8-2012

# Targeting TRAF6 for Cancer Therapeutical Development

John K. Morrow

TARGETING TRAF6 FOR CANCER THERAPEUTICAL DEVELOPMENT

By

John Kenneth Morrow, B.S.

Approved:

_____
Shuxing Zhang, Ph.D., Supervisory Professor

_____
Bryant Darnay, Ph.D.

_____
John E. Ladbury, Ph.D.

_____
Hui-Kuan Lin, Ph.D.

_____
John McMurray, Ph.D.

Approved:

_____
Dean, The University of Texas
Graduate School of Biomedical Sciences at Houston

# TARGETING TRAF6 FOR CANCER THERAPEUTICAL DEVELOPMENT

A

THESIS

Presented to the Faculty of

The University of Texas

Health Science Center at Houston

and

The University of Texas

M. D. Anderson Cancer Center

Graduate School of Biomedical Sciences

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

by

John Kenneth Morrow, B.S.

Houston, Texas

August, 2012

# RE: request for permission to use text for thesis Current Pharmaceutical Design

AMBREEN IRSHAD - BSP [ambreenirshad@benthamscience.org]

**Sent:** Tuesday, July 24, 2012 3:15 AM
**To:** Morrow,John Kenneth
**Cc:** m.ahmed@benthamscience.org

## *Grant of Permission*

Dear Dr. Morrow:

Thank you for your interest in our copyrighted material, and for requesting permission for its use.

Permission is granted for the following subject to the conditions outlined below:

Computational Prediction of Protein Hot Spot Residues
pp.1255-1265 (11) Authors: John Kenneth Morrow, Shuxing Zhang

To be used in the following manner:

1. Bentham Science Publishers grants you the right to reproduce the material indicated above on a one-time, non-exclusive basis, solely for the purpose described. Permission must be requested separately for any future or additional use.

2. For an article, the copyright notice must be printed on the first page of article or book chapter. For figures, photographs, covers, or tables, the notice may appear with the material, in a footnote, or in the reference list.

Thank you for your patience while your request was being processed. If you wish to contact us further, please use the address below.

Sincerely,

***AMBREEN IRSHAD***

***Permissions & Rights* Manager**
Bentham Science Publishers
Email: permission@benthamscience.org
URL: www.benthamscience.com

——Original Message——
From: Morrow,John Kenneth [mailto:JKMorrow@mdanderson.org]
Sent: Friday, July 20, 2012 10:44 PM
To: support@bentham-direct.org
Subject: request for permission to use text for thesis

Hello, i would like to request permission to use the text of my review article for my thesis:

> Computational Prediction of Protein Hot Spot Residues
> pp.1255-1265 (11) Authors: John Kenneth Morrow, Shuxing Zhang

How would i go about doing this?

Thanks,

John Morrow
Experimental Therapeutics
M.D. Anderson Cancer Center=

This work is dedicated to my family and my father, Dr. Philip Ross Morrow, who passed away from adenocarcinoma of the small intestine. It is to him that I owe everything that I have become, and all that I hope to achieve.

**Acknowledgements**

I would like to thank express my greatest appreciation to my mentor, Professor Shuxing Zhang, for guiding me through my continuing journey. He is an amazing mentor and an inspiration, I am very grateful to have been in his tutelage, and I look forward to continuing to work with him in the future. Many thanks are due to the other committee members. I would like to thank Bryant Darnay, for his help in conducting the experimental testing of my compounds. Also, John McMurray, John Ladbury, and Hui-Kuan Lin are owed thanks for each of their suggestions throughout this process. I also owe thanks to the other lab members, Lu Chen for his help and advice throughout in many different areas, Hoang Tran, for help with GROMACS, Sharangdhar Phatak for keeping me sane, and Lei Du-Cuny, for fostering my abilities early on in the lab.

TARGETING TRAF6 FOR CANCER THERAPEUTICAL DEVELOPMENT

John Kenneth Morrow, B.S.

Advisor: Shuxing Zhang, Ph.D.

Tumor necrosis factor (TNF)-Receptor Associated Factors (TRAFs) are a family of signal transducer proteins. TRAF6 is a unique member of this family in that it is involved in not only the TNF superfamily, but the toll-like receptor (TLR)/IL-1R (TIR) superfamily. The formation of the complex consisting of Receptor Activator of Nuclear Factor κ B (RANK), with its ligand (RANKL) results in the recruitment of TRAF6, which activates NF-κB, JNK and MAP kinase pathways. TRAF6 is critical in signaling with leading to release of various growth factors in bone, and promotes osteoclastogenesis. TRAF6 has also been implicated as an oncogene in lung cancer and as a target in multiple myeloma. In the hopes of developing small molecule inhibitors of the TRAF6-RANK interaction, multiple steps were carried out. Computational prediction of hot spot residues on the protein-protein interaction of TRAF6 and RANK were examined. Three methods were used: Robetta, KFC2, and HotPoint, each of which uses a different methodology to determine if a residue is a hot spot. These hot spot predictions were considered the basis for resolving the binding site for *in silico* high-throughput screening using GOLD and the MyriaScreen database of drug/lead-like compounds. Computationally intensive molecular dynamics simulations highlighted the binding mechanism and TRAF6 structural changes upon hit binding. Compounds identified as hits were verified using a GST-pull down assay, comparing inhibition to a RANK decoy peptide. Since many drugs

fail due to lack of efficacy and toxicity, predictive models for the evaluation of the LD50 and bioavailability of our TRAF6 hits, and these models can be used towards other drugs and small molecule therapeutics as well. Datasets of compounds and their corresponding bioavailability and LD50 values were curated based, and QSAR models were built using molecular descriptors of these compounds using the k-nearest neighbor (k-NN) method, and quality of these models were cross-validated.

**TABLE OF CONTENTS**

**List of Tables** **Page**



**List of Illustrations** **Page**

**Introduction:**

TNF-Receptors Associated Factors (TRAFs) are a family of adapter proteins, originally identified in humans and rodents by their association to the cytoplasmic tails of different members of the TNF-Receptor (TNFR) family [1]. In mammals, TRAFs contribute to the regulation of as many as 20 TNFRs [2-5]. TRAFs are primarily involved in the signals regulating the inflammatory and immune systems, as well as regulating apoptosis. There are seven members (TRAF1 through TRAF7) of the TRAF family, and all but TRAF7 are characterized by a conserved, 180 residue fold called the TRAF domain (TD). It is the TD that supports the interaction with TNFRs and other adapter proteins; as one would expect, TRAF7 does not directly interact with TNFRs [3, 6]. The TRAF domain has an α-helix segment followed by eight anti-parallel β-strands (also called the TRAF-C domain) which fold into a β-sandwich structure, and usually oligomerizes as mushroom-shaped trimers that are stabilized by the coiled-coil interactions between the α-helices of each TD monomer [2, 6, 7]. As for the other structural features in TRAFs, all contain one to seven zinc finger domains [8] and all but TRAF1 have N-terminal RING-finger (Really Interesting New Gene) domain. All mammalian TRAFs localize to the cytoplasm except TRAF4 which is found in the nucleus.

Among the TRAF family members, TRAF6 has distinct features and physiological functions that are not shared in other TRAFs. TRAF6 was first identified by a yeast two-hybrid screen using the cytoplasmic tail of CD40 as bait and independently by expressed sequence tag (EST) screening [9, 10]. Protein sequence and phylogenetic studies revealed that is the older and less conserved TRAF family member despite maintaining all of the structural characteristics of the other members of the

family. TRAF6 shares the smallest amount of homology to the prototypical TRAF domain, and has the most divergent TRAF-C domain [11]. The unique function of TRAF6 is primarily determined by its TRAF-C domain, which does not share the same peptide motif interactions as TRAF1, -2, -3 or -5 [12-14]. TRAF6 binds to a consensus sequence of: xxPxExx(Ac/Ar) (where the last residue is aromatic or acidic), whereas TRAF1, -2, -3, or -5 bind to a (P/S/A/T)x(Q/E)E motif [2, 8, 15-17]. Another distinguishing feature of TRAF6 from other TRAFs is its ability to transduce signals from the TLR/IL-1R superfamily as well as the TNFR superfamily [18]. TRAF6 does not have direct interaction with members of the TLR/IL-1R family. Instead, recruitment of adaptors occur upon its activation, such as the myeloid differentiation factors 88 (MyD88), toll/interleukin-1 receptor adaptor protein (TIRAP), and TRIF-related adaptor molecule (TRAM) [1]. Since TRAF6 is a central hub for a wide variety of signals, it is not surprising that it can regulate an array of physiologic processes, including adaptive and innate immunity, bone metabolism, and the development of structures such as mammary glands, central nervous system, and the skin.

TRAF6 has an N-terminal RING-finger domain, followed by four zinc-finger domains, the coiled-coil domain, and finally the C-terminal TRAF domain (**Figure 1**). Again, this TD is comprised of eight anti-parallel β-strands that form a stable sandwich structure, and are preceded by a single α-helix segment. The unique sequence specificity of TRAF6 does not overlap with other TRAF members; it interacts directly with a subset of the TNFR superfamily, two such examples of which are CD40 and Receptor activator of nuclear factor-kappa B (RANK, also known as TRANCE-R).

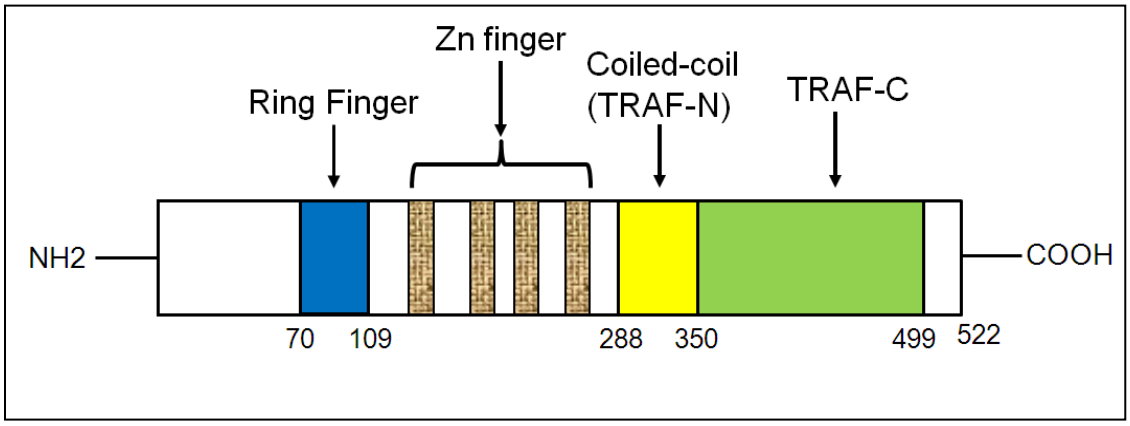



**Figure 1. Domains of the TRAF6 protein.**

**The TRAF6 Protein consists of four parts: a Really Interesting New Gene (RING) finger domain, four zinc finger (Zn finger) domains, a coiled-coil domain, and a C-terminal TRAF-C domain. The RING domain acts as an E3 ubiquitin ligase.**

In addition to the signal transduction of the TNFR superfamily, TRAF6 is also a significant transducer for the interleukin-1 receptor (IL-1R)/Toll-like receptor (TLR) superfamily. Within the IL-1Rs and TLRs resides a component called the TIR domain; it is this domain that recruits a family of TIR-domain containing signaling proteins, including MyD88, Mal/TIRAP, TRIF, and TRAM [19]. As a result, these signaling complexes recruit Serine/Threonine kinases in the IRAK family (IRAK1, IRAK2, IRAK-M and IRAK4), which then interact with TRAF6 to activate downstream signal transduction [20-23].

While TRAF6 mediates signaling in a wide spectrum of cellular physiological functions, the interaction of TRAF6 and RANK are of special interest. The pair of RANK and RANK ligand (RANKL) is essential for bone remodeling by regulation of osteoclasts development and function [24] , mammary gland development [25] and lymph node organogenesis [26, 27]. Also worthy of mention is the fact that the RANK/RANKL pair controls the incidence and onset of progestin-driven breast cancer [28], and plays a significant part in migration and metastatic behavior of cancer cells, acting as a fertile

soil for metastatic tumors going to the bone [29]. When the RANK/RANKL complex is formed, TRAF6 is recruited, which then results in the activation of the NF-κB transcription factor, as well as members of the mitogen-activated protein (MAP) kinase family, which include: MAPK, c-Jun N-terminal kinase (JNK), and p38. TRAF6-deficient mice have a defect in osteoclastogenesis, and develop osteopetrosis as a result [30-32]. TRAF6 is the critical adapter in RANK-mediated osteoclast differentiation, and the TRAF6/RANK interaction is a potential target for inhibition in the treatment of various bone diseases, and might have implications in the areas of breast/lymph node as well.

New evidence from copy number alterations shows that TRAF6 is an oncogene in lung cancer. Overexpression of TRAF6 results in tumor formation and malignant transformation of fibroblasts, and RNA interference (RNAi)-mediated knockdown of TRAF6 decreases adenocarcinoma in two lung cell lines that have *TRAF6* amplification [33]. In these two lung cell lines, RAS required TRAF6 for its oncogenic capabilities. This finding provides an explanation for constitutive NF-κB activation in RAS-driven cases of lung cancer.

TRAF6 has been shown to play a major role in the signal transduction of inflammation, cell survival and proliferation. Down-regulation of TRAF6 is beneficial in a therapeutic setting, since there have been many implications of TRAF6 in different disease states. TRAF6/RANK and their neighboring binding partners have been shown to contain valid, druggable targets in two notable areas: The RANK/RANKL interaction and TRAF6 decoy peptides. The RANK/RANKL portion of this pathway has seen monoclonal antibodies against it in the form of Denosumab, which is an anti-RANKL human monoclonal antibody developed by Amgen for use in the treatment of bone

loss/destruction due to rheumatoid arthritis, or metastatic cancers [34]. It is the first RANKL inhibitor to be approved by the FDA, and it is showing promising results for multiple myeloma with bone metastases [35]. Although germline deficiencies in RANK/RANKL show strong B cell defects and it was suggested that RANKL inhibition could trigger immunologic side effects, B cell specific RANK knockout mice in fact did not show obvious direct defects in B cell physiology or development [36]. The other notable area is that of cell-permeable peptides that bind to TRAF6 (TRAF6 decoy peptides, referred to as T6DP) and have been developed that can target the TRAF6/binding peptide interaction, and these have been shown to prevent RANK signaling [15]. These peptides display specificity for the TRAF6-binding domains, and the core motif (RKIPTEDEY) inhibited RANKL-mediated osteoclastogenesis and bone resorption. RANKL-dependent activation of NF-κB undergoes a dose-dependent inhibition upon pre-treatment with these decoy peptides, and primary mouse monocytes differentiation into functional bone resorbing osteoclasts were blocked as well [37]. While these peptides are effective in blocking the interaction in experimental protocols, their large size and potential sensitivity to endopeptidases renders them unsuitable as clinical therapeutic agents. These decoy peptides might find future use when used in liposomal or nanoparticle delivery techniques.

TRAF6's role and mechanism in signaling in its diverse pathways becomes clear when one examines its ubiquitin ligase activity. Ubiquitin ligases attach a small protein called ubiquitin to target proteins; this ubiquitin can then induce either degradation of the target protein (via the proteasome) or can promote interactions with other proteins that result in signal transduction. Two main factors of ubiquitination allow for the discrimination between degradation and signal transduction: the number of ubiquitin

proteins added (monoubiquitination versus polyubiquitination), and which specific lysine residue that gets modified. Lysine48 (K-48) linked ubiquitination is recognized by the 26S proteasome and this ubiquitination triggers the protein degradation. Lysine63 (K-63) linked ubiquitination is an important post-translational modification that can facilitate various biological processes, including: activation of kinase signaling, endocytosis of cellular receptors, protein localization and trafficking, and repair of damaged DNA [38-40]. On TRAF6, the intact RING domain and the first zinc finger are required for K-63 auto-ubiquitination of TRAF6, and subsequent activation of downstream activation of targets such as: IL-1, LPS, IKK, JNK, NF-κB, and osteoclasts differentiation via RANKL signaling [41-43]. A recent crystal structure has shown the RING domain interacts with the ubiquitin-conjugating enzyme (E2) called Ubc13, and that the first zinc finger has a structural role in binding to Ubc13 [44]. Another significant finding in the ubiquitin pathway is that Akt (the serine/threonine kinase, also called protein kinase B) ubiquitination is triggered by TRAF6 E3 ligase activity. TRAF6 adds ubiquitin to Akt and induces Akt ubiquitination *in vitro* and *in vivo* and is essential for Akt ubiquitination and localization to the membrane where Akt is then phosphorylated and activated [45]. This ubiquitination of Akt occurs via the K-63 pathway and does not trigger Akt degradation. Other kinases that are activated by TRAF6 K-63 ubiquitination are: transforming growth factor-β-activating kinase 1 (TAK1) [46, 47], mixed linage kinase 3 (MLK3) [48], and interleukin-1 receptor-associated kinase (IRAK1) [49].

TRAF6 is a prime drug target candidate to inhibit using rational drug design. To date, there are no known small molecule inhibitors of TRAF6 in the C-terminal region. This is surprising, as the above data clearly show that TRAF6 is a highly significant and valid target for the treatment of bone-related diseases. The discovery of TRAF6 as an

oncogene makes its use as a target even more appealing. The availability of high-resolution crystal structures of the C-terminus of TRAF6 bound to both RANK and CD-40 and the ineffectiveness of peptide-based treatments calls for the use of small molecule inhibitors through structure-based drug design. The effectiveness of potential hits that are discovered by this method can have their effectiveness compared to decoy peptides as a positive control. From the publication describing the crystal structures, mutational analysis was carried out to indicate which residues were required for binding to binding peptides. Three residues on TRAF6 have been shown to abolish binding to CD40 (a close analog to RANK) that shares the core motif as RANK; Arg392, Phe471, and Tyr473 (**Figure 2**); these residues can be considered hot spot residues [15]. These residues are a good starting point for which to carry out structure-based drug design, as small molecule inhibitors can interact with these hot spots and prevent formation of the TRAF6/RANK complex. Using small molecules to inhibit the formation of protein-protein interactions is a relatively recent technique, and hot spots have been used successfully to disrupt and inhibit protein-protein interactions for a multitude of targets [50].

