

The discovery of potential tumor necrosis factor alpha converting enzyme inhibitors via implementation of K Nearest Neighbor QSAR analysis

Maha Habash¹, Sami Alshakhshir¹, Shady Awwad², Mahmoud Abu-Samak³

¹ Michael Sayegh Faculty of Pharmacy, Aqaba University of Technology, Aqaba, Jordan

² Department of Pharmaceutical Chemistry & Pharmacognosy, Applied Science Private University, Amman, Jordan

³ Department of Clinical Pharmacy and Therapeutics, Applied Science Private University, Amman, Jordan

Corresponding author: Maha Habash (mhabash@aut.edu.jo)

Received 17 October 2022 ♦ Accepted 8 February 2023 ♦ Published 12 April 2023

Citation: Habash M, Alshakhshir S, Awwad S, Abu-Samak M (2023) The discovery of potential tumor necrosis factor alpha converting enzyme inhibitors via implementation of K Nearest Neighbor QSAR analysis. *Pharmacia* 70(2): 247–261. <https://doi.org/10.3897/pharmacia.70.e96423>

Abstract

Tumor necrosis factor-alpha converting enzyme (TACE) is considered as a pro-inflammatory cytokine which catalyzes the formation of TNF- α from membrane bound TNF- α precursor protein. It has important role in various pathological diseases such as inflammation, anorexia, rheumatoid arthritis, cancer and viral replication. Despite the reporting of a variety of TACE inhibitors of diverse chemical scaffolds whether peptide, peptide-like compounds or others, the bioavailability and pharmacokinetic problems are reflected strongly on its clinical effectiveness. In this effort we employed a novel combination of k-nearest neighbor QSAR modeling to select the best critical ligand-TACE contacts capable of elucidation of TACE inhibitory bioactivity among a long list of inhibitors. The recurrence of one valid pharmacophore hypothesis among the successful models emerged the pharmacophore to be used as 3D search queries to screen the National Cancer Institute's structural database for new TACE inhibitors. Eight potent TACE inhibitors were discovered with inhibition percentage exceeding 50% at 100 μ M inhibitor concentration.

Keywords

Tumor necrosis alpha converting enzyme (TACE), k nearest neighbor, Ligandfit, Pharmacophore, QSAR, In vitro assay

Introduction

Tumor Necrosis Factor (TNF)- α is a pro-inflammatory cytokine produced by different types of cells, i.e., monocytes, macrophages, neutrophils, T-cells, mast cells, epithelial cells, osteoblasts and dendritic cells. Several pathological conditions such as Crohn's disease, ulcerative colitis, diabetes, multiple sclerosis, atherosclerosis, psoriasis and stroke are caused by over-expression of TNF- α .

Furthermore, evidence exists that TNF plays a critical role in the origin and development of Rheumatoid Arthritis (RA) and other immune mediated disorders (Cherney et al. 2006; DasGupta et al. 2009; Yu Guo et al. 2010).

Coronaviruses (CoVs) included in the species of severe acute respiratory syndrome-related coronavirus (SARS38 related CoV) which are considered as highly pathogenic viruses with high rate of genetic mutations and/or recombination that render these viruses are particularly dangerous

and able to jump the species barriers from 40 animal host to humans and also to spread efficiently among humans causing severe respiratory complications, sometimes fatal (Siddell et al. 2019). In those patients, as a result of severe inflammatory-induced lung injury attributable to an uncontrolled overproduction and release of soluble markers of inflammation, as cytokines and several other mediators, referred as cytokine storm syndrome (CSS), sustaining an aberrant systemic inflammatory response, triggered by abnormal immunologic response to SARS-CoV-2 infection (Batista 2020; Gorbalenya et al. 2020; Guan et al. 2020).

Additionally, obesity is well-recognized as a major risk factor of insulin resistance and the number of obese people is increasing universally. The expression of tumor necrosis factor (TNF)- α , is increased in obese humans and animals (Hotamisligil et al. 1994).

Currently, five TNF- α inhibitors which are infliximab, etanercept, certolizumab pegol, adalimumab and golimumab are approved by Food and Drug Administration (FDA) in the USA and in other countries for the treatment of various diseases such as psoriatic arthritis, ankylosing spondylitis, and rheumatoid arthritis (RA) (Maekawa et al. 2019). The mentioned drugs bind selectively to TNF- α and inhibit its interaction with the TNF receptor. An approach for the regulation of TNF- α levels by means of the inhibition of TNF- α converting enzyme (TACE). TACE is a membrane bound zinc metalloprotease which converts the 26-kDa transmembrane bound pro-TNF- α to the mature 17-kDa soluble form of TNF (Shepeck et al. 2007a, b; Yu Guo et al. 2010). TNF- α biological response is mediated through two dissimilar receptors; namely TNF-R1 and TNF-R2. Both receptors are transmembrane glycoproteins similar in their extracellular domains sharing structural and functional homology, while their intracellular domains are distinct. TNF-R1 is constitutively expressed in most tissues while TNF-R2 is characteristically found in immune system. Experimental validation proved that under physiological conditions, TNF-R1 signaling seems to be mainly responsible for pro-inflammatory properties of TNF- α , while TNF-R2 has been hypothesized that it acts as ligand passer, at least in some cells. (Park et al. 2006; Shepeck et al. 2007c; Maekawa et al. 2019; Batista 2020).

Furthermore, TACE has been associated with hypertension by the elevation of Angiotensin II levels (De Queiroz et al. 2020). Inflammatory cytokines are accompanied with increased risk of cancer disease, so the inhibition of inflammatory mediators will inhibit immune related cancer cases (Lebrec et al. 2015).

Several drugs have been generated based on the inhibition of TACE activity or TNF- α

So our interest is the discovery of new effective small molecule leads that could modulate TNF- α levels is of high interest. (Park et al. 2006; Govinda Rao et al. 2007; Bandarage et al. 2008; Zhang et al. 2009).

The majority of known TACE inhibitors include a hydroxamic acid as the zinc chelating group. Since hy-

droxamic acids are often poorly absorbed and are susceptible to metabolic degradation and glucuronidation, the interest is discovering alternative groups to the hydroxamic acid (Levin et al. 2006; Lombart et al. 2007; Duan et al. 2008; Mazzola et al. 2008; Udechukwu et al. 2017).

Alternatively, non-hydroxamic acid TACE inhibitors, including tartrates and hydantoins. Tartaric acid and carboxylic acid based TACE inhibitors have been reported. (Shepeck et al. 2007a, b; Yu W et al. 2010; Dai et al. 2011).

As the catalytic site of TACE and matrix metalloproteinase (MMP) has high amino acid sequence similarity, the early MMP inhibitors, such as marimastat, prinomastat, and CGS27023A showed TACE inhibitory activity. TACE inhibitors or dual TACE and MMP inhibitors without MMP-1 inhibition are desirable (Janser et al. 2006; Adrian et al. 2007).

On our ongoing discovery program of potent TACE inhibitors, more exploration and understanding of the relationship between the structure and biological activity based on quantitative structure-activity relationship (QSAR) is useful tool for design of more potent inhibitors (Janser et al. 2006; Levin et al. 2006; Adrian et al. 2007; Duan et al. 2008).

Materials and methods

Molecular modeling

Molecular modeling Software and hardware

The molecular modeling software applied in the current research are:

- CS ChemDraw Ultra (Version 12.0), Cambridge Soft Corp., USA.
- Discovery Studio 2016, Accelrys Inc., San Diego, including pharmacophore generation, Ligandfit and Libdock implemented in Discovery Studio 2016
- MATLAB (Version R2007a), The MathWorks Inc., USA, Euclidean distance calculations
- Catalyst Search implemented in Discovery Studio 2016, Accelrys Inc., USA.

Data set

Structures of 115 TACE inhibitors with similar *in vitro* bioactivities (Suppl. material 1: table A) were collected from different articles (Duan et al. 2007; Duan et al. 2008; Lu et al. 2008; Ott et al. 2008a, b, c). The bioactivities were expressed as (IC₅₀), which represents the concentration of the test compound that inhibited the activity of TACE by 50%. For pharmacophore modeling and QSAR analysis purposes, the logarithm of measured IC₅₀ (nM) values were used to ensure correlating the data linear to the free energy change.

The chemical structures of TACE inhibitors (1–115, Suppl. material 1: table A) were drawn as 2D format in ChemDraw Ultra (Version 12.0). Rule based methods implemented in Discovery Studio were used for the conver-

sion into 3D structures followed by energy minimization to the closest local minimum energy by means of the molecular mechanics CHARMM force field. Subsequent saving as SD format for following docking experiments. The Log(IC50) (nM) values were applied in modeling analysis, hence linearly correlating the biological data to the free energy change. The obtained 3D structures were used as starting conformers for conformational analysis for pharmacophore modeling.

Data sets

HYPOGEN module from the CATALYST software package was employed to construct numerous plausible binding hypotheses for TACE inhibitors. (Duan et al. 2007; Duan et al. 2008; Lu et al. 2008; Ott et al. 2008a, b, c). The conformational space of each inhibitor (1–115, Suppl. material 1: table A) was explored adopting the “CAESAR” option within CATALYST (Smellie and Teig 1995; Khanfar and Taha 2013). Detailed experimental and theoretical explanations of pharmacophore modeling and conformational analysis are provided in the Suppl. material 1: Section SM-1, SM2, SM-3.

