Targeting the NF κB/IκBα Complex via Fragment-Based E-Pharmacophore Virtual Screening and Binary QSAR Models

Tarek Kanan[a,b], Duaa Kanan[a,b], Ismail Erol[b,c], Samira Yazdi[d,e], Matthias Stein[d,]*, Serdar Durdagi[b,f,]*

[a]*School of Medicine, Bahcesehir University, Istanbul, Turkey;* [b]*Computational Biology and Molecular Simulations Laboratory, Department of Biophysics, School of Medicine, Bahcesehir University, Istanbul, Turkey;* [c]*Department of Chemistry, Gebze Technical University, Kocaeli, Turkey;* [d]*Molecular Simulations and Design Group, Max Planck Institute for Dynamics and Complex Technical Systems, Magdeburg, Germany;* [e]*Current address: Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden;* [f]*Neuroscience Program, Institute of Health Sciences, Bahcesehir University, Istanbul, Turkey*

Nuclear factor-κB (NF κB) transcription factors represent a conserved family of proteins that regulate not only immune cells, but also heart cells, glial cells and neurons, playing a fundamental role in various cellular processes. Due to its dysregulation in certain cancer types as well as in chronic inflammation and autoimmune diseases, it has recently been appreciated as an important therapeutic target. The aim of this study was to investigate the binding pocket of NF κB (p50/p65) heterodimer complex in association with NF κB inhibitor IκBα to identify potent ligands via fragment-based e-pharmacophore screening. The ZINC Clean Fragments (~2 million) and the Schrodinger's medically relevant Glide fragments library (~670) were used to create the e-pharmacophore models at the potential binding site of the target which was validated by site mapping. Glide/HTVS docking was conducted followed by re-docking of the top 20% fragments by Glide/SP and Glide/XP protocols. The top-85000 Glide XP-docked fragments were used to generate the e-pharmacophore hypotheses. The Otava small molecule library (~260000 drug-like molecules) and additional 85 known NF κB inhibitors were screened against the derived e-pharmacophore models. The top-1000 high-scored molecules, which were well aligned to the e-pharmacophore models, from the Otava small molecule library, were then docked into the binding pocket. Finally, the selected 88 hit molecules and the 85 known inhibitors were analyzed by the MetaCore/MetaDrug™ platform, which uses developed binary QSAR models for therapeutic activity prediction as well as pharmacokinetic and toxicity profile predictions of screening molecules. Ligand selection criteria led to the refinement of 3 potent hit molecules using molecular dynamics (MD) simulations to better investigate their structural and dynamical profiles. The selected hit molecules had a low toxicity and a significant therapeutic potential for heart failure, antiviral activity, asthma and depression, all conditions in which NF κB plays a critical role. These hit ligands were also structurally stable at the NF κB/IκBα complex as per the MD simulations and MM/GBSA analysis. Two of these ligands (Otava IDs: *1426436* and *6248112*) were energetically more favorable and therefore are hypothesized to be more potent. Identifying new potent NF κB/IκBα inhibitors may thus present a novel therapy for inflammation-mediated conditions as well as cancer, facilitating more efficient research, and leading the way to future drug development efforts.

Introduction

The transcription factor nuclear factor kappa B (NF κB) has now been appreciated as a master regulator of the body's innate and adaptive immune systems[1]. It represents a family of transcription factors that has played a fundamental role in the defense and protection of *homo sapiens* against injury and infections throughout evolution[1]. First discovered in B cells, NF κB has proved to be an important family of conserved transcription factors that plays a critical role for the functioning of not only immune cells, but also heart cells, glial cells and neurons[2,3]. It has an essential role in various cellular processes including signaling, autophagy, cellular senescence, tissue regeneration and repair and cellular metabolism[2]. Triggering the NF κB classical/canonical signaling pathway can be activated by certain stimuli such as proinflammatory cytokines, tumor necrosis factor (TNF) and pathogen-associated molecular patterns (PAMPS) such as lipopolysaccharides in gram-negative bacteria[1,2,4]. A second alternative, the non-canonical pathway, which activates p52-RELB dimers of NF κB is triggered by different stimuli[2]. These signals lead to the phosphorylation of IκBα, targeting it for degradation by the ubiquitin-proteasome pathway[5,6]. Consequently, the p50-p65 subunits of NF κB, which are mainly found in the cytoplasm bound to IκB proteins in its latent state, are free to translocate to the nucleus and bind DNA to enhance the expression of over 500 genes[1,6,7]. The complexity of NF κB and its variable functions in different cell types is perhaps mediated by the crosstalk with other signaling pathways, signaling proteins and other transcription factors[2].

There is now clear evidence about the pivotal role of NF κB in linking chronic inflammation and persistent infections to the increased risk of cancer upregulation and development[2,8]. NF κB is dysregulated not only in certain cancers[9-11], but also in chronic inflammation, autoimmune diseases as well as neurodegenerative and heart diseases[2,12,13]. While the differential and multi-faceted role of NF κB in the heart is complex and currently under investigation, it has been found that prolonged NF κB activation is associated with heart failure by promoting chronic inflammation, endoplasmic reticulum stress responses, apoptosis and adverse remodeling post-myocardial infarction[14,15] as well as the development of inflammatory cardiomyopathy[16]. NF κB is released cyclically in the heart and thus, it is also important for cardiac pathophysiology[16].

Various research studies have, on the other hand, showed that corticosteroids, mediate their strong anti-inflammatory effect via NF κB inhibition, potentially by increasing IκBα and thus inhibiting NF κB translocation to the nucleus, however this is still not clearly understood [7,17,18]. NF κB is majorly activated in asthma in response to IL-1 and TNFα; exacerbation of asthma attacks may be related to induced activation of NF κB via the stimulation of TLRs by bacterial or viral infections[18]. A variety of antioxidants such as vitamin E derivates can also inhibit NF κB activation[17]. Getting insight into the complexity of NF κB signaling pathways thus requires an integration of many biochemical, molecular and computational studies conducted so far since its discovery about 30 years ago.

Finding potent inhibitors that could act more selectively in the NF κB pathway may present a novel therapeutic strategy for cancer as well as other inflammatory conditions and lead the way to future drug design efforts and experimental studies. The complexity of NF-κB signaling in inflammation and cancer has been comprehensively reviewed by Hoesel and Schmid[19]. NF κB inhibitors represent a new therapeutic alternative and may also play a significant role in treating corticosteroid-resistant asthma and COPD[20]. While a large number of NF-κB inhibitors are known[13], compounds specifically designed as NF-κB inhibitors are not in clinical use yet, but will be addressed as treatments for certain cancers, neurodegenerative and inflammatory diseases. NF-κB activation can be inhibited by classical chemotherapeutic IKKβ-selective inhibitors[21]. Also, certain natural compounds like plant-derived substances have been evaluated as inhibitors of the NF-κB pathway[22].

Few molecules that target NF κB have been tested so far; among these, DHMEQ, a known potent NF κB inhibitor, was shown to reduce eosinophil-mediated airway inflammation and remodeling in experimental mice models of asthma[23]. In addition to asthma and COPD, NF κB has also been found to play a critical role in depressive behavior mediated by acute and chronic stress and IL-1β and IL-6 signaling[24-26], highlighting the importance of NF κB signaling in neuronal survival, development, growth and degeneration, circadian sleep rhythms and mood-related behavior[12,26,27]. Administration of an NF κB inhibitor blocked the inhibition of neurogenesis in the hippocampus which is responsible for the pro-depressive behaviors, particularly anhedonia[24]. Recently, it is shown that inhibition of NF κB was effective at restoring the sensitivity of drug-resistant prostate cancer to current anti-androgen therapy[28].

3

NF κB dysregulation has also been associated with various other conditions such as AIDS, viral infections, inflammatory bowel disease and arthritis[29].

Targeting the NF-κB pathway by computational approaches has so far been mostly addressed by mathematical, dynamic pathway modeling using ordinary differential equations (ODEs) of the signaling pathway[30], but the structural biology has not been fully exploited yet. Computational studies have proven to play an integral role in linking theoretical biophysics to the fields of experimental cell biology, clinical pharmacology and medicine. Molecular docking has been appreciated as a powerful tool in structure-based drug discovery[31]. In addition, pharmacophore modeling has become an essential part of lead discovery, an area that is evolving rapidly and used extensively by academic and industry researchers[32]. This method provides a way to elucidate chemical features that are common and essential for the biological activity of high affinity ligands[33,34]. A fragment-based energy-optimized pharmacophore (e-pharmacophore) approach was developed which combines both the computational efficiency of virtual pharmacophore library screening as well as the accuracy of structure-based molecular docking[33]. In comparison to ligand-based methods, fragment-based pharmacophore models are built based on chemical fragments opposed to ligands[34]. This is particularly advantageous as prior information on known active ligands at the target complex is not necessary, thereby providing an unbiased way to explore challenging target proteins via thorough sampling of the chemical space[32]. The fragments are docked onto the protein complex and common chemical features which maximize the interactions that satisfy both positions and directions are discovered[35]. Once these pharmacophore models are generated, they can be used for database screening, hit identification, and as a framework for future drug discovery efforts.

Literature provide only few details on the specific mechanism of action and interaction sites of known inhibitors in the NF κB pathway. While there are now more than 785 known NF κB inhibitors, these inhibitors target different steps in the pathway including IkB kinases (IKK), IκBα, the cytoplasmic retention of the NF κB/IκBα complex, the nuclear translocation and DNA binding[13]. Therefore, a fragment-based pharmacophore model generation is an appropriate approach for investigating and analyzing small molecule binding to the NF κB/IκBα complex. Thus, herein we present an *in silico* study to investigate the NF κB/IκBα p50/p65 (RelA) heterodimer complex to discover potent ligands with strong binding affinities via fragment-based virtual e-pharmacophore screening, molecular docking and molecular dynamics (MD)

simulations followed by therapeutic activity, pharmacokinetic and toxicity predictions using a binary QSAR platform (MetaCore/MetaDrug).

Methods

Protein Preparation

There is no full-length NF κB/IκBα (p50/p65) complex structure from protein crystallography. Available structures of the protein-protein complex are missing the N-terminal signal receiving domain (SRD) which is essential in the non-canonical activation pathway by IKK. The SRD displays an additional 1.5 ankyrin repeat units which additionally stabilize the complex. The full length protein-protein complex model was taken from our previous study[36] and prepared using the protein preparation module of Maestro molecular modeling tool[37] to add hydrogen atoms, fix side chains and loops and to generate disulfide bonds[38]. The protonation states of amino acids at pH 7.4 were generated using PROPKA to mimic physiological conditions[39,40]. The OPLS3 forcefield was finally used for the structural optimization of the protein[41].

Fragment and Ligand Preparation

The ZINC Clean Fragments library (~2 million fragments) and the Schrodinger medically relevant Glide Fragment Library (~667) were downloaded from *zinc.docking.org/subsets/clean-fragments* and *www.schrodinger.com/glide#block-2974*, respectively. The Otava Drug-Like Green Collection (~176000 small molecules) was downloaded from http://www.otavachemicals.com/. These downloaded libraries were prepared using the LigPrep module of the Maestro molecular modeling package with the OPLS3 forcefield[41,42] (after ligand preparation, total number of molecules reached to ~260000 due to different combinations of enantiomers and tautomers of molecules). Epik was used to apply the protonation states to the fragments and ligands at pH 7.4[43,44]. Sampling of large rings was enhanced using MacroModel[45].