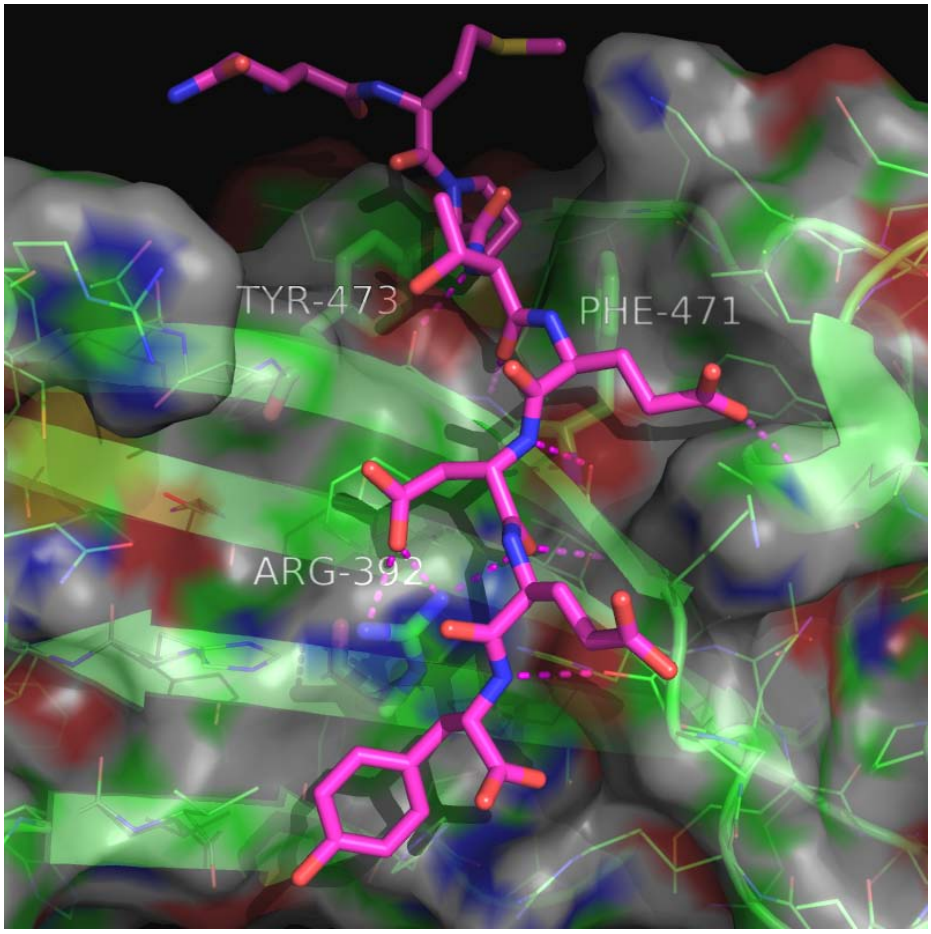



**Figure 2. Structure of TRAF6 (green) bound to RANK peptide (magenta).**

**Residues experimentally determined to be hot spots (Arg392, Phe471, and Tyr473) are displayed as green sticks. Hydrogen bonds between RANK and TRAF6 displayed as dashed magenta lines. Figure generated in PyMOL.**

In the current study, I will outline the steps that will aid in bringing a TRAF6 small molecule inhibitor closer to clinical relevance. In the hopes of developing TRAF6 as a valid target for small molecule therapeutics, computational prediction of hot spots was carried out on the protein-protein interaction of TRAF6 and its binding peptides. Three methods were used: Robetta, KFC2, and HotPoint, each of which uses a different methodology to determine if a residue is a hot spot. These hot spot predictions were considered the basis for resolving the binding site for TRAF6 *in silico* high-throughput screening. Screens on the crystal structure of TRAF6 in complex with RANK will be

carried out using combined docking programs and our database of 10,000 drug/lead-like compounds. Flexibility of side-chains that contribute significantly to the binding will be considered. During hit selection, chemical diversity and drug-like properties will be considered. Computationally intensive molecular dynamics simulations will be conducted using GROMACS to determine the binding mechanism and TRAF6 structural changes upon hit binding. Compounds identified as hits will be verified using a GST-pull down assay, comparing inhibition to a RANK decoy peptide.

Since many drugs fail due to lack of efficacy and toxicity, predictive models for the evaluation of the LD50 and bioavailability of our TRAF6 hits were developed, and these models can be used towards other drugs and small molecule therapeutics as well. Datasets of compounds and their corresponding bioavailability and LD50 values were curated, and predictive QSAR models were built using molecular descriptors of these compounds using the k-nearest neighbor (k-NN) method, and quality of these models were cross-validated using leave-one-out cross validation.

**Methods:**

1. **Computational Hot Spot Predictions of TRAF6**

Since the three hot spots of the TRAF6/binding peptide interaction have been described previously, computational prediction tools were used and compared to the actual results, in an attempt to validate these predictions. Additionally, these predictions can predict residues that are weak hot spots, or calculate potentials of residues that were not biologically tested; these predictions were then used to guide the virtual screening. The computational tools were: Robetta [51], an energy based computational alanine

scanning technique, HotPoint [52], a simple empirical model based on accessible surface area, and KFC2 [53, 54], a knowledge-based and machine learning approach.

## 2. Preparation and Analysis of Target Receptor

The crystal structures of the C-terminal region of TRAF6 in complex with the RANK and CD40 polypeptide were obtained from the Protein Data Bank [55] (1LB5 and 1LB6, respectively) [15]. These crystal structures display residues Gln347 to Thr501 of TRAF6 (the TRAF6 sequence has 522 residues in total). 1LB5 was selected as the target for virtual screening since it has the RANK peptide bound to the TRAF6 structure, and the A chain was analyzed in complex with the B chain (RANK peptide) for virtual screening. Structural water molecules were not considered to be significant to binding, and these water molecules along with the B chain were removed from this 1LB5 structure using PyMOL [56].

## 3. Chemical Library Selection

The dataset of compounds to be screened was the 10,000 compound MyriaScreen Diversity Collection from Sigma-Aldrich in collaboration with TimTec. This was selected for its high diversity and good drug-like and lead-like properties, as well as the commercial availability of high-purity compounds. MyriaScreen was assembled from an original pool of 300,000 Sigma-Aldrich and TimTec compounds, and TimTec's proprietary software was used to filter this original pool on the basis of diversity. Then, additional filters were set to consider MW (>225 and <600), cLogP, H-acceptors, H-donors, and rotatable bonds. 99.04% of compounds in the MyriaScreen collection satisfy all four rules of the Lipinsky rule of 5 [57].

The MyriaScreen chemical dataset (in .sdf file format) was minimized with root mean square gradient of 1 $x10^{-6}$ in MOE software from Chemical Computing Group [58] based on MMFF94x (Merck molecular force field 94x) [59] and partial charges. The MMFF94x force field is quite suitable for small molecules [60]. Minimization is required to reach optimal energy levels for ligands before submitting them to virtual screening.

## 4. Virtual Screening

The MyriaScreen dataset was docked using GOLD from Cambridge Crystallographic Data Centre (CCDC) [61, 62]. Hydrogen atoms were added using GOLD program before docking was performed. The protein was treated as rigid, with no amino acid side chains considered flexible. The binding site residues were manually selected, and formed the cavity file. The residues selected were based on their proximity to the RANK peptide co-crystallized with TRAF6: Arg392, Phe410, Met450, Leu456, Leu457, Ala458, Phe459, Pro468, Lys469, Gly470, Phe471, Gly472, Tyr473, and Val474. Ten genetic algorithm runs were selected for each ligand, and search efficiency was set at 200% (double the default efficiency). Early termination of the docking of a given ligand was performed (meaning GOLD advanced to the next ligand) if the top 3 solutions/conformations of a given ligand were within 1.5Å of each other. Although all scores were kept, only the top 2000 GoldScore solutions/conformations were written out as an SDF file.

## 5. Processing of Results

The results from the docking were then clustered based on structural MACCS (Molecular ACCess System) fingerprints using MOE fingerprint clustering. The top 2000 GoldScored compounds were entered into MOE, and their molecular fingerprints were

calculated. The metric used was the Tanimoto coefficient. Similarity and overlap were set at 80% to cluster similar compounds together. The result of this clustering was that compounds that shared 80% of their total molecular features with other compounds were merged into one cluster. For each cluster, the compound with the highest GoldScore in that cluster was selected as the representative for that cluster, and the others in that cluster were discarded. This clustering output contained 1619 compounds, indicating that 381 compounds had at least 80% similar molecular features but lower GoldScores than other compounds in a single cluster, and these 381 were discarded.

### 6. Selection of Compounds for Biological Testing

From these 1,619 ligands, the top 300 ligands were selected for individual molecular visualization and consideration. Each of the 300 ligands were individually visualized (using PyMOL) in their docked conformation on TRAF6, and compounds were selected for purchase and experimental testing via both pull-down and fluorescence polarization assays.

Three main criterions were considered sequentially in this final selection process. The first criterion was the degree to which the ligands' conformation occupied the interaction site between TRAF6 and RANK. If the ligands' conformation primarily existed outside of the interaction site, it was no longer considered a candidate. Second, the following qualities: conformation, binding affinity, and resulting "fit" within the TRAF6 binding site (compared to the corresponding RANK residues) were considered. Each ligand, given its orientation in the docked position, was evaluated on its ability to potentially block the binding of the RANK peptide. Along with these potential blocking effects, we also considered the potential binding affinity of the ligand in its pocket. This

included qualities such as hydrogen bonds, static and aromatic effects that could be formed from the ligand and TRAF6. The third and final consideration was the distinctiveness of the molecular scaffold when compared to other ligands in the result. The ligands were partitioned into “classes” based upon the individuality of their underlying scaffold and how this scaffold was oriented in the binding site. Ligands that shared a similar scaffold and orientation were considered to be in the same class, and would then share the same class number. We then assigned a “priority” to each molecule which was an aggregate of the second and third factors (conformation in the pocket and distinctiveness). This priority was listed as: low, medium or high. The priority then became similar to a degree of confidence for each ligand; compounds with “high” priority were regarded as most likely to yield good biological activity, and should be tested over the other low and medium confidence ligands.

## 7. Biological Testing

To test potential hits for their ability to inhibit TRAF6 functions, a GST-RANK competitive inhibition assay was used. The procedure outlined here was provided by and performed primarily by Professor Bryant Darnay from the department of Experimental Therapeutics in MD Anderson Cancer Center. Human embryonic kidney (HEK) 293 cells that stably express FLAG-tagged TRAF6 were harvested and lysed in lysis buffer (20 mM Tris, pH 7.4, 250 mM NaCl, 1 mM DTT, 1 mM sodium orthovanadate, 2 mM EDTA, 1% Triton X-100, 2 µg/ml leupeptin, and 2 µg/ml aprotinin 20 mM) for 30 min on ice. The cell lysate was centrifuged for 15 min and the supernatant was collected and protein estimated. Equal amounts (150-200 µg) of protein lysate in binding buffer (20 mM Tris, pH 7.4, 150 mM NaCl, 1 mM dithiothreitol, 0.1% Nonidet P-40 and 2 mM

EDTA) were mixed with the indicated compounds or TRAF6 decoy peptide 3 (T6DP3) and allowed to rotate for 1 h at 4□C, after which the samples were centrifuged for 5 min. The supernatant was then mixed with bacterial purified GST-RANK-(340-358) bound to glutathione-agarose beads for 2 h at 4°C with end-over-end rotation. The samples were washed three times in binding buffer and two times in low salt buffer (20 mM TRIS pH 7.4, 25 mM NaCl, and 1 mM DTT). Bound proteins were then eluted in SDS-sample buffer and boiled for 5 min, subjected to SDS-PAGE, and immunoblotted with anti-FLAG [37].

**8. Molecular Dynamics simulations**

In order to describe the interactions of the hits that were identified in the screening, I performed molecular dynamics (MD) simulations of the three hits in complex with TRAF6, as well as a simulation of the TRAF6/RANK complex as a comparison. GROMACS (version 4.5.3) was the software used, and explicit water was used for these simulations [63, 64]. The force field selected was the GROMOS96 43a1 official distribution [65]. A cubic water box was created for the protein-ligand complex, and the distance between the solute and the box was set at 0.9 Å. Charges were added to neutralize the system to a formal charge of zero, and then the system was minimized when the maximum force of the system was less than 10.0 kJ/mol. PME (Particle mesh Ewald) algorithm was used for the long-range electrostatics for all simulations. For the NVT (moles, volume, temperature) and NPT (moles, pressure, temperature) equilibration the protein and ligand were considered as one entity (coupled together), and the ions and water were also coupled together as one entity. This coupling was necessary, as the system exploded (became highly unstable) not long after production MD started when

this coupling was not present. Temperature was set at 298K, and the modified Berendsen thermostat and Parrinello-Rahman were used for NVT and NPT coupling, respectively. Periodic boundary conditions were set for x, y, and z directions. For each of these production MD simulations, the bound conformations from GOLD were used as the starting input, and the PRODRG server was used to generate topology (.itp) files for each of the ligands [66]. All simulations were run for 20 nanoseconds. Jobs were run by remote SSH terminal from my workstation into the Lonestar computing cluster in the Texas Advanced Computing Center (TACC), using 144 cores [67]. Results were extracted using VMD, and were analyzed using PyMOL [56, 68].

## 9. $LD_{50}$ Predictive QSAR Modeling

A dataset of 7385 compounds and $LD_{50}$ values for oral rat exposure was collected from a previous study by Hao Zhu *et al.* who were attempting to achieve the same endpoint; the details of how that dataset was collected are described in the following section [69]. A dataset of more than 8000 compounds and corresponding $LD_{50}$ (oral rat exposure) from the ChemID plus database was collected, and the structures of these compounds was verified using an method discussed by Young *et al*. [70]. Zhu's group removed the inorganic, organometallic compounds, salts, and compound mixtures. They also converted the $LD_{50}$ values from mol/kg to log(mol/kg) values according to standard QSAR practices [69]. Various descriptors were calculated including those from the Dragon software v5.4 [71], MOE [58], and cxcalc from ChemAxon [72]. These descriptors were then used with the kNN-QSAR method to provide the models [73]. The kNN-QSAR method was used to generate the predictive models.

### 10. Bioavailability Predictive QSAR Modeling

A dataset of 809 compounds was collected from two publications: 768 compounds from Hou *et al.* and from Veber *et al.*, these two datasets were merged despite having many similar entries [74, 75]. The values were often taken from the literature; single values were recorded as-is, whereas when a range of values was reported, a simple average was used as the percent bioavailability [76]. For example, if a value was reported as 0-20% bioavailability in Goodman and Gilman's, the value used was considered 10%. Descriptors for this data set were calculated in an identical manner to the $LD_{50}$ data set, and the kNN-QSAR method was also used to generate the predictive models.

## Results:

### 1. Computational Hot Spot Predictions of TRAF6 - Results

Robetta uses a simple physical model and a computational alanine scanning technique that includes various energy and chemical bonding parameters such as: Lennard Jones interactions, packing and solvation terms, and hydrogen bonding to calculate free energy [51]. Computational alanine scanning involves mutation of a given residue involved in a protein-protein interaction to alanine (a non-reactive and non charged side chain) of one of the interacting proteins, and computing the resulting change in binding energy of the interface. If this mutation results in a significant decrease in the binding constant (typically tenfold or greater), this residue is considered a hot spot [77]. Robetta mutates the residue to alanine for each residue, and then it locally repacks other residues in the structure within 5Å of the mutant residue, maintaining the conformation of

the remaining protein residues. The predicted changes in binding energies from this computational alanine scanning constitute the foundation for Robetta's hot spot predictions [78]. Robetta provided the most accurate (indeed a perfect score) of the hot spots on TRAF6, with all three experimental hot spot residues in the structure having a predicted ΔΔG greater than 1 kcal/mol (**Table 1**). The Arg392 position is predicted to be a very active hot spot with a ΔΔG value much higher than the other two correctly predicted hot spots; from this prediction, one can infer that rational design around this residue is likely to give a better inhibitor of the TRAF6/RANK complex, since it contributes the most to binding of the RANK peptide. Values that are positive (or more positive) for the ΔΔG of the TRAF6/RANK complex indicate that a destabilization of the complex occurs when a residue is replaced by alanine; the other predictions do not show such a change in energy; hence they are not predicted to be hot spots. It is interesting to note that the next highest (and closest to 1.0 kcal/mol) ΔΔG value predicted by Robetta is the Phe410, with a value of 0.76 kcal/mol. This residue was later shown to be important to binding by the molecular dynamics results.

| Num | Residue | int_id | ΔΔG(complex) | ΔG(partner) |
|---|---|---|---|---|
| 374 | Val | 1 | 0.36 | 1.04 |
| 376 | His | 1 | 0.44 | 0.32 |
| 392 | Arg | 1 | 3.76 | 1.51 |
| 410 | Phe | 1 | 0.76 | 3.21 |
| 412 | His | 1 | 0.4 | 2.61 |
| 448 | Glu | 0 | -0.03 | -0.12 |
| 450 | Met | 1 | 0.13 | 2.38 |
| 453 | Lys | 0 | -0.12 | 0.08 |
| 456 | Leu | 1 | 0.48 | 1.34 |
| 466 | Arg | 1 | 0.07 | 0.94 |
| 469 | Lys | 1 | 0.62 | -0.14 |
| 471 | Phe | 1 | 2.05 | 1.65 |
| 473 | Tyr | 1 | 1.73 | 1.8 |
| 474 | Val | 0 | 0.07 | 0.43 |

**Table 1. TRAF6 Hot Spot predictions from Robetta.**

**Num: residue number. Residue: three-letter code for amino acid. int_id: Binary descriptor of if a side chain of a residue is present within 4 Å of another partner's atom (1) or not having a contact directly, but is buried upon binding (0); ΔΔG(complex): Prediction of binding free energy change upon mutation of alanine; ΔG(partner): Predicted change in stability of protein of the corresponding mutated partner complex upon subsequent alanine mutation. Residues highlighted yellow are biologically proven hot spots. If ΔΔG(complex) is greater than 1.0 kcal/mol, they are considered a hot spot by Robetta.**

KFC2 did not have as successful of a prediction as Robetta. KFC2 is comprised of two models, KFC2a and KFCb, each having their own parameters. KFC2a offered no predictions for this structure, and KFC2b was adequate in providing correct hot spot predictions of Arg392 and Phe471, but provided two false positive results of Phe410 and His412. KFC2b also did not label Tyr473. Both KFC2 models label a residue a hot spot when the confidence of either method is a non-negative number (**Table 2**).

