-promoting effects and enhance the efficacy of other anti-cancer treatments. [29, 34, 60]

Furthermore, TBK1 inhibitors are also being investigated for their potential in treating autoimmune disorders and inflammatory conditions. Since TBK1 plays a role in the regulation of immune responses and inflammation, inhibiting its activity could help modulate aberrant immune activation and reduce inflammation associated with autoimmune diseases. [31, 50, 64]

Despite promising preclinical results, the development of TBK1 inhibitors faces several challenges, including achieving sufficient selectivity to minimize off-target effects, optimizing pharmacokinetic properties for effective delivery and distribution in the body, and ensuring safety and tolerability in clinical settings. Overall, while TBK1 inhibitors represent a promising avenue for therapeutic intervention in various diseases, further research and development efforts are needed to fully realize their clinical potential and address the challenges associated with their development.

Why there are no successful TBK1 inhibitors on the market

Our search failed to identify any current FDA-approved TBK1 inhibitors available on the market. The immunomodulatory drug Amlexanox which was approved by FDA for the treatment of aphthous ulcers in 2004 [65], was later discontinued in 2014 due to the associated undesired adverse reactions of the formulation. [66] Amlexanox-loaded nanoliposome formulation are being currently developed as a potential alternative for the localized oromucosal delivery of the drug. [66]

The lack of FDA-approved TBK1 inhibitors at present can be attributed to several factors:

(i) TBK1 is involved in multiple cellular processes, including immune regulation, inflammation, autophagy, and stress responses. The complexity of TBK1 signaling complicates the progress in TBK1 inhibitors. Developing inhibitors that selectively target TBK1's pathological functions while sparing its essential physiological roles is challenging.

(ii) Achieving sufficient selectivity is crucial when developing kinase inhibitors to minimize off-target effects and potential toxicities. Designing compounds that specifically inhibit TBK1 without interfering with other kinases or cellular pathways can be difficult.

(iii) The process of drug development, from discovery to approval, is lengthy and resource intensive. Developing TBK1 inhibitors with desirable pharmacokinetic properties, efficacy, and safety profiles requires substantial investment in preclinical research, clinical trials, and regulatory approval processes.

(iv) Even if promising TBK1 inhibitors are identified in preclinical studies, their clinical efficacy and safety must be rigorously evaluated in human clinical trials. Negative results or unforeseen complications in clinical trials can delay or halt the development of candidate inhibitors.

(v) The prioritization of research and funding allocation in the pharmaceutical industry and academic institutions also influences the pace of drug development. While TBK1 inhibitors hold promise for therapeutic intervention in various diseases, competing priorities and resource constraints may impact the rate of progress in this area.

Despite these mitigating factors, a few key players that are involved in research related to TBK1 are shown in Figure 2 and include well-known companies such as Merck, Novartis, Gilead Sciences and Bayer. Other key players include the PROTAC-focused biotechnology company Arvinas [67], the Drug Discovery CRO Domainex [68], Asuragen [69] which appears to be centered around molecular diagnostics in oncology and other fields as well as the China-based drug development company Shenzhen Chipscreen Biosciences [70]. Examples of patents by these organizations mostly consist of exploring various scaffolds including pyrimidine- (WO2019079375

[71]; US8962609 [72]) and heteroarylbenzimidazole-based (WO2017207534 [73]) with the aim of developing small molecule inhibitors (WO2017106556 [74]). Other examples include development of proteolysis targeting chimeras (PROTACs) against TBK1 by Arvinas (WO2016197114 [75]).

How IC50 values work for inhibitors

The IC50 (half-maximal inhibitory concentration) value is a measure used in pharmacology and biochemistry to quantify the potency of an inhibitor, particularly in enzyme inhibition studies. It represents the concentration of an inhibitor required to inhibit 50% of the activity of a biological or biochemical target, such as an enzyme or a cellular process. It represents the most widely used and informative measure of a drug's efficacy. [76, 77]

In a typical experimental setup to determine the IC50 value of an inhibitor, varying concentrations of the inhibitor are tested against a fixed concentration of the target enzyme or biological process. The activity of the target is measured in the presence of each inhibitor concentration. The data obtained from the experiment is used to plot a dose-response curve, where the concentration of the inhibitor is plotted on the x-axis, and the remaining activity of the target (expressed as a percentage of the uninhibited activity) is plotted on the y-axis. As the concentration of the inhibitor increases, the activity of the target decreases. The IC50 value is determined by finding the concentration of the inhibitor that corresponds to 50% inhibition of the target activity on the dose-response curve. This concentration is the IC50 value. It represents the potency of the inhibitor—the lower the IC50 value, the more potent the inhibitor, as it achieves significant inhibition at lower concentrations. A low IC50 value indicates that the inhibitor is effective at lower concentrations, meaning it can achieve significant inhibition of the target with relatively low doses. Conversely, a high IC50 value indicates that higher concentrations of the inhibitor are needed to achieve the same level of inhibition, suggesting lower potency. IC50 values can be used to compare the potency of different inhibitors targeting the same biological target. The inhibitor with the lower IC50 value is generally considered more potent and may be more suitable for further development as a therapeutic agent or research tool.

In brief, the IC50 value is a quantitative measure of the potency of an inhibitor, representing the concentration required to inhibit 50% of the activity of a biological target. It is an essential parameter in drug discovery and enzyme inhibition studies, helping researchers evaluate and compare the effectiveness of different inhibitors.

QSAR modeling of TBK1 inhibitors

Computer modeling of kinase inhibitors and computer-aided drug design

Computer modeling plays a significant role in the design of kinase inhibitors, which are crucial in drug discovery and development. [78, 79] In computer-aided drug design, predicting the IC50 values of potential drug candidates is crucial for assessing their potency in inhibiting the target enzyme or protein.

Quantitative structure-activity relationship (QSAR) models relate the chemical structure of compounds to their biological activity, including IC50 values. By analyzing a dataset of known inhibitors with experimental IC50 values, QSAR models can be used to predict the IC50 values of new compounds. Molecular descriptors such as molecular weight, lipophilicity, hydrogen

bonding capacity, and electronic properties are used to characterize the compounds and correlate them with their IC50 values. [80, 81]

In structure-based drug design molecular docking simulations predict the binding mode and affinity of small molecules to the target protein's active site. [82-84] Compounds with favorable docking scores are more likely to have lower IC50 values. Molecular dynamics simulations can further refine the binding poses and assess the stability of the protein-ligand complex, providing insights into the dynamic behavior that may influence IC50 values. In ligand-based drug design pharmacophore modeling identifies the essential structural features required for binding to the target protein. [85, 86] Compounds that match the pharmacophore features are likely to exhibit activity, including potency measured by IC50 values. Similarity searching compares the chemical features of potential drug candidates to known active compounds with known IC50 values, enabling the prediction of potency based on structural similarity.

Machine learning algorithms, such as support vector machines (SVM), random forest, or neural networks, can be trained on large datasets of compounds and their corresponding IC50 values to predict the potency of new compounds. Deep learning approaches, including convolutional neural networks (CNN) and recurrent neural networks (RNN), can capture complex relationships between chemical structures and biological activity, improving the accuracy of IC50 predictions. [87-90]

QSAR modeling of TBK1 inhibitors based on available IC50 experimental data.

Our main goals in this study were to develop the best possible predictive models for TBK1 inhibitors that could be used by research scientists in their quest to discover effective drugs against wide range of diseases such as cancer, viral infections, and inflammatory disorders. We also made a conscious effort to explain why the models work based on the top molecular descriptors that appear in the models which may provide an invaluable insight into the mechanism of the TBK1 inhibition. To accomplish this, we used CAS Content Collection to extract all available data associated with target protein TBK1 including IC50 and pIC50 values of inhibitors.. After removing all duplicate structures, records without structural information, salts, and mixtures we arrived at our final training set with 1,183 compounds, all single organic chemicals. Upon close examination of the data, the IC50 values of more than half of the structures have been entered as active (IC50 < 0.1uM or IC50 < 0.001uM) or as inactive (IC50 > 10uM) (Figure S1). The data distribution of pIC50 of the training set is outlined in Figure 4. These specifics of the data entry prompted us to pursue 3 distinctive types of predictive models. For the continuous distribution Figure 4. (a) regression models, for the binary distribution Figure 4. (b) – binary models, and for the 3-category distribution Figure 4. (c) – multiclassification models.

As already mentioned above, only a fraction of the data is presented with their exactly measured concentration. For the development of the regression models, we only used the 475 structures with the exact experimental measurement of IC50 and omit any data entered as active/inactive as they introduce a significant penalty for the accuracy of the regression models. For the classification models we used the entire set of 1,183 chemicals with 349/834 distribution of active/inactive and a breakpoint at pIC50=7 for the binary models and 349/446/388 of active/marginal/inactive and breakpoints at pIC50=7 and 8 for the 3-category models.