Site Mapping

SiteMap module of Maestro was used to delineate the protein binding sites of the complex target[46,47]. A grid was placed over the whole target protein by SiteMap. Vertices were allocated inside concavities, but not in the protein itself, and were called as site points. Furthermore, all the points present in the neighborhood were clustered to characterize a binding site. Hydrophilic and

hydrophobic surfaces were mapped out by electrostatic and van der Waals probes present at site points. The regions of hydrogen bonds and metal chelation were also accommodated. Druggable sites were identified according to the SiteScore. The SiteScore is an empirical function containing weighted sum of hydrogen bonding, enclosure/exposure, contact, and hydrophilic/hydrophobic terms. According to the knowledgebase metrics in SiteMap, a SiteScore of > 1.0 is associated with a site which is considered as druggable; a score between 0.8 and 1.0 indicates that the region is difficult for binding, and a score below 0.8 denotes a site which is considered as non-druggable[48]. On the basis of Sitescores (>1.0), top-binding sites were selected for screening of compounds. The potential protein binding sites were identified using SiteMap via a more restrictive definition of hydrophobicity and standard grid[47]. At least 15 site points per reported site were used as criteria for the definition of the binding sites.

Molecular Docking

Algorithms used in the docking studies included Glide/HTVS, Glide/SP and Glide/XP in Maestro with flexible ligand sampling[49-52]. A receptor grid box was generated and certain amino acids were allowed to rotate their side chains to add flexibility to the used target: Ser51, Thr78, Ser180, Ser240, Ser276, Thr219, Thr247, Tyr251, Ser252, Tyr254, Thr257, Thr273, Ser283, Ser288, Tyr289, Thr291 and Glu292.

Both the ZINC Clean Fragments (~2 million) and the Schrodinger's Glide Fragments (~670) libraries were used to create the e-pharmacophore models at the potential site found from SiteMap. Glide's high-throughput virtual screening (HTVS) docking module was conducted followed by re-docking of the top 20% scoring fragments using Glide's standard precision (Glide/SP) then extra precision (Glide/XP). Input partial charges were used; scaling factor of van der Waals radii was set as default as 0.80 with a partial charge cutoff of 0.15. Glide/XP docking used flexible ligand sampling; nitrogen inversions and ring conformations were sampled. Post-docking minimization was performed. Non-planar conformations were penalized. The Glide/XP docking was also conducted with expanded sampling with modified settings as described by Loving et al[33]. Accordingly, the number of poses per ligand for the initial phase of docking was set to 50000. The scoring window for keeping the initial poses was set as 500, and the best 1000 poses per ligand were kept for energy minimization[33]. The energy minimization protocol used a distance-dependent dielectric constant ε of 2 and 100 maximum minimization steps. OPLS3 force field was used for these calculations[41].

Fragment Library-Based e-Pharmacophore Model Generation

The use of modified Glide XP-docking allowed us to thoroughly sample all of the fragments in the binding sites while keeping the top-1000 poses for each fragment[33]. The hypotheses were then generated using PHASE based on the e-pharmacophore model generation protocol performed by the post-processing docking script which was previously described by Salam and Dixon[34,53]. Essentially, in this approach, energies derived from Glide/XP scores are mapped onto all atoms that define each pharmacophore chemical feature[53]. Accordingly, the pharmacophore sites are ranked and the best-scoring, most favorable sites are thus identified. The top-85000 Glide XP-docked fragments were used to generate the e-pharmacophore hypotheses. The hypothesis generation also applies rules to prevent overlapping features – any two features are set at least 2 Å apart, while 4 Å is set as the minimum distance if two features are of the same type. For fragment-based pharmacophore model generation, six chemical features were analyzed in the hypothetical binding pocket: hydrogen bond acceptor (A), hydrogen bond donor (D), hydrophobic (H), negative ionizable (N), positive ionizable (P), and aromatic ring (R) features. The hydrogen bond acceptors and donors were assigned based on vectors or pure projected points, leading to different vector-based and projected points-based hypotheses[53]. In the projected points approach, hydrogen bonds can be made by dissimilar active molecules regardless of their direction and point of origin, providing a more flexible approach[53].

E-Pharmacophore-based Virtual Screening

The Otava Drug-Like Green collection includes around 176000 compounds which fits Lipinski's rule of 5 (i.e., logP from -1 to 5, molecular weight from 160 to 500 Da, number of H-bond donors from 0 to 5, number of H-bond acceptors from 0 to 10, and number of NO_2 groups from 0 to 2). Compounds with reactive groups, biologically unstable compounds, and compounds containing any atom different than O, N, C, H, Br, Cl, F, or S are removed. These 176000 compounds were enriched to around ~260000 after ligand preparation (due to different combinations of enantiomers and tautomers of molecules). Thus, ~260000 small molecules from the Otava library and 85 known NF κB inhibitors were screened at each of the derived e-pharmacophore models. The known NF κB inhibitors from previous studies were included as a positive control group. The NF κB inhibitors database compiled by Boston University (available from *http://www.bu.edu/NF κB/physiological-mediators/inhibitors/*) was used for this purpose. The chemical structures were downloaded from PubChem and other open-source chemistry

online resources. All known inhibitors and molecules were prepared using Maestro's LigPrep module, and energy-minimized using the previously described algorithms[42]. Only ligands that matched at least 4 chemical features in each e-pharmacophore model were considered for the next stage in analysis. The top-1000 screened molecules based on fitness score from Otava were later docked onto the binding site using Glide/SP and Glide/XP protocols for further investigation.

MetaCore/ MetaDrug Analysis

The structure of the top 88 screened molecules from the e-pharmacophore screening from Otava, based on fitness and docking scores, and 85 known inhibitors were submitted to MetaCore/MetaDrug for analysis of therapeutic activity properties, metabolites, pharmacokinetic properties and toxicity effects. MetaCore/MetaDrug provides a comprehensive tool to analyze compounds and their biochemical and pharmacological behaviors in the field of drug design and development. MetaCore/MetaDrug can predict both the first-pass (pre-systemic elimination which describes the metabolism of drugs before it reaches the systemic circulation) and the second pass metabolites. In addition, phase 1 and phase 2 metabolites are predicted by using a database which contains more than 10000 xenobiotic reactions, and more than 2500 combined substrates and enzyme inhibitors. There are also over 89 rules used to predict whether the metabolites are reactive or not. MetaDrug uses the property of *Tanimoto Prioritization* (TP) to find the similarity between analyzed compounds and compound sets in the quantitative structure-activity relationships (QSAR) models based on elements found in the structure. These models were prepared with a diverse set of compounds based on experimental evidence of their activity/function on a certain protein of interest, and then tested with validation sets. The accuracy of each model depends on the number of compounds used to create it and can be estimated by the correlation coefficient (R^2) and root mean squared error (RMSE), where a higher R^2 and low RMSE indicate higher model accuracy. The QSAR model with the highest specificity, sensitivity, accuracy and the Matthews Correlation Coefficient (MCC) was selected in MetaDrug for each particular activity or toxicity tested. The prediction of a therapeutic activity or toxic effect is calculated based on the ChemTree ability to correlate structural descriptors to that property using the recursive partitioning algorithm. The ChemTree parameters that gave the best results were as follows: path length – 5, max segments – 3, p-value threshold Bonnferoni – 0.99, p-value multiway split – 0.99 and number of random trees – 50. The training

set used in MetaCore/MetaDrug includes molecules that possess the property (positives) and chemicals that do not have such property (negatives) in approximately equal numbers. The marketed drugs were used if their number was greater than 100 in the disease QSAR models; drug candidates in clinical trials and preclinical compounds with *in vivo* activity have been added to the training set. The drugs that have been annotated to cause a particular toxic effect were used for the prediction of toxic effects.

Molecular Dynamics (MD) simulations

The criteria to select screened ligands to advance into the MD simulations phase was not only based on docking scores but also on fitness scores, pharmacophore features, low toxicity (i.e., selecting the compounds that show no toxicity in any of the 26 different toxicity models, however we also further investigated molecules that have high therapeutic activity (>0.75) and showing low toxicity (values not greater than 0.60) in 1 to 3 models). Three candidate ligands were thus filtered from the Otava Drug-Like Green Collection (Compound IDs: *6248112, 7132624,* and *1426436*). Classical MD simulations were used to study the structural and dynamical behavior of the NF‑κB/IκBα complex bound to the known and three selected hit ligands. This exploration of the conformational behavior of the screened ligands with the complex is the key to understanding their biological, chemical and physical behaviors and structural stability at the binding cavity. MD simulations were performed using the Desmond program with the OPLS2005 force field and RESPA integrator[54,55]. Explicit water molecules (SPC) and 0.15 M NaCl ion concentration were used to prepare the system and neutralize the complex. The total number of atoms was between 82228 and 103209 atoms at the simulation boxes. The NPT ensemble at 310 K with Nose-Hoover temperature coupling[56] and at constant pressure of 1.01 bar via Martyna–Tobias–Klein pressure coupling[57] were used. Other settings were used as default. Trajectory analysis was carried out for 100 ns with 2 fs time steps from each of the MD simulations. Triplicate MD simulations were performed for the most potent ligands *1426436* and *6248112*.

The Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) Continuum Solvation Calculations

The MM/GBSA approach has been proven to be a feasible approach to predict the free binding energy of biological systems, thus allowing a post-MD evaluation of the protein-protein and protein-ligand interactions[58]. This method was implemented to study our structure with DHMEQ

and with each of the novel hit molecules. The energetic calculations were performed using Schrodinger's Prime module. The complete details and applicable thermodynamic equations were described by Miller *et al*.[58] For this purpose, 51 trajectory frames were considered from the last 50 ns of the MD simulations. The OPLS2005 forcefield and VSGB 2.0 solvation model[59] were applied. Triplicate MM/GBSA analysis were conducted for the most potent ligands *1426436* and *6248112*. The study's full methodology is summarized in Scheme 1.

Results and Discussion

Protein Structure Analysis

The NF κB/IκBα complex investigated in this study is a heterodimer complex made up of the p50 and p65 subunits mostly found in the cytoplasm (Figure 1). Its N-terminal domain (NTD) and C-terminal domain (CTD) are part of the conserved Rel homology region (RHR) necessary for dimerization, nuclear translocation and DNA and IκB binding[4]. While NTD is important for specific interactions with DNA bases and non-specific interactions with the phosphate backbone, CTD is important for dimerization and also contacts DNA non-specifically[1]. All IkB proteins have ankyrin repeats (ARs) which are key for interacting with NF κB proteins. ARs 4, 5 and 6 interact with the RHR-CTD of NF-κB, while AR-6 and the C-terminal PEST residues interact with the p65 RHR-NTD[4].

Bound IκBα to NF κB has a half-life of more than 24 hours, suggesting the extremely high binding energy, compared to less than 10 minutes for the free unbound IκBα[60]. Free IκBα degradation requires no phosphorylation or ubiquitination as is required for the bound IκBα. Upon binding of NF κB, the previously highly dynamic ARs 5 and 6 form an organized folded structure, while the middle ARs such as AR-3 go from a well-folded state to a disordered state, implying compensation for the energy required for folding[60,61]. Analysis of various IkB mutant proteins shows that the process of NF κB inhibition requires the C-terminal PEST sequence and the weakly folded ARs 5 and 6[4].

Structural and experimental studies have shown that there are two key regions at opposite ends of the interface between NF κB and IκBα which are important for their binding interaction energy[62]. The two regions that majorly contribute to the high binding affinity of IκBα and NF κB include the nuclear localization signal (NLS) sequence (residues 305–321 in NF κB

10

p65) which contacts the first ankyrin repeat of IκBα, as well as the IκBα PEST sequence (residues 276–287) which contacts the DNA-binding dimerization domain of NF κB (p65)[62,63]. Hence, the NF κB/IκBα large interface of more than 4000 Å^2 contains two hot spots at opposite ends which are particularly significant for their high binding affinity, suggesting a squeeze mechanism of binding that may add further to the stability of the ARs[62]. Mutation of residues in these regions, but not the contacting residues interestingly, leads to a significant decrease of more than 500-fold and 5000-fold in the binding energy for the NLS and PEST sequences, respectively[62]. This indicates that the PEST sequence (residues 276–287) is more important for identifying the high affinity ligands. The PEST sequence is a negatively charged acidic sequence found at the carboxyl end of IκBα and is rich in Pro, Glu, Ser and Thr residues[63]. The PEST sequence is also known to be a site for post-translational phosphorylation by casein kinase II. However, it was found that phosphorylation of the PEST sequence is not critical for the binding affinity, but only slightly increase the association rate between NF κB and IκBα by perhaps enhancing the folding mechanism of binding[64]. Casein kinase II is known to phosphorylate residues Ser283, Ser288, Thr291, Ser293, and Ser299[64].