An “Uncertainty” value of 3 was reported for the inhibitors biological data, which means that the actual bioactivity of a particular inhibitor is assumed to be within an interval ranging from one-third to three-times the reported bioactivity value of that inhibitor (Sutter et al. 2000). Afterwards, three structurally diverse training subsets were carefully selected from the collection for pharmacophore modeling: sets A, B and C (Suppl. material 1: table B) as demonstrated in Scheme 1.

Pharmacophoric hypotheses generation

Three structurally diverse training subsets were selected from the collected 115 inhibitors (Suppl. material 1: table B). For each subset group, eight modeling runs were performed to explore the pharmacophoric space of TACE inhibitors. The difference between the resulted hypotheses were based on altering the interfeature spacing and the number of allowed features in the resulting pharmacophores (Suppl. material 1: table C).

Eventually, pharmacophore exploration (8 automatic runs, Suppl. material 1: tables C, D) culminated in 240 pharmacophore models of variable qualities (See Suppl. material 1: SM-2 for details about CATALYST Pharmacophore Generation Algorithm)(Li and Hoffmann 2000; Güner et al. 2004) and the summary of the experimental part is summarized in Scheme 1.

Assessment and clustering of the generated pharmacophore hypotheses

Upon the generation of hypotheses, CATALYST seeks to reduce a cost function consisting of three terms: Weight cost, error cost and configuration cost (Sutter et al. 2000; Catalyst 4.11 User Guide 2005; Discovery Studio 2.5.5 User Guide 2010). Clustered 46 pharmacophores, out of

240 generated models, were found to possess Fisher confidence values $\geq 90\%$ (see section SM-3). Suppl. material 1: tables C, D show the success criteria of representative pharmacophores from each run. Detailed theoretical explanations of CATALYST’s assessment of binding hypotheses are provided in the Suppl. material 1: SM-3.

CATALYST based method was applied for clustering of the successful models (240) which were clustered into 46 groups applying the hierarchy average linkage. Among 240 hypotheses, the closely-related pharmacophores were gathered in three-membered clusters. Accordingly, the highest-ranking representatives based on their fit-to-bioactivity correlation r^2 -values (calculated against collected compounds 1-115), were designated to represent their corresponding clusters in subsequent QSAR modeling (Suppl. material 1: table D).

KNN-based descriptor selection

The kNN-QSAR methodology relies on a distance learning approach such that the activity value of an unknown member is calculated from the activity values of certain number (k) of nearest neighbors (kNNs) in the training set. The equations and the details of calculated distances for kNNs are explained in the Suppl. material 1: SM-4.

ROC curve analysis

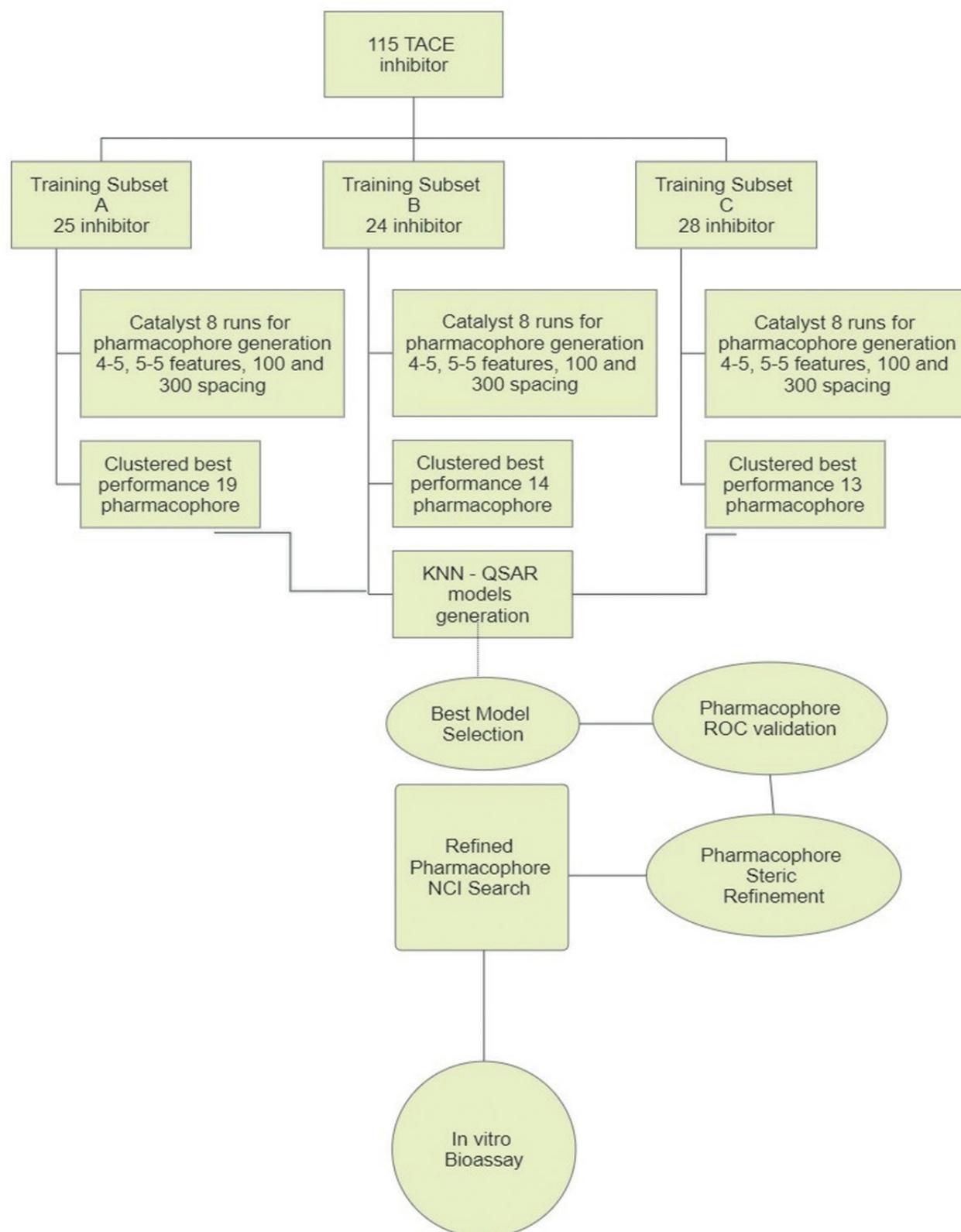
Successful GFA-kNN selected pharmacophore models were validated by assessing their abilities to selectively capture diverse TACE inhibitors from a large list of decoys employing ROC analysis as described by Verdonk and co-workers.(Verdonk et al. 2004; Triballeau et al. 2005; Kirchmair et al. 2008). For each active compound in the testing set an average of 35 decoys were randomly chosen from the ZINC database (Irwin and Shoichet 2005). See Suppl. material 1: SM-5 for detailed experimental and theoretical explanations of ROC analysis.

Addition of exclusion volumes

To account for the steric constraints of the binding pocket and to optimize the ROC curves of our QSAR-selected pharmacophores, it was decided to add exclusion volumes to the successful GFA-kNN selected pharmacophore models employing the HIPHOP-REFINE module of CATALYST. HIPHOP-REFINE uses inactive training compounds to add exclusion spheres to resemble the steric constraints of the binding pocket. It identifies spaces occupied by the conformations of inactive compounds and free from active ones. These regions are then filled with excluded volumes (Clement and Mehl 2000). More details are provided in the Suppl. material 1: SM-6).

Preparation of TACE protein structure

The 3D structure of TACE was downloaded from the Protein Data Bank (TACE, PDB code: 2i47, resolution: 1.9 Å). Addition of hydrogen atoms to the protein was



Scheme 1. The Experimental flowchart for experimental kNN-QSAR Modeling.

performed utilizing Discovery Studio 2016 templates for protein residues. No further energy minimization of the protein structure was applied before performing the docking experiments. It was very difficult to gain a selective TACE protein over various Matrix metalloproteinases

(MMPs), mainly MMP-2 and MMP-13 due to high similarity in these proteins and TACE structure. The selection of this protein structure (PDB code: 2i47) has preferences over other available protein crystallographic structures due to high resolution value, in addition to the selectivity

of this protein structure with a well-known binding pocket (Condon et al. 2007).

TACE ligands libdocking

LibDock docks the ligands into a binding site directed by binding hot spots. It aligns docked ligand conformations to polar and nonpolar receptor interactions sites, i.e., hotspots (Diller and Mertz 2001).

In the current docking experiments, certain parameters were employed for docking and scoring of the total list (115) TACE inhibitor as mentioned under Suppl. material 1: SM-7.