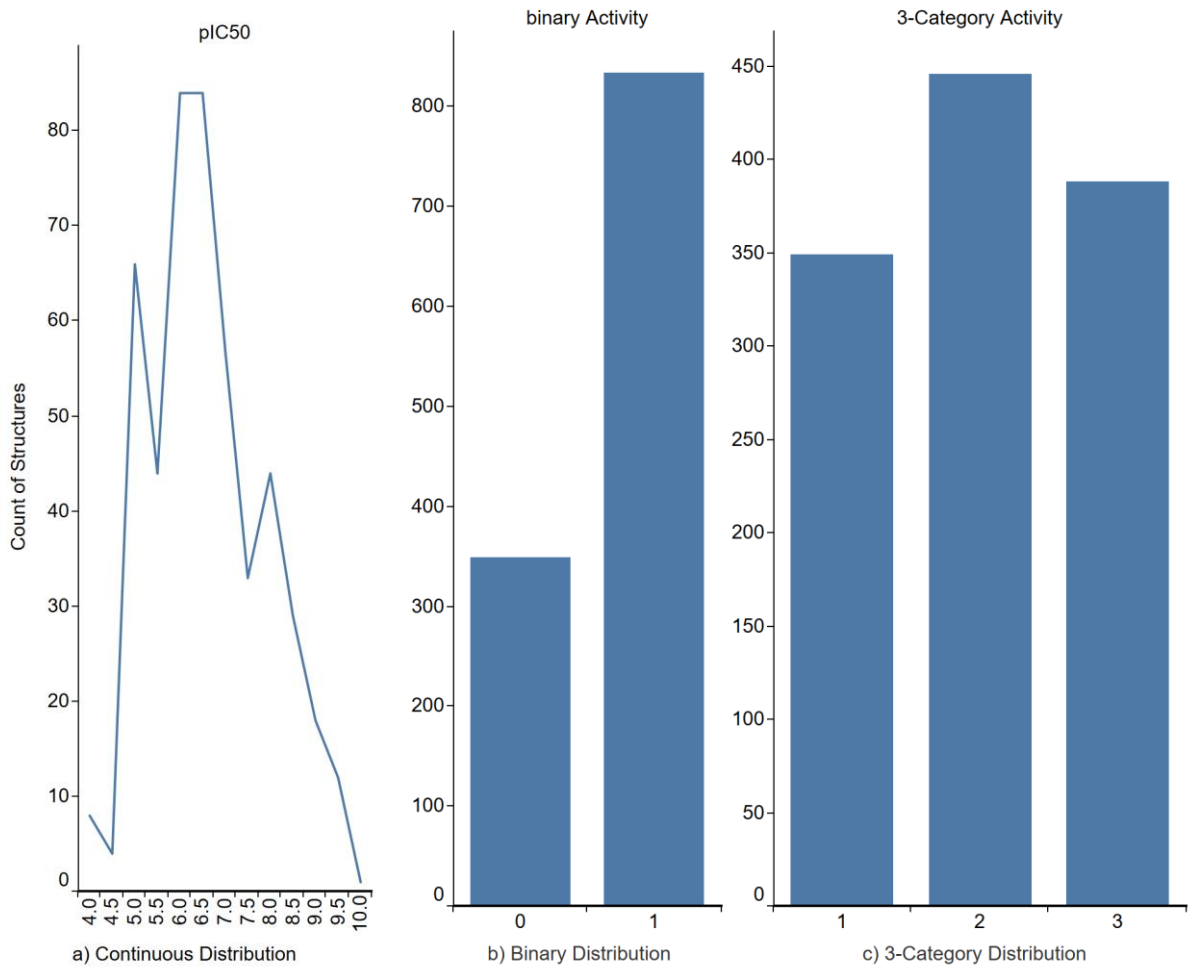


Figure 4. Distributions of pIC50 of the training set data for the three predictive QSAR models developed – (A) regression model, (B) binary model and (C) 3-category distribution.

We explored several different sets of molecular descriptors, as it is not known beforehand what features best correlate with the inhibition of TBK1. In this investigation we used our proprietary CAS fingerprint [91], the fragments generated by the Structure-Functional Analysis [92], and the available molecular descriptors in RdKit [93]. A brief summary of all molecular descriptors is given in Table 1.

Table 1. Molecular descriptors were used in the study.

Molecular descriptors	Short summary
CAS fingerprint [91]	CAS proprietary fingerprint consists of over 7k molecular features
CAS structure-functionality [92]	CAS proprietary structure-functionality analysis
Morgan fingerprint [94]	RdKit - The hashed Morgan fingerprint for a molecule (radius = 3; length =2048)
MACCS keys [95]	RdKit - 166 public MACCS keys
Atom Pairs [96]	RdKit - The atom-pair fingerprint for a molecule
Topological Torsion [97] fingerprints	RdKit - The hashed topological-torsion fingerprint for a molecule

Crippen LogP, and MR [98]	RdKit - The Wildman-Crippen logp, mr
MQNs [99]	RdKit - The Molecular Quantum Numbers
PEOE_VSA, SMR_VSA, SlogP_VSA [100]	RdKit - Atoms van der Waals surface area (VSA) descriptors
BCUT2D [101]	RdKit - Diagonal elements: atomic mass, Gasteiger charge, Crippen logP, Crippen MR
FractionCSP3 [93]	RdKit - The fraction of C atoms that are SP3 hybridized
Topological descriptors [93]	RdKit - Various topological descriptors

An automated machine learning platform, DataRobot (https://www.datarobot.com/), was used to train and evaluate performance of more than 70 different machine learning algorithms. DataRobot is also employed to build informative features selected from molecular descriptors. To counter the overfitting, a well-known problem in machine learning, and to estimate the statistical performance of the models, a common fivefold cross-validation procedure has been utilized for all models in this study.

The model building procedures is as follow: first the initial set of structures is split into two sub-sets with 80% of the structures utilized as model building set and the remainder 20% utilized as a holdout test set. The holdout test set is kept aside and is not used in the development of QSAR models. Instead, it is utilized as control set to assess the accuracy of the models as the chemicals in the holdout set are external with respect to the model building procedure. The model building set is subject to a fivefold cross-validation procedure for internal validations. In each of the five iterations of this approach, 80% of the model building set is used to build a model, and 20% is held as a test set. Across the entire process, then, every record is held out as validation in one part of the process, yet all records are made available to the model. Grid search was used as the default method for hyperparameter optimization.

As already mentioned, we pursued 3 types of models to predict the TBK1 inhibition: binary classifiers, three category classifiers, and regression models. Number of common matrices were used to evaluate the statistical performance of the predictors.

Binary classifiers:

Area Under the ROC Curve (AUC) – measures the ability to distinguish ones from zeros.

Sensitivity/Recall – measures the probability of a positive test result.

$$\text{Sensitivity} = TP / (TP + FN) \quad \text{(f1)}$$

Specificity – measures the probability of a negative test result.

$$\text{Specificity} = TP / (TN + FP) \quad \text{(f2)}$$

Precision - fraction of relevant instances among the retrieved instances.

$$\text{Precision} = TP / (TP + FP) \quad \text{(f3)}$$

Accuracy – fraction of the correctly classified samples.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad \text{(f4)}$$

F1_score – measures the predictive skill of a model by elaborating on its class-wise. It combines two competing metrics- precision and recall scores of a model.

F1_score = TP / (TP + 0.5(FP + FN)) (f5)

where: TP - true positives, TN - true negatives, FP – false positives, FN - false negatives.

Three category classifiers:

Logarithmic loss – measures the inaccuracy of the predicted probabilities.

$$LogLoss = -(y \log(p) + (1 - y) \log(1 - p)) \quad \text{(f6)}$$

where: y-Actual output, p-probability predicted by the logistic regression.

Accuracy – fraction of the correctly classified samples (f4) for each class.

Balanced Accuracy – average of Sensitivity (f1) per target class.

Area Under the ROC Curve (AUC) – measures the ability to distinguish ones from zeros.

Regression models:

R Squared - measures the proportion of total variation of outcomes explained by the model

$$R^2 = 1 - \frac{\sum_i^n (\hat{y}_i - y_i)^2}{\sum_i^n (y_i - \bar{y}_i)^2} \quad \text{(F7)}$$

Root Mean Square Error (RMSE) – measures the inaccuracy of the predicted mean values.

$$RMSE = \sqrt{\frac{\sum_i^n (\hat{y}_i - y_i)^2}{n}} \quad \text{(f8)}$$

Mean Absolute Error (MAE) - measures the inaccuracy of predicted median values.

$$MAE = \frac{\sum_i^n (\hat{y}_i - y_i)^2}{n} \quad \text{(f9)}$$

where: $\hat{y}_i$ is a predicted value, y_i is a real value, $\bar{y}_i$ is a mean value, over all samples, and n is the number of samples.