The interaction between NF κB and IκBα is limited here to the binding site containing the C-terminal PEST sequence (Figure 1). SiteMap from Maestro which has been effectively used to find target sites that have potential biological activity was used for confirmation. SiteMap has a correct identification of known binding sites in more than 96% of the cases, while more than 98% were chosen based on tight binding[46,65]. A SiteScore of more than 1 suggests a site of promising biological activity[46]. When combined with literature findings from experimental work discussed above, this suggested a highly promising potential binding site. While two sites were found to have a score of over 1, Site-2 with a score of 1.099 also contained most of the amino acid residues found in the PEST sequence (Figure 2). Also, although site-1 (SiteScore: 1.006) had a larger volume than site-2, the structure of the binding site was more diffuse opposed to a more pocket shape in site-2. Thus, site-2 is mainly considered for further investigations.

Fragment-based e-Pharmacophore Modeling Studies

Two fragment libraries were used to create the e-pharmacophore models at site-2 of NF κB/IκBα complex which was given by SiteMap. The e-pharmacophore hypothesis models were created using the top 85000 Glide/XP docked fragments. The hypothesis models were generated via both the *vector-based* and the *projected points donor* methodologies (Figures 3 and

4). Both the ZINC Clean Lead library (around 2 million) and the Schrodinger's medical fragment library (667) were used. Glide/HTVS was conducted by docking the top 20% fragments in Glide SP then XP modes. Excluded volumes which are shown as blue spheres, in Figures 3, 4 and Figure S1 in the supplementary information, are part of the derived binding site models. These excluded volumes help to reduce false positives by preventing overlapping and clashing of fragments with the protein complex, as well as eliminate inactive molecules that cannot match the features of the hypothesis[53]. Nevertheless, a loose excluded volume surface still allows small overlaps in order to account for more flexibility and better induced fit of the molecules with the binding site[53].

MetaCore/ MetaDrug Analysis

The hunt for selective NF κB inhibitors is expensive and spans various fields of research highlighting the relevance of this protein as a therapeutic target. However, these inhibitors face various challenges such as specificity and potential toxicities as they may interfere with other pathways and cellular processes regulated by NF κB. For this reason, we have decided to evaluate chemical properties, possible metabolites, potential therapeutic values and toxicities of selected potent ligands. This allowed us to get a thorough understanding and prediction of their biological activity when compared to hundreds of other drugs using MetaCore from Clarivate Analytics. With a cutoff value for therapeutic activity of 0.5, the selected hit molecules showed a significant therapeutic potential for heart failure, viral infections, asthma and depression, conditions in which NF κB plays a strong role (Figure 5). These ligands also showed the potential to be effective (predicted value (PV) >0.50) for allergy, cancer, migraine, pain, depression, schizophrenia, psoriasis, thrombosis, hypertension and Parkinson's and Alzheimer's diseases, all conditions which were not shown for the control inhibitor, DHMEQ (Figure 5). This molecule has been shown to be effectively used for asthma[23]; the PV from MetaCore was 0.64. Our selected ligands had a higher prediction for their potential use for asthma (*training set N=366, test set N=63, sensitivity= 0.92, specificity=0.86, accuracy=0.89, MCC=0.78*) with ligand *7132624* being the best with a therapeutic activity PV of 0.84, followed by molecules *1426436* (PV of 0.77) and *6248112* (PV of 0.72). When using a high cutoff value of 0.70, our selected ligands were shown to be better in asthma, depression, heart failure, hypertension, migraine, obesity and viral diseases. Ligand *6248112* had the highest therapeutic value of 0.88 for heart failure (*training set N=204, test set N=33, sensitivity= 0.78, specificity=0.87,*

accuracy=0.82, MCC=0.64). Ligand *7132624* was best for asthma (PV of 0.84). Ligand *1426436* was best equally for depression (*training set N=335, test set N=62, sensitivity= 0.93, specificity=0.82, accuracy=0.87, MCC=0.75*) and antiviral activity *(training set N=206, test set N=35, sensitivity= 0.92, specificity=0.95, accuracy=0.94, MCC=0.88)* with a predicted value of 0.83, and also showed potential against obesity with a PV of 0.96 *(training set N=472, test set N=75, sensitivity= 0.89, specificity=0.97, accuracy=0.93, MCC=0.87).*

Two of the ligands (*6248112* and *7132624*) were shown to result in metabolites that contain no reactive groups (Figure 6). Only ligand *1426436* had reactive metabolites, which suggest that it may participate in side interactions other than the intended target, possibly contributing to its activity. In comparison with these molecules, the positive control molecule DHMEQ also gave reactive metabolites. All ligands including the positive control met Lipinski's rule of 5 so they are all likely to be orally bioavailable (Figure 6). However, it should be noted that the control molecule had significantly higher toxicities, both in terms of types and values when compared to other molecules in the QSAR models, including kidney necrosis, liver necrosis, kidney weight gain, nephron injury, liver cholestasis, carcinogenicity in male rats, neurotoxicity and carcinogenicity in male mice, with values ranging from 0.51 to 0.78 in decreasing order (Figure 6). The selected potent ligands, on the other hand, had significantly lower toxicity effects and lower predicted toxicity values. These toxicities were not higher than 0.56 when excluding AMES (*Model description: N=1780, R^2=0.69, RMSE=0.29*). Ligand *6248112* had predicted cardiotoxicity (PV of 0.51) and genotoxicity (PV of 0.55) and an AMES (PV of 0.57). Ligand *7132624* had only a relatively low predicted kidney necrosis toxicity effect (PV of 0.53) and had an AMES score of 0.61. Ligand *1426436*, on the other hand, was shown to be related to carcinogenicity in male mice models (PV of 0.56) and male rat models (PV of 0.53) and liver cholestasis (PV of 0.52). In comparison with the selected hit ligands, the control molecule had a predicted higher toxicity in kidney necrosis (PV of 0.78), liver cholestasis (PV of 0.56) and carcinogenicity in male rat models (PV of 0.56). Additionally, different types of toxicities which were not common among the selected hits, but which were observed in the control molecule included liver necrosis (PV of 0.78), kidney weight gain (PV of 0.73), nephron injury (PV of 0.63) and neurotoxicity (PV of 0.51) as shown in Figure 6.

First and second pass major metabolites by two of the selected hit ligands and control molecule were also re-docked using Glide/SP and analyzed via MetaDrug (Figure 6). The metabolites

were shown to have docking scores that ranged between -5.0 to -7.9 kcal/mol for ligand *7132624*, -4.7 to -8.1 kcal/mol for ligand *1426436* and -5.4 to -7.2 to kcal/mol for ligand *6248112*. All of these ligands and their metabolites had a higher docking score compared to the control molecule, suggesting their greater biological activity at the NF κB/IκBα complex (Figure S2, supplementary material). Docking scores were highest for the ligand *1426436* (-8.68 kcal/ mol) followed by *6248112* (-6.41 kcal/ mol) and *7132624* (-6.10 kcal/ mol). The docking score of the positive control molecule was -4.26 kcal/mol (Figure S2). The pharmacokinetic properties and chemical structures of the three hit ligands as well as the reference (positive control) molecule were shown in Figure 7.

Conformational Analysis and Binding Energy Calculations

The root mean squared deviation (RMSD) is primarily used to investigate the structural stability of the biological system throughout the MD simulations. An essential part of our research efforts is to use RMSD terms to elucidate structural and dynamical properties, allowing us to explore the structural stability of binding pocket residues as well as used ligands throughout the MD simulations. The RMSD values of Cα atoms away from the initial positions are used to describe the flexibility of all possible protein conformers along the MD simulations. The "fit on protein/profit" and "fit on ligand/ligfit" modes respectively represent the ligand's translational and rotational motion in the binding pocket. The lead ligands *1426436*, *6248112* and *7132624* fluctuated with mean Cα RMSD values of 4.39, 4.21 and 5.26 Å, respectively (Figure 8). Of these ligands, it is important to note that the NF κB/IκBα protein complex was more structurally stable with ligand *1426436* for the majority of the simulation time and especially for the last 45 ns. Although the reference molecule had a mean RMSD of 4.47 Å, there were significant fluctuations evident from an increase of RMSDs from 1.54 to 5.88 Å over the 100 ns run time. Figure 8 may also suggest that ligand *713262* slightly destabilizes the protein complex throughout the simulation. Therefore, based on this analysis, it can be suggested that ligands *6248112* and *1426436* are more energetically favorable at the binding pocket.

In the *profit* mode, the mean values of RMSDs for ligands *1426436*, *6248112* and *7132624* were found to be 5.75, 7.43 and 8.85 Å as shown in Figure S3 in the supplementary information, respectively. It is evident that ligands *6248112* and *1426436* have stable conformations in the last 60 ns of the MD simulations. It can also be concluded that molecule *1426436* had the highest conformational stability compared to the other hit ligands and the reference molecule. Although

14

the reference molecule has average RMSD of 4.68 Å, it has quite sharp changes in RMSD values (8 to 16 Å) especially between 20 to 40 ns simulation time and it has larger RMSD values in the last 10 ns. Moreover, molecule *1426436* had the least fluctuations and maintained a stable conformation over the last 60 ns. Ligand *6248112* seemed more stable in the last 60 ns of the simulation, although here, the RMSDs continued to increase prior to reaching a stable position. Ligand *7132624* had the highest fluctuations in RMSD values, indicating its lower stability, which is in line with the previously discussed results. In the *ligfit* mode shown in Figure S4 supplementary information, all studied ligands have RMSD values less than 2 Å, which may represent that rotational motions of the molecules are limited at the binding pocket, implying that their internal conformation was not changed dramatically throughout the simulation. Hit ligands *6248112* and *7132624* appeared significantly stable with minimal fluctuations (respective average RMSD values were 0.50 and 0.68 Å) when in comparison to ligand *1426436* (average RMSD, 1.75 Å) and the reference molecule (average RMSD, 0.57 Å).

Ligand *1426436* made significant interactions with Leu277, Glu282, Glu284, Asp290 and Thr291 that were present for 93%, 130%, 96%, 97% and 35% of the MD simulation time respectively via hydrogen bonds as well as via water bridges (Figure 9). The fact that Glu284 maintained interactions for over 100% is due to the fact that it made multiple hydrogen bonds one for 98% and the other for 32% of the simulation time as seen in Figure 9. Another feature about this ligand is that it made ionic interactions with Glu284. The torsion of the ligand's rotatable bonds in close proximity to Glu282 and Glu284 maintained a prominent angled conformation, giving insight about the changes the ligand underwent at the binding pocket (Figure S9). The interactions with the residues predicted by the docking protocols were mostly maintained in the MD simulations (Figure 9 and Figure S5, supplementary information); these were Glu282, Glu284 and Thr291. Other residues that proved to be significant (>10% of the interaction fraction) for the ligand-protein interaction were Lys79 and Arg158, present outside of the PEST sequence, indicating the potential role that these residues may play in the mechanism of this ligand. Hydrophobic interactions were made with Met279 throughout the 100 ns run time. For ligand *7132624*, significant contacts were maintained with Glu282 for about 45% of the 100 ns MD simulations via hydrogen bonding. Interactions with the PEST sequence was primarily made via hydrogen bonding, hydrophobic contacts and water bridges (Figure 9). The ligand also maintained a prominent angled conformation, evident from the torsion of the ligand's rotatable

15

bonds in close proximity to Glu282, giving insight about the changes this ligand underwent at the binding pocket (Figure S9). Other residues which the hit ligand *7132624* made contact with for 10% to 19% of the run time were in decreasing interaction fractions: Gln255, Ser252, Arg245 and Asn244. In addition, unlike the other ligands, ligand *7132624* had intramolecular hydrogen bonds which may contribute to its binding affinity. The docking results predicted the role of Asn276, Glu284, Gln255, Asp290 and Lys79, all of which played a role in the protein-ligand interaction in the MD simulation (Figure 9 and Figure S5, supplementary information). Significant interactions (> 20% of interaction time) by this ligand with Lys221, Asn244, Val246, Ser252, Gln255 and Glu282 were evident throughout the simulations.