Hotpoint's predictions were identical (in terms of false positives and false negatives) to those from KFC2, predicting Arg392 and Phe471 correctly, but not labeling the residue Tyr473 and providing a false prediction of Phe410 as a hot spot (**Table 3**). The potential value on Arg392 was the highest among all of the residues that were considered interacting. Hotpoint did not predict His412 to be a hot spot, but did predict Met450 as one. This Met450 residue lies close to the aromatic residues of Phe471 and Tyr473, and is significantly buried.

| Num | Res | KFC2-A Class | KFC20-A Conf | KFC2-B Class | KFC2-B Conf |
|---|---|---|---|---|---|
| 374 | Val | 0 | -1.78 | 0 | -0.99 |
| 376 | His | 0 | -2.18 | 0 | -0.98 |
| 392 | Arg | 0 | -0.37 | Hotspot | 0.15 |
| 394 | His | 0 | -1.9 | 0 | -0.95 |
| 410 | Phe | 0 | -0.43 | Hotspot | 0.23 |
| 412 | His | 0 | -0.63 | Hotspot | 0.05 |
| 448 | Glu | 0 | -2.31 | 0 | -0.98 |
| 449 | Ile | 0 | -2.39 | 0 | -0.92 |
| 450 | Met | 0 | -1.42 | 0 | -0.26 |
| 451 | Asp | 0 | -2.32 | 0 | -0.96 |
| 453 | Lys | 0 | -1.77 | 0 | -0.96 |
| 456 | Leu | 0 | -1.25 | 0 | -0.83 |
| 457 | Leu | 0 | -2.37 | 0 | -0.94 |
| 458 | Ala | 0 | -0.84 | 0 | -0.76 |
| 466 | Arg | 0 | -1.8 | 0 | -0.79 |
| 468 | Pro | 0 | -0.88 | 0 | -0.8 |
| 469 | Lys | 0 | -0.79 | 0 | -0.56 |
| 470 | Gly | 0 | -0.22 | 0 | -0.46 |
| 471 | Phe | 0 | -0.17 | Hotspot | 0.28 |
| 472 | Gly | 0 | -0.38 | 0 | -0.59 |
| 473 | Tyr | 0 | -0.87 | 0 | -0.04 |
| 474 | Val | 0 | -1.24 | 0 | -0.95 |
| 475 | Thr | 0 | -1.83 | 0 | -0.99 |

**Table 2. KFC2 prediction of hot spots on TRAF6.**

**Num: residue number. Res: three-letter code for amino acid. Model classification and predictions of two KFC models (A or B model); (Conf): Confidence of prediction. Residues highlighted in yellow (Arg392, Phe471, Tyr473) are experimentally proven hot spots.**

| Num | Residue | RelCompASA | RelMonomerASA | Potential | Prediction |
|---|---|---|---|---|---|
| 374 | Val | 36.17 | 50.36 | 12.38 | 0 |
| 376 | His | 29.78 | 35.89 | 18.13 | 0 |
| 392 | Arg | 1.34 | 22.39 | 25.64 | Hotspot |
| 410 | Phe | 1.94 | 11.86 | 28.55 | Hotspot |
| 450 | Met | 1.76 | 6.09 | 33.49 | Hotspot |
| 456 | Leu | 21.77 | 35.29 | 11 | 0 |
| 457 | Leu | 61.09 | 62.37 | 7.59 | 0 |
| 458 | Ala | 2.41 | 5.39 | 13.8 | 0 |
| 468 | Pro | 37.32 | 74.16 | 4.79 | 0 |
| 469 | Lys | 25.35 | 53.96 | 8.9 | 0 |
| 470 | Gly | 0.19 | 29.14 | 12.76 | 0 |
| 471 | Phe | 1.22 | 22.9 | 37.55 | Hotspot |
| 472 | Gly | 0.45 | 35.78 | 16.16 | 0 |
| 473 | Tyr | 7.52 | 27.91 | 17.22 | 0 |

**Table 3. Predictions of TRAF6 hot spots from Hotpoint.**

**RelComp ASA: Relative ASA of complex structure. RelMonomer ASA: Monomer Relative ASA. Residues in yellow (Arg392, Phe471, Tyr473) are experimentally proven hot spots.**

KFC2 and Hotpoint both were lacking in their prediction of Tyr473 as a hot spot, despite both methods having values that were close to the cutoff values to change the prediction form non-hot spot to hot spot. The confidence of Tyr473's prediction in the KFC2b model very close to the threshold; this residue had the smallest negative value among the other residues in the list. The only reason Hotpoint's prediction did not mention Tyr473 as a hot spot was due to the potential factor; the value was not greater than 18.0. The false positives from Hotpoint (residues Met450 and Phe410) stem from the relative high potential, and not from the accessible surface area metric. As was mentioned before, Phe410 and Met450 were in fact tested at the same time as the other residues, and were not proven to be hot spots [15]. This conclusion does not mean that

they have no contribution to the complex, but that they do not meet the experimental cutoff value of a hot spot. Since both programs scored the Phe410 and Met450 residue very close to a hot spot, both these residues should not be ignored as they will likely contribute to the binding, albeit not significantly (at least to the RANK peptide).

## 2. Virtual Screening

In GOLD, the default scoring function is the GoldScore; this is comprised of four components from the equation:

$$f = S_{hb\_ext} + S_{vdw\_ext} + S_{hb\_\mathrm{int}} + S_{vdw\_\mathrm{int}}$$

Where $S_{hb\_ext}$ is the protein-ligand hydrogen bonding score and $S_{hb\_int}$ is the internal hydrogen bonding of the ligand. $S_{vdw\_ext}$ and $S_{vdw\_int}$ are the scores arising from weak external and internal Van der Waals forces, respectively. The scores from the virtual screening approached an extreme value distribution **(Figure 3)**

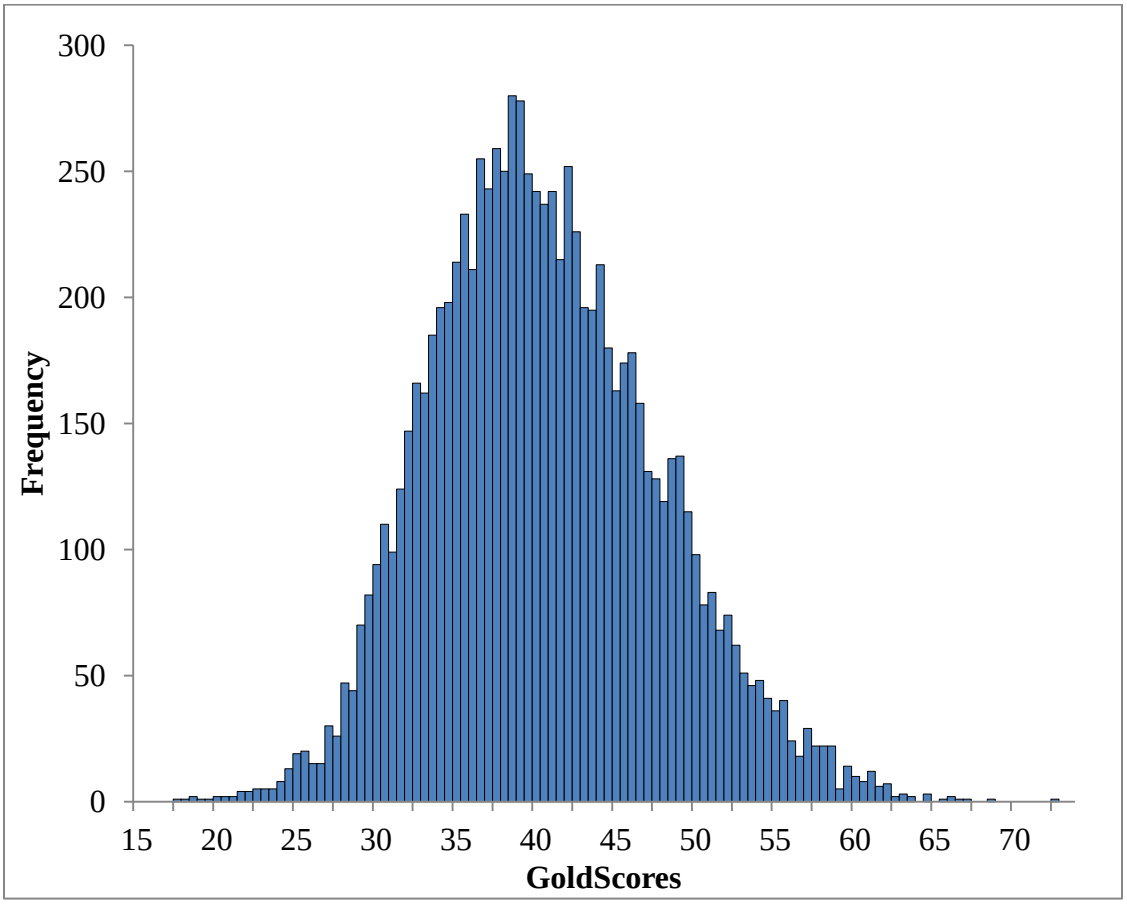



**Figure 3. Histogram of TRAF6 virtual screening results (GoldScores) from GOLD. The highest GoldScore was 75.18 and the lowest was 18.31. The average was 40.30 for the GoldScore.**

### 3. Selection of Compounds for Biological Testing, and Biological testing

The technique described above resulted in 300 compounds for consideration for biological testing. The first round of selected compounds was limited in the number of compounds that were to be tested, and only seven compounds were selected for testing **(Table 4)**. These first seven tested compounds were found to be not active when tested at 100µM and 300µM **(Figure 4)**.

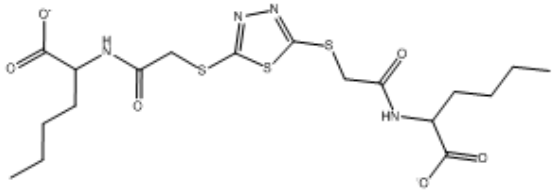
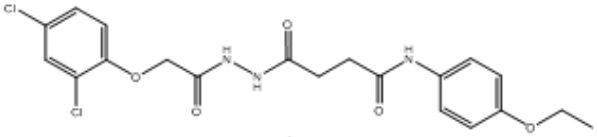
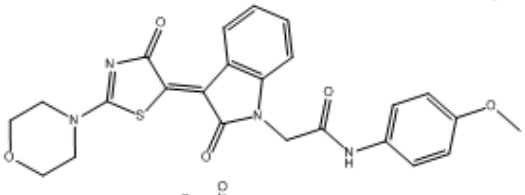
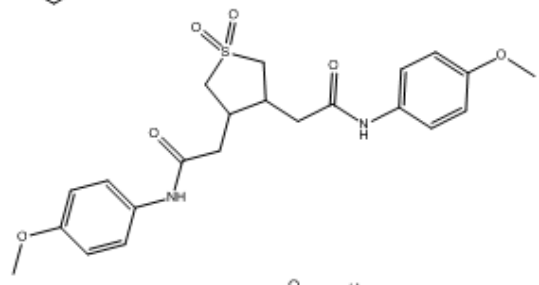
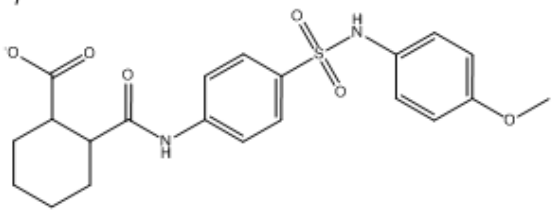
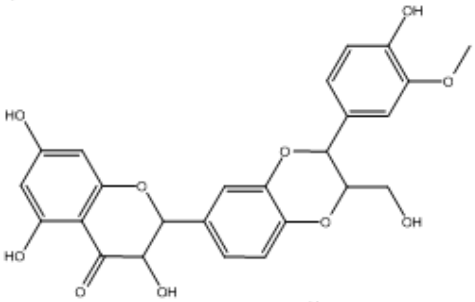
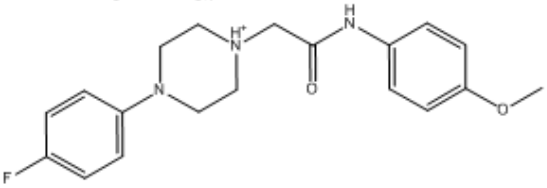
| Test ID | Structure | GOLD | logP |
|---|---|---|---|
| SZB-2 |  | 68.38 | 3.96 |
| SZB-5 |  | 63.45 | 2.99 |
| SZB-6 |  | 62.87 | 2.71 |
| SZB-9 |  | 60.27 | 1.32 |
| SZB-13 |  | 58.34 | 2.71 |
| SZB-15 |  | 58.17 | 2.64 |
| SZB-18 |  | 57.62 | 2.24 |

**Table 4. First set of selected compounds for biological testing.**

Test ID values were assigned from decreasing GoldScore from the list of 79 compounds. GOLD: GoldScore output from GOLD. logP: calculated log of octanol/water partition coefficient. All seven of these were shown to be inactive via GST pull-down assay.

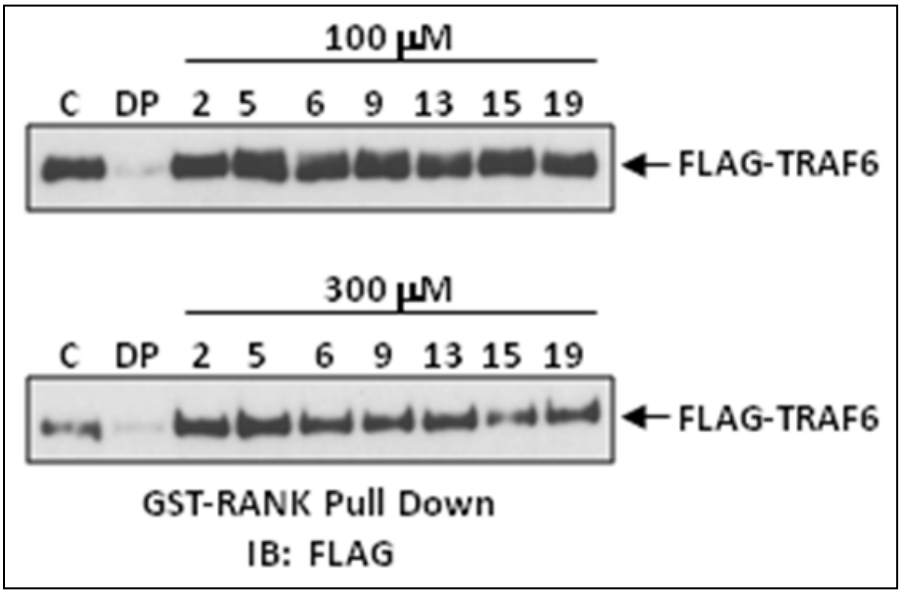



**Figure 4. GST-RANK pull down assay for first round of seven selected compounds. DP is the positive control decoy peptide, and C is the negative control (no peptide).**

After discovery that the first set of compounds were not active, the docking results were examined once again, this time with emphasis on the Arg392/Phe410 binding site. The reasoning was that it was this region (not the Phe471/Tyr473 region) that might be more significant to the binding of compounds to TRAF6. The second round of compounds were taken from the pool of 79 compounds, but compounds to be selected for the second round were those compounds that had a more favorable binding to this region. Twenty compounds were selected for the second round of testing (**Table 5)**, **(Table 6)**, and **(Table 7)** using the same technique as the first round of compounds, but were to only be tested at 200µM (**Figure 5)**. While the concentration of the ligands was high, there does seem to be some reduction in the binding of the compounds SZB-40, SZB-45, and SZB-46.

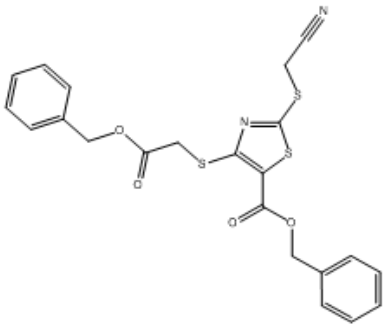
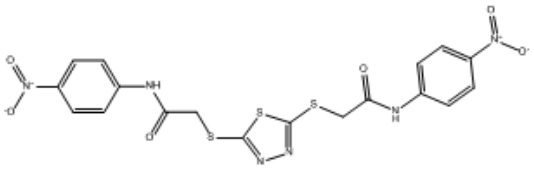
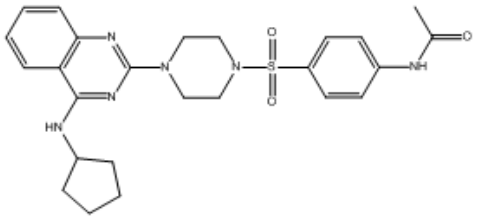
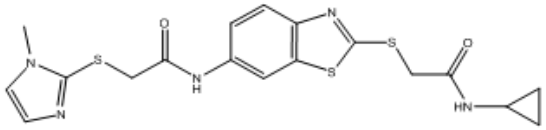
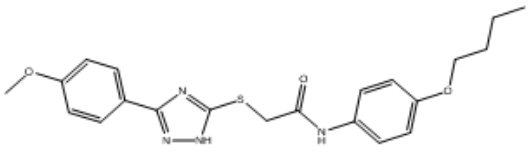
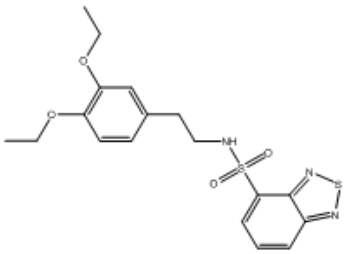
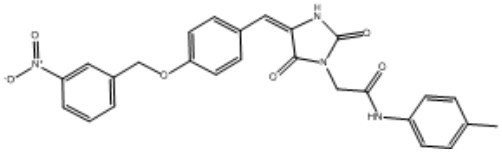
| Test ID | Structure | GOLD | LogP |
|---|---|---|---|
| SZB-1 |  | 72.18 | 5.13 |
| SZB-4 |  | 65.35 | 3.97 |
| SZB-11 |  | 59.03 | 2.13 |
| SZB-12 |  | 58.49 | 2.81 |
| SZB-14 |  | 58.27 | 4.43 |
| SZB-16 |  | 58.01 | 3.539 |
| SZB-17 |  | 57.95 | 4.23 |

**Table 5. Second set of compounds to be tested.**

**GOLD: GoldScore output from GOLD. logP: calculated log octanol/water partition coefficient. Test ID: assigned names from list of 79 compounds.**