Ligandfit docking

Ligandfit docking engine considers receptor as rigid structure and treat the ligands as flexible compounds (Venkatachalam et al. 2003). Steps of Ligandfit and the corresponding protocol settings are mentioned under Suppl. material 1: SM-8. It aligns docked ligand conformations to polar and nonpolar receptor interactions sites, i.e., hotspots. The binding site is determined from the co-crystalline ligand (KGY, PDB Code: 2i47). Scoring of the high ranking docked conformers/poses was performed using 7 scoring functions, namely: Jain, LigScore1, LigScore2, PLP1, PLP2, PMF and PMF04 (Venkatachalam et al. 2003).

In the current docking experiments, certain parameters were employed for docking and scoring of the total list (115) TACE inhibitor as mentioned under Suppl. material 1: SM-8.

In-silico screening of NCI database

The sterically-refined version of Hypo(A-T6-4) selected in the optimal kNN model was employed as 3D search query to screen the National Cancer Institute list of compounds (includes 265,242 structures). Screening was performed employing the “Best Flexible Database” search option implemented within Discovery Studio 2016. NCI hits were subsequently filtered based on Lipinski’s rule such that only hits of H-bond donors less < 5, .molecular weights < 500 dalton, H-bond acceptors less < 10, H-bond acceptors < 5, and logP < 5 were retained (Lipinski et al. 2001; Veber et al. 2002). Filtered hits were fitted against Hypo(A-T6-4) by the “Best Fit” option in Discovery Studio 2016 and their fit values together with other relevant molecular descriptors were used to predict the activity of these selected hits. High-ranking hits were ordered from National Cancer Institute (NCI) for subsequent in vitro assay (see Suppl. material 1: table E).

In vitro TACE inhibition bioassay

The assay kit was purchased from Abcam’s TACE inhibitor screening Kit, TACE hydrolyzes the specific FRET sub-

strate to release the quenched fluorescent group, which can be detected fluorometrically at Ex/Em = 318/449 nm. In the presence of the tested TACE inhibitor, the hydrolyzation of substrate will be impeded.

The assay protocol is described as follows: an aliquot of (50 µL) of diluted TACE enzyme solution was mixed with (25 µL) of a testing diluted compound, the positive control, enzyme control (without inhibitor) or the buffer solution in the case of negative control. Mix well, and incubate for 5 minutes at 37 °C. Next, add (25 µL) of diluted substrate solution into each well and mix well. Then the mixture was incubated at 37 °C for 30 minutes protected from light. The fluorescence intensity (Excitation : 318 nm; Emission : 449 nm) was read in FLX800TBI Microplate Fluorimeter (BioTek Instruments, Winooski, VT, USA). The selected hits were dissolved in DMSO yielding 10 mM stock solutions followed by dilution to the required concentration using distilled deionized water. DMSO concentration was adjusted to 0.1%. The percentage of remaining TACE activity was identified in the presence and absence of the tested molecules. TACE activity is not affected by DMSO. The negative control samples were used as a contrast background. The carried experimental protocol and measurements were in duplicates. The % inhibition of TACE by the selected hits was calculated using the following equation (1)

$$\% \text{ Inhibition} = \frac{[\Delta\text{RFU of EC} - \Delta\text{RFU of S}]}{\Delta\text{RFU of EC}} * 100\% \quad (1)$$

Where, $\Delta\text{RFU of EC}$: The difference in fluorescence generated by hydrolyzation of substrate for enzyme control, $\Delta\text{RFU of S}$: The difference in fluorescence generated by hydrolyzation of substrate in presence of tested hit. The measurements were carried out in duplicates.

Results and discussion

QSAR modeling

After the generation of high quality 240 pharmacophores from 8 various runs for each subset (Suppl. material 1: table C). Clustering of the obtained pharmacophores resulted in 46 pharmacophore used in subsequent QSAR modeling (Suppl. material 1: table D). Best representative pharmacophore yielded high quality models of comparable success criteria (Suppl. material 1: table D). Thus, selecting particular binding hypothesis(es) to explain bioactivity variations across TACE inhibitors is a challenging process. Traditional GFA-MLR methods for normal QSAR equation generation were not efficient to have a correlation between essential pharmacophore, 2D descriptors as independent variables and biological activity as dependent variable. The fit values of these hypothesis against collected inhibitors (1-115) were enrolled together with a selection of 2D descriptors as independent variables in GFA/kNN-based QSAR analysis. QSAR based on kNN regression was applied as a competition platform to

extract the best possible combination of pharmacophores and other molecular descriptors collectively capable of explaining bioactivity variations across collected TACE inhibitors (Khanfar and Taha 2013). kNN is defined as a non-linear nonparametric method that predicts a ligand's bioactivity as the weighted-average of the bioactivities of its *k* nearest neighbors. The neighborhood is defined based on certain selected descriptor(s). The measurement of nearness is based on an appropriate distance metric. (Sharaf and Kowalski 1986; Zheng and Tropsha 2000; Khanfar and Taha 2013).

kNN-based QSAR modeling

We employed kNN-based QSAR modeling for TACE target. The kNN-QSAR methodology is based on a distance-based learning approach, which calculates the activity value of an unknown member is based on the activity values of certain number (*k*) of nearest neighbors (kNNs) in the training set. The distance metric similarity is measured by Euclidean distance. We applied the following kNN workflow: (1) Compute Euclidean distances between an unknown object (*x*) and all the objects in the training set with respect to certain descriptor(s); (2) choose from the training set *k* objects which are most similar to object *x*; (3) compute the distance-weighted average bioactivities of *k* objects as predicted bioactivity of *x*. (4) Correlate between the predicted bioactivities with the experimental ones to define the optimal *k* value and illustrative descriptors via leave-20% out cross-validation (Sharaf and Kowalski 1986; Zheng and Tropsha 2000). Table 3 shows the selected descriptors, nearest neighbors and statistical criteria of the top 5 kNN-based QSAR models. We selected model number 1 (Table 3) as the best representative for subsequent virtual screening and QSAR-based predictions because it exhibits excellent overall explanatory power with the least number of descriptors and nearest neighbors (Sharaf and Kowalski 1986; Zheng and Tropsha 2000).

kNN-QSAR model (1) selected four descriptors encoding for dipole *Z* that ensures the importance of polarity and the formation of hydrogen bonding with TACE binding site. The second descriptor is *ES_Count_sssCH* which indicates the number of methanetriyl groups. Obviously, the presence of the *ES_Count_sssCH* descriptor in Model 1 indicates the participation of bulky substituents of quinoline and other heterocyclic compounds in steric interactions aromatic and olefinic carbon atoms (*ES_Count_sssCH*). The topology, the connectivity and the bulkiness of the structure represented by (*CHI_V_3_P*), in addition to *Jurs_SASA* is calculated by mapping atomic partial charges on total solvent accessible surface areas of individual atoms. This descriptor has positively influenced the binding affinity of the inhibitors (Buglak et al. 2019).

Additionally, the appearance of dipole *Z* descriptor in model (1) which indicates The components of the dipole moment along z-axis, in addition to $3\chi_P$ is path connectivity index. Moreover, it selected one pharmacophores as ad-

ditional explanatory descriptors, namely, **Hypo(A-T6-4)**. Interestingly, many of the descriptors in kNN-QSAR model (1), repeatedly emerged in other best performing kNN-QSAR models (Table 3), including the pharmacophore model which add further weight to these descriptors.

Repeated appearance of Hypo(A-T6-4) among the optimal kNN-QSAR models indicates the essential pharmacophoric features.

The repeated appearance of dipole *Z* and ES count descriptor in top ranking kNN-QSAR models, including model 1 (Table 3), is suggestive of significant role played by aromatic or aliphatic nitrogen atoms in ligand binding within TACE probably through hydrogen bond attractive forces or through ionic attractive forces to acidic amino acid moieties in the binding pocket. Similarly, aliphatic and aromatic carbon atoms count descriptors probably encode for affinity interactions connecting different training ligands and hydrophobic moieties within TACE binding pocket. TACE binding site contains several hydrophobic and aromatic moieties capable of π - π -stacking and hydrophobic interactions with various ligands including, Pro437, Lys315, Leu350, Leu318 and Tyr352 (Fig. 1).

Addition of exclusion volumes and ROC curve analysis

Fig. 2 shows **Hypo(A-T6-4)** and how it maps co-crystallized ligand within the closely-homologous protein (PDB Code: 2i47), while Table 1 shows the X, Y, and Z coordinates of **Hypo(A-T6-4)**.

Table 1. TACE based pharmacophore model selected by kNN-QSAR modeling.

Model	Definitions	Chemical features						
Hypo A-T6-4		HBD		R.A		Hbic		
	Weights	2.24	2.24	2.24	2.24	2.24		
	Tolerances	1.6	2.2	1.6	2.2	1.6	1.6	
	Coordinates X	1.78	3.91	3.17	5.28	2.24	0.39	-0.83
	Y	-1.34	-3.37	1.93	-0.19	1.06	-1.07	-3.55
Z	-0.13	0.48	3.51	3.94	7.35	6.31	9.03	

Number of Associated Exclusion Spheres (A-T6-4) at the following X, Y, Z coordinates:

Excluded volume (0.51, -2.40, 3.40) Tol: 120.00;

Excluded volume (-0.04, 1.06, 2.40) Tol: 120.00;

Excluded volume (4.97, 5.27, 4.26) Tol: 120.00.