Ligand *6248112* made interactions with the PEST sequence that were primarily made via water bridges, hydrogen bonding, hydrophobic contacts as well as ionic bonding with Asn276 (Figure 9). Additionally, the ligand made significant contacts with Arg158 and Gln255 for over 10% and 35% of the simulation time, respectively. The torsion of the ligand's rotatable bonds in close proximity to these residues maintained a prominent angled conformation, giving insight about the changes the ligand underwent at the binding pocket (Figure S6). The docking results predicted the role of Leu277, Gln278, Thr291 and Gln255 (Figure S5, supplementary information). Significant interactions (> 10% interaction fraction), however, were noted for Arg158, Asn276 and Met279 evident throughout the simulation.

DHMEQ, on the other hand, made interactions only with Leu277 for 42% of the simulation time (Figure 9). Significant interactions (> 20% interaction fraction) were noted for Lys79, Gly250, Gln255, Asn276, Leu277 and Met279 evident throughout the simulation. However, the highest interaction fraction was made for 60% of the run time compared to hit ligand *1426436* in which the interaction fraction with Glu282 was 130%.

The selected hit molecules were stable throughout the MD simulations and have shown persistent interactions with the PEST sequence. Interestingly, residues Lys79, Asp290, and Thr291 were important for binding at the identified binding site when analyzing both docking and MD simulation studies. This indicates the probable role of these residues for the binding interactions between NF☐κB and IκBα.

Analysis of the MD simulations in different parameters was shown in Figures S7-S9 in the supplementary information. Figure S7 shows RMSD evolution over time for the side chains over a 100 ns MD simulation with the NF☐κB/IκBα (p50/p65) complex. Highest side chain RMSD

16

values were obtained in the case of presence of ligand *7132624*, which may represent the perturbation of side chain residues at the binding pocket for the construction/re-construction of binding interactions with the ligand. Ligands *1426436* and *6248112* showed similar side chain RMSD values with apo-form for the last 10 ns MD simulations. Figure S8 shows root mean square fluctuations (RMSF) evolution over time for the C_α atoms and side chains over a 100 ns MD simulation with the NF□κB/IκBα (p50/p65) complex. Reference ligand DHMEQ and hit ligand *7132624* showed slightly higher fluctuations at the binding pocket and significant higher fluctuations compared to others especially for the region of residue numbers 550 to 650. Figure S9 represents solvent accessible surface area (SASA) in Å^2 of the three ligands over 100 ns MD simulations with the NF□κB/IκBα (p50/p65) complex. Reference ligand DHMEQ and hit ligand *7132624* showed higher SASA values compared to other ligands.

RMSD and RMSF graphs which belong only binding pocket were also investigated. (Figures S10 and S11) Residues in binding region shows smaller fluctuations compared to whole protein. Apo-form in the binding pocket has slightly larger RMSD values compared to other ligand-bound systems may indicates the structural stabilities of binding pocket residues with the presence of ligand.

The post-MD simulations MM/GBSA calculation were performed to determine the Gibbs free energy changes of a reference inhibitor and the 3 hit ligands from our screening protocols. Figure 10 shows that ligand *6248112* had the lowest average ΔG of −50.54± 4.30 kcal/mol over the last 50 ns MD simulations, suggesting a more stable ligand conformation and a higher affinity at the binding site. Similarly, ligand *1426436* had an average ΔG of −49.54± 2.63 kcal/mol over the last half of the MD simulations. Ligand *7132624*, on the other hand, had a lower average ΔG (−36.92 ± 4.89 kcal/mol), indicating its lower structural stability and affinity compared to the other ligands. The reference ligand DHMEQ had an average ΔG value of −16.70± 6.60 kcal/mol which shows that the selected three hit ligands were more energetically favorable. This adds another level of evidence to indicate the higher stability of our lead ligands at the NF κB/IκBα complex.

Conclusions

Inhibition of the NF□κB pathway is a highly promising strategy for therapeutic use since it presents the link between inflammation and cancer. It is a target for cancer, inflammation,

autoimmune diseases and even viral infections. A growing list of known inhibitors is available in the literature, and the number of experiments is expanding, however, there is no clear consensus as to the mechanism or detailed biochemical interactions with the NF☐κB/IκBα complex. In this study, the complex of NF☐κB and its inhibitor IκBα were targeted and analyzed in detail. To conduct an in-depth *in-situ* analysis of this complex, it was essential to utilize a full-length protein-protein complex structure and known binding ligands to generate energetic-optimized e-pharmacophore models. This method creates an energetic map emphasizing the potential interactions of the complex by utilizing fragments of molecules. By eliminating the need of active ligands of the NF☐κB complex, fragment-based e-pharmacophore model generation was the most reasonable approach to conduct our study effectively. In addition, this method also provided us with detailed information about our complex at the biochemical level including the identification of potentially important amino acid residues, chemical bonds and features which may prove to be important for future drug design efforts and further wet lab investigations.

In this study, the main conceptual innovation is the integrated approach to drug discovery that combines molecular docking, MD simulations and QSAR models under a single umbrella. Each method has its own strengths and weaknesses, and when used alone, is not likely to yield very useful results. However, when these methods are combined with positive feedback loops, they enhance each other so that one is much more likely to obtain successful drug leads. In more detail, the conceptual and methodological innovations in this study are: (i) As it is stated, 2 hits were identified and validated using molecular modeling approaches against NF-κB/IκBα complex. Thus, this information can be used by medicinal chemists for the designing of new NF-κB inhibitor analogues with enhanced activity and other tailored properties. (ii) MD simulations were performed for the identified hit molecules from small-molecules library for better understanding of their dynamical and structural profiles throughout MD simulations. Thus, the most important structural and dynamical properties of both ligands and binding pocket residues were discussed. (iii) The study does not provides only hit molecules against NF-κB/IκBα complex, but also pharmacokinetic and toxicity profiles of all identified hit molecules were analyzed using MetaCore/MetaDrug platform. Metacore/MetaDrug is an integrated software package having manually curated biological information about small molecules. 25 different common disease QSAR models as well as 26 different toxicity-QSAR models were used for the estimation of therapeutic activity and toxicity profiles of the studied molecules, which we believe

18

this is the largest amount of therapeutic activity and toxicity models that are used in the literature.

This study highlights the master role of NF κB not only in cancer, but also in the pathogenesis of inflammatory-mediated diseases including cardiomyopathy, asthma, COPD and neurodegenerative diseases. NF κB inhibitors that act more selectively at the investigated binding pocket of NF κB/IκBα complex may thus present a promising therapy for the discussed inflammatory conditions and lead the way to future drug design efforts and experimental studies. Future efforts for studying the complex requires an integration of all the biochemical and computational studies in order to get an insight into the complexity of the NF κB pathways and molecular interactions. We developed e-pharmacophore models for the NF κB/IκBα complex based on 85000 fragments. The derived e-pharmacophore models were used in lead compound discovery and screening efforts. The Otava chemicals library containing around 260000 compounds was screened against the constructed e-pharmacophore models. The constructed models are a valuable computational tool that may facilitate more efficient research efforts in the search for potent NF κB inhibitors. The ligands with new scaffolds were identified with promising high therapeutic values for inflammatory conditions along with low toxicity effects when compared to known inhibitors from the literature. These compounds can now move to the next stage of drug design to be tested preclinically.

Acknowledgments

References

1. Napetschnig, J.; Wu, H., Molecular basis of NF-kappaB signaling. *Annu Rev Biophys* **2013,** *42*, 443-68.
2. Taniguchi, K.; Karin, M., NF-κB, inflammation, immunity and cancer: coming of age. *Nature Reviews Immunology* **2018**.
3. Gilmore, T. D.; Garbati, M. R., Inhibition of NF-kappaB signaling as a strategy in disease therapy. *Curr Top Microbiol Immunol* **2011,** *349*, 245-63.
4. Zheng, C.; Yin, Q.; Wu, H., Structural studies of NF-kappaB signaling. *Cell Res* **2011,** *21* (1), 183-95.
5. Huxford, T.; Ghosh, G., A structural guide to proteins of the NF-kappaB signaling module. *Cold Spring Harb Perspect Biol* **2009,** *1* (3), a000075.
6. Ferreiro, D. U.; Komives, E. A., Molecular mechanisms of system control of NF-kappaB signaling by IkappaBalpha. *Biochemistry* **2010,** *49* (8), 1560-7.
7. Gupta, S. C.; Sundaram, C.; Reuter, S.; Aggarwal, B. B., Inhibiting NF-kappaB activation by small molecules as a therapeutic strategy. *Biochim Biophys Acta* **2010,** *1799* (10-12), 775-87.
8. Ben-Neriah, Y.; Karin, M., Inflammation meets cancer, with NF-kappaB as the matchmaker. *Nat Immunol* **2011,** *12* (8), 715-23.
9. Vaisitti, T.; Gaudino, F.; Ouk, S.; Moscvin, M.; Vitale, N.; Serra, S.; Arruga, F.; Zakrzewski, J. L.; Liou, H.-C.; Allan, J. N.; Furman, R. R.; Deaglio, S., Targeting metabolism and survival in chronic lymphocytic leukemia and Richter syndrome cells by a novel NF-κB inhibitor. *Haematologica* **2017,** *102* (11), 1878-1889.
10. Lerebours, F.; Vacher, S.; Andrieu, C.; Espie, M.; Marty, M.; Lidereau, R.; Bieche, I., NF-kappa B genes have a major role in Inflammatory Breast Cancer. *BMC Cancer* **2008,** *8*, 41-41.
11. De Simone, V.; Franzè, E.; Ronchetti, G.; Colantoni, A.; Fantini, M. C.; Di Fusco, D.; Sica, G. S.; Sileri, P.; MacDonald, T. T.; Pallone, F.; Monteleone, G.; Stolfi, C., Th17-type cytokines, IL-6 and TNF-α synergistically activate STAT3 and NF-kB to promote colorectal cancer cell growth. *Oncogene* **2014,** *34*, 3493.
12. Mincheva-Tasheva, S.; Soler, R. M., NF-kappaB signaling pathways: role in nervous system physiology and pathology. *Neuroscientist* **2013,** *19* (2), 175-94.
13. Gilmore, T. D.; Herscovitch, M., Inhibitors of NF-kappaB signaling: 785 and counting. *Oncogene* **2006,** *25* (51), 6887-99.
14. Gordon, J. W.; Shaw, J. A.; Kirshenbaum, L. A., Multiple facets of NF-kappaB in the heart: to be or not to NF-kappaB. *Circ Res* **2011,** *108* (9), 1122-32.

15. Hamid, T.; Guo, S. Z.; Kingery, J. R.; Xiang, X.; Dawn, B.; Prabhu, S. D., Cardiomyocyte NF-κB p65 promotes adverse remodelling, apoptosis, and endoplasmic reticulum stress in heart failure. *Cardiovascular Research* **2011,** *89* (1), 129-138.

16. Maier, H. J.; Schips, T. G.; Wietelmann, A.; Krüger, M.; Brunner, C.; Sauter, M.; Klingel, K.; Böttger, T.; Braun, T.; Wirth, T., Cardiomyocyte-specific IκB kinase (IKK)/NF-κB activation induces reversible inflammatory cardiomyopathy and heart failure. *Proceedings of the National Academy of Sciences of the United States of America* **2012,** *109* (29), 11794-11799.