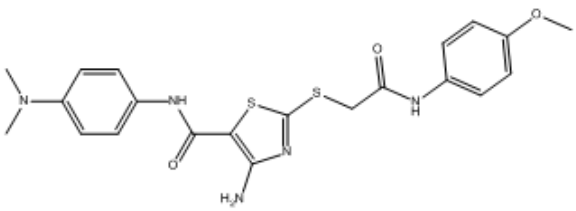
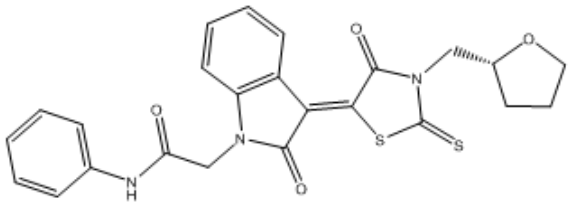
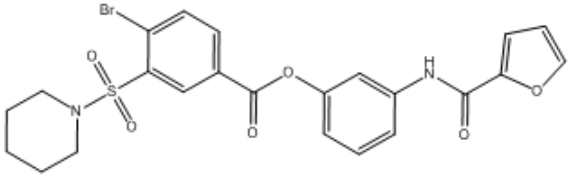
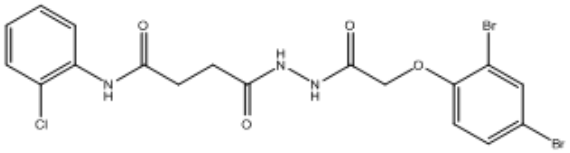
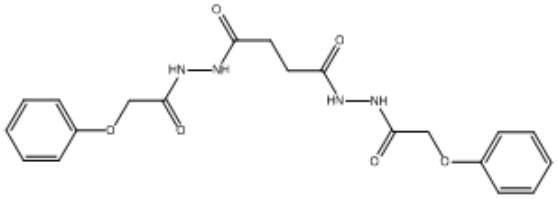
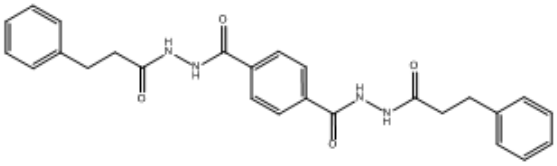
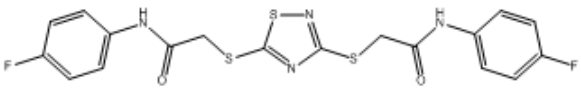
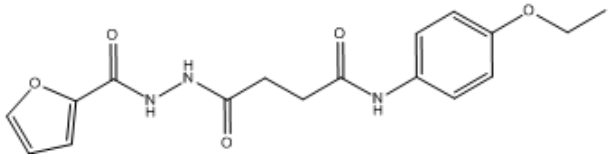
| Test ID | Structure | GOLD | LogP |
|---|---|---|---|
| SZB-19 |  | 57.18 | 2.69 |
| SZB-25 |  | 55.49 | 2.91 |
| SZB-26 |  | 63.23 | 3.76 |
| SZB-30 |  | 60.79 | 3.69 |
| SZB-31 |  | 60.30 | 0.91 |
| SZB-32 |  | 60.22 | 3.48 |
| SZB-34 |  | 59.74 | 3.89 |
| SZB-35 |  | 59.49 | 0.70 |

**Table 6. Second set of biologically tested compounds (continued).**

**GOLD: GoldScore output from GOLD. logP: calculated log octanol/water partition coefficient. Test ID: assigned names from list of 79 compounds.**

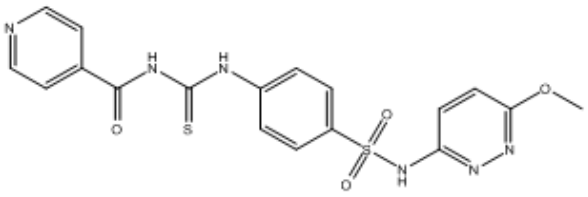
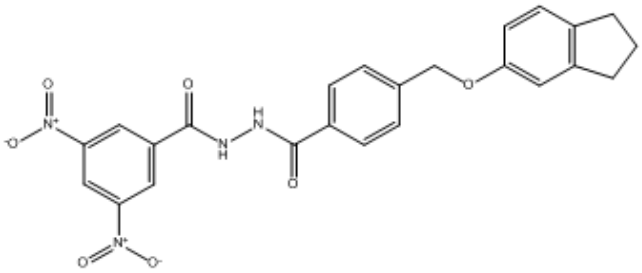
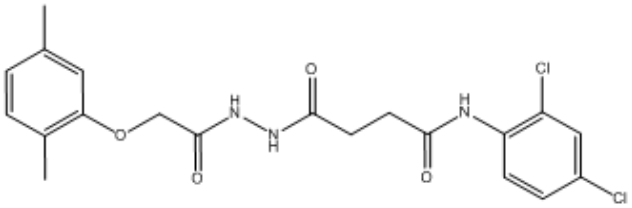
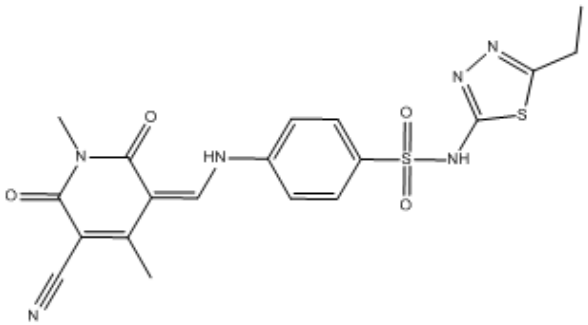
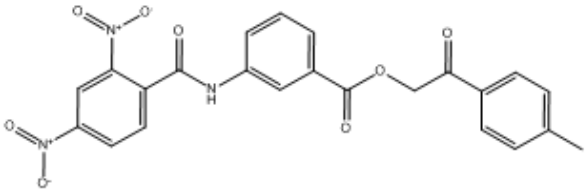
| Test ID | Structure | GOLD | LogP |
|---|---|---|---|
| SZB-36 |  | 59.46 | 2.04 |
| SZB-39 |  | 57.73 | 4.89 |
| SZB-40 |  | 57.45 | 3.32 |
| SZB-45 |  | 56.05 | 1.45 |
| SZB-46 |  | 55.45 | 4.38 |

**Table 7: Second set of biologically tested compounds (continued).**

**GOLD: GoldScore output from GOLD. logP: calculated log octanol/water partition coefficient. Test ID: assigned names from list of 79 compounds. SZB-40, SZB-45, and SZB-46 showed some activity.**

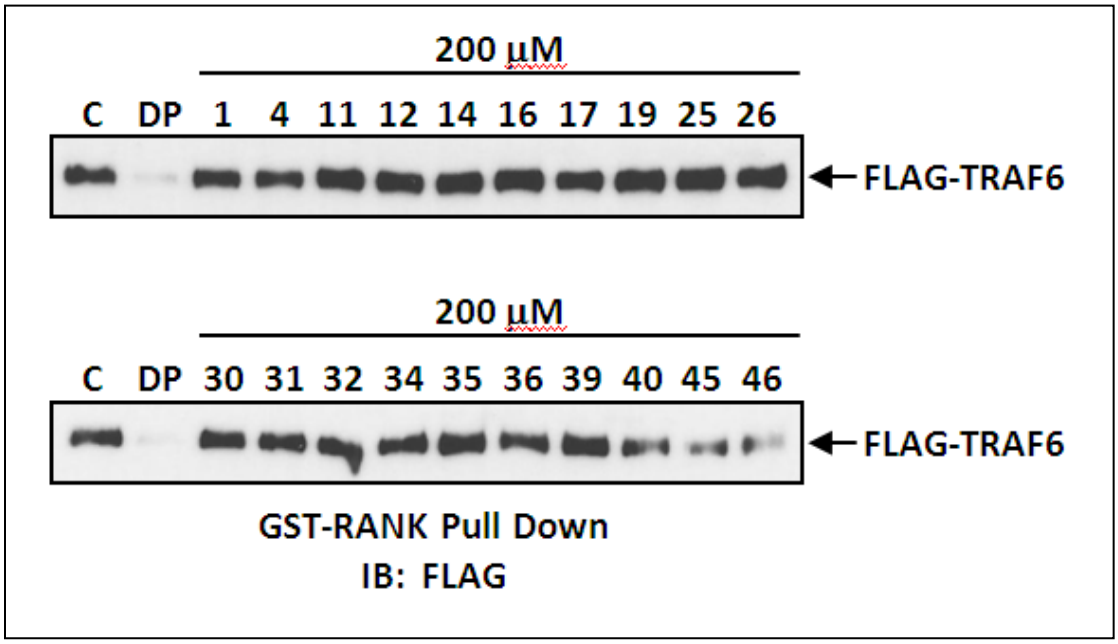



**Figure 5. GST-RANK pull down assay for first round of selected compounds.**

**DP is the positive control decoy peptide, and C is the negative control (no peptide). Compounds 40, 45, and 46 appeared to have some activity.**

### Molecular Dynamics simulations

GROMACS simulations were run for 20 nanoseconds compound SZB-40, SZB-46, and for the RANK peptide (as a control and comparison). SZB-40, which appeared to have the best binding score from GOLD, did not have as significant of a binding energy when compared with the RANK peptide **(Figure 6)**. This figure shows the short range Lennard-Jones energy of the protein-ligand complex. The average energy of the run of SZB-40 was -177kJ/mol, while RANK had energy of -257 kJ/mol; standard deviations were 20.86 and 28.10 for SZB-40 and RANK, respectively.

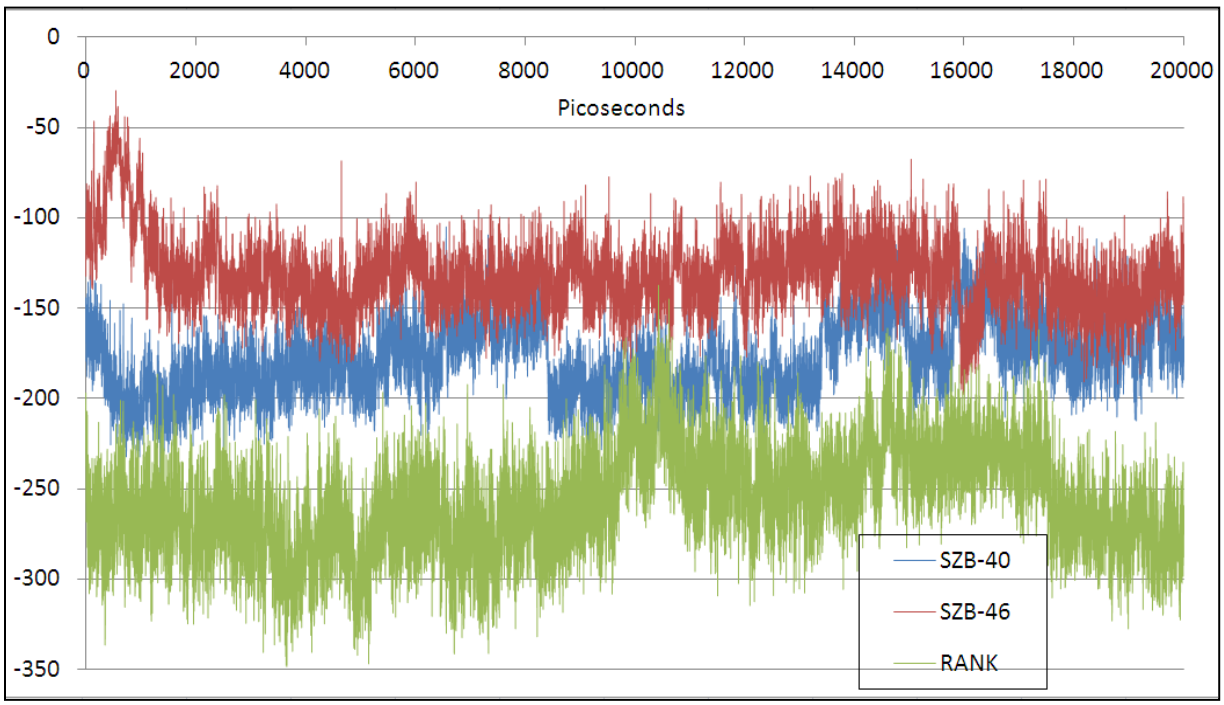



**Figure 6. Lennard-Jones energies for molecular dynamics simulations.**

**TRAF6-ligand complex (kJ/mol) of SZB-40 (blue), SZB-46 (red), and RANK peptide (green) for the entire 20 nanosecond molecular dynamics simulation. Output of energies was performed for each trajectory, and the output was graphed in Excel.**

The two simulations produced roughly similar RMSD values, the control RANK simulation deviating at an average of 0.239Å from their main chain, and the SZB-40 RMSD equal to 0.196Å, indicating that the overall stability of the structure was not significantly affected by the ligands. The standard deviation of main chain of the structures was 0.0399 Å for RANK, 0.0288 Å for SZB-40.

The region of the 9,000 to 13,000 picoseconds (9 to 13 nanoseconds) is interesting, as it shows better energy values. Output of trajectories at every 2,000 picoseconds was carried out for the entire simulation, in order to evaluate the differences that occur between SZB-40 and the TRAF6 structure throughout the MD run. The trajectory at 10 nanoseconds shows a very interesting conformation **(Figure 7)**. It appears that the ligand has achieved pi-stacking with the aromatic residues Phe470, and there is

also a pi-stacking interaction to the Phe410 residue. There are hydrogen bonds to two points on the main chain of Gly469, similar to that of RANK. It is interesting to note that many portions of the MD simulation run showed the ligand on the side of TRAF6 closer to the Phe471 residue. This conformation was not possible in the GOLD screening/docking, as only the side chains are flexible; the GROMACS MD simulation shows a more favorable position to those that were seen from the GOLD docking.

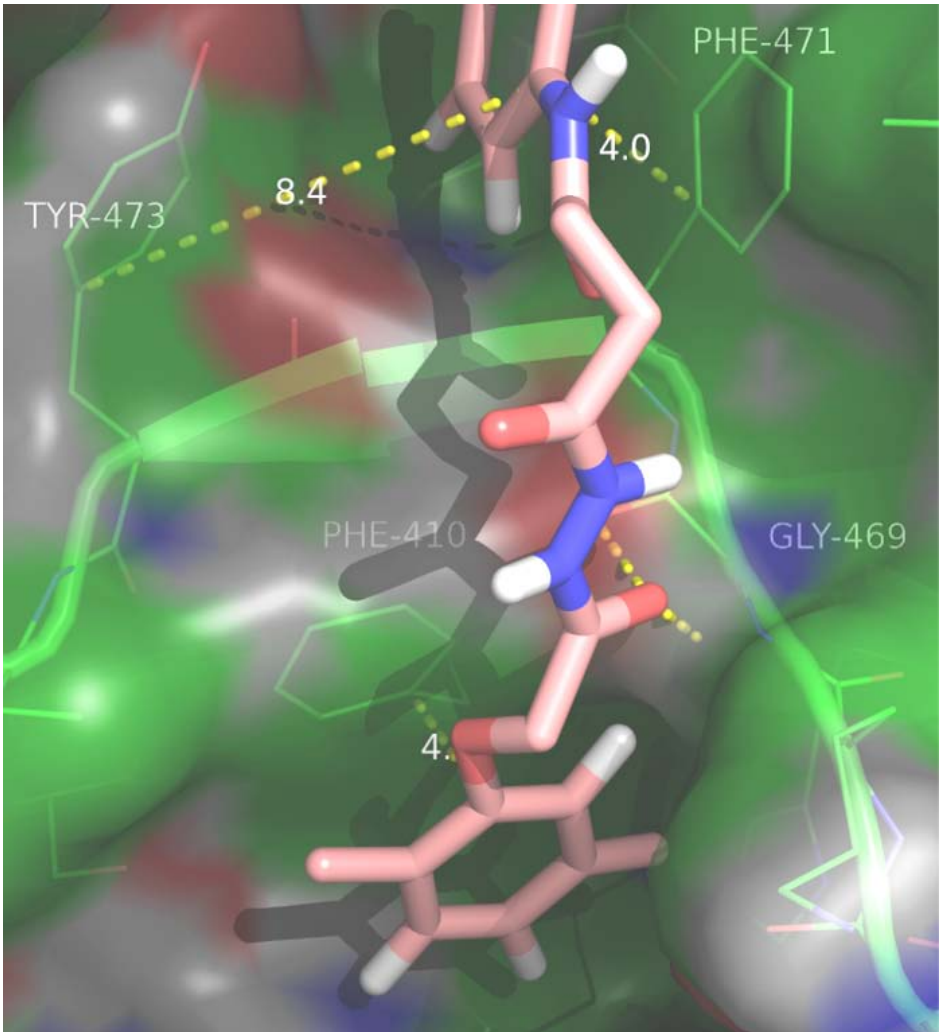



**Figure 7. Sample trajectory conformation at mid-point (10 ns) of MD simulation of SZB-40 (pink) and TRAF6 (green). Yellow dashed lines with numerical labels denote distances to neighboring aromatic groups. Hydrogen bonds to main chain atoms on Gly469 to SZB-40 are shown in unlabeled yellow dashed lines. This conformation was close to the average structure over the 9-11 nanosecond range, and it shows a more favorable pose to docked positions, with pi-stacking of both aromatic ends of SZB-40 to other aromatic groups on TRAF6. The distance from SZB-40 to Phe410 is 4.0 angstroms, and Phe410 was predicted as a hot spot by HotPoint and KFC2.**

Another sample output was taken at the 12,000 picoseconds (12 nanoseconds), which is also in the region of lower energy. One would assume that this lower energy state would correspond to a more favorable docked position similar to that of the 10 nanosecond point, and this is indeed the case **(Figure 8.)**. This is similar to that of the 10 nanosecond point, but there is additional hydrogen bonding to Gly471 (which lies in

between Tyr472 and Phe470. This state should have an even tighter binding to TRAF6, and this might explain why SZB-40 was shown to be active.