To further validate our kNN-QSAR-selected pharmacophores, we subjected them to ROC curve analysis to assess their abilities to selectively capture diverse TACE inhibitors from a large list of decoys as clearly mentioned in Suppl. material 1: SM-5. The ROC performance was accepted for the selected pharmacophore with ROC-AUC value of approximately 0.8 as shown in Table 2. Consequently, for the improvement of the ROC-AUC curve parameter, addition of exclusion spheres was performed using HipHop refining list that is selected in Table 4 and according to the procedure mentioned under Suppl. mate-

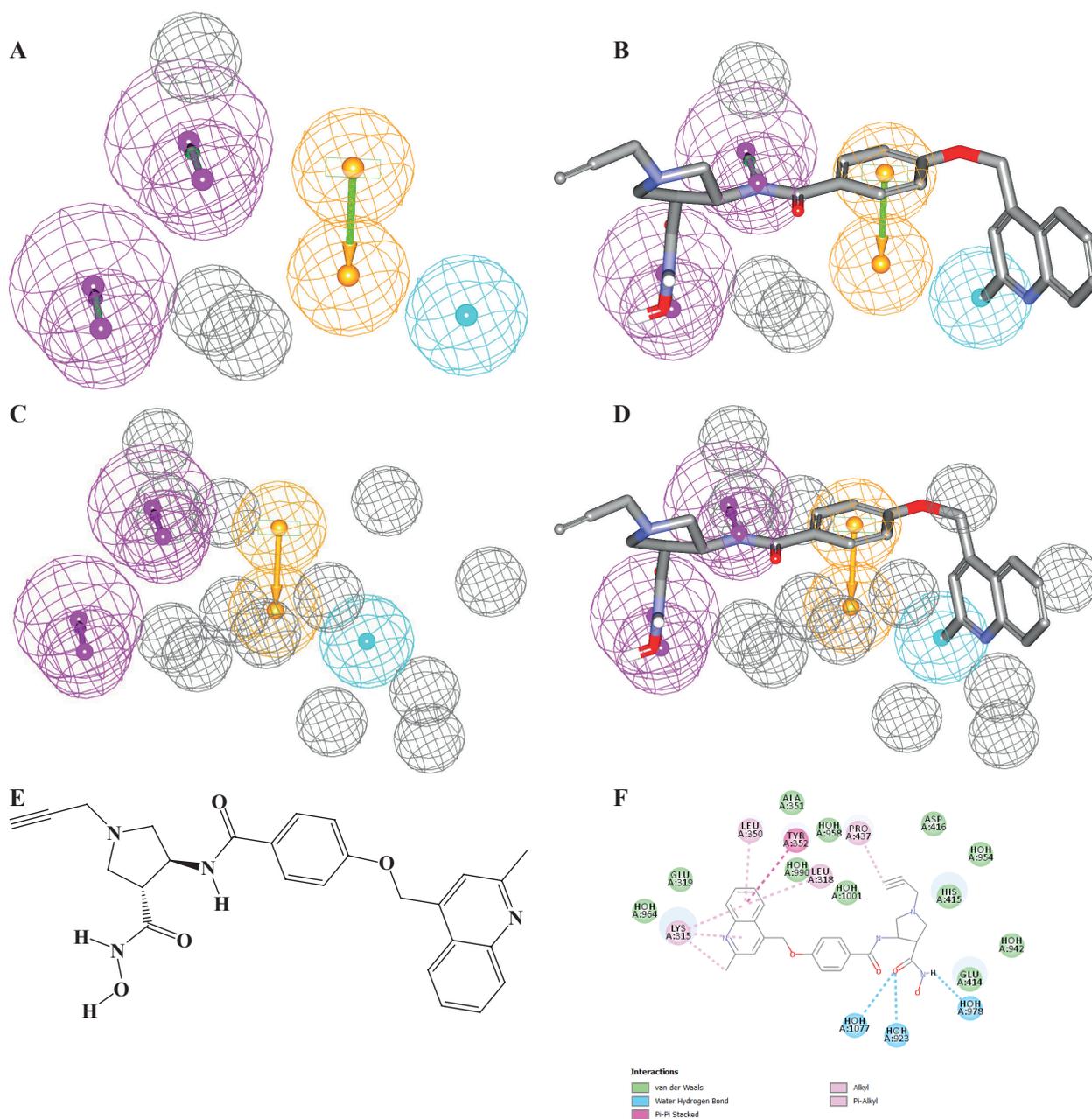


Figure 1. Hypo(A-T6-4). **A.** Pharmacophoric features of the binding model: HBD as violet vectored spheres, Ring aromatic as orange spheres, Hbic as blue spheres and Exclusion volumes as gray spheres); **B.** Hypo(A-T6-4) mapped against potent TACE inhibitor 90 (Suppl. material 1: table A); **C.** Refined Hypo(A-T6-4); **D.** Refined Hypo(A-T6-4) mapped against potent TACE inhibitor 90, IC_{50n} = 0.14nM (Suppl. material 1: table A); **E.** Chemical structure of TACE inhibitor 90; **F.** Libdock docked pose of inhibitor 90 within binding site of TACE binding site(PDB code: 2i47).

Table 2. ROC^a performances of KNN-selected pharmacophores and their sterically refined versions as 3D search queries.

Pharmacophore model	ROC ^a -AUC ^b	ACC ^c	SPC ^d	TPR ^e	FNR ^f
Hypo A-T6-4	0.798	0.972	0.99	0.25	0.007
Hypo A-T6-4 Refined	0.929	0.972	0.99	0.25	0.007

^aROC: Receiver operating characteristic; ^bAUC: Area under the curve; ^cACC: Overall accuracy; ^dSPC: Overall specificity; ^eTPR: Overall true positive rate; ^fFNR: Overall false negative rate.

rial 1: SM-6. Ten exclusion spheres were added and the obtained **Refined Hypo (A-T6-4)** ROC-AUC was increased to 0.929 as obviously illustrated in Table 2 and Suppl. ma-

terial 1: fig. A, SM-6, which represents the shapes of ROC curves for the unrefined and refined **Hypo (A-T6-4)**. Despite low values of TPR = 0.25, selectivity (SCC = 0.99) is high which increases the hits inhibition against TACE over MMPs. Also low false negative rate (FNR = 0.007) decreases the loss of potential active hits.

Ligandfit and libdock

Additionally, Ligandfit and Libdock docking experiments performed for the most active hits are shown obviously in Figs 4, 5. The binding site of TACE protein (PDB