In addition to the common internal and external QSAR validations, the models were also evaluated by making predictions against a completely independent external validation set of 21 TBK1 inhibitors with experimental IC50 data, found in recent publications [26, 102] that are not part of the initial data set of 1,183 structures.

Results and discussions

To derive the best predictive models, we first computed all molecular descriptors. For CAS fingerprint [91] and Fragment-functional analysis [92] we employed our own proprietary software. For Fragment-functional analysis [92] over 66,000 single fragments were generated after splitting the chemical structures of the training set into sub-fragments. The RdKit [93] molecular descriptors were computed utilizing the python implementation of RdKit and the SMILES [103] strings of the chemicals in the training set. As scientists and regulatory agencies around the world may have diverse needs for how the predictions are presented as well as the particulars of the input data (as explained in section "Data distribution" above) we probed 3 distinct types of predictive models: Regression, Binary, and 3-Category models.

Considering both the performance of the common statistical matrices and the predictions of the holdout test set and the independent external set, the best overall predictive models were derived from the fragment-functional analysis descriptors [92]. Graphical representation of the statistical performance of these models is outlined in Figure 5.

To assess the predictive abilities of our models we used internal and external validation tests. [104-106] For the regression model we obtained $R^2_{holdout} = 0.822$ (Figure 5. a.) and $R^2_{cros\text{-}validation} = 0.796$ (Figure 5. b.) for the external and internal validations, respectively. The difference $R^2_{holdout} - R^2_{cros\text{-}validation}$ of 0.026 clearly indicates that there are low overfitting ramifications with the regression model. The same holds true for the classification models, where we consider $AUC_{holdout}$ and $AUC_{cros\text{-}validation}$. We achieved AUC differences of 0.0015 and 0.0212 for the binary (Figure 5. c. & d.) and 3-category (Figure 5. e. & f.) models respectively.

We also utilized Tropsha's statistical characteristics [104] to assess the external predictability of the regression model. We computed (as described in [104]) k, k', R^2_0, and R'^2_0, the slopes and correlation coefficients between predicted vs observed (and vice versa: observed vs predicted) activities of the structures in the holdout test set.

A QSAR model is acceptable if the following conditions are met [104]:

1. $R^2_{cross\text{-}validation} > 0.5$
2. $R^2_{test\ set} > 0.6$
3. $| (R^2_{test\ set} - R^2_0)/ R^2_{test\ set} | < 0.1$ or $| (R^2_{test\ set} - R'^2_0)/ R^2_{test\ set} | < 0.1$
4. $0.85 \leq k \leq 1.15$ or $0.85 \leq k' \leq 1.15$

For the regression model we have:

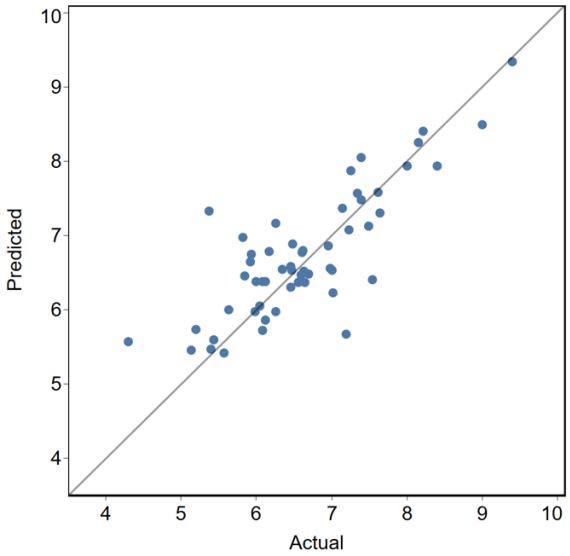
$R^2_{cross\text{-}validation} = 0.796$ (which is > 0.5)

$R^2_{holdout} = 0.822$ (which is > 0.6)

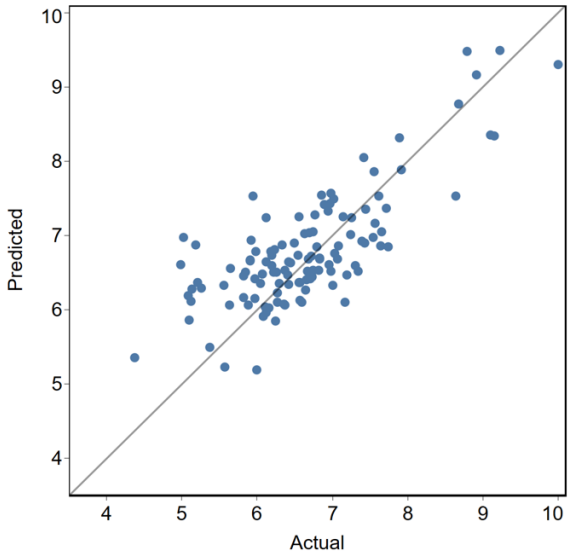
$R^2_0 = 0.8879$ and $R'^2_0 = 0.9518$ -> $| (0.822 - 0.8879)/0.822 | = 0.08$ (which is < 0.1)

$k = 1.0547$ and $k' = 0.9502$ - both are in the range of 0.85 – 1.15.

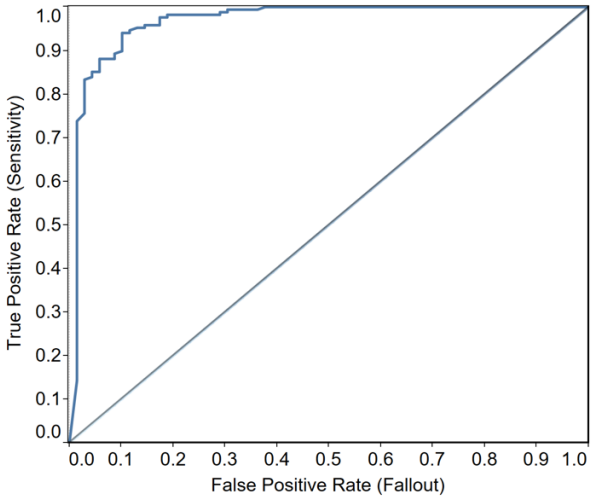
Thus, satisfying all the above conditions.



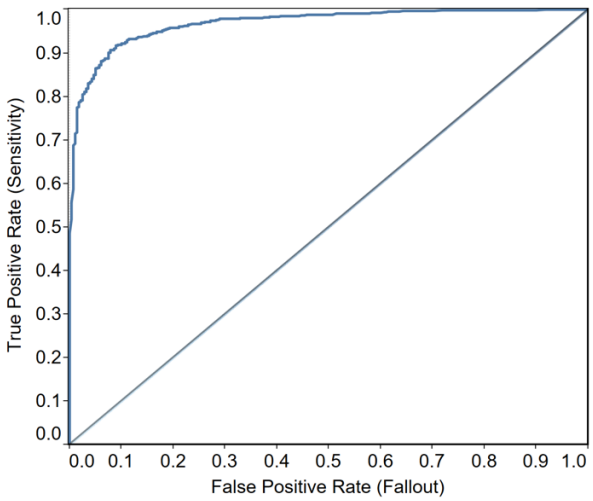
a) Holdout set - Regression Model: R^2=0.8220



b) Training set - Regression Model: R^2=0.7960



c) Fragments - Binary Model ROC-AUC (holdout set): 0.9672



d) Fragments - Binary Model ROC-AUC (training set): 0.9687

Confusion Matrix - holdout set

	Predicted Class		
Actual Class	1	2	3
1	56	7	0
2	10	75	8
3	4	11	66

AUC	= 0.9454
Accuracy	= 0.8312
Balanced Accuracy	= 0.8367
LogLoss	= 0.4282

e) Fragments - 3-Category Model

Confusion Matrix - training set

	Predicted Class		
Actual Class	1	2	3
1	251	31	4
2	34	308	11
3	9	24	274

AUC	= 0.9666
Accuracy	= 0.8806
Balanced Accuracy	= 0.8812
LogLoss	= 0.3323

f) Fragments - 3-Category Model

Figure 5. Statistical performance of the models derived by the Fragments-functional descriptors.