17. D'Acquisto, F.; May, M. J.; Ghosh, S., Inhibition of nuclear factor kappa B (NF-B): an emerging theme in anti-inflammatory therapies. *Mol Interv* **2002,** *2* (1), 22-35.

18. Schuliga, M., NF-kappaB Signaling in Chronic Inflammatory Airway Disease. *Biomolecules* **2015,** *5* (3), 1266-83.

19. Hoesel, B.; Schmid, J. A., The complexity of NF-κB signaling in inflammation and cancer. *Molecular Cancer* **2013,** *12*, 86-86.

20. Edwards, M. R.; Bartlett, N. W.; Clarke, D.; Birrell, M.; Belvisi, M.; Johnston, S. L., Targeting the NF-kappaB pathway in asthma and chronic obstructive pulmonary disease. *Pharmacol Ther* **2009,** *121* (1), 1-13.

21. Ota, F., Transcription Factor NF-κB Inhibitors as Single Therapeutic Agents or in Combination with Classical Chemotherapeutic Agents for the Treatment of Hematologic Malignancies. *Current Molecular Pharmacology* **2010,** *3* (3), 98-122.

22. Nguyen-Hai, N., Naturally Occurring NF-κB Inhibitors. *Mini-Reviews in Medicinal Chemistry* **2006,** *6* (8), 945-951.

23. Shimizu, K.; Konno, S.; Ozaki, M.; Umezawa, K.; Yamashita, K.; Todo, S.; Nishimura, M., Dehydroxymethylepoxyquinomicin (DHMEQ), a novel NF-kappaB inhibitor, inhibits allergic inflammation and airway remodelling in murine models of asthma. *Clin Exp Allergy* **2012,** *42* (8), 1273-81.

24. Koo, J. W.; Russo, S. J.; Ferguson, D.; Nestler, E. J.; Duman, R. S., Nuclear factor-kappaB is a critical mediator of stress-impaired neurogenesis and depressive behavior. *Proc Natl Acad Sci U S A* **2010,** *107* (6), 2669-74.

25. LaPlant, Q.; Chakravarty, S.; Vialou, V.; Mukherjee, S.; Koo, J. W.; Kalahasti, G.; Bradbury, K. R.; Taylor, S. V.; Maze, I.; Kumar, A.; Graham, A.; Birnbaum, S. G.; Krishnan, V.; Truong, H. T.; Neve, R. L.; Nestler, E. J.; Russo, S. J., Role of nuclear factor kappaB in ovarian hormone-mediated stress hypersensitivity in female mice. *Biol Psychiatry* **2009,** *65* (10), 874-80.

26. Monje, F. J.; Cabatic, M.; Divisch, I.; Kim, E.-J.; Herkner, K. R.; Binder, B. R.; Pollak, D. D., Constant Darkness Induces IL-6-Dependent Depression-Like Behavior through the NF-κB Signaling Pathway. *The Journal of Neuroscience* **2011,** *31* (25), 9075-9083.

27. Gutierrez, H.; Davies, A. M., Regulation of neural process growth, elaboration and structural plasticity by NF-kappaB. *Trends Neurosci* **2011,** *34* (6), 316-25.

28. Jin, R.; Yamashita, H.; Yu, X.; Wang, J.; Franco, O. E.; Wang, Y.; Hayward, S. W.; Matusik, R. J., Inhibition of NF-kappa B signaling restores responsiveness of castrate-resistant prostate cancer cells to anti-androgen treatment by decreasing androgen receptor-variant expression. *Oncogene* **2014,** *34*, 3700.

29. Gupta, S. C.; Sundaram, C.; Reuter, S.; Aggarwal, B. B., Inhibiting NF-κB Activation by Small Molecules As a Therapeutic Strategy. *Biochimica et biophysica acta* **2010,** *1799* (10-12), 775-787.

30. Oppelt, A.; Kaschek, D.; Huppelschoten, S.; Sison-Young, R.; Zhang, F.; Buck-Wiese, M.; Herrmann, F.; Malkusch, S.; Krüger, C. L.; Meub, M.; Merkt, B.; Zimmermann, L.; Schofield, A.; Jones, R. P.; Malik, H.; Schilling, M.; Heilemann, M.; van de Water, B.; Goldring, C. E.; Park, B. K.; Timmer, J.; Klingmüller, U., Model-based identification of TNFα-induced IKKβ-mediated and IκBα-mediated regulation of NFκB signal transduction as a tool to quantify the impact of drug-induced liver injury compounds. *NPJ Systems Biology and Applications* **2018,** *4*, 23.

31. Meng, X.-Y.; Zhang, H.-X.; Mezei, M.; Cui, M., Molecular Docking: A powerful approach for structure-based drug discovery. *Current computer-aided drug design* **2011,** *7* (2), 146-157.
32. Keseru, G. M.; Erlanson, D. A.; Ferenczy, G. G.; Hann, M. M.; Murray, C. W.; Pickett, S. D., Design Principles for Fragment Libraries: Maximizing the Value of Learnings from Pharma Fragment-Based Drug Discovery (FBDD) Programs for Use in Academia. *J Med Chem* **2016,** *59* (18), 8189-206.
33. Loving, K.; Salam, N. K.; Sherman, W., Energetic analysis of fragment docking and application to structure-based pharmacophore hypothesis generation. *J Comput Aided Mol Des* **2009,** *23* (8), 541-54.
34. Dixon, S. L.; Smondyrev, A. M.; Rao, S. N., PHASE: a novel approach to pharmacophore modeling and 3D database searching. *Chem Biol Drug Des* **2006,** *67* (5), 370-2.
35. Dixon, S. L.; Smondyrev, A. M.; Knoll, E. H.; Rao, S. N.; Shaw, D. E.; Friesner, R. A., PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J Comput Aided Mol Des* **2006,** *20* (10-11), 647-71.
36. Yazdi, S.; Durdagi, S.; Naumann, M.; Stein, M., Structural modeling of the N-terminal signal–receiving domain of IκBα. *Frontiers in Molecular Biosciences* **2015,** *2*, 32.
37. Schrodinger *Maestro*, LLC: New York, NY, 2016.
38. Sastry, G. M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W., Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des* **2013,** *27* (3), 221-34.
39. Bas, D. C.; Rogers, D. M.; Jensen, J. H., Very fast prediction and rationalization of pKa values for protein–ligand complexes. *Proteins: Structure, Function, and Bioinformatics* **2008,** *73* (3), 765-783.
40. Li, H.; Robertson, A. D.; Jensen, J. H., Very fast empirical prediction and rationalization of protein pKa values. *Proteins: Structure, Function, and Bioinformatics* **2005,** *61* (4), 704-721.
41. Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A., OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *Journal of Chemical Theory and Computation* **2016,** *12* (1), 281-296.
42. Schrodinger *LigPrep*, LLC: New York, NY, 2016.
43. Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M., Epik: a software program for pK(a) prediction and protonation state generation for drug-like molecules. *J Comput Aided Mol Des* **2007,** *21* (12), 681-91.
44. Schrodinger *Epik*, LLC: New York, NY, 2016.
45. Schrodinger *MacroModel*, LLC: New York, NY, 2016.
46. Halgren, T. A., Identifying and Characterizing Binding Sites and Assessing Druggability. *Journal of Chemical Information and Modeling* **2009,** *49* (2), 377-389.
47. Schrodinger *SiteMap*, LLC: New York, NY, 2016.
48. Halgren, T. A., Identifying and characterizing binding sites and assessing druggability. *J Chem Inf Model* **2009,** *49* (2), 377-89.
49. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S., Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry* **2004,** *47* (7), 1739-1749.
50. Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T., Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein–Ligand Complexes. *Journal of Medicinal Chemistry* **2006,** *49* (21), 6177-6196.

51.   Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L., Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *Journal of Medicinal Chemistry* **2004,** *47* (7), 1750-1759.

52.   Schrodinger *Glide*, LLC: New York, NY, 2016.

53.   Salam, N. K.; Nuti, R.; Sherman, W., Novel Method for Generating Structure-Based Pharmacophores Using Energetic Analysis. *Journal of Chemical Information and Modeling* **2009,** *49* (10), 2356-2368.

54.   Research, D. E. S. *Desmond Molecular Dynamics System*, LLC: New York, NY, 2016.

55.   Schrodinger *Maestro-Desmond Interoperability Tools*, LLC: New York, NY, 2016.

56.   Hoover, W. G., Canonical dynamics: Equilibrium phase-space distributions. *Phys Rev A Gen Phys* **1985,** *31* (3), 1695-1697.

57.   Martyna, G. J.; Tobias, D. J.; Klein, M. L., Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics* **1994,** *101* (5), 4177-4189.

58.   Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E., MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *Journal of Chemical Theory and Computation* **2012,** *8* (9), 3314-3321.

59.   Jianing, L.; Robert, A.; Kai, Z.; Yixiang, C.; Suwen, Z.; A., F. R., The VSGB 2.0 model: A next generation energy model for high resolution protein structure modeling. *Proteins: Structure, Function, and Bioinformatics* **2011,** *79* (10), 2794-2812.

60.   Bergqvist, S.; Alverdi, V.; Mengel, B.; Hoffmann, A.; Ghosh, G.; Komives, E. A., Kinetic enhancement of NF-kappaBxDNA dissociation by IkappaBalpha. *Proc Natl Acad Sci U S A* **2009,** *106* (46), 19328-33.

61.   Sue, S. C.; Cervantes, C.; Komives, E. A.; Dyson, H. J., Transfer of flexibility between ankyrin repeats in IkappaB* upon formation of the NF-kappaB complex. *J Mol Biol* **2008,** *380* (5), 917-31.

62.   Bergqvist, S.; Ghosh, G.; Komives, E. A., The IkappaBalpha/NF-kappaB complex has two hot spots, one at either end of the interface. *Protein Sci* **2008,** *17* (12), 2051-8.

63.   Sue, S. C.; Dyson, H. J., Interaction of the IkappaBalpha C-terminal PEST sequence with NF-kappaB: insights into the inhibition of NF-kappaB DNA binding by IkappaBalpha. *J Mol Biol* **2009,** *388* (4), 824-38.

64.   Yazdi, S.; Naumann, M.; Stein, M., Double phosphorylation-induced structural changes in the signal-receiving domain of IkappaBalpha in complex with NF-kappaB. *Proteins* **2017,** *85* (1), 17-29.