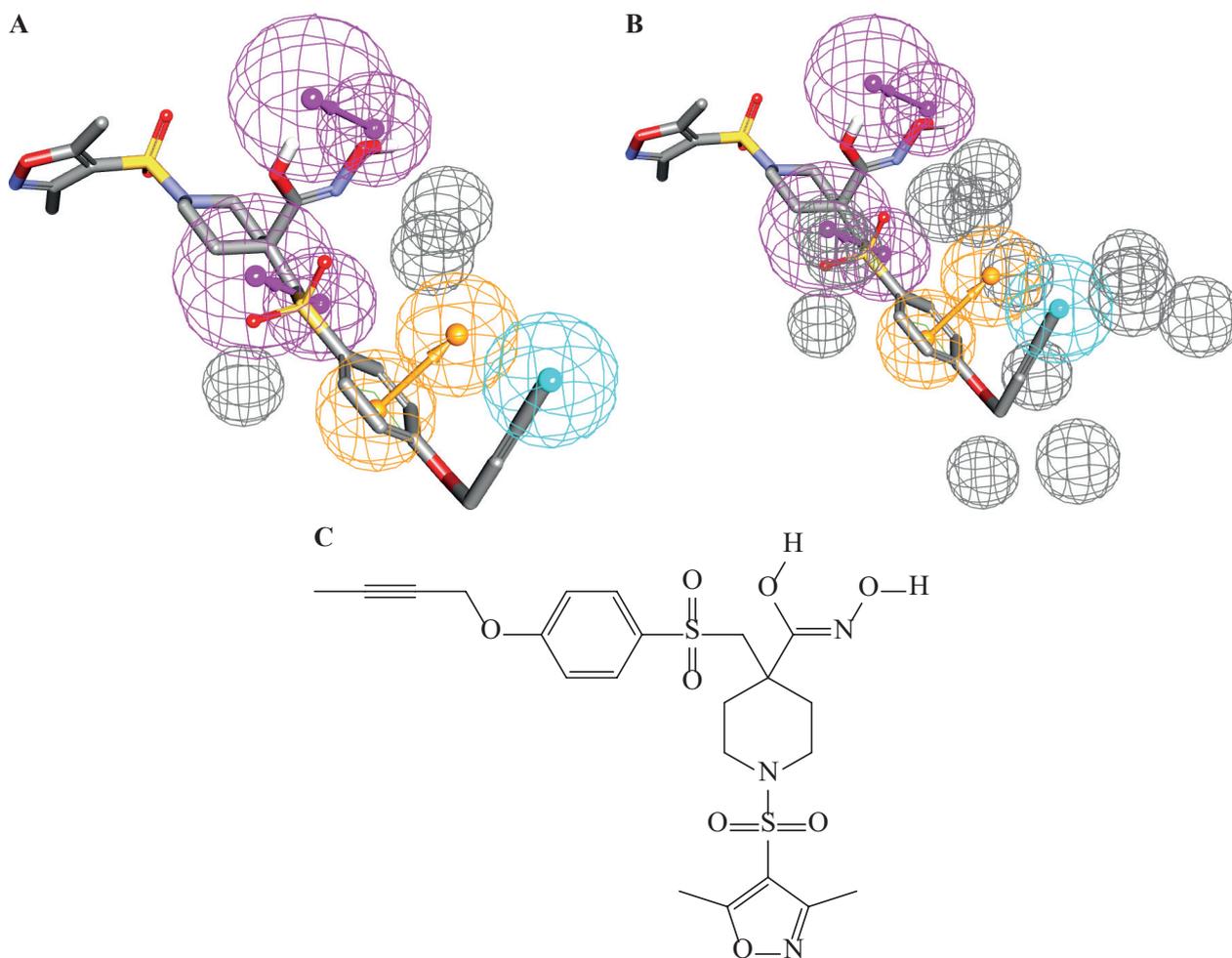


Figure 2. A. Hypo(A-T6-4) mapped against co-crystalline ligand (KGY, PDB code: 2i47) Pharmacophoric features of the binding model: HBD as violet vectored spheres, Ring aromatic as orange spheres, Hbic as blue spheres and Exclusion volumes as gray spheres); B. Refined Hypo(A-T6-4) mapped against co-crystalline ligand (KGY, PDB code: 2i47); C. Chemical structure of co-crystalline ligand (KGY, PDB code: 2i47).

code: 2i47) simulates the pharmacophoric features of **Hypo(A-T6-4)** as hydrogen bonding was formed with various amino acids according to the binding mode of the NCI hit such as Glu 406 and H₂O 927 as in the case of hit 118 (Fig. 4B), hydrophobic attractive forces with His 405, Lys 315, Gly 346 as presented with several hits (Fig. 4A, E, F). Also aromatic π - π stacking is shown with aromatic amino acid residues as shown with His 405, His 409 and His 415 with hit **140** (Fig. 4G).

According to the settings of Libdock (Suppl. material 1: SM-7), the docked active NCI hits poses are presented in Fig. 5, for example hit **128** forms hydrogen bonding with amino acid residues His 415 and Asp 418, hydrophobic forces are formed with Glu319, Leu 318 and Leu 350 (Fig. 5C). Additionally, the most potent hit **131** has more obvious attractive forces with TACE binding site (Fig. 5E), that hydrogen bonding is clear with H₂O 923, H₂O 990 and Ala 351, aromatic π - π stacking is presented with Tyr 351 and His 415 which simulates the aromatic ring pharmacophoric feature of **Hypo(A-T6-4)**. Finally, hydrophobic attractive forces with Leu350, Lys315 and Leu318.

Furthermore, to ensure the correct selection of docking settings performed above, the co-crystalline ligand KGY (PDB: 2i47) was re-docked again within TACE binding site and according to the previous scoring functions. Clearly from the figure, the applied docking settings closely reproduced the bound structure with RMSD value of 1.14 Å° which support the confidence in our docking experiment.

NCI search query and *in vitro* bioassay

The filtered hits activity was predicted according to selected kNN model, 28 out of the best predicted NCI compounds were tested as TACE inhibitors using the bioassay kit as mentioned in the experimental part. The inhibitory activity of the selected NCI hits was performed at 100 μ M concentration and the obtained results were tabulated with the corresponding structure for each hit as shown in Suppl. material 1: table E.

Mapping of the most potent hits against **Hypo(A-T6-4)** are presented in Fig. 3 which shows that 8 active hits with

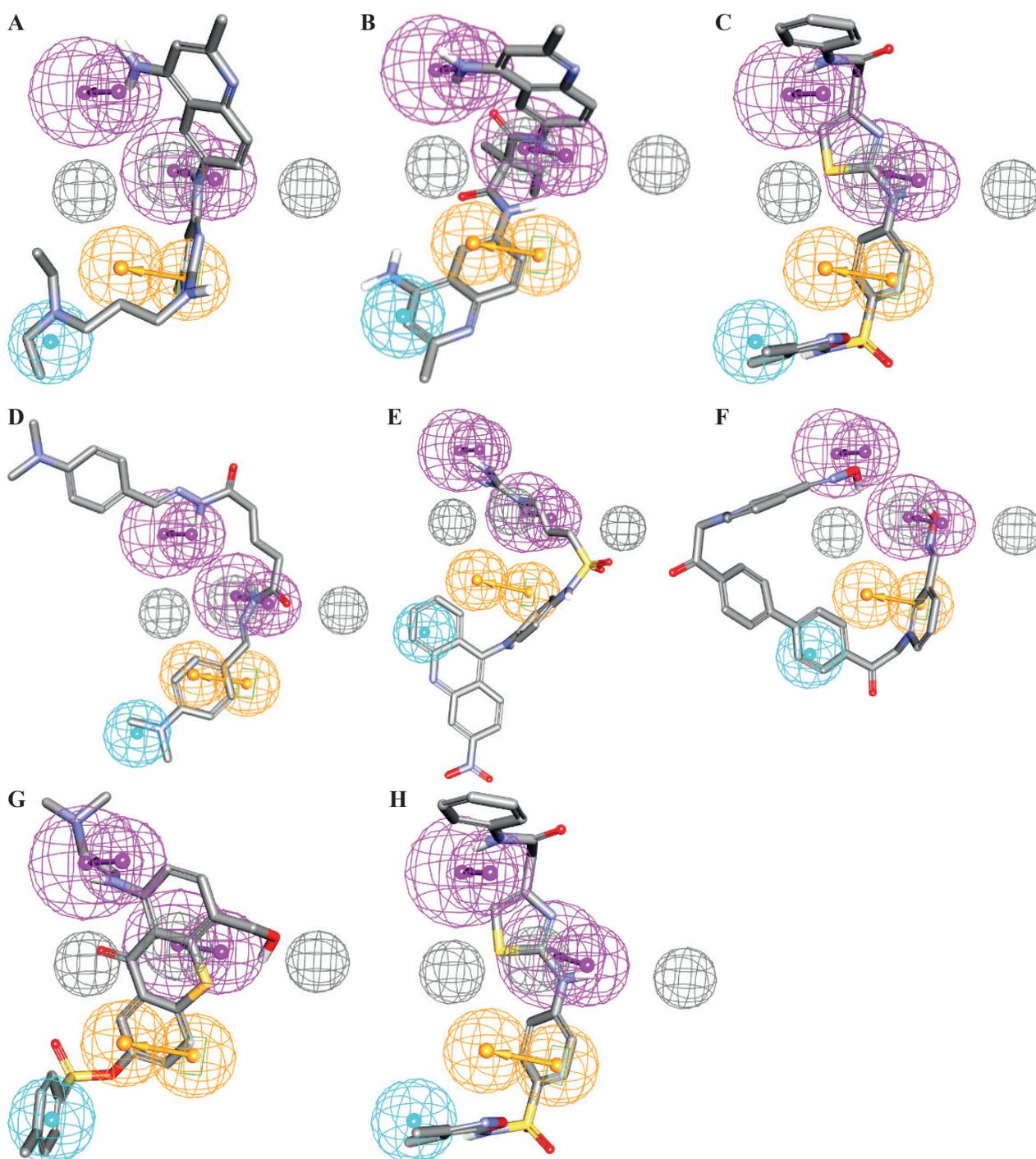


Figure 3. A. Hypo(A-T6-4) mapped against 117^a(51%^b); B. Hypo(A-T6-4) mapped against 118^a(81%^b); C. Hypo(A-T6-4) mapped against 125^a(51.9%^b); D. Hypo(A-T6-4) mapped against 128^a(53.3%^b); E. Hypo(A-T6-4) mapped against 130^a(51.2%^b); F. Hypo(A-T6-4) mapped against 131^a(81.4%^b); G. Hypo(A-T6-4) mapped against 140^a(52.2%^b); H. Hypo(A-T6-4) mapped against 143^a(54.7%^b); ^aThe number of the active NCI hit (Suppl. material 1: table E); ^bThe inhibition percent of the active NCI hit at 100 μ M concentration (Suppl. material 1: table E).

TACE inhibition activity exceeding 50% at 100 μ M concentration. Obviously, the active hits mapped with Hypo(A-T6-4) without any missing features which explains the activity of the captured hits.