Y-scrambling validation is a widely used technique [107-109] to evaluate the robustness of QSAR models and to ensure that the developed models are not derived due to chance. In this test, the dependent variable (observed activity) randomly shuffles while keeping the independent variables (molecular descriptors) unchanged, and a new model is derived. This process is repeated several times and the values of $R^2_{test,}$ and $R^2_{cross\text{-}validation}$ are recorded. The values of R^2_{test} and $R^2_{cross\text{-}validation}$ are expected to be low, ensuring the developed QSAR is robust and not derived due to chance. For the current study we run 10 Y-scrambling tests, and the results are presented on Table 2. All values of R^2_{test} and $R^2_{cross\text{-}validation}$ are below 0.1 thus confirming that the regression model is not derived due to chance.

Table 2. Y-scrambling test results.

#	$R^2_{cross\text{-}validation}$	R^2_{test}
1	0.0205	0.00084
2	0.0119	0.00877
3	0.0072	0.00356
4	0.0187	0.00442
5	0.0013	0.00919
6	0.0101	0.0118
7	0.0200	0.0347
8	0.0200	0.00222
9	0.0227	0.00625
10	0.0225	0.00062

The results for the predictions of the external test chemicals are presented in Table 3. The results obtained from all molecular descriptor sets and models in this investigation are available in the Supplementary information material (Figures S4-S24, and Tables 2-4).

A quick glance at the results in Table 3 reveals that while the statistical criteria used to validate the models meet or exceed the common standards for good predictors, R^2 of 0.69 for the predictions of the independent external set with the regression model, although acceptable, is noticeable lower than the R^2 of the holdout test set (0.822). The lower value of R^2 for the external set can be attributed to the fact that these structures are external in the true sense of the word. While the chemicals in the holdout test set are external for the model building procedure, they are not entirely external for the model building data with respect to structural similarity. While most of the chemicals in the initial training set originate from different studies, however, a single study often has series of chemical compounds that may bear some structural similarities. In contrast, the TBK1 inhibitors in our independent external test set are not part of the initial set and any similarities that might exist can be attributed to chance. With much more chemical diversity for the classification models where the initial set is 1,183 chemicals strong, both the binary and 3-category predictors perform well in the validation of the holdout set as well as the predictions of the independent external set.

Table 3. Predictions for the independent external set of 21 compounds.

#	CAS RN	pIC50-observed	Binary-observed Breakpoint at pIC50=7	3-Category-observed*	pIC50-predicted	Binary-predicted	3-Category-predicted*
1	68301-99-5	4.26	0	1	5.36	0	1
2	2116443-03-7	4.62	0	1	4.67	0	1
3	2116445-80-6	5.54	0	1	6.22	0	1
4	2116445-76-0	4.00	0	1	5.67	0	1
5	2116445-77-1	4.44	0	1	5.34	0	1
6	2116445-78-2	4.14	0	1	5.33	0	1
7	unknown**	5.35	0	1	5.67	0	1
8	2116445-85-1	4.48	0	1	4.73	0	1
9	70529-18-9	6.40	0	1	5.89	0	1
10	2116445-81-7	5.54	0	1	5.54	0	1
11	2116445-82-8	4.00	0	1	5.80	0	1
12	1056634-68-4	7.24	1	2	7.16	1	3
13	2116443-62-8	6.70	0	1	5.75	0	1
14	2243281-75-4	8.05	1	3	7.80	1	2
15	2243281-77-6	7.70	1	2	7.80	1	2
16	2322365-47-7	8.30	1	3	8.07	0	3
17	81267-65-4	6.02	0	1	5.92	0	1
18	1835675-67-6	8.55	1	3	8.45	0	1
19	2101906-58-3	7.86	1	2	7.76†	1	2
20	1903773-70-5	7.70	1	2	6.61†	0	1
21	2020003-22-7	6.89	0	1	5.21	0	1
					R^2 = 0.69	Accuracy=0.86	Accuracy=0.81

*Category 1 (pIC50<7); Category 2 (pIC50<8); Category 3 (pIC50>=8).

**SMILES: CC(C)C1C=CC(=C2C=1)OC3N=C(N)C(C(=O)OCCN(C)C)=CC=3C2=O

†Warning: The prediction might be incorrect as the chemical lies otside of the applicability domain

One particularly important aspect of QSAR modeling is defining the applicability domain [105, 107-111] of the predictors. Or in other words to determine the chemical space where the models are suitable to make quality predictions and avoid a potential misuse of the results. Thus, the predictions for new molecules obtained from a QSAR model are acceptable only if the new molecules fall inside the applicability domain of that model. The applicability domain is the chemical space defined by all molecular descriptors used to build the QSAR model. For a fragment-based methodology the molecular descriptors are the structural alerts obtained by splitting the chemical structures of the training set into all possible sub-fragments. In the current

study our algorithm generated over 66K unique fragments from the compounds in the training set which define the domain of applicability for our TBK1 predictive models.

There are several ways to graphically illustrate the applicability domain of QSAR models. In the current study we utilized the Williams plot which is commonly used [107-111] and represents the applicability domain as a plot of Leverage (h) vs Standardized Residuals (σ).

Leverage of a given chemical structure h_i is defined as:

$$h_i = x_i^T (X^T X)^{-1} x_i \quad \text{(F10)}$$

Where:

$\boldsymbol{x_i}$ is the descriptor vector of the i^{-th} structure

$\boldsymbol{X}$ is the descriptor matrix of the training set used to build the model

The warning leverage h^* is defined as:

$$h^* = 3(p+1)/n \quad \text{(F11)}$$

Where:

$\boldsymbol{p}$ is the number of descriptors in the model

$\boldsymbol{n}$ is the number of chemicals in the training set

Test chemicals with $x_i < h^*$ are considered reliably predicted.

On Figure 6. is shown the Williams plot for the regression model, where the applicability domain is defined within ±3σ and a leverage threshold $h^* = 0.25$. From the plot is apparent that all compounds of the training set are within the applicability domain except 3 with leverage values greater than the warning h*. These 3 chemicals could influence the performance of the model, however, like [108] their standard residuals are well within the established limits and thus, not model outliers to be considered for removal from the training set. There are also 2 structures from the external test set with h > h* marked (with †) on Table 3 as warning predictions. There are also 3 external set compounds outside of the established limits of the standard residuals, which explains the lower R^2 of the independent external set.