65.   Halgren, T., New method for fast and accurate binding-site identification and analysis. *Chem Biol Drug Des* **2007,** *69* (2), 146-8.
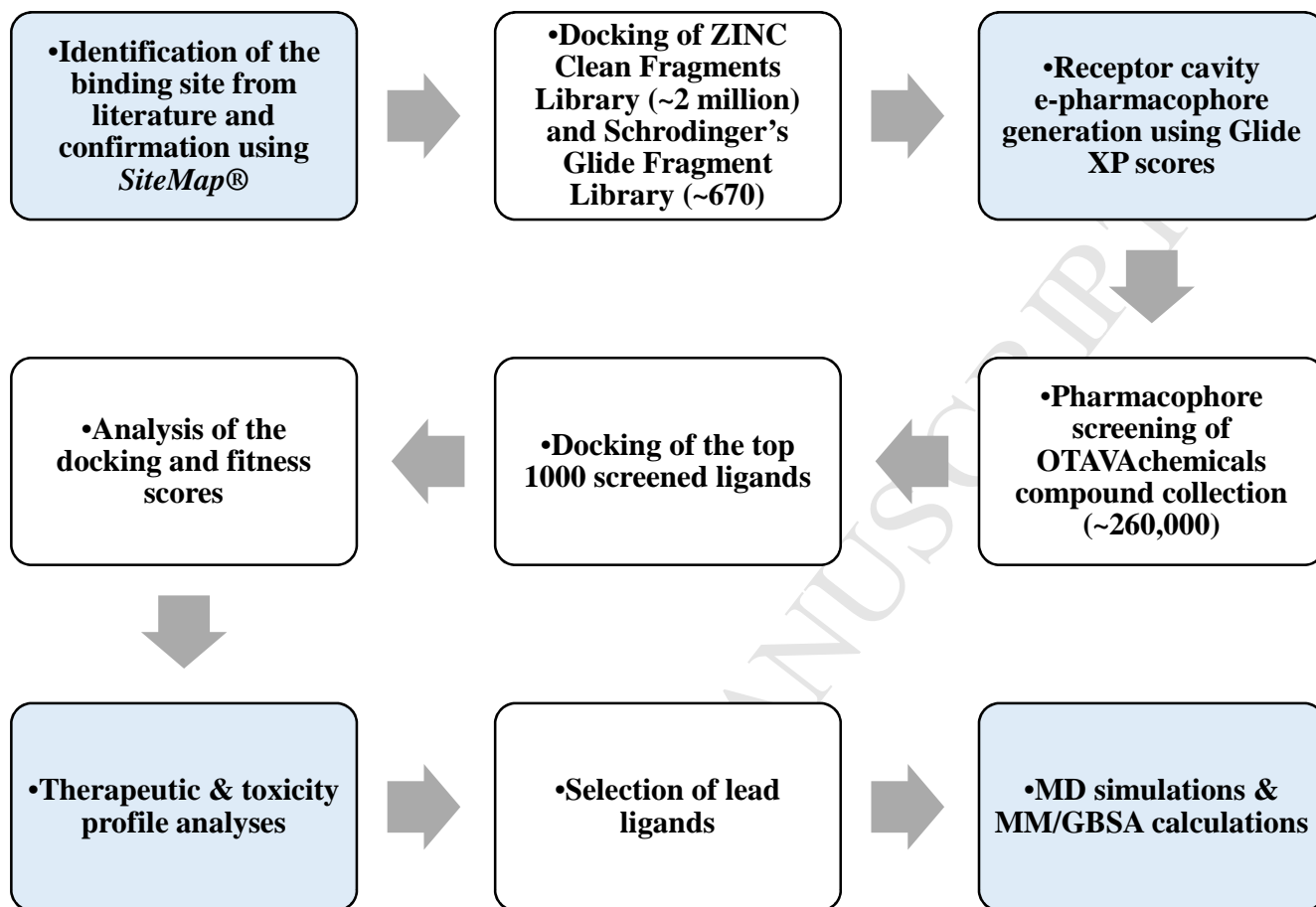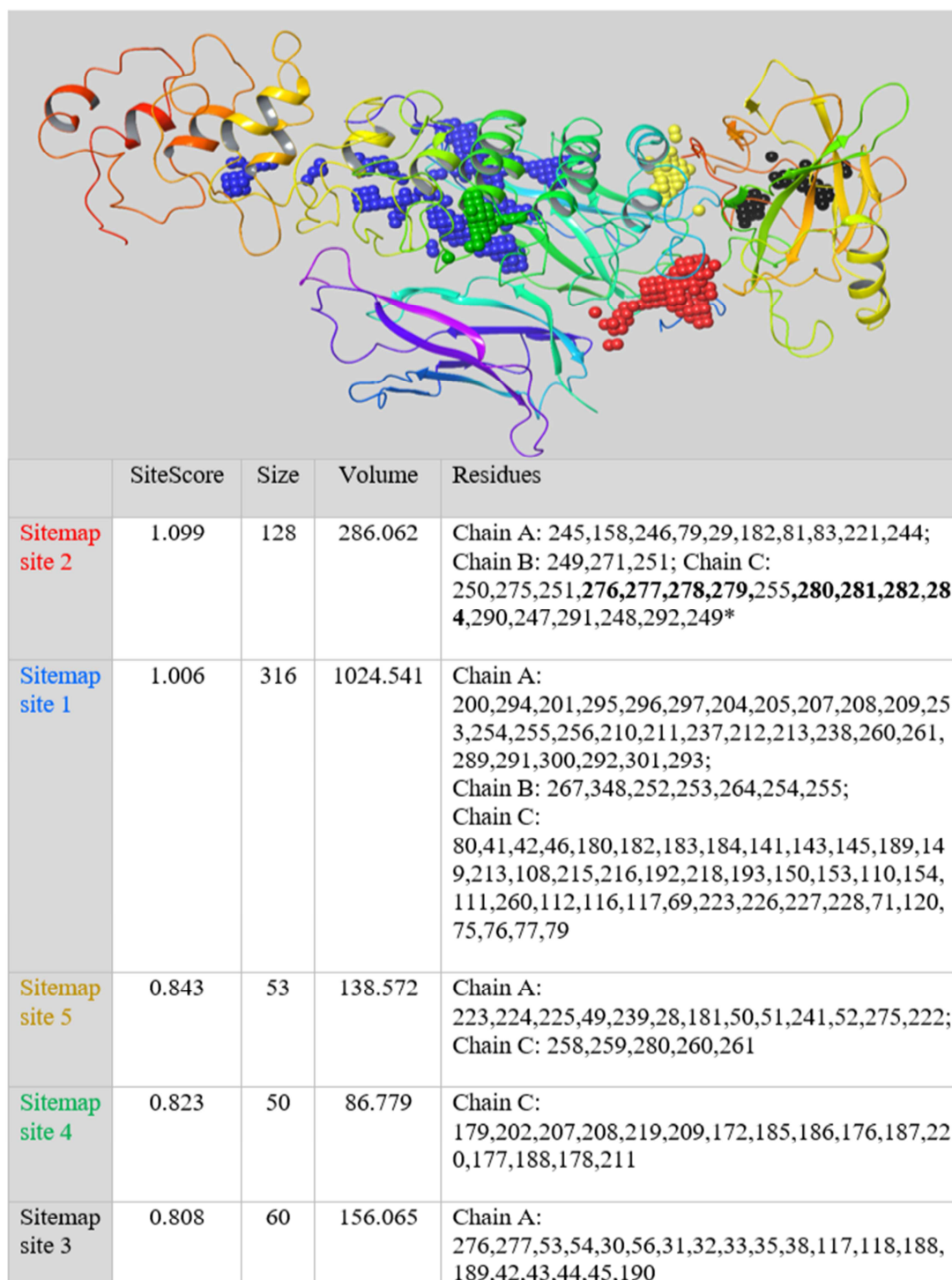
Scheme 1. Fragment-based e-pharmacophore virtual library screening flowchart followed by molecular modeling, ADMET and therapeutic analyses for the discovery of potent ligands of the NF κB/IkBα complex.

24
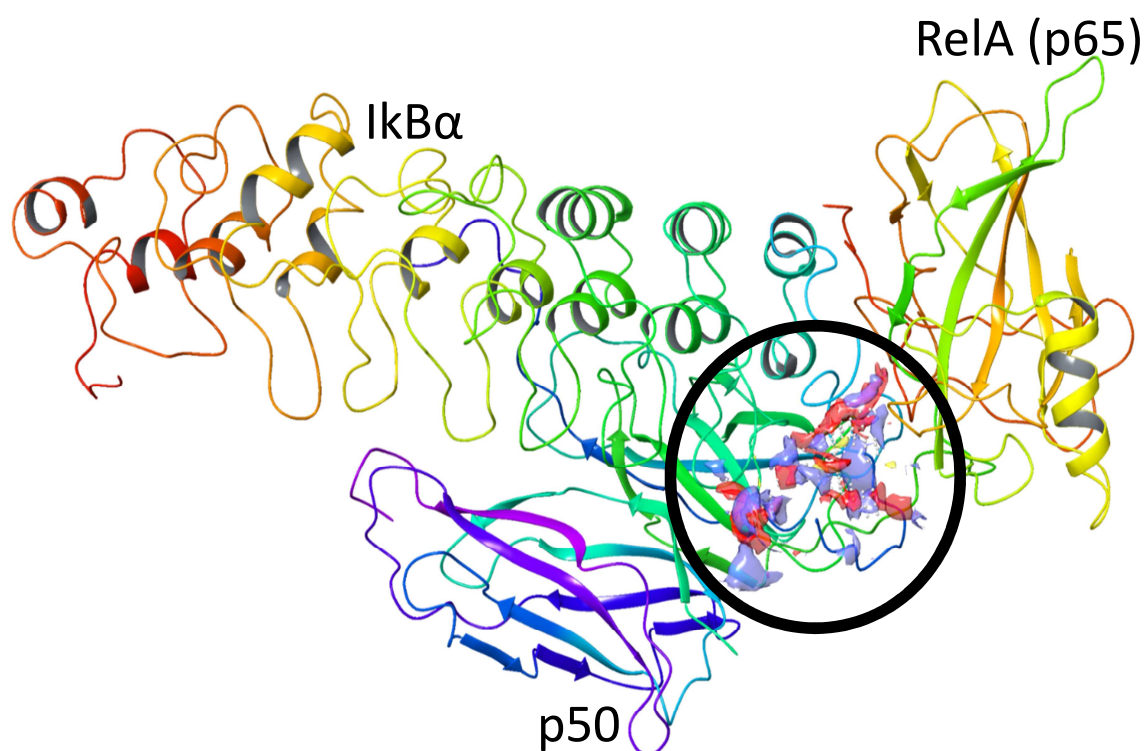
Figures

**Figure 1.** A ribbon representation of the NF κB/IkBα complex and the binding site (Site-2) which was used for molecular docking and e-pharmacophore model hypotheses generation. IkBα (barrel-like shaped protein on left) is in contact with the NF κB complex formed by the p50 (blue) and RelA/p65 (yellow) subunits. The binding site is composed majorly of ligand acceptor (red) and ligand donor (blue) maps as well as few hydrophobic maps (yellow). This binding site was computationally generated by SiteMap and supported by literature findings.

	SiteScore	Size	Volume	Residues
Sitemap site 2	1.099	128	286.062	Chain A: 245,158,246,79,29,182,81,83,221,244; Chain B: 249,271,251; Chain C: 250,275,251,**276,277,278,279**,255,**280,281,282,284**,290,247,291,248,292,249*
Sitemap site 1	1.006	316	1024.541	Chain A: 200,294,201,295,296,297,204,205,207,208,209,253,254,255,256,210,211,237,212,213,238,260,261,289,291,300,292,301,293; Chain B: 267,348,252,253,264,254,255; Chain C: 80,41,42,46,180,182,183,184,141,143,145,189,149,213,108,215,216,192,218,193,150,153,110,154,111,260,112,116,117,69,223,226,227,228,71,120,75,76,77,79
Sitemap site 5	0.843	53	138.572	Chain A: 223,224,225,49,239,28,181,50,51,241,52,275,222; Chain C: 258,259,280,260,261
Sitemap site 4	0.823	50	86.779	Chain C: 179,202,207,208,219,209,172,185,186,176,187,220,177,188,178,211
Sitemap site 3	0.808	60	156.065	Chain A: 276,277,53,54,30,56,31,32,33,35,38,117,118,188,189,42,43,44,45,190