As explained before, the active hits mapped with the selected pharmacophore in our successful model (Hypo(A-T6-4) as obviously illustrated in Fig. 3. The dif-

ferences in the obtained activity of the NCI hits can be explained by the variations in the docking poses of these hits and the types of the forces formed with amino acid residues as presented in Figs 4, 5. The better the matching between the docking hit forces with the pharmacophoric feature of Hypo(A-T6-4), the higher the activity of our captured hit as mentioned for hit 118 and 131

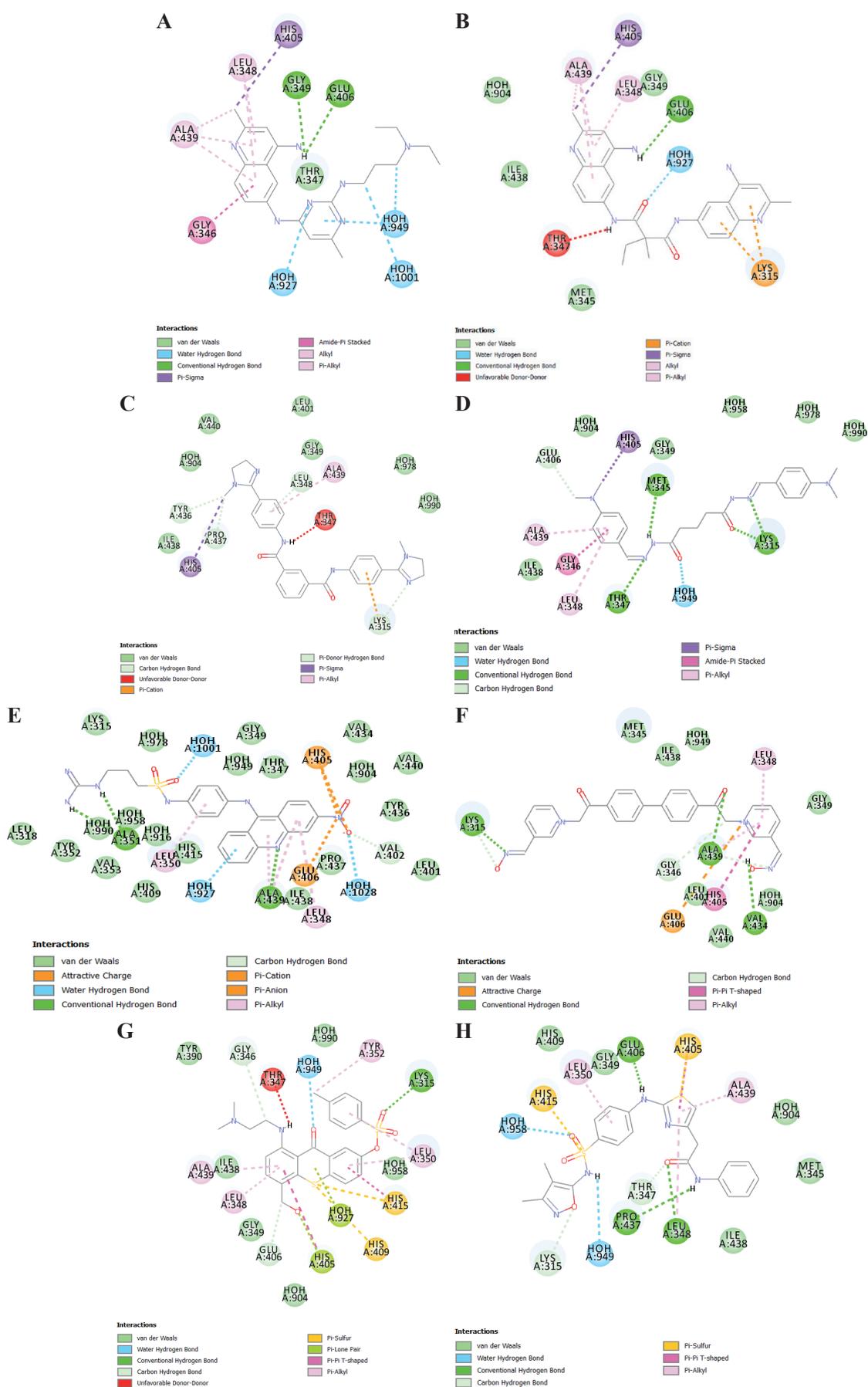


Figure 4. Ligand docked poses of the most potent hits **A**. 117^a(51%^b); **B**. 118^a(81%^b); **C**. 125^a(51.9%^b); **D**. 128^a(53.3%^b); **E**. 130^a(51.2%^b); **F**. 131^a(81.4%^b); **G**. 140^a(52.2%^b); **H**. 143^a(54.7%^b); ^a: The number of the active NCI hit (Suppl. material 1: table E); ^b: The inhibition percent of the active NCI hit at 100 μ M concentration (Suppl. material 1: table E).

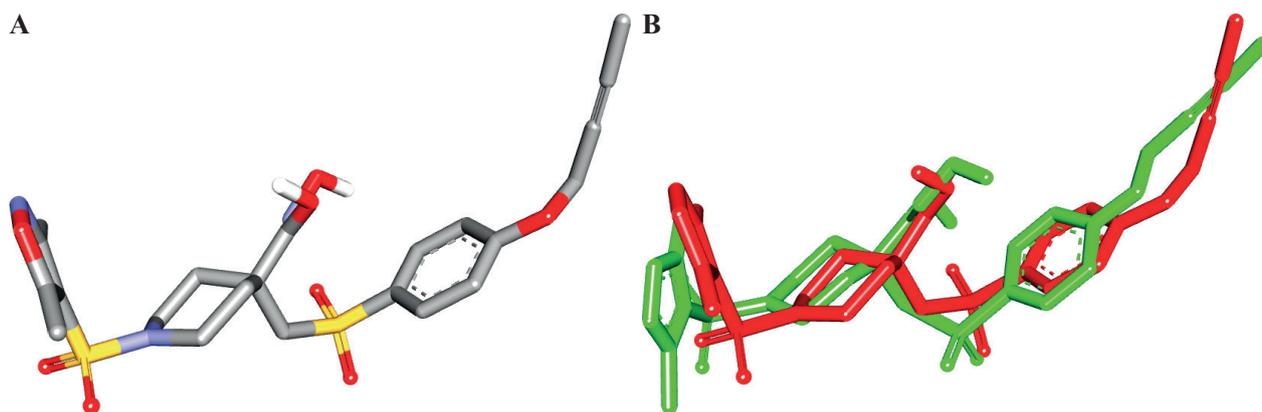


Figure 6. A. Structure of co-crystalline compound KGY against B. Comparison between the docked pose of KGY (red) as produced by docking simulation and the crystallographic structure of this inhibitor within TACE (green, PDB code: 2i47).

Table 3. Optimal kNN-QSAR models including their corresponding descriptors, nearest neighbors, and statistical criteria.

Model	Selected descriptors	Number of nearest neighbors	Statistical criteria		
			r^{2a}	r^2LOO^b	$r^2L20\%Out^c$
1	Hypo(A-T6-4) ^d	4	0.952096	0.857226	0.843372083
	CHI_V_3_P				
	Dipole_Z				
	ES_Count_sssCH				
	Jurs_SASA				
2	Hypo(A-T6-4)	4	0.941066	0.863068	0.846875968
	Hypo(DIV-T4-10) ^d				
	CIC				
	LogD				
3	Hypo(A-T6-4)	4	0.957687	0.875218	0.864507451
	Hypo(A-T2-9) ^d				
	Hypo(B-T2-8) ^d				
	Dipole_Z				
	ES_Count_dO				
4	Hypo(A-T6-4)	3	0.962255	0.87031	0.855564671
	Hypo(A-T2-9)				
	ES_Count_sOH				
	ES_Sum_ssCH2				
	Shadow_XYfrac				
5	Hypo(A-T6-4)	3	0.930032	0.860158	0.85490382
	Hypo(A-T5-1) ^d				
	Hypo(A-T7-4) ^d				
	Hypo(DIV-T8-3)				
	BIC				
SIC					

Table 4. The training compounds used for adding excluded spheres for **Hypo(A-T6-4)** using HIPHOP-REFINE module of CATALYST.

Compd ^a	IC ₅₀ (nM)	Principal value	MaxOmitFeat ^b
1	1,030	0	2
2	1,300	0	2
3	2,200	0	2
4	49,000	0	1
5	100,000	0	2
6	11,000	0	2
7	800	0	1
9	100,000	0	1
10	100,000	0	1
11	32,000	0	1
12	2,360	0	2
44	1	2	0
46	1	2	0
56	1	2	0
58	1	2	0
63	1	2	0
71	1	2	0
76	1	2	0
86	0.4	2	0
87	1.1	2	0
88	0.94	2	0
90	0.14	2	0
96	1	2	0
97	1	2	0
101	1	2	0
103	1	2	0
110	1	2	0
115	1	2	0

^aCompounds' numbers are as in Suppl. material 1: fig. A, table A; ^bMaxOmitFeat: maximum omitted features.

Conclusion

TNF- α converting enzyme is considered as a main cause of various serious diseases, so the discovery of new TACE inhibitors is remarkable step especially with drawbacks of the marketed known inhibitors. Ordinary QSAR modeling process was not efficient in the creation of normal QSAR equation comprising the most significant pharmacophores and 2D descriptors which manipulates the inhibitory activity of TACE inhibitors. Another approach via kNN-QSAR modeling procedure

was a successful alternate for obtaining a validated models which clearly selected the essential pharmacophoric features needed for TACE inhibitors among 240 generated pharmacophores using three various subsets, in addition to properly picking out of crucial 2D descriptors in the model. ROC validated confirms the success of the selected phamacophore upon its validation and the subsequent active Refined **Hypo(A-T6-4)** was applied as search query on NCI compound. Selected NCI com-

pounds inhibitory activity was examined and achieved TACE inhibitory activity exceeding 80% at 100 μ M concentration for some NCI hits.