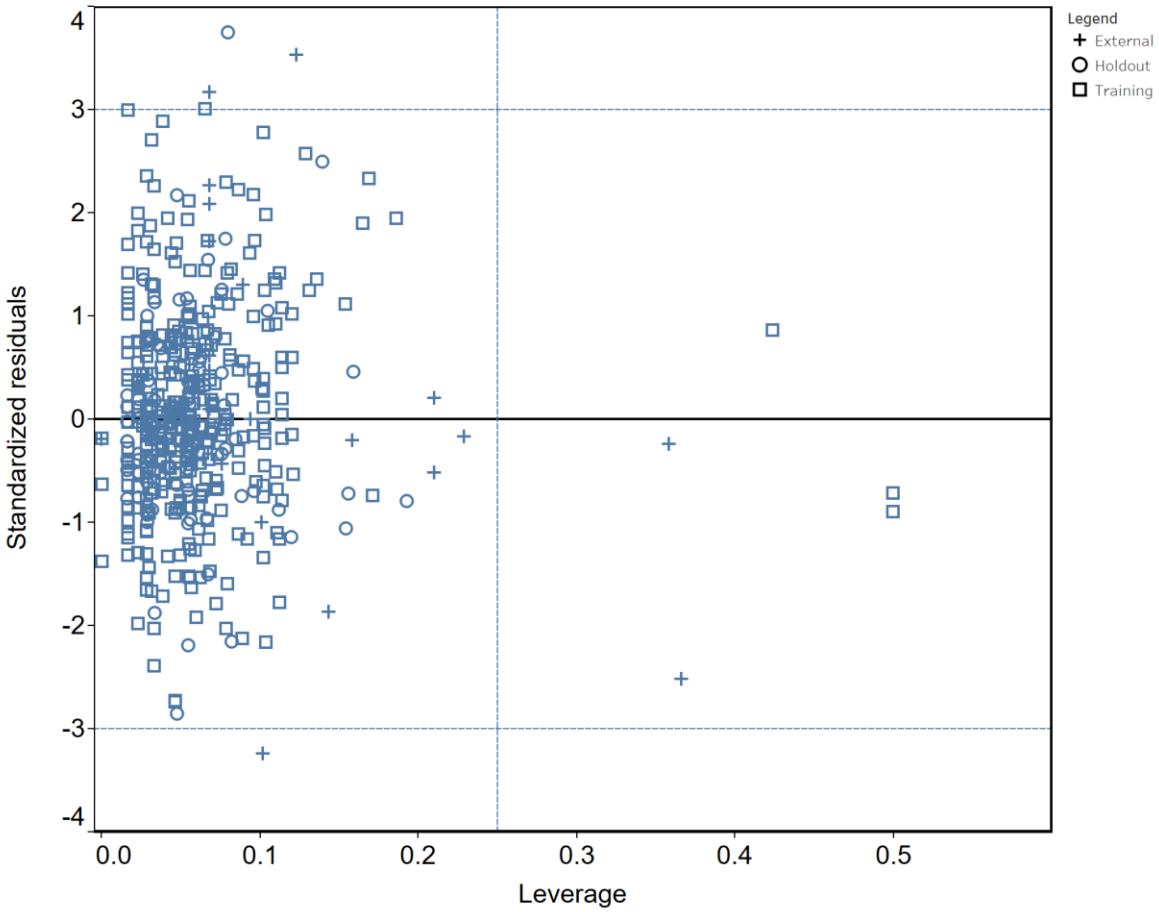
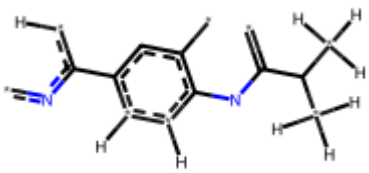
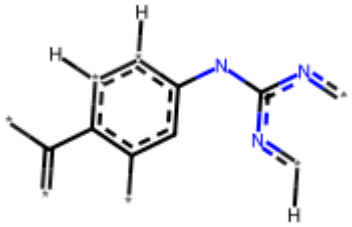
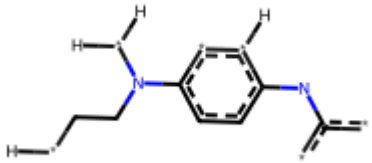
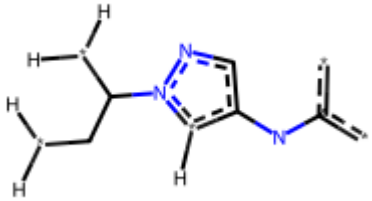
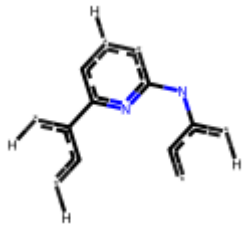
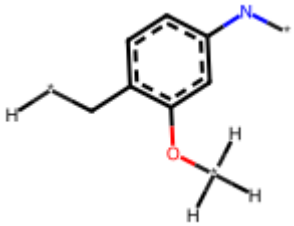
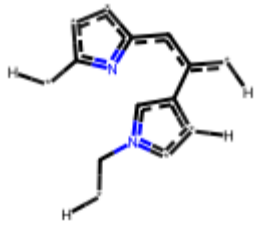
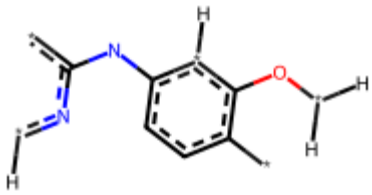
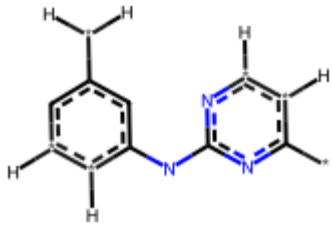
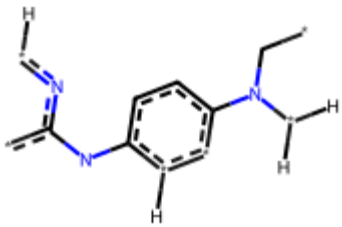


Figure 6. Williams plot of the applicability domain.

In our study we did not try to define upfront what structural alerts are statistically significant/important. Instead, we let the machine learning algorithms decide the importance of the molecular descriptors. In Table 4. are listed some of the most significant structural alerts. In fragment-based drug design, structural alerts serve as critical tools for identifying promising fragment scaffolds with potential to bind to a target protein. By recognizing specific molecular features associated with desired or undesired properties, these alerts enable early prediction of a fragment's suitability for progression into lead optimization. This knowledge helps focus drug discovery efforts on compounds with higher likelihood of success while minimizing the risk of developing molecules with liabilities such as toxicity or poor pharmacokinetics.

Table 4. Structural alerts for the Fragments binary model.

Structural alert	Average pIC50	# of Actives	# of Inactives
	8.0	10	0
	8.0	4	0
	8.0	6	0
	8.003	11	0
	8.068	215	1
	8.0	3	0

	8.667	3	0
	8.0	21	0
	8.654	16	0
	8.0	8	0

As has been described briefly in the introduction, TBK1 is composed of four domains – kinase domain, ubiquitin-like domain, scaffold/dimerization domain and TANK-binding domain. The structure of TBK1 has been determined with several structures from across different methodologies (X-ray, cryo-EM) and species and are available in the Protein Data Bank (PDB). Many of the reported structures of TBK1 are in the presence of an inhibitor such as the withdrawn drug amlexanox (CAS RN: 68302-57-8) [102] and a highly selective small molecule inhibitor in development, BAY-985 (CAS RN: 2409479-29-2) [112]. Both inhibitors appear to be competitive in nature, binding around the same area as ATP would.

A few key features emerge:

1. The binding site itself appears to consist of a mixture of charged residues and polar residues (Glu87, Arg25, Thr156, Cys89) as well as hydrophobic residues (Leu15, Val23, Ala33, Gly92).
2. Ability to H bond with Cys89 appears to be important.
3. Besides this, H bond interactions with other residues such as Glu87 and Thr156 also appear to be crucial for the inhibitory effect.

4. There also appears to be a size limit to the binding site wherein addition of bulk beyond a certain point is not tolerated and leads to a sharp decline in inhibitory activity.
5. Pushing the gatekeeper residue, Met86, resulted in increased potency.

Based on the available information, it appears that the ability to H bond may be crucial for inhibitory effect at TBK1. Consequently, structural features capable of H bonding (donors and acceptors) are likely important. Also, inhibitors having variations of the purine moiety are likely to be effective since inhibitors have to compete with ATP for the binding site. Structural elements capable of engaging in Van der Waals interactions with hydrophobic residues likely help boost inhibitory potency. Finally, structural elements that aid in pushing the gatekeeper residue Met 86, such as bulky substituents (5 or 6 membered rings) are likely to also increase inhibitory potency.

Conclusions

TBK1 is a serine/threonine kinase involved in various signaling pathways, particularly those regulating immune responses, which pinpoints it as an important player in innate immunity, particularly in the regulation of type I interferon responses. Understanding its key molecular fragments and their interactions with receptors can help in the development of inhibitors or modulators for therapeutic purposes. In the last two decades, there has been a steady increase in interest in TBK1 from the scientific community, especially evident after 2019. Despite this continued and sustained interest and TBK1's obvious involvement in a wide variety of diseases, targeting TBK1 effectively has been challenging. Notwithstanding the challenges, the pursuit of a TBK1 selective inhibitor is of interest to the scientific community and pharmaceutical entities.

In this study we developed a few machine learning models capable of predicting IC50 of small molecules inhibiting TBK1, a promising therapeutic target for the development of drugs modulating immune responses and treating wide range of diseases. We used the CAS Content Collection to assemble a training set of 1,183 chemical structures with experimentally measured IC50 toward TBK1. We explored several machine learning techniques combined with various molecular descriptors to derive and select the best TBK1 inhibitor QSAR models. Several structural alerts responsible for the mechanism of inhibition of TBK1 are also outlined and discussed. In the context of fragment-based drug design, such structural alerts can help identify promising fragments and predict potential liabilities, in order to guide lead optimization in drug discovery.

Notes

The authors declare no competing financial interest.

Acknowledgement

The authors sincerely appreciate Dharmini Patel for project coordination and are grateful to Manuel Guzman, Michael Dennis, Tim Wahlberg, Dawn Riedel, Dawn George, and Hong Xie for executive sponsorship. The authors also appreciate the rest of the Science Connect team at CAS for their support and insightful discussions.

Funding: None

Data and Software Availability

Publications of TBK1 inhibitors were identified by optimizing a search of relevant terms on the CAS Content Collection using CAS STN. While the full dataset is considered proprietary by CAS, the search string used for retrieval is included in the Supporting Information. TBK1 chemical inhibitors, the associated structural activity relationship data were extracted from these publications in CAS Content Collection.