Figure 2. The top five potential binding sites derived by SiteMap and their properties. The coloring code of the site maps in the table corresponds to the protein illustration above. Chain A refers to p65/ RelA; Chain B refers to p50; Chain C refers to IkBα. *The bolded residues in site 2 indicate residues in the PEST sequence.
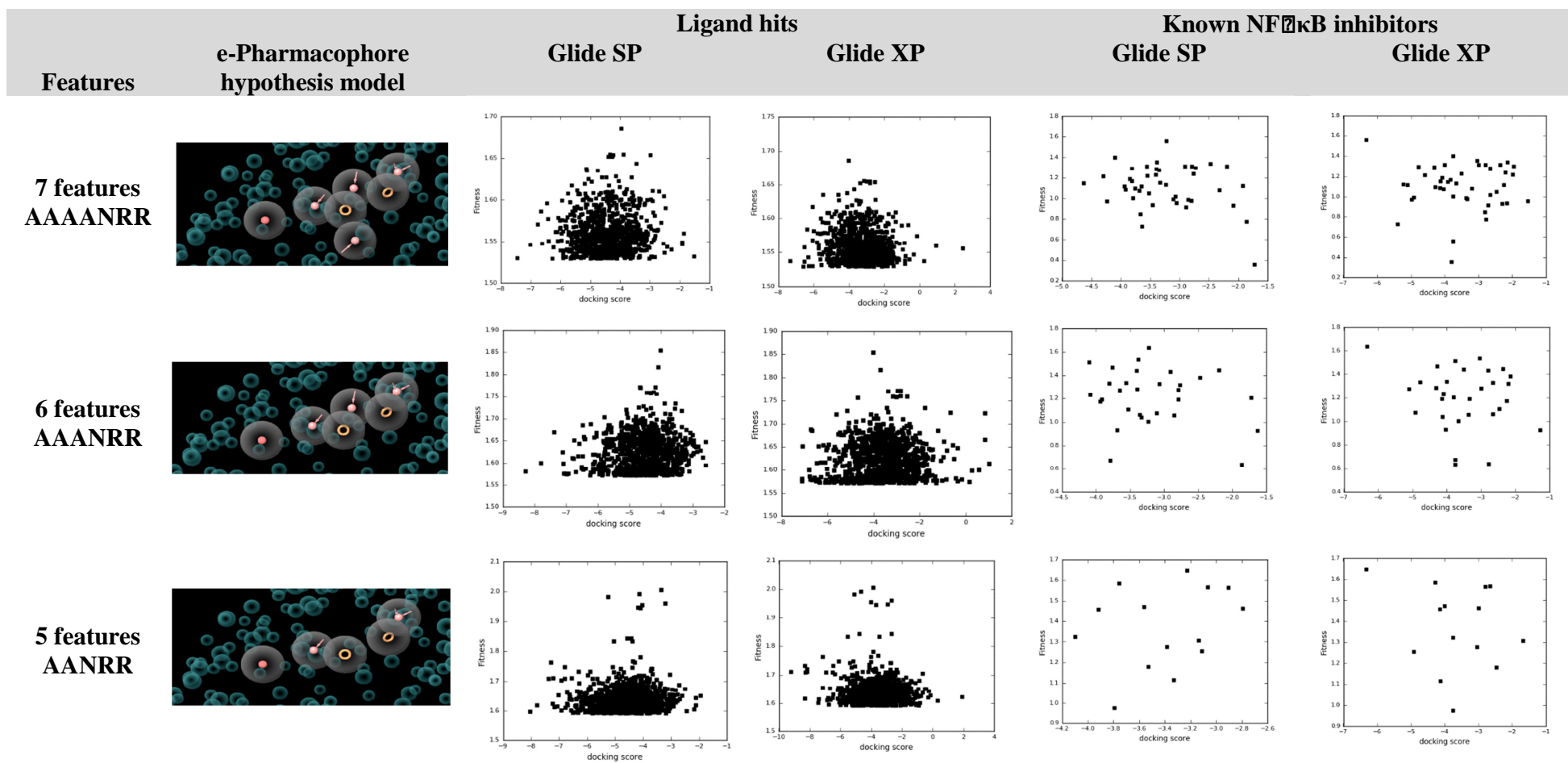
Figure 3. Fragment-based e-pharmacophore model hypotheses derived for the NF κB (p50/p65)/IkBα complex. Graphs show the fitness and Glide/ XP docking scores of molecules that have successfully met the minimum requirement of matching at least 4 featured ligand sites for the known NF κB inhibitors (right) and compounds from the OTAVAchemicals library collection (left). The hypothesis models were generated via the *vector-based* methodology. A/pink = H-bond acceptor; D/blue= H-bond donor; R/orange= aromatic ring; N/red= negative ionizable group. Blue circles in the background indicate receptor-based excluded volumes.
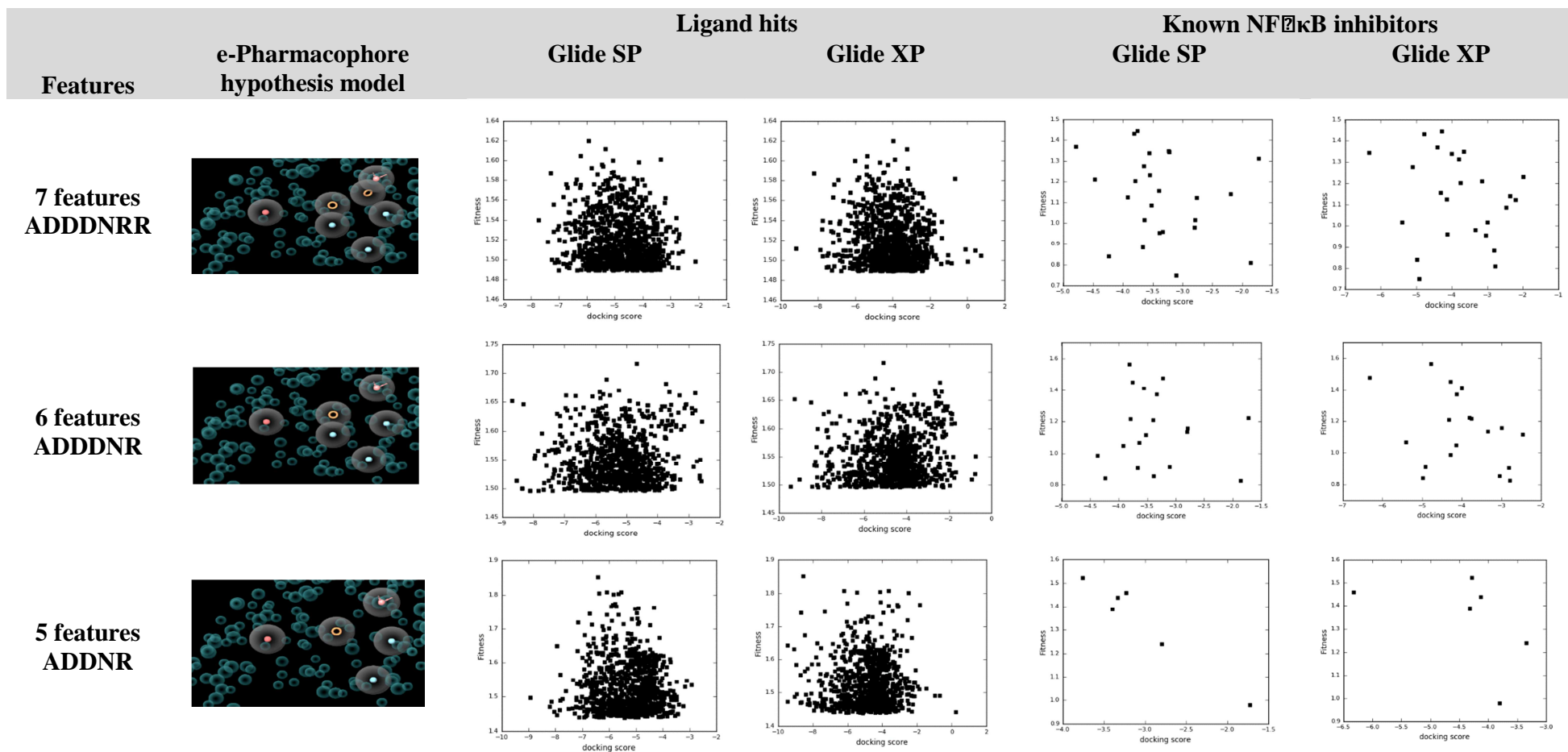
Figure 4. Fragment-based e-pharmacophore model hypotheses derived for the NF κB (p50/p65)/IkBα complex. Graphs show the fitness and Glide/ XP docking scores of molecules that have successfully met the minimum requirement of matching at least 4 featured ligand sites for the known NF κB inhibitors (right) and compounds from the OTAVAchemicals library collection (left). The hypothesis models were generated via the *projected points donor* methodology. A/pink = H-bond acceptor; D/blue= H-bond donor; R/orange= aromatic ring; N/red= negative ionizable group. Blue circles in the background indicate receptor-based excluded volume.

	Ligand hits			Control
	6248112	7132624	1426436	DHMEQ
Allergy	0.5	0.53		
Alzheimer	0.53	0.57	0.51	
Angina				
Arthritis	0.58	0.56	0.67	0.58
Asthma	0.72	0.84	0.77	0.64
Bacterial				
Cancer	0.55	0.59		
Depression	0.79	0.74	0.83	
Diabetes		0.67		0.53
HIV	0.67	0.54	0.65	0.6
Heart failure	0.88	0.66	0.57	0.54
Hyperlipidemia			0.56	0.57
Hypertension	0.63	0.76	0.53	
Inflammation				
Migraine	0.72	0.71	0.58	
Mycosis				0.56
Obesity	0.65		0.96	0.9
Osteoporosis				
Pain			0.57	
Parkinson	0.52	0.62		
Psoriasis		0.53	0.58	
Schizophrenia			0.5	
Skin diseases			0.64	0.69
Thrombosis			0.59	
Viral	0.84	0.77	0.83	

Figure 5. The therapeutic activity and docking scores of ligands selected from the virtual screening compared to a known inhibitor from literature. The ligands were submitted to MetaDrug/ MetaCore™. In general, values greater than 0.50 indicate that the molecules are active as per the scores from the QSAR models. Only numerical values indicating activity scores over 0.50 were presented in the table for clarity and simplicity. The coloring scheme increases in intensity with greater therapeutic values.
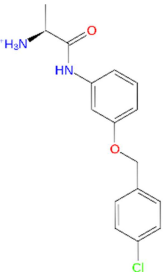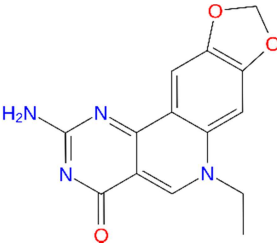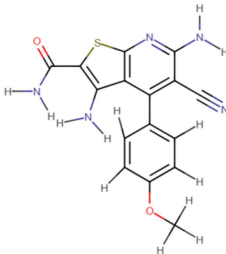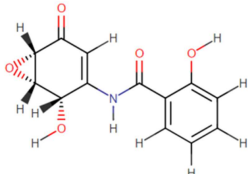
	6248112	Ligand hits 7132624	1426436	Control DHMEQ
AMES	0.57	0.61		
Anemia				
carcinogenicity				
carcinogenicity mouse female				
carcinogenicity mouse male			0.56	0.51
carcinogenicity rat female				
carcinogenicity rat male			0.53	0.56
cardiotoxicity	0.51			
Cytotoxicity model, -log GI50 (M)				
Epididymis toxicity				
Genotoxicity	0.55			
Hepatotoxicity				
Kidney Necrosis		0.53		0.78
Kidney Weight Gain				0.73
Liver Cholestasis			0.52	0.56
Liver Lipid Accumulation				
Liver Necrosis				0.78
Liver Weight Gain				
MRTD*				0.66
Nasal pathology				
Nephron Injury				0.63
Nephrotoxicity				
Neurotoxicity				0.51
Pulmonary toxicity				
SkinSens, EC3				
Testicular toxicity				
Reactive**	OK	OK	R	R
Ruleof5***	OK	OK	OK	OK

* Maximum Recommended Therapeutic Dose, log mg/kg-bm/day, range is from -5 to 3. Cutoff is 0.5. Chemicals with high log MRTDs can be classified as mildly toxic compounds, chemicals with low log MRTDs as highly toxic compounds.
** Metabolites contain reactive groups
** Ruleof5 (likely to be orally bioavailable)

Figure 6. The predicted toxicity effects of ligands selected from the virtual screening compared to a known NF☐κB inhibitor from literature. The ligands were submitted to MetaDrug/ MetaCore™. In general, values greater than 0.50 indicate that the molecules are toxic as per the scores from the QSAR models. Only numerical values indicating toxicity scores over 0.50 were presented in the table for clarity and simplicity. Gray boxes indicate no toxicity or toxicity lower than 0.50 which is nonsignificant. The coloring scheme increases in intensity with greater toxicity values.