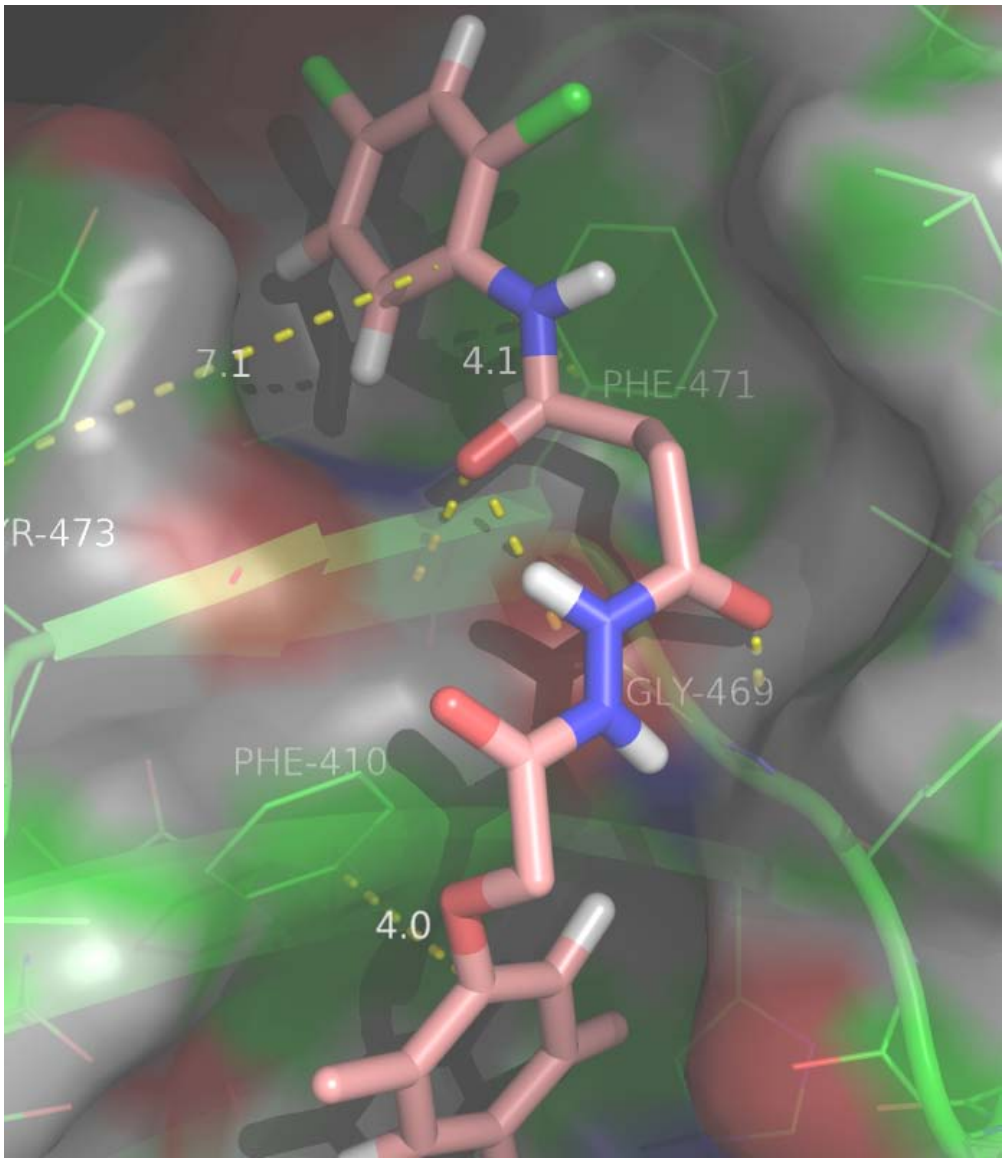



**Figure 8. 12 nanosecond sample trajectory of SZB-40 (pink) and TRAF6 (green). Yellow dashed lines denote distances to neighboring aromatic groups. Hydrogen bonds are shown in yellow dashed lines without distance labels. Note the more extensive hydrogen bonding to both Gly469 but also to main chain atoms on Phe471. This interaction is also more favorable to docked positions seen from GOLD.**

## 4. $LD_{50}$ Predictive QSAR Modeling

A histogram of the compounds shows a good distribution of values, and approaches an extreme value distribution **(Figure 9)**. After the kNN-QSAR models were generated at various parameters, the predictive model with the best cross validated $r^2$ ($q^2$) had eight descriptors and were as follows: logPWeighted, smallestRingSize, acceptorCount, ASAPolar, topologicalPolarSurfaceArea, chainAtomCount, atomCount, and hararyIndex. These descriptors are primarily associated with size, logP, and accessible surface area. The number of nearest neighbors for this highest rated model was three (k = 3), and this model had a $q^2$ value of 0.324 in the training set, and a corresponding $r^2$ of 0.6252 for the 385 compounds in the test set. The distribution of the actual versus predicted values can be seen when plotted (**Figure 10**).

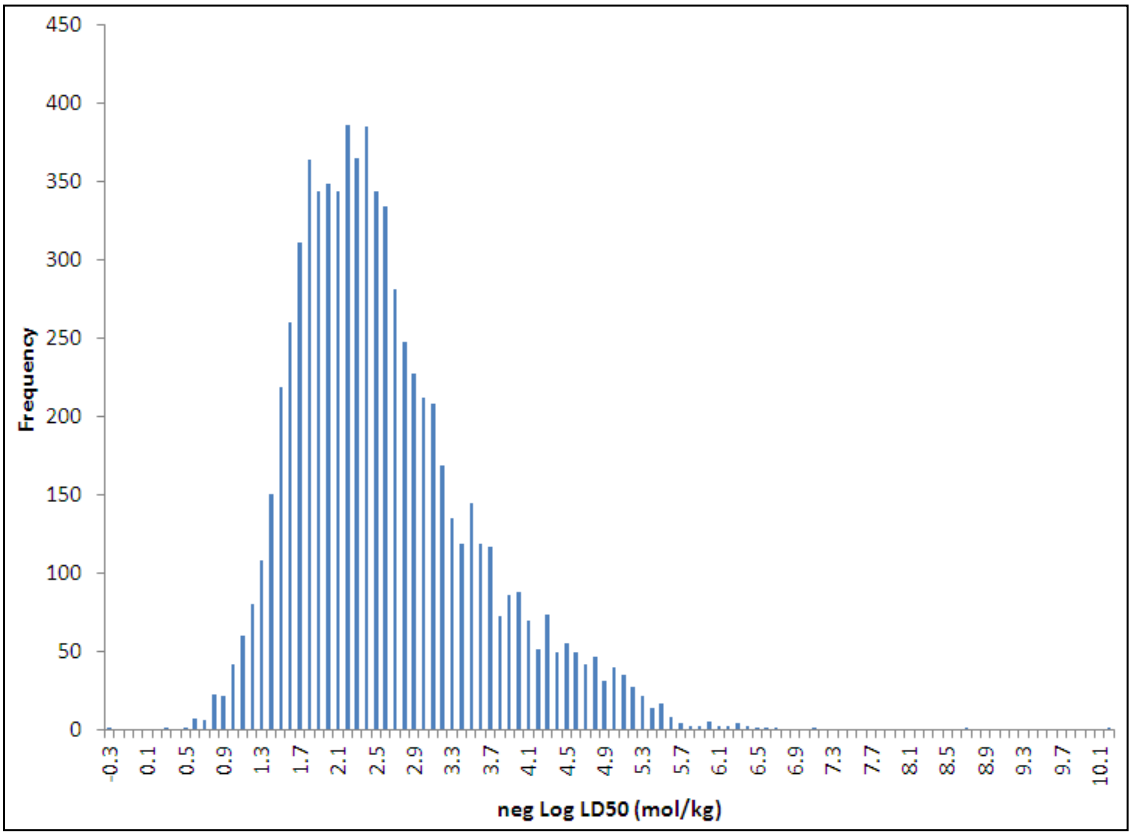



**Figure 9. Histogram of negative log of $LD_{50}$ values (mol/kg) of 3472 compounds.**

**The highest value was that of tetrachlorodibenzo-p-dioxin (TCDD), with a value of 10.2, while the lowest was 6-methyl uracil, with a value of 0.29. The average value was 2.466. Data was transformed and graphed using Excel.**

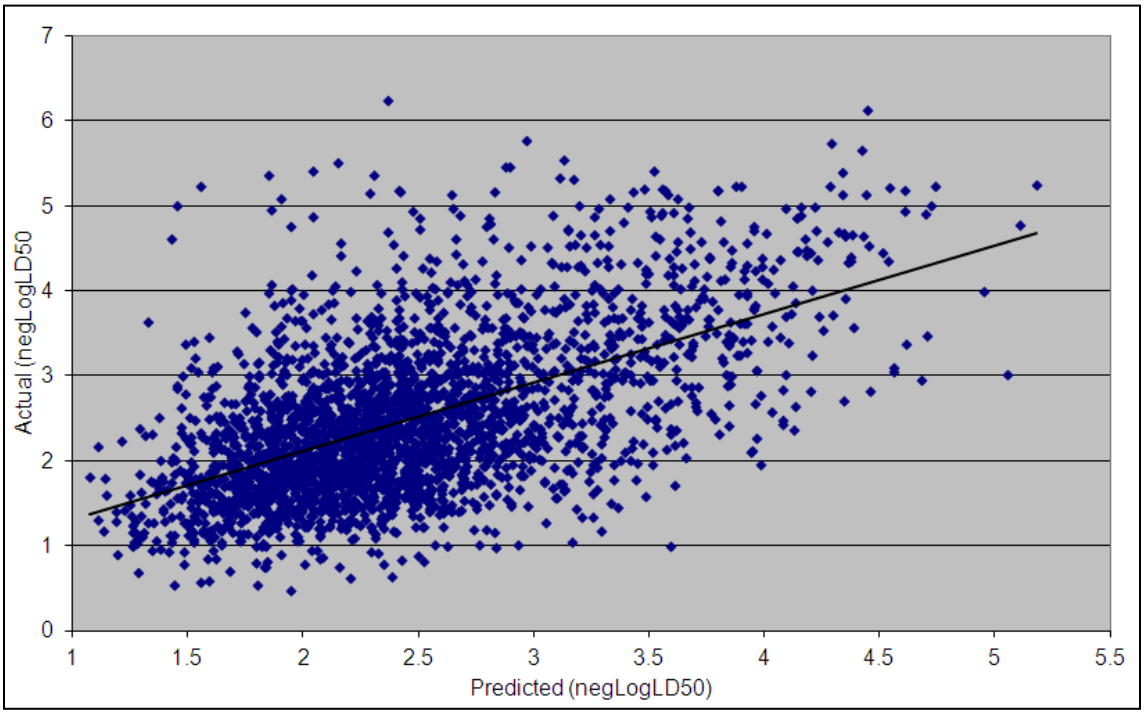



**Figure 10. Actual vs. Predicted LD50 values from best kNN-QSAR model.**

## 5. Bioavailability Predictive QSAR Modeling

While the values for bioavailability were well distributed **(Figure 11)**, the bioavailability model did not perform as well. There are 22 descriptors for this model, these were all generated from MOE, and they are as follows: randicIndex, FASA_H, PEOE_VSA_FPPOS, pmiY, vdw_vol, SMR_VSA2, Q_RPC-, PEOE_VSA_FPOL, PEOE_VSA-2, vsurf_HB2, PEOE_VSA+3, b_1rotR, carboRingCountOfSize, logDPHYS, SMR_VSA1, aliphaticRingCount, chi1v_C, a_ICM, SlogP_VSA0, BCUT_SMR_3, GCUT_SLOGP_2 mr, and BCUT_SLOGP_1. There are four nearest neighbors, and the $q^2$ is 0.363177 for 414 compounds, and the $r^2$ is 0.123 for 306 compounds **(Figure 12)**.

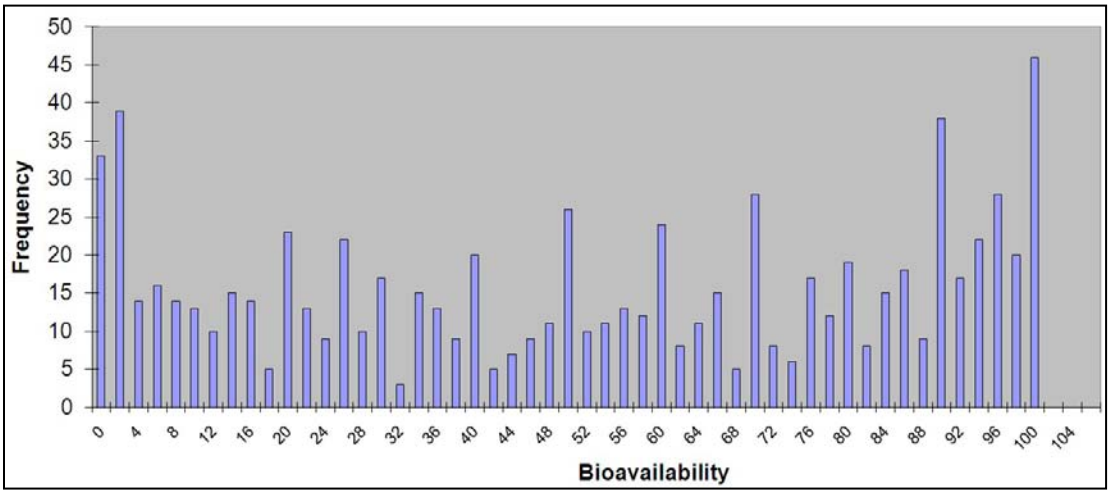



**Figure 11. Histogram of 809 bioavailability values used in the predictive model.**

**The data set has good distribution over the range of possible values with a very slight bias towards high values: the average value is 52.0 and median value is 54. Graph was generated in Excel.**

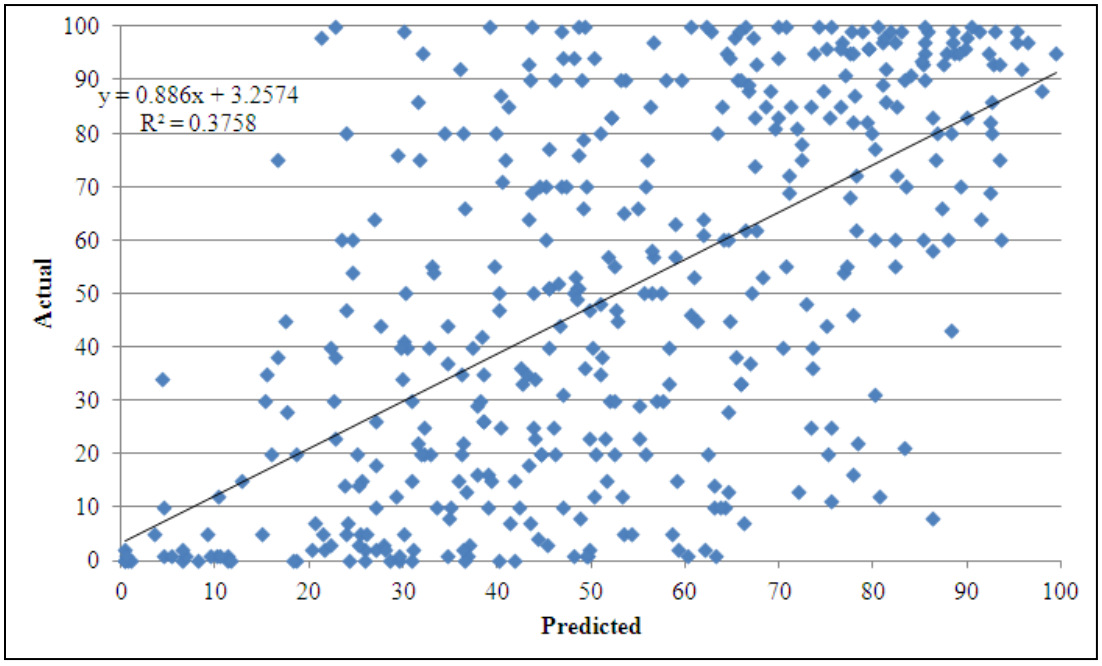



**Figure 12. Actual vs. Predicted bioavailability values from kNN-QSAR model.**

**This is the $r^2$ test set, with 414 compounds. Values were taken from the output of the best quality model, and were graphed in Excel. The low quality of the data used in this model adversely affected the predictive quality of the model itself.**

**Discussion:**

## 1. Computational Hot Spot Predictions of TRAF6

The TRAF6/RANK pathway is typical of most biological cellular functions, in that a protein-protein interaction is required to retain functionality [79]. Broadly speaking, formation of protein-protein interactions necessitate particular contacts and interactions which form the foundation for almost all biological functions, most notable of these are events are signal transduction events. Protein-protein interactions are seen as having locally primed features optimal for the binding of their partners, and they form highly conserved and dense networked clusters contributing cooperatively to the complex's stability [80]. The interaction sites of these protein-protein pairings are made up of surfaces that contain electrostatic and shape complementarity properties [81-84]. Protein-protein interactions can also be drastically influenced by the inherent properties of their residues, such as hydrophobicity and flexibility [82-86]. Some protein-protein interfaces have been described with relatively small areas, with surfaces as small as 1150-1200 $Å^2$ for structures such as complexes forming low-stability and short-lived regions. So called "standard-size" surfaces of interaction [87] fall into the region of 1600 $Å^2$ (± 400 $Å^2$) and involve relatively small conformational fluctuations and side-chain movements upon forming said complexes[88]. "Large" interfaces involve 2000 to 4600 $Å^2$, and require significant, global changes in the interacting members, and are seen primarily in G-protein receptors and family members, as well as information processing elements of signaling proteins. Using the DMBL-EBI PDBePISA (Protein Interfaces, Surfaces and Assemblies) server [89], the TRAF6/RANK complex is shown to have a very small interface area, being only 557.8 $Å^2$ in size. The TRAF6/CD40 complex (1LB6), has a slightly larger surface area, with 573.1 $Å^2$. This means that the

TRAF6/RANK complex is very small and unstable, and very little changes in either structure are expected to occur when binding takes place.

The three methods used to predict hot spot residues used a variety of techniques, and each has been successful in other test sets of proteins. Robetta achieved a perfect score in its prediction of the TRAF6 residues, and in another study, correctly predicted 79% of residues with a 1.0 kcal/mol cutoff (indicating they are in fact hot spots) with an error averaging to the value of 1.06 kcal/mol [78, 90]. While this method was the most successful in predicting hot spots of the TRAF6/RANK complex, there is one potential source of error in the implicit solvation model used by Robetta and for other predictions: hot spots that are mediated from water molecules can potentially compensate for particular alanine mutations, and can potentially give a false positive result. Since the waters were removed from this complex before submitting to Robetta, this was not an issue.

HotPoint is based on work that was done by Tuncbag *et al.* it uses primarily accessible surface area (ASA) and includes factors for conservation and pair-wise potentials of the residues to form an empirical model. This model has been recorded at 70% accuracy, when using 150 residues from ASEdb as the training set; this accuracy is higher than many machine learning based methods [91]. This technique was implemented into a publicly available web server, also named HotPoint, where visitors can register and submit a job and easily visualize the output prediction on any compatible browser [52]. Accessibilities to the solvent for each residue are computed for each form (bound and unbound), and the energetic pair-wise potential contributions from the solvent were provided by Keskin *et al* [92]. If and only if the following two endpoints are satisfied: the relative ASA in complex is equal to less than 20%, and the summation of the potential for

a given pair is greater or equal to 18.0, then that residue is classified as a hot spot [52]. The mutation of solvent accessible surface area (ΔASA) when a complex is formed for two proteins has a high degree of correlation to the energy of solvation. This factor was previously identified to be highly predictive versus other factors towards that of the binding of protein partners [93]. The same group also developed a database from this technique, called HotSprint, and was successful in predicting hot spots on the p53 binder, Mdm2. An accuracy of 76% was achieved, which can out-compete many popular methods that rely on machine-learning approaches [94].