Supporting Information

Data modeling description

References:

1. Kinase. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/kinase (accessed Feb 23, 2024).
2. Ardito, F., Giuliani, M., Perrone, D., Troiano, G., and Lo Muzio, L. (2017) The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). Int J Mol Med 40, 271-280.
3. Seok, S. H. (2021) Structural Insights into Protein Regulation by Phosphorylation and Substrate Recognition of Protein Kinases/Phosphatases. Life (Basel) 11.
4. Zhu, L., Li, Y., Xie, X., Zhou, X., Gu, M., Jie, Z., Ko, C.-J., Gao, T., Hernandez, B. E., Cheng, X., et al. (2019) TBKBP1 and TBK1 form a growth factor signalling axis mediating immunosuppression and tumourigenesis. Nature Cell Biology 21, 1604-1614.
5. Shaw, L. M. (2021) TBK1 has a new Akt. Journal of Biological Chemistry 297.
6. Paul, S. The Non-canonical Function and Regulation of TBK1 in the Cell Cycle. Doctor of Philosophy, Virginia Polytechnic Institute, 2023.
7. Yu, T., Yi, Y.-S., Yang, Y., Oh, J., Jeong, D., and Cho, J. Y. (2012) The pivotal role of TBK1 in inflammatory responses mediated by macrophages. Mediators of Inflammation 2012.
8. Yu, J., Zhou, X., Chang, M., Nakaya, M., Chang, J.-H., Xiao, Y., William Lindsey, J., Dorta-Estremera, S., Cao, W., and Zal, A. (2015) Regulation of T-cell activation and migration by the kinase TBK1 during neuroinflammation. Nature communications 6, 6074.
9. Matsumoto, G., Shimogori, T., Hattori, N., and Nukina, N. (2015) TBK1 controls autophagosomal engulfment of polyubiquitinated mitochondria through p62/SQSTM1 phosphorylation. Human molecular genetics 24, 4429-4442.

10. Pilli, M., Arko-Mensah, J., Ponpuak, M., Roberts, E., Master, S., Mandell, M. A., Dupont, N., Ornatowski, W., Jiang, S., and Bradfute, S. B. (2012) TBK-1 promotes autophagy-mediated antimicrobial defense by controlling autophagosome maturation. Immunity 37, 223-234.
11. Sarraf, S. A., Sideris, D. P., Giagtzoglou, N., Ni, L., Kankel, M. W., Sen, A., Bochicchio, L. E., Huang, C. H., Nussenzweig, S. C., Worley, S. H., et al. (2019) PINK1/Parkin Influences Cell Cycle by Sequestering TBK1 at Damaged Mitochondria, Inhibiting Mitosis. Cell Rep 29, 225-235.e225.
12. Thurston, T. L., Boyle, K. B., Allen, M., Ravenhill, B. J., Karpiyevich, M., Bloor, S., Kaul, A., Noad, J., Foeglein, A., Matthews, S. A., et al. (2016) Recruitment of TBK1 to cytosol-invading Salmonella induces WIPI2-dependent antibacterial autophagy. Embo j 35, 1779-1792.
13. Ruyuan, Z., Zhang, Q., and Xu, P. (2020) TBK1, a central kinase in innate immune sensing of nucleic acids and beyond. Acta biochimica et biophysica Sinica 52.
14. Herhaus, L. (2021) TBK1 (TANK-binding kinase 1)-mediated regulation of autophagy in health and disease. Matrix Biology 100-101, 84-98.
15. Ahmad, L., Zhang, S. Y., Casanova, J. L., and Sancho-Shimizu, V. (2016) Human TBK1: A Gatekeeper of Neuroinflammation. Trends in molecular medicine 22, 511-527.
16. Li, D., and Wu, M. (2021) Pattern recognition receptors in health and diseases. Signal Transduction and Targeted Therapy 6, 291.
17. Mogensen, T. H. (2009) Pathogen recognition and inflammatory signaling in innate immune defenses. Clin Microbiol Rev 22, 240-273, Table of Contents.
18. Suresh, R., and Mosser, D. M. (2013) Pattern recognition receptors in innate immunity, host defense, and immunopathology. Adv Physiol Educ 37, 284-291.
19. Tooley, A. S., Kazyken, D., Bodur, C., Gonzalez, I. E., and Fingar, D. C. (2021) The innate immune kinase TBK1 directly increases mTORC2 activity and downstream signaling to Akt. Journal of Biological Chemistry 297.
20. Ye, J., Dephoure, N., and Maniatis, T. (2022) Complex Intracellular Mechanisms of TBK1 Kinase Activation Revealed by a Specific Small Molecule Inhibitor. bioRxiv, 2022.2010.2011.511671.
21. Zhou, J., Rasmussen, N. L., Olsvik, H. L., Akimov, V., Hu, Z., Evjen, G., Kaeser-Pebernard, S., Sankar, D. S., Roubaty, C., Verlhac, P., et al. (2022) TBK1 phosphorylation activates LIR-dependent degradation of the inflammation repressor TNIP1. Journal of Cell Biology 222.
22. Hasan, M., and Yan, N. (2016) Therapeutic potential of targeting TBK1 in autoimmune diseases and interferonopathies. Pharmacological Research 111, 336-342.
23. Taft, J., Markson, M., Legarda, D., Patel, R., Chan, M., Malle, L., Richardson, A., Gruber, C., Martín-Fernández, M., Mancini, G. M. S., et al. (2021) Human TBK1 deficiency leads to autoinflammation driven by TNF-induced cell death. Cell 184, 4447-4463.e4420.
24. Papageorgiou, L., Mangana, E., Papakonstantinou, E., Diakou, I., Pierouli, K., Dragoumani, K., Bacopoulou, F., Chrousos, G. P., Exarchos, T. P., Vlamos, P., et al. (2022) An updated evolutionary and structural study of TBK1 reveals highly conserved motifs as potential pharmacological targets in neurodegenerative diseases. bioRxiv, 2022.2009.2016.508274.
25. Thomson, D. W., and Bergamini, G. (2021) Recent progress in small molecule TBK1 inhibitors: a patent review (2015- 2020). Expert Opin Ther Pat 31, 785-794.
26. Xiang, S., Song, S., Tang, H., Smaill, J. B., Wang, A., Xie, H., and Lu, X. (2021) TANK-binding kinase 1 (TBK1): An emerging therapeutic target for drug discovery. Drug Discovery Today 26, 2445-2455.
27. Revach, O.-Y., Liu, S., and Jenkins, R. W. (2020) Targeting TANK-binding kinase 1 (TBK1) in cancer. Expert Opinion on Therapeutic Targets 24, 1065-1078.
28. CAS Content Collection. https://www.cas.org/about/cas-content (accessed Mar 31, 2024).
29. Alam, M., Hasan, G. M., and Hassan, M. I. (2021) A review on the role of TANK-binding kinase 1 signaling in cancer. Int J Biol Macromol 183, 2364-2375.