30

| | Ligand hits | | | Reference |
	1426436	6248112	7132624	DHMEQ
Chemical formula	$C_{16}H_{18}ClN_2O_2$	$C_{14}H_{12}N_4O_3$	$C_{16}H_{13}N_5O_2S$	$C_{13}H_{11}NO_5$
BBB, log ratio	-0.24 (44.72)	-0.33 (36.33)	-0.22 (35.87)	-0.61 (40.11)
Prot-bind, log t	-0.25 (44.72)	-0.17 (36.33)	-0.08 (35.87)	-0.33 (40.11)
Prot-bind, %	38.02 (50.70)	43.08 (36.33)	40.59 (35.87)	56.20 (46.13)
G-logP	2.26	0.75	1.78	2.77
WSol, log mg/L	1.34	1.92	3.28	2.06

Figure 7. The 2D chemical structure and chemical properties of three potential NF☐κB Ikbα inhibitors. The ligands were selected after the e-pharmacophore virtual screening and predicted ADMET/ therapeutic analyses. Two-digit numbers in brackets indicate the property of Tanimoto Prioritization (TP), a score for similarity between the analyzed compounds and compound sets in the quantitative structure-activity relationships (QSAR) models.
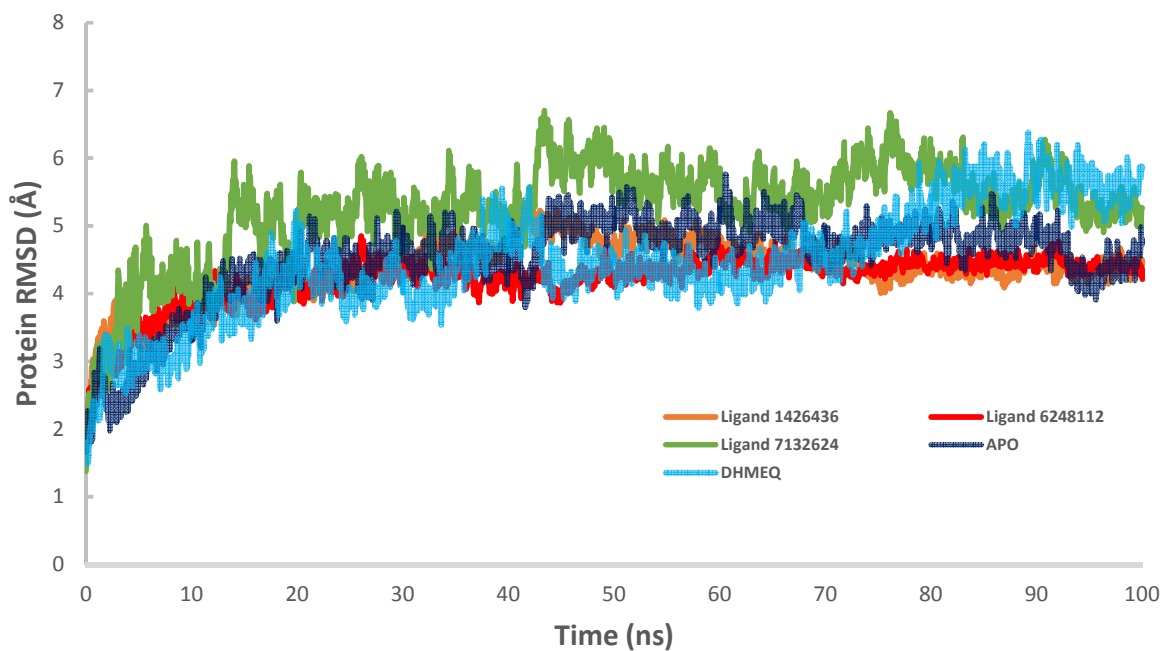
Figure 8. RMSD evolution over time of the C-alpha atoms over a 100 ns MD simulation with the NF☐κB-Ikbα (p50/p65) complex. The three ligands from the Otava library were selected after e-pharmacophore virtual screening and post-docking MetaCore analysis. DHMEQ is used as a control NF☐κB inhibitor. Average values were plotted for the potent ligands *1426436* and *6248112* from triplicate MD simulations.
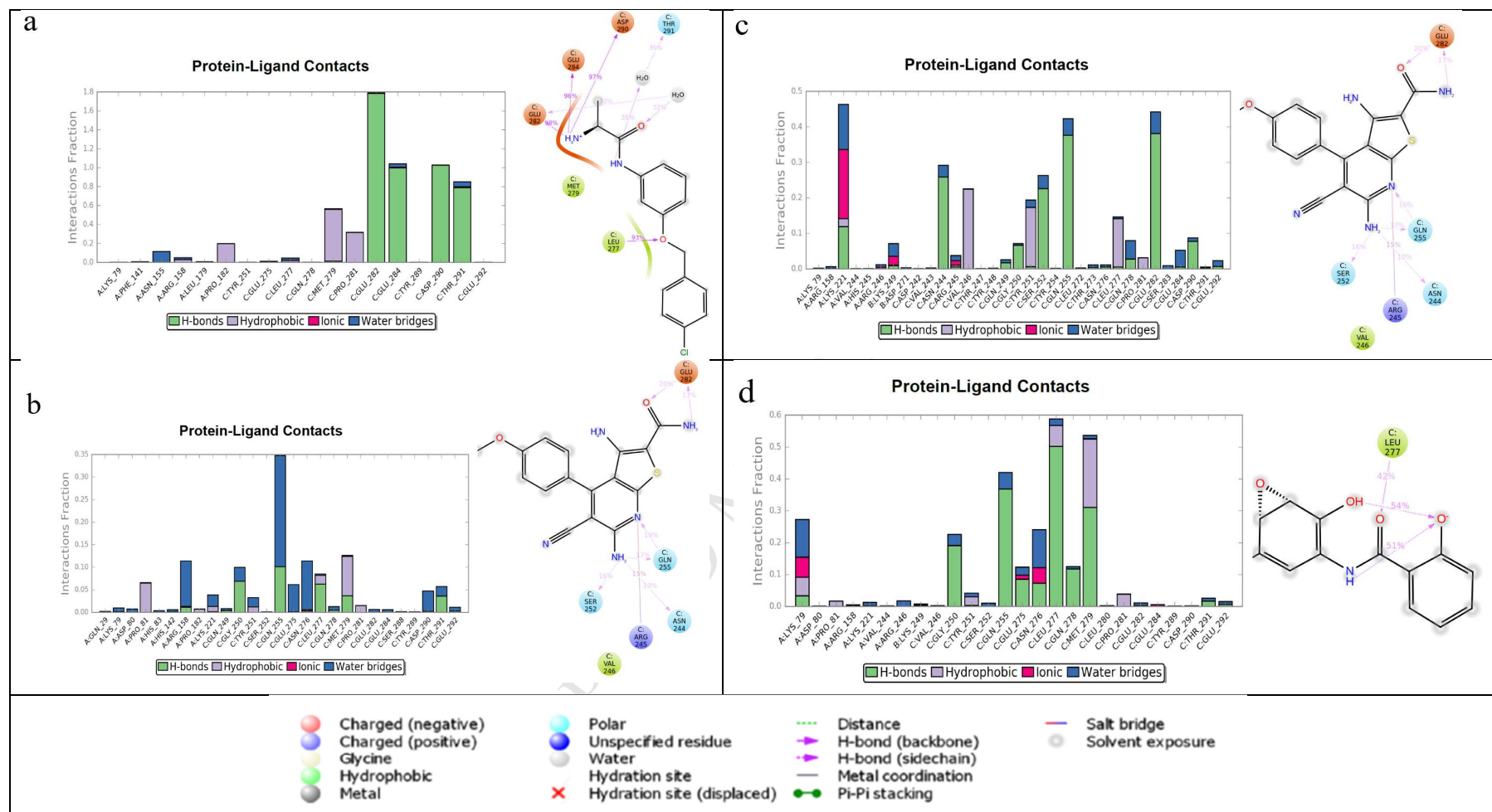
Figure 9. The protein-ligand interaction profiles of the discovered three ligands and the control molecule in the binding pocket of the NF κB/IkBα complex throughout the 100 ns MD simulations. The three ligands from the Otava library were selected after e-pharmacophore virtual screening and post-docking MetaCore/ MetaDrug analysis. The interaction fraction indicates the percentage of time the contact is made throughout the simulation runtime. The profiles are displayed for a) ligand *1426436* b) ligand *6248112* c) ligand *7132624* and d) *DHMEQ*, the control molecule.

33
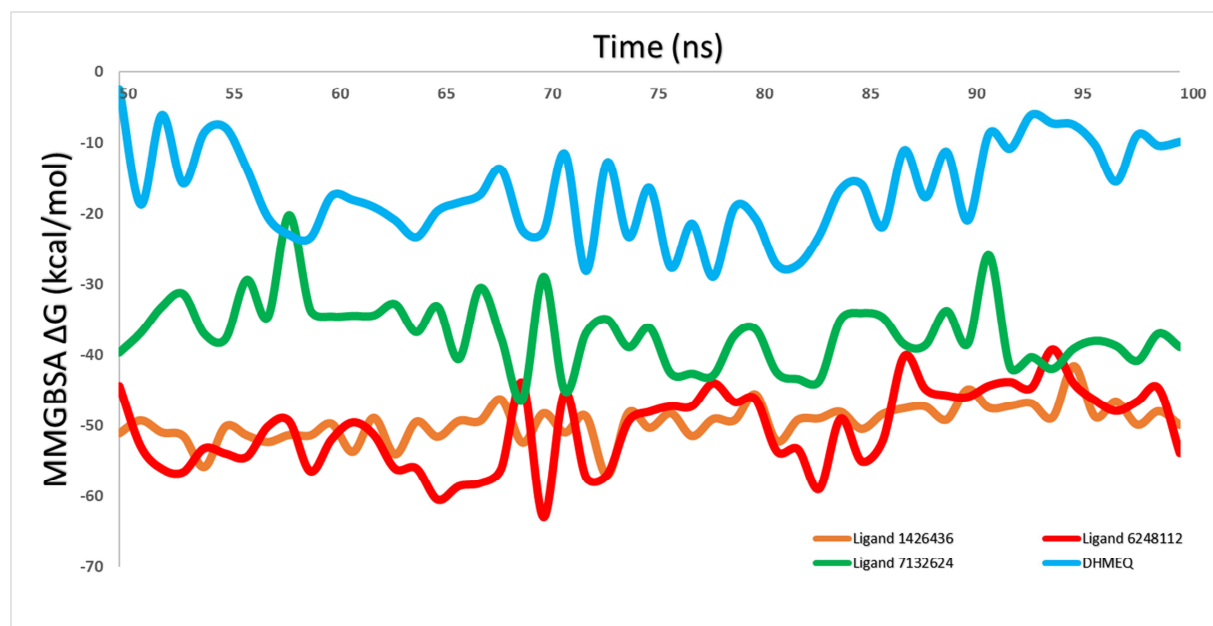
Figure 10. MM/GBSA free energy analysis for the hit ligands and known control at the binding pocket of NFκB/ Ikbα throughout the last half of the MD simulations. Average values were plotted for the potent ligands *1426436* and *6248112* from triplicate MD simulations. Average values in kcal/mol were: ligand 6248112: −50.54± 4.30; ligand *1426436*: −49.54± 2.63; ligand *7132624*: −36.92 ± 4.89; and control ligand: −16.70± 6.60.

Highlights

- e-pharmacophore models were developed for the NF κB/IκBα complex based on 85000 fragments.

- The Otava chemicals library was screened against the constructed e-pharmacophore models.

- Three ligands were identified with promising high therapeutic values.