KFC2 (Knowledge-based FADE and Contacts, by Darnell, Zhu, Page, and Mitchell [53, 95] ) uses a support vector machine (SVM) [96], machine-learning method to build two models, each of which are decision tree-based. Each model is built on two techniques: the first one is named K-FADE, it uses the size of a given residue and the specificity of its shape that are calculated by Fast Atomic Density Evaluation (hence the name FADE) [97], and K-CON uses the residue's hydrogen bonds, chemical types, interface points, and intermolecular atomic contacts [54]. These features are combined to output an answer in binary form to the question of if a residue is a hot spot or not, and provide a corresponding confidence score with this binary prediction. KFC2 has two models trained on SVM: KFC2a and KFC2b. The KFCa model was shown to have 0.85 as its rate of predicting hot spots correctly. This value was an improvement from the previous KFC method, as well as other predictions from HotPoint, Robetta, and FOLDEF, despite KFCa possessing a greater rate of incorrectly labeled hot spots than these competitors. KFC2b uses a total of seven features (overlapping two of the features from KFCa), and is seen to output a specificity greater than KFCa [53]. It is interesting that despite KFC2a's higher reported sensitivity and accuracy when compared to KFC2b,

no hot spots were predicted for the TRAF/RANK complex with KFC2a. It is also interesting that while the KFC2a model predicted no hot spots on TRAF6, the KFC2b model appears to have had both a high true positive, and a higher false positive rate.

## 2. Preparation and Analysis of Target Receptor

Selection of crystal structures 1LB5/1LB6 was necessary since both have the RANK/CD40 peptides bound to them. The remaining C-terminal 21 residues at the end of the crystal structure that are not included but exist in the TRAF6 sequence do not lie close to the RANK/CD40 binding site. The shortest distance from Thr501 to the RANK peptide is 16.4Å. Also, Thr501 ends a β-sheet, so the remaining residues of the structure are likely in a loop region that either interacts with the other loop regions in this area, or with the long helix that forms in the TRAF6 trimer **(Figure13)**

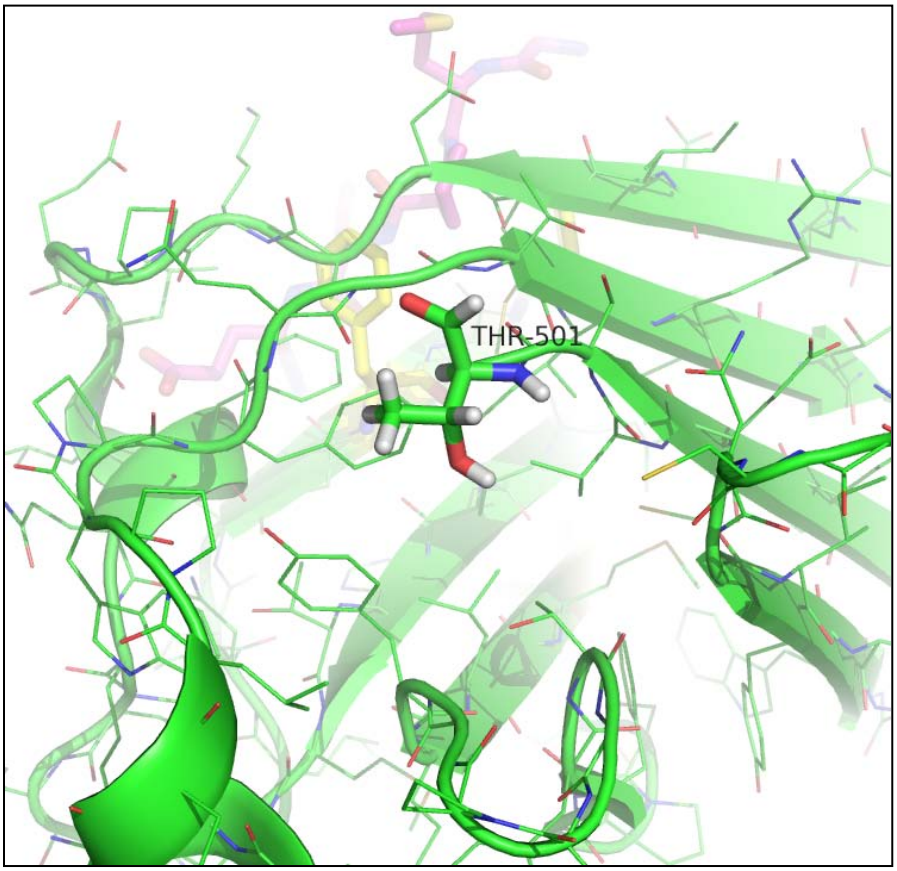



**Figure 13. C-terminal Thr501 residue in TRAF6 1LB5 crystal structure (green).**

**RANK residue shown in magenta sticks. This threonine residue lies at least 19 angstroms to the RANK binding site; remaining residues on this tail are unlikely to have an effect on the binding site.**

1LB5 was selected as the structure for which to do docking, since it contains the RANK peptide bound to TRAF6. [15]. While the TRAF6-CD40 complex of 1LB6 has a slightly higher resolution (1.80Å), it is very similar to 1LB5 (resolution of 2.40 Å). The RMS of the two structures (when superimposed) is only 0.294 Å, and this difference is mainly in the loop regions of the two structures. The only TRAF6 residue in the binding site that is significantly different in its conformation is the Arg392 residue (**Figure 14)**.

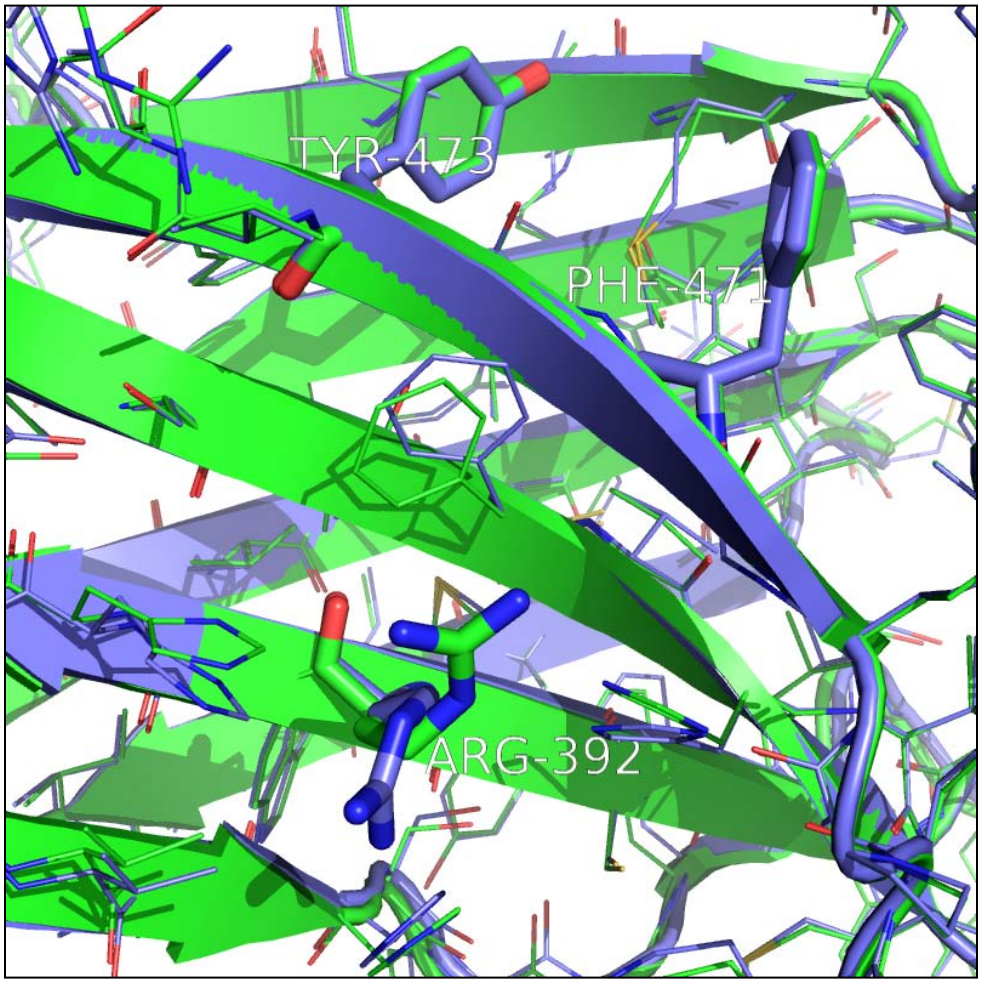



**Figure 14. Superimposed overlay of 1LB5 (green) and 1LB6 (blue) structures.**

**Hotspot residues are displayed in sticks. Tyr473 and Phe471 are nearly identical in their conformation, while Arg392 is rotated significantly.**

### 3. Chemical Library Selection

The choice of MyriaScreen as the primary ligand screening set ensured that adequate chemical diversity was present in the virtual screening, while retaining good drug-like and lead-like structures. Additionally, MyriaScreen compounds are easily commercially available, making for easy procurement and testing of the selected compounds.

### 4. Virtual Screening

GoldScore fitness function was selected as the scoring function, as it has been created and optimized for small molecule docking into the prediction of ligand binding energy. GOLD has proven to be a more than acceptable program for various techniques when compared with other docking programs [98-100], and it has been shown to be capable of reproducing reliable docked conformations of ligands for many of my other projects. The program is able to run on a multi-core environment, which is favorable to other methods that are more computationally expensive.

### 5. Processing of Results

The clustering of the GOLD results based on MACCS fingerprints removes similar compounds and increases the diversity of the pool of molecules by removing structurally similar compounds. Retaining the highest scoring representative structure of each cluster ensures the quality of the docked conformations is retained, but repetitive structures are discarded. This technique also ensures that the highest scoring representatives of each of the compounds is retained and selected as the best representative of the respective chemical core and binding pose.

### 6. Selection of Compounds for Biological Testing

As for the selection of the final hits, the technique used here has been used in various other projects. The cluster centers were individually visualized by hand in PyMOL to maximize the potential of finding ligands that would prove to be active biologically. This higher scrutiny at this stage not only ensured that there was more than

adequate diversity in the final compounds; it confirmed that the compounds resided in the desired interaction site and therefore had a high potential of inhibiting the interaction with RANK. Also of note, our analysis was independent of the docking scores obtained; this made for our analysis to be more qualitative than quantitative.

**Flexible Docking of three active compounds**

The three compounds that showed affinity (SZB: 40, 45, and 46) were submitted to a more rigorous GOLD docking, in which the ten closest residues to their original conformation were allowed to have their side chains be fully flexible; this process permits a deeper understanding of the potential interactions between the hits and TRAF6. For these docking jobs, GOLD was again used as the docking program, and the number of operations and population size of the genetic algorithm were doubled from the original virtual screening protocol.

Although the SZB-40 compound contains a ring structure that can gain stability by pi-pi stacking in between Tyr473 and Phe471, the flexible docking did not show such a conformation **(Figure 15)**. Instead, there was main chain hydrogen bonding seen between Gly470 and His412. But the most significant difference in the side chains (due to the flexible docking) is seen in the Arg392 and Arg466 residues. There are two hydrogen bonded interactions to the Arg392 and this is stabilized by the pi-stacking of the terminal (non-chlorinated) benzene structure of SZB-40 to Arg466. The movement of these two structures between the original structures creates a more favorable pocket.

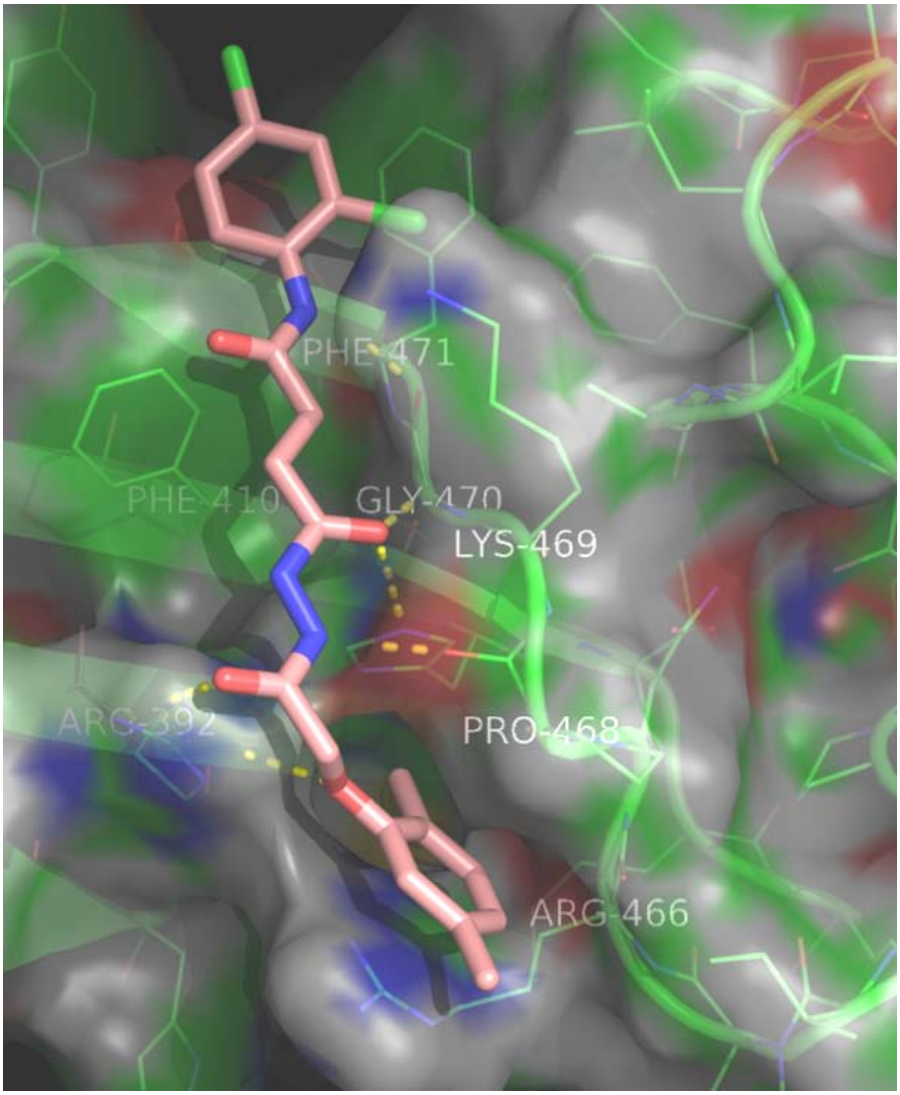



**Figure 15. Flexible GOLD docking of SZB-40 (pink) to TRAF6 (green).**

**Hydrogen bonds are shown as dashed yellow lines. This conformation show a weak pi-stacking to the Arg466 residue, and main chain hydrogen bonds to the Gly470 and Phe471, but no significant interactions to the aromatic Phe410, Phe471, or Tyr473 residues.**

SZB-45 was similar to SZB-40 in that it also had a moiety by which to mimic the proline structure on RANK in the pi-pi stacking interaction between the Tyr473 and Phe471 residues, but this conformation was not seen in the docking results **(Figure 16)**. This structure had much more extensive binding to the Arg392 and Arg466 structures (from the highly charged sulfonamide region) at multiple locations on each of these

residues. There is also a hydrogen bonding interaction with the Asn467 main chain. There appears to be a pi-stacking interaction with the Phe410 moiety; this somewhat validates the prediction of Hotpoint and KFC2, which both predicted this residue as a hot spot. Robetta calculated this residue's ΔΔG of the complex with a value of 0.76 kcal/mol.

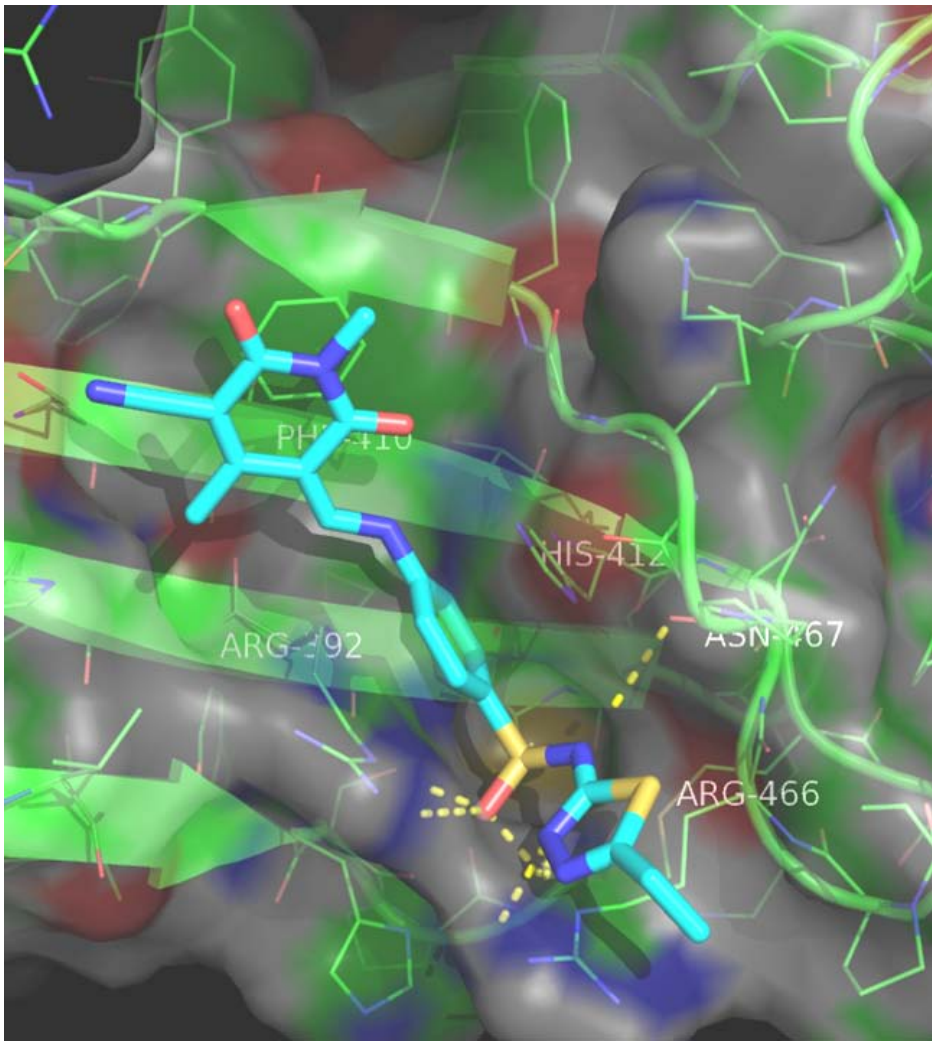



**Figure 16. Flexible docking of SZB-45 (blue) to TRAF6 (green).**

**Hydrogen bonds shown as dashed yellow lines. Main chain hydrogen bonding interactions are seen to Asn467 and side chain hydrogen bonds are seen to the Arg466 and Arg392 residues. There is also a good pi-stacking interaction with the Phe410 residue.**

As for SZB-46, this conformation seems to have a better opportunity for pi-stacking in between the two aromatic structures, and it also is stabilized by the two

arginine groups **(Figure 17)**. Additionally, the hydrogen bonding interactions with the Gly470 are similar to that of RANK, and there are two hydrogen atoms on the Lys469 structure that share hydrogen bonds with two different carbonyl regions of the SZB-46 ligand.