30. Helgason, E., Phung, Q. T., and Dueber, E. C. (2013) Recent insights into the complexity of Tank-binding kinase 1 signaling networks: The emerging role of cellular localization in the activation and substrate specificity of TBK1. FEBS Letters 587, 1230-1237.
31. Louis, C., Burns, C., and Wicks, I. (2018) TANK-Binding Kinase 1-Dependent Responses in Health and Autoimmunity. Frontiers in Immunology 9.
32. Cruz, V. H., and Brekken, R. A. (2018) Assessment of TANK-binding kinase 1 as a therapeutic target in cancer. Journal of Cell Communication and Signaling 12, 83-90.
33. Zhao, W. (2013) Negative regulation of TBK1-mediated antiviral immunity. FEBS Lett 587, 542-548.
34. Runde, A. P., Mack, R., S, J. P., and Zhang, J. (2022) The role of TBK1 in cancer pathogenesis and anticancer immunity. J Exp Clin Cancer Res 41, 135.
35. Chen, W., Luo, K., Ke, Z., Kuai, B., He, S., Jiang, W., Huang, W., and Cai, Z. (2017) TBK1 Promote Bladder Cancer Cell Proliferation and Migration via Akt Signaling. J Cancer 8, 1892-1899.
36. Delhase, M., Kim, S.-Y., Lee, H., Naiki-Ito, A., Chen, Y., Ahn, E.-R., Murata, K., Kim, S.-J., Lautsch, N., Kobayashi, K. S., et al. (2012) TANK-binding kinase 1 (TBK1) controls cell survival through PAI-2/serpinB2 and transglutaminase 2. Proceedings of the National Academy of Sciences 109, E177-E186.
37. Zhao, P., Wong, K. I., Sun, X., Reilly, S. M., Uhm, M., Liao, Z., Skorobogatko, Y., and Saltiel, A. R. (2018) TBK1 at the Crossroads of Inflammation and Energy Homeostasis in Adipose Tissue. Cell 172, 731-743.e712.
38. Zhao, P., and Saltiel, A. R. (2020) Interaction of Adipocyte Metabolic and Immune Functions Through TBK1. Front Immunol 11, 592949.
39. Huh, J. Y., and Saltiel, A. R. (2021) Roles of IκB kinases and TANK-binding kinase 1 in hepatic lipid metabolism and nonalcoholic fatty liver disease. Experimental & Molecular Medicine 53, 1697-1705.
40. Pied, N., Daussy, C. F., Denis, Z., Ragues, J., Faure, M., Iggo, R., Tschan, M. P., Roger, B., Rayne, F., and Wodrich, H. (2022) TBK1 is part of a galectin 8 dependent membrane damage recognition complex and drives autophagy upon Adenovirus endosomal escape. PLoS Pathog 18, e1010736.
41. Runde, A. P., Mack, R., S.J, P. B., and Zhang, J. (2022) The role of TBK1 in cancer pathogenesis and anticancer immunity. Journal of Experimental & Clinical Cancer Research 41, 135.
42. Oakes, J. A., Davies, M. C., and Collins, M. O. (2017) TBK1: a new player in ALS linking autophagy and neuroinflammation. Molecular Brain 10, 5.
43. Shu, C., Sankaran, B., Chaton, Catherine T., Herr, Andrew B., Mishra, A., Peng, J., and Li, P. (2013) Structural Insights into the Functions of TBK1 in Innate Antimicrobial Immunity. Structure 21, 1137-1148.
44. Ryzhakov, G., and Randow, F. (2007) SINTBAD, a novel component of innate antiviral immunity, shares a TBK1-binding domain with NAP1 and TANK. Embo j 26, 3180-3190.
45. Israël, A. (2010) The IKK complex, a central regulator of NF-kappaB activation. Cold Spring Harb Perspect Biol 2, a000158.
46. Goncalves, A., Bürckstümmer, T., Dixit, E., Scheicher, R., Górna, M. W., Karayel, E., Sugar, C., Stukalov, A., Berg, T., Kralovics, R., et al. (2011) Functional dissection of the TBK1 molecular network. PloS one 6, e23971.
47. TBK1 TANK binding kinase 1 [Homo sapiens (human)]. https://www.ncbi.nlm.nih.gov/gene/29110 (accessed Apr 6, 2024).
48. Zhang, C., Shang, G., Gui, X., Zhang, X., Bai, X.-c., and Chen, Z. J. (2019) Structural basis of STING binding with and phosphorylation by TBK1. Nature 567, 394-398.
49. Panne, D., and Larabi, A., Crystal structure and mechanism of activation of TBK1. Jan 24, 2013 ed.; Protein Data Bank, 2013.
50. Hasan, M., and Yan, N. (2016) Therapeutic potential of targeting TBK1 in autoimmune diseases and interferonopathies. Pharmacol Res 111, 336-342.

51. Hasan, M., Gonugunta, V. K., Dobbs, N., Ali, A., Palchik, G., Calvaruso, M. A., DeBerardinis, R. J., and Yan, N. (2017) Chronic innate immune activation of TBK1 suppresses mTORC1 activity and dysregulates cellular metabolism. Proceedings of the National Academy of Sciences 114, 746-751.
52. Oakes, J. A., Davies, M. C., and Collins, M. O. (2017) TBK1: a new player in ALS linking autophagy and neuroinflammation. Mol Brain 10, 5.
53. Wang, D., Huo, H., Werid, G. M., Ibrahim, Y. M., Tang, L., Wang, Y., and Chen, H. (2022) TBK1 Mediates Innate Antiviral Immune Response against Duck Enteritis Virus. Viruses 14.
54. Miyahira, A. K., Shahangian, A., Hwang, S., Sun, R., and Cheng, G. (2009) TANK-binding kinase-1 plays an important role during in vitro and in vivo type I IFN responses to DNA virus infections. Journal of immunology (Baltimore, Md. : 1950) 182, 2248-2257.
55. Hagan, R. S., Torres-Castillo, J., and Doerschuk, C. M. (2019) Myeloid TBK1 Signaling Contributes to the Immune Response to Influenza. American Journal of Respiratory Cell and Molecular Biology 60, 335-345.
56. Jiang, Q., Guan, Y., Zheng, J., and Lu, H. (2023) TBK1 promotes thyroid cancer progress by activating the PI3K/Akt/mTOR signaling pathway. Immun Inflamm Dis 11, e796.
57. Gao, C.-Q., Chu, Z.-Z., Zhang, D., Xiao, Y., Zhou, X.-Y., Wu, J.-R., Yuan, H., Jiang, Y.-C., Chen, D., Zhang, J.-C., et al. (2023) Serine/threonine kinase TBK1 promotes cholangiocarcinoma progression via direct regulation of β-catenin. Oncogene 42, 1492-1507.
58. Xiao, Q.-A., He, Q., Li, L., Song, Y., Chen, Y.-r., Zeng, J., and Xia, X. (2022) Role of IKKε in the Metabolic Diseases: Physiology, Pathophysiology, and Pharmacology. Frontiers in Pharmacology 13.
59. Ding, C., Song, Z., Shen, A., Chen, T., and Zhang, A. (2020) Small molecules targeting the innate immune cGAS–STING–TBK1 signaling pathway. Acta Pharm Sin B 10, 2272-2298.
60. Revach, O. Y., Liu, S., and Jenkins, R. W. (2020) Targeting TANK-binding kinase 1 (TBK1) in cancer. Expert Opin Ther Targets 24, 1065-1078.
61. Zhang, M., Zou, Y., Zhou, X., and Zhou, J. (2022) Inhibitory targeting cGAS-STING-TBK1 axis: Emerging strategies for autoimmune diseases therapy. Frontiers in Immunology 13.
62. Tong, Z., Fleming, J. L., Grozdic, I., McElroy, J., Beyer, S., Fabian, D., Becker, A., Bell, E. H., Mahler, K. L., Popp, I., et al. (2022) Therapeutic Potential of Small Molecule Inhibitors of TBK1 in Glioma. International Journal of Radiation Oncology*Biology*Physics 114, e80.
63. Schlicher, L., Green, L. G., Romagnani, A., and Renner, F. (2023) Small molecule inhibitors for cancer immunotherapy and associated biomarkers – the current status. Frontiers in Immunology 14.
64. Zarrin, A. A., Bao, K., Lupardus, P., and Vucic, D. (2021) Kinase inhibition in autoimmunity and inflammation. Nature Reviews Drug Discovery 20, 39-63.
65. OraDisc (Amlexanox) Mucoadhesive Patch. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2004/021727s000_OraDiscTOC.cfm#:~:text=Approval%20Date%3A%2009%2F29%2F2004 (accessed Feb 23, 2024).
66. Abouzid, A., Moustafa, A. Y., Allcock, N., Najlah, M., Elhissi, A., Stanley, C. W., Ahmed, W., Seville, P., Crean, S., Forbes, R. T., et al. (2023) Amlexanox-loaded nanoliposomes showing enhanced anti-inflammatory activity in cultured macrophages: A potential formulation for treatment of oral aphthous stomatitis. Journal of Drug Delivery Science and Technology 79, 104052.
67. Arvinas. https://www.arvinas.com/ (accessed Apr 4, 2024).
68. Domainex. https://www.domainex.co.uk/Integrated-drug-discovery-services-USA?gad_source=1&gclid=Cj0KCQjwn7mwBhCiARIsAGoxjaLuIH7OE9VflO21qX31cJQHZRUSp1wfZrv3Z9rWQcxJPYlK_Fdgs98aAkiWEALw_wcB (accessed Apr 4, 2024).
69. Asuragen. https://asuragen.com/ (accessed Apr 4, 2024).
70. Chipscreen Biosciences. https://www.chipscreen.com/en/ (accessed Apr 4, 2024).
71. KARRA, S. R., XIAO, Y., SHERER, B. A., and CHEKLER, E. PYRIMIDINE TBK/IKKε INHIBITOR COMPOUNDS AND USES THEREOF. WO2019079375, 2019.