Abbreviations used

GFA, genetic function algorithm; Hbic, hydrophobic; kNN, k nearest neighbor; TACE, TNF- α converting enzyme; NCI, national Cancer Institute; RingArom, ring aromatic; ROC, receiver operating characteristic; AUC: area under the curve; ACC: overall accuracy; SPC: overall

specificity; TPR: overall true positive rate; f FNR: overall false negative rate; MMPs: Matrix metalloproteinases

Acknowledgments

The authors thank the head of Drug Discovery Unit at the University of Jordan (Professor Mutasem Taha) for software supply and beneficial help. The authors are also thankful to the National Cancer Institute for freely providing NCI hits.

References

- Adrian Huang DJM, Lovering F, Sun L, Wang W, Xu W, Zhu Y, Cui J, Zhang Y, Levinc J (2007) Structure-based design of TACE selective inhibitors: Manipulations in the S10–S30 pocket. *Bioorganic & Medicinal Chemistry* 15: 6170–6181. <https://doi.org/10.1016/j.bmc.2007.06.031>
- Batista J (2020) Tumor necrosis factor Alpha converting enzyme (TACE) as possible therapeutic target in SARS-CoV-2 induced Acute Respiratory Distress Syndrome (ARDS). *ChemRxiv*. Cambridge Open Engage, Cambridge.
- Bandarage UK, Wang T, Come JH, Perola E, Wei Y, Rao BG (2008) Novel thiol-based TACE inhibitors. Part 2: Rational design, synthesis, and SAR of thiol-containing aryl sulfones. *Bioorganic & Medicinal Chemistry Letters* 18(1): 44–48. <https://doi.org/10.1016/j.bmcl.2007.11.014>
- Buglak AA, Zherdev AV, Lei HT, Dzantiev BB (2019) QSAR analysis of immune recognition for triazine herbicides based on immunoassay data for polyclonal and monoclonal antibodies. *PLoS ONE* 14(4): e0214879. <https://doi.org/10.1371/journal.pone.0214879>
- Catalyst 4.11 User Guide (2005) Accelrys Software Inc. San Diego, CA.
- Cherney RJ, King BW, Gilmore L, Liu RQ, Covington MB, Duan JJ, Decicco CP (2006) Conversion of potent MMP inhibitors into selective TACE inhibitors. *Bioorg Med Chem Lett* 16(4): 1028–1031. <https://doi.org/10.1016/j.bmcl.2005.10.078>
- Clement OO, Mehl AT (2000) HipHop: Pharmacophores based on multiple common-feature alignment. In: Güner OF (Ed.) *Pharmacophore Perception, Development, and Use in Drug Design*. International University Line, San Diego.
- Condon JS, Joseph-McCarthy D, Levin JI, Lombart HG, Lovering FE, Sun L, Wang W, Xu W, Zhang Y (2007) Identification of potent and selective TACE inhibitors via the S1 pocket. *Bioorganic & Medicinal Chemistry Letters* 17(1): 34–39. <https://doi.org/10.1016/j.bmcl.2006.10.004> [PMID: 17064892]
- Dai C, Li D, Popovici-Muller J, Zhao L, Girijavallabhan VM, Rosner KE, Siddiqui MA (2011) 2-(2-Aminothiazol-4-yl)pyrrolidine-based tartrate diamides as potent, selective and orally bioavailable TACE inhibitors. *Bioorganic Medicinal Chemistry Letter* 21(10): 3172–3176. <https://doi.org/10.1016/j.bmcl.2011.01.002>
- DasGupta S, Murumkar PR, Giridhar R, Yadav MR (2009) Current perspective of TACE inhibitors: a review. *Bioorganic and Medicinal Chemistry Letter* 17(2): 444–459. <https://doi.org/10.1016/j.bmc.2008.11.067>
- De Queiroz TM, Lakkappa N, Lazartigues E (2020) ADAM17-mediated shedding of inflammatory cytokines in hypertension. *Front Pharmacol* 11: e1154. <https://doi.org/10.3389/fphar.2020.01154>
- Diller DJ, Merz KMJ (2001) High throughput docking for library design and library prioritization. *Proteins* 43(2): 113–124. [https://doi.org/10.1002/1097-0134\(20010501\)43:2<113::AID-PROT1023>3.0.CO;2-T](https://doi.org/10.1002/1097-0134(20010501)43:2<113::AID-PROT1023>3.0.CO;2-T)
- Discovery Studio 2.5.5 User Guide (2010) Discovery Studio 2.5.5 User Guide. Accelrys Inc., San Diego.
- Duan JJ, Chen L, Lu Z, Jiang B, Asakawa N, Sheppeck JE, Decicco CP (2007) Discovery of low nanomolar non-hydroxamate inhibitors of tumor necrosis factor-alpha converting enzyme (TACE). *Bioorganic and Medicinal Chemistry Letter* 17(1): 266–271. <https://doi.org/10.1016/j.bmcl.2006.09.048>
- Duan JJ, Chen L, Lu Z, Xue CB, Liu RQ, Covington MB, Decicco CP (2008) Discovery of β -benzamido hydroxamic acids as potent, selective, and orally bioavailable TACE inhibitors. *Bioorganic and Medicinal Chemistry Letters* 18(1): 241–246. <https://doi.org/10.1016/j.bmcl.2007.10.093>
- Gorbalenya AE, Susan C, Ralph S, Raoul J, Christian D, Anastasia A, Bart L, Lauber C, Andrey M, Benjamin W, Penzar D, Perlman S, Poon LLM, Dmitry V, Igor A, Sola I, Ziebuhr J (2020) The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature. Microbiology* 5(4): 536–544. <https://doi.org/10.1038/s41564-020-0695-z>
- Rao G, Bandarage UK, Wang T, Come JH, Perola E, Wei Y, Saunders JO (2007) Novel thiol-based TACE inhibitors: rational design, synthesis, and SAR of thiol-containing aryl sulfonamides. *Bioorganic and Medicinal Chemistry Letters* 17(8): 2250–2253. <https://doi.org/10.1016/j.bmcl.2007.01.064>
- Guan W-j, Ni Z-y, Hu Y, Liang W-h, Ou C-q, He J-x, Liu L, Shan H, Lei C-l, Hui DSC, Du B, Li L-j, Zeng G, Yuen K-Y, Chen R-c, Tang C-l, Wang T, Chen P-y, Xiang J, Li S-y, Wang J-l, Liang Z-j, Peng Y-x, Wei L, Liu Y, Hu Y-h, Peng P, Wang J-m, Liu J-y, Chen Z, Li G, Zheng Z-j, Qiu S-q, Luo J, Ye C-j, Zhu S-y, Zhong N-s (2020) Clinical Characteristics of Coronavirus Disease 2019 in China. *New England Journal of Medicine* 382(18): 1708–1720. <https://doi.org/10.1056/NEJMoa2002032>
- Güner O, Clement O, Kurogi Y (2004) Pharmacophore modeling and three dimensional database searching for drug design using catalyst: recent advances. *Current Medicinal Chemistry* 11(22): 2991–3005. <https://doi.org/10.2174/0929867043364036>
- Hotamisligil GS, Budavari A, Murray D, Spiegelman BM (1994) Reduced tyrosine kinase activity of the insulin receptor in obesity-diabetes. Central role of tumor necrosis factor-alpha. *Journal of Clinical Investigation* 94(4): 1543–1549. <https://doi.org/10.1172/JCI117495>
- Irwin JJ, Shoichet BK (2005) ZINC – A Free Database of Commercially Available Compounds for Virtual Screening. *Journal of Chemical*