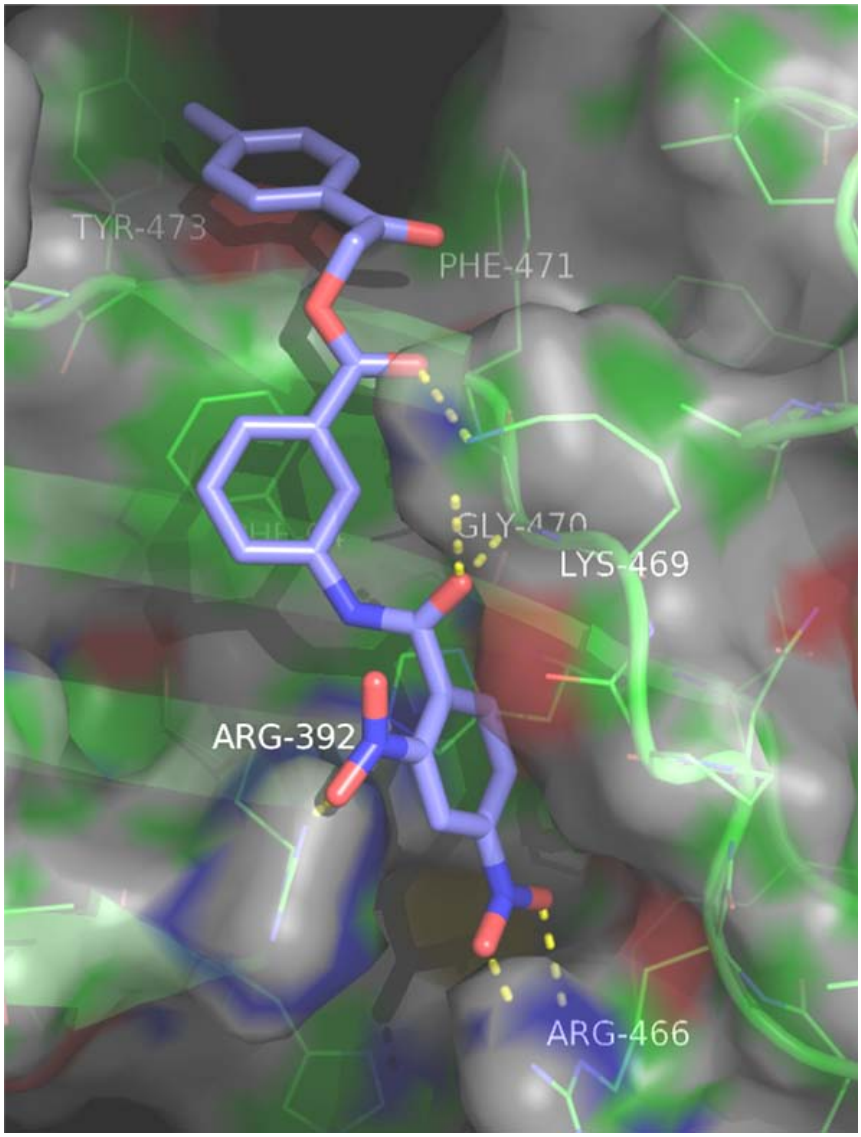



**Figure 17. Flexible docking of SZB-46 (blue). Hydrogen bonds shown in yellow.**

**This conformation shows side chain hydrogen bonds to Arf466, Arg392, and main chain bonds to Gly470 and Phe471. There is also a good pi-stacking interaction to Phe410.**

## 7. Biological Testing

The GST-RANK pull down assay did manage to show an inhibition of three of the hits on the second round of testing, but as this method is more qualitative than quantitative, it is difficult to elaborate on the results beyond what is apparent from the Western blot. Further testing is required (preferably one that can be quantized) using another technique, possibly an ITC or luciferase assay.

## 8. Molecular Dynamics simulations

The choice of GROMACS as the molecular dynamics (MD) package was an easy one, as GROMACS is very flexible and widely used software [64]. GROMACS has been shown to be much more efficient than other MD programs, and scales very well on large, parallel computing clusters [101]. It has been shown that the topologies resulting from the PRODRG server can deviate from simulations from GROMOS parameters, and reasonable configurations can be achieved using this method [102]. When the average back bone structure of the MD simulation is output from the MD simulation and then aligned to the original 1LB5 crystal structure, the RMSD is only 1.346Å, and this difference lies only in the loop regions, not in the β-sheet region **(Figure 18)**. It is reasonable to assume that the stability of the β-sheet sandwich in TRAF6 is highly stable in the explicit water solvent used here. Hence, the GROMACS molecular dynamics method appears to be a good system on which to compare the result of this and future TRAF6 inhibitors.

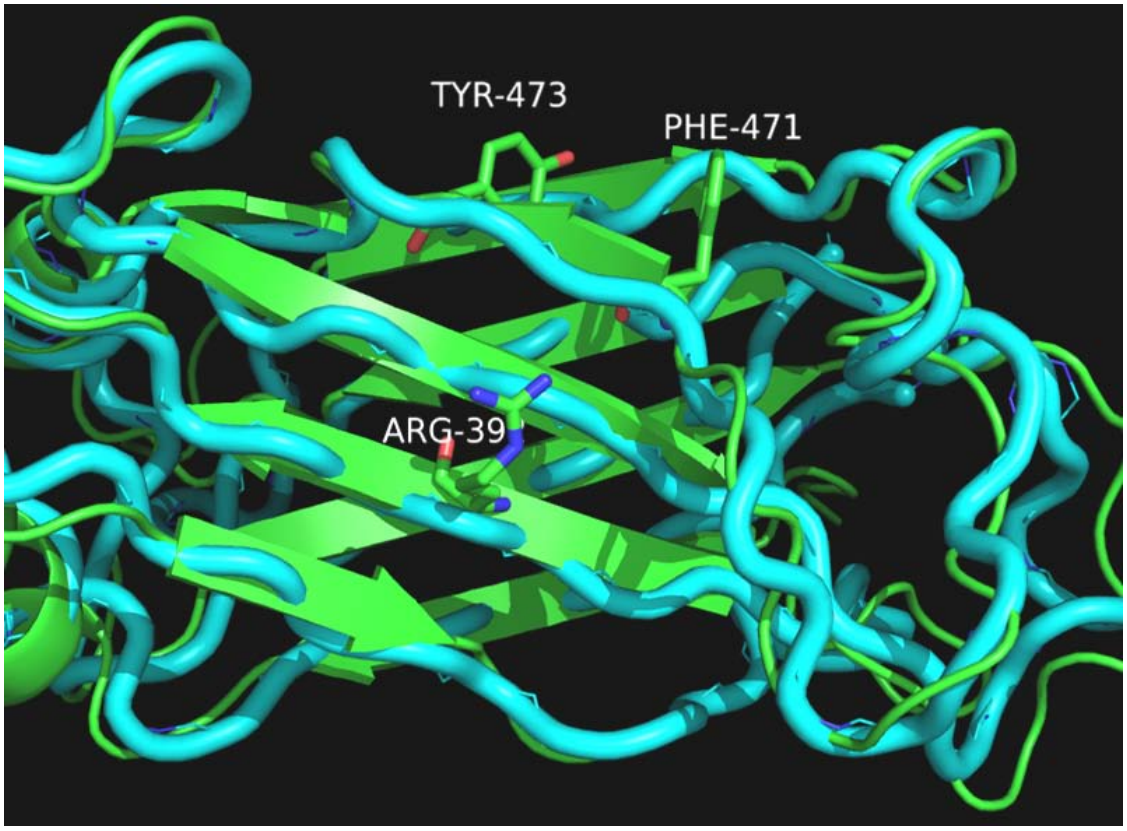



**Figure 18. Alignment of cartoon structures of MD simulation of TRAF/RANK.**

**Average structure (backbone atoms) of results of MD simulation (blue), 1LB5 structure TRAF6/RANK complex (green). RMSD of these structures is 1.346Å. The three hot spot residues are shown as sticks as a point of reference. Output for the simulation was prepared using Gromacs, and the structures were aligned using the "align" feature in PyMOL.**

### 9. $LD_{50}$ Predictive QSAR Modeling

The large size of the dataset was a potential hindrance in the production of better models, but considering the variability of the data, the models were able to achieve a general prediction as to the magnitude of the endpoint which was attempted to be modeled. While most of the data is from oral testing of rats, the species and weights of these rats is not available. What is interesting (and perhaps contradictory to the previous statement) is that selecting only the less toxic compounds (removing the top 10% most lethal compounds from the data set) actually produced worse models than the one

presented here. The model presented includes highly lethal compounds like several forms of polychlorinated dibenzodioxins (dioxins), and the nerve agent O-ethyl-S-(2-dimethylaminoethyl)-methyl-phosphonothioate (EDMM). Since lethality can manifest in different forms (inhibition of protein synthesis, mitochondrial damage, blocking of critical ion channels, DNA damage, to name a few), there are many different structural cores that are present in the dataset. These different structural cores and their corresponding mechanisms of lethality are likely creating confusion in the models. A data set that only contained ion channel disruptors is more likely to produce a good model of lethality, but would not be representative (or applicable) to other, more diverse compounds. Despite the variability of the data, logP is known to be a major factor in toxicity, as non-polar structures are much more likely to be metabolized and excreted (as well as be handled by the acidic gut) before their toxicity can become apparent.

### 10. Bioavailability Predictive QSAR Modeling

The inconsistencies in the original data set definitely hurt the quality of the models generated, it is apparent that there is over fitting of the data, and the number of descriptors seen in the model (especially when compared with the $LD_{50}$ model) backs this assumption up. Variability in the dosing, formulation, pharmacogenomics (polymorphisms in efficacy and other pharmacokinetic properties), health of the subjects and sampling error when recording the bioavailability data will all lead to a corresponding decrease of the quality of the models, and this is most likely what is seen here. It appears that such a complex endpoint in humans is difficult to model with simple tools, and classification schemes will likely lead to better predictive models. The strength

of using a kNN based model such as the one that was used here is that it is non-parametric; there are no assumptions made on the distribution of the data set.

**Conclusion:**

In this study novel inhibitors of the TRAF6/binding peptide interaction site are presented that were discovered through a rational design process incorporating various *in silico* techniques. From an initial set of 10,000 compounds, 26 of them were tested. Three of these 26 compounds were shown to be biologically active. These results validate the virtual screening and biological methodology described here, as well as validates the TRAF6/binding peptide interaction as a potential target for small molecule inhibition.

The results here lay the framework for future study of inhibition of the TRAF6/binding peptide interaction, and further study of the binding mechanism (either by crystallization of the inhibitors in complex with TRAF6, or by other biological experiments) will reveal means of optimizing the inhibitors for use in clinical settings. Future studies of the generation of TRAF6 inhibitors will surely benefit from additional assays for which to test the activity and mechanism of binding. Techniques such as a luciferase assay or isothermal titration calorimetry (ITC) would bolster the evidence of a possible inhibitor from the initial GST assay described above. Once a decent inhibitor has been found, optimization of the respective chemical groups can increase the binding affinity even further. While the TRAF6/binding peptide protein-protein interaction is a very valuable target, it is a difficult one. The binding site is not a prototypical pocket that is conducive to traditional drug design efforts, and the interaction site, despite being small for protein-protein interactions, is large for small molecule inhibitors to bind to. In

general, targeting protein-protein interactions are high-risk and high-reward, but the TRAF6/binding peptide interaction is beyond the exemplary target and has even higher potential value. Computational techniques such as the ones that I have outlined here (hot spot prediction, virtual screening, and molecular dynamics simulations) will help to ameliorate the risk involved by guiding the decisions throughout the drug design process.

Given that TRAF6 plays many essential roles in immunity and other diverse biological functions, TRAF6 inhibitors are most certainly in high demand. Not only will they serve to advance the understanding of the TRAF6 pathway, but has the potential to improve the lives of patients through controlling inflammation and treating a wide range of diseases, such as: osteoporosis, cancer-induced bone lesions and other bone diseases, postmenopausal osteoporosis, multiple myeloma, periodontitis, connective tissue destruction, bladder outlet obstruction, Paget's disease, and viral infections [103-105]. On the front of cancer therapeutics, it has been postulated that TRAF6 inhibitors should be able to inhibit the spread of multiple myeloma and prevent bone loss (the most significant clinical manifestation of MM). These inhibitors will mean the inclusion of a new target for the treatment of cancer, and will be able to overcome resistance to chemotherapy [106]. There are still more questions as to the TRAF6 pathway, most notably the specific activation mechanism of TRAF6. It is unknown if TRAF6 is monomeric before recruitment to its receptors; additionally it is unclear if it is the TRAF6 oligomerization or the resulting conformational changes from oligomerization that guide the subsequent activation of TRAF6 [107]. These questions, as well as understanding how ubiquitination plays a role in TRAF6 activation, need to be answered to complete the puzzle. As the understanding of TRAF6 and its effectors continues to be understood, the use and affinity of TRAF6 inhibitors will prove to me more efficacious as our understanding improves.

## References


1. Zapata J.M., Martinez-Garcia V., and Lefebvre S., Phylogeny of the TRAF/MATH domain. Adv Exp.Med Biol 2007; 597: 1-24.
2. Chung J.Y., Park Y.C., Ye H., and Wu H., All TRAFs are not created equal: common and distinct molecular mechanisms of TRAF-mediated signal transduction. J Cell Sci 2002; 115: 679-688.
3. Xu L.G., Li L.Y., and Shu H.B., TRAF7 potentiates MEKK3-induced AP1 and CHOP activation and induces apoptosis. J Biol Chem 2004; 279: 17278-17282.
4. Zapata J.M., Lefebvre S., and Reed J.C., Targeting TRAfs for therapeutic intervention. Adv Exp.Med Biol 2007; 597: 188-201.
5. Wu H., Assembly of post-receptor signaling complexes for the tumor necrosis factor receptor superfamily. Adv Protein Chem 2004; 68: 225-79.
6. Bradley J.R. and Pober J.S., Tumor necrosis factor receptor-associated factors (TRAFs). Oncogene 2001; 20: 6482-6491.
7. Ely K.R. and Li C., Structurally adaptive hot spots at a protein interaction interface on TRAF3. J Mol Recognit 2002; 15: 286-90.
8. Inoue J., Ishida T., Tsukamoto N., Kobayashi N., Naito A., Azuma S., and Yamamoto T., Tumor necrosis factor receptor-associated factor (TRAF) family: adapter proteins that mediate cytokine signaling. Exp.Cell Res 2000; 254: 14-24.
9. Ishida T., Mizushima S., Azuma S., Kobayashi N., Tojo T., Suzuki K., Aizawa S., Watanabe T., Mosialos G., Kieff E., Yamamoto T., and Inoue J., Identification of TRAF6, a novel tumor necrosis factor receptor-associated factor protein that mediates signaling from an amino-terminal domain of the CD40 cytoplasmic region. J Biol Chem 1996; 271: 28745-28748.

10. Cao Z., Xiong J., Takeuchi M., Kurama T., and Goeddel D.V., TRAF6 is a signal transducer for interleukin-1. Nature 1996; 383: 443-446.
11. Lee N.K. and Lee S.Y., Modulation of life and death by the tumor necrosis factor receptor-associated factors (TRAFs). J Biochem Mol Biol 2002; 35: 61-6.
12. Park Y.C., Burkitt V., Villa A.R., Tong L., and Wu H., Structural basis for self-association and receptor recognition of human TRAF2. Nature 1999; 398: 533-538.
13. Ye H., Park Y.C., Kreishman M., Kieff E., and Wu H., The structural basis for the recognition of diverse receptor sequences by TRAF2. Mol Cell 1999; 4: 321-330.
14. McWhirter S.M., Pullen S.S., Holton J.M., Crute J.J., Kehry M.R., and Alber T., Crystallographic analysis of CD40 recognition and signaling by human TRAF2. Proc Natl Acad Sci U S A 1999; 96: 8408-8413.
15. Ye H., Arron J.R., Lamothe B., Cirilli M., Kobayashi T., Shevde N.K., Segal D., Dzivenu O.K., Vologodskaia M., Yim M., Du K., Singh S., Pike J.W., Darnay B.G., Choi Y., and Wu H., Distinct molecular mechanism for initiating TRAF6 signalling. Nature 2002; 418: 443-447.
16. Zapata J.M., TNF-receptor-associated factors as targets for drug development. Expert Opin Ther Targets 2003; 7: 411-425.
17. Darnay B.G., Ni J., Moore P.A., and Aggarwal B.B., Activation of NF-kappaB by RANK requires tumor necrosis factor receptor-associated factor (TRAF) 6 and NF-kappaB-inducing kinase. Identification of a novel TRAF6 interaction motif. J Biol Chem 1999; 274: 7724-7731.
18. Wu H. and Arron J.R., TRAF6, a molecular bridge spanning adaptive immunity, innate immunity and osteoimmunology. Bioessays 2003; 25: 1096-1105.
19. O'Neill L.A., Fitzgerald K.A., and Bowie A.G., The Toll-IL-1 receptor adaptor family grows to five members. Trends Immunol 2003; 24: 286-90.

20. Cao Z., Henzel W.J., and Gao X., IRAK: a kinase associated with the interleukin-1 receptor. Science 1996; 271: 1128-31.

21. Muzio M., Ni J., Feng P., and Dixit V.M., IRAK (Pelle) family member IRAK-2 and MyD88 as proximal mediators of IL-1 signaling. Science 1997; 278: 1612-5.

22. Suzuki N., Suzuki S., Duncan G.S., Millar D.G., Wada T., Mirtsos C., Takada H., Wakeham A., Itie A., Li S., Penninger J.M., Wesche H., Ohashi P.S., Mak T.W., and Yeh W.C., Severe impairment of interleukin-1 and Toll-like receptor signalling in mice lacking IRAK-4. Nature 2002; 416: 750-6.

23. Wesche H., Gao X., Li X., Kirschning C.J., Stark G.R., and Cao Z., IRAK-M is a novel member of the Pelle/interleukin-1 receptor-associated kinase (IRAK) family. J Biol Chem 1999; 274: 19403-10.

24. Leibbrandt A. and Penninger J.M., RANK/RANKL: regulators of immune responses and bone physiology. Ann N Y Acad Sci 2008; 1143: 123-50.