72. Perrior, T. R., Newton, G. K., Stewart, M. R., and Aqil, R. Pyrimidine compounds as inhibitors of protein kinases IKK epsilon and/or TBK-1, processes for their preparation, and pharmaceutical compositions containing them. US8962609B2, 2015.
73. Lefranc, J., Mengel, A., Schulze, V., Christ, C., Prinz, F., Wengner, A. M., Stoeckigt, D., Boemer, U., and Bader, B. Substituted heteroarylbenzimidazole compounds as TBK1 kinase inhibitors and their preparation. WO2017207534, 2017.
74. Du, Z., Dornan, D., Guerrero, J. A., Kaplan, J. A., Knox, J. E., Naduthambi, D., Phillips, B. W., Stinson, S. Y., Venkataramani, C., Wang, P., et al. Preparation of TANK-binding kinase inhibitor compounds. WO2017106556, 2017.
75. Crew, A. P., Wang, J., Dong, H., Qian, Y., Siu, K., Ferraro, C., and Crews, C. M. TANK-BINDING KINASE-1 PROTACS AND ASSOCIATED METHODS OF USE. WO2016197114, 2016.
76. Aykul, S., and Martinez-Hackert, E. (2016) Determination of half-maximal inhibitory concentration using biosensor-based protein interaction analysis. Anal Biochem 508, 97-103.
77. Swinney, D. C., Chapter 18 - Molecular Mechanism of Action (MMoA) in Drug Discovery. In *Annual Reports in Medicinal Chemistry*, Macor, J. E., Ed. Academic Press: 2011; Vol. 46, pp 301-317.
78. Sliwoski, G., Kothiwale, S., Meiler, J., and Lowe, E. W., Jr. (2014) Computational methods in drug discovery. Pharmacol Rev 66, 334-395.
79. Osakwe, O., Chapter 5 - The Significance of Discovery Screening and Structure Optimization Studies. In *Social Aspects of Drug Discovery, Development and Commercialization*, Osakwe, O.; Rizvi, S. A. A., Eds. Academic Press: Boston, 2016; pp 109-128.
80. Bastikar, V., Bastikar, A., and Gupta, P., *Quantitative structure–activity relationship-based computational approaches*. Computational Approaches for Novel Therapeutic and Diagnostic Designing to Mitigate SARS-CoV-2 Infection. 2022:191-205. doi: 10.1016/B978-0-323-91172-6.00001-7. Epub 2022 Jul 15.
81. Kwon, S., Bae, H., Jo, J., and Yoon, S. (2019) Comprehensive ensemble in QSAR prediction for drug discovery. BMC Bioinformatics 20, 521.
82. Ferreira, L. G., Dos Santos, R. N., Oliva, G., and Andricopulo, A. D. (2015) Molecular docking and structure-based drug design strategies. Molecules 20, 13384-13421.
83. Meng, X. Y., Zhang, H. X., Mezei, M., and Cui, M. (2011) Molecular docking: a powerful approach for structure-based drug discovery. Curr Comput Aided Drug Des 7, 146-157.
84. Agu, P. C., Afiukwa, C. A., Orji, O. U., Ezeh, E. M., Ofoke, I. H., Ogbu, C. O., Ugwuja, E. I., and Aja, P. M. (2023) Molecular docking as a tool for the discovery of molecular targets of nutraceuticals in diseases management. Scientific reports 13, 13398.
85. Aparoy, P., Reddy, K. K., and Reddanna, P. (2012) Structure and ligand based drug design strategies in the development of novel 5- LOX inhibitors. Curr Med Chem 19, 3763-3778.
86. Acharya, C., Coop, A., Polli, J. E., and Mackerell, A. D., Jr. (2011) Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. Curr Comput Aided Drug Des 7, 10-22.
87. Nasser, M., Yusof, U. K., and Salim, N. (2023) Deep Learning Based Methods for Molecular Similarity Searching: A Systematic Review. Processes 11, 1340.
88. Dara, S., Dhamercherla, S., Jadav, S. S., Babu, C. M., and Ahsan, M. J. (2022) Machine Learning in Drug Discovery: A Review. Artif Intell Rev 55, 1947-1999.
89. Patel, L., Shukla, T., Huang, X., Ussery, D. W., and Wang, S. (2020) Machine Learning Methods in Drug Discovery. Molecules 25.
90. Staszak, M., Staszak, K., Wieszczycka, K., Bajek, A., Roszkowski, K., and Tylkowski, B. (2022) Machine learning in drug design: Use of artificial intelligence to explore the chemical structure–biological activity relationship. WIREs Computational Molecular Science 12, e1568.

91. Wills, T. J., Polshakov, D. A., Robinson, M. C., and Lee, A. A. (2020) Impact of Chemist-In-The-Loop Molecular Representations on Machine Learning Outcomes. Journal of Chemical Information and Modeling 60, 4449-4456.
92. Ivanov, J., Polshakov, D., Kato-Weinstein, J., Zhou, Q., Li, Y., Granet, R., Garner, L., Deng, Y., Liu, C., Albaiu, D., et al. (2020) Quantitative Structure–Activity Relationship Machine Learning Models and their Applications for Identifying Viral 3CLpro- and RdRp-Targeting Compounds as Potential Therapeutics for COVID-19 and Related Viral Infections. ACS Omega 5, 27344-27358.
93. RDKit: Open-Source Cheminformatics Software. http://www.rdkit.org/ (accessed Apr 4, 2024).
94. Rogers, D., and Hahn, M. (2010) Extended-Connectivity Fingerprints. Journal of Chemical Information and Modeling 50, 742-754.
95. Durant, J. L., Leland, B. A., Henry, D. R., and Nourse, J. G. (2002) Reoptimization of MDL Keys for Use in Drug Discovery. Journal of Chemical Information and Computer Sciences 42, 1273-1280.
96. Carhart, R. E., Smith, D. H., and Venkataraghavan, R. (1985) Atom pairs as molecular features in structure-activity studies: definition and applications. J. Chem. Inf. Comput. Sci. 25, 64-73.
97. Nilakantan, R., Bauman, N., Dixon, J. S., and Venkataraghavan, R. (1987) Topological torsion a new molecular descriptor for sar applications comparison with other descriptors. Journal of Chemical Information and Computer Sciences 27, 82-85.
98. Wildman, S. A., and Crippen, G. M. (1999) Prediction of Physicochemical Parameters by Atomic Contributions. J. Chem. Inf. Comput. Sci. 39, 868-873.
99. Nguyen, K. T., Blum, L. C., van Deursen, R., and Reymond, J.-L. (2009) Classification of Organic Molecules by Molecular Quantum Numbers. ChemMedChem 4, 1803-1805.
100. Labute, P. (2000) A widely applicable set of descriptors. Journal of molecular graphics & modelling 18, 464-477.
101. Bayada, D. M., Hamersma, H., and van Geerestein, V. J. (1999) Molecular Diversity and Representativity in Chemical Databases. Journal of Chemical Information and Computer Sciences 39, 1-10.
102. Beyett, T. S., Gan, X., Reilly, S. M., Chang, L., Gomez, A. V., Saltiel, A. R., Showalter, H. D., and Tesmer, J. J. G. (2018) Carboxylic Acid Derivatives of Amlexanox Display Enhanced Potency toward TBK1 and IKKε and Reveal Mechanisms for Selective Inhibition. Molecular pharmacology 94, 1210-1219.
103. Weininger, D. (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. Journal of Chemical Information and Computer Sciences 28, 31-36.
104. Golbraikh, A., and Tropsha, A. (2002) Beware of q2! Journal of Molecular Graphics and Modelling 20, 269-276.
105. Tropsha, A. (2010) Best Practices for QSAR Model Development, Validation, and Exploitation. Molecular Informatics 29, 476-488.
106. Gramatica, P., Chirico, N., Papa, E., Cassani, S., and Kovarich, S. (2013) QSARINS: A new software for the development, analysis, and validation of QSAR MLR models. Journal of Computational Chemistry 34, 2121-2132.
107. Tropsha, A., Gramatica, P., and Gombar, V. K. (2003) The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. QSAR & Combinatorial Science 22, 69-77.
108. Garg, R., and Smith, C. J. (2014) Predicting the bioconcentration factor of highly hydrophobic organic chemicals. Food and Chemical Toxicology 69, 252-259.
109. Pourbasheer, E., Vahdani, S., Malekzadeh, D., Aalizadeh, R., and Ebadi, A. (2017) QSAR Study of 17β-HSD3 Inhibitors by Genetic Algorithm-Support Vector Machine as a Target Receptor for the Treatment of Prostate Cancer. Iran J Pharm Res 16, 966-980.

110. Netzeva, T. I., Worth, A., Aldenberg, T., Benigni, R., Cronin, M. T., Gramatica, P., Jaworska, J. S., Kahn, S., Klopman, G., Marchant, C. A., et al. (2005) Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. The report and recommendations of ECVAM Workshop 52. Altern Lab Anim 33, 155-173.
111. Eriksson, L., Jaworska, J., Worth, A. P., Cronin, M. T., McDowell, R. M., and Gramatica, P. (2003) Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. Environ Health Perspect 111, 1361-1375.
112. Lefranc, J., Schulze, V. K., Hillig, R. C., Briem, H., Prinz, F., Mengel, A., Heinrich, T., Balint, J., Rengachari, S., Irlbacher, H., et al. (2020) Discovery of BAY-985, a Highly Selective TBK1/IKKε Inhibitor. Journal of Medicinal Chemistry 63, 601-612.

TOC Graphic