- Information and Modeling 45(1): 177–182. <https://doi.org/10.1021/ci049714>
- Janser P, Neumann U, Miltz W, Feifel R, Buhl T (2006) A cassette-dosing approach for improvement of oral bioavailability of dual TACE/MMP inhibitors. *Bioorganic and Medicinal Chemistry Letters* 16(10): 2632–2636. <https://doi.org/10.1016/j.bmcl.2006.02.042>
- Khanfar MA, Taha MO (2013) Elaborate ligand-based modeling coupled with multiple linear regression and k nearest neighbor QSAR analyses unveiled new nanomolar mTOR inhibitors. *Journal of Chemical Information Model* 53(10): 2587–2612. <https://doi.org/10.1021/ci4003798>
- Kirchmair J, Markt P, Distinto S, Wolber G, Langer T (2008) Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection—what can we learn from earlier mistakes? *Journal of Computer-Aided Molecular Design* 22(3–4): 213–228. <https://doi.org/10.1007/s10822-007-9163-6>
- Lebrech H, Ponce R, Preston BD, Iles J, Born TL, Hooper M (2015) Tumor necrosis factor, tumor necrosis factor inhibition, and cancer risk. *Current Medical Research Opinion* 31(3): 557–574. <https://doi.org/10.1185/03007995.2015.1011778>
- Levin JI, Chen JM, Laakso LM, Du M, Schmid J, Xu W, Skotnicki JS (2006) Acetylenic TACE inhibitors. Part 3: Thiomorpholine sulfonamide hydroxamates. *Bioorganic and Medicinal Chemistry Letters* 16(6): 1605–1609. <https://doi.org/10.1016/j.bmcl.2005.12.020>
- Li HS, Hoffmann R (2000) HypoGen: An Automated System for Generating 3D Predictive Pharmacophore Models. In: Güner OF (Ed.) *Pharmacophore Perception, Development, and Use in Drug Design*. International University Line, San Diego.
- Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Review* 46(1–3): 3–26. [https://doi.org/10.1016/S0169-409X\(00\)00129-0](https://doi.org/10.1016/S0169-409X(00)00129-0)
- Lombart HG, Feyfant E, Joseph-McCarthy D, Huang A, Lovering F, Sun L, Levin J (2007) Design and synthesis of 3,3-piperidine hydroxamate analogs as selective TACE inhibitors. *Bioorganic & Medicinal Chemistry Letters* 17(15): 4333–4337. <https://doi.org/10.1016/j.bmcl.2007.05.022>
- Lu Z, Ott GR, Anand R, Liu RQ, Covington MB, Vaddi K, Duan JJ (2008) Potent, selective, orally bioavailable inhibitors of tumor necrosis factor- α converting enzyme (TACE): discovery of indole, benzofuran, imidazopyridine and pyrazolopyridine P1' substituents. *Bioorganic and Medicinal Chemistry Letters* 18(6): 1958–1962. <https://doi.org/10.1016/j.bmcl.2008.01.120>
- Maekawa M, Tadaki H, Tomimoto D, Okuma C, Sano R, Ishii Y, Sasase T (2019) A Novel TNF- α Converting Enzyme (TACE) Selective Inhibitor JTP-96193 Prevents Insulin Resistance in KK-A(y) Type 2 Diabetic Mice and Diabetic Peripheral Neuropathy in Type 1 Diabetic Mice. *Biological and Pharmaceutical Bulletin* 42(11): 1906–1912. <https://doi.org/10.1248/bpb.b19-00526>
- Mazzola RD, Zhu Z, Sinning L, McKittrick B, Lavey B, Spitler J, Niu X (2008) Discovery of novel hydroxamates as highly potent tumor necrosis factor- α converting enzyme inhibitors. Part II: optimization of the S3' pocket. *Bioorganic and Medicinal Chemistry Letters* 18(21): 5809–5814. <https://doi.org/10.1016/j.bmcl.2008.09.045>
- Park K, Aplasca A, Du MT, Sun L, Zhu Y, Zhang Y, Levin JI (2006) Design and synthesis of butyryloxyphenyl beta-sulfone piperidine hydroxamates as TACE inhibitors. *Bioorganic and Medicinal Chemistry Letters* 16(15): 3927–3931. <https://doi.org/10.1016/j.bmcl.2006.05.026>
- Sharaf MA, Kowalski BR (1986) *Chemometrics*. Wiley, New York.
- Sheppeck JE, Gilmore JL, Tebben A, Xue CB, Liu RQ, Decicco CP, Duan JJ (2007a) Hydantoins, triazolones, and imidazolones as selective non-hydroxamate inhibitors of tumor necrosis factor- α converting enzyme (TACE). *Bioorganic and Medicinal Chemistry Letters* 17(10): 2769–2774. <https://doi.org/10.1016/j.bmcl.2007.02.076>
- Sheppeck JE, Gilmore JL, Yang A, Chen XT, Xue CB, Roderick J, Duan JJ (2007b) Discovery of novel hydantoins as selective non-hydroxamate inhibitors of tumor necrosis factor- α converting enzyme (TACE). *Bioorganic and Medicinal Chemistry Letters* 17(5): 1413–1417. <https://doi.org/10.1016/j.bmcl.2006.11.089>
- Sheppeck JE, Tebben A, Gilmore JL, Yang A, Wasserman ZR, Decicco CP, Duan JJW (2007c) A molecular modeling analysis of novel non-hydroxamate inhibitors of TACE. *Bioorganic & Medicinal Chemistry Letters* 17(5): 1408–1412. <https://doi.org/10.1016/j.bmcl.2006.11.082>
- Siddell SG, Walker PJ, Lefkowitz EJ, Mushegian AR, Adams MJ, Dutilh BE, Davison AJ (2019) Additional changes to taxonomy ratified in a special vote by the International Committee on Taxonomy of Viruses (October 2018). *Archive of Virology* 164(3): 943–946. <https://doi.org/10.1007/s00705-018-04136-2>
- Smellie AK, Teig SL (1995) Analysis of Conformational Coverage 0.2. Application of Conformational Models. *Journal of chemical information and computer sciences* 35: 295–304. <https://doi.org/10.1021/ci00024a019>
- Sutter J, Güner OF, Hoffman R, Li H, Waldman M (2000) Effect of Variable Weights and Tolerances on Predictive Model Generation. 499–512.
- Triballeau N, Acher F, Brabet I, Pin JP, Bertrand HO (2005) Virtual screening workflow development guided by the “receiver operating characteristic” curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *Journal of Medicinal Chemistry* 48(7): 2534–2547. <https://doi.org/10.1021/jm049092j>
- Udechukwu MC, Tsopmo A, Mawhinney H, He R, Kienesberger PC, Udenigwe CC (2017) Inhibition of ADAM17/TACE activity by zinc-chelating rye secalin-derived tripeptides and analogues. *RSC Advances* 7(42): 26361–26369. <https://doi.org/10.1039/C6RA26678A>
- Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD (2002) Molecular properties that influence the oral bioavailability of drug candidates. *Journal of Medicinal Chemistry* 45(12): 2615–2623. <https://doi.org/10.1021/jm020017n>
- Venkatachalam CM, Jiang X, Oldfield T, Waldman M (2003) LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. *Journal Molecule Graph Model* 21: 289–307. [https://doi.org/10.1016/S1093-3263\(02\)00164-X](https://doi.org/10.1016/S1093-3263(02)00164-X)
- Verdonk ML, Berdini V, Hartshorn MJ, Mooij WT, Murray CW, Taylor RD, Watson P (2004) Virtual screening using protein-ligand docking: avoiding artificial enrichment. *Journal of chemical information and computer sciences* 44(3): 793–806. <https://doi.org/10.1021/ci034289q>
- Yu Guo Z, Orth P, Madison V, Chen L, Dai C, Zhou G (2010) Discovery and SAR of hydantoin TACE inhibitors. *Bioorganic & Medicinal Chemistry Letters* 20(6): 1877–1880. <https://doi.org/10.1016/j.bmcl.2010.01.148>
- Yu W, Tong L, Kim SH, Wong MK, Chen L, Yang DY, Kozłowski JA (2010) Biaryl substituted hydantoin compounds as TACE inhibitors. *Bioorganic & Medicinal Chemistry Letters* 20(17): 5286–5289. <https://doi.org/10.1016/j.bmcl.2010.06.134>
- Zhang C, Lovering F, Behnke M, Zask A, Sandanayaka V, Sun L, Levin J (2009) Synthesis and activity of quinolinylmethyl P1' α -sulfone piperidine hydroxamate inhibitors of TACE. *Bioorganic & Medicinal Chemistry Letters* 19(13): 3445–3448. <https://doi.org/10.1016/j.bmcl.2009.05.020>
- Zheng W, Tropsha A (2000) Novel variable selection quantitative structure–property relationship approach based on the k-nearest-neighbor principle. *Journal of chemical information and computer sciences* 40(1): 185–194. <https://doi.org/10.1021/ci980033m>

Supplementary material 1

The discovery of potential tumor necrosis factor alpha converting enzyme inhibitors via implementation of K Nearest Neighbor QSAR analysis

Authors: Maha Habash, Sami Alshakhshir, Shady Awwad, Mahmoud Abu-Samak

Data type: Word document including equations and tables

Explanation note: SM-1. Conformational Analysis. SM-2. CATALYST pharmacophore generation algorithm. SM-3. Pharmacophore Assessment in CATALYST. SM-4. KNN-based descriptor selection. SM-5. ROC analysis. SM-6. Addition of Exclusion Volumes. SM-7. LibDock docking. SM 8. LigandFit docking. Figure A: Receiver operating characteristic (ROC) curves of kNN-QSAR selected pharmacophores: (A) Hypo(A-T6-4), (B) sterically refined Hypo(A-T6-4). Table A. The structures of TACE inhibitors utilized in modeling. Table B. The training subsets employed in exploring the pharmacophoric space of TACE inhibitors, numbers correspond to compounds in figure A and table A. Table C: Training sets and CATALYST run parameters employed for exploring TACE pharmacophoric space. Table D: The performance of the best representatives (cluster centers) of clustered pharmacophore hypotheses generated for TACE. Table E. High-ranking captured hits by kNN-dbcICA model 1 and in vitro bioactivities.

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: <https://doi.org/10.3897/pharmacia.70.e96423.suppl1>