25. Fata J.E., Kong Y.Y., Li J., Sasaki T., Irie-Sasaki J., Moorehead R.A., Elliott R., Scully S., Voura E.B., Lacey D.L., Boyle W.J., Khokha R., and Penninger J.M., The osteoclast differentiation factor osteoprotegerin-ligand is essential for mammary gland development. Cell 2000; 103: 41-50.

26. Dougall W.C., Glaccum M., Charrier K., Rohrbach K., Brasel K., De Smedt T., Daro E., Smith J., Tometsko M.E., Maliszewski C.R., Armstrong A., Shen V., Bain S., Cosman D., Anderson D., Morrissey P.J., Peschon J.J., and Schuh J., RANK is essential for osteoclast and lymph node development. Genes Dev 1999; 13: 2412-24.

27. Kong Y.Y., Yoshida H., Sarosi I., Tan H.L., Timms E., Capparelli C., Morony S., Oliveira-dos-Santos A.J., Van G., Itie A., Khoo W., Wakeham A., Dunstan C.R., Lacey D.L., Mak T.W., Boyle W.J., and Penninger J.M., OPGL is a key regulator of osteoclastogenesis, lymphocyte development and lymph-node organogenesis. Nature 1999; 397: 315-23.

28. Schramek D., Leibbrandt A., Sigl V., Kenner L., Pospisilik J.A., Lee H.J., Hanada R., Joshi P.A., Aliprantis A., Glimcher L., Pasparakis M., Khokha R., Ormandy C.J., Widschwendter M., Schett G., and Penninger J.M., Osteoclast differentiation factor RANKL controls development of progestin-driven mammary cancer. Nature 2010; 468: 98-102.

29. Jones D.H., Nakashima T., Sanchez O.H., Kozieradzki I., Komarova S.V., Sarosi I., Morony S., Rubin E., Sarao R., Hojilla C.V., Komnenovic V., Kong Y.Y., Schreiber M., Dixon S.J., Sims S.M., Khokha R., Wada T., and Penninger J.M., Regulation of cancer cell migration and bone metastasis by RANKL. Nature 2006; 440: 692-6.

30. Naito A., Yoshida H., Nishioka E., Satoh M., Azuma S., Yamamoto T., Nishikawa S., and Inoue J., TRAF6-deficient mice display hypohidrotic ectodermal dysplasia. Proc Natl Acad Sci U S A 2002; 99: 8766-8771.

31. Lomaga M.A., Yeh W.C., Sarosi I., Duncan G.S., Furlonger C., Ho A., Morony S., Capparelli C., Van G., Kaufman S., van der Heiden A., Itie A., Wakeham A., Khoo W., Sasaki T., Cao Z., Penninger J.M., Paige C.J., Lacey D.L., Dunstan C.R., Boyle W.J., Goeddel D.V., and Mak T.W., TRAF6 deficiency results in osteopetrosis and defective interleukin-1, CD40, and LPS signaling. Genes Dev 1999; 13: 1015-1024.

32. Naito A., Azuma S., Tanaka S., Miyazaki T., Takaki S., Takatsu K., Nakao K., Nakamura K., Katsuki M., Yamamoto T., and Inoue J., Severe osteopetrosis, defective interleukin-1 signalling and lymph node organogenesis in TRAF6-deficient mice. Genes Cells 1999; 4: 353-362.

33. Starczynowski D.T., Lockwood W.W., Delehouzee S., Chari R., Wegrzyn J., Fuller M., Tsao M.S., Lam S., Gazdar A.F., Lam W.L., and Karsan A., TRAF6 is an amplified oncogene bridging the RAS and NF-kappaB pathways in human lung cancer. J Clin Invest 2011/10; 121: 4095-4105.

34. Pageau S.C., Denosumab. MAbs 2009; 1: 210-5.

35. Terpos E., Efstathiou E., Christoulas D., Roussou M., Katodritou E., and Dimopoulos M.A., RANKL inhibition: clinical implications for the management of patients with multiple myeloma and solid tumors with bone metastases. Expert Opin Biol Ther 2009; 9: 465-79.

36. Perlot T. and Penninger J.M., Development and function of murine B cells lacking RANK. J Immunol 2012; 188: 1201-5.

37. Poblenz A.T., Jacoby J.J., Singh S., and Darnay B.G., Inhibition of RANKL-mediated osteoclast differentiation by selective TRAF6 decoy peptides. Biochem Biophys.Res Commun. 2007; 359: 510-515.

38. Yang W.L., Zhang X., and Lin H.K., Emerging role of Lys-63 ubiquitination in protein kinase and phosphatase activation and cancer development. Oncogene 2010; 29: 4493-503.

39. Mukhopadhyay D. and Riezman H., Proteasome-independent functions of ubiquitin in endocytosis and signaling. Science 2007; 315: 201-5.

40. Pickart C.M., Mechanisms underlying ubiquitination. Annu Rev Biochem 2001; 70: 503-33.

41. Lamothe B., Besse A., Campos A.D., Webster W.K., Wu H., and Darnay B.G., Site-specific Lys-63-linked tumor necrosis factor receptor-associated factor 6 auto-ubiquitination is a critical determinant of I kappa B kinase activation. J Biol Chem 2007; 282: 4102-4112.

42. Lamothe B., Campos A.D., Webster W.K., Gopinathan A., Hur L., and Darnay B.G., The RING domain and first zinc finger of TRAF6 coordinate signaling by interleukin-1, lipopolysaccharide, and RANKL. J Biol Chem 2008; 283: 24871-80.

43. Lamothe B., Webster W.K., Gopinathan A., Besse A., Campos A.D., and Darnay B.G., TRAF6 ubiquitin ligase is essential for RANKL signaling and osteoclast differentiation. Biochem Biophys Res Commun 2007; 359: 1044-9.

44. Yin Q., Lin S.C., Lamothe B., Lu M., Lo Y.C., Hura G., Zheng L., Rich R.L., Campos A.D., Myszka D.G., Lenardo M.J., Darnay B.G., and Wu H., E2 interaction and dimerization in the crystal structure of TRAF6. Nat Struct Mol Biol 2009; 16: 658-66.

45. Yang W.L., Wang J., Chan C.H., Lee S.W., Campos A.D., Lamothe B., Hur L., Grabiner B.C., Lin X., Darnay B.G., and Lin H.K., The E3 ligase TRAF6 regulates Akt ubiquitination and activation. Science 2009; 325: 1134-1138.

46. Sorrentino A., Thakur N., Grimsby S., Marcusson A., von Bulow V., Schuster N., Zhang S., Heldin C.H., and Landstrom M., The type I TGF-beta receptor engages TRAF6 to activate TAK1 in a receptor kinase-independent manner. Nat Cell Biol 2008; 10: 1199-207.

47. Yamashita M., Fatyol K., Jin C., Wang X., Liu Z., and Zhang Y.E., TRAF6 mediates Smad-independent activation of JNK and p38 by TGF-beta. Mol Cell 2008; 31: 918-24.

48. Korchnak A.C., Zhan Y., Aguilar M.T., and Chadee D.N., Cytokine-induced activation of mixed lineage kinase 3 requires TRAF2 and TRAF6. Cell Signal 2009; 21: 1620-5.

49. Conze D.B., Wu C.J., Thomas J.A., Landstrom A., and Ashwell J.D., Lys63-linked polyubiquitination of IRAK-1 is required for interleukin-1 receptor- and toll-like receptor-mediated NF-kappaB activation. Mol Cell Biol 2008; 28: 3538-47.

50. Bienstock R.J., Computational drug design targeting protein-protein interactions. Curr Pharm Des 2012; 18: 1240-54.

51. Kim D.E., Chivian D., and Baker D., Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res 2004; 32: W526-31.

52. Tuncbag N., Keskin O., and Gursoy A., HotPoint: hot spot prediction server for protein interfaces. Nucleic Acids Res 2010; 38: W402-6.

53. Zhu X. and Mitchell J.C., KFC2: A knowledge-based hot spot prediction method based on interface solvation, atomic density, and plasticity features. Proteins 2011; 79: 2671-83.

54. Darnell S.J., LeGault L., and Mitchell J.C., KFC Server: interactive forecasting of protein interaction hot spots. Nucleic Acids Res 2008; 36: W265-9.

55. Sussman J.L., Lin D., Jiang J., Manning N.O., Prilusky J., Ritter O., and Abola E.E., Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules. Acta Crystallogr D Biol Crystallogr 1998; 54: 1078-84.

56. *The PyMOL Molecular Graphics System, Version 1.2r1, Schr"dinger, LLC.* 2009.

57. Lipinski C.A., Lombardo F., Dominy B.W., and Feeney P.J., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 2001; 46: 3-26.

58. *Molecular Operating Environment (MOE), developed and distributed by Chemical Computing Group* *http://www.chemcomp.com/software-moe2009.htm* 2009.

59. Halgren T.A. and Bush B.L., The Merck molecular force field (MMFF94). Extension and application. Abstracts of Papers of the American Chemical Society 1996; 212: 2-Comp.

60. Courcot B. and Bridgeman A.J., Modeling the interactions between polyoxometalates and their environment. J Comput Chem 2011; 32: 3143-53.

61. Verdonk M.L., Cole J.C., Hartshorn M.J., Murray C.W., and Taylor R.D., Improved protein-ligand docking using GOLD. Proteins 2003; 52: 609-623.

62. Jones G., Willett P., Glen R.C., Leach A.R., and Taylor R., Development and validation of a genetic algorithm for flexible docking. J Mol Biol 1997; 267: 727-48.

63. Hess B., Kutzner C., van der Spoel D., and Lindahl E., GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. Journal of Chemical Theory and Computation 2008; 4: 435-447.

64. Van Der Spoel D., Lindahl E., Hess B., Groenhof G., Mark A.E., and Berendsen H.J., GROMACS: fast, flexible, and free. J Comput Chem 2005; 26: 1701-18.

65. Scott W.R.P., Hunenberger P.H., Tironi I.G., Mark A.E., Billeter S.R., Fennen J., Torda A.E., Huber T., Kruger P., and van Gunsteren W.F., The GROMOS biomolecular simulation program package. Journal of Physical Chemistry A 1999; 103: 3596-3607.

66. Schuttelkopf A.W. and van Aalten D.M., PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr D Biol Crystallogr 2004; 60: 1355-63.

67. (TACC) T.A.C.C., *http://www.tacc.utexas.edu*. The University of Texas at Austin.

68. Humphrey W., Dalke A., and Schulten K., VMD: visual molecular dynamics. J Mol Graph 1996; 14: 33-8, 27-8.

69. Zhu H., Martin T.M., Ye L., Sedykh A., Young D.M., and Tropsha A., Quantitative structure-activity relationship modeling of rat acute toxicity by oral exposure. Chem Res Toxicol 2009; 22: 1913-21.

70. Young D., Martin T., Venkatapathy R., and Harten P., Are the Chemical Structures in Your QSAR Correct? Qsar & Combinatorial Science 2008; 27: 1337-1345.

71. Tetko I.V., Gasteiger J., Todeschini R., Mauri A., Livingstone D., Ertl P., Palyulin V.A., Radchenko E.V., Zefirov N.S., Makarenko A.S., Tanchuk V.Y., and Prokopenko V.V., Virtual computational chemistry laboratory--design and description. J Comput Aided Mol Des 2005; 19: 453-63.

72. *Version 5.1 ChemAxon (http://www.chemaxon.com)*. 2010.

73. Zheng W. and Tropsha A., Novel variable selection quantitative structure--property relationship approach based on the k-nearest-neighbor principle. J Chem Inf Comput Sci 2000; 40: 185-94.

74. Hou T., Wang J., Zhang W., and Xu X., ADME evaluation in drug discovery. 6. Can oral bioavailability in humans be effectively predicted by simple molecular property-based rules? J Chem Inf Model 2007; 47: 460-3.

75. Veber D.F., Johnson S.R., Cheng H.Y., Smith B.R., Ward K.W., and Kopple K.D., Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem 2002; 45: 2615-23.

76. Hardman J.G., Limbird L.E., Gilman A.G., and Data. P., *Goodman & Gilman's The Pharmacological Basis of Therapeutics*. 2001.

77. DeLano W.L., Unraveling hot spots in binding interfaces: progress and challenges. Curr Opin Struct Biol 2002; 12: 14-20.

78. Kortemme T., Kim D.E., and Baker D., Computational alanine scanning of protein-protein interfaces. Sci STKE 2004; 2004: pl2.

79. Gavin A.C., Bosche M., Krause R., Grandi P., Marzioch M., Bauer A., Schultz J., Rick J.M., Michon A.M., Cruciat C.M., Remor M., Hofert C., Schelder M., Brajenovic M., Ruffner H., Merino A., Klein K., Hudak M., Dickson D., Rudi T., Gnau V., Bauch A., Bastuck S., Huhse B., Leutwein C., Heurtier M.A., Copley R.R., Edelmann A., Querfurth E., Rybin V., Drewes G., Raida M., Bouwmeester T., Bork P., Seraphin B., Kuster B., Neubauer G., and Superti-Furga G., Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 2002; 415: 141-7.

80. Keskin O., Ma B., and Nussinov R., Hot regions in protein--protein interactions: the organization and contribution of structurally conserved hot spot residues. J Mol Biol 2005; 345: 1281-94.

81. Janin J. and Chothia C., The structure of protein-protein recognition sites. J Biol Chem 1990; 265: 16027-30.

82. Janin J., Elusive affinities. Proteins 1995; 21: 30-9.

83. Janin J., Protein-protein recognition. Prog Biophys Mol Biol 1995; 64: 145-66.

84. Jones S. and Thornton J.M., Principles of protein-protein interactions. Proc Natl Acad Sci U S A 1996; 93: 13-20.

85. Chothia C. and Janin J., Principles of protein-protein recognition. Nature 1975; 256: 705-8.

86. Tsai C.J. and Nussinov R., Hydrophobic folding units at protein-protein interfaces: implications to protein folding and to protein-protein association. Protein Sci 1997; 6: 1426-37.

87. Horton N. and Lewis M., Calculation of the free energy of association for protein complexes. Protein Sci 1992; 1: 169-81.

88. Lo Conte L., Chothia C., and Janin J., The atomic structure of protein-protein recognition sites. J Mol Biol 1999; 285: 2177-98.

89. Krissinel E. and Henrick K., Inference of macromolecular assemblies from crystalline state. J Mol Biol 2007; 372: 774-97.

90. Kortemme T. and Baker D., A simple physical model for binding energy hot spots in protein-protein complexes. Proc Natl Acad Sci U S A 2002; 99: 14116-21.

91. Tuncbag N., Gursoy A., and Keskin O., Identification of computational hot spots in protein interfaces: combining solvent accessibility and inter-residue potentials improves the accuracy. Bioinformatics 2009; 25: 1513-20.

92. Keskin O., Bahar I., Badretdinov A.Y., Ptitsyn O.B., and Jernigan R.L., Empirical solvent-mediated potentials hold for both intra-molecular and inter-molecular inter-residue interactions. Protein Sci 1998; 7: 2578-86.

93. Eisenberg D. and McLachlan A.D., Solvation energy in protein folding and binding. Nature 1986; 319: 199-203.

94. Guney E., Tuncbag N., Keskin O., and Gursoy A., HotSprint: database of computational hot spots in protein interfaces. Nucleic Acids Res 2008; 36: D662-6.

95. Darnell S.J., Page D., and Mitchell J.C., An automated decision-tree approach to predicting protein interaction hot spots. Proteins 2007; 68: 813-23.

96. Noble W.S., What is a support vector machine? Nat Biotechnol 2006; 24: 1565-7.

97. Mitchell J.C., Kerr R., and Ten Eyck L.F., Rapid atomic density methods for molecular shape characterization. J Mol Graph Model 2001; 19: 325-30, 388-90.

98. Hui-fang L., Qing S., Jian Z., and Wei F., Evaluation of various inverse docking schemes in multiple targets identification. J Mol Graph Model 2010; 29: 326-30.

99. Plewczynski D., Lazniewski M., Augustyniak R., and Ginalski K., Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. J Comput Chem 2011; 32: 742-55.

100. Li X., Li Y., Cheng T., Liu Z., and Wang R., Evaluation of the performance of four molecular docking programs on a diverse set of protein-ligand complexes. J Comput Chem 2010; 31: 2109-25.

101. Kutzner C., van der Spoel D., Fechner M., Lindahl E., Schmitt U.W., de Groot B.L., and Grubmuller H., Speeding up parallel GROMACS on high-latency networks. J Comput Chem 2007; 28: 2075-84.

102. Lemkul J.A., Allen W.J., and Bevan D.R., Practical considerations for building GROMOS-compatible small-molecule topologies. J Chem Inf Model 2010; 50: 2221-35.

103. Walsh D.E., Greene C.M., Carroll T.P., Taggart C.C., Gallagher P.M., O'Neill S.J., and McElvaney N.G., Interleukin-8 up-regulation by neutrophil elastase is mediated by MyD88/IRAK/TRAF-6 in human bronchial epithelium. J Biol Chem 2001; 276: 35494-9.

104. Boch J.A., Wara-aswapati N., and Auron P.E., Interleukin 1 signal transduction--current concepts and relevance to periodontitis. J Dent Res 2001; 80: 400-7.

105. Burkhard F.C., Lemack G.E., Alcorn M.D., Zimmern P.E., Lin V.K., and Connell J.D., Up-regulation of a gene homologous to the human tumor necrosis factor receptor associated factor 6 gene in the obstructed rabbit bladder determined by differential display polymerase chain reaction. J Urol 2001; 165: 1289-93.

106. Liu H., Tamashiro S., Baritaki S., Penichet M., Yu Y., Chen H., Berenson J., and Bonavida B., TRAF6 Activation in Multiple Myeloma: A Potential Therapeutic Target. Clin Lymphoma Myeloma Leuk 2012.

107. Chung J.Y., Lu M., Yin Q., Lin S.C., and Wu H., Molecular basis for the unique specificity of TRAF6. Adv Exp.Med Biol 2007; 597: 122-130.

**Vita**

John Kenneth Morrow was born in Davis, California to Jeanne and Philip Ross Morrow. He graduated high school at University of San Diego in 1996, and obtained his Bachelors of Science in Molecular Toxicology at University of California, Berkeley. In August 2009, he joined the University of Texas Health Science Center and University of Texas MD Anderson Cancer Center Graduate School of Biomedical Sciences. He then joined the department of Experimental Therapeutics in the laboratory of Shuxing Zhang, Ph.D. to carry out his thesis research. John has published six papers on aspects of rational drug design since joining this lab, and he was awarded the Howard Hughes Medical Institute Med-Into-Grad Fellowship in 2012.